IT a anatomie firmy

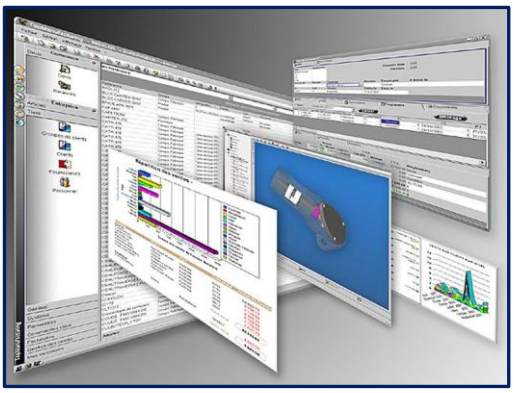
Prediktivní analytika

(pracovní dokument)



MBI tým

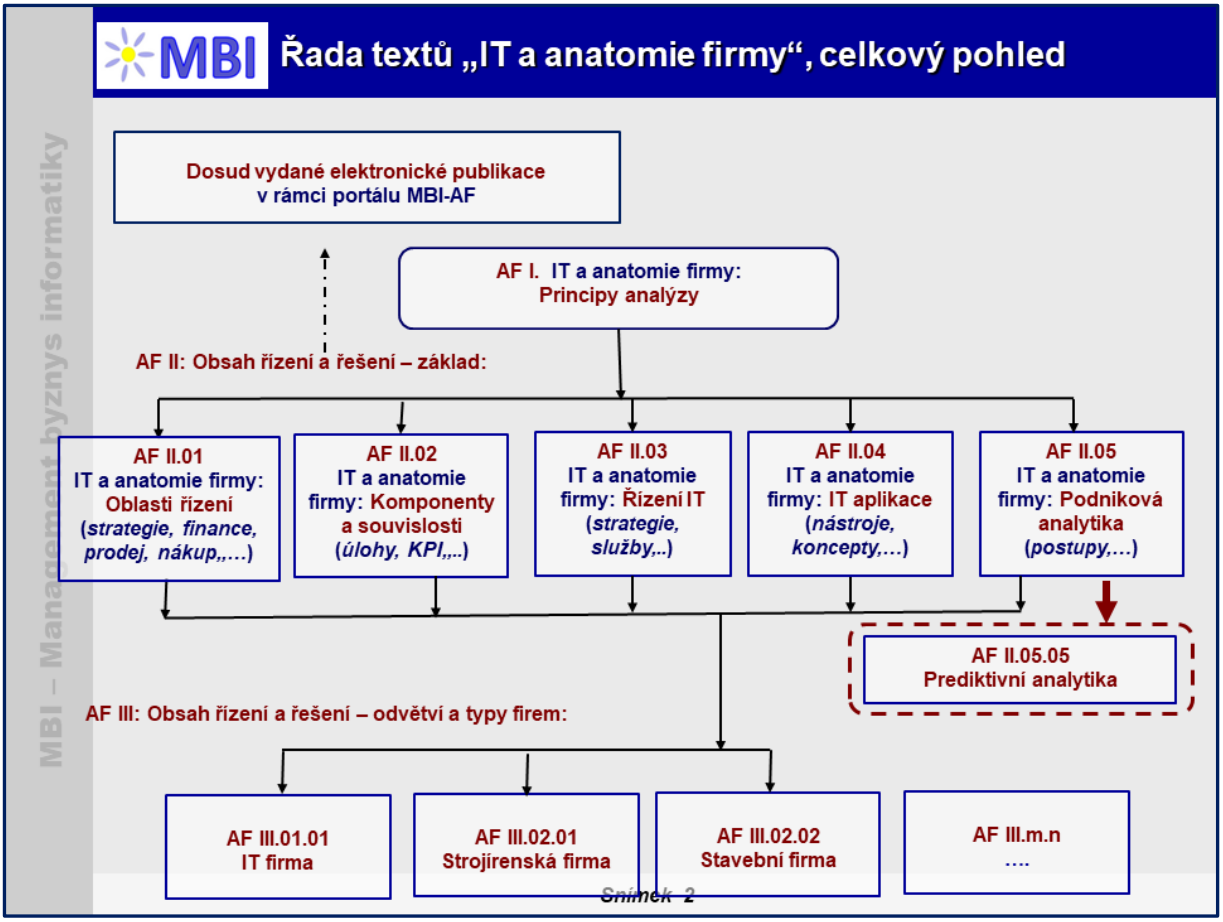
VŠE Praha, 2023



[A)] Prediktivní analytika: obsah, principy, komponenty ***(Úlohy v řešení prediktivní analytiky, potřeba prediktivní analytiky, scénáře, analytické otázky, role v prediktivní analytice)***
[B)] Postupy a organizace řešení prediktivní analytiky ***(Jednotlivé fáze řešení podle metodiky CRISP DM)***
[C)] Uplatnění prediktivní analytiky podle oblastí řízení firmy ***(Strategie firmy, finanční řízení, obchod,…)***
[D)] Prediktivní analytika v kontextu podnikové analytiky, funkce a metody ***(Vazby prediktivní analytiky na ostatní disciplíny, základní funkce prediktivní analytik, hlavní využívané metody, např. regresní analýzy, rozhodovací stromy apod.)***
[D)] Machine Learning ***(Principy machine learning (ML), typy ML podle typů učení, řešení ML, faktory ovlivňující ML, využití ML, prediktivní analytika a ML.)***

Prediktivní analytika představuje jednu z klíčových disciplín pro podporu byznysu na základě analytických metod a nástrojů IT. Jejím hlavním smyslem je to, že ***využívá dostupná data k předpovědi budoucích jevů***. ***Účelem*** tohoto ***dokumentu*** je specifikovat obsah a přístupy k prediktivní analytice a analyzovat její možnosti a omezení ve vztahu k reálnému byznysu.

Dokument představuje doplnění k základnímu dokumentu orientovanému na podnikovou analytiku: „*AF_II_05_Podnikova_Analytika.pdf*“. Jeho místo ve struktuře portálu MBI-AF ukazuje další obrázek:


MBI
Řada textů „IT a anatomie firmy“, celkový pohled
MBI – Management byznys informatiky
Dosud vydané elektronické publikace v rámci portálu MBI-AF
AF I. IT a anatomie firmy: Principy analýzy
AF II: Obsah řízení a řešení – základ:
AF II.01 IT a anatomie firmy: Oblasti řízení (strategie, finance, prodej, nákup,,...)
AF II.02 IT a anatomie firmy: Komponenty a souvislosti (úlohy, KPI,,..)
AF II.03 IT a anatomie firmy: Řízení IT (strategie, služby,..)
AF II.04 IT a anatomie firmy: IT aplikace (nástroje, koncepty,...)
AF II.05 IT a anatomie firmy: Podniková analytika (postupy,...)
AF II.05.05 Prediktivní analytika
AF III: Obsah řízení a řešení – odvětví a typy firem:
AF III.01.01 IT firma
AF III.02.01 Strojírenská firma
AF III.02.02 Stavební firma
AF III.m.n ....
Snímek 2

Obsah

1. Úvod ... 11
1.1 Podniková analytika ... 12
1.2 Oblasti řízení firmy ... 13
1.3 Řízení IT ... 14
1.4 Komponenty řízení firmy ... 15
1.5 IT v řízení firmy ... 16
A) Prediktivní analytika: obsah, principy ... 18
2. Analytické úlohy ... 19
2.1 Deskriptivní analytika ... 19
2.2 Diagnostická analytika ... 19
2.3 Prediktivní analytika ... 20
2.4 Preskriptivní analytika ... 20
2.5 Kognitivní analytika ... 20
2.6 Závěry ... 21
3. Potřeba a obsah prediktivní analytiky ... 23
3.1 Potřeba prediktivní analytiky: prognózování a plánování ... 23
3.1.1 Prognózování v byznysu ... 23
3.1.2 Podstata a principy plánování ... 24
3.1.3 Podstata plánovacích úloh ... 25
3.1.4 Principy plánovacích úloh ... 25
3.1.5 Plánovací úlohy v kontextu řízení firmy ... 25
3.2 Rekapitulace vybraných ukazatelů ... 26
3.3 Podstatné principy prediktivní analytiky ... 26
3.4 Řešení prediktivní analytiky ... 27
3.5 Efekty uplatnění prediktivní analytiky ... 28
3.6 Omezení, problémy, předpoklady řešení prediktivní analytiky ... 28
3.7 Příklady využití prediktivní analytiky ... 29
3.8 Závěry ... 29
4. Role v prediktivní analytice ... 31
4.1 Doménový expert ... 31
4.2 Data a databázový expert ... 31
4.3 Expert prediktivního modelování ... 32
4.4 Závěry ... 32
5. Faktory prediktivní analytiky ... 33
5.1 Byznys prostředí ... 33
5.1.1 Velikost firmy ... 33

5.1.2 Původ a vlastnictví firmy ... 34
5.1.3 Konkurenční prostředí ... 34
5.1.4 Odvětví působnosti firmy ... 34
5.2 Řízení a organizace firmy ... 34
5.2.1 Firemní kultura ... 34
5.2.2 Organizace firmy ... 35
5.2.3 Business model ... 35
5.2.4 Vztah podniku ke svým obchodním a dalším partnerům ... 35
5.2.5 Personální zdroje, úroveň jejich znalostí ... 35
5.3 Řízení prediktivní analytiky ... 35
5.3.1 Strategie řešení prediktivní analytiky ... 36
5.3.2 Identifikace potenciálních efektů prediktivní analytiky ... 36
5.3.3 Existence silného sponzora ... 36
5.4 Kvalita informačního systému ... 36
5.4.1 Kvalita datových zdrojů ... 36
5.4.2 Kvalita podnikových aplikací ... 37
5.5 Závěry ... 37
6. Scénáře, analytické otázky ... 39
6.1 Scénář 1: Jak řešit plánovací úlohy vzhledem k hlavním potřebám byznysu ... 39
6.2 Scénář 2: Jak řešit prediktivní analytiku vzhledem k obsahu byznysu ... 39
6.3 Scénář 3: Jak řešit a zajistit datové zdroje pro prediktivní analytiku ... 40
6.4 Scénář 4: Jak zajistit personální zdroje pro prediktivní analytiku ... 40
6.5 Scénář 5: Jaké metody volit pro prediktivní analytiku ... 40
6.6 Závěry ... 40
B) Postupy řešení a organizace prediktivní analytiky ... 42
7. Pochopení byznysu („Business Understanding“) ... 43
7.1 Definování vstupních požadavků na prediktivní modely ... 43
7.2 Byznys cíle ... 43
7.3 Návrh dat pro prediktivní modelování ... 44
7.4 Definování jednotek analýzy ... 44
7.5 Definování cílových proměnných ... 44
7.6 Závěry ... 44
8. Porozumění datům („Data Understanding“) ... 46
8.1 Definování dat pro prediktivní modelování ... 46
8.2 Analýza dostupných zdrojů dat ... 46
8.3 Analýza datové kvality ... 47
8.4 Závěry, doporučení ... 47
9. Příprava dat („Data Preparation“) ... 48
9.1 Organizace a shromáždění dat ... 48
9.2 Čištění dat ... 48
9.3 Vytvoření odvozených proměnných ... 49
9.4 Závěry, doporučení ... 49

10. Modelování („Modeling“) *50*
10.1.1 Typy prediktivních modelů 50
10.1.2 Hledání vztahů – data mining, cluster analýza 50
10.1.3 Tvorba prediktivního modelu 51
10.1.4 Soubor modelů (ensemble) 51
10.2 Závěry, doporučení **51**
11. Vyhodnocení výsledků („Assessing Models“) *53*
11.1 Závěry **53**
12. Využití výsledků („Model Deployment“) *54*
12.1 Prezentace výsledků **54**
12.2 Konsolidace informací **54**
12.3 Závěry **55**
C) Uplatnění prediktivní analytiky podle oblastí řízení *56*
13. Prediktivní analytika strategického řízení firmy *57*
13.1 Vymezení obsahu oblasti řízení, datové zdroje **57**
13.2 Účel prediktivní analytiky ve strategickém řízeni **58**
13.3 Cílové proměnné **58**
13.4 Analytické jednotky **58**
14. Prediktivní analytika finančního řízení *59*
14.1 Vymezení obsahu oblasti řízení **59**
14.2 Datové zdroje **60**
14.3 Účel prediktivní analytiky ve finančním řízeni **61**
14.4 Cílové proměnné **61**
14.5 Analytické jednotky **61**
14.6 Prediktivní funkcionalita **61**
15. Prediktivní analytika v řízení prodeje *63*
15.1 Vymezení obsahu oblasti řízení **63**
15.2 Datové zdroje **64**
15.3 Účel prediktivní analytiky v řízeni prodeje **65**
15.4 Cílové proměnné **65**
15.5 Analytické jednotky **65**
15.6 Prediktivní funkcionalita **65**
16. Prediktivní analytika v řízení nákupu *67*
16.1 Vymezení obsahu oblasti řízení, ... **67**
16.2 Datové zdroje **68**
16.3 Účel prediktivní analytiky v řízeni nákupu **69**
16.4 Cílové proměnné **69**
16.5 Jednotky analýzy **69**

16.6 Prediktivní funkcionalita ... 69
17. Prediktivní analytika v řízení personálních zdrojů ... 71
17.1 Vymezení obsahu oblastí řízení ... 71
17.2 Datové zdroje ... 72
17.3 Účel prediktivní analytiky v řízeni lidských zdrojů ... 73
17.4 Cílové proměnné ... 73
17.5 Jednotky analýzy ... 74
17.6 Prediktivní metody a funkcionalita ... 74
18. Prediktivní analytika v řízení majetku ... 75
18.1 Vymezení obsahu oblasti řízení, datové zdroje ... 75
18.2 Datové zdroje ... 76
18.3 Účel prediktivní analytiky v řízeni majetku ... 76
18.4 Cílové proměnné ... 77
18.5 Jednotky analýzy ... 77
18.6 Prediktivní funkcionalita ... 77
18.7 Příklad: Prediktivní údržba ... 77
18.7.1 Úvod do prediktivní údržby ... 78
18.7.2 Křivka vzniku potencionální poruchy (P-F) ... 78
18.7.3 Prediktivní údržba pomáhá optimalizovat plánovanou dobu odstávky ... 79
18.7.4 Prediktivní údržba může pomoci optimalizovat produktivitu zaměstnanců. ... 80
18.7.5 Zvýšení příjmů ... 80
18.7.6 Principy preventivní údržby ... 80
18.7.7 Techniky prediktivní údržby ... 81
18.7.8 Omezení prediktivní analytiky ... 82
18.7.9 Zdroje ... 82
19. Prediktivní analytika v řízení marketingu ... 84
19.1 Vymezení obsahu oblasti řízení ... 84
19.2 Datové zdroje ... 85
19.3 Účel prediktivní analytiky v řízeni marketingu ... 85
19.4 Cílové proměnné ... 86
19.5 Jednotky analýzy ... 86
19.6 Prediktivní funkcionalita ... 86
20. Prediktivní analytika v řízení dopravy ... 88
20.1 Vymezení obsahu oblasti řízení ... 88
20.2 Datové zdroje ... 89
20.3 Účel prediktivní analytiky v řízeni dopravy ... 89
20.4 Cílové proměnné ... 89
20.5 Jednotky analýzy ... 90
20.6 Prediktivní funkcionalita ... 90
21. Prediktivní analytika v řízení energií ... 91
21.1 Vymezení obsahu oblasti řízení ... 91

21.2 Datové zdroje ... 91
21.3 Účel prediktivní analytiky ve strategickém řízeni ... 92
21.4 Cílové proměnné ... 92
21.5 Jednotky analýzy ... 93
21.6 Prediktivní funkcionalita ... 93
D) Kontext, funkce, metody prediktivní analytiky ... 94
22. Prediktivní analytika v kontextu podnikové analytiky ... 95
22.1 Machine Learning, ML ... 95
22.2 Business Intelligence, BI ... 95
22.2.1 Principy business intelligence ... 96
22.2.2 Prediktivní analytika a business Intelligence ... 96
22.3 Competitive intelligence, CI ... 97
22.3.1 Principy competitive intelligence ... 97
22.3.2 Prediktivní analytika a competitive Intelligence ... 97
22.4 Data Science ... 97
22.4.1 Principy Data Science ... 97
22.4.2 Prediktivní analytika a Data Science ... 98
22.5 Data mining, DMI ... 98
22.5.1 Principy data miningu ... 98
22.5.2 Prediktivní analytika a data mining ... 99
22.6 Text mining ... 100
22.6.1 Principy text miningu ... 100
22.6.2 Využití text miningu ... 100
22.6.3 Prediktivní analytika a text mining ... 100
22.7 Umělá inteligence (Artificial Intelligence) ... 101
22.7.1 Principy umělé inteligence ... 101
22.7.2 Prediktivní analytika a umělá inteligence ... 101
22.8 Big Data Analytics ... 101
22.8.1 Principy *Big Data Analytics* ... 101
22.8.2 Prediktivní analytika a Big Data Analytics ... 102
22.9 Závěry ... 102
23. Základní funkce uplatňované v prediktivní analytice ... 103
23.1 Clustering ... 103
23.2 Klasifikace a hodnocení pravděpodobnosti tříd (Classification and class probability estimation): ... 103
23.3 Rozhodovací pravidla ... 103
23.4 Asociační pravidla ... 104
23.5 Přiřazování podobností (Similarity matching): ... 105
23.6 Seskupování podle výskytů (Co-occurrence grouping): ... 105
23.7 Profilování (Profiling): ... 105
23.8 Predikce vazeb (Link prediction): ... 105
23.9 Redukce dat (Data reduction): ... 105
23.10 Kauzální modelování (Causal modeling): ... 105

23.11 Závěry, doporučení ... 105

24. Metody uplatňované v prediktivní analytice ... 107

24.1 Regresní analýzy ... 107
24.1.1 Základní principy regresní analýzy ... 107
24.1.2 Výhody regresní analýzy ... 110
24.1.3 Omezení regresní analýzy: ... 111

24.2 Rozhodovací stromy ... 111
24.2.1 Principy rozhodovacích stromů ... 111
24.2.2 Řešení rozhodovacích stromů ... 112
24.2.3 Využití v prediktivní analytice ... 113
24.2.4 Efekty rozhodovacích stromů ... 113
24.2.5 Omezení rozhodovacích stromů ... 114

24.3 Neuronové sítě ... 114
24.3.1 Výhody neuronových sítí ... 117
24.3.2 Omezení neuronových sítí ... 117

24.4 Bayesovská klasifikace ... 117
24.4.1 Principy Bayesovské klasifikace ... 117
24.4.2 Efekty Bayesovské klasifikace ... 118
24.4.3 Omezení Bayesovské klasifikace ... 118

24.5 Závěry, doporučení ... 118

E) Machine Learning ... 119

25. Principy machine learning ... 121

26. Typy ML systémů podle typů učení ... 122

26.1 Klasifikace dle řízení učení ... 122
26.1.1 Řízené učení („supervised learning“) ... 122
26.1.2 Neřízené učení („unsupervised learning“) ... 122
26.1.3 Semi-supervised learning ... 122
26.1.4 Self-supervised learning ... 122
26.1.5 Reinforcement learning ... 122

26.2 Dávkové a online učení ... 122
26.2.1 Dávkové učení („batch learning“) ... 123
26.2.2 Učení online („online learning“) ... 123

26.3 Učení založené na instancích nebo modelech („instance-based / model-based learning“) ... 123
26.3.1 Instance-based learning ... 123
26.3.2 Model-based learning ... 123

27. Řešení ML ... 124

28. Faktory machine learning ... 125

28.1 Nedostatečné množství trénovacích dat ... 125

28.2 Nerepresentativní trénovací data ... 125

28.3 Špatná kvalita dat ... 125

28.4 Validace a ladění modelů ... 125

28.5 Vyhodnocení modelu ... 126

29. Využití machine learning ... 127

30. Prediktivní analytika a machine learning ... 128

31. Závěry ... 129

Terminologie ***130***
Zdroje ***136***

1. Úvod

Otázky prediktivní analytiky představují ***velmi široký rozsah informací, přístupů, analytických úvah a nástrojů***, které mají jak obecnou, společnou povahu, tak zcela konkrétní vazbu na odvětví, prostředí a potřeby jednotlivých typů firem. Dokument „Prediktivní analytika“ vychází dále ***z celé řady zdrojů***, které jsou souhrnně uvedeny v závěru dokumentu. Za ***hlavní*** pro účely tohoto dokumentu považujeme tyto:

- ABBOTT, D.: ***Applied Predictive Analytics. Principlec and Techniques for the Professional Data Analyst***. John Wiley & Sons, Indianoplolis, 2014. ISBN: 978-1-118-72796-6. K této publikaci s ohledem na její význam doplníme ***několik podstatných poznámek autora***:
  - *Kniha popisuje prediktivní analytiku primárně pohledem praktika než teoretika,*
  - *Dobrý praktik nemusí rozumět matematickým principům algoritmů, aby byl schopen je úspěšně aplikovat.*
  - *Na druhé straně dobrý praktik má rozumět dopadům změn parametrů pro modely, vlivům předpokladů na algoritmy predikcí a omezení algoritmů – zejména v prostředí velmi náročných projektů.*
  - *V praxi je často mnoho způsobů, jak řešit úspěšně problémy a výběr vždy závisí na mnoha faktorech, které je ovlivňují a je proto třeba je jasně identifikovat a vyhodnocovat.*
  - *Úspěšné řešení prediktivních modelů závisí více na dobrém pochopení dat (a jejich obsahu) než na využití sofistikovaných algoritmů.*
- FITZ-ENZ J., MATTOX II J.,R,: *Predictive Analytics for Human Resources,* Wiley and SAS Business Series, 2014.ISBN: 978-1-118-89367-8.
- KUHN, M., JOHNSON, K.: ***Applied Predictive Modeling***, Groton, Saline, USA, Springer, 2016. ISBN 978-1-4614-6848-6.
- SIEGEL, E: ***Predictive Analytics***. New York, John Wiley & Sons, 2016. ISBN 978-1-119-14567-7.
- WILSON, J., E.: ***Predictive Analytics for Business Forecasting and Planning***. Graceway Publishing Company, 2021. ISBN 978-0-9839413-8-5. V rámci publikace jsou z pohledu významu a pojetí prediktivní analytiky ***vyjádření několika špičkových expertů***, např.:

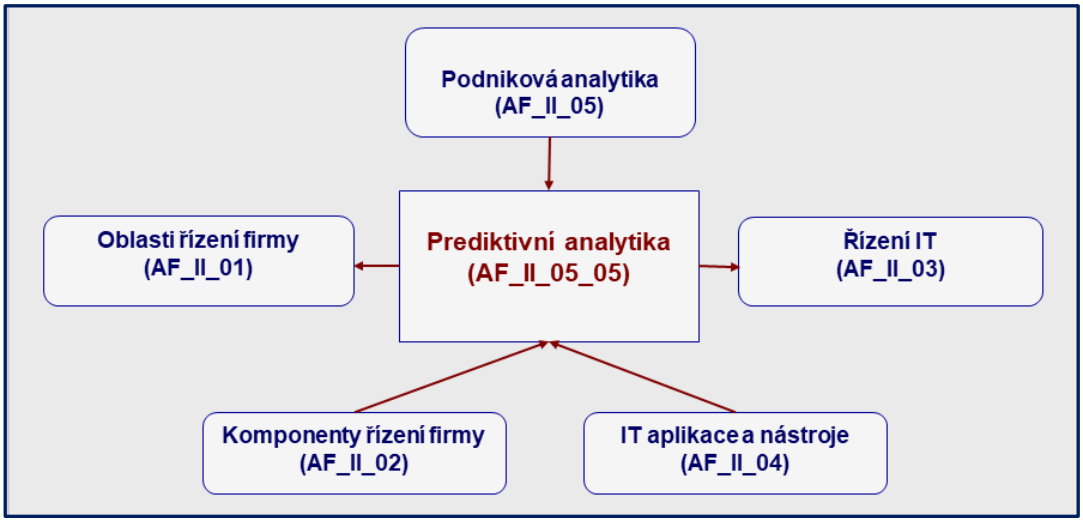
  „*Predictive analytics is the most important business topic of our time extending an organization´s ability to optimize efficiency while improving speed to market….*“
  Joe Eschenbrenner, ACPF
  Director, Demand & Supply Planning
  PUMA Group

  “*Most works on predictive analytics discuss esoteric algorithms and techniques, all too often focusing on buzzwords instead of how practitioners can benefit from their use…*”
  Dr. Larry Lapide
  Research Affiliate
  MIT Center for Transport & Logistics

  “*Cause-and-effect techniques have transformed business forecasting and analytics…*”
  Dr. Barry Keating
  Professor of Business Economics & predictive Analytics
  University of Notre Dame

  “*Integrating predictive data analytics and scalable new techniques is the future of demand planning…*”
  Tim Hotze
  SVP, Global SCM & Logistics,
  Network Planning & Global Intelligence
  Target

S ohledem na značný rozsah informací k prediktivní analytice jsou ***specifické otázky***, které s prediktivní analytikou souvisejí nebo pro ni vytvářejí potřebné předpoklady, ***obsahem dalších pracovních dokumentů*** na portále MBI-AF a v rámci tohoto dokumentu se na ně budeme odvolávat. ***Vazby*** na tyto další dokumenty prezentuje Obrázek 1-1.



Obrázek 1-1: Vazby dokumentu na ostatní dokumenty v rámci MBI-AF

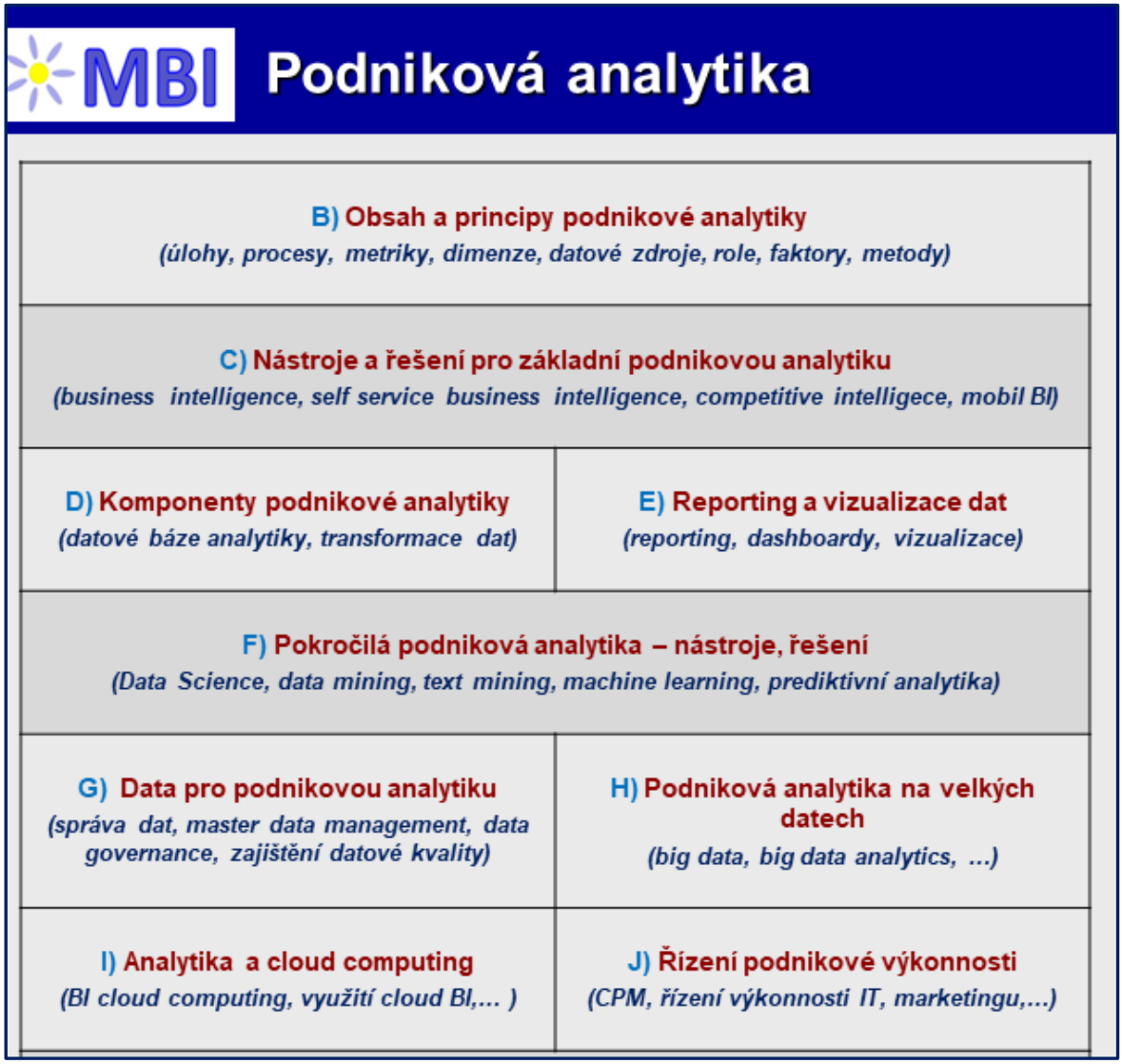
V dalších podkapitolách je vymezen obsah uvedených dokumentů a souvislosti s prediktivní analytikou.

1.1 Podniková analytika

Dokument „*Podniková analytika*“ je ***základním dokumentem***, na který „Prediktivní analytika“ navazuje. Snaží se poskytnout ***celkový***, byť relativně stručný, ***přehled*** o principech, postupech, produktech, problémech i řešeních podnikové analytiky v praxi. Zahrnuje otázky jak „***základní analytiky***“ postavené většinou na nástrojích a přístupech business intelligence, self sertvice business intelligence nebo competitive intelligence, tak ***pokročilé analytiky***, jejíž součástí je i prediktivní analytika. Pro prediktivní analytiku tak ***definuje prostředí*** a výchozí otázky a problémy, které i prediktivní analytika musí řešit

→ ***Dokument***: „*AF_II_05_Podniková_analytika.pdf*“: [Podniková analytika]

Strukturu dokumentu, a tedy i témata, na která bude dokument Prediktivní analytiky navazovat, představuje Obrázek 1-2:



Obrázek 1-2: Podniková analytika, obecně

1.2 Oblasti řízení firmy

Uplatnění prediktivní analytiky se obvykle ***podle jednotlivých oblastí řízení (finance, obchod atd.) liší,*** mají různý dopad na obsah a řešení projektů a je nezbytné potřeby těchto oblastí řízení v rámci firmy identifikovat a analyzovat. Podstatným vstupem jsou tak ***informace a pochopení obsahu řízení firmy*** podle jednotlivých oblastí. Ty jsou obsahem dokumentu, jehož ***odkaz a strukturu*** představuje další část.

→ ***Dokument***: „*AF_II_01_Oblasti.pdf*“: [Oblasti řízení]



Obrázek 1-3: Oblasti řízení firmy

Rekapitulace oblastí řízení firmy je v následujícím přehledu:

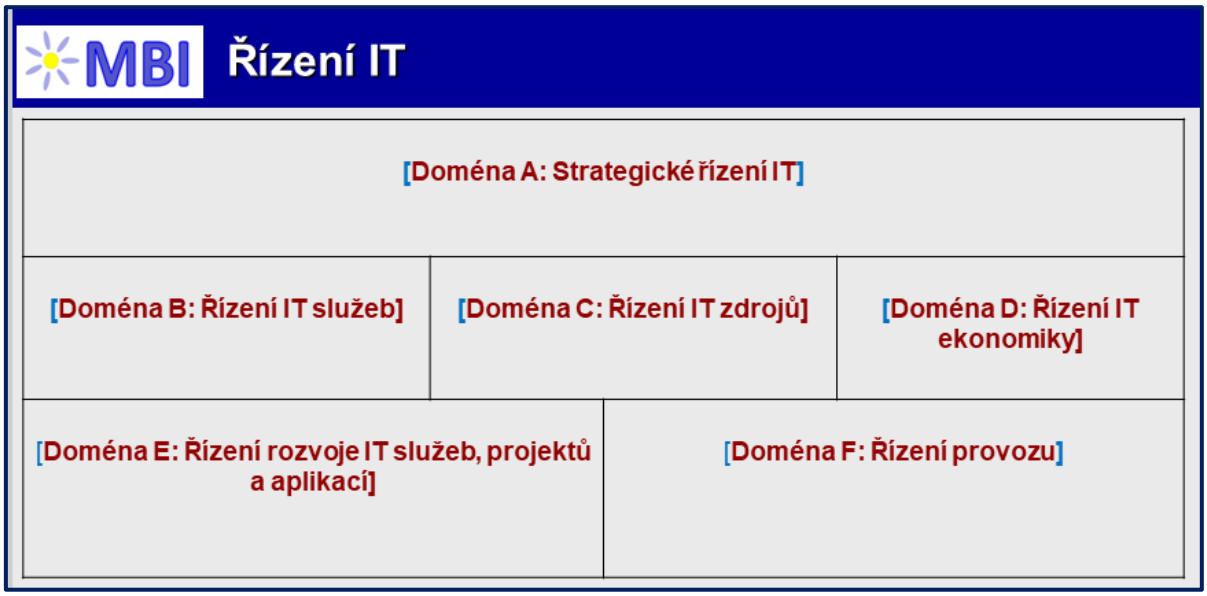
- ***Strategické řízení firmy,*** tj. strategické analýzy, formulace strategie, vytvoření byznys modelu, řízení inovací atd.
- ***Finanční řízení strojírenské firmy,*** tj. účetní evidence, finanční transakce, finanční reporting, finanční analýzy, plánování, zpracování rozpočtů.
- ***Řízení závazků***, tj. evidence závazků a transakce s nimi, reporting závazků, analýzy závazků.
- ***Řízení pohledávek,*** tj. evidence pohledávek a transakce s nimi, reporting pohledávek, analýzy pohledávek.
- ***Práce a mzdy, PAM,*** tj. evidence mzdových složek, evidence a zpracování mezd, mzdový reporting, mzdové analýzy, plánování mzdového vývoje.
- ***Controlling,*** tj. analýzy na bázi controllingu, zpracování controllingových plánů.
- ***Řízení prodeje výrobků, zboží a služeb,*** tj. evidence a řízení obchodních případů „Prodej“, řízení poprodejního servisu, reklamací, reporting prodejů, prodejní analýzy, plánování a prognózování prodeje.
- ***Řízení nákupu materiálů, kooperací a služeb,*** tj. evidence a řízení obchodních případů „Nákup“, reporting nákupů, analýzy nákupů, specifikace potřeb a plánování nákupů.
- ***Řízení skladů,*** tj. evidence skladů a skladových zásob, řízení skladových transakcí, reporting zásob (regleta a další), analýzy zásob.
- ***Personální řízení,*** tj. personální evidence, řízení personálu, přijímání a propouštění zaměstnanců, řízení kvalifikačního rozvoje, personální reporting, personální analýzy, personální plánování.
- ***Řízení a správa majetku,*** tj. evidence majetku, řízení majetkových transakcí, řízení odpisů, reporting majetku, analýzy majetku, plánování rozvoje majetku a investic.
- ***Řízení marketingu,*** tj. evidence, příprava a řízení marketingových akcí, marketingové analýzy, plánování marketingových akcí.
- ***Řízení interní dopravy,*** tj. evidence dopravy a dopravních prostředků, řízení požadavků na dopravu a jejich zajištění, reporting dopravy, dopravní analýzy, plánování dopravních kapacit.
- ***Řízení energií,*** tj. evidence měřidel, řízení spotřeby energií a výroby tepla, analýzy energií, plánování potřeby energií.

1.3 Řízení IT

Řešení úloh a projektů prediktivní analytiky musí nezbytně ***respektovat obsah a pravidla řízení IT*** v dané firmě, např. vychází z IT strategie, respektuje pravidla řízení IT služeb, datových zdrojů nebo IT

ekonomiky, např. na úrovni nákladů na řešení prediktivní analytiky a na druhé straně definování očekávaných efektů. Strukturu dokumentu nabízí následující dokument.

→ ***Dokument***: *„AF_II_03_Rizeni_IT.pdf“*. [Řízení IT]

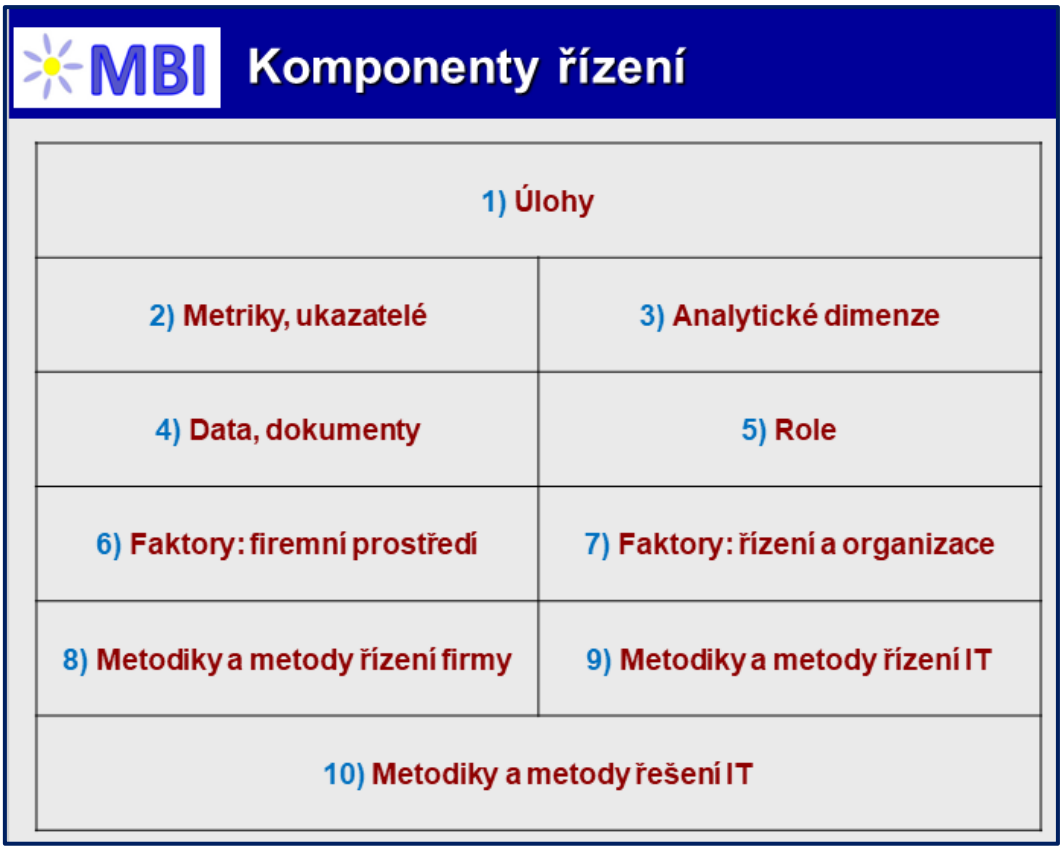


Obrázek 1-4: Domény a oblasti řízení IT ve firmě

1.4 Komponenty řízení firmy

Dokumentace jednotlivých komponent řízení, ***zejména metrik, datových zdrojů, rolí, faktorů a metodik a metod*** je relativně rozsáhlá, a proto je zde vyčleněna do zvláštního dokumentu. Prediktivní analytika na těchto komponentách řízení nezbytně staví. Odkaz a struktura dokumentu jsou:

→ ***Dokument***: *„AF_II_02_Komponenty.pdf“*: [Komponenty a souvislosti]

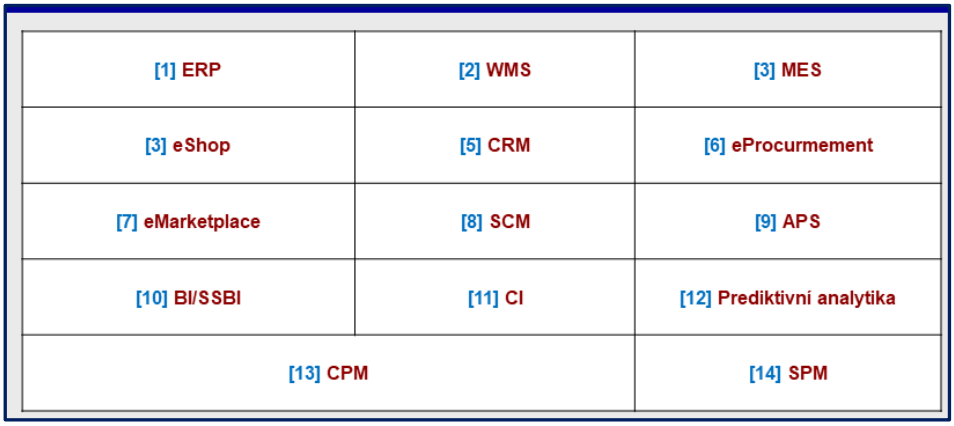


Obrázek 1-5: Komponenty řízení firmy a jejich souvislosti

1.5 IT v řízení firmy

Dokument „*IT v řízení firmy*“ pokrývá ***IT aplikace a nástroje využívané v řízení a fungování firem***, a to jak transakční aplikace (ERP, WMS, CRM a další), tak také aplikace a nástroje podnikové analytiky. Ty jsou především zdroji dat pro funkce prediktivní analytiky. Kromě základního ***vymezení funkcionality*** obsahuje i analytické zhodnocení jejich pozitiv a problémů. Odkaz na dokument a jeho struktura jsou následující:

→ ***Dokument***: „*AF_II_04_IT_aplikace.pdf*“. [IT aplikace]



Obrázek 1-6: IT aplikace a nástroje v řízení firmy

Na úvod k vymezení prediktivní analytiky uvádíme odkazy na další dvě videa:

- [*The Fundamentals of Predictive Analytics - Data Science Wednesday - YouTube*],

- [*What is predictive analytics? Transforming data into future insights - YouTube*],



Další text dokumentu zahrnuje ***3 základní oddíly***:

A. Vymezení obsahu a hlavních principů prediktivní analytiky
B. Specifikace postupu řešení úloh prediktivní analytiky na bázi metodiky CRISP-DM
C. Určení základní náplně řešení prediktivní analytiky podle vymezených oblastí řízení firmy.

A) Prediktivní analytika: obsah, principy



[2] Analytické úlohy ***(deskriptivní, diagnostická, prediktivní, preskriptivní, kognitivní)***	**[3] Potřeba a obsah prediktivní analytiky** ***(potřeba, podstatné charakteristiky, principy, využití, efekty, omezení)***
[4] Role v prediktivní analytice ***(manažer projektu, byznys analytik, datový analytik, klíčový uživatel)***	**[5] Faktory prediktivní analytiky** ***(byznys prostředí, řízení a organizace firmy, řízení prediktivní analytiky, kvalita informačního systému)***
[6] Scénáře, analytické otázky, pro řešení prediktivní analytiky ***(ve vztahu k byznysu ve vztahu k datům, ve vztahu k modelům)***	

Vstupní oddíl A pokrývá zejména její ***obsahovou stránku***, tedy co je nebo by mělo být prvotním zájmem analytika při řešení projektů, případně manažera při řízení byznysu. Patří sem tak všechny ***aspekty tvořící logiku prediktivní analytiky a současně základnu*** pro využití nejrůznějších ***technologií a nástrojů***.

2. Analytické úlohy

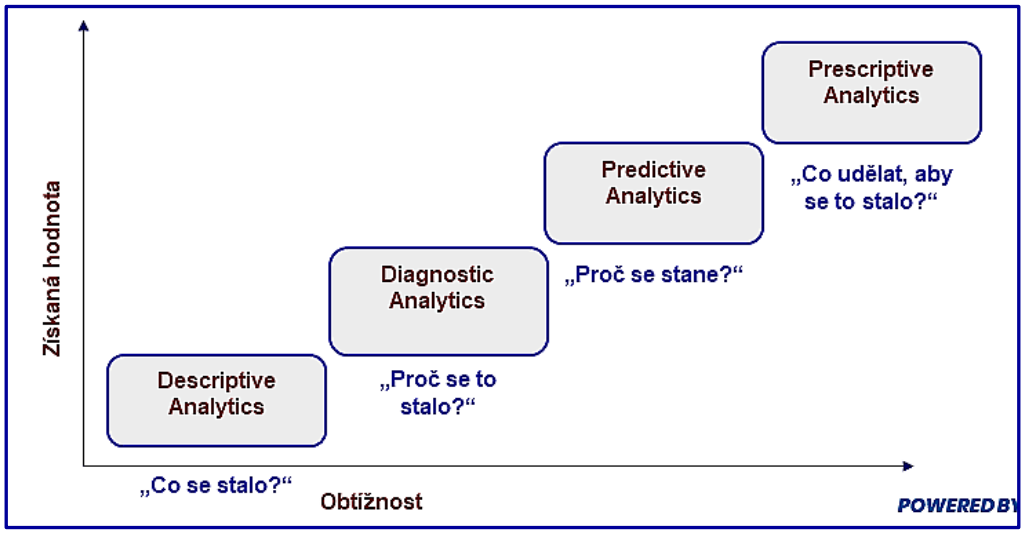
Prediktivní analytika je součástí podnikové analytiky a řadí se ***do komplexu analytických úloh***. ***Účelem*** kapitoly je vymezit ***podstatu těchto úloh*** tak, aby je bylo možné dát do souvislostí s prediktivní analytikou a vzájemně je odlišit.

Analytiku obecně lze vymezit jako proces využívající výpočetní metody k odhalování a prezentování významných vzorů v datech. Cílem je získat kvalitní vědomosti o datech a vytvářet podklady pro efektivní rozhodování.

Jako východisko dalších detailnějších charakteristik prediktivní analytiky je proto účelné využít ***schéma společnosti Gartner*** (Obrázek 2-1), které definuje jednotlivé typy analytických úloh, a to:

- ***deskriptivní analytiku*** („*descriptive analytics*“), odpovídající na otázky „Co se stalo?“,
- ***diagnostickou analytiku*** („*diagnostic analytics*“), odpovídající na otázky „Proč se to stalo?“,
- ***prediktivní analytiku*** („*predictive analytics*“) odpovídající na otázky „Proč a co se stane?“,
- ***preskriptivní analytiku*** („*presriptive analytics*“) odpovídající na otázky „Co udělat, aby se něco stalo?

Vedle těchto 4 typů uvádí E. Wilson ve (Wilson, 2021) i tzv. ***kognitivní analytiku*** („*cognitive analytics*“), která kombinuje několik analytických metod a nástrojů v jeden celek.



Obrázek 2-1: Úrovně řešení a typy úloh analytiky (Zdroj: Gartner Analytic Ascendency Model)

Další podkapitoly podávají stručnou charakteristiku jednotlivých typů úloh (Wilson, 2021).

2.1 Deskriptivní analytika

Deskriptivní analytika je proces ***shromažďování a interpretací dat*** k popisu ***událostí, které nastaly*** v minulosti. Ve skutečnosti naprostá ***většina reportů*** v praxi se vztahuje ***k minulosti***, které se pokouší objasnit minulé události a aktivity a popsat, jak se liší jedna od druhé. Ve většině případů jde tak o deskriptivní analytiku, i když na této úrovni se ***může vyskytovat „data mining“, nebo „machine learning***“, ale většinou k tomu, aby připravovaly data pro vyšší typy analytiky. Deskriptivní analytika ukazuje události minulosti i s problémy, které se u nich vyskytly, avšak bez vysvětlení důvodů, proč k takovým problémům došlo.

2.2 Diagnostická analytika

Diagnostická analytika je proces shromažďování a interpretace dat a ***identifikuje anomálie, detekuje vzory událostí a určuje souvislosti***, které se v rámci nastalých událostí objevily. V rámci diagnostické analytiky se uplatňují funkce alertů, dril-down, data miningu a nejrůznější korelace. Využívá algoritmy ***„machine learning***“ pro funkce klasifikace a korelace.

2.3 Prediktivní analytika

Prediktivní analytika, PA představuje typ analýzy ***využívající data a prediktivní modely pro předpověď jevů*** na mikroekonomické úrovni. Učí se ze zkušeností (dat) a předvídá budoucí chování jedinců, zahrnuje v sobě množství statistických a analytických technik.

Prediktivní analytika je proces a strategie, která využívá ***pokročilé statistické algoritmy k identifikaci vzorů a podmínek***, které se mohou v budoucnosti vyskytnout k tomu, aby bylo možné určovat, co se následně stane. ***Cílem*** je jít za poznání toho, co se stalo, k lepšímu vyhodnocení, co se stane, resp. jaké faktory budou ovlivňovat budoucnost firmy (viz dále). Sleduje a predikuje takové hodnoty, které jsou dosud neznámé, ale mimořádně užitečné.

Otázkou je, ***co odlišuje*** prediktivní analytiku od ostatních analytických úloh, např.:

- prediktivní analytika vychází z dat, „*data driven*“, to znamená, že klíčové charakteristiky modelů jsou odvozovány z dat spíše než z předpokladů vytvářených analytikem. To zahrnuje identifikaci proměnných, parametrů, vah, koeficientů a celkovou komplexitu modelu,
- prediktivní analytika automatizuje proces objevování vzorů v datech, identifikuje např. které hodnoty proměnných budou pro predikce nejlepší.

2.4 Preskriptivní analytika

Prescriptive analytics se zabývá tím, ***co je třeba v budoucnosti dělat***. Vychází tedy z předchozích dvou kroků, kdy ***z descriptive analytics*** a ***predictive analytics***. Závěry vytvořené predictive analytics, tedy to, co se stane, jsou doplněny o to, co s tím můžeme udělat (Ševčík, 2020).

Zároveň prescriptive analytics přináší uživateli ***možnost podívat se na pravděpodobnou budoucnost v rámci různých scénářů***. Tyto scénáře ***vychází z odlišných možných kroků společnosti***, díky kterým může každý usoudit, co je pro ni nejvhodnější z hlediska jejích priorit. V podstatě jsou prezentovány ***možné závěry a důsledky různých akcí s doporučeními*** a je tedy snazší se správně rozhodnout, ale také objevit možnosti, které ani nemusely nikoho napadnout a optimalizovat tak chod společnosti. Snazší, respektive informovanější rozhodování je podpořeno navíc tím, že ***konkrétní scénáře mají také ohodnocení*** z hlediska šance, že nastane právě daný scénář (Logility Blog, 2020).

Jde o to, že pomocí těchto závěrů prescriptive analytics radí uživateli, jak by se měl chovat, aby byly jeho ***procesy a chování, pokud možno, co nejvíce optimální*** a měl usnadněné rozhodování. Může se také stát, že společnost umožní softwaru a umělé inteligenci automaticky provádět určité změny, je-li například nezbytné, aby byly provedeny co nejdříve kvůli efektivitě a množství takto provedených změn. (Gurobi, 2020).

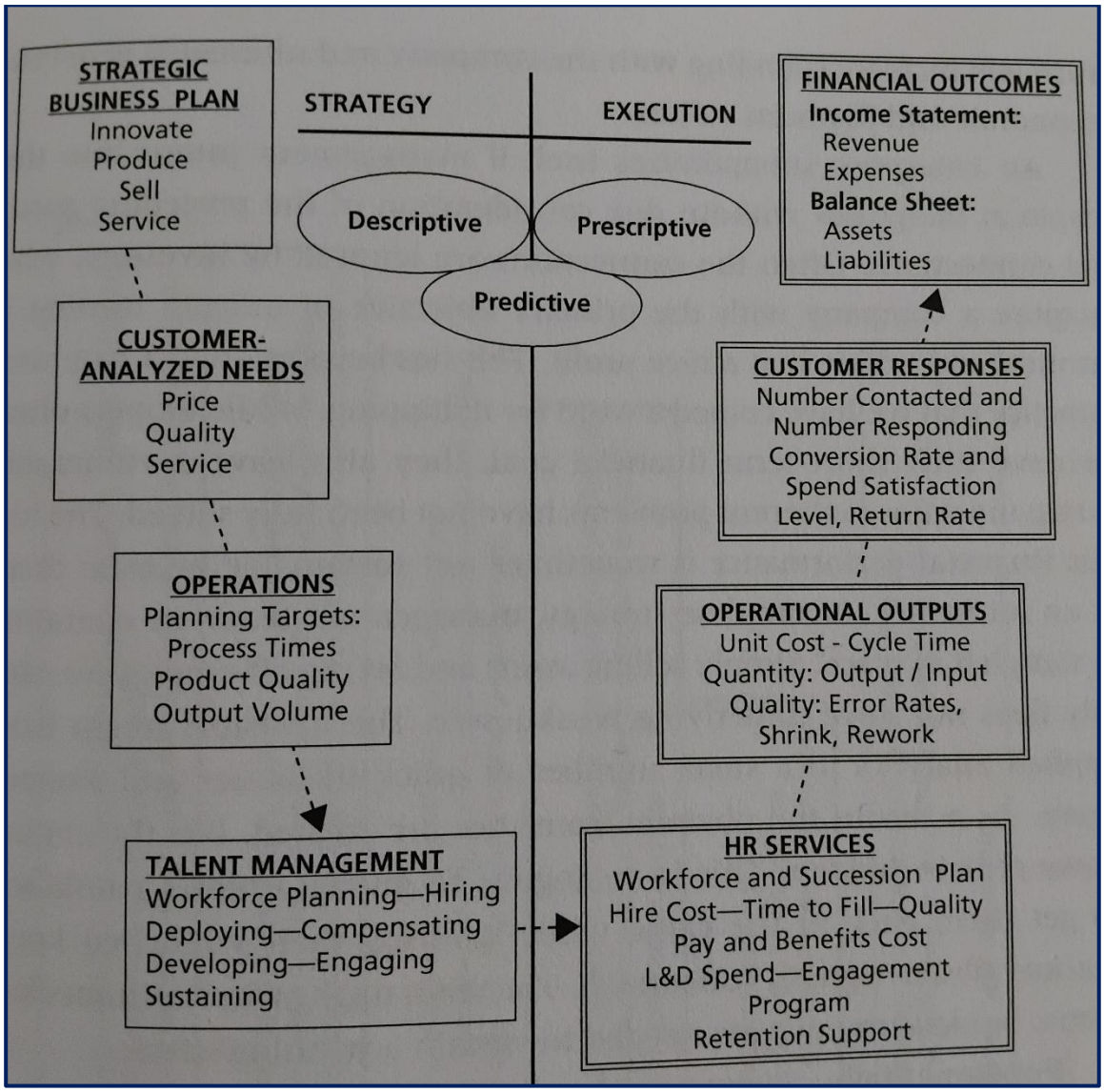
Toto vše funguje díky využití kombinace technik a nástrojů. ***Nejdůležitější*** pro prescriptive analytics je ***strojové učení***, které umožňuje počítači pochopit ohromné množství dat a ***najít mezi nimi souvislosti*** bez potřeby lidského zásahu. Dále je využíváno specifických algoritmů, minulých i současných dat, analytických metod, statistik a modelů pro vytvoření představy a možného scénáře pravděpodobného vývoje nebo důsledku konkrétní akce (Segal, 2020).

Jako konkrétní ***příklad využití*** prescriptive analytics je využití v leteckém průmyslu. Pokud by chtěla společnost maximalizovat své zisky, tak prescriptive analytics toto umožní například pomocí ***doporučení cen a automatických úprav cen*** lístků. Při registraci zvýšeného zájmu o lístky z jednoho místa do druhého by mohlo dojít k automatickému navýšení ceny těchto lístků, ale ne zase o tolik, aby byly lístky dražší než lístky konkurence. Naopak pokud by docházelo k poklesu zájmu o lístky i přes automatické snížení ceny lístků a už by nebylo možné nadále snižovat cenu kvůli cenám paliva, mohlo by dojít k návrhu, třeba použít menší letadlo pro tuto linku, snížit frekvenci letů či linku úplně zrušit. Všechny tyto možnosti by měly také připravené vyvozené závěry plynoucí z provedení konkrétního opatření a míru pravděpodobnosti naplnění tohoto závěru (Segal, 2020).

2.5 Kognitivní analytika

Kognitivní analytika je proces, který zahrnuje ***kombinace sémantiky, umělé inteligence, strojového učení*** a další škálu komplexních algoritmů k podpoře vysoce efektivního prognózování a plánování v řídící praxi.

Pro práci analytika jsou z uvedených typů úloh ***nejvýznamnější deskriptivní, prediktivní a preskriptivní***. To, jak se ***promítají do hlavních fází řízení*** ukazuje podle (Fitz-enz, Mattox II, 2014) následující obrázek:



Obrázek 2-2: Uplatnění analytických úloh v řízení, resp. "Analytic Value Chain" (Zdroj: Fitz-enz, Mattox II, 2014)

2.6 Závěry

- V rámci analytických úloh se rozlišuje jejich ***pět základních typů:***
  - ***deskriptivní*** – analyzují události a aktivity v minulosti a připravují data pro vyšší typy úloh,
  - ***diagnostické*** – analyzují události a aktivity rovněž v minulosti, ale s vyhodnocením důvodů, které k nim vedly,
  - ***prediktivní*** – určují události a aktivity, které s jistou pravděpodobností nastanou v budoucnosti,
  - ***preskriptivní*** – určují události a aktivity, které pravděpodobností nastanou v budoucnosti, ale i s určením řešení, které k nim povedou,

- **kognitivní** – jsou založené na kombinacích různých metod a přístupů, které vedou k efektivnímu určování jevů v budoucnosti.

- ***Úlohy prediktivní analytiky*** tak vytvářejí ***přechod*** mezi úlohami zaměřenými na analýzy minulosti k prognózování a plánování budoucnosti.

3. Potřeba a obsah prediktivní analytiky



Obsah řízení a prediktivní analytiky je ***velmi variantní*** podle odvětví, typu a zaměření firem a jeho zvládnutí je pro analytiky hodně náročné. Z pohledu řízení firmy je ale samotným základem pro prediktivní analytiku plánování, resp. plánování poptávky („*demand planning*“) a prognózování („*forecasting*“), a to s uplatněním pokročilých principů, např. umělé inteligence a „*machine learning*“.

Účelem kapitoly je:

- definovat ***potřeby byznysu*** vzhledem k prediktivní analytice, tedy potřeby a požadavky plánování a prognózování ve firmě,
- ***vymezit*** prediktivní analytiku, ***kategorizovat součásti*** jejího obsahu, a to ve vztahu potřebám prognózování a plánování,
- specifikovat ***klíčové principy***, na nichž je prediktivní analytika založena,
- zhodnotit podstatné ***efekty a omezení*** prediktivní analytiky,
- určit ***oblasti využití*** prediktivní analytiky.

Na úvod ke kapitole je účelné rozlišit úlohy reportingu a analytické úlohy, včetně prediktivních z pohledu dopadu na užití a řízení, jak ukazuje další tabulka.

Tabulka 3-1: Porovnání úloh reportingu a analytických úloh (Zdroj: Fitz-enz, Mattox II, 2014, upraveno)

Úlohy reportingu	Analytické úlohy, včetně prediktivních
Poskytují data, upravená, strukturovaná, komentovaná	Poskytují odpovědi na analytické a plánovací otázky a problémy
Poskytují odpovědi na to, co je požadováno.	Poskytují informace k tomu, co je reálně potřeba
Jsou typicky velmi standardizované.	Jsou silně customizované podle potřeb jednotlivých uživatelů, především manažerů a specialistů firmy.
Nemusí zahrnovat aktivní součinnost uživatele na řešení.	Předpokládají intenzivní součinnost uživatelů na řešení analytických a plánovacích úloh
Jsou omezené z hlediska flexibility vzhledem ke změnám v potřebách a zdrojích.	Jsou vysoce flexibilní vzhledem ke změnám.

3.1 Potřeba prediktivní analytiky: prognózování a plánování

Prognózování a ***plánování*** v řízení firmy jsou aktivity, které ***spolu souvisejí***, i když reálně jsou v některých případech realizovány nezávisle na sobě. Platí ale, že čím kvalitnější a přesnější jsou prognózy, tím kvalitnější jsou plány a plánování.

Rozvoj plánování a prognózování, jejich funkcionality ve vazbě na prediktivní analytiku je možné považovat za jednu z ***klíčových oblastí rozvoje byznysu*** a investic ve firmách bez ohledu na to, v kterém odvětví působí, jaká je jejich velikost, složitost apod. Současné firmy disponují obrovskými objemy dat a smyslem konceptu, metod a nástrojů prediktivní analytiky je z nich vytěžit právě takové informace, které povedou k ***posilování kvality plánování prognózování*** a tím i jejich konkurenceschopnosti a získávání nových konkurenčních výhod.

3.1.1 Prognózování v byznysu

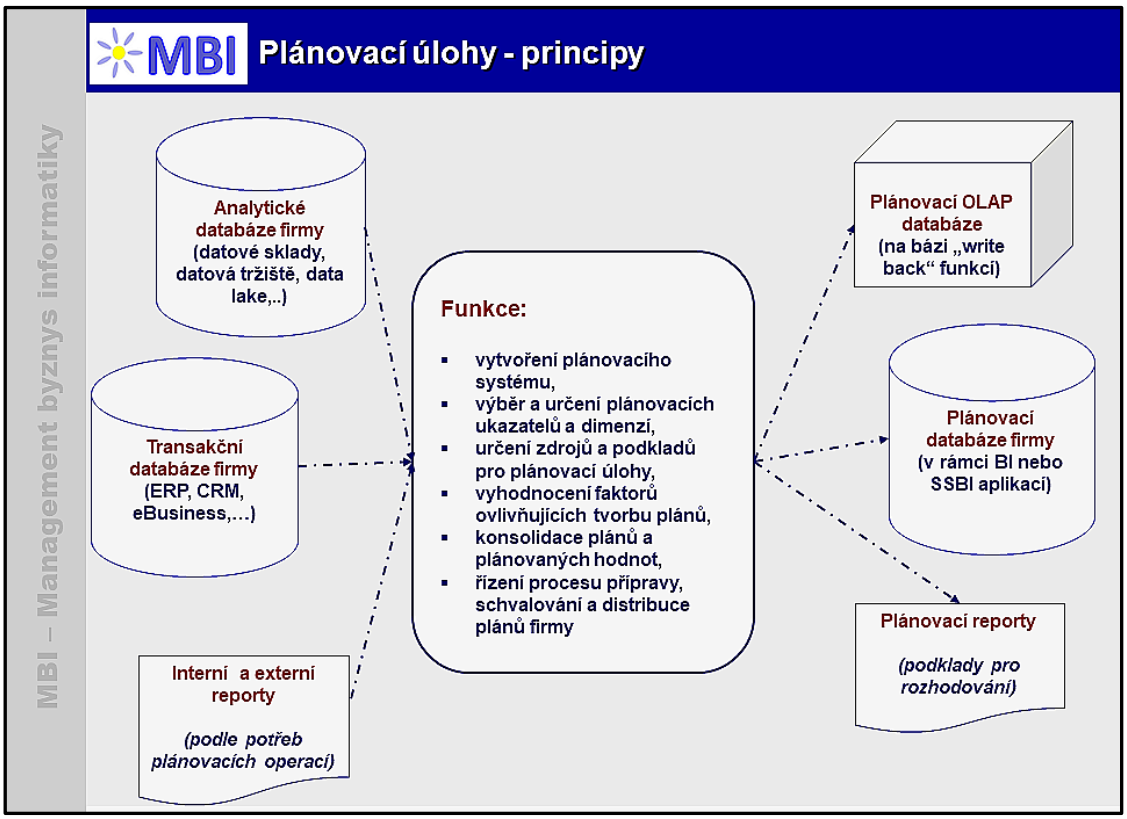
Prognózování v byznysu („*business forecasting*“) lze chápat jako proces, který využívá data, analytiku a vyhodnocování zkušeností pro vytváření predikcí podle potřeb byznysu. Prognózování ***lze rozdělit,*** obdobně jako jiné řídící aktivity, na (Wilson, 2021):

- ***strategické,*** na dlouhé období, ale s rizikem větších chyb,
- ***taktické***, se střední dobou odhadů, obvykle s vysokou přesností,
- ***operativní***, real time umožňují bezprostředně reagovat na změny tržního prostředí, změny poptávky, změny v disponibilních zdrojích apod.

Prognózování je většinou založené na ***sledování trendů, sezónních vlivů, nejrůznějších faktorů***, na jejichž základě se formulují odhady, prognózy budoucího vývoje, např. obchodu, ekonomiky, technologií ve firmě.

3.1.2 Podstata a principy plánování

Plánování v byznysu se přímo váže na prognózování a v některých případech staví na výstupech prognózování, resp. jednotlivých prognóz. Je to de facto projekce budoucnosti, která kombinuje znalost minulosti s vyhodnocením budoucích potřeb produktů a služeb. Podstatu a principy plánovacích úloh dokumentuje Obrázek 3-1.



Obrázek 3-1: Principy plánování a plánovacích úloh

Hlavním smyslem plánování je ***snížení nejistoty budoucího vývoje,*** tj. plány a rozpočty poskytují kontrolní mechanismus, zdali jsou naplánované cíle plněny s určitými povolenými odchylkami (Žůrková 2007, s. 9). Klíčovým aspektem systému plánů a rozpočtů.je ***časový předstih***, s nímž jsou možná rizika a úzká místa plnění cílů a řízení firmy identifikována.

Úrovně plánování zahrnují (v návaznosti na úrovně prognóz) 3 základní úrovně plánů:

- strategické, cca na 10 let,
- taktické (manažerské), 1 – 3 roky,
- operativní, denní, týdenní, měsíční či kvartální.

Klouzavé plánování zahrnuje na ně navázané klouzavé rozpočty. Funguje na principu neustálé aktualizace plánů a rozpočtů, a to na základě skutečně sledovaného vývoje. Umožňuje průběžně porovnávat plány a rozpočty se skutečností a vyhodnocovat jejich dosavadní přesnost a příslušně je upravovat pro další období.

3.1.3 Podstata plánovacích úloh

Pro řešení plánovacích úloh (plánů tržeb, nákladů, investic atd.) je ***třeba nejen data zpřístupňovat podle nejrůznějších definovaných dimenzí*** a analyzovat je podle nich, ale i nová data, tedy ***plány tvořit, rovněž s respektováním těchto dimenzí*** a jejich hierarchických struktur. Plánovací úlohy sledují ***několik základních cílů***, zejména:

- vytvořit a ***využít plánovací systém*** respektující ve firmě uplatňované plánovací a rozvrhové metody,
- ***zajistit konsolidace vytvářených plánů*** vznikajících na různých organizačních jednotkách, tj. závodech, divizích, odděleních, nebo naopak rozpouštění centrálně stanovených plánů na tyto jednotky,
- ***zajistit konsolidace hodnot z různých druhů plánů***, např. plánu investičního, výrobního nákupního, prodejního, personálního apod. do výsledného, obvykle finančního plánu,
- ***zajistit konsolidace plánů z pohledu různých měn*** a přepočet na výslednou požadovanou měnu,
- automatizovat ***řízení pracovního toku (workflow)*** při přípravě plánu, resp. plánů, na kterém se podílejí různí manažeři, plánovači a další pracovníci podniku,
- efektivně ***zpřístupňovat sestavené plány*** zainteresovaným pracovníkům podniku,
- zajistit ***potřebnou bezpečnost a nastavení přístupových práv*** pro zpracování plánů i pro jejich presentaci ve firmě, případně mimo ni, kde jde o možnosti jejich čtení, zápisu a schvalování.

3.1.4 Principy plánovacích úloh

Data do analytických databází se obvykle ***nahrávají z primárních zdrojů***. Jinou možností je však tzv. ***zpětný zápis***, což je funkce, kdy data do analytické databáze může zadávat přímo uživatel, nebo určitá, většinou plánovací aplikace. To znamená, že uživatel, resp. uživatelská aplikace zde může měnit hodnoty ukazatelů a dimenzí. Takové možnosti vytváření plánů a realizace změn prováděných uživatelem zpětným zápisem jsou tak základem řešení efektivních a výkonných plánovacích aplikací, které přinášejí ***několik podstatných efektů***:

- ***využití již existujících hodnot*** dříve vytvořených plánů, nebo hodnot o skutečnosti z minulých období, které se tak stávají základem pro tvorbu nových plánů,
- ***využití časových řad,*** vyhodnocování podstatných faktorů a proměnných a na jejich základě vytvoření jejich extrapolací,
- ***pružné zapracování změn plánů***, kdy je třeba promítnout změny v celé plánovací struktuře, např. pokud se změní plán v jednom oddělení, je nutné měnit i konsolidované plány na vyšších úrovních podnikového řízení, a naopak kdy je nutné promítat změny z vyšších organizačních jednotek na nižší,
- ***promítání změn zpět do zdrojových databází*** transakčních aplikací lze efektivně využívat k distribuci aktuálních plánů do aplikací nebo databází podřízených útvarů, divizí, jednotlivých poboček apod.,
- při využití efektivních ***rozvrhových algoritmů aplikovaných na struktury jednotlivých dimenzí*** a jejich kombinace (např. procentuálních nebo obdobných rozvrhových schémat) lze podnikové plány sestavovat podstatně efektivněji a zajistit jejich konsistenci,
- je možné definovat a využívat tzv. ***podniková pravidla*** (*business rules*) nebo omezení (*business constraints*), tj. pravidla pro kontroly plánů, např. objem investic nepřesáhne 20 % plánovaných nákladů celého podniku apod.,
- podnikové plány jsou ***centrálně uložené v analytické databázi***, a to znamená jejich rychlé a konzistentní vytváření s rozpadem, či agregací plánovacích ukazatelů dle definovaných dimenzí a možnost rychlého porovnání skutečnosti s plánem, pokud je použita stejná struktura ukazatelů a dimenzí.

3.1.5 Plánovací úlohy v kontextu řízení firmy

Plánovací úlohy jsou rovněž ***součástí naprosté většiny oblastí řízení*** firmy i firmy jako celku. Na základě definovaných principů lze určit jejich ***uplatnění v oblastech řízení firmy,*** což je pak

konkretizováno dle oblastí řízení (v dokumentu „*AF.II.01: Oblasti řízení*“). ***Pracovně*** lze přijmout následující ***klasifikaci***:

- ***Sestavení plánu*** za vybranou oblast ***podle vybraných ukazatelů*** a s využitím relevantních dimenzí, např. plán tržeb v daném období, plán obejmu nákupů vybraných materiálů, plán počtu servisních zásahů na dodaných zboží apod.
- ***Zpracování výhledu podle vybraných ukazatelů a s využitím relevantních dimenzí***, např. na 12 měsíců klouzavě v průběhu roku.
- ***Sestavení rozpočtu,*** tedy finančních hodnot ukazatelů, které silně závisí na prostředí konkrétní organizace, charakteristice trhu, druhu nabízeného produktu a dalších.

V každém případě jsou dílčí charakteristiky prognózování a plánování a plánovacích úloh podstatným předpokladem ***pochopení potřeb prediktivní analytiky***, jejího řešení a uplatnění v manažerské praxi.

3.2 Rekapitulace vybraných ukazatelů

Na závěr této úvodní části je dobré připomenout, že prakticky ve všech uvedených typech úloh lze využívat alespoň základní ***ukazatelé***, jejichž základ je ve statistice. Pouze jako upozornění a rekapitulace je uveden přehled vybraných z nich bez dalšího vysvětlení, neboť k tomu jsou k dispozici podstatně kvalifikovanější statistické texty. K tomu ***několik poznámek***:

- statistika nabízí celou ***škálu ukazatelů (statistických znaků, resp. proměnných)*** a jejich přesné vymezení, které mohou být a jsou v praxi analytických úloh často velmi efektivně využitelné, a to:
  - v rozlišení na ***kvantitativní a kvalitativní***, resp. ***kategoriální***,
  - kvantitativní proměnné se pak člení na ***měřitelné*** (metriky) a ***pořadové*** (ordinální), nebo nespojité a spojité,
- jednou z možností uspořádání ukazatelů je rozdělení četností získaných hodnot, resp. ***intervalové rozdělení četností***, kde variační rozpětí celého souboru, což je rozdíl mezi maximální a minimální hodnotou, se rozdělí na určitý počet intervalů s uvedením odpovídajících počtů hodnot,
- u rozdělení četností významná hodnota „vrcholu“, tedy nejčetnější hodnoty, označované jako ***modus***,
- ***kvantil*** je hodnota, která soubor rozděluje na části, jedna obsahuje hodnoty nebo stejné než kvantil, druhá větší nebo stejné, využívají se medián, kvartil, decil, percentil,
- ***medián*** je prostřední hodnota, resp. 50 % kvantil,
- střední hodnota ze všech hodnot ukazatelů je ***průměr***, a může být ***aritmetický, harmonický, geometrický, kvadratický***,
- průměry se rozlišují na ***prosté a vážené***,
- statistika nabízí také ***míry variability***, jejichž velikost je závislá na všech hodnotách souboru, zejména:
  - ***rozptyl*** – měří současně variabilitu hodnot kolem aritmetického průměru a současně variabilitu ve smyslu vzájemných odchylek jednotlivých hodnot,
  - ***směrodatná odchylka*** – jako kladná odmocnina rozptylu a vyjadřuje se ve stejných jednotkách jako základní proměnné.

3.3 Podstatné principy prediktivní analytiky

Výchozím principem a současně předpokladem řešení a užití prediktivní analytiky je to, že jejím hlavním účelem je poskytovat ***kvalitní a sofistikovanou podporu prognózování a plánování*** v řízení firmy (viz výše) a tím přispívat k posilování její pozice na trhu a celkové ekonomické a obchodní úspěšnosti.

Prediktivní analytika staví na jiných analyticky orientovaných disciplínách jako data mining, machine learning, statistika nebo rozpoznávání vzorů. Prediktivní analytika (Praus, 2013) je ***typ analýzy využívající data a prediktivní modely pro předpověď jevů na mikroekonomické úrovni.*** Prediktivní analytika využívá technologie, které se učí ze zkušeností (dat), aby předvídala budoucí chování jedinců k lepšímu rozhodování (Siegel, 2013).

Hlavním předpokladem, se kterým pracuje, je to, že ***v lidském chování lze nalézt podobnosti***, vztahy a vzorce. Z dostupných dat je prediktivní analytika schopná, ***pomocí technik dolování dat***, identifikovat důležité podobnosti a vztahy a využít je k předpovědi nejrůznějších jevů událostí a aktivit.

Hlavní principy, na nichž je založeno řešení a využití prediktivní analytiky v praxi, je možné shrnout ***do následujících bodů*** (Abbott, 2014, Siegel, 2013, Wilson, 2021):

- Řešení prediktivní analytiky často závisí na potřebě ***analyzovat hodnoty, které nejsou přesně známé***, ale mimořádně užitečné. V některých případech jsou takové hodnoty známé historicky, ale nejsou známé, kdy má docházet k rozhodnutí, resp. predikovanému rozhodnutí.
- ***Využití prediktivní analytiky*** se váže k různým sférám zajmu, např. k odhadům prodeje zboží, vývoje cen, ale i makroekonomických nebo demografických trendů, počasí a dalších.
- Jde o proces ***využití výpočetních metod*** k nalézání a reportování významných vzorů, resp. schémat v datech. Prediktivní analytika vyhodnocuje historická data a je ***realizována ve vztahu k disciplínám a metodám*** jako je např. business intelligence, competitive intelligence, data science a další..
- Zvyšování pozornosti prediktivní analytice je založeno na tom, že ***firmy disponují stále většími objemy dat*** a přirozená cesta je využívat je ke zlepšování odhadů, prognóz, rozhodovacích činností a zvyšování celkové efektivity, a to v podstatně kratším čase, než tomu bylo dosud.
- Prediktivní analytika ***zahrnuje nejistotu***: prognózy nemohou být nikdy přesné, výsledky na 100 % neexistují. Do odpovědí musí být vždy zahrnuta pravděpodobnost a ohodnocení chyb.
- Na druhé straně, zatímco tradiční metody prognózování a plánování vyžadovaly relativně přesně otázky a požadavky, metody prediktivní analytiky ***hledají odpovědi i na špatně formulované nebo i žádné otázky***.
- ***Řešení*** prediktivní analytiky ***zahrnuje*** jak „***vědu***“, resp. vědecké disciplíny, tak „***umění***“ reprezentované zejména praktickými zkušenostmi řešitelů projektů. Přístup jeden bez druhého, ale nevede obvykle k úspěchu.
- Na prediktivní analytice se podílejí zejména ***tyto role*** doménový expert, data a databázový expert a expert prediktivního modelování. Jejich funkční náplň a požadavky jsou obsahem ***kapitoly 4.***
- Podmínkou efektivního řešení a využití prediktivní analytiky je schopnost ***správně formulovat analytické otázky*** ve vztahu k jejímu řešení. Vstupní návrhy k přípravě takových otázek obsahuje ***kapitola 5***.
- Prediktivní analytika je ***méně přesná, pokud jde do většího detailu***: vysoká granularita v atributech, vlastnostech produktů apod. vždy povede k větším nepřesnostem, než když bude úroveň detailu nízká. I v dimenzi času, pokud se prognózy pohybují od roků, měsíců, např. ke dnům klesá jejich přesnost.
- Prediktivní analytika je ***přesnější, čím je blíže k události, k níž se vztahuje***: pokud se bude prognóza realizovat s relativně větším časovým zpožděním ve vztahu k nastalé situaci, pak se přesnost snižuje. Cílem např. je se dostat co nejblíže k aktuální zákaznické poptávce.
- Prediktivní analytika ***se zlepšuje s objemem informací a znalostí*** o daném jevu: cílem je získat co největší a nejkomplexnější obrázek o situaci, která má být předmětem analytiky.
- Nejvýznamnější součástí prognózování a plánování, a tedy i prediktivní analytiky jsou ***předpoklady*** a jsou důležitější než vlastní čísla. Předpoklady jsou velmi četné a významně přispívají k pochopení komplexity a nejistoty reality.
- Dalším podstatným principem jsou ***scénáře***. Představují potenciální okolnosti a kombinace předpokladů, které mohou mít významný dopad na firmu. Při větším počtu definovaných předpokladů je velmi pravděpodobné, že prediktivní analytika povede i k většímu počtu výsledků. Je pak nezbytné vyhodnotit faktory ty, které budou ovlivňovat, které z výsledků budou mít v dané situaci vyšší ***pravděpodobnost*** uplatnění a které méně.

3.4 Řešení prediktivní analytiky

Celkový ***detailnější postup*** řešení úloh prediktivní analytiky podle metodiky CRISP-DM je ***náplní celého oddílu B).*** Na tomto místě shrneme pouze základní aspekty řešení s odkazy na další části textu.

Vstupem prediktivní analytiky jsou data se ***znalostí cílové proměnné a její hodnoty (targetu)***, která má být v budoucnu predikována. Data se v rámci prediktivní analytiky uspořádají, pročistí a

vytvoří se nový datový zdroj. Na něj jsou poté ***aplikovány funkce a metody***, jejich základní charakteristiky jsou náplní oddílu D). ***Výstupem*** je nejúspěšnější prediktivní model, schopný s určitou pravděpodobností hodnotu cílové proměnné předpovídat.

Řešení, výsledky i využití úloh prediktivní analytiky je významně ovlivňováno celou ***soustavou faktorů***, které jsou pro kvalitu řešení nutné dobře pochopit a vyhodnotit. Jejich podstata je ***náplní kapitoly 5***.

3.5 Efekty uplatnění prediktivní analytiky

Efekty uplatnění prediktivní analytiky se vztahují ke zkvalitnění řízení a rozhodovacích aktivit firmy, a to primárně v následujících bodech:

- Schopnost úspěšné predikce je ***využita ke zlepšení rozhodnutí***, které je tak více postavené na faktech (vztazích, trendech) nalezených v datech než na intuici. Postupně se realizuje ***promítání aktuálních událostí do prediktivních modelů*** v reálném čase.
- Předpovědi poskytnuté prediktivní analytikou se týkají převážně ***mikroekonomických efektů***, vyhodnocuje se chování jednoho člověka, a ne masy lidí. Například předvídá, kdo si, s jakou pravděpodobností koupí nějaký produktů (Siegel, 2013).
- ***Prediktivní analytika a plánovací scénáře*** umožňují firmě řešit alternativní situace rychleji a efektivněji. Prediktivní analytika, která je zaměřená na ***identifikaci parametrů a faktorů*** ovlivňujících prognózy přispívá k určování aktivit, které mají být realizovány pro dosažení požadovaných budoucích výsledků.

3.6 Omezení, problémy, předpoklady řešení prediktivní analytiky

Problémy a předpoklady řešení prediktivní analytiky se promítají do jednotlivých fází řešení (Abbott, 2014), které jsou pak konkretizovány v ***oddíle B).***

- ***Problémy v řízení:***
  - Obvyklým problémem je zajištění potřebné kvality dat, tedy vyloučení redundancí, duplicit, chyb, řešení absence unifikovaných dat.
  - Obdobně určité zdroje vyžaduje i nasazení prediktivní analytiky do provozu, a proto mnohé modely a aplikace nejsou vůbec využity
- ***Problémy v datech:***
  - Obvyklým problémem je zajištění potřebné kvality dat, redundance, duplicity, chyby, absence unifikovaných dat.
  - Specifickým problémem je potřeba transformovat data ze zdrojových systémů (např. ERP) do pouze dvourozměrných tabulek pro prediktivní analytiku. To může být při složitých datových strukturách problém.
  - Problém je často spojen i s nedostatkem dat pro vyvinutí úspěšných prediktivních modelů, i s potřebnou aktualizací dat v konzistenci se zdrojovými systémy.
- ***Problémy s modely***:
  - Hlavní častý problém modelů je přeučení („*overfitting*“), modely jsou příliš komplikované.
  - To znamená, že model špatně funguje na nových datech a interpretace výsledků je nespolehlivá. Často pro přeučení se v průběhu řešení ani neprojeví, ale až při předání do provozu se ukáže jeho slabší výkon.
  - Další problém je špatný odhad náročnosti řešení modelu vzhledem k dostupnosti dat i komplikovanosti funkcí a tím problém s dodržením časových termínů pro předání do používání. Je lepší proto začít s řešením relativně jednodušším a následně přidávat další funkce.
- ***Problémy s distribucí modelů:***
  - Problémy mohou být spojeny s integrací modelů se zdrojovými systémy a jejich dostupností, případně s jejich upgrady.
  - Realizace prediktivní analytiky je časově náročná s nejistým výsledkem.

3.7 Příklady využití prediktivní analytiky

V současné době se využití prediktivní analytiky ***posunulo od velkých společností i k menším*** nebo středně velkým firmám. Současně se analytici často posunují od business intelligence k prediktivní analytice, jak stoupá tlak na co nejvyšší využití disponibilních dat. Existuje ***velké množství aplikací*** prediktivní analytiky napříč mnoha obory lidské činnosti. Od bankovnictví, pojišťovnictví, marketingu po medicínu, bezpečnost a další, jako např. (SIEGEL, 2013), např.:

- Americká města Chicago, Los Angeles, Memphis (TN), Richmond (VA), Santa Cruz (CA) a Vineland (NJ) směřují policejní hlídky do oblastí, kde je predikováno nejvyšší riziko kriminality.
- FedEx predikuje s 65 % až 90 % úspěšností zákazníky, kteří přejdou ke konkurenci.
- Odhadem 40 % obchodů na londýnské burze je prováděno algoritmy.

Prediktivní analytiku ***využívají především podniky*** jako např.:

- ***výrobní a obchodní firmy***:
  - odhady objemu financí a finančních zdrojů,
  - predikce objemu tržeb podle zákazníků,
  - prediktivní údržba,
- ***banky***:
  - hodnocení klienta a kredibility,
  - identifikace a hodnocení rizikovosti půjček,
  - vyhodnocování trendů,
- ***pojišťovny***:
  - při odhalování podvodných pojistných událostí,
  - pří upisování,
  - při cenění rizik a pojistného,
  - pro zlepšení efektivity marketingových kampaní,
- ***telekomunikační společnosti***:
  - vyhodnocování trendů,
  - lepší cílení marketingových kampaní,
- ***burzy***:
  - predikce vývoje akcií a komodit,
  - obecně se najde využití v marketingu, prodeji, finančních službách, medicíně

Obecně se ale dá předpokládat, že se úlohy prediktivní analytiky využívají nebo budou využívat ***v různých oblastech řízení*** napříč odvětvími ekonomiky. Základní ***specifikaci řešení, přes oblasti řízení*** vymezené v dokumentu „*AF_II_01_Oblasti.pdf*“ obsahuje ***oddíl C).***

3.8 Závěry



- ***Potřeba*** prediktivní analytiky se váže převážně k ***řešení prognóz*** v rámci firemního řízení a následně ***přípravě plánů a rozpočtů***. Musí respektovat principy a specifické nároky na funkcionalitu a data těchto úloh.
- Řešení a využití prediktivní analytiky je ***založeno na řadě principů*** účelných pro pochopení jejích možností, zejména *„řešení prediktivní analytiky zahrnuje jak „vědu“, resp. vědecké disciplíny, tak „umění“ reprezentované zejména praktickými zkušenostmi řešitelů projektů“* a tomu musí odpovídat i role expertů podílejících se na těchto řešeních apod.
- Obdobně je při řešení prediktivní analytiky nezbytné kvalifikovaně ***vyhodnotit efekty***, které má a může přinést a ***omezení***, s nimiž je třeba počítat. To je podstatným předpokladem pro úspěšnou realizaci projektů.

- Využití prediktivní analytiky pokrývá již nejen ***velké společnosti, ale i střední a menší***, přes nejrůznější odvětví a typy firem a napříč oblastmi jejich řízení.
- Prediktivní analytika je součástí celého ***komplexu disciplín*** tvořících ***podnikovou analytiku***, které jsou náplní další kapitoly.

4. Role v prediktivní analytice



Role v prediktivní analytice a při řešení projektů v této oblasti zahrnují ***obdobnou škálu rolí***, jako v jiných typech projektů, které jsou uvedeny zejména v dokumentu „*AF_II_02_Komponenty.pdf*".

Na druhé straně se zde zdůrazňují role na ***principu stoličky o třech nohách*** („*The Three-Legged Stool*"), a to:

- doménový expert,
- data a databázový expert,
- expert prediktivního modelování.

Účelem kapitoly je vymezit hlavní ***funkční náplň*** uvedených rolí.

Další podkapitoly obsahují hlavní úkoly a funkce uvedených expertů.

4.1 Doménový expert

Doménový expert ***rozumí byznys problematice***, a tedy potřebám byznysu, jednotlivých oblastí řízení vzhledem k možnostem prediktivní analytiky*:*

- specifikuje charakter a ***nároky na prognózování a plánování*** v rámci firmy i ve vztahu k jednotlivým oblastem jejích aktivit,
- formuluje ***zadání nových plánovacích aplikací*** a reportů, specifikuje nároky na jejich kvalitu, tj. funkcionalitu, dostupnost, uživatelské rozhraní apod.,
- ***konzultuje požadavky*** na řešení prediktivní analytiky v rámci své oblasti řízení, a to z pohledu hlavních potřeb firmy,
- zajišťuje ***hodnocení funkcionality*** stávajících řešení a aplikací prediktivní analytiky vzhledem k vývoji potřeb firmy, případně k vývoji požadavků partnerů,
- určuje ***požadavky na funkce operací*** prediktivní analytiky,
- ***analyzuje požadavky*** uživatelů na prediktivní funkce, řeší jejich konsolidaci,
- poskytuje informace a zkušenosti o ***stavu a kvalitě relevantních zdrojových databází***,
- definuje specifikace ***cílových proměnných***, resp. ukazatelů,
- konzultuje specifické otázky ***transformací dat***,
- podílí se na řešení ***problémů spojených s kvalitou dat***,
- podílí se se ***kvalifikační přípravě*** všech ostatních uživatelů.

4.2 Data a databázový expert

Data a databázový expert se orientuje v datových zdrojích a datových strukturách. Zajišťuje podle jednotlivých úloh řešení prediktivní analytiky tyto ***činnosti***:

- zajišťuje ***analýzu datových struktur zdrojových systémů***, jejich kvalitu a dostupnost,
- spolupracuje na ***vyhodnocení stavu a kvality dat*** na formulaci principů řízení jejich kvality,
- konzultuje a posuzuje možné ***varianty přístupů k řešení*** prediktivní analytiky z pohledu datové architektury,
- spolupracuje na ***specifikaci nástrojů pro transformace dat***,
- spolupracuje s doménovým expertem na ***určení oblastí***, jež budou v rámci datových zdrojů ***prioritní,***
- definuje ***detailní pravidla a procedury transformací dat***, popis transformací polí mezi zdrojovými a cílovým systémem (datové typy, délka polí, plnění konstantami).

4.3 Expert prediktivního modelování

Expert prediktivního modelování specializovaný na vytváření modulů a využívání jednotlivých metod a algoritmů. Zajišťuje podle jednotlivých oblastí řešení tyto ***činnosti:***

- ***konzultuje s doménovými experty*** jejich problémy a požadavky na aplikace pro podporu prognózování a plánování ve firmě,
- kooperuje na ***analýzách požadavků*** na prediktivní modely a na jejich konsolidaci,
- posuzuje ***kvalitu zdrojových databází***,
- specifikuje rámcový ***obsah a strukturu výstupů***, cílových proměnných, jejich hodnot a vizualizace,
- definuje základní ***funkcionalitu*** prediktivních modelů,
- zajišťuje ***analýzu současného stavu*** disponibilních zdrojových databází a aplikací,
- řeší ***analytické otázky transformací dat***, tj. vlastní transformace dat,
- ***navrhuje kontrolní a opravné procedury*** v souvislosti s čištěním, resp. zajištěním kvality dat,
- podílí se na ***řešení problémů nebo chyb*** vyplývajících s přípravou provozu prediktivní analytiky.

4.4 Závěry



Z kapitoly vyplývají následující ***závěry***:

- Řešení prediktivní analytiky je charakteristické ***vysokými nároky na kooperaci doménových expertů datových expertů i expertů prediktivního modelování***. S rostoucí komplexností a složitostí prediktivních systémů tyto nároky dále rostou. Proto ***vymezení struktury a náplně jednotlivých rolí*** je z tohoto pohledu velmi významná.
- Výše uvedené role tvoří pouze ***podmnožinu***, které se úzce váže na jednotlivé součásti řešení prediktivní analytiky. Je nezbytné pak ***brát v úvahu i další role, zejména manažerské***, které jsou v základní rovině vymezeny v dokumentu „*AF II.02: Komponenty*“ v kapitole 5.
- Náplň jednotlivých rolí je kromě zmíněné kooperace, účelná i pro systematické plánování a ***řízení kvalifikačních programů*** pro prediktivní analytiku.
- Náplň jednotlivých rolí je nezbytné ***modifikovat podle potřeb*** a možností firmy a zejména struktury jejích personálních zdrojů.

5. Faktory prediktivní analytiky



Faktor představuje podstatnou součást vymezení prostředí, v němž se mají realizovat úlohy prediktivní analytiky. To znamená, že jde o podmínky, které ovlivňují jak řešení a využití prediktivní analytiky v reálné praxi.

Účelem kapitoly je:

- ***vymezit*** podstatu faktorů a uplatnění v prediktivní analytice,
- ***kategorizovat*** faktory, resp. najít jejich členění odpovídající potřebám řešení prediktivní analytiky,
- formulovat ***obsah*** nevýznamnějších faktorů a jejich vliv.

Faktory prediktivní analytiky jsou svým zaměřením i strukturou obdobné těm, které jsou definovány pro podnikovou analytiku. Na druhé straně mají svá specifika vycházející z podstaty prediktivní analytiky. Při řešení úloh prediktivní analytiky jsou významné „předpoklady“, za nichž řešení probíhá (viz kapitola 3.2 a 3.4). Pro určení a naplnění těchto předpokladů může být právě sada faktorů podstatným vstupem.

Kromě toho každý z faktorů představuje vymezení konkrétního prostředí, kde se analytické úlohy realizují, a to se musí ve většině případů důsledně respektovat, v opačném případě je možné připravovat kvalitní prediktivní modely, ale které nikdo nechce nebo nepotřebuje. Jak je rozmanité prostředí, tak jsou poměrně rozmanité i sady faktorů. Je proto účelné je na této obecné úrovni ***klasifikovat*** tak, aby analytici se v nich mohli rychle orientovat a vyhodnocovat i to do jaké míry jim mají věnovat pozornost.

Další podkapitoly obsahují specifikace vybraných faktorů rozdělených do těchto skupin:

- ***byznys prostředí*** – vyjadřuje vnější podmínky a vlivy pro uplatnění prediktivní analytiky,
- ***řízení a organizace firmy*** představují vnitřní podmínky a kvalitu řízení firmy a její organizační uspořádání,
- ***řízení prediktivní analytiky*** zahrnuje podmínky řízení spojené s rozvojem a provozem úloh prediktivní analytiky,
- ***kvalita informačního systému,*** jeho služeb a zdrojů***.***

5.1 Byznys prostředí

Skupina faktorů označených jako „*byznys prostředí*“ vyjadřuje převážně ***vnější podmínky*** pro uplatnění prediktivní analytiky, zahrnuje:

- velikost firmy,
- původ a vlastnictví firmy,
- konkurenční prostředí,
- odvětví působnosti firmy.

5.1.1 Velikost firmy

Jak bylo již uvedeno na začátku textu, v minulosti byla řešení prediktivní analytiky záležitostí převážně velkých, v současnosti se ***posouvá i do středních a menších firem*** a uplatňuje se prakticky na všech úrovních jejich velikosti.

Je nutné ale zdůraznit, že ve velkých firmách jsou obchodní, ekonomické i další aktivity většinou podstatně komplexnější a složitější, a tedy i jejich prognózování a plánování je náročnější vyvolávající tak i i silnější potřebu prediktivních úloh. Na druhé straně disponují obvykle dostatečnými ***finančními a personálním zdroji*** pro jejich realizaci. Jsou zde většinou ***složitější zdrojové systémy***, a tedy i náročnější transformace dat do datových souborů prediktivní analytiky.

5.1.2 Původ a vlastnictví firmy

Vlastnictví firmy je jeden z faktorů, který ***vyjadřuje formy a složitost vlastnických vztahů*** a v tomto případě zejména také to, zda je firma v českém vlastnictví, nebo je pobočkou nějaké nadnárodní společnosti, nebo je výlučně zahraniční firmou působící v českém prostředí a na českých trzích. Je zřejmé, že firmy se složitějšími vlastnickými vztahy budou znamenat větší nároky na řešení prediktivních úloh.

Jiným problémem je i to, zda zahraniční nebo ***nadnárodní firma má zájem na řešení***, a tedy i financování těchto úloh. Případně jaké jsou záměry vedení firmy vzhledem k rozvoji řízení místní pobočky včetně prediktivní analytiky.

5.1.3 Konkurenční prostředí

Potřeby řešení prediktivní analytiky jsou výrazně ovlivněny i ***tržním, zejména konkurenčním prostředím***, kde firma působí. V segmentech trhu s relativně nízkým konkurenčním tlakem, např. v hutnictví nebo těžkém strojírenství a případně ve veřejné správě je potřeba těchto projektů většinou omezena:

- Konkurence, zákazníci, dodavatelé a další partneři, jejich hodnocení, jejich očekávaný vývoj, jejich ***nároky a očekávání se stávají velmi významnou součástí podnikové analytiky*** a obvykle tvoří také jádro prediktivních aplikací zaměřených na strategické řízení, marketing, nebo prodej.
- Síla konkurence je faktor, který, zejména u velkých firem, ***posiluje potřeby investic zejména do pokročilé analytiky,*** jako např. prediktivní analytiky.
- Faktor konkurenčního prostředí znamená ***výraznější potřebu a uplatnění externích datových zdrojů*** a s tím spojené nezbytné aktivity, jako např. jejich kvalifikované hodnocení kvality, dostupnosti, finanční náročnosti.

5.1.4 Odvětví působnosti firmy

Společné charakteristiky ve vztahu k odvětví, kde firma působí jsou následující:

- Odvětví působnosti firmy je ***významné zejména pro obsahové zaměření prognózování a plánovacích úloh,*** a tedy i prediktivní analytiky, s ohledem na náročnost a složitost jejich funkcionality.
- Firmy v různých odvětvích ekonomiky svojí složitostí řízení, tlakem na efektivitu, rozsahem různorodých datových zdrojů, již existující IT infrastrukturou ***vytvářejí jak různé předpoklady pro rozvoj a uplatňování analytiky, tak i vyšší či nižší objektivní náročností na funkcionalitu*** a technologickou úroveň.

5.2 Řízení a organizace firmy

Skupina faktorů označených jako „*řízení a organizace firmy*“ vyjadřuje převážně ***vnitřní podmínky řízení*** pro uplatnění prediktivní analytiky ve firmě, patří sem:

- firemní kultura,
- organizace firmy,
- existence business modelu,
- vztah podniku ke svým obchodním a dalším partnerům,
- personální zdroje.

5.2.1 Firemní kultura

Zřejmě výchozím a nejpodstatnějším faktorem je ***potřeba*** kvalitního prognózování a plánování (viz kapitola 3.1) a samotný ***zájem vedení společnosti*** o úlohy a řešení prediktivní analytiky. Ty jsou určeny primárně ***pro manažersky nebo analyticky orientované uživatele***. Z toho vyplývá, že kvalita jejich řešení a zejména užití není primárně dána předpisy, metodikami, resp. disciplínou pracovníků, ***ale zájmem, motivací a invencí*** na straně zejména managementu.

Kultura firmy, tj. ***systém hodnot, které podnik vyznává***, zaběhnutá schémata jednání a rozhodování atd., má vliv na styl řízení a na úroveň detailu, na jakém se otázky v podniku řeší:

- Úroveň kultury firmy má také ***vliv na iniciativu a kreativitu pracovníků*****,** která je pro účast na řešení úloh prediktivní analytiky často rozhodující.
- Vysoká podniková kultura umožňuje obvykle ***méně problémů se zaváděním*** prediktivní analytiky ve větším rozsahu.

5.2.2 Organizace firmy

Organizace firmy je prostředí pro racionální kooperaci pracovníků a pracovních týmů, výrazem efektivní dělby práce. ***Organizační struktura*** řeší problém ***přijatelného rozpětí řízení***, tj. počtu pracovníků, který je schopen daný řídící pracovník efektivně řídit:

- Jasně definovaná a dokumentovaná podniková organizace ***přispívá k efektivnosti řízení*** a k uplatnění prediktivní analytiky.
- Řešení organizace ***navazuje na strategii firmy***, tedy i záměry nebo překážky v řešení prediktivní analytiky.
- Podniková organizace má být natolik ***flexibilní, aby byla schopna rychle reagovat na vývoj*** podnikatelského prostředí a vytvářet prostor pro řešení operativní prediktivní analytiky.
- Podniková organizace ***má respektovat i prostor uplatnění prediktivní analytiky***.

5.2.3 Business model

Business model je ***forma a přístup pro kvalitní pochopení základního fungování firmy***.

- Podporuje ***uvědomění si souvislostí jednotlivých částí a aspektů firmy*** a souvislostí v prognózování a plánovacích úlohách.
- Umožňuje ***kvalifikovanou aplikaci tohoto přístupu při řešení strategických plánovacích úloh*** v řízení firmy.
- Vyžaduje ale ***motivaci a zájem ze strany vedení firmy*** pro uplatnění business modelu v analytice a plánování.

5.2.4 Vztah podniku ke svým obchodním a dalším partnerům

Úlohy prediktivní analytiky se silně váží i na hodnocení a plánování aktivit ve vztahu k externím partnerům. Nabídka této prediktivní funkcionality s pochopením a respektováním potřeb i externích subjektů může ovlivnit výsledný obchodní a ekonomický úspěch firmy.

5.2.5 Personální zdroje, úroveň jejich znalostí

Úlohy prediktivní analytiky ***předpokládají velmi intenzivní a kvalifikovanou kooperaci*** všech tří zmíněných expertů, případně i dalších lidí zúčastněných na těchto úlohách. Čím kvalifikovanější a motivovanější budou tito pracovníci, tím lze očekávat kvalitnější výsledky. A to zde platí v podstatně větší míře než u jiných typů aplikací.

5.3 Řízení prediktivní analytiky

Skupina faktorů označených jako „*řízení prediktivní analytiky*“ vyjadřuje převážně ***podmínky řízení*** vyvolané charakterem a nároky ve firmě. Řízení prediktivní analytiky je v tomto textu vázáno na metodiku CRISP-DM a celý postup je obsažen v oddíle B). Dále uvedené faktory tak vycházejí i z uvedené metodiky ale i pokrývají některé aspekty za jejími hranicemi. ***Zahrnují***:

- strategie řešení prediktivní analytiky,
- identifikace potenciálních efektů prediktivní analytiky,
- existence silného sponzora.

5.3.1 Strategie řešení prediktivní analytiky

Strategické otázky řešení prediktivní analytiky se promítají, (obdobně jako u jiných nástrojů a aplikací) na nejvyšší úrovni ***do informační strategie*** firmy vztahující se k informatice firmy jako celku a na nižší úrovni do strategie jednotlivých projektů. Při převládajícím ***značném rozsahu*** těchto projektů, jejich ***finanční a časové náročnosti***, mimořádně rychlém ***rozvoji technologií***, na nichž jsou založeny, je určení správné strategie odpovídající potřebám a možnostem podniku ***velmi podstatným faktorem*** řešení. Stanovení strategie by tak mělo zohlednit i vyhodnocení všech dílčích faktorů, a to zejména:

- určení ***priorit*** řešení ve vztahu k oblastem řízení, resp. podnikovým procesům,
- určení ***přístupu*** k řešení a využití různých disponibilních funkcí a metod (kapitola 23),
- správný výběr technologií a produktů.

5.3.2 Identifikace potenciálních efektů prediktivní analytiky

K tomu, aby ***potřeba úloh*** prediktivní analytiky byla kvalifikovaně posuzována, je nezbytné ***formulovat jejich potenciální efekty*** s ohledem na danou situaci firmy. Určování a ***posuzování efektů*** prediktivní analytiky je v porovnání s ostatními typy aplikací poněkud ***specifické***. V každém případě je vymezení očekávaných efektů a sledování jejich naplnění podstatné s ohledem na to, že analytické přípravě a využití těchto aplikací musí ***věnovat čas*** již zmínění experti a případně další specialisté, jejichž časové možnosti jsou vesměs omezené. Musí proto, pokud možno, přesně vědět, co jim takto vynaložená časová i finanční investice přinese.

Na druhé straně však efekty prediktivní analytiky ***nemusí*** být vždy zcela ***přesně kvantifikovatelné***, resp. ve finančním vyjádření. Často se v těchto případech jedná o ***kvalitativní efekty*** znamenající dosažení vyšší konkurenceschopnosti firmy, získání lepší pozice na trhu, poskytování kvalitnějších informačních služeb obchodním partnerům atd. (viz kapitola 3). Je dobré si v tomto kontextu i položit ***otázku*** ***„****jaký bude mít dopad na podnik situace, kdy nebude investovat do prediktivní analytiky, zatímco konkurence ano****?“***.

5.3.3 Existence silného sponzora

Pravidlo ***silného sponzora*** je v oblasti podnikové analytiky a prediktivní analytiky rovněž již všeobecně známé. S ohledem na konečný úspěch by tyto projekty měly být uvnitř firmy vždy podporovány ***osobností se značnou mírou vlivu a s nezbytnými rozhodovacími pravomocemi***. Jde o osobnost, která navíc vedle svého vlivu a presentovaného zájmu o řešení, je schopna vidět podnik a jeho aktivity ve všech podstatných souvislostech, ve vztahu k podnikovému okolí, je schopna formulovat a rozhodovat o klíčových prioritách řešení a samozřejmě je schopna řešit finanční zajištění projektu a dalšího provozu.

5.4 Kvalita informačního systému

Skupina faktorů spojených s informačním systémem firmy představuje zejména ***hodnocení kvality jeho zdrojů***, resp. z čeho může řešení analytiky vycházet, tj. zejména úrovně řešení kvality dat a kvality podnikových aplikací. ***Spadají sem tyto faktory***:

- kvalita ***datových zdrojů***, včetně dostupnosti dokumentace datových zdrojů,
- kvalita ***provozovaných aplikací*** ve firmě.

5.4.1 Kvalita datových zdrojů

Význam kvality datových zdrojů pro řešení prediktivní i celé podnikové analytiky byl již několikrát zdůrazněn. Do této kapitoly je tento faktor zahrnut pouze jako konstatování jeho klíčového významu pro úspěšnost řešení a užití prediktivních úloh. ***Dílčí faktory*** ovlivňující datovou kvalitu lze vymezit v následujících ***třech skupinách***:

- ***technické prostředí*** zahrnující celopodnikový slovník dat, centralizaci aplikací a jejich datových zdrojů, např. jednotná identifikace zákazníků, kontroly definovaných business pravidel,
- úroveň použité ***metodiky***, tj. metodiky a směrnice definované v rámci firmy, kvalita číselníků a kódových tabulek, systém řízení změn,
- ***přístupy k řešení*** informačního systému, tj. způsob přípravy dat, přípravy uživatelů, systém motivačních kritérií.

S tím souvisí ještě další podstatné aspekty, a to ***dostupnost dokumentace datových zdrojů***, případně možnost poskytování potřebných dat řešitelům prediktivní analytiky. V některých případech ***není zcela jednoduché tyto dokumentace, resp. data získat*** s ohledem na autorská práva nebo smlouvy mezi zákazníkem a poskytovateli těchto primárních systémů. V každém případě je dobré si tyto podmínky a možnosti ještě před zahájením projektu ověřit a podle možností je začít řešit.

5.4.2 Kvalita podnikových aplikací

Otázka kvality provozovaných aplikací, zejména transakčních, je v souvislosti s úlohami prediktivní analytiky ***posuzována v několika úhlech pohledu***:

- do jaké míry jsou tyto aplikace ***schopné poskytovat úplná, konsistentní a přesná data***, tj. jak je navržena jejich ***datová základna***, jaký ***systém kontrol*** zahrnuje jejich ***funkcionalita***, jak odpovídají potřebám firmy z pohledu poskytovaných funkcí i vytvářených a zpracovávaných dat,
- zda zahrnují ***vlastní plánovací funkcionalitu***, do jaké míry je využívána, jak může posloužit i jako vstup pro řešení prediktivní analytiky.

5.5 Závěry



- Řešení prediktivní analytiky je charakteristické ***širokým spektrem faktorů,*** které úspěšnost jejích výsledných řešení ovlivňují.
- Je tedy nezbytné již na počátku řešení tyto faktory v konkrétním firemním prostředí ***kvalifikovaně vyhodnotit*** a ***vyhnout se*** tak následným ***nepříjemným zjištěním***, že vykonávaná práce se úplně nepromítala do potřeb a možností firmy, případně byla zbytečná.
- Sady faktorů se ***v různých odvětvích a firmách různí***. Je ale účelné vycházet z obecněji pojatého přehledu a současně i jejich ***základní klasifikace***. Na tomto místě byly zvoleny ***tyto skupiny***:
  - faktory byznys prostředí,
  - faktory kvality řízení a organizace firmy,
  - přístupy a kvalita řízení prediktivní analytiky,
  - podstatné aspekty kvality informačního systému firmy, zejména datových zdrojů a aplikací.

6. Scénáře, analytické otázky



Při řešení jakéhokoli projektu, u prediktivní analytiky včetně, je vždy dobré si postavit sady analytických otázek, resp. scénářů, které specifikují aktuální problémy a na které se pak v průběhu řešení hledají odpovědi.

Účelem kapitoly je:

- vymezit princip takových analytických otázek,
- seskupit otázky do skupin, scénářů jako vstupy do jednotlivých částí řešení úloh prediktivní analytiky,
- navrhnout obsah analytických otázek v následujících scénářích, a to:
  - jak přistupovat k řešení plánovacích úloh z pohledu potřeb byznysu,
  - základní otázky prediktivní analytiky ve vztahu k obsahu byznysu, podle (Wilson, 2021),
  - jak řešit a zajistit datové zdroje pro prediktivní analytiku,
  - jak zajistit personální zdroje pro prediktivní analytiku,
  - jaké metody volit pro prediktivní analytiku.

Další podkapitoly obsahují souhrnné přehledy navrhovaných otázek, které jsou pak dále předmětem určité konkretizace podle jednotlivých oblastí řízení firmy.

6.1 Scénář 1: Jak řešit plánovací úlohy vzhledem k hlavním potřebám byznysu

- Jak zvýšit ***úspěšnost a výkonnost*** byznysu díky kvalitě a využití plánovacích úloh?
- Jak specifikovat ***potenciální efekty*** užití plánovacích úloh, a to ***ekonomické i mimoekonomické***?
- Které ***oblasti řízení*** jsou a budou z hlediska uplatnění plánovacích úloh ***prioritní***?
- Které ***metriky*** budou pro plánovací úlohy významné, které budou mít charakter KPI, které ***dimenze*** ve vztahu k metrikám budou relevantní?
- Které ***faktory*** budou pro řešení a uplatnění plánovacích úloh ***rozhodující***?
- Jak zajistit požadovanou ***komplexnost a kvalitu*** plánovacích úloh?
- Jak dosáhnout požadované ***flexibility*** plánovacích úloh vzhledem k momentálním potřebám manažerů a specialistů firmy?
- Jak zajistit kvalitní a systematické ***shromažďování podkladů*** z jednotlivých útvarů firmy pro přípravu plánů?
- Jak nastavit ***efektivní postupy projednávání*** a schvalování připravovaných plánů?
- Jaké ***vlivy*** má úroveň a využití plánovacích úloh ***na řízení a organizaci firmy***, jaké jsou hlavní problémy nebo omezení?

6.2 Scénář 2: Jak řešit prediktivní analytiku vzhledem k obsahu byznysu

Otázky, které lze efektivně řešit s pomocí prediktivní analytiky, jsou zde formulovány podle (Wilson, 2021):

- Jak a jaké ***služby*** zajišťovat ***zákazníkům***?
- Jak dosáhnout požadovanou úroveň „***OTIF, on-time in-full delivery***“?
- Co je potřeba udržovat ***v zásobách*** a potřebných zdrojích?
- Co budou ***zákazníci požadovat***?

- Jak lépe prezentovat vlastní ***nabídku***?
- Jak lze ***zákazníky udržovat***? Jak řešit „*Churn Analysis*“ – hodnocení zákazníků firmy z pohledu rizika, že je firma ztratí?
- Co si ***zákazníci skutečně myslí*** o zboží a službách dané firmy?
- Jak je možné ***zachovávat nebo opustit určité položky zboží***, které charakteristiky zboží z hlediska prodeje nefungují?
- Jak využívat „***Conjoint Analysis***“ – statistické vyhodnocování průzkumů a výzkumů trhu se zaměřením na zjišťování hodnoty, kterou zákazníci přiřazují jednotlivým vlastnostem zboží nebo služeb?
- Do jaké míry má firma ***snižovat ceny***, jak využívat ***cenovou dynamiku*** a elasticitu?
- Na koho je třeba zaměřovat ***marketingové aktivity***?
- Jak využívat „***MBA, Market Basket Analysis***“ – která napomáhá firmě uplatňovat „cross-sell“ a „up-sell“ přístupy?
- Odkud ***zákazníci firmy pocházejí***, jaké jsou demografické, geografické, socioekonomické a další atributy zákaznického portfolia?
- Jak využívat „***SNA, Social Network Analysis***“ – mapování a zjišťování vazeb mezi různými atributy, místy, a informačními entitami k hodnocení sociálních struktur?
- Jaké ***množství nových produktů*** bude firma prodávat?
- Jak řešit ***prognózy jednotlivých nových produktů***?
- Jaká je ***hodnota jednotlivých zákazníků***? Jak využívat „***CLV, Customer Lifetime Value***“ – představuje hodnotu zákazníka pro firmu po celou dobu jeho vztahu s firmou?
- ***Co se má předpovídat*** – akce, chování, událost? (Siegel, 2013)
- Jak budou ***predikce využity***? Je známé rozhodnutí nebo akce, uskutečněné na základě predikce? (Siegel, 2013)

6.3 Scénář 3: Jak řešit a zajistit datové zdroje pro prediktivní analytiku

- Jaký je aktuální ***obsah datových zdrojů*** vzhledem k potřebám prediktivní analytiky?
- Jaká je ***kvalita datových zdrojů*** vzhledem k prediktivní analytice?
- Jak dosáhnout požadovanou ***granularitu dat*** v rámci jednotlivých operací prediktivních úloh?
- Jak zajistit ***potřebné datové zdroje*** na požadované úrovni granularity?

6.4 Scénář 4: Jak zajistit personální zdroje pro prediktivní analytiku

- Jak dosáhnout potřebné ***kvalifikace a motivace*** expertů, manažerů a specialistů v oblasti využití a případně i řešení prediktivní analytiky?
- Jak ***připravovat pracovníky*** firmy na uplatnění nových řešení v oblasti prediktivní analytiky?
- Jak nastavit efektivní ***kooperační schémata*** mezi jednotlivými typy expertů?

6.5 Scénář 5: Jaké metody volit pro prediktivní analytiku

- Jaké ***metody a funkce zvolit*** pro potřeby modelování v rámci prediktivní analytiky?
- Jak nastavit efektivní vazby k ***ostatním disciplínám podnikové analytiky,*** např. business intelligence, data mining, text mining a další?
- Jaká je k dispozici ***podpora*** pro využití jednotlivých metod a případně souvisejících disciplín?
- Jak dosahovat ***efektivní doby a časové náročnosti*** při přípravě modelů prediktivní analytiky?

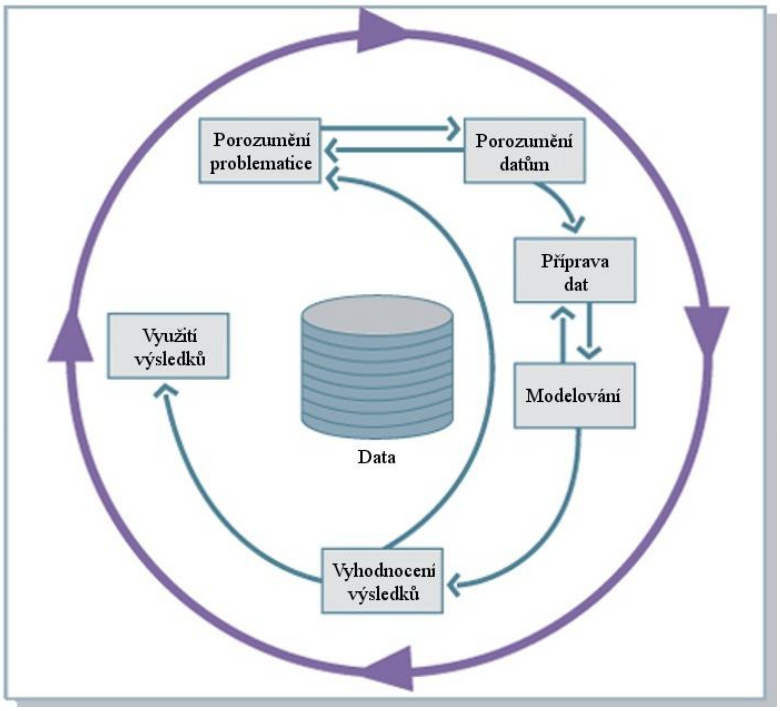
6.6 Závěry

	Z kapitoly vyplývají následující ***závěry***:

- Analytické otázky jsou de facto ***formulací problémů*** nebo potenciálních problémů v řízení firmy, které mají předmětem řešení prediktivní analytiky.
- Analytické otázky se podle zaměření ***seskupují do skupin***, které jsou zde pracovně chápany jako ***scénáře***.
- Scénáře jsou jedním z možných ***vstupů do diskusí a kooperací*** mezi experty i managementem firmy při řešení úloh prediktivní analytiky.
- Scénáře a analytické otázky je nezbytné ***před jednáním konkretizovat*** podle podmínek a potřeb dané firmy.
- Do otázek scénářů je potřeba i promítat ***specifika použitých metod*** a přístupů k řešení modelů v rámci prediktivní analytiky.

Další kapitoly představují ***oddíl B*** dokumentu zaměřený na jednotlivé ***fáze postupu řešení*** prediktivní analytiky na bázi metodiky CRISP-DM.

B) Postupy řešení a organizace prediktivní analytiky



Obrázek: Fáze procesního modelu CRISP-DM (CRISP-DM), (CRISP-DM, n.d.) (Berka, 2003, stránky 24-28)

[7] Pochopení byznysu	**[8] Porozumění datům**
[9] Příprava dat	**[10] Modelování**
[11] Vyhodnocení modelů	**[12] Využití výsledků**

Nejčastěji užívanou metodiku je CRISP-DM již od roku 1990, která se primárně váže k projektům data miningu. Je to dáno tím, že prediktivní analytika a data mining mají k sobě velmi blízko, používají i obdobných algoritmů. Prediktivní analytika tvoří jistou nadstavbu nad data miningem.

CRISP-DM značí „***Cross-Industry Standard Process for Data Mining***“ (CRISP-DM, n.d.). ***Cílem*** je vytvoření standardního ***procesního modelu, bez vazby na konkrétní prostředí***, zdarma a veřejně dostupného, se zacílením na praktickou aplikaci prediktivní analytiky (povětšinou komerční). Obsahuje nejen jednotlivé kroky a postupy, ale i nezbytnou dokumentaci, která má být s projektem k dispozici. Metodika je užitečná pro analytiky i proto, že jim poskytuje argumentaci pro manažery, co v rámci projektu dělají a proč to dělají.

Návraty na předchozí fáze jsou dány např. nedostatečnými daty a jejich kvality pro vytváření modelů, nebo při vyhodnocení modelů se dojde k tomu, že výsledky nejsou adekvátní předpokladům apod.

CRISP-DM zahrnuje přímo kroky věnující se ***formulaci zadání společně s pochopením problematiky a finálnímu využití výsledku***.

7. Pochopení byznysu („Business Understanding“)



Nejdůležitější částí každého projektu prediktivní analytiky je úvodní část ***vymezující celý projekt***, kde hlavním předmětem zájmu ve fázi je vymezení a pochopení byznysu dané firmy a analýza potřeb v oblasti prediktivní analytiky.

Účelem fáze je:

- specifikovat hlavní ***požadavky*** na prediktivní modely,
- definovat ***potřeby a cíle byznysu*** v souvislostech plánování a prognózování tak, aby bylo zřejmé, jaké jsou důvody pro řešení prediktivní analytiky,
- definovat ***základní charakteristiky dat*** vzhledem k potřebám analytiky, zejména cílové proměnné, jednotky analýzy a další,
- formulovat ***zadání*** vhodné pro úlohy prediktivní analytiky.

Na úvod jde o velmi náročnou fázi vyžadující ***znalosti dané domény byznysu***, ***znalosti dat*** a databází a ***znalosti prediktivních modelů***. Z tohoto důvodu musí být většinou ***řešena týmově***. Klíčová role je zde přisuzována praktikům.

7.1 Definování vstupních požadavků na prediktivní modely

V této úvodní fázi procesu prediktivní analytiky se analyzují všechny ***problémy a požadavky na řešení*** podle konkrétních podmínek firmy. ***Cílem*** je přesně určit a poznat uživatele a účel prediktivního modelu. Musí být známo, ***k čemu model bude sloužit***, na co bude aplikován a jak svojí predikcí pomůže. Určí se odpovědi na ***dvě základní otázky*** (SIEGEL, 2013):

- Co má být predikováno?
- Jak bude s predikcí naloženo?

Definují se ***metriky vyhodnocení modelu a postup řešení*** a implementace prediktivního modelu.

7.2 Byznys cíle

Fáze spočívá v ***identifikaci cílů*** úlohy, ***porozumění zadání*** formulovaného manažery (vycházející z business cílů). Na identifikaci cílů se podílejí všichni výše uvedení 3 experti. Prediktivní modelování pokrývá široký rozsah byznys cílů. ***Existuje 6 klíčových problémů***, které je třeba ve fázi „*Pochopení byznysu*“ řešit:

- Jaké jsou hlavní („*core*“) byznys cíle, na které se mají zaměřit prediktivní modely?
- Jak mohou být byznys cíle kvantifikovány?
- Jaká data jsou k dispozici pro kvantifikaci byznys cílů?
- Jaké metody modelování mohou být využity pro popis nebo predikci byznys cílů?
- Jak lze kvantifikovat správnost modelu vzhledem k byznys cílům, jak výsledky modelu odpovídají potřebám byznysu?
- Jak mohou být prediktivní modely distribuovány a dodávány uživatelům?

Byznys cíle vycházejí primárně z kvalifikovaného pochopení obsahu řízení firmy a nároků na analytiku podle jednotlivých oblastí řízení. Základní orientaci poskytují jednotlivé podkapitoly oddílu C).

Příklady úloh, resp. projektů prediktivní analytiky podle (Abbott, 2014, upraveno) nabízí další přehled:

- Získávání nových zákazníků.
- Prognózy obchodních příležitostí.
- Cross-sell / Up-sell úlohy.
- Produkty, které zákazníci budou napříště kupovat.
- Navštěvování webových stránek.
- „*Market-basket*“ analýza.

- Hodnota zákazníka.
- Profitabilita zákazníka.
- Segmentace zákazníků.
- A další.

7.3 Návrh dat pro prediktivní modelování

Data pro prediktivní modelování ***musí být dvou dimenzionální*** představující řádky a sloupce. Každá ***řádka znamená jednotku analýzy*** („*unit of analysis*“), např. pro analýzu zákazníků je to zákazní, pro analýzu podvodů je to transakce, pro analýzu call center je to jeden individuální hovor, pro analýzu průzkumů je to jeden průzkum (Abbott, 2014). Sloupce v tabulce se označují různě: atributy, deskriptory, proměnné („*variables*“), pole („fields“), vlastnosti („features“) nebo jednoduše sloupce („*columns*“). Proměnné se chápou jako míry (ukazatelé), které se váží k záznamu. Atributy jsou např. id. zákazníka, věk, profese apod.

Pokud jde o formát dat, není významný, protože software pro analýzy umožňuje většinou zpracovávat různé formáty, např. flat files, soubory s pevnou strukturou záznamu, binární soubory ze specializovaných software (SAS, SPSS a další).

Většina analytických software ***nabízí přímou konektivitu do databází***, např. pomocí ODBC driverů a umožňují zpřístupňovat data přímo na základě dotazů nebo pohledů z databází. To má řadu výhod:

- Data nemusí být ukládána na disk a do analytických balíků.
- Data jsou automaticky spravována a udržována v databázi.
- Analytik má lepší přehled o dostupných datech, které může z databáze získávat.

7.4 Definování jednotek analýzy

Prediktivní modelování předpokládá, že každý ***záznam*** (řádka) je ***nezávislé pozorování.*** To znamená, že algoritmy neuvažují přímé vazby mezi záznamy. ***Není ale vždy jasné***, které jednotky analýzy použít. Obvyklým příkladem řešení jednotky analýzy je ***osoba: zákazník***.

Pokud ale např. zákazník – host ***navštíví hotel vícekrát***, pak každá návštěva bude jednotka analýzy. Záleží na požadovaném detailu analýzy. V některých případech se vytvářejí ***variantní modely*** podle jednotek analýzy (zákazník, návštěva, sumarizace návštěv), podle potřeb rozhodování a byznysu.

7.5 Definování cílových proměnných

Důležitá součást fáze je definování ***jedné nebo více cílových proměnných***. Cílová proměnná je sloupec tabulky, jehož ***hodnota je predikována nebo hodnocena*** v prediktivním modelu na základě stanovených byznys cílů. Může být kontinuální ***numerická nebo kategoriální*** (případně binární), na základě typu modelu. ***Příklady*** mohou být následující:

- Objem odměn pracovníka, hodnota zákazníka, počet prodejů produktu apod.: kontinuální numerická.
- Získání pracovníka, ztráta zákazníka apod.: binární
- Podvod při nákupu služby podle typů podvodu: kategoriální.

7.6 Závěry



Z kapitoly vyplývají následující ***závěry***:

- Základem je ***identifikace cílů*** úlohy, ***porozumění zadání*** formulovaného manažery (vycházející z business cílů).
- Identifikace cílů se řeší ***týmově***, podílejí se na ni všichni ***3 experti***, tedy doménový expert, datový expert i expert pro tvorbu modelů.
- ***Existuje 6 klíčových problémů*** „*Pochopení byznysu*“, jedním z hlavních je „*Jaké jsou hlavní („core“) byznys cíle, na které se mají zaměřit prediktivní modely?“*

- Na úvod se analyzují všechny ***problémy a požadavky na řešení*** podle konkrétních podmínek firmy.
- Data pro prediktivní modelování musí být ***dvoudimenzionální – řádky, sloupce***.
- Každá ***řádka*** reprezentuje ***jednotku analýzy*** („*unit of analysis*“).
- Sloupce v tabulkách se často označují jako ***atributy, deskriptory, proměnné, pole, vlastnosti***.
- ***Proměnné*** se chápou jako ***ukazatelé (measures)****,* které se vztahují nebo popisují záznam. Podstatnou součástí je definování ***jedné nebo více cílových proměnných***.
- Cílová proměnná je taková, jejíž ***hodnota je predikována nebo hodnocena*** v prediktivním modelu.
- Prediktivní modelování předpokládá, že každý ***záznam*** (řádka) je ***nezávislé pozorování.*** To znamená, že algoritmy neuvažují přímé vazby mezi záznamy.

8. Porozumění datům („Data Understanding“)



Fáze se zabývá ***získáním dat*** a jejich následnou ***analýzou.***

Účelem fáze je:

- zajistit identifikaci a ***vyhodnocení datových zdrojů*** použitelných pro úlohy prediktivní analytiky,
- analyzovat jak ***interní***, tak ***externí*** datové zdroje,
- připravit a realizovat potřebné aktivity v souvislosti se zajištěním ***potřebné kvality dat***.

V návaznosti na pochopení byznysu a jeho potřeb z předchozí fáze se řeší ***problematika „pochopení dat“*** a datových zdrojů, na který je řízení firmy založeno a současně i ***identifikace hlavních problémů v datech***, které je třeba v souvislosti s přípravou prediktivních modelů řešit.

8.1 Definování dat pro prediktivní modelování

Platí zde tyto ***základní principy*** (Abbott, 2014):

- Data pro prediktivní modelování musí být ***dvoudimenzionální – řádky, sloupce***.
- Každá ***řádka*** reprezentuje ***jednotku analýzy*** („*unit of analysis*“). Např. pro zákaznickou analytiku je to *zákazník*. Pro analytiku call center to může být jeden *hovor*. Pro analytiku podvodů jedna transakce. Jednotky analýzy jsou úzce vázány na byznys obsah prediktivní analytiky.
- Sloupce v tabulkách se často označují jako ***atributy, deskriptory, proměnné, pole, vlastnosti***.
- ***Proměnné*** se chápou jako ***ukazatelé (measures)***, které se vztahují nebo popisují záznam. Prediktivní modelování považuje sloupec za separátní ukazatel, pořadí struktura musí být standardní.
- Data jsou obvykle získávána ***ze souborů, různých formátů***, např.:
  - „*Flat files*“ (csv soubory), položky mohou být odděleny čárkou, specifickým znakem, tabulátorem.
  - „*Flat files* s pevnou délkou záznamu“ s pevným počtem znaků na položku.
  - Ostatní kastomizované „*Flat files*“.
  - Binární soubory na bázi specifických formátů odpovídajících softwarovým balíkům (SPSS, SAS, Matlab,…).
- Data mohou být zpřístupněné ***přímo z databází*** (tabulky, views) na bázi nativních driverů nebo ODBC. Některé software umožňují komplexní dotazy pro získávání data pro predikce, což umožňuje:
  - data nemusí být speciálně ukládána na disky,
  - data se mohou udržovat přímo v databázi a nemusí být udržována ve flat souborech,
  - analytici mají větší kontrolu a flexibilitu ve využití dat.
- Musí být zajištěno, že tabulky a pohledy ***musí zůstat v průběhu predikcí stejné***.

Důležitá součást fáze je definování ***jedné nebo více cílových proměnných***. Cílová proměnná je taková, jejíž ***hodnota je predikována nebo hodnocena*** v prediktivním modelu a může být ***numerická nebo kategoriální*** (případně binární).

8.2 Analýza dostupných zdrojů dat

Druhá fáze procesu zahrnuje ***identifikaci*** všech relevantních potencionálně ***využitelných zdrojů dat***. ***Cílem*** je identifikovat, ***které datové zdroje*** a v jaké míře poslouží jako ***základ pro tvorbu modelu*** a zdali jsou dostupné datové zdroje dostačující pro účely prediktivního modelování. Ideální je využít všechna dostupná relevantní data.

Co nejvyšší úplnost relevantních dat zajistí přesnější odhady postavené na prediktivním modelu. Je tedy vhodné zahrnout jak data ***z interních systémů*** (upisování, marketing, platební systémy, rating) tak ***z externích systémů*** (např. dostupné demografické údaje, kreditní informace, informace o inflaci atp.). Je důležité, aby se nevynechala žádná, potencionálně relevantní a důležitá data (ROOSEVELT, 2005).

8.3 Analýza datové kvality

Analyzuje se ***datová kvalita***, identifikují se slabá místa a omezení dostupných dat. Příkladem může být ***analýza úplnosti dat, redundance*** dat. Hledá se odpověď na ***otázky***:

- Je vůbec prediktivní analýza na dostupných datech proveditelná?
- Jaké zdroje dat budou využity?
- Budou pro realizaci potřeba změny v samotných datových zdrojích?
- Jaké nástroje budou využity?

Zpravidla se jako primární zdroj dat využívá ***datový sklad podniku, či datamarty*** (datová tržiště) vyvinuté nad datovým skladem. Především proto, že tyto datové struktury svými vlastnostmi nejlépe odpovídají potřebám prediktivní analýzy – eliminují či omezují duplicity a chyby dat, sdružují většinu / všechny podnikové systémy.

Sledují se charakteristiky atributů jako četnosti hodnot, extrémní hodnoty, rozdělení četností hodnot apod. Je vhodné pro tuto činnost využívat ***vizualizačních nástrojů***. Podstatnou součástí řešení je zde ***identifikace chybějících hodnot*** a celková ***analýza kvality*** dat.

Výsledkem je základní představa o datech, která máme k dispozici.

8.4 Závěry, doporučení



Z kapitoly vyplývají následující ***závěry***:

- Úplnost relevantních dat zajistí přesnější odhady prediktivní analytiky.
- Je účelné přitom využívat jak data ***z interních*** systémů***,*** tak ***z externích*** systémů.
- Jako primární zdroj dat se často využívá ***datový sklad podniku, či*** datová tržiště.
- Klíčovou činností v rámci dané fáze je ***analýzy kvality dat*** a řešení takových otázek jako, zda budou pro potřeba změny v samotných datových zdrojích.
- Pro porozumění datům a řešení jejich kvality se ve větší míře využívají i ***vizualizační nástroje.***
- Při získávání a sběru nových dat je účelné ***identifikovat*** ty, které lze efektivněji získat ***z existujících databází*** a dostupných datových zdrojů.
- Při sběru nových dat je nejprve třeba je ***vyhodnotit z pohledu jejich důvěryhodnosti*** a využitelnosti v řešení prediktivních modelů.
- Při čištění dat je účelné se zaměřovat i na ***anomálie v datech***.
- Testování a čištění dat je dobré v průběhu řešení realizovat ***co nejdříve***.

9. Příprava dat („Data Preparation“)



Úkolem je ***připravit finální datový soubor***, který bude zpracováván konkrétními algoritmy.
Účelem fáze je:

- realizovat přípravu finálního datového souboru,
- zajistit ***kontroly, čištění dat,***
- realizovat ***doplňující výpočty.***

Příprava dat je fáze časově i pracovně mimořádně náročná a podle různých studií představuje až ***80 % celkového času řešení*** prediktivní analytiky. Znamená ***shromáždění dat, jejich kontroly, čistění a vytváření doplňujících dat*** na základě odvozovaných proměnných. Navíc obvykle platí, čím více dat se pro prediktivní modely zajišťuje, tím větší riziko chyb a nekvality je s nimi spojeno, a tedy i větší náročnost na jejich čištění.

9.1 Organizace a shromáždění dat

Jedná se většinou o ***nejpracnější část*** celého procesu prediktivní analytiky. Tato fáze zahrnuje už vlastní operace s daty nad datovým zdrojem /zdroji. ***Hlavní činností*** je především ***shlukování, třídění a řazení dat k sobě podle byznys významu a potřeb predikce***. Tato část vyžaduje výbornou technickou a business znalost dat. Je nutné, aby data byla obsáhlá a úplná, protože i zdánlivě nerelevantní proměnné a vztahy mezi nimi se mohou ukázat jako důležité.

Nerelevantní proměnné (například s nulovými hodnotami, nebo náhodně generované hodnoty a identifikátory) se poté ***vyřadí ve fázi tvorby modelu***. Data je nutné mít co nejúplnější a bezchybná, protože ***kvalita datového základu*** přímo ovlivňuje úspěšnost predikce prediktivních modelů.

9.2 Čištění dat

Čištění dat v sobě zahrnuje ***procházení záznamů a hledání chybných, neaktuálních nebo nekompletních dat***. Takové záznamy mohou mít dopad na přesnost prediktivního modelu a ideálně by měly být opraveny nebo doplněny, aby byl konečný model co nejpřesnější (NYCE, 2007). Předpokladem pro efektivní řešení kvality data je ***jasná specifikace zdrojů dat***, odkud data pocházejí a s jakými riziky je třeba vzhledem k existujícím zdrojům počítat (např. externí zdroje, zcela nové zdroje dat, struktura a kvalifikace pracovníků, pořizujících data apod.). Na specifikaci a analýzu zdrojů dat pak váže ***profilování dat***, tj. jejich struktur, vlastníků a zodpovědností, potenciálních problémů apod.

Problémy nebo chyby, které je třeba řešit jsou obvykle tyto (Wilson, 2021):

- ***chybějící data*** – naskýtají se tyto možnosti řešení:
  - vynechání celého řádku nebo sloupce,
  - doplnění předpokládaných nebo fixních hodnot,
  - doplnění průměrné hodnoty, mediánu nebo modu, podle konkrétní situace.
- ***data mimo obvyklé hodnoty („outliers“)*** – jsou to hodnoty, které jsou extrémně velké, malé nebo leží mimo obvyklé tolerance. Problém ale je v tom, že tyto hodnoty nemusí být nakonec „outliers“ a záleží na jejich kvalifikovaném prověření. Řešení mohou být:
  - eliminace takových hodnot,
  - využití průměrné hodnoty, mediánu nebo modu, nebo nahrazení obvyklou hodnotou,
  - řešení pomocí agregace a úpravy stávajících hodnot.
- ***duplicitní data*** představují častý problém a jejich řešení je následující:
  - vynechání celého řádku nebo sloupce,
  - využití klíčů pro identifikaci duplicit a jejich následné řešení,
  - využití vizualizace k určení míst duplikátů a určení jejich řešení,
- ***problémy formátování dat*** mohou být nejvíce časově náročné na jejich řešení, možnosti jsou např. následující:
  - rozdělení sloupce na více sloupců podle různých formátů dat,

- zjištění, kde jsou numerická a kategoriální data a podle to jejich rozdělení,
- přesná identifikace formátovacích problémů (např. s využitím klíčových hodnot) a jejich následné řešení.

9.3 Vytvoření odvozených proměnných

V této fázi prediktivní analytiky se dále vytváří ***odvozené vypočítané hodnoty / ukazatele*** na základě analýz, zkušeností a známých faktů z daného oboru. Tyto hodnoty mohou mít v prediktivních modelech potencionálně velice ***vysoké váhy***, proto je potřeba věnovat jim zvýšenou pozornost a konzultovat je s ostatními pracovníky.

Finální datový soubor je připravován s přihlédnutím ***ke konkrétnímu zvolenému algoritmu*** (schopnost algoritmu zacházet s chybějícími hodnotami, vyžadované datové typy na vstupu, rozsah datového souboru).

Poměrně často dochází ***k návratu do této fáze*** z kroku modelování, ať již z důvodů odhalených nepokrytých chyb v datech, či v případě nutných změn souvisejících s konkrétním algoritmem.

9.4 Závěry, doporučení



Z kapitoly vyplývají následující ***závěry***:

- Příprava dat představuje většinou ***nejpracnější část*** celého řešení prediktivní analytiky.
- Zahrnuje ***shlukování, třídění a řazení dat k sobě podle byznys významu a potřeb predikce***.
- Čištění dat v sobě zahrnuje ***procházení záznamů a hledání chybných, neaktuálních nebo nekompletních dat***.
- Obvykle se musí ***řešit tyto problémy***:
  - chybějící data,
  - data mimo obvyklé hodnoty („outliers“),
  - duplicitní data,
  - problémy formátování dat.
- Součástí této fáze je i to, že se dále vytváří ***odvozené vypočítané hodnoty / ukazatele*** na základě analýz, zkušeností a známých faktů z daného oboru.

10. Modelování („Modeling“)

V rámci fáze jsou aplikovány ***analytické metody*** (algoritmy pro úlohy prediktivní analytiky), dochází k hledání optimálních ***nastavení parametrů*** pro jednotlivé algoritmy.

Účelem fáze je:

- specifikovat ***vhodnou metodu*** pro modelování úloh prediktivní analytiky,
- vytvářet ***prediktivní model***,
- určování a vytváření ***větších celků*** pro řešení modelů.

„*Prediktivní model je kombinace metod, kde primárním cílem je predikovat pravděpodobnost kategoriálních nebo kontinuálních hodnot s využitím hodnot prediktorů*“ (Wilson, 2021). Prediktivní modelování představuje ***algoritmy typu „řízené učení“*** („*supervised learning*“), kde algoritmus se snaží nají vztahy mezi vstupy k jedné nebo více cílových proměnných. Správné modely se váží k cílovým proměnným nejlépe odpovídajících byznys potřebám.

Vzhledem k faktu, že ***neexistuje optimální metoda*** pro všechny úlohy, doporučuje se hledat vhodnou metodu a vhodné nastavení parametrů, a pro definitivní výběr nejlepší metody porovnat výsledky různých nastavení a různých metod.

Pokud jsou data připravena a zkontrolována, ***je součástí této fáze*** výběr přístupů a metod, specifikace odhadů, specifikace předpokladů, generování predikcí i ověřování získaných výsledků.

10.1.1 Typy prediktivních modelů

Moderní prediktivní modely a jejich algoritmy jsou ***postavené na principech strojového učení*** (machine learning) (SIEGEL, 2013). Modely se různými způsoby ***učí z historických dat***, v nichž ***hledají významné vztahy a proměnné*** vztahující se k cílové proměnné (proměnným), která je v historických datech známa, a kterou je v konkrétním případě žádoucí predikovat.

Nejdůležitější vlastností prediktivních modelů je ***generalizace*** – schopnost naučit se v dostupných datech ***jen to, co je důležité*** a správně ***vyhodnotit náhodné jevy a šum*** v datech jako nevýznamné. Problémy přeučení a nedoučení, zapříčiňující špatnou generalizaci, se týkají všech typů modelů. Řeší je různé techniky, od prořezávání větví rozhodovacího stromu, až po skladbu několika i různých modelů do jednoho velkého modelu (ELDER IV, 2003). V současnosti existuje množství prediktivních modelů.

10.1.2 Hledání vztahů – data mining, cluster analýza

Data mining je definován jako: “*analysis of … observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner*.” (HAND, a další, 2001).

Podstatou je ***rozbor množiny dat vybraných pro prediktivní analytiku***. Je to analýza dat sloužící k ***identifikování skrytých vazeb, vzorů a vztahů***. ***Data mining*** je ***důležitá část prediktivní analytiky***, protože data a vztahy, která identifikuje jako relevantní, mohou být použity při vývoji prediktivního modelu. Data mining v procesu prediktivní analytiky představuje ***získávání znalostí o vztazích*** a výsledný ***prediktivní model je aplikací těchto znalostí***. Hlavní předností data miningu je to, že zaznamenává všechny vztahy (nebo korelace), které jsou v datech přítomny, bez ohledu a znalosti toho, co je zapříčinilo (NYCE, 2007).

Cluster analýza je v prediktivní analytice využitá pro ***hledání podobností v datech***. Pomocí algoritmů a metod seskupuje ***objekty podobných vlastností do skupin***. Tato analýza může být využita k odhalení struktur v datech, ale neposkytuje interpretaci nebo vysvětlení proč tyto struktury existují (STATSOFT, 2013).

10.1.3 Tvorba prediktivního modelu

„The machine actually learns more about your next likely action by studying others than by studying you.“ (SIEGEL, 2013).

Prediktivní modely jsou jedním ***z nejdůležitějších základů prediktivní analytiky***, která na nich přímo staví. Modely z dostupných dat ***analyzují historická chování*** k posouzení ***pravděpodobnosti výskytu predikovaného jevu***, např. že zákazník s určitými vlastnostmi zakoupí nějaký produkt, využije nějakou službu, onemocní, pokusí se o podvod, klikne na reklamní banner na webové stránce atp.

Nejdůležitější vlastností prediktivních modelů je ***generalizace***. Ta zaručuje to, že model naučený z historických dat (in-sample) ***dokáže správně vyhodnotit data nová*** (out-of-sample), která do tvorby a určení modelu nevstoupila.

V nejjednodušší podobě je prediktivní model v podstatě (i jediná) matematická rovnice (TEASLEY, 2005). V současnosti se používají komplexní prediktivní modely využívající ***principy strojového učení*** (SIEGEL, 2013). Modeláři je tvoří jednoduše a rychle v moderních nástrojích, určených pro dolování dat.

Vytvořené modely se v programech ***sami validují, optimalizují a vyhodnocují***. Existuje celá řada typů modelů, které se snaží o co nejpřesnější předpověď daného jevu a jsou vhodné na různé situace.

Vývoj a výběr prediktivního modelu znamená, že modely mohou mít různou velikost a tvar v závislosti na jejich složitosti a využití, pro které jsou navrženy (NYCE, 2007). Pro co nejvyšší přesnost může být použito ***více modelů***, které jsou následně ***porovnávány a kombinovány***.

Vývoj a výběr modelu se skládá ***z následujících kroků***:

- Definují se ***vstupní proměnné a jejich váhy***. Váhy proměnných jsou určeny automatizovaně pomocí algoritmů a strojového učení nebo je modelář nastaví sám.
- Vybere se ***co nejvhodnější model***, který co nejlépe „sedí“ na historická data (insample) a na nová a testovací data (out-of-sample).
- Model se ***ladí a vylepšuje, probíhá úprava proměnných, jejich vah a samotného modelu*** na základě testování, zpětné validace a nových dat.

10.1.4 Soubor modelů (ensemble)

Modely a jejich výstupy (predikce) je možné ***skládat do většího celku***. Vznikne tak jeden velký a robustní model postavený na několika různých prediktivních modelech. Z povahy tohoto modelu se může zdát, že bude náchylnější k přeučení a na out-of-sample datech bude vykazovat horší výsledky než samostatné modely, ze kterých je složený. Je-li složený z více modelů, vstupuje do něj více proměnných, což popírá princip generalizace prediktivních modelů. Paradoxně ale ***výsledky praxe ukazují, že tento model dosahuje lepší predikce*** než samostatné modely, ze kterých je složený.

10.2 Závěry, doporučení



Z kapitoly vyplývají následující ***závěry***:

- Prediktivní modely a jejich algoritmy jsou ***postavené na principech strojového učení*** (machine learning).
- Je nezbytné dobře pochopit klíčové ***faktory, proměnné a parametry***.
- Modely se různými způsoby ***učí z historických dat***, v nichž ***hledají významné vztahy a proměnné*** vztahující se k cílové proměnné.
- Základem je zde ***rozbor množiny dat vybraných pro prediktivní analytiku***, tedy analýza dat sloužící k ***identifikování skrytých vazeb, vzorů a vztahů***.
- Modely z dostupných dat ***analyzují historická chování*** k posouzení ***pravděpodobnosti výskytu predikovaného jevu***.
- Používají se komplexní prediktivní modely využívající ***principy strojového učení.***

- Existuje ***celá řada typů modelů***, které se snaží o co nejpřesnější předpověď daného jevu a jsou vhodné na různé situace.
- ***Vývoj a výběr prediktivního modelu*** znamená, že modely mohou mít různou velikost a tvar v závislosti na jejich složitosti a využití,
- ***Modely a jejich výstupy (predikce)*** je možné ***skládat do většího celku***. Vznikne tak jeden velký a robustní model.
- Klíčovou součástí návrhu modelu, je to, že se definují ***vstupní proměnné a jejich váhy***. Váhy proměnných jsou určeny automatizovaně pomocí algoritmů a strojového učení nebo je modelář nastaví sám.

11. Vyhodnocení výsledků („Assessing Models“)

V této fázi dochází ***ke kontrole dosažených výsledků*** oproti manažerskému zadání. Jsou již k dispozici výsledné znalosti (modely, vzory) dosažené analytickými metodami.

Účelem je:

- definovat ***principy validace a ladění*** prediktivních modelů,
- analyzovat ***kvalitu a obsah*** navržených prediktivních modelů,
- definovat metriky pro hodnocení úspěšnosti modelu,
- vymezit problémy ***overfitting, tedy přeučení modelu*** a na druhé straně ***underfitting, resp. nedoučení modelu,***
- specifikovat celkový ***přístup k hodnocení*** modelu.

Ve fázi je třeba zpětně zkontrolovat ***celý proces*** v souvislosti s manažerským zadáním a případně identifikovat nedostatečně zohledněné obchodní aspekty. Jako součást tohoto kroku je i ***rozhodnutí o využití získaných*** znalostí.

11.1 Závěry



Z kapitoly vyplývají následující ***závěry***:

- Prediktivní model co nejpřesnější, ***musí být testován pomocí skupiny dat out-of-sample, testovacích dat***, která nijak nevstoupila do vývoje a učení modelu.
- Hodnotí se dílčí charakteristiky modelu (overfittig, underfitting), detailněji kapitola 28.4.
- Hlavní ***metriky modelu,*** pomocí kterých se vyhodnocuje ***úspěšnost modelu, jsou Lift, ROC, Missclassification Rate***.
- ***Overfitting,*** resp. přeučení modelu znamená, že model ***špatně vyhodnocuje náhodný šum v datech***, určuje důležité vztahy na základě náhodných proměnných a postrádá schopnost generalizace.
- ***Problémy overfitting,*** mohou být způsobeny následujícími případy:
  - špatné nastavení modelu, například příliš velký (hluboký, rozvětvený),
  - příliš malý vzorek učících dat,
  - chyby ve vstupních proměnných.
- ***Underfitting*** je dáno tím, že ***učení modelu bylo chybou nastavení modelu***, nebo nedostatkem dat ***zastaveno příliš brzy*** a nebyly odhaleny všechny důležité vztahy.
- Při vyhodnocování modelu se na základě predikce ***uskuteční rozhodnutí***. Poté se vyhodnocuje, jak moc predikce odpovídá realitě, ***vyhodnocuje se úspěšnost modelu*** v praxi.

12. Využití výsledků („Model Deployment“)



V rámci fáze se specifikuje ***další využití*** získaných znalostí.

Účelem kapitoly je:

- specifikovat a zajistit dodání výsledků modelů u konkrétních uživatelů,
- realizovat efektivní a názorné prezentace výsledků pro uživatele,
- specifikovat potřeby konsolidace výsledků modelu pro uživatele.

Výstup této fáze může mít podobu sahající od seznamu pravidel, přes formulace získaných znalostí či analytické zprávy, až po praktická nasazení získaných modelů (např. pro klasifikaci zákazníka) do produkčních systémů podniku.

12.1 Prezentace výsledků

Prezentace získaných výsledků z modelů směřuje primárně na uživatele, tj. manažery, specialisty v oblasti prognózování nebo plánování přes různé oblasti byznysu. Je zřejmé, že prediktivní analytika může přinést firmě efekty, pokud tito pracovníci výsledky dobře pochopí a budou je kvalifikovaně využívat. Proto právě prezentace výsledných informací patří k velmi významným součástem řešení úloh prediktivní analytiky. K tomu ***několik poznámek a doporučení*** (Wilson, 2021):

- prezentace by měla být orientována především ***na byznys výsledky, získané efekty*** a řešení stávajících byznys problémů, podstatně méně na technické aspekty řešení,
- základem efektivní prezentace má být ***poznání účastníků prezentace*** na uživatelské straně, zejména jejich potřeb,
- jádrem prezentace by mělo být ***manažerské shrnutí*** („*management sumary*“),
- i když manažeři a plánovači pracují převážně s čísly, v případě prezentací je vhodnější se zaměřit převážně na „příběh“, tedy ***řešení konkrétních problémů***, které výsledky modelů přinášejí,
- ***cílem*** má být nejen přinášet čísla, ale oslovit a přesvědčit uživatele, aby je využívali,
- velmi podstatnou, i když dnes už běžnou, má být ***kvalitní vizualizace*** prezentovaných výsledků.

12.2 Konsolidace informací

Kvalitní informace pro uživatele jsou ty, ***které potřebuje, ale ne zbytečně víc***, se kterými se bude obtížně vyrovnávat a orientovat se v nich. Rovněž zde ***několik poznámek***:

- obvyklá kritika se vztahuje k poskytovaným informacím, které dostatečně ***rychle nevedou k požadovaným výsledkům***,
- ***příliš mnoho času se věnuje způsobu***, resp. postupu analýzy oproti byznys podstatě výsledných informací,
- obvykle se doporučuje ***samotnému postupu*** řešení věnovat při prezentacích pouze ***cca 10 % času*** (oproti vlastnímu obsahu),
- naopak je účelné konsolidovat výstupní informace do logických celků, tedy v kontextu řízení celé firmy,
- konsolidované informace v širokém kontextu řešení ***přispějí nejen k efektivnímu řízení***, ale svým způsobem budou podporovat zvyšování manažerské a analytické kvalifikace klíčových uživatelů.

12.3 Závěry



Z kapitoly vyplývají následující ***závěry***:

- ***Prezentace výsledných informací*** získaných na základě prediktivních modelů patří ke konečné, ale ***velmi významné součásti*** celého postupu řešení prediktivní analytiky.
- Důležitým předpokladem je ***poznání očekávání a potřeb zúčastněných uživatelů***, tedy manažerů a specialistů firmy.
- Efektivní využití poskytovaných výstupů z prediktivní analytiky závisí i na jejich ***konsolidaci v širokém kontextu vzhledem k řízení*** firmy.

C) Uplatnění prediktivní analytiky podle oblastí řízení



[13] Strategické řízení		[14] Finanční řízení, ekonomika
[15] Řízení prodeje		[16] Řízení nákupu
[17] Personální řízení		[18] Řízení majetku
[19] Řízení marketingu	[20] Řízení dopravy	[21] Řízení energií

Účelem oddílu C dokumentu je ***promítnout vymezené principy*** a přístupy prediktivní analytiky ***do vybraných oblastí řízení*** firmy. Je nutné zdůraznit, že prezentované návrhy lze považovat převážně pouze za ***vstupní náměty*** na uplatnění prediktivní analytiky ve firmě a její ***reálné využití se nutně liší*** podle odvětví a typů firem a jejich konkrétních podmínek a potřeb.

Není zde smyslem opakovat veškeré charakteristiky řízení, které jsou již detailněji ***uvedeny v souvisejících dokumentech na portálu MBI-AF,*** a proto se v tomto oddíle často na tyto další dokumenty odkazujeme a zde vybíráme pouze klíčové informace významné pro řešení prediktivní analytiky.

Smyslem každé kapitoly je poskytnout vstupní představu o možnostech prediktivní analytiky vzhledem k prognózování a plánování v přes jednotlivé oblasti řízení. Je zde rovněž snahou, s ohledem na rychlou orientaci, uplatňovat ***standardní strukturu*** jednotlivých kapitol, která obsahuje:

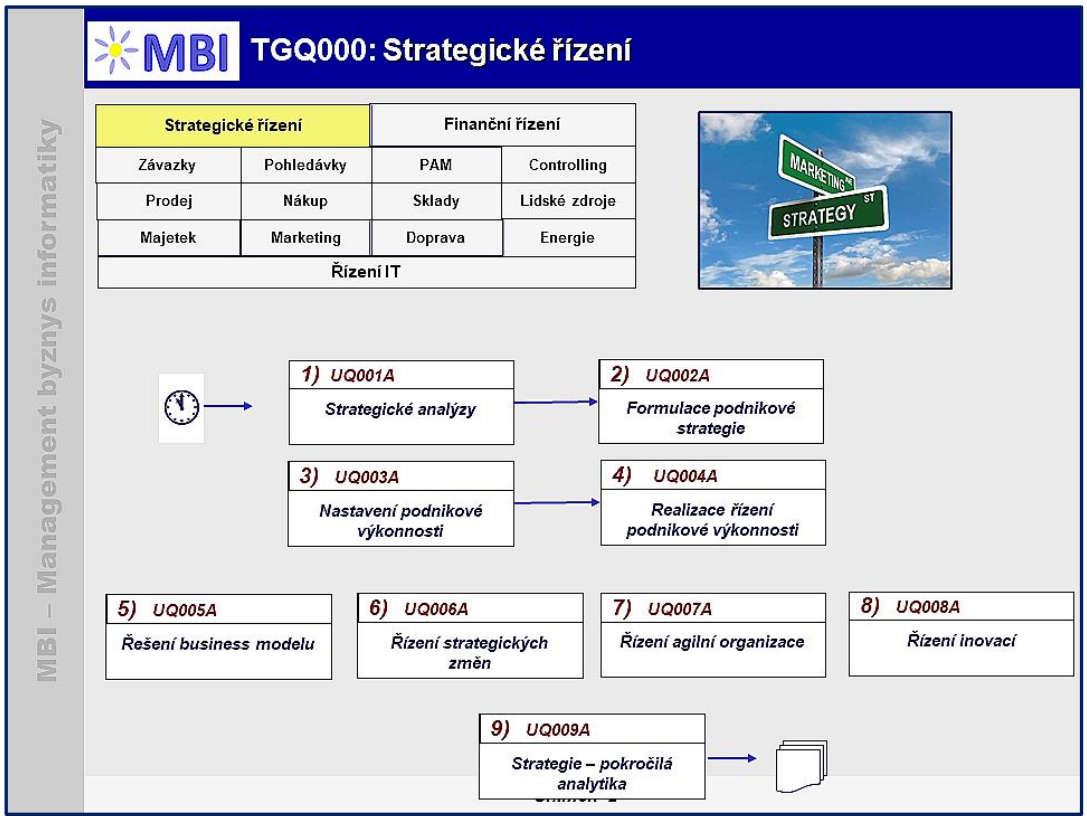
- vymezení ***obsahu řízení*** v dané oblastí řízení a přehled hlavních datových zdrojů jako vstupy do prediktivní analytiky, resp. specifikace prostředí jako základ pro pochopení potřeb prediktivní analytiky:
  - odkaz na dokument „***AF_II_01_Oblasti.pdf***“,
- ***účel a předmět*** prediktivní analytiky v oblasti řízení
- specifikace ***cílových proměnných***, co má být hlavními výstupy prediktivní analytiky pro řízení v dané oblasti a ***analytické jednotky,*** které mají být předmětem analýz,
  - odkaz na dokument „***AF_II_02_Komponenty.pdf***“,
- specifikace „***prediktivní funkcionality***“, to je přehled hlavních funkcí, metod, případně efektů, které mají úlohy a modely prediktivní analytiky firmě poskytovat,

13. Prediktivní analytika strategického řízení firmy

Strategické řízení firmy lze chápat jako základ a ***vstup pro formulaci všech plánů a projektů*** ve firmě**.** Do strategického řízení, které je primárně ***záležitostí nejvyššího vedení firmy***, patří ***definování jejího poslání (mission),*** tj. smysl existence ve vztahu k vlastníkům a dalším zainteresovaným skupinám lidí, dále zpracování ***vize firmy***, formulace ***byznys modelu a*** specifikace ***hlavních a dílčích cílů*****.**

13.1 Vymezení obsahu oblasti řízení, datové zdroje

Součástí strategie jsou jednotlivé ***strategie pro dílčí oblasti řízení***, tj. marketingová strategie, výrobní strategie, obchodní strategie, strategie nákupu, personální strategie atd. včetně řešení jejich vzájemné provázanosti.



Obrázek 13-1: Strategické řízení – přehled úloh

Do strategického řízení firmy ***spadají tyto úlohy***:

- ***Strategické analýzy*** – monitorování a analýzy změn prostředí na trhu, analýzy konkurence a vliv na podnik apod.
- ***Formulace podnikové strategie*** – specifikace podnikových cílů a formulace rozvojových záměrů firmy.
- ***Nastavení systému řízení podnikové výkonnosti*** – vymezení objektů pro řízení výkonnosti a hlavních oblastí podnikového řízení ovlivňujících výkonnost.
- ***Řízení podnikové výkonnosti*** – vlastní průběh sledování a vyhodnocování výkonnosti a řešení omezení a problémů.
- ***Řešení business modelu*** – definování základního fungování firmy, uvědomění si souvislostí jednotlivých částí a aspektů firmy.
- ***Řízení strategických změn*** – definování rozsahu a úskalí strategických změn.

- ***Řízení agilní organizace*** – využití pojetí agilní organizace, které je inspirováno výhodami start-upů.
- ***Řízení inovací*** – vymezení smyslu inovací pro existenci firmy a způsobu jejich řízení.
- ***Strategie – pokročilá analytika*** – využití pokročilých analytických metod v řízení firmy, např. prediktivní analytiky.

→ Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***1***.

13.2 Účel prediktivní analytiky ve strategickém řízeni

Účelem prediktivní analytiky ve strategickém řízení je zejména vytvářet podklady pro přípravu prognóz vývoje a fungování firmy. Většinou jde o predikce klíčových ukazatelů z pohledu celé firmy a jejich využití ve zmíněných prognózách nebo i při řízení výkonnosti firmy, formulování nebo aktualizace byznys modelu a při dalších operacích na úrovni strategického řízení.

13.3 Cílové proměnné

Náměty na specifikaci ***vybraných*** cílových proměnných („*target variables*“) v oblasti strategického řízení obsahuje další přehled:

- Hospodářský výsledek,
- Obrat.
- Objem nákladů firmy.
- MVA (Market Value Added), EVA, Economic Value Added.
- EBITDA (Earnings Before Interest, Taxes, Depreciation and Amortization).
- Počet zákazníků firmy
- Objem investic.
- Tržní podíl je podíl tržeb (obratu) firmy na celkovém tržním obratu vztažený k relevantnímu trhu.

13.4 Analytické jednotky

Jednotky analýzy („*unit of analysis*“) mohou v případě strategického řízení zahrnovat:

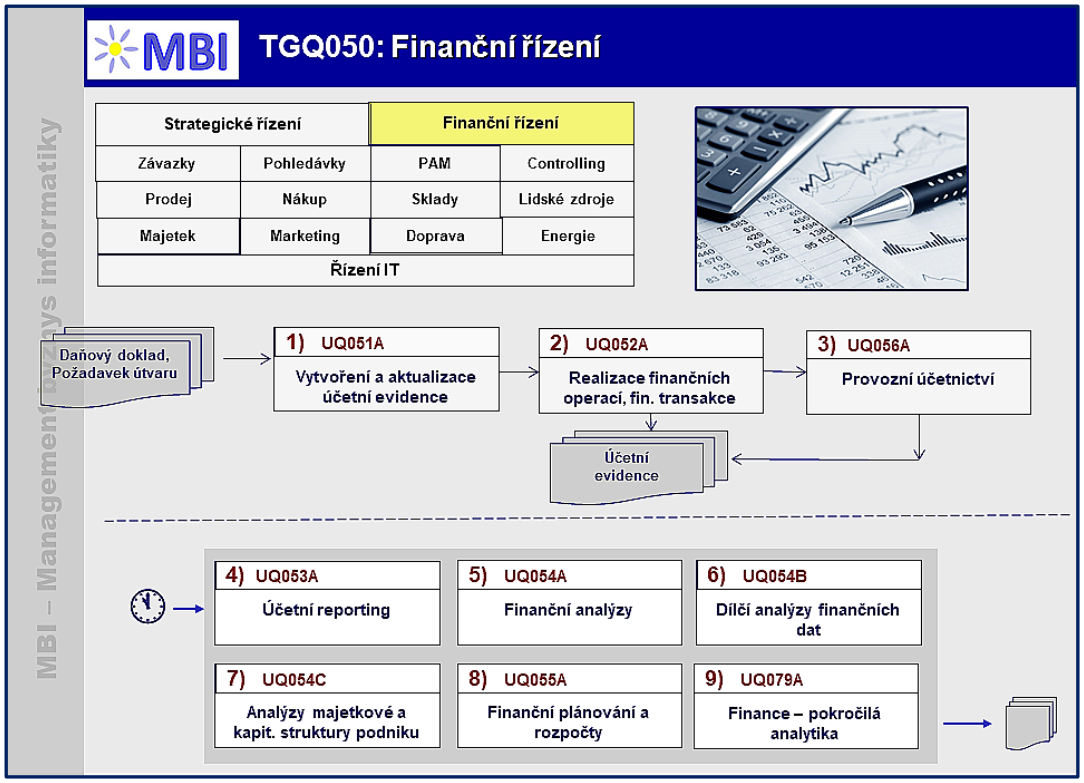
- Obchodní jednotka (business unit).
- Nákladové středisko.
- Nákladový druh.
- Segment trhu.
- Zákazník.
- Investice.

14. Prediktivní analytika finančního řízení

Účelem finančního řízení je:

- zajistit efektivní a spolehlivé ***vykonávání všech účetních operací***,
- realizovat úspěšné ***finanční (úvěrové a další) operace*** vedoucí k udržení finanční stability firmy a jeho konkurenceschopnosti,
- zajistit ***maximalizaci hodnoty majetku*** (pro akcionáře),
- poskytovat pracovníkům finančního řízení (manažerům, finančním analytikům) kvalitní ***analytické a plánovací podklady*** pro řízení finančních zdrojů a zamýšlené finanční transakce,
- ***konsolidovat finanční informace*** z jednotlivých organizačních jednotek a oblastí řízení,
- dosahovat ***stálé likvidity***, tj. schopnosti splácet vklady věřitelům (u bank) a ***likvidnosti***, tj. schopnosti přeměnit finanční instrument (akcie) na peníze,
- zajistit ***solventnost firmy***, tj. schopnost firmy splácet své krátkodobé a dlouhodobé závazky v době jejich splatnosti,
- zajistit ***rentabilitu, tj. ziskovost*** firmy, která měří efektivnost, s níž využívá kapitál, tj. podíl zisku k vloženému kapitálu.

14.1 Vymezení obsahu oblasti řízení



Obrázek 14-1: Finanční řízení – přehled úloh

Do finančního řízení spadají ***tyto úlohy***:

- ***Evidence*** – vytvoření a aktualizace účetní evidence, vytvoření, aktualizace a dotazy do účetní databáze.
- ***Transakce*** – účetní transakce, zpracování daňového dokladu, provedení účetních transakcí.
- ***Reporting*** – účetní reporting, vytváření standardních finančních výkazů a reportů.

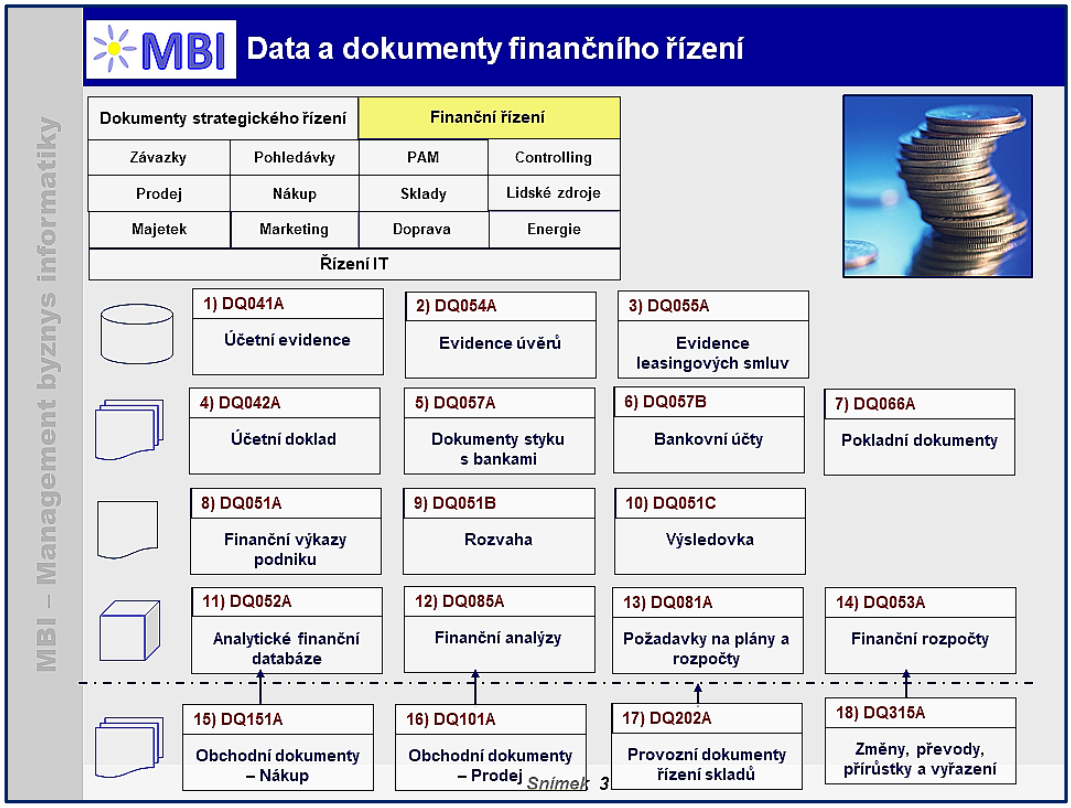
- ***Finanční analýzy*** – analýzy globálních finančních ukazatelů (likvidita apod.) a jejich časového vývoje.
- ***Dílčí analýzy finančních dat*** – základní analýzy stavu účtů, jejich pohybů podle různých dimenzí (účtů, poboček, útvarů atd.) a jejich časového vývoje.
- ***Analýzy majetkové a kapitálové struktury podniku*** – analýzy na základě informací z rozvahy.
- ***Analýzy výnosů podniku.***
- ***Analýzy nákladů podniku.***
- ***Plánování*** – finanční plánování a rozpočty.
- ***Provozní účetnictví –*** sleduje hospodaření jednotlivých středisek.
- ***Finance – pokročilá analytika –*** řeší vývoj, problémy a predikce v oblasti finančního řízení.



Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***2***.

14.2 Datové zdroje

Souhrnný pohled na data a dokumenty finančního řízení a jejich základní členění prezentuje další obrázek. ***Datové zdroje*** představují zejména účetní evidenci, evidenci úvěrů, evidenci leasingových smluv a další.



Obrázek 14-2: Přehled a členění dat a dokumentů finančního řízení



Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.2***.

14.3 Účel prediktivní analytiky ve finančním řízeni

Účelem prediktivní analytiky ve finančním řízení firmy je poskytovat podklady pro přípravu finančních prognóz a zejména finančních plánů různých typů a zaměření. K těm patří zejména:

- plánování externích zdrojů financování, úvěrování atd.
- hlavní podnikový rozpočet,
- predikce cash flow,
- rozpočtová rozvaha,
- rozpočtová výsledovka.

S tím souvisí zpracování predikcí na bázi hlavních finančních ukazatelů, jak ukazují cílové proměnné v další podkapitole. Ukazatelů je v tomto případě značné množství, a tedy je nezbytné z nich vybírat právě ty, které mají pro finanční plánování firmy zásadní význam.

14.4 Cílové proměnné

Náměty na specifikaci ***vybraných*** cílových proměnných („*target variables*“) v oblasti finančního řízení obsahuje další přehled:

- Hospodářský výsledek,
- Obrat.
- Objem nákladů firmy.
- MVA (Market Value Added), EVA, Economic Value Added.
- EBITDA (Earnings Before Interest, Taxes, Depreciation and Amortization).
- Ukazatelé cash-flow.
- Ukazatelé rentability a nákladovosti.
- Ukazatelé aktivity.
- Ukazatelé zadluženosti a finanční struktury
- Ukazatelé likvidity.
- Ukazatelé kapitálového trhu.
- Ukazatelé finančních fondů
- Ukazatelé majetkové struktury firmy.
- Ukazatelé kapitálové struktury firmy

→ Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***2.2***.

14.5 Analytické jednotky

Jednotky analýzy („*unit of analysis*“) mohou v případě finančního řízení zahrnovat:

- Obchodní jednotka (business unit).
- Nákladové středisko.
- Nákladový druh.
- Druh příjmů.
- Druh výdajů

14.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace finančních zdrojů – podle jejich objemu a dostupnosti,

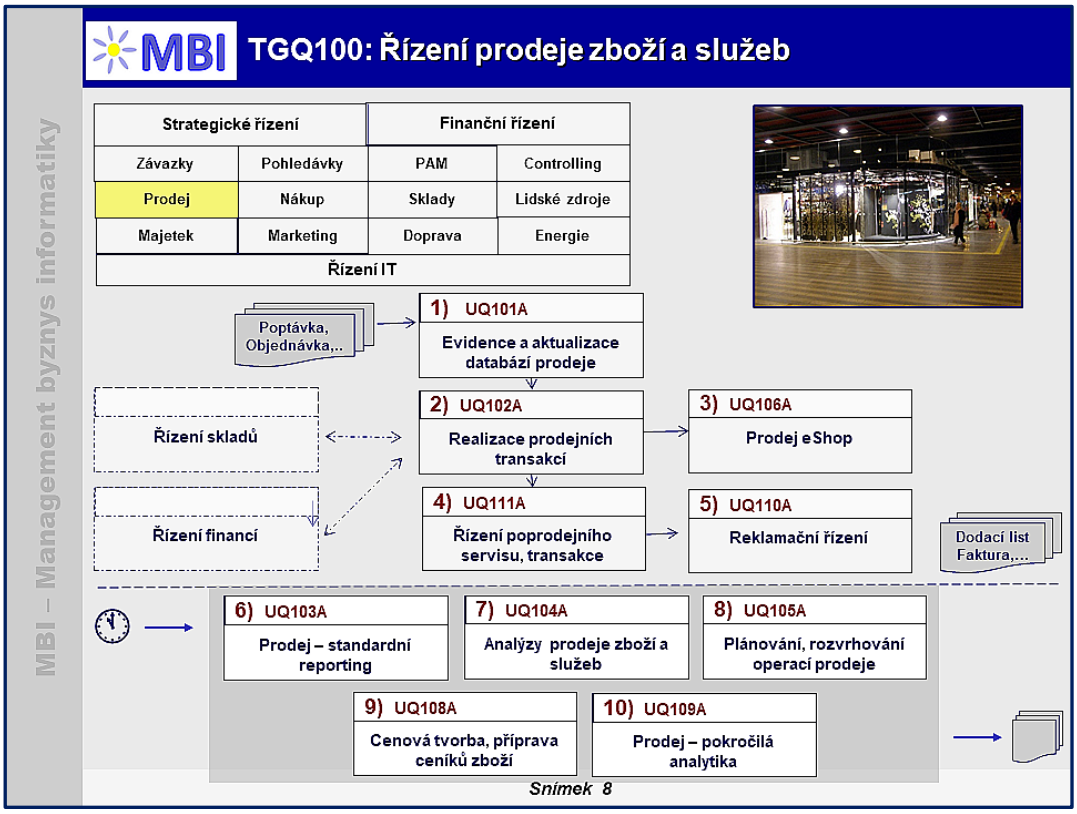
- klasifikace nákladů firmy,
- klasifikace finančních výkazů.

15. Prediktivní analytika v řízení prodeje

Účelem řízení prodeje je:

- dosahovat požadované ***ekonomické výsledky***, tj. tržby, obrat, ziskovost atd.,
- zajistit efektivní, rychlé ***provádění prodejních operací*** s vysokou flexibilitou vzhledem k požadavkům zákazníka,
- poskytovat pracovníkům prodeje (manažerům, obchodníkům) kvalitní ***analytické a plánovací podklady*** pro rozhodování o celkové orientaci prodejních aktivit firmy i o realizaci jednotlivých zakázek.

15.1 Vymezení obsahu oblasti řízení



Obrázek 15-1: Řízení prodeje – přehled úloh

Do řízení prodeje spadají ***tyto úlohy***:

- ***Evidence*** – evidence prodejů a prodejních operací, aktualizace základních údajů databází prodeje, pokud požadavek představuje nového zákazníka, nové zboží nebo službu apod..
- ***Transakce*** – realizace ***obchodního případu Prodej,*** řízení prodejních transakcí, na základě vyvolané událostí např. příchod objednávky, zařazení a vyhodnocení objednávky, poskytnutí vyjádření a dalších informací zákazníkovi atd.
- ***Reporting*** – vytváření standardních reportů prodeje.
- ***Analýzy*** – analýzy prodeje podle různých hledisek (zboží, služeb, zákazníků atd.) a jeho časového vývoje.
- ***Plánování*** – plánování, rozvrhování, prognózování prodeje.
- ***eShop*** – realizace prodejních transakcí prostřednictvím eShopu.

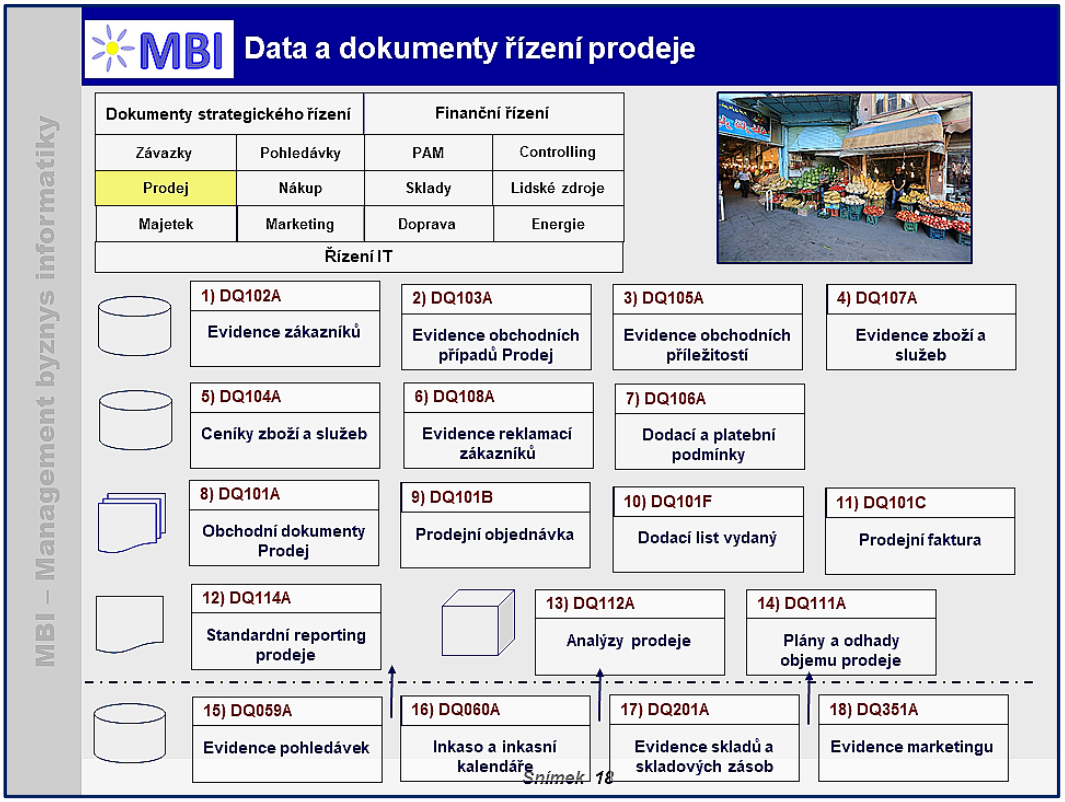
- ***Cenová tvorba*** – příprava ceníků zboží.
- ***Prodej – pokročilá analytika.***
- ***Reklamační řízení.***
- ***Servis*** – řízení poprodejního servisu, servisní zásahy a transakce.

→ Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***7***.

15.2 Datové zdroje

Souhrnný pohled na data, dokumenty v řízení prodeje dokumentuje další obrázek. ***Datové zdroje*** představují zejména:

- Evidence zákazníků.
- Evidence obchodních případů „*Prodej*“.
- Evidence obchodních příležitostí.
- Evidence zboží a služeb.
- Ceníky zboží a služeb.
- Evidence reklamací zákazníků.
- Dodací a platební podmínky.



Obrázek 15-2: Souhrnný přehled dat a dokumentů v řízení prodeje

→ Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.7***.

15.3 Účel prediktivní analytiky v řízeni prodeje

Účelem prediktivní analytiky v řízení prodeje zboží a služeb je podpora a příprava podkladů pro následující prognostické a plánovací aktivity firmy:

- ***Sestavení plánu prodeje*** obvykle na časová období rok, pololetí, kvartál, dekáda, případně týden a den.
- Zajištění propojení ***mezi obchodními cíli organizace a ostatními dílčími plány*** a přímo ovlivňuje firmou nabízený sortiment zboží a služeb.
- ***Zpracování výhledu prodeje***, např. na 12 měsíců klouzavě v průběhu roku.
- ***Sestavení prodejního rozpočtu*** silně závisí na prostředí konkrétní firmy, charakteristice trhu, druhu nabízeného produktu a dalších.
- Prodejní rozpočet ***v detailu na konkrétní skupiny výrobků a služeb***, kde se vybírá konkrétní prvek skupiny s nejobecnějšími parametry. Zvolený detail záleží vždy na situaci dané organizace.
- Rozpočet ***podle detailu na jednotlivé odběratele*** (typické pro velkoobchodní prodej), kde se s jednotlivými partnery mohou aplikovat různá pravidla vzájemného obchodního vztahu.
- Předpokládaný ***vývoj prodeje***, kde se zohledňuje sezónnost daného odvětví a zkušenosti s cyklickým chováním typického zákazníka dané organizace.
- ***Odhady prodeje*** se stanovením minimálního nutného objemu prodeje k pokrytí nákladů firmy.

15.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables*“) v oblasti řízení prodeje obsahuje další přehled:

- ***Počet zákazníků firmy*** zahrnuje všechny zákazníky firmy včetně zákazníků IT služeb.
- ***Tržby z prodeje zboží a služeb*** jsou tržby podle sortimentních položek, regionů, podnikových útvarů, prodejců a podíly jednotlivých položek na celkových tržbách
- ***Tržní podíl*** je podíl tržeb (obratu) firmy na celkovém tržním obratu vztažený k relevantnímu trhu.
- ***Náklady prodeje zboží a služeb*** jsou náklady na prodej podle nákladových druhů,
- ***Prodejní marže*** je marže podle sortimentních položek.
- ***Dodací lhůta*** vyjadřuje dobu, která uplyne od předání objednávky odběratelem až po okamžik dostupnosti (pohotovosti) zboží u odběratele vyjádřené ve stanoveném čase
- ***Počet prodaných produktů přes eShop*** je celkový součet prodaných kusů jednotlivých produktů (kategorií produktů).

15.5 Analytické jednotky

Jednotky analýzy („*unit of analysis*“) mohou v případě řízení prodeje zahrnovat:

- Zákazník.
- Zboží
- Služba.
- Obchodní případ Prodej.
- Region.
- Objednávka.

15.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace zákazníků – jejich rozřazení do stejnorodých tříd a odhad pravděpodobnosti správnosti takového zařazení pro každého zákazníka,

- klasifikace prodávaných zboží – obvykle podle katalogu zboží,
- klasifikace poskytovaných služeb – posouzení správnosti klasifikace uvedením pravděpodobnosti správného členění.

- *Přiřazování podobností:*
  - určování podstatných charakteristik zákazníků a na tomto základě výběr firem, které jsou obdobné našim nejlepším zákazníkům,
- *Profilování, „Popis chování":*
  - specifikace klíčových charakteristik zákazníků z pohledu přístupů k zajištění a řízení obchodních případů Prodej,
- *Predikce vazeb:*
  - vyhodnocování vazeb kvality a včasnosti zajišťovaných dodávek firmou.

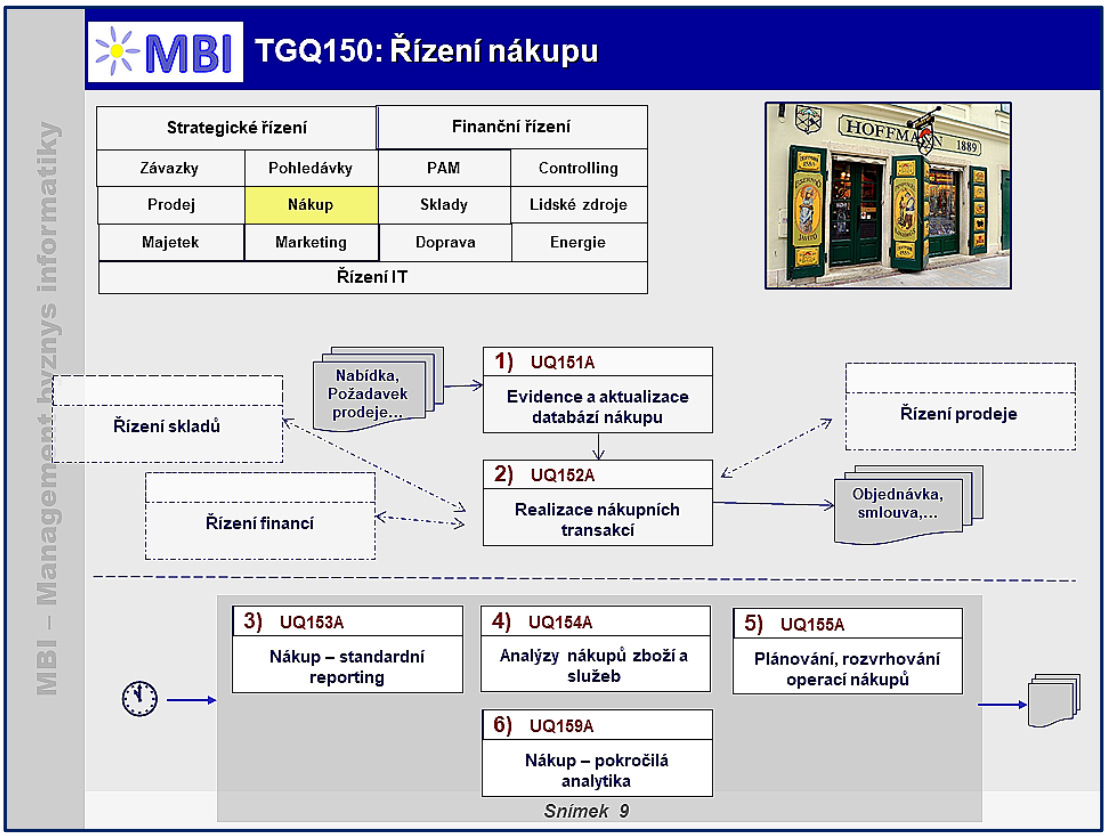
16. Prediktivní analytika v řízení nákupu

Účelem řízení nákupu je:

- dosahovat požadované ***ekonomické výhodnosti realizovaných nákupů***, při požadované kvalitě, objemu a sortimentu nakupovaných zboží, materiálů a služeb,
- zajistit efektivní, rychlé ***provádění nákupních operací*** s vysokou flexibilitou vzhledem k požadavkům firmy a možnostem dodavatelů,
- poskytovat pracovníkům nákupu (manažerům, obchodníkům) kvalitní ***analytické a plánovací podklady*** pro rozhodování o celkové orientaci nákupů firmy i o realizaci jednotlivých nákupních operací.

16.1 Vymezení obsahu oblasti řízení,

Celkový přehled úloh řízení nákupu obsahuje další obrázek.



Obrázek 16-1: Řízení nákupu – přehled úloh

Do řízení nákupu spadají ***tyto úlohy***:

- ***Evidence*** – nákupů a nákupních operací, aktualizace základních údajů databází nákupů, pokud jde o nového dodavatele, nové zboží nebo materiál apod.
- ***Transakce*** – realizace ***obchodního případu Nákup,*** jednotlivých nákupních transakcí vyvolané událostí, např. požadavkem na nákup, příchodem nové nabídky dodavatele, následně zařazení a vyhodnocení požadavku, nebo nabídky (ekonomické, vzhledem k zásobám), vyjádření podnikovému útvaru nebo dodavateli atd.
- ***Reporting*** – nákupů, vytváření standardních reportů nákupu.
- ***Analýzy*** – nákupů materiálů, zboží a služeb, analýzy nákupů podle různých hledisek (zboží, dodavatelů atd.) a jejich časového vývoje.

- ***Plánování*** – plánování, rozvrhování nákupních operací, plánování nákupů zboží a služeb a zjišťování budoucí, resp. předpokládané jejich potřeby.
- ***Nákup – pokročilá analytika***.

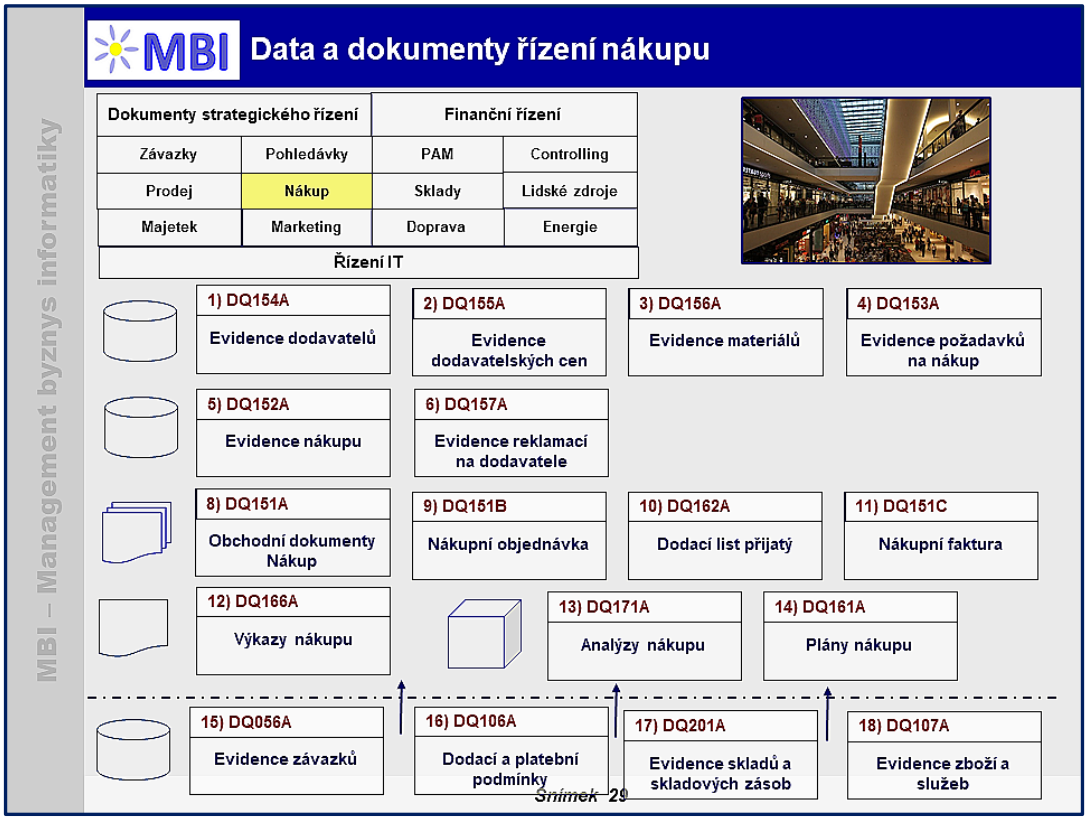


Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***8***.

16.2 Datové zdroje

Souhrnný přehled hlavních dat a dokumentů řízení nákupu dokumentuje další obrázek. ***Datové zdroje*** představují zejména:

- Evidence dodavatelů.
- Evidence dodavatelských cen.
- Evidence nakupovaných materiálů.
- Evidence nakupovaných zboží a služeb.
- Evidence požadavků na nákup.
- Evidence nákupu.
- Evidence reklamací na dodavatele.



Obrázek 16-2: Souhrnný přehled dat a dokumentů řízení nákupu



Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.8***.

16.3 Účel prediktivní analytiky v řízeni nákupu

Prediktivní analytika v řízení nákupů podporuje prognózování a plánování nákupů materiálů, zboží a služeb pro potřeby firmy. V oblasti plánování nákupů má zejména připravovat podklady pro tyto operace:

- sestavení výhledu a plánu nákupu,
- propočet spotřeby materiálu, resp. zboží na dané období,
- bilancování potřeby dodávek materiálu je základní metrika s následnými úpravami podle možností dodavatele, představuje pak vstup do poptávek nebo objednávek,
- sestavení nákupního rozpočtu,
- operativní plánování nákupu se připravuje ve vazbě na plánování prodeje a plán výroby, resp. plánování výrobních zakázek.

16.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables*“) v oblasti řízení nákupu obsahuje další přehled:

- ***Objem nákupů za stanovený čas*** je celkový objem nákupů zboží a služeb v rozlišení např. podle sortimentu, dodavatelů a regionů.
- ***Náklady na zajištění nákupu zboží*** je objem nákladů na nákup podle nákladových druhů, případně útvarů.
- ***Počet dodavatelů*** je počet aktuálně využívaných, potenciálních i v minulosti využívaných dodavatelů firmou***.***
- ***Dodací lhůta nákupů*** vyjadřuje dobu, která uplyne od předání objednávky firmou až po okamžik dostupnosti (pohotovosti) zboží nebo materiálu.

16.5 Jednotky analýzy

Jednotky analýzy („*unit of analysis*“) mohou v případě řízení nákupu zahrnovat:

- Zboží
- Služba.
- Materiál.
- Dodavatel.
- Obchodní případ Nákup.
- Region.

16.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace dodavatelů – jejich rozřazení do stejnorodých tříd a odhad pravděpodobnosti správnosti takového zařazení pro každého dodavatele,
  - klasifikace dodávaných materiálů – obvykle vychází z oficiální nomenklatury materiálů,
  - klasifikace poskytovaných a nakupovaných služeb – posouzení správnosti klasifikace uvedením pravděpodobnosti správného členění.
- *Přiřazování podobností:*
  - určování podstatných charakteristik našich obchodních partnerů firem a na tomto základě výběr firem, které jsou obdobné našim nejlepším dodavatelům.
- *Profilování, „Popis chování“:*
  - specifikace klíčových charakteristik dodavatelů z pohledu přístupů k zajištění a řízení obchodních případů.

- *Predikce vazeb:*
  - vyhodnocování vazeb kvality a včasnosti dodávek vzhledem k velikosti, ekonomické a personální síle firmy.

17. Prediktivní analytika v řízení personálních zdrojů

Účelem řízení lidských zdrojů je:

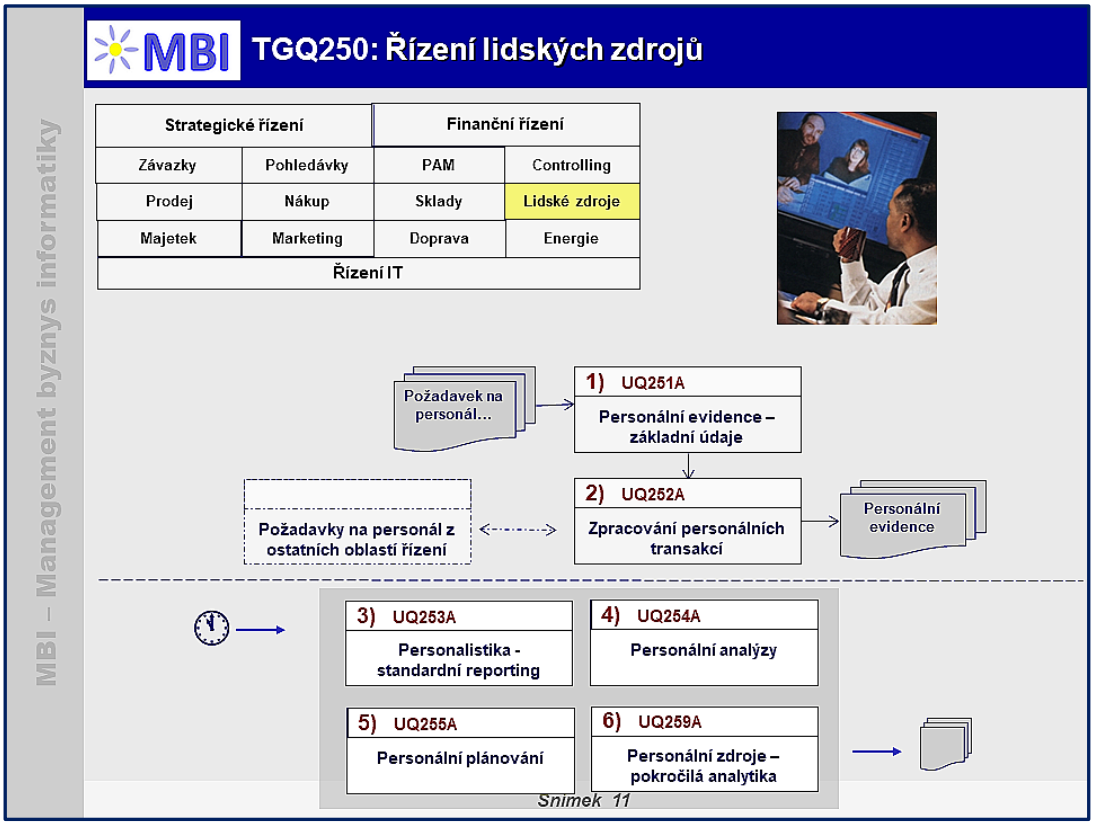
- dosahovat ***počtu a profesní a věkové struktury pracovníků*** podle potřeb firmy,
- zajistit efektivní, rychlé ***provádění operací v personálním řízení*** s vysokou flexibilitou vzhledem k požadavkům firmy,
- realizovat potřebný ***kvalifikační rozvoj pracovníků*** vzhledem k aktuálnímu stavu obchodních a dalších aktivit firmy a k jejich očekávanému vývoji,
- poskytovat vedoucím pracovníkům kvalitní ***analytické a plánovací*** podklady pro řešení požadavků na personální kapacity a jejich struktury.

Hlavní zdroje:

- FITZ-ENZ J., MATTOX II J.,R,: *Predictive Analytics for Human Resources,* Wiley and SAS Business Series, 2014.ISBN: 978-1-118-89367-8.
- EDWADS M., EDWARDS, K.: *Predictive HR Analytics: Mastering the HR Metric.* Kogan Page, 2019. ISBN: 978-0-7494-8444-6.

17.1 Vymezení obsahu oblastí řízení

Celkový přehled úloh řízení lidských zdrojů obsahuje další obrázek:



Obrázek 17-1: Řízení lidských zdrojů – přehled úloh

Do řízení lidských zdrojů spadají ***tyto úlohy***:

- ***Personální evidence*** – aktualizace základních údajů personálních databází.
- ***Transakce*** – zpracování personální evidence, transakce, přijímání pracovníka, propuštění pracovníka, požadavky na školení, zpracování dílčích personálních dokumentů pracovníka.

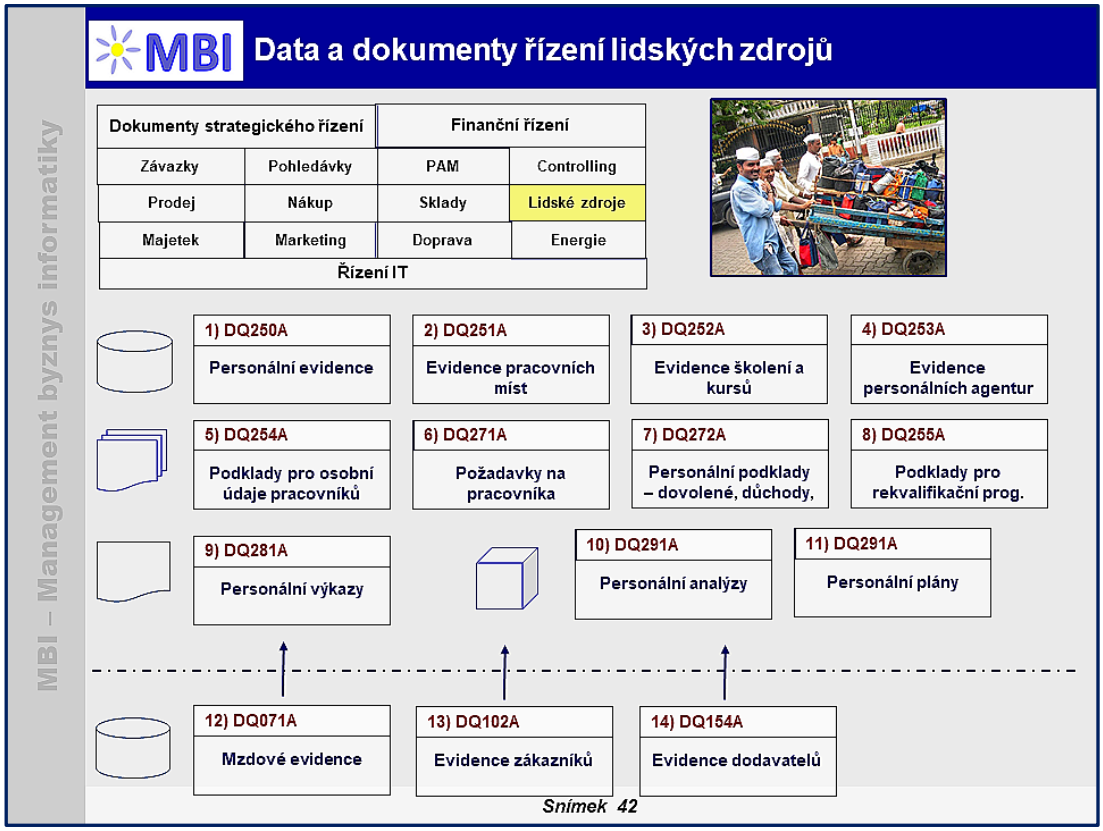
- ***Reporting*** – standardní reporting, vytváření standardních personálních reportů a výkazů.
- ***Personální analýzy.***
- ***Personální plánování*** včetně plánování kvalifikačního rozvoje.
- ***Personální zdroje – pokročilá analytika***.

→ Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***10***.

17.2 Datové zdroje

Souhrnný přehled dat a dokumentů řízení lidských zdrojů dokumentuje další obrázek. ***Datové zdroje*** představují zejména:

- Podniková personální evidence.
- Evidence pracovníků se změněnou pracovní schopností.
- Evidence závazků za zaměstnancem a zaměstnavatelem.
- Evidence vypůjčeného nebo přiděleného materiálu zaměstnanci.
- Evidence uchazečů o zaměstnání.
- Evidence nároků na dovolenou a sociálních výhod.
- Evidence pracovních míst.
- Evidence školení a kursů.
- Evidence personálních a školicích agentur.



Obrázek 17-2: Souhrnný přehled dat a dokumentů řízení lidských zdrojů

→ Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.10***.

17.3 Účel prediktivní analytiky v řízeni lidských zdrojů

Účelem je na základě současných a historických faktů predikovat očekávaný vývoj potřeb a možností zajištění personálních zdrojů ve firmě a zvyšovat pravděpodobnost výběru správných lidí na správná místa, zvyšovat efektivitu získávání nových lidí a efektivitu jejich kvalifikační přípravy. Úloha tak představuje uplatnění metod prediktivní analytiky v personálním řízení

V oblasti personálního plánování má zejména připravovat podklady pro tyto operace:

Zahrnuje např. tyto ***funkce:***

- ***Podpora plánování*** pracovní síly, řízení kompetencí a řízení změn v personálních strukturách s dopady na změny organizace.
- ***Predikce potřeby*** pracovníků podle vybraných charakteristik:
  - předpokládaný vývoj potřeby personálních zdrojů v dalších létech podle prediktorů objem výroby nebo služeb, počet zákazníků, počet a velikost dislokovaných poboček,
  - určování předpokládaných nákladů na získávání, uplatňování a udržování lidského kapitálu.
- ***Hodnocení pracovníků*** a pracovních týmů.
- ***Očekávaná potřeba*** kvalifikačních programů.
- Predikce ***ekonomických výsledků investic*** do personálních zdrojů.
- ***Očekávaný vývoj*** na pracovním trhu, předpokládaný vývoj nabídky na trhu práce podle dále upřesňovaných prediktorů z oblasti demografie

Poznámky:

- Podle (Fitz-Enz, Mattox, 2014) se uplatnění prediktivní analytiky v personálním řízení projevuje v celkovém zvýšení výkonu firmy, někdy až o 4 %.

17.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables“*) v oblasti personálního řízení obsahuje další přehled:

- ***Počet otevřených pracovních pozic*** ve firmě.
- ***Počet nových pracovníků*** podle útvarů, pozic a úrovní řízení.
- ***Čas na získání nového pracovníka*** podle pozic, útvarů, regionů.
- ***Čas na zapracování nového pracovníka*** podle pozic, útvarů, regionů.
- ***Průměrné náklady na získání*** nového pracovníka, průměrná mzda nových pracovníků, náklady na zapracování nového pracovníka.
- Celkový ***objem plánovaných nebo vyplacených mezd***, a to podle pozic a různých druhů***.***
- ***Počty pracovníků*** jsou fyzické počty pracovníků firmy, tj. nepřepočítané podle úvazků.
- ***Pracovní fond v člověkodnech*** znamená přepočítaný objem pracovní doby pracovníků firmy.
- ***Utilizace pracovníků,*** procentní vyjádření využití času, např. na projektech.
- ***Profitabilita pracovníků***
- ***Spokojenost pracovníků ve firmě – eNPS Net Promoter Score***: „Ohodnoťte na stupnici od 0 - 10, na kolik je pravděpodobné, že doporučíte firmu nejlepšímu kamarádovi jako vhodného zaměstnavatele?”: Promotér (9 – 10), Neutrální (7 – 8), Kritik (0 – 6).“
- ***Fluktuace zaměstnanců*** je ukazatel, který je zaměřen na měření změn ve stavu zaměstnanců.

- ***Objem kursů a programů*** je objem plánovaných i absolvovaných školení v člověkodnech, např. manažerských školení, školení metod v různých oblastech podnikového řízení, školení bezpečnosti práce atd.
- ***Počty získaných certifikátů.***
- ***Náklady na kursy a kvalifikační programy*** jsou náklady na jednotlivé vzdělávací programy, kursy a odborná školení.
- ***Výkonnost pracovníků*** podle pozic, měrných jednotek, kvality výkonů.
- ***Hodnocení spokojenosti manažerů s pracovníky*** podle útvarů, pracovníků.

17.5 Jednotky analýzy

Jednotky analýzy („*unit of analysis*“) mohou v případě personálního řízení zahrnovat:

- Pracovník.
- Pracovní pozice.
- Pracovní úkoly.
- Podnikový útvar.
- Nákladové druhy.
- Druhy mezd.
- Kurs, kvalifikační program.
- Poskytovatel kursů.
- Certifikát.

17.6 Prediktivní metody a funkcionalita

Metody aplikovatelné v prediktivní analytice lidský zdrojů jsou obvykle:

- Regresní analýzy.
- Rozhodovací stromy.
- Neuronové sítě.

Významné ***prediktivní funkce***:

- *Klasifikace*:
  - klasifikace pracovníků – jejich rozdělení do skupin a odhad pravděpodobnosti správnosti takového zařazení pro každého pracovníka,
  - klasifikace poskytovaných školení a kvalifikačních programů – obvykle vychází z oficiální nabídek školicích firem,
  - klasifikace kvalifikačních tříd – posouzení nároků na znalosti a zkušenosti u každé kvalifikační třídy,
- *Profilování,* „*Popis chování*“*:*
  - specifikace klíčových charakteristik jednotlivých pracovníků a jejich skupin z pohledu přístupů k zajištění pracovních úkolů.

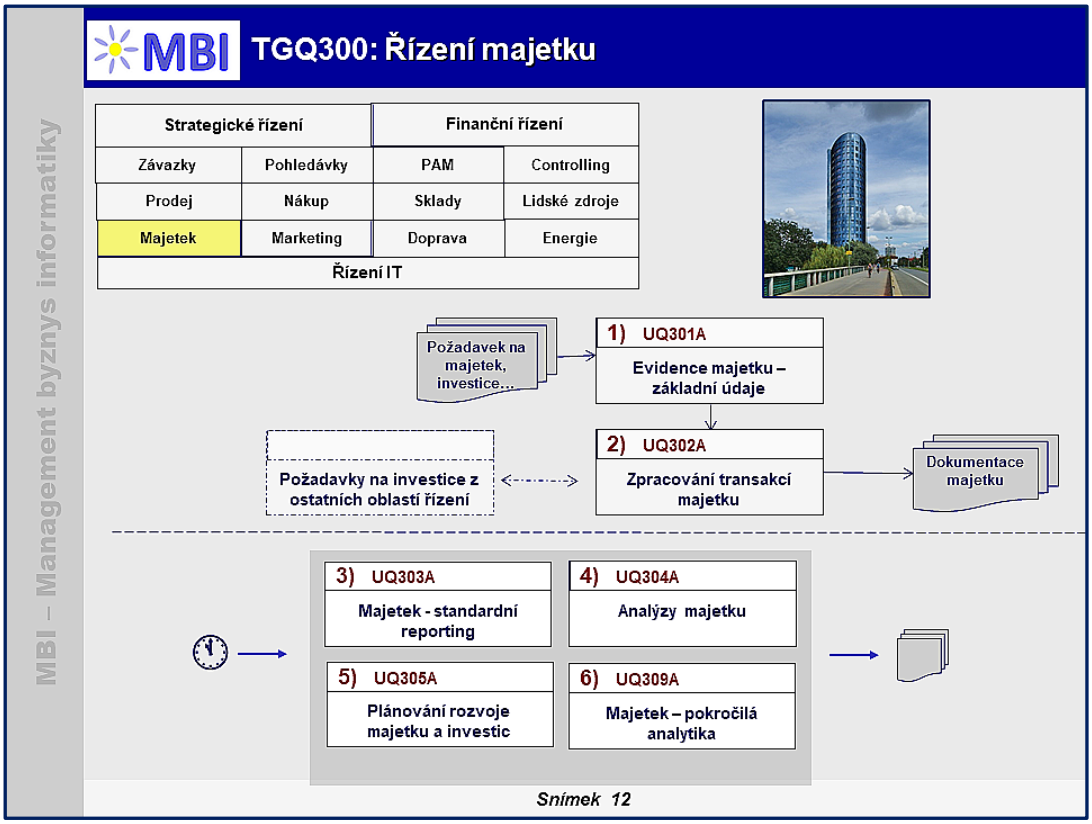
18. Prediktivní analytika v řízení majetku

Účelem je:

- dosahovat takové ***struktury a rozmístění majetku*****,** které budou nejlépe odpovídat aktuálním i budoucím potřebám firmy,
- zajistit ***kvalitní dokumentaci majetku*** na úrovni jeho pasportizace, specifické technické dokumentace, ekonomické dokumentace,
- realizovat efektivní ***plánování a průběh investičních akcí***,
- poskytovat pracovníkům v oblasti řízení majetku kvalitní ***analytické a plánovací podklady*** pro rozhodování o stavu majetku nárocích na údržbu a na přípravu nových investičních akcí, pro uplatnění optimálních odpisových modelů a další.

18.1 Vymezení obsahu oblasti řízení, datové zdroje

Celkový přehled úloh řízení majetku obsahuje další obrázek:



Obrázek 18-1: Řízení majetku – přehled úloh

Do řízení majetku spadají ***tyto úlohy***:

- ***Evidence majetku*** – aktualizace základních údajů majetkových databází, klasifikace majetku, technických údajů apod.
- ***Transakce*** – zpracování majetku, zpracování jednotlivých majetkových transakcí, požadavku na investice, řešení oprav, zpracování odpisů apod.
- ***Reporting*** – vytváření standardních majetkových reportů a výkazů.
- ***Analýzy majetku*** podle různých dimenzí (druhy majetku, zodpovědnost, umístění apod.) a sledování vývoje majetku v čase.
- ***Plánování*** rozvoje majetku a investic.

- ***Majetek – pokročilá analytika.***

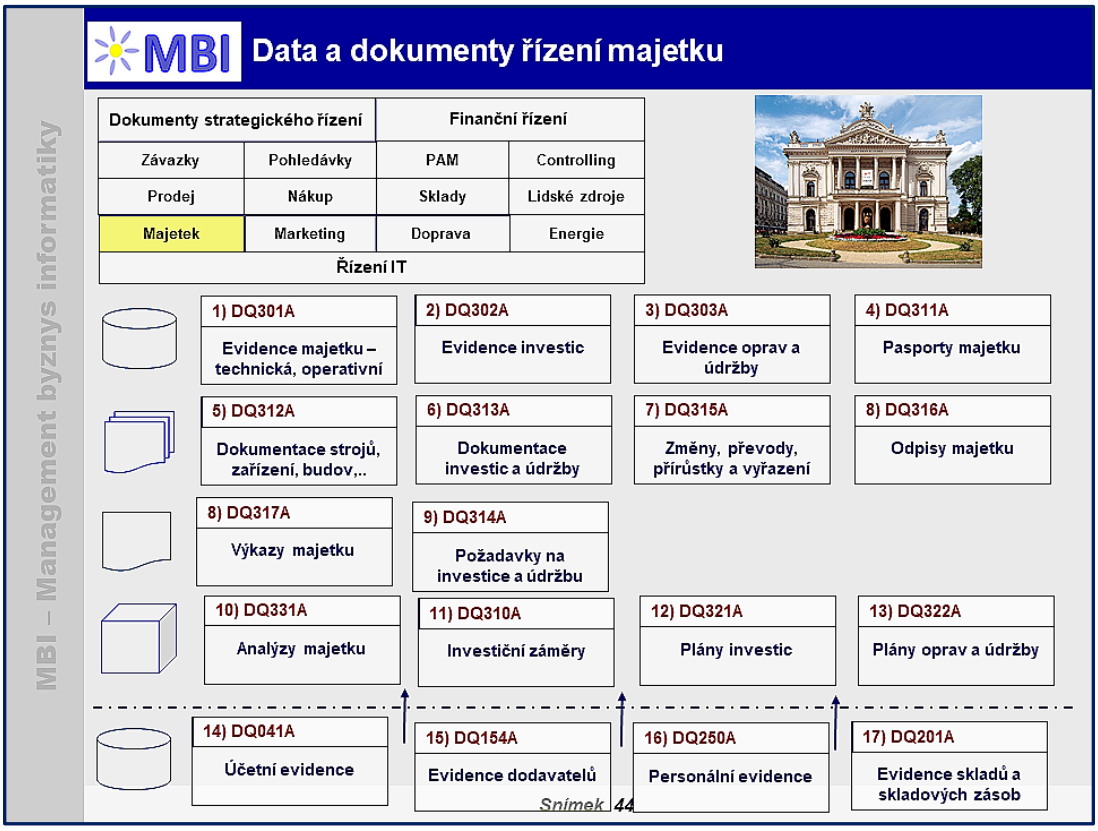


Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***11***.

18.2 Datové zdroje

Souhrnný přehled dat a dokumentů řízení majetku dokumentuje další obrázek. ***Datové zdroje*** představují zejména:

- Evidence majetku – technická, operativní, účetní.
- Evidence investic.
- Evidence vlastních / dodavatelských oprav a údržby.
- Pasporty majetku.



Obrázek 18-2: Data a dokumenty v Řízení majetku



Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.11***.

18.3 Účel prediktivní analytiky v řízeni majetku

Prediktivní analytika v řízení majetku podporuje prognózování a plánování rozvoje a údržby majetku. V oblasti řízení majetku, investic a údržby má zejména připravovat podklady pro tyto operace:

- ***plánování oprav a údržby*** plánování celkového objemu oprav a jednotlivých akcí.
- ***příprava plánu rozvoje majetku,*** plánování celkového objemu investic.

- ***schválení a realizace investičního plánu,*** tj. souhrnné schválení na konci předcházejícího roku, nejprve na úrovni útvaru, schválení plánu představenstvem na úrovni jednotlivých položek a termínů nákupu.
- ***plánování služeb*** souvisejících s majetkem, tj. odborných prohlídek a zkoušek jednotlivých typů zařízení.

18.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables*“) v oblasti řízení majetku obsahuje další přehled:

- ***Objem majetku firmy*** podle druhů majetku, vyjádřený v měrných jednotkách a Kč.
- ***Objem odpisů*** majetku podle odpisových kategorií.
- ***Objem investic*** představuje objem plánovaných a realizovaných investic do majetku firmy.
- ***Objem oprav a údržby*** je objem plánovaných a realizovaných oprav a údržby vyjádřený v Kč a normohodinách.
- ***Kapacita*** výrobních zařízení podle typů.
- ***Počet*** dopravních prostředků podle typů.

18.5 Jednotky analýzy

Jednotky analýzy („*unit of analysis*“) mohou v případě řízení majetku zahrnovat:

- Druhy majetku.
- Investice.
- Podnikový útvar.
- Kategorie odpisů.
- Dopravní prostředek.
- Druh oprav.
- Druh údržby.
- Dodavatel oprav.

18.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace majetku podle jednotlivých druhů – jejich rozdělení do skupin a odhad pravděpodobnosti správnosti takového zařazení pro každého druhu majetku,
  - klasifikace majetku podle úrovně využití – obvykle vychází z provozní evidence firmy,
  - klasifikace odpisových tříd – podle běžných standardů a legislativy,
  - klasifikace připravovaných investic,
  - klasifikace údržbových činností,
- *Profilování, „Popis chování“:*
  - specifikace dodavatelů investic, jejich kvality a poskytovaných služb.

18.7 Příklad: Prediktivní údržba

(Adámek, 2019)

„***Náklady na údržbu*** jsou jednou z hlavních částí celkových operačních nákladů ve výrobních podnicích. Podle specifik průmyslu, se mohou jednotlivé náklady na údržbu pohybovat mezi 15 % a až 60 %

z ceny vyráběného produktu. Pro příklad v potravinářském odvětví se průměrná suma nákladů na údržbu pohybuje okolo 15 procent v oblasti těžkého průmyslu mohou tyto náklady lehce dosáhnout až na 60 procent celkové produkční ceny. Průzkumy ukazují, že například v USA utratí každoročně za údržbu 200 miliard dolarů, což ukazuje důležitost údržby v úplně jiném ekonomickém světle.“ (Mobley, 2002)

Literatura týkající se tématu údržby uvádí ***tři různé druhy***:

- ***Reaktivní údržba***, jedná se o údržbu, která spočívá v opravě systému pouze po poruše.
- ***Preventivní údržba***, spočívá v pravidelném udržování systému, tak aby nedocházelo k poruchám. Délka intervalu kontrol je stanovená na základě statisticky poruch. Například po každých 100 hodinách provozu je nezbytné vyměnit těsnění.
- ***Prediktivní údržba***, která spočívá v tom, že se údržba provádí jen v tom případě, kdy to systém vyžaduje tzn. Když je potencionální porucha detekována.

18.7.1 Úvod do prediktivní údržby

Průmyslové stroje generují denně velké množství dat, které firmy skladují na jednom místě. V této záplavě dat se ovšem skrývají cenné informace, které mohou firmám ušetřit nemalé prostředky. Identifikace série datových bodů, která označuje, že stroj za miliony korun se chystá rozbít, může pomoci ***zabránit případným prostojům ve výrobě, dodatečným nákladům*** a dlouhodobému poškození. Nicméně bez automatizovaného řešení, je to jako hledání jehly v kupce sena.

S rozšířením IoT senzorů a nástupu cloudových řešení se přímo nabízelo využít tyto dnešní technologie i ve výrobě a ***pomocí strojového učení zavést prediktivní údržbu***, která by podniku usnadňovala plánování odstávky stojů a tím pomůže snížit náklady. Tím pádem správný systém sběru dat nemusí být pouhou investicí, ale i naopak možností, jak ušetřit nemalé prostředky, které se dají následně využít i jinak.

Koncept prediktivní údržby není nikterak nový. Každý z techniků už dnes chápe, že spoléhat se na reaktivní údržbu strojů či zavést pouze plán na preventivní údržbu v dnešní době nestačí. ***Zavádění prediktivní údržby*** do výroby má ***řadu úskalí***, např. ne u každého stroje má smyl predikovat jeho chování. S účinným pohledem na stav strojů ***je možné předcházet problémům, než se stanou závažnými***, a tím potencionálně ušetřit až miliony korun ročně na nákladech na opravy a zároveň prodloužit životnost.

Pokud bychom na příklad implementovaly ***postupy z prediktivní údržby***, založené na efektivní prediktivní analytice, inženýr může s jistotou říci, ***které díly budou potřeba vyměnit a kdy***. Selhání je téměř vždy ***výsledkem řetězce událostí***. Jak jeden problém vede k jinému, je vytvořen digitální signál – ten je považován za jakýsi příznak nemoci. U strojů připojených k IIoT jsou tyto příznaky rozptýleny v milionech datových bodů, pocházejí z různých senzorů v různých časech a jsou uloženy v samostatných silech. Nalezení kritického signálu uprostřed těchto milionů rozptýlených datových bodů není lidsky možné.

18.7.2 Křivka vzniku potencionální poruchy (P-F)

V souvislosti s lepším pochopením ***metodiky prediktivní údržby*** a jejího ***rozdílu od preventivní údržby*** je důležité pochopit ***P-F křivku*** vzniku potencionální poruchy (P-F je převzato z angličtiny, kdy P znamená „Potential“ = potencionální a F znamená „Failure“ = selhání, porucha).

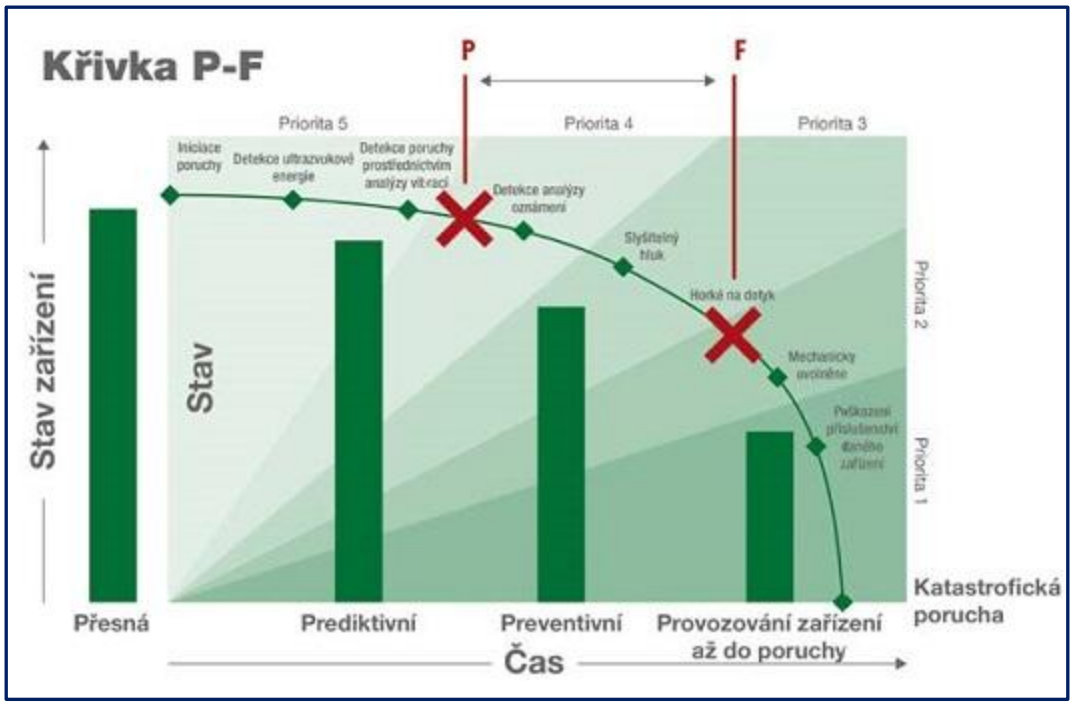
„***P***“ v P-F křivce ***označuje potencionální selhání***, kdy by zařízení na základě historických dat mohlo selhat, nebo by se mohly objevit ***první známky poškození***, které by mohlo vést k poruše. Jako příklad lze použít sledování teploty ložisek, na základě dlouhodobého měření se zjistilo, že pokud teplota ložiska přesáhne 60 stupňů dojde s vysokou pravděpodobností k poruše.

Naopak „***F***“ ***označuje funkční selhání přístroje***, kdy přístroj opravdu selže. Pokud budeme pokračovat se stejným příkladem je zřejmé, že obvykle jsou přibližně čtyři dny mezi okamžikem, kdy teplota ložiska překročí 60 stupňů (P) a kdy selže (F). Proto by se mělo kontrolovat v tomto intervalu (možná každé dva dny). Tím ***se zjistí, že dojde k selhání dříve***, než k němu skutečně dojde.

„Oprava porouchaného přístroje není tak jednoduchým úkonem jako je provádění pravidelné inspekce ve stanoveném intervalu. Tato metoda předpokládá, že ***selhání je poměrně náhodná a***

nepředvídatelná událost způsobená působením několika faktorů. ***Pokud*** jsou, ale ***správně stanoveny známky selhání a správně určená doba***, která následovala po indikování potencionálního selhání, lze ***stanovit P-F interval***. Je snahou, aby byl tento interval ***co možná nejdelší s méně častými, ale více přesnými kontrolami***. Je nezbytné, aby rozmezí kontrol bylo menší, než je P-F interval, aby bylo možné zachytit poruchu poté co byla detekována, ale dříve, než k ní skutečně dojde.“ (Bellstedt, 2019)

Osa X na grafu P-F křivky ***znázorňuje čas***, svislá ***osa Y představuje stav zařízení*** (viz další obrázek). K degradaci daného přístroje dochází postupně, ***potencionální porucha (P) představuje bod na křivce***, kdy je poprvé možné zaznamenat, že hrozí porucha daného zařízení. Pokud vada zůstane neodstraněna nebo neodhalena, ***proces degradace se urychluje***, až se zařízení dostane ***do bodu funkční poruchy (F)***. Tento interval P-F může být rozdílně dlouhý, mohou to být dny, týdny dokonce až i měsíce.



Obrázek 18-3 Křivka P-F; Zdroj: (Karchová, 2017)

18.7.3 Prediktivní údržba pomáhá optimalizovat plánovanou dobu odstávky

Plánované odstávky mohou zahrnovat vše od čištění, promazaní stroje až po výměnu součástek, o kterých je známo, že často selhávají. Tento ***druh preventivní údržby snižuje riziko neplánovaných odstávek***. Díky datům shromážděným ve strojních operacích ***lze preventivní údržbu naplánovat pravidelně*** a v době, která bude mít minimální dopad na výrobu zakázek. Je zde také přidaná výhoda, že adekvátní údržba této povahy ***trvale prodlouží životnost*** stroje, který by byl obtížný a nákladný nahradit. Výsledkem maximalizace doby provozuschopnosti a životnosti komponenty bude nakonec výrazná úspora nákladů.

„Vzhledem k tomu, že plánovaná preventivní údržba může zajistit, že stroje budou běžet hladce, ***monitorovací stroje*** digitálně shromažďují ***množství dat, která při analýze zobrazí vzory*** na daném stroji. Tento druh detekce vzorů, založený na historických datech, může pomoci i***dentifikovat přístroj, u něhož je pravděpodobné, že nastane výpadek***, a pro který lze proaktivně plánovat údržbu.“ (Immerman, 2018)

Díky možnosti sledovat efektivitu, výkon a kvalitu stroje v průběhu času budou objevena data, která budou identifikovat, kdy bude stroj vyžadovat údržbu, jak už bylo uvedeno výše, ale ***také pomůže identifikovat, kdy stroj dosáhne konce své životnosti***.

18.7.4 Prediktivní údržba může pomoci optimalizovat produktivitu zaměstnanců.

„Existuje mnoho způsobů, jak prediktivní údržba může optimalizovat produktivitu zaměstnanců. Nejprve je nutné se podívat na náklady samotné práce. Když jsou naplánovány opravy, ***doba potřebná k opravě se sníží*** kvůli menšímu počtu vyměněných dílů namísto, aby bylo vyměněno celé zařízení. Zároveň je i ***snížen počet hlášených kritických selhání*** strojů, a počet urgentních požadavků na opravu." (Immerman, 2018)

18.7.5 Zvýšení příjmů

Díky efektivní údržbě, která neplýtvá zdroji na opravě fungujících komponent a rychlejší opravou vadných komponent mohou být ***opravy efektivněji zpracovány***, čímž se ***zkracuje doba opravy***. V roce 2015 provedla společnost McKinsey studii o potenciálu průmyslové analytiky, která odhalila možnost následujících vylepšení:

- „***Snížení nákladů na údržbu o 10-40 %.*** Vzhledem k tomu, že většina údržby je zatím plánovaná, tudíž existují případy, kdy jsou prováděny zbytečné opravy, které ještě nebyly zdaleka potřeba. (Burghin Jacques, 2016)
- ***Snížení množství odpadu o 10-20 %.*** Neoptimalizované procesy, které zatím nebyly odhaleny, mohou vést k nehospodárné výrobě. V takových případech dochází k plýtvání surovinami, energií, pracovními náklady a strojním časem. Systémy prediktivní údržby mohou odhalit problémy, které mohou mít za následek vznik odpadu. (Burghin Jacques, 2016)
- Jakmile se sběr dat stane automatizovaným, ***počet optimalizací procesů může vzrůst denně až o 10-50 %,*** právě prostřednictvím pokročilých analytických procesů" (Burghin Jacques, 2016)

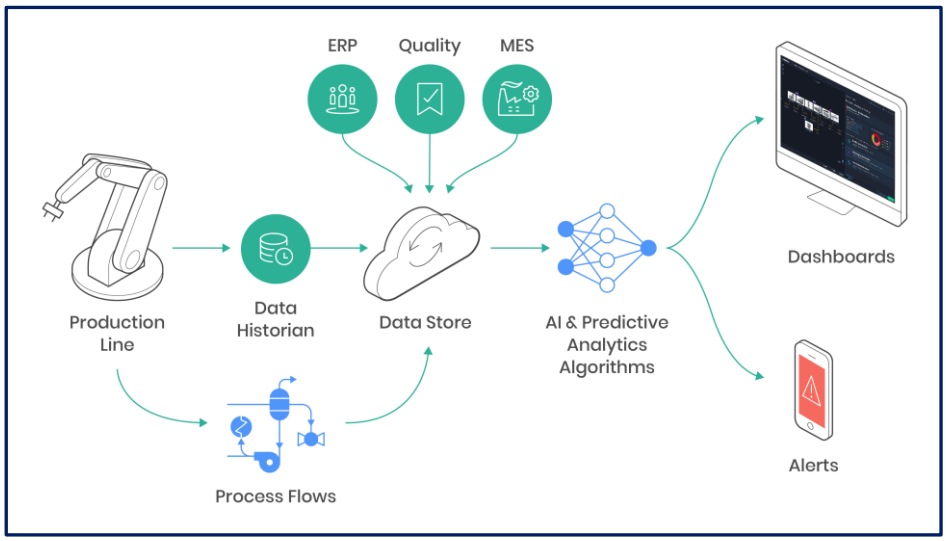
18.7.6 Principy preventivní údržby

Prediktivní údržba je v podstatě ***zdokonalená preventivní údržba řízená podmínkami***. Prediktivní údržba ***sleduje mechanický stav, účinnost zařízení a další parametry*** a pokouší se odvodit přibližnou dobu funkční poruchy. ***Komplexní program prediktivní údržby*** využívá kombinaci nákladově nejefektivnějších nástrojů pro získání skutečných provozních podmínek zařízení a výrobních systémů. Na základě těchto ***shromážděných údajů jsou vybrány plány údržby***. Prediktivní údržba používá různé techniky, jako je analýza vibrací, analýza nečistot v oleji a opotřebení, ultrazvuk, termografie, hodnocení výkonu a další techniky pro posouzení stavu zařízení.

U technik prediktivní analytiky lze spatřovat ***analogii k lékařským diagnostickým technikám***. Protože kdykoliv má lidské tělo problém projevuje se nějakým symptomem, který je možný zaznamenat. Nervový systém poskytuje informace – to je fáze detekce. Dále jsou v případě potřeby provedeny patologické testy pro diagnostiku problému. Na tomto základě se doporučuje vhodná léčba.

Podobným způsobem lze sledovat poškození na stroji, ***když se začne opotřebovávat, začnou se projevovat symptomy*** jako jsou například vibrace nebo jiný druh signálu. Tyto symptomy však nemusí být postřehnutelné lidskými smysly, proto se na to v dnešní době ***nasazují systémy IIoT***. Právě zde přicházejí na pomoc techniky prediktivní údržby. Tyto techniky ***detekují symptomy defektů***, které se vyskytly ve strojích, a pomáhají při diagnostice přesných defektů, které se vyskytly. V mnoha případech je také ***možné odhadnout závažnost závad***.

„Konkrétní použité ***techniky závisí na typu výrobního zařízení***, ***jejich dopadu*** na výrobu nebo na jiné klíčové parametry provozu zařízení. Aby byl dosažen ideální stav fungování prediktivní údržby (viz další obrázek) je nezbytné ***do procesu predikce vložit co možná nejvíce relevantních informací***, tedy jak přímo naměřené provozní hodnoty každého přístroje, tak i všechny dostupné výsledky ***jiných zařízení podobných vlastností***, ideální by bylo pak i nasbírat naměřené hodnoty zařízení podobného typu po celém světě, pro správnou predikci je zkrátka nezbytné operovat s velkým množstvím relevantních dat. Z toho je tedy patrné, že ideálně fungující prediktivní údržbu ***nelze vykonávat na úrovni výpočetních jednotek jednotlivých přístrojů či linek***, ale ideálně propojením všech strojů na centrální datové úložiště, kam se budou odesílat všechna naměřená data a následně se zde bude provádět i analýza a vyhodnocení těchto dat." (Vojáček, 2018)



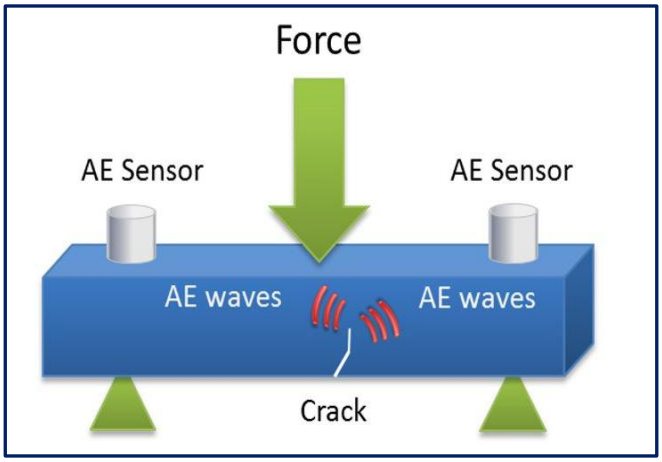
Obrázek 18-4 Schéma prediktivní údržby; Zdroj: (Seebo, 2018)

18.7.7 Techniky prediktivní údržby

Existují četné techniky prediktivní údržby:

a) ***Monitorování vibrací***: Jedná se o nejefektivnější techniku pro detekci závad v rotujících strojích

b) ***„Akustická emise: Slouží k detekci a lokalizaci trhlin*** (viz další obrázek). Někdy označována jako pasivní ultrazvuk. Výhodou této metody je, že se jedná o objemovou metodu. V reálném čase je zkoušeno celé těleso (například potrubí). Metoda dokáže zachytit prvotní porušení struktury a tím je možné předejít případnému poškození. Na základě fyzikálních principů šíření vlny lze lokalizovat zdroj, který tyto vlny generuje.“ (Solař, 2018)



Obrázek 18-5 Princip akustické emise; Zdroj: (Solař, 2018)

c) ***Analýza oleje a částic***: V oleji je zkoumaná přítomnost mikroskopických částic, které mohou indikovat opotřebení ložisek a ozubených kol. Opotřebované součásti strojů, ať už v pístových strojích, převodovkách nebo hydraulických systémech, uvolňují nečistoty. Sběr a analýza těchto nečistot poskytuje důležité informace o zhoršení těchto složek.

d) ***Monitorování koroze***: Měření ultrazvukové tloušťky se provádí na potrubích, pobřežních konstrukcích a technologických zařízeních pro sledování výskytu korozního opotřebení.

e) ***Termografie se používá k analýze aktivního elektrického a mechanického zařízení.*** Metoda může detekovat tepelné nebo mechanické vady v generátorech, nadzemních vedeních,

kotlích, nesouosých spojkách a mnoha dalších vadách. Může také detekovat poškození buněk ve strukturách uhlíkových vláken na letadlech.

f) „***Sledování výkonnosti***: Jedná se o velmi účinnou techniku pro určení provozních problémů zařízení. Účinnost strojů poskytuje dobrý přehled o jejich vnitřních podmínkách.“ (Girdhar, 2004)

„Navzdory všem těmto metodám jsou známé ***příklady, kdy prediktivní údržba nebyla schopna přinést hmatatelné výsledky*** pro organizaci. Příčinou byla většinou nedostatečná podpora managementu, špatné plánování nebo nedostatek zkušených a vyškolených zaměstnanců.“ (Friedman, 2019)

„Po začlenění procesů prediktivní údržby do provozu je velmi důležité ***rozhodnout o specifických metodách monitorování výrobních zařízení***. Metody jsou také závislé na typu průmyslového odvětví, typu strojního zařízení a do značné míry i dostupností kvalifikované pracovní síly. Techniky prediktivní údržby ***vyžadují technicky sofistikované nástroje*** pro provádění detekce a diagnostiky strojního zařízení. Tyto nástroje jsou obecně velmi drahé a potřebují technicky kvalifikované lidi k analýze jejich výstupu. Výše počátečních nákladů na pořízení sofistikovaných nástrojů, softwaru anebo náboru a vyškolení kvalifikované pracovní síly často vedou k otázce, jestli se to společnosti vyplatí. Díky podpoře managementu, adekvátním investicím do lidí a vybavení však může prediktivní údržba po krátké době přinést velmi dobré výsledky.“ (Girdhar, 2004)

18.7.8 Omezení prediktivní analytiky

Bezpečnost je jednou z hlavních překážek, která brání v masovém nasazení do výroby. Pokud by byla prediktivní údržba byla nasazena v celém svém rozsahu znamenalo by to pro společnosti poskytovat velké množství detailních výrobních dat externí firmě, která se stará o cloudové úložiště. Ne každá společnost si může dovolit postavit vlastní datové centrum. U menších společností se proto nabízí ***model pronájmu takovéhoto úložiště***, ale je potřebné zvážit určitá bezpečností rizika, protože se jedná o sdílení vysoce citlivých dat z výroby v reálném čase.

18.7.9 Zdroje

Bellstedt, Sarah. 2019. What is the P-F Curve? | P-F Interval. *Fiix.* [Online] Fiix Inc., 12. Únor 2019. [Citace: 10. Duben 2019.] https://www.fiixsoftware.com/blog/what-is-the-p-f-curve-p-f-interval/.

Burghin Jacques, Nicolaus Henke, Michael Chui. 2016. THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD. *McKinsey Global Institute.* [Online] Prosinec 2016. [Citace: 15. Březen 2019.] https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx.

Friedman, Alan. 2019. Why Do Predictive Maintenance Programs Fail? *Reliabilityweb.* [Online] Reliabilityweb, 2019. [Citace: 1. Duben 2019.] https://reliabilityweb.com/articles/entry/why_do_predictive_maintenance_programs_fail/.

Girdhar, Paresh. 2004. *Practical Machinery Vibration Analysis and Predictive Maintenance.* Burlington : Elsevier, 2004. 0750662751.

Immerman, Graham. 2018. The Impact of Predictive Maintenance on Manufacturing. *MachineMetrics Blog.* [Online] MachineMetrics, 24. Červenec 2018. [Citace: 1. Duben 2019.] https://www.machinemetrics.com/blog/the-impact-of-predictive-maintenance-on-manufacturing.

Karchová, Barbora. 2017. Údržba podniku. *Údržba podniku.* [Online] Září 2017. [Citace: 9. Duben 2019.] http://udrzbapodniku.cz/fileadmin/grafika/Barca_Karchova/casopis_zari_2017/us.jpg.

Mobley, R. Keith. 2002. *An Introduction to Predictive Maintenance.* Knoxville : Elsevier, 2002. 9780080478692.

Sanghavi, Sundeep. 2016. Revolutionizing Manufacturing With Predictive Maintenance Analytics. *Manufacturing Business Technology; Rockaway.* [Online] 22. Srpen 2016. [Citace: 15. Únor 2019.] Revolutionizing Manufacturing With Predictive Maintenance Analytics. 15543404.

Solař, Josef. 2018. Metoda akustické emise. *PTS Josef Solnař s.r.o.* [Online] PTS, 2018. [Citace: 10. Bžezen 2019.] https://www.ptsndt.com/cs/prodej/at-zkouseni-akustickou-emisi.

The Wall Street Journal. 2018. How Manufacturers Achieve Top Quartile Performance. *The Wall Street Journal.* [Online] The Wall Street Journal, 24. Červenec 2018. [Citace: 25. Březen 2019.] https://partners.wsj.com/emerson/unlocking-performance/how-manufacturers-can-achieve-top-quartile-performance/.

Vojáček, Antonín. 2016. Co se skrývá pod výrazy Industry 4.0 / Průmysl 4.0 ? *Automatizace.hw.cz.* [Online] 19. Březen 2016. [Citace: 1. Duben 2019.] https://automatizace.hw.cz/mimochodem/co-je-se-skryva-pod-vyrazy-industry-40-prumysl-40.html.

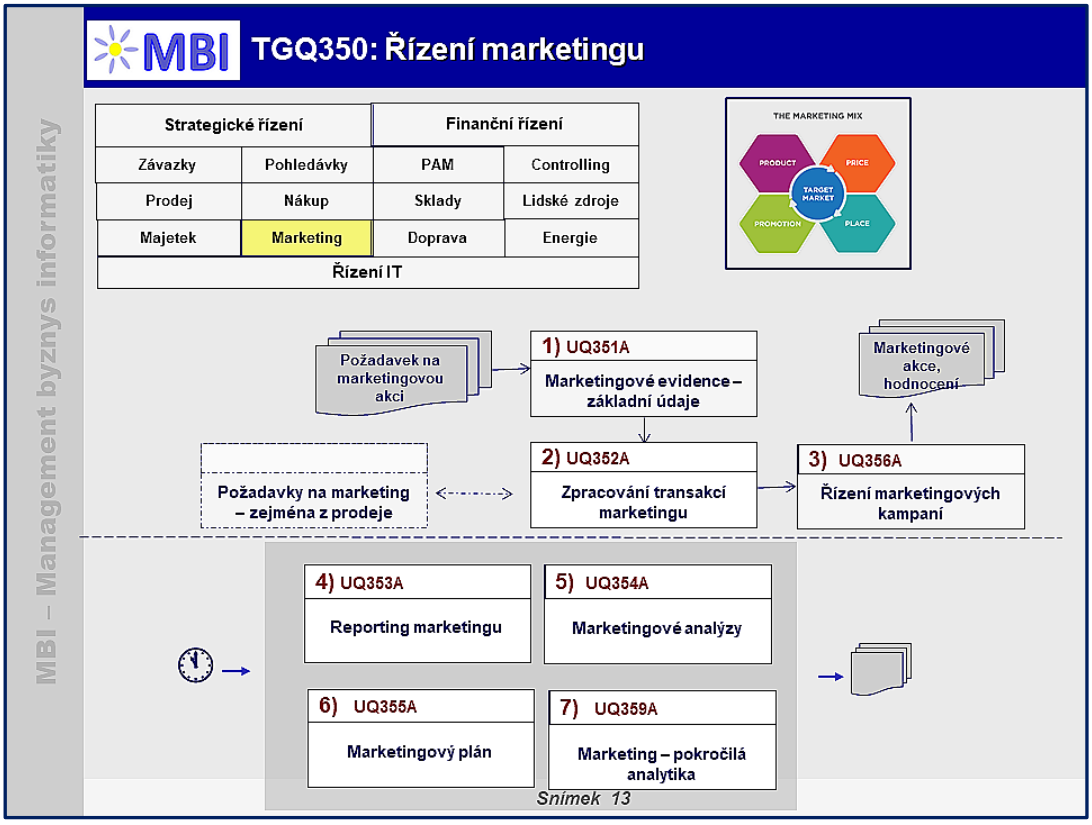
19. Prediktivní analytika v řízení marketingu

Účelem je:

- podporovat ***obchodní aktivity*** firmy,
- přispívat k ***posilování pozice firmy*** na trhu, její konkurenceschopnosti a získávání konkurenčních výhod,
- poskytovat pracovníkům marketingu, prodeje, nákupu (manažerům, obchodníkům) kvalitní ***analytické a plánovací podklady*** pro rozhodování o celkové orientaci obchodních aktivit firmy,
- realizovat ***úspěšné marketingové akce a kampaně*** s vysokou návratností.

19.1 Vymezení obsahu oblasti řízení

Celkový pohled na úlohy řízení marketingu obsahuje další obrázek:



Obrázek 19-1: Marketing – přehled úloh

Do řízení marketingu spadají ***tyto úlohy***:

- ***Marketingové evidence*** – aktualizace základních údajů marketingu, např. klasifikace obchodních partnerů, konkurence apod.
- ***Transakce*** – řízení marketingu, zpracování jednotlivých marketingových transakcí spojených s přípravou realizací a vyhodnocením marketingových akcí.
- ***Reporting*** marketingu.
- ***Marketingové analýzy,*** tj. analýzy trhu, konkurence, požadavků zákazníků apod.
- ***Plánování*** – marketingový plán, plánování marketingových a promo akcí.
- ***Řízení marketingových kampaní*** – operativní řízení jednotlivých marketingových kampaní.

- ***Marketing – pokročilá analytika***.

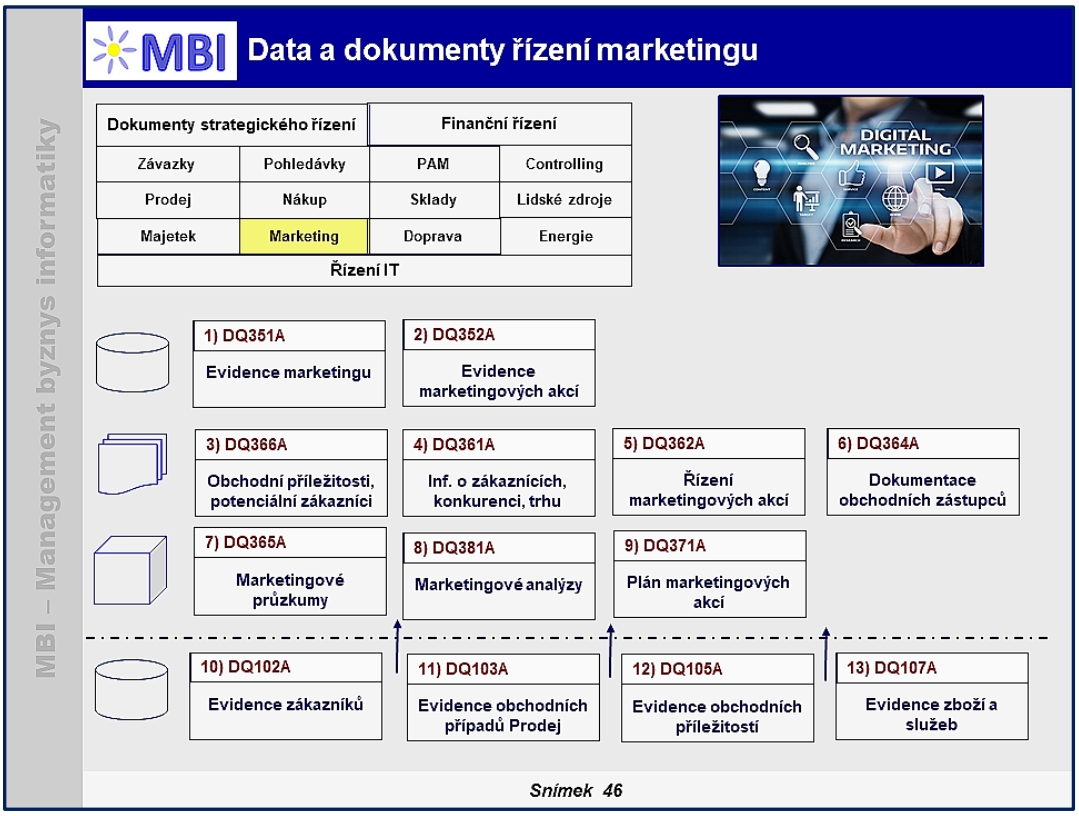


Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***12***.

19.2 Datové zdroje

Souhrnný přehled dat a dokumentů řízení marketingu dokumentuje další obrázek. ***Datové zdroje*** představují zejména:

- Evidence marketingu.
- Evidence marketingových akcí.



Obrázek 19-2: Vstupy a výstupy úloh řízení marketingu



Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.12***.

19.3 Účel prediktivní analytiky v řízeni marketingu

Prediktivní analytika v řízení marketingu má podporovat následující aktivity:

- formulaci tržní diagnózy – tržní situace firmy, analýza skupin zákazníků, segmentů trhu, prodejních možností,
- tržní prognózu – odhad trendů na trhu, změny trhu, pozice konkurence,
- plánování cílů marketingu – zaměření marketingových akcí,
- plánování marketingového mixu – plánování v oblasti výrobkové, cenové, distribuční, komunikační,
- sestavení marketingového rozpočtu.

- Získávání nových zákazníků.
- Prognózy obchodních příležitostí.
- Cross-sell / Up-sell úlohy.
- Produkty, které zákazníci budou napříště kupovat.
- Navštěvování webových stránek.
- „*Market-basket*“ analýza.
- Hodnota zákazníka.
- Profitabilita zákazníka.
- Segmentace zákazníků.
- A další.
-

19.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables“*) v oblasti řízení marketingu obsahuje další přehled:

- ***Pozice firmy na trhu*** vyjadřuje podíl firmy na daném segmentu, resp. segmentech trhu v %.
- ***Úspěšnost prodejní kampaně*** je úspěšnost kampaně v procentech a je stanovena jako poměr počtu zákazníků, kterým jsou prodány nové služby, děleno počtem všech oslovených zákazníků v kampani.
- ***Marketingové náklady na objednávku*** je podíl celkových marketingových výdajů na celkovém počtu objednávek ukazuje náročnost investic do marketingu, aby byla vygenerována jedna objednávka.
- ***Podíl marketingu na celkových tržbách*** je procentuální podíl marketingových výdajů na tržbách

19.5 Jednotky analýzy

Jednotky analýzy („*unit of analysis*“) mohou v případě řízení marketingu zahrnovat:

- Marketingová kampaň
- Zákazník.
- Segment trhu.
- Objednávka.
- Region.

19.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace zákazníků – jejich rozřazení do stejnorodých tříd a odhad pravděpodobnosti správnosti takového zařazení pro každého zákazníka,
  - klasifikace prodávaných zboží – obvykle podle katalogu zboží,
  - klasifikace marketingových akcí – posouzení správnosti klasifikace uvedením pravděpodobnosti správného členění.
- *Přiřazování podobností:*
  - určování podstatných charakteristik zákazníků a na tomto základě výběr firem, které jsou obdobné našim nejlepším zákazníkům,
- *Profilování, „Popis chování“:*
  - specifikace klíčových charakteristik zákazníků z pohledu přístupů k zajištění a řízení obchodních případů Prodej,
- *Predikce vazeb:*

- vyhodnocování vazeb marketingových aktivit vzhledem k dosahovaným obchodním výsledkům.

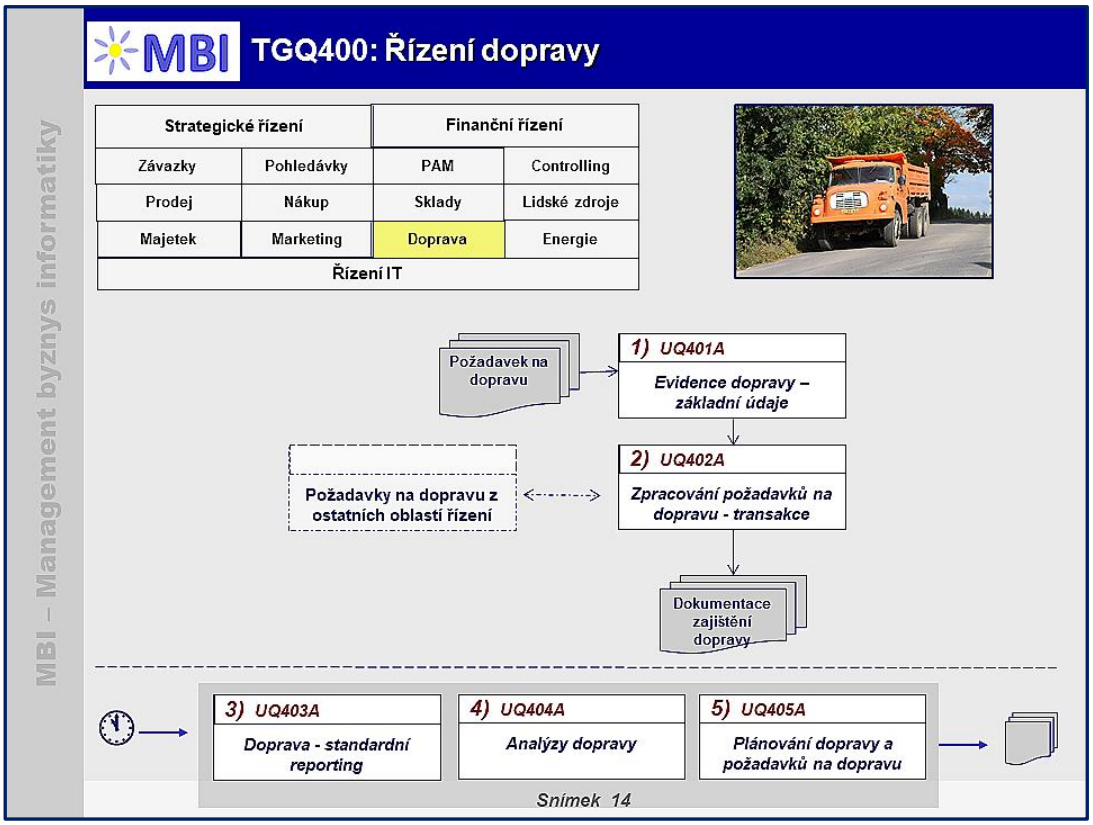
20. Prediktivní analytika v řízení dopravy

Účelem úlohy je:

- dosahovat co nejlepší zajištění ***dopravních potřeb firmy*** při optimalizaci nákladů na dopravu,
- zajistit efektivní poměr v zajištění dopravních požadavků ***vlastními prostředky a externími dopravci***,
- poskytovat pracovníkům v oblasti řízení dopravy ve firmě kvalitní ***analytické a plánovací podklady*** pro rozhodování o disponibilních vlastních a externích dopravních kapacitách, dopravcích a jejich kvalitě a spolehlivosti.

20.1 Vymezení obsahu oblasti řízení

Celkový přehled úloh řízení dopravy obsahuje další obrázek:



Obrázek 20-1: Řízení dopravy – přehled úloh

Do řízení dopravy spadají ***tyto úlohy***:

- ***Evidence dopravy*** – aktualizace základních údajů o dopravcích, dopravních prostředcích, dopravních trasách apod.
- ***Transakce*** – řízení dopravy, příjem požadavku na dopravu, zpracování požadavků na dopravu apod.
- ***Reporting dopravy*** – vytváření standardních reportů dopravy a zajištění požadavků na dopravu.
- ***Dopravní analýzy*** – analýzy dopravy a využití dopravních prostředků.
- ***Plánování dopravy*** – plánování požadavků na dopravu, rozvoje dopravních kapacit.

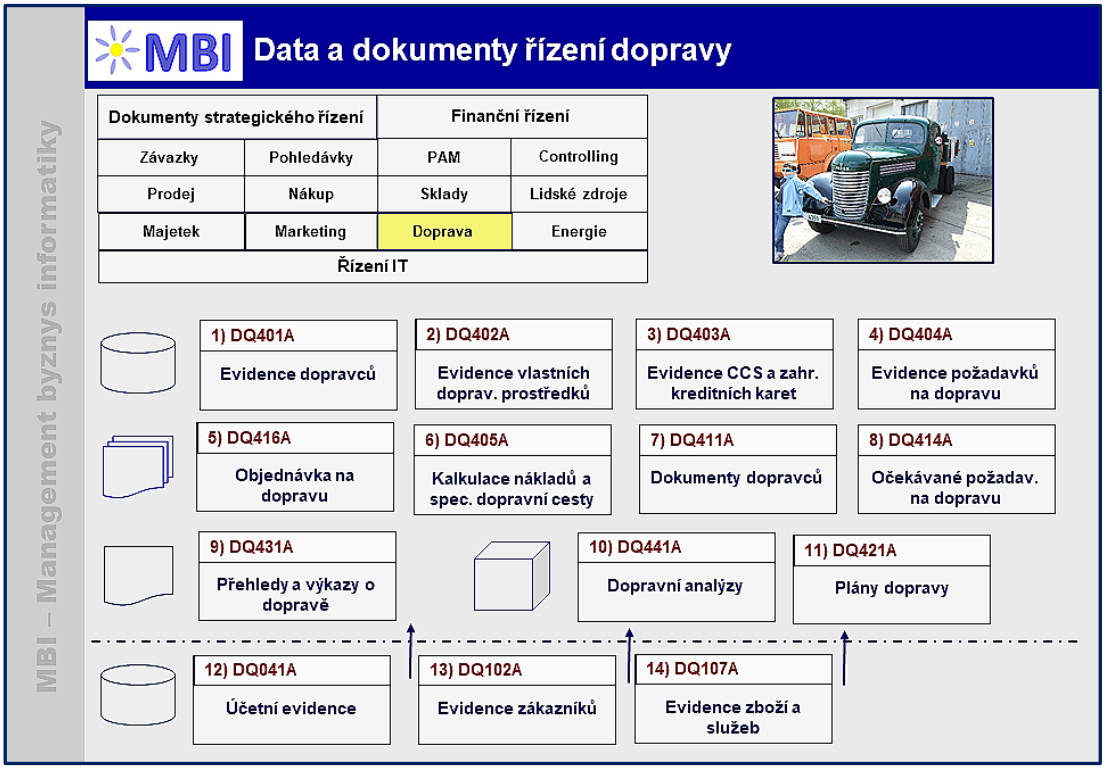


Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***13***.

20.2 Datové zdroje

Souhrnný přehled dat a dokumentů řízení dopravy prezentuje další obrázek. Datové zdroje představují zejména:

- Evidence dopravců.
- Evidence vlastních dopravních prostředků.
- Evidence CCS a zahraničních kreditních karet.
- Evidence požadavků na dopravu.



Obrázek 20-2: Souhrnný přehled dat a dokumentů řízení dopravy



Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.13***.

20.3 Účel prediktivní analytiky v řízeni dopravy

Prediktivní analytika v řízení dopravy firmy má podporovat následující aktivity:

- ***zpracování plánů pro řízení a optimalizaci dopravy***
- zpracování ***požadavků na dopravu*** podle různých hledisek (dimenzí) pro různé časové horizonty

20.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables*“) v oblasti řízení dopravy obsahuje další přehled:

- ***Náklady na logistiku*** jsou náklady na zajištění celého procesu logistiky ve firmě.

- ***Náklady na dopravu*** jsou náklady na dopravu podle jednotlivých přepravních a dopravních zakázek + pojistné
- ***Výkon dopravy*** vyjádřený v tunokilometrech
- ***Podíl využití nákladního auta*** počítá se jako podíl váhy nakládaného zboží na celkové kapacitě.

20.5 Jednotky analýzy

Jednotky analýzy („*unit of analysis*“) mohou v případě řízení dopravy zahrnovat:

- Poskytovatel dopravy.
- Požadavek na dopravu.
- Dopravní prostředek.
- Region.

20.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace poskytovatelů dopravy – jejich rozřazení do stejnorodých tříd a odhad pravděpodobnosti správnosti takového zařazení pro každého dodavatele,
  - klasifikace vlastních dopravních prostředků,
  - klasifikace vlastních a nakupovaných dopravních služeb.
- *Přiřazování podobností:*
  - určování podstatných charakteristik poskytovatelů dopravy a na tomto základě výběr firem, které jsou obdobné našim nejlepším poskytovatelům.
- *Profilování,* „*Popis chování*“*:*
  - specifikace klíčových charakteristik poskytovatelů z pohledu přístupů k zajištění a řízení požadavků na dopravu.
- *Predikce vazeb:*
  - vyhodnocování vazeb včasnosti dodávek vzhledem k obchodním výsledkům firmy.

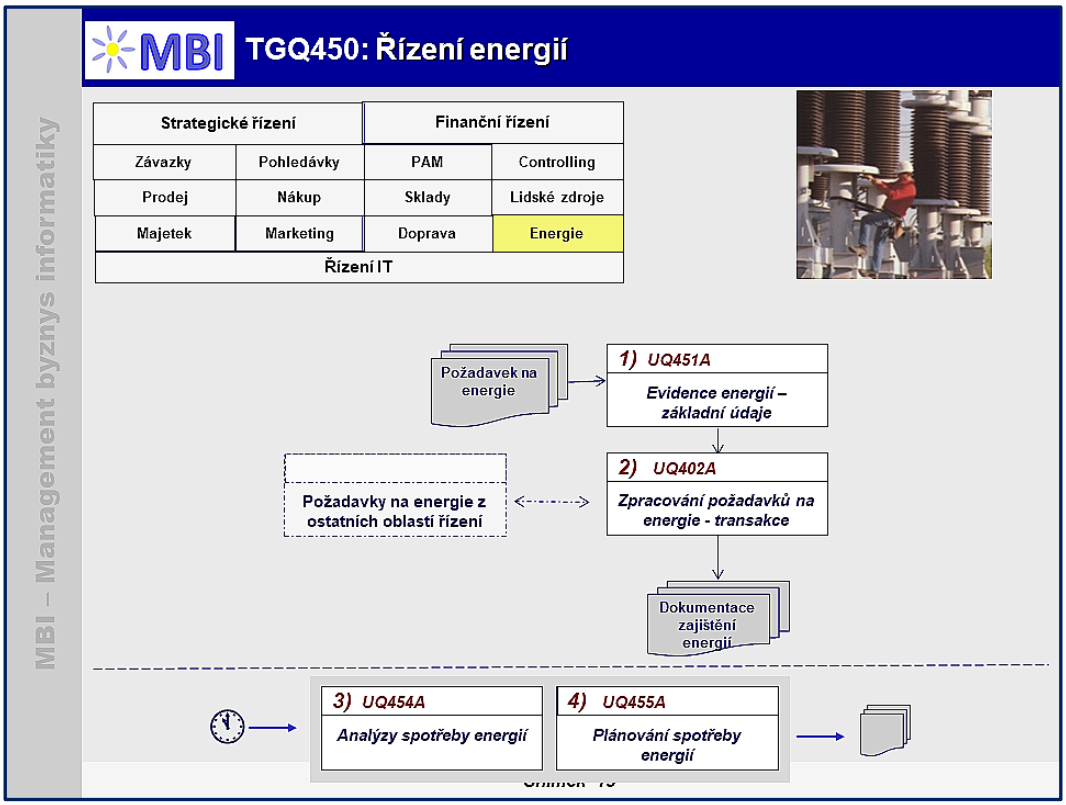
21. Prediktivní analytika v řízení energií

Účelem úloh je:

- dosahovat efektivního ***plánování potřeb a řízení spotřeby energií***,
- poskytovat pracovníkům v oblasti řízení energií v firmy kvalitní ***analytické a plánovací podklady*** pro rozhodování o energetických zdrojích a jejich poskytovatelích.

21.1 Vymezení obsahu oblasti řízení

Celkový pohled na úlohy řízení energií obsahuje další obrázek:



Obrázek 21-1: Řízení energií – přehled úloh

Do řízení energií spadají ***tyto úlohy***:

- ***Evidence energií*** – aktualizace základních údajů o spotřebičích, zdrojích apod.
- ***Transakce*** – řízení energií, příjem požadavku na energie, odečet spotřeby.
- ***Analýzy*** – spotřeby energií.
- ***Plánování*** – spotřeby energií a energetických zdrojů.

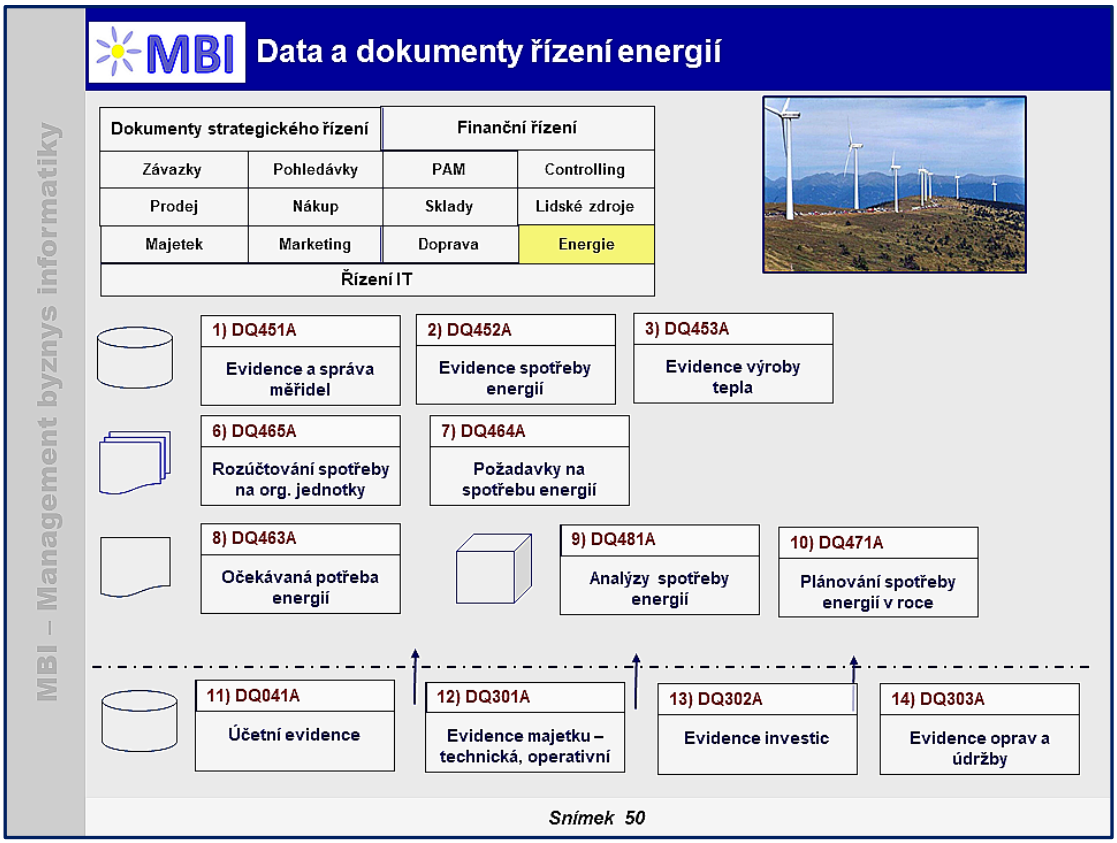
→ Detailněji – v dokumentu „*AF II.01: Oblasti*“, kapitola ***14***.

21.2 Datové zdroje

Souhrnný přehled dat a dokumentů řízení energií prezentuje další obrázek. ***Datové zdroje*** představují zejména:

- Evidence a správa měřidel.

- Evidence spotřeby energií.
- Evidence výroby tepla.



Obrázek 21-2: Souhrnný přehled dat a dokumentů řízení energií

→ Detailněji – v dokumentu „*AF II.02: Komponenty*“, kapitola ***4.14***.

21.3 Účel prediktivní analytiky ve strategickém řízení

Prediktivní analytika v řízení energií má poskytovat podporu v těchto plánovacích aktivitách:

- zpracování ***plánů řízení spotřeby energií***
- hodnocení ***požadavků na energie*** podle různých hledisek (dimenzí) pro různé časové horizonty

21.4 Cílové proměnné

Náměty na specifikaci cílových proměnných („*target variables*“) v oblasti řízení energií obsahuje další přehled:

- ***Objem spotřeby elektrické energie*** sleduje a rozlišuje spotřebu elektrické energie v MWh, např. podle útvarů, poskytovatelů, případně měřidel.
- ***Objem spotřeby zemního plynu*** v GJ, resp. MWh.
- ***Objem spotřeby nafty*** v litrech.
- ***Objem spotřeby vody*** v m^3***.***
- ***Objem spotřeby tepla,*** metrika sleduje spotřebu tepla v prostorách firmy a podle regionů včetně sezónních výkyvů.

- ***Náklady na energie*** podle druhů v tisících Kč.
- Objem ***vlastní výroby*** el. energie v MWh.

21.5 Jednotky analýzy

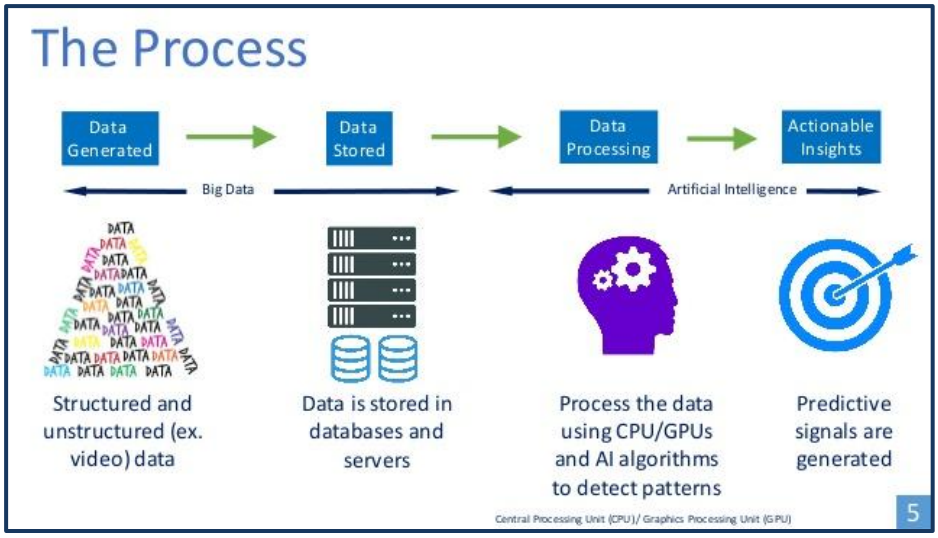
Jednotky analýzy („*unit of analysis*“) mohou v případě řízení energií zahrnovat:

- Druhy energií.
- Dodavatel energií.
- Požadavek na energie.
- Region.

21.6 Prediktivní funkcionalita

- *Klasifikace*:
  - klasifikace dodavatelů energií – jejich rozřazení do stejnorodých tříd a odhad pravděpodobnosti správnosti takového zařazení pro každého dodavatele,
  - klasifikace využívaných energií,
  - klasifikace vlastních zdrojů a výroby energií.
- *Přiřazování podobností:*
  - určování podstatných charakteristik dodavatelů energií a na tomto základě výběr firem, které jsou obdobné našim nejlepším dodavatelům.
- *Profilování,* „*Popis chování*“*:*
  - specifikace klíčových charakteristik dodavatelů energií z pohledu přístupů k zajištění a řízení požadavků na energie.
- *Predikce vazeb:*
  - vyhodnocování vazeb kvality dodávek energií vzhledem k obchodním výsledkům firmy.

D) Kontext, funkce, metody prediktivní analytiky



[22] Prediktivní analytika v kontextu podnikové analytiky ***(BI, CI, data science, data mining, text mining, machine learning, big data, statistika)***	**[23] Funkce prediktivní analytiky** ***(clustering, klasifikace, asociační pravidla, přiřazování podobnosti, seskupování podle výskytů, predikce vazeb, redukce dat, Kauzální modelování)***
[24] Metody prediktivní analytiky ***(regresní analýzy, rozhodovací stromy, neuronové sítě, Bayesovské klasifikace)***	

22. Prediktivní analytika v kontextu podnikové analytiky



Jak již bylo výše uvedeno, je prediktivní analytika podstatnou ***součástí podnikové analytiky*** a podnikového řízení. Podniková analytika představuje v současné době celou ***škálu disciplín, metod, přístupů a nástrojů***, které se různě ***doplňují***, ale i ***překrývají*** a zejména se velmi rychle významně ***rozvíjejí***. To je důvodem i pro to, že najít nějaké sjednocující schéma s potřebnými vazbami je dnes velmi obtížné.

Účelem kapitoly je specifikovat ***alespoň rámcově podstatu těchto disciplín, metod a nástrojů*** tak, aby bylo zřejmé kde a jak prediktivní analytiku v komplexu těchto ostatních disciplín, metod a nástrojů v praxi řešit a v řízení využívat. Jejich ***detailnější charakteristiky*** jsou náplní ***dokumentu*** „*AF_II_05_Podniková_analytika.pdf*“. V této kapitole je uvedeno pouze stručné shrnutí a vazby na prediktivní analytiku.

Pro vymezení kontextu prediktivní analytiky ***v rámci podnikové analytiky*** byly do dalších podkapitol zařazeny tyto její součásti:

- ***Machine Learning*** obsahuje vymezení podstaty a kroků řešení úloh ML, vymezení možností produkcionalizace a governance řešení.
- ***Business Intelligence (BI)***, tj. řešení založená na celém spektru velmi výkonných komponent (zejména databázových systémů, transformačních nástrojů, analytických a dalších nástrojů).
- ***Competitive Intelligence (CI)*** se zabývá sběrem, zpracováním a ochranou informací s cílem získat konkurenční výhodu pro firmu.
- ***Data Science*** zahrnuje porozumění business logiky dat, přípravu dat, modelování / optimalizace / simulace, vyhodnocení a nasazení analytického modelu.
- ***Data Mining,*** dolování dat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných informací z velmi rozsáhlých databází.
- ***Text Mining,*** resp. textová analytika představuje analýzu textových zdrojů a získávání z nich nových informací, kde zdroje mohou být velmi různorodé od knižních nebo novinových publikací, přes blogy a další.
- ***Umělá inteligence*** je schopnost stroje nebo systému zvýšit nebo rozšířit jakýkoli proces nebo výstup, které by normálně vyžadovaly lidské znalosti a inteligenci.
- ***Big Data a Big Data Analytics:*** “ je výrazný fenomén současné doby, který je dán extrémním nárůstem dat v souvislosti s uplatňováním IoT, sociálními sítěmi a dalšími. „Big Data Analytics“ představuje i uplatňování principů prediktivní analytiky na těchto objemech dat.

Další podkapitoly se věnují stručné charakteristice uvedených disciplín, metod a nástrojů i s návaznostmi na prediktivní analytiku.

22.1 Machine Learning, ML

Je jednou z klíčových souvisejících disciplín s prediktivní analytikou. S ohledem na rozsah podkladů je jí věnován speciální oddíl E).

22.2 Business Intelligence, BI

Business intelligence zejména sleduje podstatné charakteristiky dat, KPI, vytváří reporty a KPI reporty, určované uživateli („*user driven*“) (Abbott, 2014).

22.2.1 Principy business intelligence

Pokud odhlédneme od realizačních charakteristik a jednotlivých produktů, pak ***základní principy*** řešení business intelligence můžeme shrnout do následujících bodů:

- zatímco transakční systémy jsou primárně určeny pro pořizování a aktualizace dat a tomu odpovídá i organizace dat v databázi, pak řešení business intelligence jsou určené pro analytické aplikace a tomu musí odpovídat i ***výběr dat*** ze zdrojových databází a jejich ***organizace*** v analytických databázích;
- zatímco transakční systémy udržují data na maximální úrovni detailu (většinou na úrovni jedné transakce se všemi jejími detailními atributy), BI řešení ukládají pouze data relevantní pro analýzy, tedy na potřebné ***úrovni detailu (granularity)***, tedy ***detailní i agregovaná*** podle požadovaných hledisek podnikového řízení;
- analytické aplikace pracují primárně s daty podnikových ukazatelů a ty vyhodnocují podle nejrůznějších hledisek, dimenzí a jejich kombinací. To znamená, že BI řešení jsou vesměs založena na ***multidimenzionalitě*** uložení a zpracování dat;
- zatímco transakční systémy obvykle udržují databáze většinou s aktuálními daty, BI řešení jsou založeny na využití ***časové dimenze***, to znamená, že ukládají data do analytických databází postupně v jednotlivých časových snímcích;
- uvedené předchozí principy (agregace, multidimenzionalita, časová dimenze) vedou ke specifickým, podstatně vyšším nárokům business intelligence na ***kvalitu dat***.
- business intelligence je postavena na celé řadě dalších dílčích pravidel a principů a současně i jejich ***různých modifikací*** podle konkrétního technologického prostředí,
- významnou modifikaci v současné době představuje „***self service business intelligence***“ poskytují efektivní a operativní prostředí pro řešení analytických úloh relativně menšího rozsahu a s tomu odpovídajícím menším rozsahem potřebných zdrojů.

22.2.2 Prediktivní analytika a business Intelligence

Zatímco aplikace ***business intelligence*** se orientují na funkce a proměnné, které jsou určované uživateli „*user driven*“, ***prediktivní analytika*** specifikuje funkce a hledané vzory určované daty „*data driven*“. Navíc ***prediktivní modely*** zajišťují více než jen identifikovat atributy, které jsou spojené s cílovou proměnnou, ***identifikují*** také ***nejlepší cestu, jak predikovat cíl***. Neidentifikují jen, které proměnné jsou predikované, ale také, jak dobře predikují cíl. Rozdíl mezi oběma disciplínami dobře dokumentují i rozdíly ***v otázkách zákaznické analytiky***, které obě disciplíny řeší (Abbott, 2014):

- ***Business intelligence***:
  - Jaký je poměr návratnosti emailů, odpovědí apod.?
  - Jaké regiony mají jakou návratnost odpovědí v průzkumech?
  - Jaké jsou poměry návštěvnosti webu podle produktů?
  - Kolik bylo opakujících se nakupujících podle regionů, produktů apod.?
  - Kolik je účastníků loajality programů podle regionů, skupin zákazníků apod.?
- ***Prediktivní analytika***:
  - Jaká je pravděpodobnost návratnosti emailů, odpovědí apod.?
  - Které produkty se budou pravděpodobně nejvíce prodávat?
  - Kolik emailů je pravděpodobně poslat zákazníkům pro zvýšení objemu prodeje?
  - Jaká bude očekávaná návštěvnost webu v příštím týdnu?
  - Jaká je pravděpodobnost vyprodání produktu v příštím měsíci?
  - Jaké je odhadované CLV každého, resp. vybraných zákazníků?

Zatímco funkcionalita ***business intelligence*** konstatuje a analyzuje stav „co se stalo a proč se to stalo?“, funkcionalita ***prediktivní analytiky*** zahrnuje odhady budoucích stavů na základě pravděpodobnosti.

22.3 Competitive intelligence, CI

Competitive intelligence se využívá ***externí informační zdroje*** za účelem ***vyhodnocovat tržní prostředí*** a trendy, které zde probíhají, ***konkurenci***, konkurenční subjekty a jejich významné charakteristiky.

22.3.1 Principy competitive intelligence

Řešení a využití competitive intelligence ve firemní praxi je obvykle spojeno s některými z následujících principů:

- competitive intelligence pracuje výhradně ***s legálními informačními zdroji***, nejedná se o průmyslovou špionáž,
- ***konkurenční zpravodajství*** (competitive Intelligence) je, na rozdíl od průmyslové špionáže, ***etická a legální činnost*** využívající veřejně dostupných zdrojů s cílem zvýšit konkurenceschopnost společnosti za pomoci podpory rozhodování, analýzy trhu, identifikace rizik a příležitostí, ať už současných, nebo budoucích, a to v nekončícím systematickém procesu,
- má primární vazbu na ***strategické řízení*** firmy a její IT, ale ***podporuje analytickou, plánovací a rozhodovací činnost*** organizace na všech úrovních a ve všech oblastech řízení, tj. prodeje, nákupu, marketingu, finančního řízení, controllingu, majetku, řízení lidských zdrojů, výroby a dalších,
- forma ***aktivního*** competitive intelligence se primárně zaměřuje na ***externí informační zdroje a tržní okolí*** firmy (odběratelé, dodavatelé, partneři, konkurence, legislativní rámec, profesní a zájmové komunity, zpravodajské portály, sociální sítě a další),
- competitive intelligence ***umožňuje*** zmapovat tržní prostředí a aktivity konkurence, zhodnotit pozici podniku vůči konkurentům, odhalit případné vnější hrozby, identifikovat možné příležitosti pro další růst a rozvoj,
- ***pasivní*** competitive intelligence se zaměřuje na ochranu vlastních interních zdrojů před aktivitami konkurence.
- ***pasivní*** competitive intelligence ***směřuje k*** tomu, jak omezit efekt aktivit konkurence v oblasti competitive intelligence, upevnit, resp. zlepšit stávající pozici firmy na trhu, identifikovat klíčové nositele znalosti z řad zaměstnanců, odhalit a využít možný skrytý potenciál uvnitř vlastní firmy,
- zahrnuje i proaktivní vyhledávání ***příležitostí pro inovace*** a tím i zlepšování pozice firmy na trhu.

22.3.2 Prediktivní analytika a competitive Intelligence

Je zřejmé, že competitive intelligence se, obdobně jako business intelligence zaměřuje na především na získávání a ***vyhodnocování informací o aktuálním stavu***, v tomto případě převážně ***externích informací*** o externím tržním prostředí.

Na druhé straně díky aktuálním informacím o stavu trhu a aktivitách konkurence představuje podporu ***prognózování a plánování*** v řízení firmy. To znamená, že v tomto místě se obě disciplíny propojují a informace v rámci competitive intelligence mohou představovat i významný datový zdroj pro řešení prediktivních modelů.

22.4 Data Science

Data Science je přístup k řešení IT, který ***z velkého množství komplexních dat,*** které navíc i velmi rychle narůstá***, získává informace a znalosti*** (F. Provost a T. Fawcet, 2013). Z pohledu tohoto textu představuje Data Science také ***společnou nadstavbu*** nad řadou dílčích a rozdílných disciplín, metod a nástrojů.

22.4.1 Principy Data Science

Data Science, datová věda představuje ***široký koncept pro rozvoj a užití dalších dílčích disciplín*** spadajících do podnikové analytiky a v tomto smyslu je postavena na těchto principech:

- v praxi je možné definovat ***sadu základních konceptů***, které jsou ***podstatou pro formování základních disciplín*** pro získávání informací a znalostí z dostupných dat. Tyto koncepty

pokrývají celý proces od pochopení byznys problémů, přes využití technik data science až po poskytování výsledků manažerům a analytikům s cílem zkvalitňovat jejich rozhodování,

- v první řadě jsou to ***koncepty ukazující***:
  - jak data science začlenit do firmy a organizace v rámci konkurenčního prostředí,
  - jak efektivně sestavovat a připravovat týmy pro data science,
  - jak formovat způsob analytického myšlení zaměření na získávání konkurenčních výhod,
  - jak efektivně postupovat při řešení projektů orientovaných na data science.
- ***koncepty zaměřené na způsob analytického myšlení*** umožňují:
  - identifikovat vhodná data a metody pro řešení úloh data science,
  - aplikovat různé procesy data miningu a dalších přístupů a efektivně přispívat k řešení úloh.
- uplatnění ***konceptů pro získávání znalostí z dat*** je další součásti řešení úloh data science,
- zásadním a výchozím ***předpokladem*** pro efektivní využívání uvedených ***konceptů*** Data Science je jejich zasazení do kontextu řízení celé firmy, tj. do soustavy typů aplikací a odpovídajících nástrojů, které se vzájemně doplňují, ale i překrývají, tj. respektování takových souvislostí a jejich efektivní využití vzhledem k potřebám dané firmy.

22.4.2 Prediktivní analytika a Data Science

Je zřejmé, že objasňování principů Data Science na základě uvedených konceptů výrazně ***usnadní komunikaci*** mezi specialisty v Data Science a současně i v prediktivní analytice, tj. mezi doménovými experty, datovými experty a experty prediktivního modelování. To znamená, že se vytváří předpoklady pro vzájemné pochopení a ***podstatně přesnější formulaci problémů*** firmy řešitelných na bázi úloh Data Science a prediktivní analytiky.

Využití prediktivní analytiky na bázi konceptů Data Science ***v rámci jednotlivých oblastí řízení*** firmy znamená předpoklad jejího efektivního řešení v přímých vazbách na potřeby byznysu.

Příkladem takového konceptu je ***určování podobnosti*** dvou entit, k nimž se vztahují určitá data. Takový koncept ***vytváří základnu pro řešení nejrůznějších úloh***, např. pro hledání podobnosti zákazníka s jinými zákazníky. To pak vytváří podklady pro prediktivní úlohy zaměřené na řešení cíleného marketingu, hodnocení zákazníka z hlediska obchodních možností (tržeb), pravděpodobnosti, že právě on využije presentovaných nabídek firmou.

Tyto koncepty jsou také např. základnou pro ***úlohy clusteringu*** sdružující objekty (např. zákazníky, zboží apod.) do skupin na bázi jejich sdílených vlastností. S tím souvisí i ***využívání vzorů („pattern“)*** pro analýzy vlastností zkoumaných objektů, a to v různém kontextu, což je rovněž součástí prediktivní analytiky.

22.5 Data mining, DMI

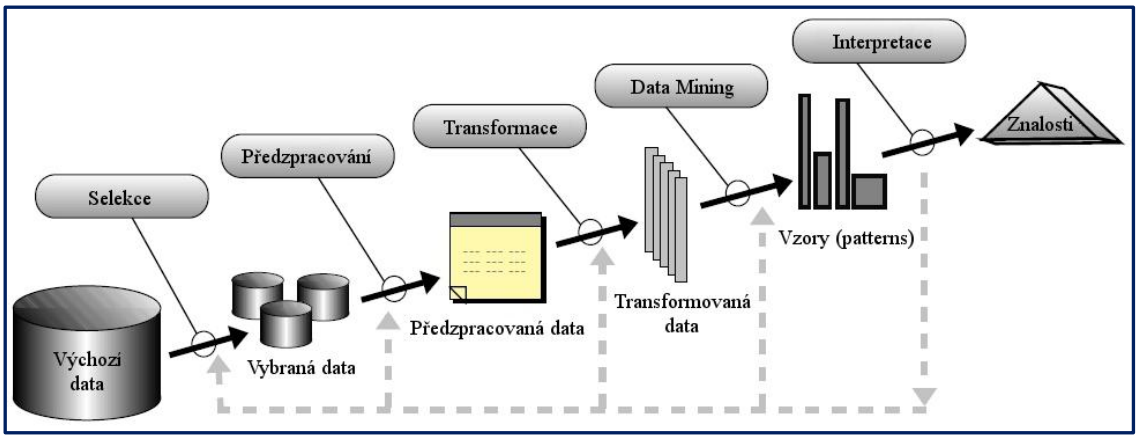
Data mining, DMI (dolování dat) je proces ***extrakce relevantních, předem neznámých nebo nedefinovaných informací*** z velmi rozsáhlých databází.

22.5.1 Principy data miningu

Data mining v analytice a pokročilé analytice představuje „klasickou“ disciplínu s dlouhou historií, jejíž principy a přístupy využívají i další analytické disciplíny:

- důležitou vlastností dolování dat je, že se jedná o ***analýzy odvozované z obsahu dat***, nikoli předem specifikované uživatelem nebo implementátorem, a jedná se především o odvozování prediktivních informací,
- data mining je součástí ***procesu*** „***dobývání znalostí***“, který představuje ***5 fází***, které jsou prováděny opakovaně. Každá fáze tak může (a většinou je) prováděna opakovaně, s cílem nalezení nejlepšího možného výsledku dané fáze, a jeho předání jako vstupu fáze následující.
- ***fáze dobývání znalostí*** jsou (Berka, 2003), viz Obrázek 22-1:
  - ***selekce*** – vytvoření či shromáždění cílových dat, tj. těch dat, která poslouží pro dobývání znalostí

- ***předzpracování*** – „čištění“ špatných dat, zpracování chybějících hodnot, příprava odvozených atributů apod. Předzpracování dat je pokládáno za nejdůležitější část celého procesu dobývání znalostí,
- ***transformace*** – převedení dat do formátů vhodných pro různé algoritmy data miningu či různé softwarové nástroje. Tato fáze zároveň může obsahovat různé agregace a výběry vzorků dat,
- data mining – aplikace analytických metod. Výstupem této fáze jsou nalezené vzory a modely,
- ***interpretace*** – fáze s cílem vyhodnocení nalezených vzorů a modelů. Tato fáze vyžaduje znalost odborníka na řešenou oblast, neboť identifikuje výsledky obsahující znalosti nové, předem známé, samozřejmé, nezajímavé apod. Výstupem se pak stávají analytické zprávy, vizualizace či rozhodnutí provést vhodné akce.



Obrázek 22-1: Přehled fází tvořících proces Dobývání znalostí z databází (Fayyad, a další, 1996 str. 41)

- data mining využívá celou ***škálu metod*** jako rozhodovací stromy, rozhodovací pravidla, asociační pravidla, neuronové sítě a další. Jejich charakteristiky jsou uvedeny ***v kapitole 23,***
- ***aplikace*** data miningu lze rozdělit dle mnoha kritérií a do různých skupin (Máša, 2008). Podle potřeb řízení a uplatnění ve firmách zahrnují:
  - ***segmentace*** je rozdělení objektů do skupin, které mají podobné charakteristiky. Nejčastěji se jedná o segmentaci zákazníků, a to dle různých charakteristik,
  - ***predikce odchodu zákazníků,*** kde účelem je předpovědět klienty, kteří hodlají odejít ke konkurenci, přestat využívat určitého produktu apod.
  - ***cílený marketing,*** tj. marketingové akce specializující se pouze na určitý okruh klientů, takových, u kterých je oproti ostatním klientům vyšší pravděpodobnost, že zareagují na konkrétní kampaň,
  - ***credit scoring*** je ohodnocením klienta dle úvěrového rizika. Tedy rizika, zda budou splácet např. svůj úvěr bance,
  - ***fraud detection,*** aplikace, kdy jsou hledány podvody (ať již v rámci pojistných událostí, bankovních transakcí či v systémech telekomunikačních operátorů),
  - ***analýza nákupního košíku*** zjišťuje souvislosti mezi produkty, které kupují zákazníci společně. (např. párky společně s hořčicí). Díky tomu lze zákazníkovi vhodně nabízet kombinace produktů (cross selling), pracovat s rozmístěním produktů v rámci prodejní plochy prodejny či stránky s produktem v rámci eShopu.

22.5.2 Prediktivní analytika a data mining

Provázání prediktivní analytiky a data miningu je ***velmi silné***. Využívá stejných nebo obdobných metod a obvykle i datových zdrojů.

Rozdíl (a současně shoda) je v tom, že ***cíle data miningu*** jsou predikce ***i*** deskripce, kde z pohledu data miningu je:

- ***„predikce*** funkcí, která využívá některých proměnných k předpovězení neznámých či budoucích hodnot jiných proměnných, které nás zajímají“ (těch, které nejsou vstupy funkce,
- ***deskripce*** „se zaměřuje na hledání lidem srozumitelných vzorů popisujících data“.

V každém případě se řešení úloh prediktivní analytiky silně propojuje s postupy a metodami řešení úloh data miningu.

22.6 Text mining

Text Mining, resp. textová analytika představuje analýzu textových zdrojů a získávání nových informací.

22.6.1 Principy text miningu

Text mining se stává stále významnější součástí podnikové analytiky a je založen na těchto principech:

- ***zdroje*** data mohou být ***velmi různorodé*** od knižních nebo novinových publikací, přes blogy, pracovních poznámek zaměstnanců, např. o kontaktech se zákazníky, o situaci na trhu apod.,
- jako nestrukturovaná data, která jsou předmětem text miningu zahrnují:
    - standardní texty (Word, PDF apod.), prezentace (Powerpoint),
    - texty vyjádřené v HTML, XML apod.,
    - emaily,
    - obrázky, videa,
    - audio soubory
- funkce text miningu vyžadují ***pochopení*** daného jazyka, struktur vět, gramatiky, slangu, souvislostí v textech,

22.6.2 Využití text miningu

Využití text miningu se váže např. k následujícím oblastem:

- význam text miningu se uplatňuje ***s využitím sociálních sítí*** zejména pro oblast ***marketingu*** a marketingové analýzy informací, které se např. na sociálních sítích k firmě váží,
- text mining nachází silné uplatnění i ve ***vyhodnocování nejrůznějších průzkumů***, kdy dotazníky a otázky mají otevřený charakter a respondent na ně má odpovídat volným textem, obdobně při analýzách požadavků a dotazů na help desku,
- text mining se velmi silně ***uplatňuje v aplikacích Competitive Intelligence***,
- do určité míry s text mining souvisí i oblast analýzy obsahu, ***Content Analytics***, která rozšiřuje zdroje o fotografie, multimédia, hlas, případně další,
- významnou aplikační oblast text miningu představuje ***zdravotnická analytika***, kde většina informací jako diagnóz apod. má textový charakter.

22.6.3 Prediktivní analytika a text mining

Text mining ***není*** zatím přímo ***ve středu zájmu*** firem ve spojení s prediktivní analytikou. Na druhé straně jeho ***význam v tomto kontextu rychle stoupá*** tak, jak se ukazuje bohatství informací obsažených v textových zdrojích a využitelných v prediktivních úlohách. K tomu je třeba uvést:

- ***řešení*** úloh text miningu je relativně ***hodně náročné*** vyžaduje si od analytiků prediktivní analytiky značné penzum dodatečných znalostí.
- úspěšnost text miningu v souvislosti s prediktivní analytikou dokumentuje i ***příklad IBM počítače Watson***, který kombinuje techniky text miningu, prediktivní analytiky a umělé inteligence.
- ***analýza textů***, která je často velmi komplexní obvykle ***přináší*** do úloh prediktivní analytiky významná zlepšení a ***vyšší přesnost***,

- zatímco prediktivní analytika staví převážně na čtyřech hlavních druzích analýzy, naproti tomu text mining využívá sedm a více druhů analýz, z nichž následující 5 jsou využívány i v úlohách prediktivní analytiiky a to (Abbott, 2014):
    - „*information retrieval*“ – vyhodnocuje dokumenty podle přehledu klíčových slov,
    - „*document clustering*“ – sdružuje dokumenty na základě klíčových slov a frází z každého dokumentu,
    - „*document classification*“ – označuje dokumenty podle klíčových slov a frází z každého dokumentu,
    - „*sentiment analysis*“ – je speciální případ klasifikace dokumentů, predikuje pozitivní, negativní nebo neutrální pocity, dojmy z různých zákaznických poznámek nebo hodnocení nebo informací na sociálních sítích,
    - „*information extraction*“ – vybírá a sumarizuje klíčová slova a fráze, které odpovídají dokumentům.

22.7 Umělá inteligence (Artificial Intelligence)

22.7.1 Principy umělé inteligence

Umělá inteligence je schopnost stroje nebo systému zvýšit nebo rozšířit jakýkoli proces nebo výstup, které by normálně vyžadovaly lidské znalosti a inteligenci.

Umělá inteligence se chápe ***jako věda*** studující inteligentní agenty, což je jakákoliv entita, která vnímá prostředí a podniká kroky, které maximalizují svou šanci na úspěšné dosažení stanovených cílů. Často však bývá ztotožňována s termínem „*strojové učení*“, resp. „*Machine Learning*“

22.7.2 Prediktivní analytika a umělá inteligence

Umělá inteligence přinesla do prediktivní analytiky obrovský posun, protože před jejím uplatněním bylo možné pracovat pouze s určitým omezeným počtem proměnných, zatímco s jejím vstupem do prediktivní analytiky se tyto počty enormně zvýšily.

Technologie spojené s umělou inteligencí představují významný pokrok v prognózování byznysu a obdobně v celém komplexu plánovacích úloh.

22.8 Big Data Analytics

Příklady použití Big Data Analytics lze objevit téměř v každém sektoru podnikání, kde se pracuje s rozsáhlými objemy dat.

22.8.1 Principy *Big Data Analytics*

Jednou z hlavních předností Big Data Analytics je fakt, že umožňují hledat odpovědi na otázky, na které se organizace v minulosti ani nenapadlo ptát (Halama, 2021).

Výsledkem mohou být nové a lepší produkty nebo optimalizace stávajících služeb. Analytik Jeff Kelly ve svém Big Data manifestu (Kelly, 2017) uvedl několik příkladů použití Big Data nástrojů a technologií v současnosti v praxi již zavedených a osvědčených:

- ***systém doporučení:*** Řada elektronických obchodů využívá Big Data nástroje pro ***doporučení produktu nebo služby*** spotřebiteli na základě analýzy jeho chování, historie a profilu v reálném čase. Amazon tímto způsobem nabízí související produkty a sociální sít LinkedIn využívá Big Data nástroje pro doporučení možných přátel a známých.
- ***názorová analýza:*** Big Data nástroje umožňují pokročilé ***analýzy textu i nestrukturovaných dat např. ze sociálních médií*** jako jsou Twitter a Facebook. Jednou z pokročilých funkcí je názorová analýza, která v daném textu hledá a analyzuje slova vyjadřující názor (pozitivní a negativní indikátory) tykající se zkoumaného produktu nebo služby.
- ***detekce podvodů:*** V oblasti bankovnictví se Big Data nástroje používají k ***detekci podvodů a krádeží***. Analýzou chování uživatele a historií jeho transakcí lze odhalit transakce, u kterých je vysoká pravděpodobnost podvodu, například kompromitace účtu nebo krádež kreditní karty.

- ***analýza marketingových kampaní:*** V současnosti existuje řada nástrojů pro ***monitoring a kvantifikaci efektivnosti marketingových kampaní***. Big Data kromě standardních funkcí umožňuje tyto analýzy provádět v reálném čase a na rozsáhlejších objemech dat umožňující vyšší granularitu dat.
- ***retenční analýza: Analýza zákaznické věrnosti***, respektive pravděpodobnosti, že ***zákazník přejde ke konkurenci*** je další z oblastí, kde mnoho podniků využívá Big Data technologie. Pravidelným rozborem zákaznického chování, využívání služeb, výdajů a dalších vzorů chování lze proaktivně zabraňovat zákaznické retenci.
- ***analýza sociálních vztahů:*** Analýza sociálních vztahů a jejich ***vizualizace pomocí sociálních grafů a sociogramů*** pomáhá určit vztahy ***uvnitř sociální skupiny*** a vliv jednotlivců na ostatní členy skupiny. Tyto informace o spotřebitelích jsou pro podniky cenné z toho důvodu, že určují nejdůležitější zákazníky jako ty, kteří nejvíce ovlivňují ostatní ke koupi, a ne ty, kteří sami kupují nejvíce zboží. Pro dosažení úspěchu produktu je tedy nutné přesvědčit právě spotřebitele s velkým vlivem na ostatní.
- ***analýza zákaznické zkušenosti:*** Big Data technologie se také používají k integraci a ***analýze řady různých datových zdrojů o interakci se zákazníkem*** jako jsou call centra, web a sociální média. Kombinace těchto zdrojů dává ucelený a komplexní pohled na zákaznickou zkušenost a umožňuje lepší porozumění a optimalizaci zákaznické komunikace.
- ***monitoring logů:*** Big Data nástroje a technologie jsou vhodné pro skladování a ***analýzu velkých objemů automaticky generovaných strojových dat*** při provozu počítačové sítě jako jsou síťové a serverové logy a data generovaná dalším hardwarem. Analýza a monitoring těchto dat pomáhá diagnozovat problémy v síti a eliminovat slabé stránky sítě.
- ***věda a výzkum:*** Big Data technologie se často používají k analýze obrovských objemů dat ***v oblasti zdravotnictví a farmacie***, zejména pro prediktivní modelování při vývoji léků a analýze dat z klinických testů.

22.8.2 Prediktivní analytika a Big Data Analytics

Big Data Analytics a prediktivní analytika staví na v zásadě společných metodách. I v rámci Big Data Analytics se realizuje prediktivní modelování, ale v prostředí obrovských objemů dat.

22.9 Závěry



- Existuje celá ***řada disciplín***, které se orientují na analytiku a analytické úlohy. Problém je v tom, že tyto disciplíny se nejen ***doplňují***, ale i ***vzájemně překrývají***, poskytují často obdobnou funkcionalitu, využívají stejné nebo obdobné metody. Specifikovat přesně jejich vzájemné vazby je proto hodně obtížné.
- Pro jejich využití v praxi je proto důležité ***definovat jejich možnosti*** vzhledem k definovaným a vyhodnoceným potřebám byznysu.
- Z pohledu využití prediktivní analytiky je dobré vedle specifikace základních charakteristik těchto disciplín ***vyhodnotit*** alespoň rámcově jejich v***ztahy a rozdíly*** k jejím podstatným charakteristikám. V případě této kapitoly byla tato hodnocení uvedena k následujícím disciplínám:
  - business intelligence,
  - competitive intelligence,
  - data science,
  - data mining,
  - text mining,
  - machine learning,
  - big data analytics.

23. Základní funkce uplatňované v prediktivní analytice



Jako hlavní ***pracovní prostředky*** při řešení úloh prediktivní analytiky jsou ***základní analytické metody a dílčí analytické funkce. Účelem*** kapitoly je:

- vymezit ***podstatu analytických funkcí a základních metod*** vzhledem k potřebám prediktivní analytiky,
- v případě ***funkcí*** uvést velmi stručně jejich obsah a příklad využití,

Je účelné definovat ***základní typy funkcí*** prediktivní analytiky ve vztahu k problémům byznysu, neboť to podporuje systematičnost řešení a zvyšuje jeho kvalitu. To zdůrazňují F. Provost a T. Fawcet v.(Provost, Fawcett, 2016) nebo E. Wilson ve (Wilson, 2021). K hlavním funkcím, podle uvedených autorů, patří ty uvedené v dalších podkapitolách.

23.1 Clustering

Je to přístup nebo kombinace metod, kde hlavním ***cílem je seskupovat data***, která se k sobě váží do jednotlivých skupin. Seskupuje tak objekty na ***základě jejich podobnosti***. Pomáhá, jak pochopit podobnosti v datech, tak jejich rozdíly. Vznikají tak klastry zákazníků, dodavatelů, služeb apod.

Příklad: „Klastry zákazníků založené na demografických charakteristikách, jako je věk, pohlaví, nebo jak často nakupují v daném obchodě nebo na eShopu“

S clusteringem je spojena i metoda ***segmentace***. Ta představuje proces definování a rozdělení velkých homogenních datových sad do jasně identifikovaných segmentů, které mají podobné charakteristiky.

Existuje 5 základních typů clusteringu a segmentace, z nichž nejpoužívanější je ***založený na omezeních*** („*Constraint-Based*“) je nejvíce využívaná v plánování a staví na předdefinovaných pravidlech nebo omezení ve vztahu k určitému atributu. Příkladem je rozdělení dat zboží podle objemu prodeje, např nad 100 000,- Kč. K dalším typům pak patří:

- ***založené na propojení*** („*Conectivity-Based*“) využívá úroveň těsnosti vazeb mezi daty, a tedy i jejich vyšší podobnosti, většinou vedou k hierarchickým klastrům,
- ***založené na vztahu k centru („Centroid-Based“)***, kde každý objekt a data je součástí klastrů a jejich dělení je odvozené od vztahu k centrálnímu klastru,
- ***založené na distribuci*** („*Distribution-Based*“) je postavené na tom, jak všechna data v klastru patří ke stejnému rozdělení (normální Gussovo atd.),
- ***založené na hustotě*** („*Density-Based*“) vytváří klastry podle vysoké hustoty jejich členů v urité části datové sady.

23.2 Klasifikace a hodnocení pravděpodobnosti tříd (Classification and class probability estimation):

Představuje přístup nebo kombinace metod, kde primárním ***cílem je určovat cílovou třídu*** s využitím sledovaných hodnot. Predikuje pro každý objekt, do které sady tříd patří, přičemž obvykle jsou třídy vzájemně disjunktní.

Určuje, do které třídy objekt patří a s tím související ***scoring***, tj. odhad pravděpodobnosti správnosti takového zařazení,

Příklad: „Jak klasifikovat zákazníky do jednotlivých tříd, jak vyhodnotit pravděpodobnost, že tam zákazník skutečně patří?“

23.3 Rozhodovací pravidla

(Kulhavý, 2011)

Použití rozhodovacích pravidel je podobné jako u rozhodovacích stromů. Rozhodovací pravidla jsou ***zapisována ve tvaru „IF Ant THEN Class“*** a umožňují klasifikovat kategorii cílového atributu z kombinace kategorií vstupních atributů. ***Ant*** je nazýván jako ***antecedent neboli předpoklad***, ***Class*** reprezentuje ***třídu neboli kategorii cílového atributu***. Stejně jako v případě rozhodovacích stromů jsou pro rozhodovací pravidla problematické numerické atributy a chybějící hodnoty.

Podobnost metod se odráží i v možnosti ***tvorby rozhodovacích pravidel z rozhodovacího stromu*** (teoreticky existují i postupy, jak získat z rozhodovacích pravidel rozhodovací strom). Každé pravidlo reprezentuje cestu stromem od kořene až do listu. Jednotlivé uzly (včetně kořenového, avšak bez listového uzlu) tvoří předpoklad pravidla (pro každý uzel jako atribut a jeho hodnotu jako příslušnou hranu), závěrem pravidla je listový uzel. ***Celý strom lze reprezentovat jako sadu rozhodovacích pravidel.*** Důvodem pro tento převod může například být snadnější zpracovatelnost rozhodovacích pravidel v rámci automatizovaných systémů.

V případě sady rozhodovacích pravidel chápeme ***jednotlivá pravidla jako navzájem nezávislá***, kdy při klasifikaci hledáme pravidlo odpovídající příkladu. Jinou možností jsou takzvané ***rozhodovací seznamy,*** které jsou uspořádaným seznam pravidel. Uspořádaným ve smyslu zápisu

IF Ant_1 THEN $Class_i$,
ELSE IF Ant_2 THEN $Class_j$,
ELSE IF Ant_3 THEN $Class_k$…

V případě rozhodovacího seznamu každé další ***pravidlo již nezohledňuje příklady, které by splnily pravidla předcházející*** („Uspořádání spočívá v tom, že v každé podmínce ELSE IF se implicitně skrývá negace všech podmínek předcházejících pravidel.“ (Berka, 2003, str. 134)).

Pro rozhodovací pravidla platí ***obdobná omezení jako pro rozhodovací stromy***. I řešení těchto omezení jsou podobná. Chybějící hodnoty je třeba nahradit (nabízí se ignorování objektu s chybějící hodnotou, vytvoření hodnoty specifické pro chybějící hodnoty, nahrazení některou z existujících hodnot.

23.4 Asociační pravidla

(Kulhavý, 2011)

V minulé podkapitole byla uvedena pravidla zvaná rozhodovací. Pravidla zapisovaná ve tvaru ***IF-THEN***. ***Tento tvar mají asociační a rozhodovací pravidla shodný***, liší se však možnostmi v závěru pravidla. ***U asociačních pravidel není*** na rozdíl od pravidel rozhodovacích definován ***cílový atribut pro zařazení příkladu do určité třídy*** (v rámci nastavení konkrétního algoritmu lze nastavit omezení na vybrané atributy jako možné závěry pravidla), naopak nás zajímají ***vazby mezi různými atributy v předpokladu a v závěru***. S tím souvisí problematičtější vyhodnocování těchto pravidel.

Asociační pravidla jsou ***zapisována jako „Ant => Suc“,*** kde ***Ant*** značí ***předpoklad pravidla***, takzvaný antecedent a ***Suc značí závěr pravidla***, takzvaný sukcedent. Antecedent i sukcedent jsou kombinací kategorií (tedy pravidlo může obsahovat více kategorií více atributů jak na straně předpokladu, tak i na straně závěru pravidla).

Hodnocení jednotlivých asociačních pravidel vychází z kontingenční tabulky (též nazývané ***čtyřpolní tabulka***) reprezentující četnosti objektů, kde:

Tabulka 23-1: Čtyřpolní tabulka

	Suc	**¬ Suc**	∑
Ant	a	b	r
¬ Ant	c	d	s
∑	k	l	n

a je počet objektů, splňujících předpoklad i závěr pravidla, tedy splňují celé pravidlo
b je počet objektů, splňujících předpoklad, avšak nesplňujících závěr
c je počet objektů, nesplňujících předpoklad, avšak splňujících závěr
d je počet objektů, které nesplňují ani předpoklad ani závěr
r,s je počet objektů splňujících, resp. nesplňujících předpoklad
k,l je počet objektů splňujících, resp. nesplňujících závěr
n je počet všech objektů

Z hodnot uvedených ve čtyřpolní tabulce pak vycházejí charakteristiky asociačních pravidel, zmíníme základní dvě, podporu (support) a spolehlivost (confidence):

- **Podpora** P(Ant ∧ Suc) = $\frac{a}{a+b+c+d}$

  tedy relativní počet objektů splňujících předpoklad i závěr

- **Spolehlivost** P(Ant | Suc) = $\frac{a}{a+b}$

  tedy podmíněná platnost předpokladu, pokud platí závěr

Další charakteristiky lze nalézt v knize Dobývání znalostí z databází (Berka, 2003, stránky 104-105).

23.5 Přiřazování podobností (Similarity matching):

Metoda „*Přiřazování podobností*“ ***identifikuje podobné objekty*** na základě o nich známých dat. V této souvislosti se využívá tzv. „***firmographic***“, což jsou podstatné charakteristiky firem a od nich odvíjených objektů.

Příklad: „Jaké firmy jsou obdobné našim nejlepším zákazníkům?“

23.6 Seskupování podle výskytů (Co-occurrence grouping):

Seskupování podle výskytů pokouší se najít ***asociace mezi entitami na základě transakcí***, které se k nim váží. Patří sem také „***Odhalování asociačních pravidel***“ nebo „***Analýza nákupního košíku***“.

Příklad: „Jaké položky se obvykle nakupují dohromady?“

23.7 Profilování (Profiling):

Profilování charakterizuje ***typické chování jednotlivce*** nebo skupiny, označuje se také jako „*Popis chování*“,

Příklad: „Jaké je typické použití mobilního telefonu v tomto segmentu zákazníků?“

23.8 Predikce vazeb (Link prediction):

Predikce vazeb predikuje ***vazby mezi datovými položkami*** a obvykle navrhuje, že taková vazba by měla existovat a jaká je síla této vazby.

Příklad: „Jestliže úloha analýza prodeje je ovlivněna faktorem úroveň managementu, pak by měla být ovlivněna i faktorem kultura firmy, a to s obdobnou sílou tohoto vztahu“.

23.9 Redukce dat (Data reduction):

Redukce dat ***redukuje velké objemy dat pro analýzy*** pouze na ty nejpodstatnější informace. Menší objemy vybraných dat jsou snáze využitelné, nakonec i lépe poskytují právě relevantní informace.

Příklad: „Disponujeme daty o několika milionech zákazníků. Bude účelné objem dat omezit podle vybraných kritérií, např. podle objemu tržeb, významu zákazníka apod.“

23.10 Kauzální modelování (Causal modeling):

Kauzální modelování pokouší se ***pochopit, které události nebo akce aktuálně ovlivňují*** ty ostatní. Často je založené na náhodně vybraných a řešených experimentech.

Příklad: „Snahou je vyhodnotit kvalitu dodávek jednotlivých dodavatelů založenou na analýze vybraných dodávek, nebo vybraných parametrech dodavatelů, např. ekonomická síla, personální síla, lokalita, kde působí apod.“

23.11 Závěry, doporučení

Z kapitoly vyplývají následující ***závěry***:

- Jako základní funkce a metody pro prediktivní analytiku lze považovat:
  - clustering,
  - klasifikace a segmentace,

- rozhodovací pravidla,
- asosiační pravidla,
- přiřazování podobnosti,
- seskupování podle výskytů,
- profilování,
- predikce vazeb,
- redukce dat,
- Kauzální modelování.

- ***Jeden z fundamentálních principů prediktivní analytiky*** je podle (Provost, F., Fawcett, T., 2013) správně vybrat, kombinovat a využívat uvedené úlohy pro řešení konkrétních problémů byznysu.

24. Metody uplatňované v prediktivní analytice



Jako hlavní ***pracovní prostředky*** při řešení úloh prediktivní analytiky jsou ***dílčí analytické funkce a pracovní metody***, které, kromě jiného, tyto funkce využívají, resp. jsou jejich náplní.

Účelem kapitoly je pouze ***účelově vymezit vybrané následující metody,*** zejména ve vztahu k prediktivní analytice:

- lineární regrese,
- rozhodovací stromy,
- neuronové sítě,
- Bayesovská klasifikace.

Základními zdroji pro jejich specifikaci jsou:

- ABBOTT, D.: Applied Predictive Analytics. Principlec and Techniques for the Professional Data Analyst. John Wiley & Sons, Indianoplolis, 2014. ISBN: 978-1-118-72796-6.
- BERKA, P. 2003. Dobývání znalostí z databází. Praha: Academia, 2003. str. 366. ISBN 80-200-1062-9,
- GÉRON, A.: Hands-On Machine Leraning with Scikit-Learn, Keras and TensotFlow. O´Reilly, 2023. ISBN: 978-1-098-12597-4.
- KULHAVÝ, L. - Praktické uplatnění technologií Data Mining v pojišťovnictví, VŠE, Praha, 2011.
- WILSON, J., E.: Predictive Analytics for Business Forecasting and Planning. Graceway Publishing Company, 2021. ISBN 978-0-9839413-8-5.

24.1 Regresní analýzy

Představuje třídu modelů typu „příčina – následek“, kde je definován vztah mezi příčinou (nezávisle proměnná) a následkem (závisle proměnná), které jsou využity k řešení predikcí, resp. mezi vstupy, prediktory (nezávisle proměnné) a cílovými proměnnými (závisle proměnné). Odhaduje nebo predikuje pro každý objekt numerickou hodnotu určité proměnné.

24.1.1 Základní principy regresní analýzy

Regresní analýza představuje souhrn statistických postupů a metod sloužících k analýze vztahu středních hodnot numerické proměnné ***y*** a hodnot numerické proměnné ***x*** nebo většího počtu takových proměnných (Hindls a další, 1999).

K principům regresní analýzy patří (podle Hindls a další, 1999, podrobněji statistická literatura):

- směřuje k ***nalezení vhodných funkcí***, pomocí nichž bude možné odhadovat neznámé střední nebo individuální hodnoty ***proměnné y*** pomocí známých hodnot ***proměnné x***, označované jako ***regresní odhady***,
- ***měřitelné činitelé***, které se při odhadech mají brát v úvahu, představují ***okruh vysvětlujících proměnných***,
- pokud se využívá pouze ***jedna vysvětlující proměnná***, jde o ***jednoduchou regresní analýzu***, pokud se využívá ***více*** proměnných, jde o ***vícenásobnou regresní analýzu***,
- podkladem pro jednoduchou regresní analýzu je ***n dvojic***, pro vícenásobnou regresní analýzu ***n trojic, čtveřic*** atd. hodnot,
- ***úkolem*** regresní analýzy je specifikace funkce přiřazující hodnotám vysvětlujících proměnných střední hodnoty vysvětlované proměnné – ***regresní funkce***, tj. přímkové, parabolické, hyperbolické regresní funkce,

- při jednoduché regresní analýze se využívá ***bodový diagram***, kde každá z ***n dvojic hodnot*** x_i, y_j je znázorněna jako bod v pravoúhlé soustavě souřadnic,
- v případě ***více vysvětlujících proměnných*** je nutné brát v úvahu jejich ***vzájemná závislost***. Pokud ***jsou nezávislé***, pak každá z nich přispívá k vysvětlení variability hodnot ***proměnné y***,
- při regresní analýze se vychází z předpokladů formulovaných ***v pravděpodobnostních regresních modelech***,
- nejjednodušším typem regresního modelu je ***klasický lineární model***, model s lineární regresní funkcí vysvětlujících proměnných, jež nejsou náhodnými veličinami.

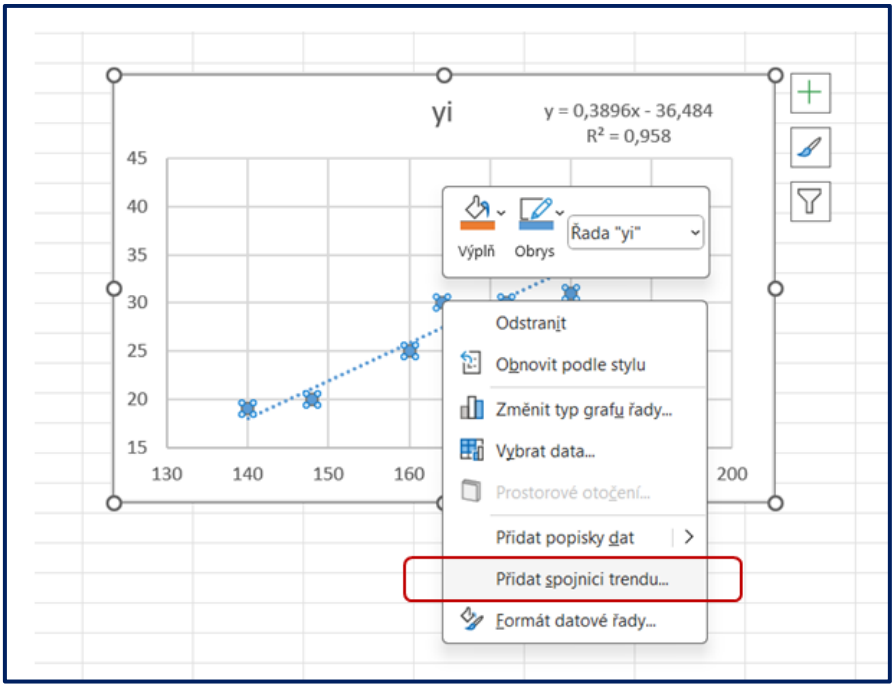
Princip jednoduché regresní analýzy dokumentuje ***následující příklad*** predikce mezd:

Je třeba zjistit ***závislost mzdy pracovníka (y_i v tis. Kč)*** na počtu jeho ***počtu odpracovaných hodin (x_{2i}),*** a to s využitím funkcí Excel. Hodnoty y_i představují vysvětlované, pozorované hodnoty. Hodnoty x_{2i} představují vysvětlující hodnoty, tj. objemy odpracovaných hodin za dané období.

yi	31	40	35	30	19	20	25	30	35	25	19	20	25
x2i	180	196	180	172	140	148	160	164	180	160	140	148	160

Na dalších obrázcích jsou uvedeny ***jednotlivé kroky pro řešení regresní analýzy***:

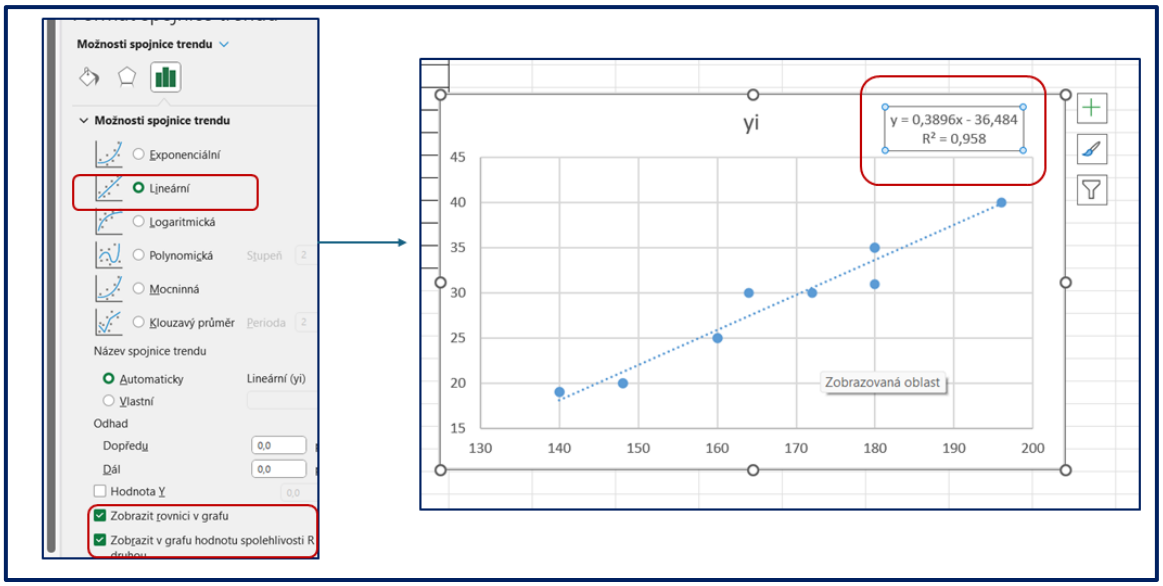
Po kliknutí na libovolný bod se zobrazí nabídka s volbou „*Přidat spojnici trendu*“ na obrázku dole.



Obrázek 24-1: Přidání spojnice trendu (grafu) k jednotlivým bodům

V dalším kroku se v tomto případě volí:

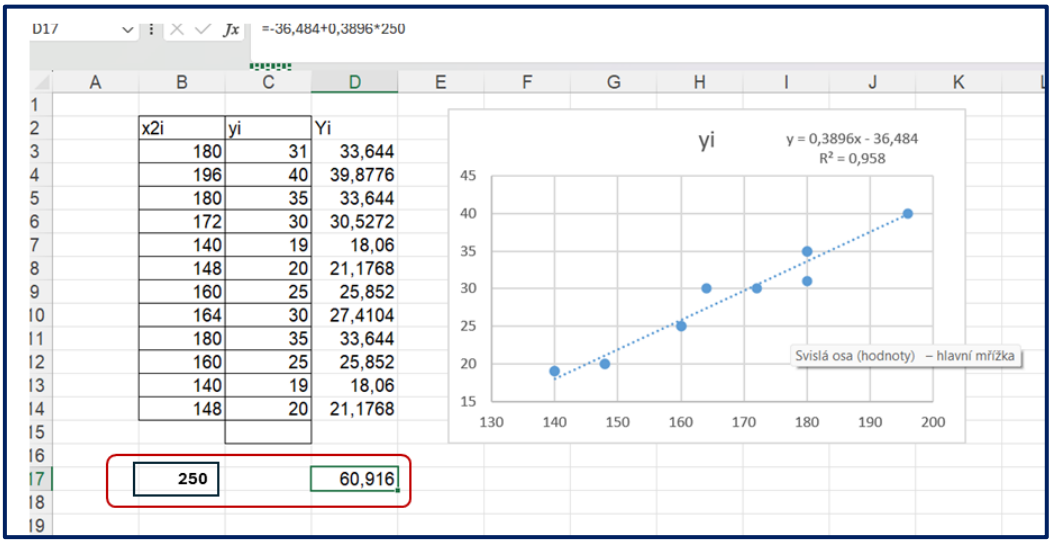
- typ funkce (grafu): „*Lineární*“ (obrázek vlevo nahoře),
- rovnice grafu: y = 0,3896x – 36,484 (obrázek vlevo dole, s výsledkem vpravo nahoře)
- hodnota spolehlivosti grafu: R^2 = 0,958, vysoká spolehlivost (obrázek vlevo dole, s výsledkem vpravo nahoře)



Obrázek 24-2: Určení zobrazení rovnice v grafu a hodnoty spolehlivosti grafu R^2

Výsledné hodnoty cílové proměnné podle výpočtu v rovnici jsou ve sloupci Y_i.

Pro novou hodnotu 250 odpracovaných hodin je mzda podle výpočtu v rovnici 60,916,- Kč.



Obrázek 24-3: Hodnoty vysvětlující, vysvětlované i cílové proměnné

K uvedeným základním principům doplníme několik poznámek podle (Abbott, 2014):

- Regresní modely predikují kontinuální cílové proměnné spíše než kategoriální cílové proměnné. Regresní modely musí predikovat každou hodnotu cílové proměnné s vysokou přesností.
- Nejpoužívanější algoritmus v regresních modelech je lineární regrese (viz výše). Jiný využívaný algoritmus pro regresní modely jsou neuronové sítě a většina software tak využívá jak lineární regresi, tak neuronové sítě.
- Forma vyjádření pro lineární regresní model je obdobná požívaná pro neuronové sítě, tedy:

  $Y_i = w_0 + w_1.X_1 + w_2.X_2 + \ldots + w_n.X_n$. kde:

  Y_i je cílová proměnná, tedy výstup,

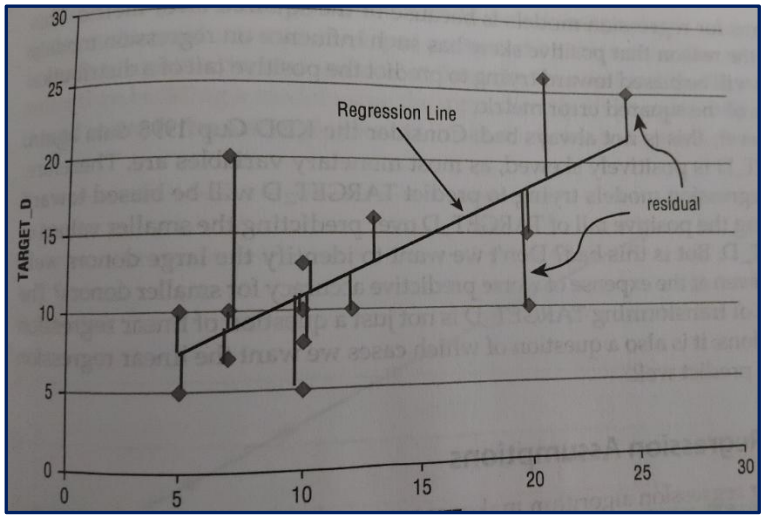
  X_1, X_2 atd. jsou vstupy,

w_0, w_1 atd. jsou váhy, resp. koeficienty,

indexy i se vztahují k jednotlivým záznamům vstupní tabulky, to znamená, že každý záznam bude mít svou predikovanou hodnotu odpovídající vstupu.

- Rozdíl mezi cílovou hodnotou „pozorovanou“ (viz výše) a hodnotou predikovanou regresním modelem představuje rezidua a obvykle se označuje písmenem *e* (Obrázek 24-4):

$$e = y_i - Y_i$$



Obrázek 24-4: Zobrazení reziduí mezi pozorovanými a predikovanými hodnotami cílové proměnné (Zdroj: Abbott, 2014).

- Lineární regrese je založena na několika předpokladech, uvádíme podle (Abbott, 2014) nejvýznamnější:
  - Vazby mezi vstupními proměnnými a výstupními proměnnými se předpokládají lineární.
  - Vstupy nemají navzájem korelace, resp. korelace mezi vstupními proměnnými se rovnají 0.
- Lineární regrese je založená na tom, že vstupy odpovídají cílové proměnné. To v řadě případů vyžaduje, aby analytik vybral v rámci návrhu modelu nejlépe odpovídající proměnné.
- Zatímco regresní algoritmy predikují kontinuální cílové proměnné, pak klasifikační algoritmy predikují kategoriální cílové proměnné.
- Každý algoritmus je založen na své sadě předpokladů, které je třeba identifikovat a přiřešení modelu aplikovat.

24.1.2 Výhody regresní analýzy

- V případě ***lineární regrese***:
  - Je ***relativně jednoduché vytvořit*** na této bázi model a rovněž jednoduché je ***interpretovat*** výsledky.
  - Může pracovat s téměř ***jakýmkoli objemem dat*** i při jednoduchém jejich začlenění do modelu.
  - Pomáhá určit, ***které faktory jsou velmi významné*** a které lze naopak ignorovat a jak tyto faktory mezi sebou ***souvisí***.
  - Funguje velmi efektivně, kde ***vlastnosti jsou lineární***.
- V případě ***vícenásobné regrese***:
  - Je široce využívaná, protože je ***velmi efektivní*** a nevyžaduje příliš mnoho výpočetních zdrojů.
  - Nevyžaduje ***normalizované vstupní vlastnosti***.
  - Může být implementovaná ***jednoduše a rychle***.

24.1.3 Omezení regresní analýzy:

- V případě ***lineární regrese***:
  - Předpokládá, že ***data jsou nezávislá***, což nemusí vždy platit.
  - Může být velmi citlivé na ***abnormální data*** (outliers).
  - Předpokládá, že ***vazba „příčina – následek" zůstává neměnná*** a lineární, což rovněž nemusí vždy platit.
- V případě ***vícenásobné regrese***:
  - ***Není jedinou metodou*** pro řešení komplexních problémů.
  - ***Nefunguje dobře s nezávislými proměnnými***, které nejsou korelovány k cílové proměnné.

24.2 Rozhodovací stromy

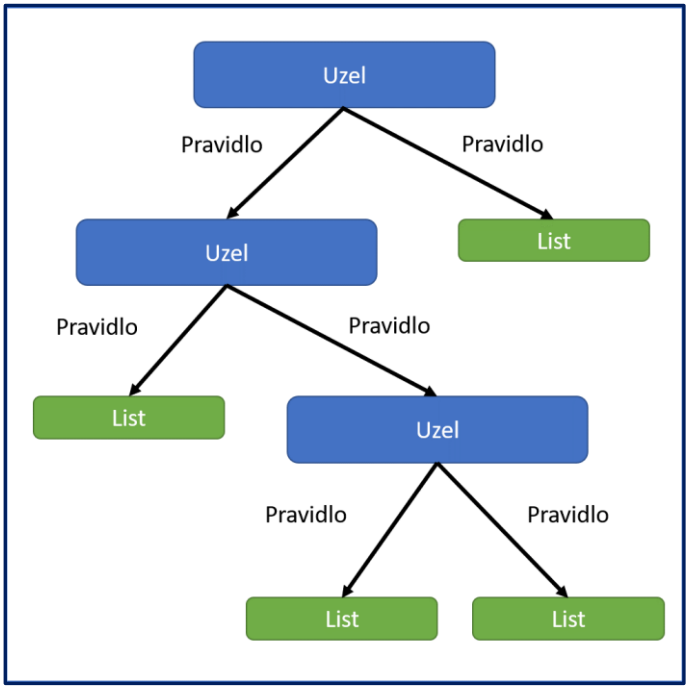
24.2.1 Principy rozhodovacích stromů

Rozhodovací stromy jsou prakticky nejpopulárnější metodou v prediktivní analytice, pouze regresní analýza je podle průzkumů využívaná častěji. Důvody popularity jsou uvedeny v kapitole jejich efektů, kapitola 24.2.4. Rozhodovací strom je grafickou reprezentací logického vývoje časově na sebe navazujících alternativních rozhodnutí. (Friebelová, 2006). Jde o zvláštní případ grafu, kdy rozhodovací strom je tvořen z: (Vomlelová, 2009)

- ***kořene a vnitřních uzlů*** – označených atributem, ze kterých vede jedna hrana pro každou možnou hodnotu tohoto atributu,
- **listů,** které jsou označeny predikovanou hodnotou cílového atributu G, za předpokladu, že ostatní atributy nabývají hodnot na cestě od kořene do listu (pokud se některé atributy na cestě nevyskytují, pak na jejich hodnotě v rámci klasifikace nezáleží),
- ***podmínek*** („*split*"), na jejichž základě se data dělí na 1 nebo více podskupin, resp. větví („*branch*"),
- ***hloubkou*** („*depth*"), která je vyjádřena počtem podmínek mezi kořenem a nejhlubším koncovým uzlem.

Klasifikaci příkladů na základě rozhodovacích stromů lze jednoduše provést tak, že z výchozího uzlu postupujeme přes vnitřní uzly (z uzlu volíme hranu odpovídající hodnotě daného atributu) až do listu, který je klasifikací daného příkladu.

Princip rozhodovacích stromů dokumentuje Obrázek 24-5:



Obrázek 24-5: Rozhodovací strom (Zdroj: Rešl, 2020)

24.2.2 Řešení rozhodovacích stromů

Podstatným ***úkolem*** v případě metod rozhodovacích stromů, je ***jejich konstrukce***. Postupuje se metodou „***rozděluj a panuj***“ („*Divide and Conquer*“). Tato metoda znamená, že jsou data postupně rozdělována na stále menší podmnožiny, s cílem nalézt tu podmnožinu, ve které převládají příklady jedné třídy (hodnoty cílového atributu).

Základní algoritmus vycházející z této metody je ***TDIDT („****Top down induction of decision trees*”). Hlavním cílem je nalézt strom, jakožto reprezentaci získané znalosti, konsistentní s trénovacími daty. Konsistentních s trénovacími daty může být více různých stromů, v tom případě je dávána přednost stromům menším a jednodušším. Algoritmus zahrnuje tyto kroky (Berka, 2003 str. 86):

- zvol jeden atribut jako kořen dílčího stromu,
- rozděl data v tomto uzlu na podmnožiny podle hodnot zvoleného atributu a přidej uzel pro každou podmnožinu,
- existuje-li uzel, pro který nepatří všechna data do téže třídy, pro tento uzel opakuj postup od bodu 1, jinak skonči.

Algoritmus TDIDT tak má snahu ***vytvořit strom co možná nejlépe reprezentující trénovací data***. Existují však situace, kdy dosažení takové reprezentace není zcela možné (např. v případě existence šumu[1] v datech), či tato snaha vede k příliš rozsáhlým stromům a tím i ke zvýšenému riziko přeučení.

Při řešení rozhodovacích stromů se (s vazbou na příslušný software) ***určují následující volby***:

- maximální hloubka, tj. maximální počet úrovní stromu, omezuje komplexitu stromu,
- minimální počet záznamů na koncových uzlech,
- minimální počet záznamů na rodičovských uzlech.

Proto je v praktických algoritmech často aplikováno takzvané ***prořezávání stromů***, kdy dojde k nahrazení určitého podstromu jediným listem. Volba, ve kterém uzlu je vhodné uvažovat o aplikaci prořezávání stromů, vychází buď ze statistického testu na trénovacích datech, či jsou využita tzv. validační data testující chování prořezaného stromu při klasifikaci nových případů v porovnání se stromem neprořezaným. Výsledkem je ***jednodušší a srozumitelnější podoba rozhodovacího stromu***, strom

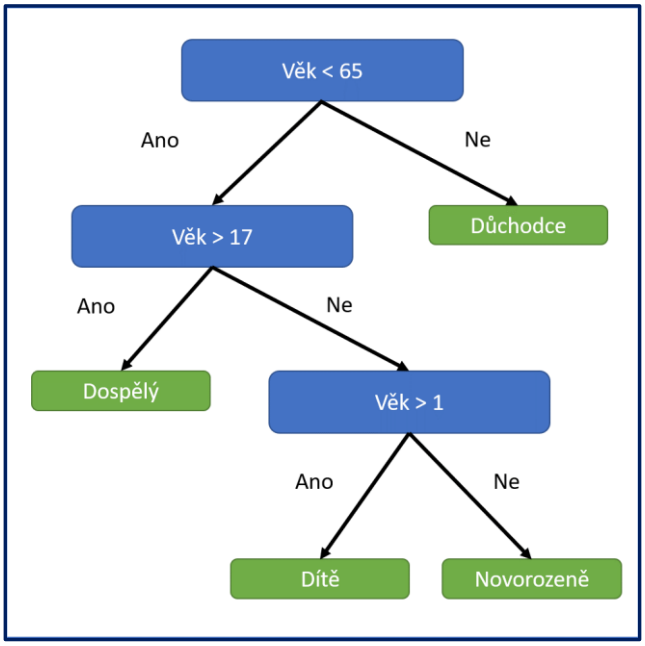
[1] Pod pojmem šum se rozumí především tzv. odlehlé hodnoty, tedy hodnoty značně vzdálené od hodnot pro daný atribut běžných, či hodnoty zjevně nesmyslné.

méně zatížený případným šumem v datech, avšak většinou za cenu zhoršeného chování při klasifikaci nových příkladů.

Klíčovou otázkou procesu tvorby rozhodovacích stromů je, ***který atribut zvolit pro větvení v kořenovém uzlu*** (následně každý vnitřní uzel je kořenovým uzlem pro podstrom, tvořící danou podmnožinu). Snahou je zvolit takový atribut, který od sebe nejlépe oddělí příklady patřící do různých tříd.

Při aplikaci metod rozhodovacích stromů je třeba uvažovat ***omezení v práci s numerickými a chybějícími hodnotami atributů.*** Pro numerické hodnoty atributů je omezením nekonečný počet možných hodnot, a tedy i nekonečná větvení v uzlech stromu. Řešením je ***diskretizace dat*** do konečného (malého) počtu intervalů. Tuto diskretizaci lze provádět v rámci předzpracování dat, alternativou je volba algoritmu, který dokáže s numerickými daty pracovat (především se jedná o možnosti online diskretizace v rámci běhu algoritmu, pokud tuto možnost algoritmus podporuje).

V případě numerického cílového atributu hovoříme o takzvaných ***regresních stromech***, které namísto klasické klasifikace objektů do tříd (klasifikační stromy), provádějí odhad numerické hodnoty cílového atributu.



Obrázek 24-6: Příklad rozhodovacích stromů na zjištění věkové kategorie (Zdroj: Rešl, 2020)

24.2.3 Využití v prediktivní analytice

Tento typ modelu je ***jedním z nejoblíbenějších*** prediktivních modelů pro svoji ***jednoduchost a dobré výsledky*** (SIEGEL, 2013). ***Ze vstupních proměnných vybírá*** model na základě algoritmů ty, které jsou statisticky ***nejdůležitější a vytváří pravidla, kterými segmentuje bázi dat***. Poskládaná pravidla naučeného modelu se dají schematicky zobrazit jako strom s kořenem nahoře a listy dole.

Rozhodovací stromy tak umožňují ***zpětnou interpretaci a vyvození dalších závěrů***. Oproti jiným modelům je možné do nich nahlédnout a ***zkoumat jednotlivá rozhodnutí a pravidla*** z nich vyplývající. Každý rozhodovací strom vychází z jednoho ***kořenového uzlu (root node), který představuje všechna data***. Kořenový uzel je ***generace 0*** (SAS Institute Inc., 2011). Jeho ***přímí potomci*** (uzly) jsou ***generace 1*** a každý další uzel obsahuje podmnožinu báze omezenou na pravidla předcházející danému uzlu. Strom je zakončen ***listy*** – uzly, které se už dále nevětví. Způsob dělení uzlů a výběr proměnných probíhá ***na základě statistických metod***, které určují důležitost každé proměnné. Pro dělení je vždy vybrána v danou chvíli nejdůležitější proměnná.

24.2.4 Efekty rozhodovacích stromů

- Jsou ***jednoduché*** na pochopení a interpretaci. Mohou být interpretovány jako posloupnosti „*if-then-else*“ pravidel, které generují nakonec prediktivní hodnoty buď jako pravděpodobné

poměry kategoriálních proměnných nebo jako průměrná hodnota cílové proměnné. Mohou být realizována na bázi na pravidlech založených systémech nebo SQL.

- Jsou ***jednoduché na vytváření***. Jsou lehce rozšiřitelné, pokud se rozsah záznamů nebo polí zvyšuje.
- Mohou pracovat jak ***s nominálními, tak kontinuálními*** vstupy.
- Mají zabudovaný ***výběr proměnných***, což je výhodné v situacích se stovkami proměnných.
- Jsou ***neparametrické, tedy nepracují s žádnými předpoklady*** o rozdělení vstupů nebo o cílových proměnných.
- Mohou automaticky řešit ***chybějící data***.
- Jsou flexibilní vzhledem k ***užití numerických i kategoriálních dat***.
- ***Nevyžadují velké objemy času*** v souvislosti s přípravou dat.

24.2.5 Omezení rozhodovacích stromů

- ***Chybná rozhodnutí*** na základě rozhodovacího stromu mohou mít ***značný dosah*** a dopad.
- I ***malé změny*** ve vstupních datech vyvolávají ***velké dopady*** na výsledky.
- Jsou velmi ***náchylné na „overfitting“*** a tedy vyžadují často prořezáváni stromu.
- Postupem doby a s rozvojem se mohou stávat ***velmi složitými***, vyžadující ***značný čas na řešení*** a aktualizace.
- Pokud je hloubka (počet úrovní) rozhodovacího stromu větší než 3, pak se většinou ***hůře interpretuje***.
- Rozhodovací stromy zahrnují ***pouze jednu proměnnou*** na každou podmínku.
- Rozhodovací stromy představují spíše nestabilní modely, ***malé změny dat znamenají významné změny v rozhodovacím stromu.***

Detailně: Berka, 2003

24.3 Neuronové sítě

Myšlenka „umělých neuronových sítí“ vychází ze znalostí o biologických neuronových sítích. Umělé neuronové sítě (*Artificial Neural Networks, ANN*) ***se skládají z jednotlivých navzájem propojených neuronů***. ***Neuron*** je jednotka, která přijímá na vstupu podněty, vstupující do součtového členu (provádí vážený součet skalárních součinů podnětu a jeho váhy). Forma vyjádření pro neuronové sítě, je obdobná jako pro lineární regresní model (viz kapitola 24.1):

$$Y_i = w_0 + w_1.X_1 + w_2.X_2 + \ldots + w_n.X_n.$$ kde:

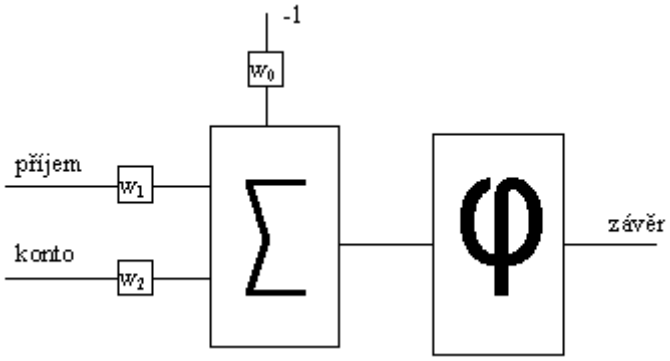
Y_i je cílová proměnná, tedy výstup,

X_1, X_2 atd. jsou vstupy,

w_0, w_1 atd. jsou váhy, resp. koeficienty,

Váhy přiřazené každému vstupu jsou využívány v procesu učení (v okamžiku počátku procesu učení jsou váhy nastaveny na náhodná, malá čísla, blízká nule (Berka, 2003, str. 160) a učením se modifikují).

Hodnotu vzniklou v součtovém členu následně aktivační funkce transformuje ***na výstup***, pokud jeho minimální hodnota překročí práh.

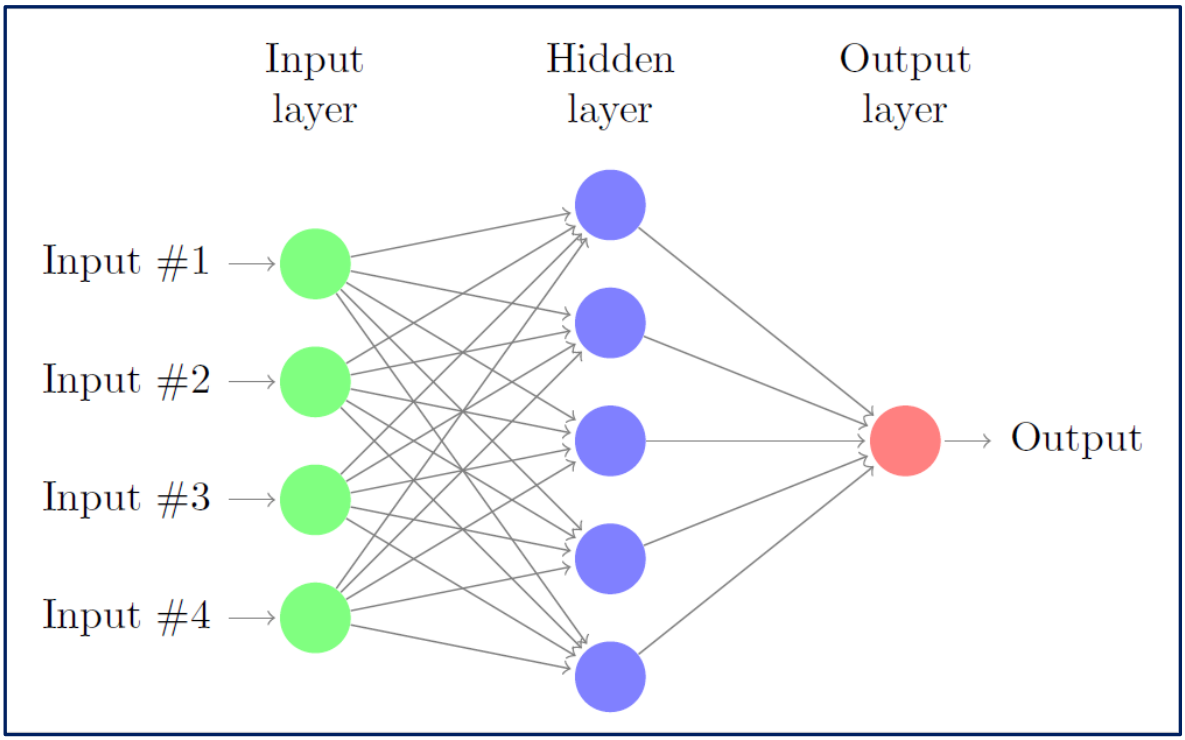


Obrázek 24-7: Schéma neuronu, kde w1 a w2 jsou vstupy, w0 je prahem. Suma reprezentuje součtový člen,φ aktivační funkci. (Berka, 2003, str. 158)

V současnosti asi ***nejpoužívanějším algoritmem*** mezi metodami neuronových sítí je ***algoritmus SVM (Support Vector Machine).*** „Hlavní myšlenkou je, převést pomocí vhodné datové transformace úlohu klasifikace do tříd, které nejsou lineárně separovatelné, na úlohu klasifikace do tříd lineárně separabilních“ (Berka, 2003, str. 171).

Tento druh modelu ***se inspiruje architekturou mozku*** organismů a jeho schopností učit se a myslet. Lidský mozek je složený z miliard specializovaných buněk, neuronů, které jsou navzájem propojeny. Tyto buňky jsou schopné přijímat, zpracovávat a přenášet elektrochemické signály (STATSOFT, 2013).

V prediktivním modelování se ***umělá neuronová síť také skládá z neuronů***, které jsou navzájem propojeny a jsou ***schopné přijímat vstupy a odesílat výstupy***. Každý neuron je aktivován, tedy produkuje výstup, ***pouze pokud hodnoty do něj vstupující*** (po vynásobení s váhami a sečtení) ***překročí definovanou, prahovou hodnotu*** (STATSOFT, 2013).



Obrázek 24-8: Struktura běžné neuronové sítě (Zdroj: Rešl, 2020)

Neurony jsou složené v několika vrstvách:

- ***Vstupní vrstva (input layer)*** – představuje proměnné, které vstupující do modelu.

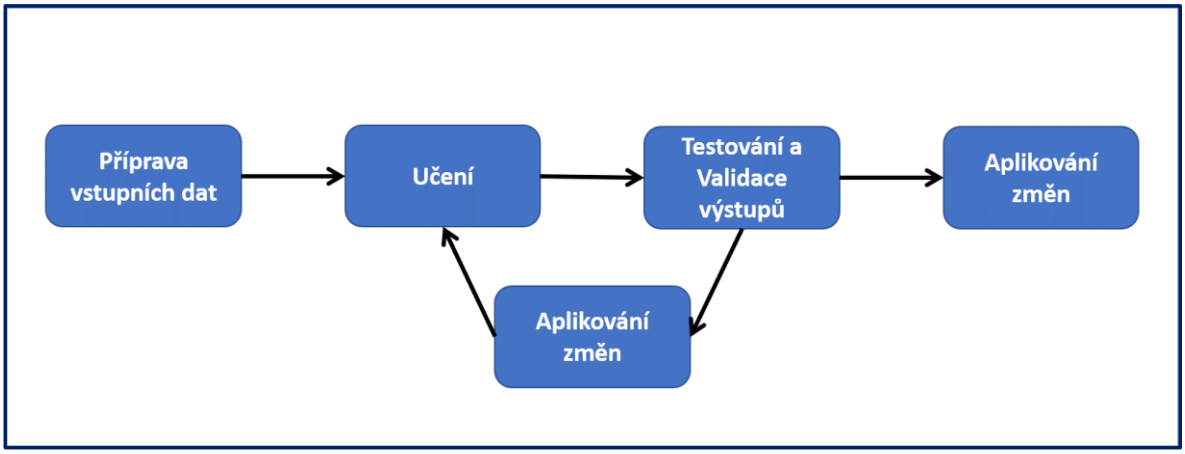
- ***Vnitřní schovaná vrstva (hidden layer)*** – může být složena z 0 až N vrstev. Hodnoty ze vstupní vrstvy jsou v této vrstvě propagovány dále, jsou násobeny se svými váhami, sečteny a aplikovány na nelineární funkci.
- ***Výstupní vrstva (output layer)*** – představuje modelem predikované hodnoty.

Všechny neurony vnitřní a výstupní vrstvy jsou ***propojeny se všemi neurony*** vrstvy předchozí (STATSOFT, 2013). Určení počtu skrytých vrstev a počtu neuronů v nich je jedním z nejdůležitějších rozhodnutí, které ovlivňuje schopnost predikce a generalizaci neuronové sítě. ***Rozhodnutí závisí na počtu vstupních proměnných a*** vlastnostech a velikosti učících dat. Příliš nízký počet má za následek nedoučení modelů, příliš vysoký zvyšuje riziko přeučení sítě (MATIGNON, 2005).

Existují ***různé druhy*** umělých neuronových sítí. V praxi ***nejvyužívanějším typem neuronové sítě je Multilayer Perceptron (MLP)*** (SAS, 2013). Jedná se o typ sítě s feed-forward topologií (hodnoty jsou sítí propagované pouze jediným směrem – vpřed) s možností specifikace počtu skrytých vrstev, počtu jednotek ve skrytých vrstvách a dalšími možnostmi (STATSOFT, 2013). ***Nevýhodou*** neuronových sítí je fakt, že produkované výstupy nejsou zpětně refaktorovatelné – ***není možné s určitostí říci, proč je výsledek takový, jaký je***.

Po trénovacích fázích dochází na fáze strojového učení. Síť se sama učí, respektive upravuje váhy tak, aby zpřesnila výstupy. Pro učení jsou pak fáze následující (Rešl, 2020):

- ***Příprava vstupních dat –*** Zde se jedná jen o přípravnou fázi celého učícího modelu. Dochází v ní totiž k přípravě datové sady, která bude pro následné učení využita. Datová sada by měla být rozdílná oproti trénovacím sadám.
- ***Učení –*** Síť obdrží datovou sadu, na které se pokouší dosáhnout požadovaného výsledku.
- ***Testování a Validace výstupů –*** V rámci testování síť pracuje s datovou sadou, kdy zná výsledky, kterých má dosáhnout. Po aplikování naučeného postupu pak dochází k validaci jednotlivých výstupů se známými výsledky.
- ***Aplikování změn nebo přeučování modelu –*** Pokud tedy dosažené výstupy splňují kritéria známých výsledků, pak může dojít na aplikování postupu vůči reálným datům. Ne vždy je ale dosazeno pozitivních výsledků. V takovém případě dochází k přeučování modelu. Model se tedy vrací do své fáze učení a upravuje algoritmy, pomocí kterých dosáhl předešlých výstupů.



Obrázek 24-9: Schéma učení u neuronových sítí (Zdroj: Rešl, 2020)

Neuronové sítě patří mezi ***metody subsymbolické***, možnou nevýhodou může být ***horší čitelnost*** v nalezených znalostech pro člověka. Tento fakt však nemusí být překážkou v okamžiku, kdy jsou neuronové sítě využívány pro klasifikaci či predikci pomocí počítačových systémů. Zároveň se neuronové sítě liší oproti dříve zmíněným metodám v problematice zpracování numerických hodnot.

Neuronové sítě naopak ***využívají numerické atributy a problematické pro ně jsou atributy kategoriální*** (řešením může být binarizace). Neuronové sítě neumí pracovat s chybějícími daty a vyřešení této problematiky je opět směřováno do fáze přípravy dat (ačkoli existují i myšlenky práce například

s několika různými sítěmi v rámci jedné úlohy či možnosti predikce chybějících hodnot v rámci běhu algoritmu (Sharpe & Solly, 1995)).

24.3.1 Výhody neuronových sítí

- Mohou generalizovat ***vazby*** v datech, které ***nejsou zjistitelné*** tradičními metodami.
- ***Neexistují omezení***, pokud jde ***o vstupní proměnné***, mohou zpracovávat výstupy i na základě nekompletních znalostí.
- Mohou být ***provozně efektivnější*** než ostatní modely, zejména při využití paralelního zpracování.

24.3.2 Omezení neuronových sítí

- ***Efektivní architektura neuronová sítě*** může být často navržena pouze na základě metody pokus – omyl.
- ***Nedostatek vysvětlení*** výstupních řešení může u uživatelů způsobovat nedůvěru k získaným výsledkům.

24.4 Bayesovská klasifikace

(Kulhavý, 2011, Berka, 2003)

24.4.1 Principy Bayesovské klasifikace

Základem metod bayesovské klasifikace je ***Bayesova věta o podmíněných pravděpodobnostech***:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

kde:

- P(H|E) aposteriorní pravděpodobnost, pravděpodobnost hypotézy H, pokud víme že nastala evidence E,
- P(H) apriorní pravděpodobnost hypotézy H (bez ohledu na evidence)
- P(E|H) podmíněná pravděpodobnost E při nastání H,
- P(E) pravděpodobnost, že nastane evidence E.

Bayesova věta popisuje vliv pouze jedné evidence E na uvažovanou hypotézu H. V reálných úlohách se však uvažuje vliv většinou více než pouze jediné na hypotézu H. V případě vlivu více evidencí, které jsou podmíněně nezávislé při platnosti hypotézy H, lze využít Naivní Bayesovský klasifikátor (naivní, protože „naivně“ předpokládá absolutní nezávislost evidencí). V případě, kdy ***evidence nejsou podmíněně nezávislé***, využívají se ***Bayesovské sítě.***

Naivní bayesovský klasifikátor vychází z výše uvedené Bayesovy věty, upravené o existenci více evidencí, tedy:

$$P(H|E_1, \dots, E_K) = \frac{P(E_1, \dots, E_K|H)P(H)}{P(E_1, \dots, E_K)} = \frac{P(H)}{P(E_1, \dots, E_K)} \prod_{k=1}^{K} P(E_k|H)$$

Bayesovské sítě pracují s předpokladem, že ***evidence nemusí být navzájem nezávislé***. „Bayesovská síť je acyklický orientovaný graf, zachycující pomocí hran pravděpodobnostní závislosti mezi náhodnými veličinami. Ke každému uzlu u (náhodné veličině) je přiřazena pravděpodobnostní distribuce tvaru P(u|rodiče(u)), kde rodiče(u) jsou uzly, ze kterých vycházejí hrany do uzlu u. Uspořádejme (a očíslujme) všechny uzly sítě tak, že rodiče jsou před svými dětmi (mají nižší pořadové číslo). Potom pro každý uzel u_i platí, že je podmíněně nezávislý na všech uzlech s nižším pořadovým číslem s výjimkou svých rodičů podmíněno rodiči rodičů.“ (Berka, 2003, str. 188) Sdružená pravděpodobnostní distribuce celé sítě je pak vyjádřena jako:

$$P(u_1, \dots, u_n) = \prod_{i=1}^{n} P(u_i | rodiče(u_i))$$

24.4.2 Efekty Bayesovské klasifikace

- ***Algoritmus naivního bayesovského klasifikátoru*** se těší velké oblíbenosti v rámci praktických aplikací technologie data mining a tedy i prediktivní analytiky.
- Výhodou je jeho ***rychlost v rámci rozsáhlých datových souborů*** (učení neprobíhá jako prohledávání prostoru hypotéz, ale prostým výpočtem pravděpodobností na základě četností hodnot jednotlivých atributů).
- Úspěšný je i ***v úlohách klasifikace***, kde není naplněn předpoklad podmíněné nezávislosti evidencí.
- ***Nevyžaduje tolik trénovacích dat*** jako ostatní metody.
- Výhodné je použití zejména v případě ***kategoriálních dat***.

24.4.3 Omezení Bayesovské klasifikace

- Jako nevýhoda je uváděna ***nižší srozumitelnost vytvářených znalostí***, jež jsou prezentovány za využití hodnot pravděpodobností (opět lze tuto nevýhodu opomenout, pokud využíváme získané znalosti v automatických systémech).
- V některých případech může být až ***příliš zjednodušený.***
- Vyžaduje velmi dobré pochopení ***nezávislých prediktorů***, jejichž existence je v praxi téměř nulová.

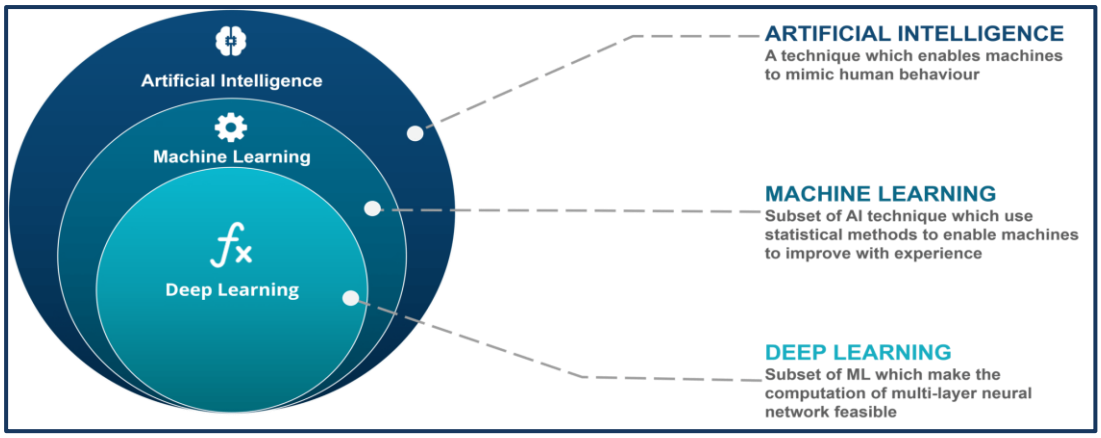
24.5 Závěry, doporučení



Z kapitoly vyplývají následující ***závěry***:

- Do základní specifikace byly v tomto dokumentu ***zařazeny tyto metody*** pro prediktivní analytiku:
  - rozhodovací stromy,
  - regresní analýzy,
  - neuronové sítě,
  - Bayesovská klasifikace.
- Uvedené metody tvoří ***pouze výběr***, přičemž i každá z nich ***má nejrůznější modifikace*** a vlastní specifika.
- Je proto velmi důležité již v počátečních fází řešení ***vybrat takovou nebo takové*** z nich, které budou ***nejlépe odpovídat potřebám prediktivní analytiky v dané firmě***. K tomu patří ***tato doporučení*** (podle Wilson, 2021):
  - nepředpokládat, že jednou metodou se vyřeší všechny problémy a úlohy,
  - udržovat při výběru a vlastním řešení přiměřenou jednoduchost,
  - metody musí odpovídat reálným potřebám a situaci ve firmě, „výzkum“ je v těchto případech zbytečný luxus,
  - metody musí odpovídat i možnostem a kapacitám (např. časovým) jejich budoucích uživatelů,
  - metody musí respektovat reálně dostupné aktuální datové zdroje a provozované aplikace.

E) Machine Learning



[25] Principy machine lerarning *(základní termíny a přístupy)*	**[26] Typy machine learning podle typů učení** *(řízení učení, dávkové / online, instance / modely)*
[27] Řešení machine learning *()*	**[28] Faktory machine learning** *()*
[29] Využití machine learning *(oblasti uplatnění, využití podle typů analytických úloh)*	
[30] Prediktivní analytika a machine learning *()*	

Machine Learning (Wilson, 2021) je ***podmnožina umělé inteligence***. Je to soubor různých technik, metod, modelování a programování, které dovolují systému se automaticky učit. Další vymezení: Arthur Samuel: Je to disciplína, která dává počítačům schopnost se učit, aniž by byly explicitně programovány.

Hlavní zdroje:

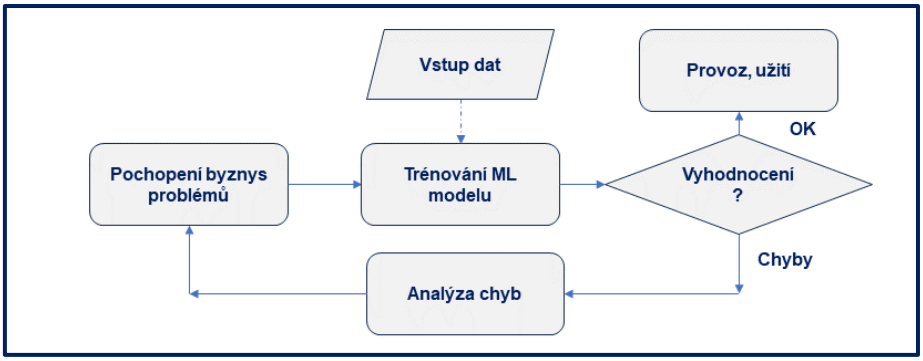
- GÉRON, A.: Hands-On Machine Leraning with Scikit-Learn, Keras and TensotFlow. O´Reilly, 2023. ISBN: 978-1-098-12597-4.
- Příklady kódů k řešení úloh: https://github.com/ageron/handson-ml3
- Tutoriály k publikaci: https://homl.info/tutorials
- Scikit-Learn: https://scikit-learn.org : jednoduché užití, zahrnuje implementace algoritmů, představuje efektivní vstup do ML prostředí.
- TensorFlow: https://tensorflow.org : zejména pro rozsáhlé neuronové sítě, distribuované zpracování modelů
- Keras: https://keras.io : deep learning s využitím neuronových sítí

Další text je založen na (Géron, 2023) a částečně upraven.

25. Principy machine learning

Machine learning je postaven na následujících principech:

- Je to ***algoritmus nebo technika***, která dovoluje systému být ***trénován a učit se na vzorech*** ze vstupů a následně je upravovat podle zkušeností, aniž by musely být explicitně programovány.
- ***Základní termíny:***
  - trénovací sada („*training set*“): případy, které systém využívá k učení,
  - jednotlivé trénovací případy: trénovací instance („*training instance*“),
  - model: část ML systému, která se učí a vytváří predikce (např. neuronová síť),
  - přesnost („*accuracy*“): míra kvality a výkonu systému
- ***Přístup k řešení*** ML, viz obrázek:



Obrázek 25-1: Přístup k řešení ML (Zdroj: Géron, 2023)

26. Typy ML systémů podle typů učení

26.1 Klasifikace dle řízení učení

26.1.1 Řízené učení („supervised learning“)

- Tréninková sada na vstupu algoritmu obsahuje požadované strukturované řešení („*labels*“).
- Typickou úlohou je zde klasifikace, nebo predikování cílové ceny produktu, požadovaných vlastností produktu, a to na bázi regrese apod.
- Pokouší se modelovat vztahy a závislosti mezi cílovou predikcí, prognózou a vstupními vlastnostmi tak, že mohou predikovat výstupní hodnoty pro nová data založené na vztazích, které byly „naučeny“ na předchozích datových sadách. Spoléhají se zde na experty, kteří fungují jako učitelé k naplnění trénovacích dat, která pomáhají modelu získávat správné odpovědi.

26.1.2 Neřízené učení („unsupervised learning“)

- Učí se z prostých, jasných příkladů bez připojených odpovědí, kde algoritmu je ponechána možnost určit, které datové vzory využije.
- Sytém nevyužívá přítomnost učitele.
- Typickými úlohami jsou:
    - clustering, resp. hierarchický clustering, který dělí jednotlivé skupiny prvků na podskupiny („*hierarchical clustering*“),
    - redukce dimenzí pro zjednodušení obsahu dat („*dimensionality reduction*“),
    - výběr, resp. vyloučení určených vlastností („*feature extraction*“),
    - identifikace anomálií („*anomaly detection*“),
    - identifikace nových instancí odlišujících se od ostatních („*novelty detection*“),
    - zjišťování asociačních pravidel („*association rule learning*“).

26.1.3 Semi-supervised learning

Vytváření požadovaných strukturovaných řešení („*labels*“) je často časově a nákladově náročné, proto se v některých případech kombinují „*labels*“ a instance bez strukturovaných řešení. Ty se označují jako „*semi-supervised learning*“ a jsou kombinací „*supervised learning*“ a „un*supervised learning*“ algoritmů.

26.1.4 Self-supervised learning

Přístup, který zahrnuje generování plně „labeled“ datového setu z plně „unlabeled“ datového setu. Pak lze využít kompletně „*supervised learning*“. Tento přístup se označuje jako „*self-supervised learning*“.

26.1.5 Reinforcement learning

„*Reinforcement learning*“, resp. posilované učení se označuje jako agent, který může vyhodnocovat prostředí, vybírat a vykonávat akce a získávat odměny nebo negativní odměny (penalizaci). Učí se, co je nejlepší strategie, resp. politika, jak získávat odměny. Politika určuje, jakou akci má agent vybrat pro danou situaci.

26.2 Dávkové a online učení

Dané rozlišení zahrnuje situace, zda je systém schopen se postupně učit z průběžně přicházejících dat, či nikoli.

26.2.1 Dávkové učení („batch learning“)

Systém se nemůže učit průběžně, musí být trénován na základě všech dostupných dat. Současně to představuje „*offline learning*“, to je systém je nejprve trénován a pak jako celek jde do provozu bez možnosti dalšího trénování. Systém postupně ztrácí na hodnotě vzhledem k tomu, že zatímco se prostředí mění, systém zůstává stejný. Tento stav se označuje jako „*model rot*“ nebo „*data drift*“. Sytém musí v pravidelných fázích procházet dalším trénováním, což závisí na úlohách, pro které je využíván. Musí se počítat s tím, že trénování systému na kompletním setu dat je obvykle časově i nákladově vysoce náročné.

26.2.2 Učení online („online learning“)

Systém je trénován průběžně na základě aktuálních vstupů dat. Trénování je tak výrazně časově méně náročné. Využívá se v prostředích častých a rychlých změn a při omezeném rozsahu výpočetních zdrojů. Může být rovněž využit na velkých objemech dat, které však nemusí být najednou v paměti – „*out-of-core learning*“.

Podstatný parametr těchto systémů je, jak rychle se může adaptovat na změněná data, a označuje se jako „*learning rate*“. Problém je v případě chybných dat, kdy systém reaguje velmi rychle a snižuje se jeho výkon. Je tak nutné průběžně systém monitorovat.

26.3 Učení založené na instancích nebo modelech („instance-based / model-based learning“)

Klasifikace je založená na tom, do jaké míry je systém schopen generalizovat, to znamená, jak je schopen poskytovat dobré predikce, generalizovat příklady.

26.3.1 Instance-based learning

Systém se učí na základě příkladů zpaměti, které generalizuje do nových případů s využitím úrovně podobnosti a porovnává je s naučenými příklady.

26.3.2 Model-based learning

Je přístup generalizace setu příkladů a vytvoření modelu příkladů, který je následně využíván k predikci. Prvním krokem je výběr modelu a určení míry jeho výkonu. Definuje se tzv. „*utility function*“ (resp. „*fitness function*“) určující, jak je model dobrý. Na druhé straně slouží „*cost function*“, tj. jak je model špatný.

27. Řešení ML

Řešení machine learning typicky obsahuje ***tyto kroky*** (Žydyk, 2020):

- ***získání dat*** – pro výběr a naučení modelu budou potřeba podkladová data.
- ***explorace a validace*** – prvotní prozkoumání a tzv. profiling dat – zjišťuje se, jak data vypadají jako celek, co v nich je, co by mohlo chybět, jaká je struktura atp.
- ***čištění (transformace, předzpracování) dat*** – nejnáročnější operace procesu – data je třeba řádně připravit (vyřešit chybějící či odlehlé hodnoty, provést kódování textových proměnných, vytvořit nové prediktory atp.).
- ***rozdělení dat*** – když jsou data připravena, rozdělí se na trénovací a testovací množinu pro učení a ověření kvality modelu.
- ***výběr a trénování modelu*** – dle povahy úlohy je třeba vybrat druh modelu (regrese, klasifikace…) a provést trénování (učení).
- ***vyhodnocení výkonu modelů*** – pro dosažení co nejlepšího výsledku se zpravidla testují různé druhy modelů s různými hyperparametry.
- ***uložení nejlepšího modelu*** – nejlepší model je uložen a zaverzován pro další použití.
- ***převedení do produkce*** – model je třeba zpřístupnit pro použití v konzumujících aplikacích.
- ***monitoring (správa) modelu v produkci*** – model v produkčním prostředí je třeba dále monitorovat (chyby, rychlost, zátěž…

28. Faktory machine learning

Hlavní faktory z pohledu řešení představují nekvalitní modely a nekvalitní data. Další části se nejprve věnují faktoru nekvalitních dat.

28.1 Nedostatečné množství trénovacích dat

Obvykle je potřeba pro jednoduché problémy tisíce příkladů v datech, pro komplexní problémy (např. rozpoznávání obrazů, či hlasu) se vyžadují miliony příkladů.

28.2 Nerepresentativní trénovací data

Pro generalizování jsou nezbytná trénovací data, která jsou reprezentativní vzhledem k novým případům, které se mají generalizovat. Pokud se využívají nereprezentativní data, pak modely, které je pro trénování používají poskytují chybné predikce.

28.3 Špatná kvalita dat

Může zahrnovat:

- Případy, kdy instance zahrnují hodnoty mimo obvyklé limity („*outliers*“).
- Některé instance neobsahují potřebné vlastnosti, parametry („features“), např. věk zákazníka. „*Features engineering*“ zahrnuje tyto kroky:
  - „*Features selection“* – výběr nejpoužívanějších vlastností.
  - „*Features extraction“* – kombinování existujících vlastností.
  - Vytváření nových vlastností na základě nových dat.

28.4 Validace a ladění modelů

Aby se zajistilo, že je naučený prediktivní model co nejpřesnější, ***musí být testován pomocí skupiny dat out-of-sample, testovacích dat***, která nijak nevstoupila do vývoje a učení modelu. Ověřuje se schopnost predikce modelu na této skupině dat a ***porovnává se, jak moc se odchyluje od výsledků z učících dat***. Pokud je odchylka velká, znamená to, že model není optimálně generalizován. Nabyté znalosti jsou poté aplikovány na model, který se podle nich upraví a následně opět testuje.

Z důvodu vlastnosti generalizace ***nejsou prediktivní modely při predikci úspěšné na 100 %***. V každém prediktivním modelu se objevují ***2 základní typy chyb*** (NYCE, 2007):

- Typ I. – chybně predikovaný výskyt, ve skutečnosti není výskyt,
- Typ II. – chybně predikovaný nevýskyt, je ve skutečnosti výskyt.

Využívané metriky modelu, pomocí kterých se vyhodnocuje ***úspěšnost modelu, jsou Lift, ROC, Missclassification Rate***.

Lift je ***ukazatel popisující úspěšnost predikce*** a použitelnost prediktivního modelu nad bází dat. Definuje, jak moc je model úspěšnější než náhodný výběr (SIEGEL, 2013). Úspěšnost prediktivních modelů se ***nad měnící se hloubkou báze dat různí***. Pro usnadnění porovnání modelů nad různými hloubkami báze se zavedl pojem decil. ***Jeden decil obsahuje 10 % záznamů skórované báze*** seřazené sestupně dle predikovaného skóre. V prvním decilu obsahuje náhodný výběr 10 % všech pozitivních targetů, ve dvou decilech 20 % a v celé bázi dat jsou odhaleny všechny pozitivní targety. Nejvyššího liftu modely zpravidla dosahují na začátku v prvním decilu a má klesající tendenci. Lift náhodného výběru je 1 nad celou bází dat. ***Úspěšný model správně určí v prvních čtyřech decilech 80 % pozitivních targetů*** (ČIČO, 2013).

ROC (Receiver operating characteristic) je dalším typem zobrazení úspěšnosti predikce modelu. Využívá se pro vyhodnocení úspěšnosti predikce binárního targetu.

Misclassification Rate (míra chybovosti) je poměr špatně vyhodnocených případů k celku. Výpočet hodnoty misclassification rate = počet špatně klasifikovaných (pozitivních i negativních targetů) / celkový počet klasifikací (ZHONG, 2007).

Overfitting (také overlearning) přeučení modelu znamená, že model ***špatně vyhodnocuje náhodný šum v datech***, určuje důležité vztahy na základě náhodných proměnných a postrádá

schopnost generalizace. Úspěšnost predikce modelu na nových datech je oproti ideálnímu stavu snížena. Tento nežádoucí stav může být mimo jiné ***způsobený následujícími případy***:

- špatné nastavení modelu, například příliš velký (hluboký, rozvětvený) rozhodovací strom, příliš komplexní neuronová síť,
- příliš malý vzorek učících dat,
- chyby ve vstupních proměnných, které nebyly řádně očištěny.

Přeučení znamená, že ***model z dostupných dat předpokládá příliš mnoho***. Pokud bude například rozhodovací strom příliš hluboký a každý jeho list bude mít pouze jeden záznam, pak je strom, se všemi svými pravidly, pouze jinou interpretací tréninkové množiny dat. (SIEGEL, 2013). Na tréninkových datech bude tento model vykazovat 100% úspěšnost predikce, ale na testovacích, a hlavně nových datech bude model méně úspěšný, než kdyby byl optimálně generalizovaný.

Underfitting (nedoučení modelu) znamená, že ***učení modelu bylo chybou nastavení modelu***, nebo nedostatkem dat ***zastaveno příliš brzy*** a nebyly odhaleny všechny důležité vztahy. Model je příliš obecný a jednoduchý. Například nedoučený strom se skládá z příliš malého počtu pravidel a má málo listů.

Pruning (prořezávání) je jedna z metod ***optimalizace modelů***. Účelem je ***snížení komplexity*** modelu. Například u rozhodovacích stromů představuje pruning „prořezávání“ větví modelu, za účelem snížení počtu větví a listů a tím snížení rizika overfittingu (SIEGEL, 2013), u neuronových sítí zase snížení počtu neuronů a vrstev Hidden Layer (MATIGNON, 2005). V praxi se k automatizaci tohoto procesu využívá validační skupina dat (SIEGEL, 2013).

28.5 Vyhodnocení modelu

Po vytvoření prvních modelů se ***výsledky těch nejúspěšnějších testují v praxi na nových datech***. Na základě predikce se ***uskuteční rozhodnutí*** a nastane nějaká akce. Poté se vyhodnocuje, jak moc predikce odpovídá realitě, ***vyhodnocuje se úspěšnost modelu*** v praxi. Ta bývá zpravidla řádově nižší, než model vykazuje na učících, či testovacích datech. V některých případech může být ***objektivní vyhodnocení problematické***, protože provedená akce ovlivní chování jedince a není tak možné zjistit jeho chování, když by akce nenastala.

Příkladem může být ***marketingová kampaň*** telekomunikační společnosti zaměřená na odcházející klienty, kde prediktivní modely vyhodnocují pravděpodobnost odstoupení klienta od smlouvy a jeho přechodu ke konkurenci. Klienti s nejvyšším skóre jsou společností osloveni a je jim nabídnuta nějaká výhoda. V tomto případě se však nedá přesně ověřit, zdali klienti, vyhodnocení jako nejnáchylnější k odstoupení, skutečně odstoupit chtěli, tedy zdali je model vyhodnotil správně jako nejnáchylnější k odstoupení. Dá se pouze měřit, kolik lidí skutečně odstoupilo a kolik ne. Řešení tohoto problému spočívá ve ***využití kontrolních skupin***, kdy se záměrně část klientů s vysokým skóre neosloví a sleduje se u nich jejich chování neovlivněné prediktivním modelem.

29. *Využití* machine learning

Hlavní oblasti uplatnění ML podle (Géron, 2023) jsou tyto:

- v problémech, kde současná řešení vyžadují obrovské množství času a ladění, resp. Jsou založena ne neúměrném počtu pravidel,
- při řešení velmi komplexních problémů, kde tradiční postupy neposkytují odpovídající výsledky,
- u řešení problémů v nestabilním prostředí, kde se data velmi rychle měním
- pro řešení komplexních problémů založených na velkých objemech dat.

Využití machine learning v jednotlivých ***typech analytických úloh*** je následující:

- ***deskriptivní analytika*** – využívá obdobných metod jako data mining. Využívají metody clusteringu a asociační pravidla. Příkladem klastrů je vytváření zákaznických skupin podle nákupních zvyklostí. Příkladem asociačního pravidla může být „kdo nakupuje zboží A, obvykle nakupuje i zboží B“.
- ***diagnostická analytika*** – využívá machine learning pro identifikace různých anomálií např. u detekcí podvodů nebo hledají vzory dat pro klasifikace událostí nebo vlastností produktů, nebo pro řešení úloh regrese.
- ***preskriptivní a kognitivní analytika*** – zahrnuje i „deep learning“, což je podmnožina machine learning založená na algoritmech, které se pokouší emulovat lidské myšlení a chování při řešení komplexních problémů. Deep learning nastavuje základní parametry dat a počítač má s využitím mnoha vrstev zpracování vytvářet potřebné vzory.

30. Prediktivní analytika a machine learning

Pro prediktivní analytiku machine learning nabízí nové možnosti a pokročilé metody při jejím využití a dosahuje se lepších prognóz. Představuje dnes ***součást všech typů analytických úloh***, od deskriptivní po kognitivní včetně prediktivní analytiky. ML je ***využíváno v mnoha aplikacích***, jako např. Amazon, YouTube, Netflix v souvislosti s dalším doporučováním obsahu podle předchozích preferencí a nákupů zákazníka. Např. Amazon podle předchozích nákupů doporučuje další tituly knih nebo publikací souvisejících s přechozími nakupovanými tématy. ***Business intelligence*** aplikace využívají machine learning algoritmy k identifikování dalších použitelných dat.

Prediktivní analytika – využívá machine learning a deskriptivní proměnné pro prediktivní modelování jako neuronové sítě, extrapolace, regresní modely, případně pro spojování několika modelů do jednoho celku. Např. u neuronových sítí lze využít data z předchozích marketingových kampaní (jako věk, příjem atd.) a určovat příležitosti a rizika spojená s novými kampaněmi.

31. Závěry

Text tvoří jeden z doplňujících textů k základnímu textu „AF_II_05_Podnikova analytika“ řady textů „IT a anatomie firmy“ v tomto případě zaměřený na prediktivní analytiku a její řízení. V tomto případě ***bylo cílem presentovat celkový pohled na prediktivní analytiku firmy***, s tím, že hlavní pozornost byla zaměřena na její využití v byznysu firmy a vazby k obsahu jednotlivých oblastí řízení. Proto se aspekty modelů a metod chápou jako sekundární s nezbytnými odkazy na specializovanou literaturu.

Právě ***podniková a současně i prediktivní analytika*** v rámci řízení firem a vstupující do řešení projektů realizovaných v tomto prostředí je ***pro jejich konečný úspěch velmi podstatná***. Kromě toho, jak bylo zmíněno v úvodu „*Prediktivní analytika je v současnosti nejdůležitější téma byznysu a na významu trvale nabývá vzhledem k potřebám firem udržovat svoji konkurenceschopnost a získávat konkurenční výhody na trhu”*

V souvislosti s ostatními texty jsme uvedli, že smyslem uvedeného pojetí a přístupu k analýze je přispět ke ***zvyšování kvality a výkonu*** práce analytiků, manažerů a analytiků vývojářů v reálné praxi. V případě tohoto textu to platí nemalou měrou. Jestli i tento text takový příspěvek představuje, pak se jeho smysl podařilo naplnit.

Terminologie

Activity Based Costing, ABC	**Activity Based Costing** – ABC – metoda, jejímž cílem je analyzovat informace o nákladech na jednotlivé služby a produkty v detailnějším členění
Architecture Development Method, ADM	**Architecture Development Method** – ADM – podporuje využívání jiných metodik (jako např. Zachman Framework, Federal Enterprise Architecture Framework, U.S. Department of Defense Architecture Framework). Popisuje, jak vybudovat pro daný podnik specifickou podnikovou architekturu, která reflektuje požadavky byznysu.
Aktivum	**Aktivum** – představuje cokoli, co má pro organizaci hodnotu (ISO/IEC 27001:2005, 3.1)
Advanced Planning and Scheduling, APS	**Advanced Planning and Scheduling** – APS – systémy zajišťující plánování výroby s uvažováním všech možných druhů omezení výrobního systému (materiál, kapacity, pracovníci, trh) pro jeden podnik, resp. výrobní jednotku. Systémy, které zajišťují koordinaci a optimalizaci procesů sítě organizací v reálném čase.
Architektura	**Architektura** – fundamentální **uspořádání systému**, které tvoří komponenty a vztahy mezi nimi, včetně vztahu k prostředí, a principy, které řídí jeho návrh a rozvoj. (ISO42010, 2007)
Architektura IS	**Architektura IS - celkový koncept informačního systému** – vize budoucího IS/IT, která zachycuje jednotlivé komponenty IS/IT a jejich vazby, architektura IS/IT se podle úplnosti pokrytí jednotlivých dimenzí (aspektů) IS/IT rozlišuje na architekturu služeb, aplikační architekturu a technologickou (Hindls,R.: Ekonomická encyklopedie, 2003)
Artificial Intelligence, AI, umělá inteligence	**Artificial Intelligence, AI, umělá inteligence** – schopnost systému nebo stroje zvýšit nebo automatizovat každý proces nebo výstup, který normálně vyžaduje lidskou inteligenci nebo intervenci (Wilson, 2021)
Balanced Scorecard	**Balanced Scorecard** – BSC – systém strategického měření podniku. Systém pro popis, implementaci a údržbu (úpravy) strategie na všech úrovních podniku prostřednictvím propojení cílů, iniciativ a ukazatelů do dynamicky modifikovatelné struktury
BASEL II	**BASEL II** – výstup Basilejské komise pro bankovní dohled s názvem International Convergence of Capital Measurement and Capital Standards
Benchmarking	**Benchmarking** – proces opakovaného srovnávání a měření vybrané společnosti/společností s referenčními organizacemi, a to jak v dané zemi, tak i ve světě.
Bezpečnost	**Bezpečnost** – vlastnost podnikové informatiky, jejíž úroveň je ovlivněna všemi aspekty, které souvisí s definováním, dosažením a udržováním vhodného stupně bezpečnostních požadavků (důvěrnost, integrita, dostupnost, individuální zodpovědnost, autenticita a spolehlivost) respektující kulturu a odvětví, ve kterém je IS/IT provozován.
Bezpečnostní incident	**Bezpečnostní incident** – událost nebo více událostí, u kterých existuje vysoká pravděpodobnost ohrožení bezpečnosti informací (ISO/IEC 27001:2005, 3.6)
Bezpečnostní politika	**Bezpečnostní politik**a – dokument, který v písemné podobě uvádí, jakým způsobem plánuje organizace ochranu svých aktiv. Může zahrnovat politiku přípustného užívání aktiv, specifikaci vzdělávacího procesu svých zaměstnanců v oblasti ochrany aktiv, objasnění způsobu uskutečňování a vynucování bezpečnostních opatření a proceduru vyhodnocení účinnosti politiky.
Business Intelligence BI	**Business Intelligence** – BI – souhrn technologií, a nástrojů, které umožňují na základě aplikace multidimenzionálního přístupu k měření vytvářet a optimalizovat aplikace na podporu analytických a plánovacích aktivit podniku
Business Process Modeling, BPM	**Business Process Modeling** – BPM – metoda procesního modelování
Business Process Reengineering	**Business Process Reengineering** – BPR – změny podnikových procesů – postup, který optimalizuje podnikové procesy tak, aby přinášely optimální efekty při minimální spotřebě podnikových zdrojů, důsledkem jsou změny v organizační struktuře podniku.
CAPEX	**CAPEX** – Capital expenses – investiční náklady, zpravidla jde o náklady jednorázového charakteru, které je nutné vynaložit na pořízení určitého hmotného, nehmotného i finančního majetku

Cloud computing	**Cloud computing** – metoda přístupu k využití výpočetní techniky, která je založena na poskytování sdílených výpočetních prostředků a jejich využívání formou služby. Existují nejrůznější modely služeb a možnosti jejich poskytování, ale všem typům cloud computingu je společná schopnost poskytovat prostředky na vyžádání, elasticky, samoobslužně a prostřednictvím přístupu z rozsáhlé sítě a také schopnost měřit spotřebované služby v rámci sdíleného fondu prostředků.
CLV	**CLV** – Customer Lifetime Value – životní cyklus hodnoty zákazníka
CMMI	**CMMI** – Capability Maturity Model Integration – model pro zlepšování procesů v oblasti rozvoje služeb
CMO	**CMO** – Chief Marketing Officer – ředitel pro marketing
CobiT	**CobiT** – Control Objectives for Information and related Technology – metodika vyvinutá a publikovaná neziskovou nezávislou organizací Information Systems Audit and Control Association (ISACA) s cílem využít mezinárodní standardy a nejlepší zkušenosti pro řízení a audit v oblasti IT.
Competitive Intelligence, CI	**Competitive Intelligence** – CI – systematický a etický program pro získávání, analyzování a řízení externích informací o trhu, konkurenci a trendech, které mohou mít vliv na podnikové plány, rozhodování a jeho činnost.
Corporate Performance Management, CPM	**Corporate Performance Management**, CPM – řízení výkonnosti podniku je souhrnným termínem, který popisuje všechny procesy, metodiky, metriky a systémy potřebné k měření a řízení výkonnosti organizace**.**
Critical Success Factor	**Critical Success Factor** – CSF– taková vlastnost (řešení informačního systému, procesu či měření podnikové informatiky), která výrazně ovlivňuje celkové efekty řešení, resp. může přispět k celkové úspěšnosti podniku
Customer Relationship Management, CRM	Customer Relationship Management, CRM – aplikace podnikové informatiky pro řízení vztahů k zákazníkům.
Cross docking	**Cross docking** – princip přijetí dodávky do distribučního centra, kde následně dochází k její dekonsolidaci a konsolidaci, tj. kompletaci dle požadavků odběratele, doba skladování však nepřesahuje 24 hodin. Jednotlivé dodávky v cross-docking centru mají již předem známého odběratele, je známa lokalita a čas, kde má být zásilka doručena.
Capacity Requirements Planning, CRP	Capacity Requirements Planning, CRP – určování úrovně kapacit a jejich obsazení, stanovování množství práce a strojového času potřebného ke splnění výrobních zakázek
Data Mart, DMA	**Data Mart** – DMA, datové tržiště, analytická databáze určená pouze pro určitou část podniku (útvar, oblast řízení apod.).
Data Science	**Data Science** – interdisciplinární obor, využívající celou řadu metod, jako je statistika, strojové učení, datová analýza k analyzování a porozumění strukturovaným i nestrukturovaným datům. (Chudán, 2023)
Data Warehouse, DWH, datový sklad	**Data Warehouse** – DWH – datový sklad je integrovaný, subjektově orientovaný, stálý a časově rozlišený souhrn dat, uspořádaný pro podporu potřeb managementu.
Datová pumpa, ETL, Extract, Transform, Load	**Datová pumpa** – ETL– Extract, Transform, Load - datová pumpa - software pro realizaci transformací dat mezi různými datovými zdroji
Diagram tříd	**Diagram tříd** – Class diagram – jeden z diagramů UML, vyjadřuje třídy a vztahy mezi nimi
Diagram užití	**Diagram užití -** Use Case diagram – jeden z diagramů UML, vyjadřuje chování systému a vztah k okolí
Dimenze	**Dimenze** – analytické hledisko pro identifikaci a hodnocení sledovaných ukazatelů a je tak součástí de facto každé metriky. Z informatického pohledu se jeví jako struktura dat, případně jako databázová tabulka obsahující záznamy o jednotlivých prvcích dimenze, tj. např. o jednotlivých službách informatiky, dodavatelích IT apod.
Dimenzionální modelování	**Dimenzionální modelování** – vymezení všech dimenzí, jejich obsahu, včetně vnitřní hierarchie prvků, a dílčích charakteristik jednotlivých dimenzí, určení soustavy sledovaných ukazatelů (faktů) a jejich dílčích charakteristik, specifikace vazeb mezi ukazateli a odpovídajícími dimenzemi.

Dimenzionální tabulka	**Dimenzionální tabulka** – tabulka obsahující kontext k datům fakt tabulky, obsahuje obvykle textové popisné údaje.
Data Mining, DMI, dolování dat	**Data Mining, DMI, dolování dat** – proces získávání vzorů ve velkých datasetech, který zahrnuje metody ze strojového učení, statistiky a databázových systémů. Jeho cílem je extrakce informací z dat a jejich transformace takovým způsobem, aby byly použitelné a srozumitelné pro další využití. (Chudán, 2023)
Data Staging Area, DSA	Data Staging Area, DSA – jde o místo v architektuře datového skladu, kde se data ze zdrojových systémů zpracovávají a transformují do podoby dat definovaných dimenzionálním (cílovým) modelem.
e_Business	**e_Business** – podpora (některých) oblastí podnikání prostředky IS/ICT, zejména internetu a příslušnými aplikačními programy.
e_Commerce, elektronický obchod	**e_Commerce** – elektronický obchod (B2C) se realizuje www aplikacemi v prostředí internetu a poskytuje spotřebitelům např. služby inzerce a speciální nabídky zboží, elektronické katalogy nabízeného zboží a s nimi spojené vyhledávací služby a další.
e_Marketplace, elektronické tržiště	**e_Marketplace** – aplikace elektronického podnikání, které v prostředí internetu vytvářejí prostor pro uskutečňování mnohostranných elektronicky realizovaných obchodních transakcí. Transakce se zde uskutečňují mezi mnoha obchodními partnery ve vazbách M:N. Vytváří se virtuální obchodní komunita s optimalizovanými řídícími a obchodními procesy.
Enterprise Application Integration, EAI	**Enterprise Application Integration, EAI** – souhrnný integrační rámec umožňující integrovat jednotlivé aplikace informačního systému na základě požadavků podniku a výběru vhodného technologického řešení.
Efficient Consumer Response, ECR	**Efficient Consumer Response, ECR** – efektivní reakce na požadavky zákazníků, vedoucí k efektům vyplývající z eliminace činností nepřidávajících hodnotu. Podstata této technologie spočívá v intenzivní spolupráci mezi obchodem a průmyslem s cílem splnění přání konečných spotřebitelů s co nejpříznivějšími logistickými náklady.
Electronic Data Interchange, EDI, Elektronická výměna dat	**Electronic Data Interchange, EDI** – elektronická výměna dat založená na světově uznávaných standardech
EDI standard	**EDI standard** – dohodnutá representace informace (položky, jejich struktura a kompletace), která se má přenášet z jedné počítačové aplikace do jiné. Standard EDI definuje způsob vyjádření jednotlivých částí a jejich seskupování do obchodních dokumentů nebo zpráv.
Elektronický obchod	**Elektronický obchod** – B2C – se realizuje www aplikacemi v prostředí Internetu a poskytuje spotřebitelům např. služby inzerce a speciální nabídky zboží, elektronické katalogy nabízeného zboží a s nimi spojené vyhledávací služby a další.
Enterprise Resource Planning, ERP	Enterprise Resource Planning, ERP – aplikace mající celopodnikový charakter a pokrývající funkcionalitou většinu, resp. velkou část funkcí podnikového řízení
Fakt	**Fakt** – metrika (zpravidla číselná hodnota), která je v podniku sledována.
Fakt tabulka	**Fakt tabulka** – tabulka obsahující obchodní fakta a hodnoty, zpravidla číselné a aditivní.
Granularita	**Granularita** – úroveň podrobnosti faktů uložených ve fakt tabulce.
In-memory computing	**In-memory computing** – výkonná databáze umožňující real-time zpracování enormního množství operací s cílem zrychlení jednotlivých transakcí jak na úrovni procesů, tak celých modulů
Incident	**Incident** – každá událost, která narušuje průběh poskytování služby nebo snižuje její kvalitu.
Insourcing	**Insourcing** – proces, který zajistí převod zodpovědnosti za danou službu/proces/zdroj z externího poskytovatele na podnik.
IT Balanced Scorecard	**IT Balanced Scorecard** – nástroj pro řízení podnikové informatiky a vztahu mezi informatikou a byznysem.
IT Governance	**IT Governance** – tvoří procesy, které ve firmě zajišťují efektivní a účinné rozhodování o IT a jeho využití k realizaci cílů byznysu. Procesy IT Governance mají své definované vstupy, výstupy, role a zodpovědnosti za rozhodování o IT.
IT Infrastructure Library, ITIL	IT Infrastructure Library, ITIL – jde o sadu publikací popisujících nejlepší praktiky řízení IT služeb a o systém certifikací a školení jednotlivců ve znalostech těchto praktik.

Just in Time, JIT	Just in Time, JIT – logistická technologie „právě včas“ – tj. umožňuje podniku vyrábět výrobky v přesně dohodnutých a dodržovaných termínech dle potřeb zákazníka. Smyslem je dodávat velmi často, v malém množství, a v co možná nejpozdějším okamžiku.
Kanban	**Kanban** – bezzásobová technologie vyvinutá společností Toyota Motors. Její princip spočívá v synchronizaci činnosti dodavatele a odběratele za předpokladu stability dodavavatelského řetězce.
Key Goal Indicator, KGI	**Key Goal Indicator** – KGI – cílově orientovaná metrika, tj. měří výsledek procesu, např. počet úspěšně vyřešených incidentů v procesu řízení incidentů.
Key Performance Indicator, KPI	**Key Performance Indicator** – KPI – metrika výkonnosti jednotlivých složek procesu (činností a zdrojů procesu), tj. měří jednotlivé činnosti a zdroje procesu, např. dobu trvání činnosti, objem zpracovaných dokumentů, % využití doby operátora service desku, apod.
Klíčová aktivita	**Klíčová aktivita** – činnost, která je součástí úlohy a má na výsledek úlohy podstatný vliv.
Kontaktní centrum	**Kontaktní centrum** – aplikace v rámci CRM. Kontaktní centra jsou vybavena počítači s přístupem k centrální zákaznické databázi. V ní se uchovávají a pravidelně aktualizují informace o jakémkoliv kontaktu se zákazníky, např. vyřizování stížností.
Kontinuita	**Kontinuita** – strategická a taktická způsobilost organizace plánovat reakci a reagovat na incidenty a přerušení činnosti za účelem pokračování provozu organizace na přijatelné, předem stanovené úrovni (BS 25999-1:2006, 2.2)
Konzistence	**Konzistence** – bezrozpornost modelů - stav, kdy mezi skutečnostmi vyjadřovanými jednotlivými modely není rozpor
Logistika	**Logistika** – disciplína, která se zabývá slaďováním, koordinací, synchronizací a celkovou optimalizací všech aktivit v rámci samoorganizujících se systémů, jejichž zřetězení je nezbytné k pružnému a hospodárnému dosažení daného synergického efektu.
Logistický řetězec	**Logistický řetězec** – propojení trhu surovin a trhu spotřeby vycházející od konečného spotřebitele. Jednotlivé procesy v řetězci mají hodnototvorný charakter. Má stránku hmotnou (hmotné toky) a stránku nehmotnou (informační toky).
m-Business	**m-Business** – Mobile Business – mobilní podnikání – souhrn veškerých aktivit, procesů a aplikací, které jsou uskutečňovány nebo podporovány mobilními technologiemi
m-Commerce	**m-Commerce** – Mobile Commerce – mobilní obchodování – každá transakce vykonaná prostřednictvím mobilních koncových zařízení, resp. použití mobilní komunikace s jakoukoliv aplikací dostupnou a vytvořenou pro mobilní koncová zařízení, jejíž využití slouží k obchodním účelům
Machine Learning, ML	**Machine Learning, ML** – Podoblast umělé inteligence, která se zabývá tvorbou algoritmů, které dokážou zpracovávat data a vytvářet modely v datech, aniž by k tomu byly explicitně naprogramovány. (Chudán, 2023)
Metadata	**Metadata** – informace o datech, tj. názvy tabulek, názvy atributů, datové typy, primární klíče, vazby, komentáře atd.
Metoda	**Metoda** – představuje známou a v rámci úloh modelu MBI aplikovatelnou metodu manažerského (např. BSC, RPZ atd.), ekonomického (např. ABC, ROI atd.) nebo informatického charakteru (datové modelování, dimenzionální modelování atd.).
Metodika	**Metodika** – doporučený souhrn fází, etap, přístupů, zásad, postupů, pravidel, dokumentů, řízení, metod, technik a nástrojů pro tvůrce informačních systémů, který pokrývá celý životní cyklus informačního systému.
Metrika	**Metrika** – sledovaná a měřená hodnota ukazatele pro potřeby řízení podniku a informatiky. K ukazatelům se váží dimenze pro jejich identifikaci, analýzy a plánování. Vychází se přitom z principů dimenzionálního modelování.
Multidimenzionální databáze	**Multidimenzionální databáze** – databáze, kde jsou data uložena na principu vícerozměrové matice. Hodnoty jsou přístupné přímo pro danou kombinaci prvků dimenzí.
NACE norma	**NACE (Nomenclature statistique des activités économiques dans la Communauté européenne)** je zkratka pro klasifikaci ekonomických činností vydávanou Evropskou komisí od roku 1970. V Česku se začala používat od 1. ledna 2008, kdy

	nahradila Odvětvovou klasifikaci ekonomických činností (OKEČ). Předpona CZ (CZ NACE) určuje, že se týká činností prováděných v České republice, v Česku má sběr těchto dat na starosti Český statistický úřad. Klasifikace činností NACE se používá například pro vydávání oprávnění k podnikání (např. živnostenské listy) a pro zařazení předmětů podnikání. Klasifikace se také používá pro zpracování statistických dat a jejich srovnání mezi státy. Díky tomu lze činnosti porovnávat mezi jednotlivými evropskými státy, protože je její používání ve státech Evropské unie povinné. (Zdroj: Wikipedia)
ODBC	**ODBC** - Open Database Connectivity – standard datového rozhraní databází
Operational Level Agreement, OLA	**OLA** – Operational Level Agreement je formální mechanismus, který zajišťuje spolupráci interních a externích dodavatelů služeb IT při plnění jejich nejčastěji outsourcingového kontraktu.
Online Analytical Processing, OLAP	**OLAP** – Online Analytical Processing – informační technologie založená především na koncepci multidimenzionálních databází. Jejím hlavním principem je několikadimenzionální tabulka umožňující rychle a pružně měnit jednotlivé dimenze a měnit tak pohledy uživatele na modelovanou ekonomickou realitu.
Operative expenses, OPEX	**OPEX** – Operative expenses – neinvestiční, tj. provozní náklady - kontinuální (nepřetržitě vznikající), které je nutné vynaložit na správu, provoz, údržbu a rozvoj prostředků, rovněž náklady, které vnikají v souvislosti s odběrem určitých služeb
Optimised Production Technology, OPT	**OPT** – Optimised Production Technology – metoda podnikového řízení orientovaná na optimalizaci výrobních procesů
Outsourcing	**Outsourcing** – zajišťování vybraných činností a IT služeb externími dodavateli. Podle předmětu se rozlišuje outsourcing rozvoje IT a outsourcing provozu IT. Totální outsourcing znamená, že dodavatel zajišťuje provoz a rozvoj zákazníkovi kompletně.
Project Management Body of Knowledge, PMBOK	**Project Management Body of Knowledge, PMBOK** – metodika řízení projektů vytvářená organizací PMI (Project Management Institute).
Portál	**Portál** – množina technologií a aplikací, tvořící universální rozhraní, jehož prostřednictvím je umožněno účastnit se procesů organizace, přistupovat ke všem relevantním informacím, komunikovat s ostatními participujícími lidmi a realizovat adekvátní aktivity spojené s podnikovými procesy.
PRINCE2	**PRINCE2** – Projects in Controlled Environments **-** metodika řízení projektů vytvářená britským úřadem OGC (Office of Government Commerce).
RACI	**RACI** – matice RACI přiřazuje a zobrazuje odpovědnosti jednotlivých osob či pracovních míst za danou oblast v organizaci: R - Responsible (vykonává), A - Accountable (zodpovídá), C - Consulted (konzultuje), I - Informed (má být informován).
Referenční model	**Referenční model** – model, který je, vedle metodické stránky řešení, naplněn předdefinovaným obsahem. Tento obsah, který tvoří například podnikové procesy a jejich prvky, vzniká a postupně se rozvíjí na základě poznatků a zkušeností z dosud realizovaných projektů pro různé zákazníky v praxi.
Results Indicators, RI	**Results Indicators**, **RI** – se vztahují k dějům, co byly ukončeny. Může jít o děje dlouhodobé např. fáze, dlouhodobý projekt, nebo o děje krátkodobé např. aktivita nebo proces.
Rozhodovací strom	**Rozhodovací strom** (decision tree) – strom, jehož uzly obsahují testy hodnot atributů a větve z uzlů vycházející reprezentují jednotlivé hodnoty daného atributu.
Řízení dodavatelských řetězců, Supply Chain Management – SCM	**Řízení dodavatelských řetězců** – Supply Chain Management – SCM – řízení všech procesů v rámci dodavatelského řetězce
Řízení pracovních toků, work flow	**Řízení pracovních toků** – WorkFlow – automatizace celého nebo části podnikového procesu, během kterého jsou dokumenty, informace nebo úkoly předávány od jednoho účastníka procesu ke druhému podle sady procedurálních pravidel.
Řízení výkonnosti	**Řízení výkonnosti** – kombinace managementu, metodik a metrik podporovaná aplikacemi, nástroji a infrastrukturou, která umožňuje uživatelům definovat, monitorovat a optimalizovat výsledky a výstupy tak, aby bylo dosaženo cílů osobních či cílů organizační jednotky v souladu se strategickými cíli stanovenými na různých úrovních řízení podniku (osobní, procesní, skupinové a korporátní cíle podniku nebo podnikatelského ekosystému).

Software-as-a-Service, SaaS	**SaaS** – Software-as-a-Service – služba, resp. model poskytování aplikací, kde jejich funkcionalita je zajišťována poskytovatelem této služby značnému počtu uživatelů prostřednictvím internetu. Specializovaný poskytovatel tak udržuje a provozuje jak samotnou aplikaci, tak IT infrastrukturu nutnou k jejímu provozu a případné doplňkové služby.
Service Level Agreement, SLA	**Service Level Agreement** – SLA – smlouva specifikuje, co je poskytovatel povinen zákazníkovi dodat, v jakém rozsahu, v jaké kvalitě a kolik za to odběratel zaplatí.
Service Oriented Architecture, SOA	**Service Oriented Architecture** – SOA – politiky, praktiky a rámce, které umožňují, aby funkcionalita aplikací byla poskytována a spotřebována jako množina služeb, a to na takové úrovni granularity, kterou potřebuje příjemce služby. Ten je oddělen od implementace služby a používá pouze jednoduché na standardech založené rozhraní.
Snowflake schéma	**Snowflake schéma** – Schéma sněhová vločka – typ dimenzionálního databázového modelu tvořeného více normalizovanými a nenormalizovanými tabulkami dimenzí.
Sourcing	**Sourcing** – podnikový proces, jehož cílem je rozhodnutí o tom, které služby, procesy a zdroje má podnik zajišťovat sám a které přenechat externím poskytovatelům, výběr nejvhodnějších poskytovatelů externích služeb, sepsání smluv s poskytovateli o obsahu a úrovni poskytovaných služeb a kontrola poskytovaných služeb a řízení vztahů s externími poskytovateli.
Star schéma	**Star schéma** – schéma hvězdy je typ dimenzionálního databázového modelu tvořeného pouze nenormalizovanými tabulkami dimenzí.
Systém pro podporu rozhodování	**Systém pro podporu rozhodování** – Decision Support Systems – systém pro podporu rozhodování. Aplikace podporující operativní, dispečerskou úroveň rozhodování v podniku, často založené na využití matematických modelů
TOGAF	**TOGAF** – The Open Group Architecture Framework – rámec podnikové architektury.
Total Quality Management, TQM	**TQM** – Total Quality Management – metoda založená na řízení kvality všemi zaměstnanci organizace, sledující dlouhodobý úspěch založený na uspokojení zákaznických potřeb
Transakční zpracování	**Transakční zpracování** – metoda zpracování dat, při které se každý požadavek zpracuje okamžitě po svém příchodu. Uživatel může v průběhu zpracování svůj požadavek upřesňovat.
Ukazatel	**Ukazatel** – sledovaný údaj (Fact,. Measure) v organizaci, například objem prodeje, počet reklamací, obrat atd.
Webové služby	**Webové služby** – WS – Web services - webové služby - volně spojené, znovupoužitelné softwarové komponenty, které sémanticky zapouzdřují oddělenou funkcionalitu a které jsou distribuovány a programově přístupné přes standardní internetové protokoly
XML	**XML** – eXtensible Markup Language – značkovací jazyk, jehož aplikací na textové soubory vznikají jednotlivé XML dokumenty. Vlastní specifikace jazyka uvádí způsob zápisu struktury dokumentu, mechanismus vytváření logických struktur v dokumentu, pravidla deklarace elementů a vlastností, apod.

Zdroje

AALST, WIL van der. Process Mining Data Science v akci. Berlni : Springer, 2016. 978-3-662-49850-7.

ADÁMEK, P.: *Prediktivní analytika ve výrobě,* DP, VŠE. 2019.

ABBOTT, D.: Applied Predictive Analytics. Principlec and Techniques for the Professional Data Analyst. John Wiley & Sons, Indianoplolis, 2014. ISBN: 978-1-118-72796-6.

ANGEL, I. O., SMITHSON, S.: Information Systems Management - Opportunities and Risks, Macmillan, 1991.

BACAL, R.: *Manager´s Guide to Performance Management*. New York, McGraw-Hill 2012. ISBN 978-0-07-177225-9.

BERKA, P. 2003. Dobývání znalostí z databází. Praha: Academia, 2003. str. 366. ISBN 80-200-1062-9,

BOTHE, O., KUBERA, O., BEDNÁŘ, D., POTANČOK, M., NOVOTNÝ, O.: Managing Analytics for Success, CRC Press, 2022. ISBN 978-1-032-20851-0

BRUCKNER, T. VOŘÍŠEK, J., BUCHALCEVOVÁ, A. a kolektiv: Tvorba informačních systémů: Principy, metodiky, architektury, Grada, 2012, ISBN 978-802477-9027.

BUCHALCEVOVÁ, A.: Metodiky vývoje a údržby informačních systémů. Praha, Grada 2004. ISBN 80-247-1075-7.

BUCHALCEVOVÁ, A., PAVLÍČKOVÁ, J., PAVLÍČEK, L.: Základy softwarového inženýrství, materiály ke cvičením. Praha, VŠE 2008.

CAO, L.: Data Science Thinking. Springer, 2018. ISBN 978-3-319-95091-4.

CARLBERG, C.: *Predictive Analytics: Microsoft Excel.* QUE, 2014, ISBN: 978-0-7897-4941-3

CRISP-DM - Home. *CRISP-DM - Home.* [Online] CRISP-DM. [Cited: 11 06, 2010.] http://www.crisp-dm.org/.

CIMLER, P., ZADRAŽILOVÁ, D. a kol.: Retail management. Praha, Management Press, 2007. ISBN: 978-80-7261-167-6.

ČIČO, M. 2013. Konzultace prediktivní analýzy. Praha, 2013.

DOHNAL, J., POUR, J.: IT v řízení podniku, Praha, Professional publishing 2016. ISBN 978-80-7431-160-4.

DOHNAL, J., PŘÍKLENK, O.: CIO a podpora byznysu. Praha, Grada, 2011. ISBN 978-80-247-4050-8.

DOHNAL, J.: Řízení vztahů se zákazníky – procesy, pracovníci, technologie. Praha, Grada 2002.

DRESNER, H.: Profiles in Performance. New York, John Wiley and Sons, 2010. ISBN: 978-0-470-40886-5.

ECKERSON, W., W.: Performance Dashboards. New Jersey, Johh Wiley & Sons 2006.

EDWADS M.: *Predictive HR Analytics: Mastering the HR Metric.* Kogan Page, 2019. ISBN: 978-0-7494-8444-6

ENGLISH, L. P.: *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing profits.* New York, John Wiley & Sons 2003.ISBN 0-471-25383-9.

FAYYAD, USAMA M., a další: *Advances in Knowledge Discovery and Data Mining.* Cambridge : AAAI Press/MIT Press, 1996. str. 560. ISBN: 0-26-256097-6.

FIBÍROVÁ, J., ŠOLJAKOVÁ, L., WAGNER, J., PETERA, P.: Manažerské účetnictví. Nástroje a metody. Praha, Wolters Kluwer, 2015. ISBN: 978-80-7478-743-0.

FITZ-ENZ J., MATTOX II J.,R,: *Predictive Analytics for Human Resources.* Wiley and SAS Business Series, 2014. ISBN: 978-1-118-89367-8.

FRIEBELOVÁ, JANA: Rozhodovací stromy. *Osobní stránka - Jana Friebelová.* [Online] 28. 12 2006. [Citace: 10. 11 2010.] http://www2.ef.jcu.cz/~jfrieb/rmp/data/teorie_oa/STROMY.pdf.

GÉRON, A.: Hands-On Machine Leraning with Scikit-Learn, Keras and TensotFlow. O´Reilly, 2023. ISBN: 978-1-098-12597-4.

HALAMA, J.: Řízení datové kvality v Hadoop Ecosystem, DP, VŠE, Praha, 2021.

HINDLS, R., HRONOVÁ, S., NOVÁK, I.: Analýza dat v manažerském rozhodování. Grada, 1999. ISBN: 80-7169-255-7.

HOLTSNIDER, B., JAFFE, B.D.: IT Managener´s Handbook. Amsterdam, Elsevier 2012. ISBN 978-0-12-415949-5.

CHANDLER, N.: The CPM Scenario. Gartner BI Summit 2008.

CHRAMOSTOVÁ, V., POTANČOK, M., POUR, J.: Byznys analytika pro manažery, Oeconomia, Praha, 2020.

CHUDÁN, D.: Vymezení pojmů pokročilé analytiky, VŠE, 2023

ISO certifikace – (MANages, a.s., [Online] @2009. [Citace: 26. září 2014.] 2009),

JANSEN, S.: *Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python*. Packt Publishing, 2020. ISBN: 978-1-83921-771-5

KELLY, J.: *Big Data Vendor Revenue and Market Forecast 2012 - 2017*. [online] 2012 [cit. 18. září 2013]. Dostupné z: http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017.

KIMBALL, R., ROSS, M.: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Indianopolis, John Wiley Publishing 2010. ISBN 978-0-470-56310-6.

KOTTER, J.P. – Vedení procesu změny: osm kroků úspěšné transformace – (Management Press 2000) – ISBN8072610155,

KOTU, V.: *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner.* Elsevier, 2015. ISBN: 978-0-12-801460-8

KRÁL, B. a kol.: Manažerské účetnictví. Praha, Management Press 2006. ISBN 80-7261-141-0.

KRÁL, Bohumil, Jana FIBÍROVÁ, Ondřej MATYÁŠ, Michal MENŠÍK, Jakub STRÁNSKÝ, Libuše ŠOLJAKOVÁ, Jaroslav WAGNER, Martin ZRALÝ a Zbyněk HALÍŘ, 2012. Manažerské účetnictví. 3. doplněné a aktualizované vydání. Praha: Management Press. ISBN 978-80-7261-217-8.

KULHAVÝ, L. - Praktické uplatnění technologií Data Mining v pojišťovnictví, VŠE, Praha, 2011,

KUNSTOVÁ, R.: Efektivní správa dokumentů. Praha, Grada, 2009. ISBN 978-80-247-3257-2.

LABERGER, R.: The Datawarehouse Mentor. New York, McGraw Hill, 2011. ISBN: 978-0-07-174532-1.

LABERGER, R.: Datové sklady. Agilní metod y business intelligence. Praha, Computer Press, McGraaw Hill, 2012. ISBN 978-80-251-3729-1.

LANEY, D., B.: Infonomics, Bibliomotion, Inc., New York, 2018. ISBN 978-1-138-09038-5.

MÁŠA, P.: Data mining – praktické aplikace. 2008.

NOVOTNÝ, O., POUR, J., BASL, J., MARYŠKA, M.: Řízení výkonnosti podnikové informatiky. Professional Publishing, Praha, 2010. ISBN 978-80-7431-040-9.

NOVOTNÝ, O., VOŘÍŠEK, J. a kol.: Digitální cesta k prosperitě. Praha: Professional Publishing 2011. ISBN 978-80-7431-047-8.

NYCE, C. 2007. *Predictive Analytics White Paper.* American Institute for CPCU/Insurance Institute of America, autor neznámý, 2007.

PALADINO, B.: Innovative Corporate Performance Management: Five Key Principles to Accelerate Results. Indianopolis, Wiley Publishing, 2011. ISBN: 978-0-470-62773-0.

PARMENTER, D.: Key Performance Indicators (KPI): Developing, Implementing,and Using Winning KPIs,

POUR, J., MARYŠKA, M., STANOVSKÁ, I., ŠEDIVÁ, Z.: Self Service Business Intelligence, Praha. Grada, 2018. ISBN 978-80-271-0616-5.

PRAUS, O.: Prediktivní analýza – postup a tvorba prediktivních modelů. DP, VŠE, 2013.

PROVOST, F., FAWCETT, T.: Data Science for Business. What You Need to Know About Data Mining and Data-Analytic Thinking. O´Reilly Media. Sebastopol. 2013. ISBN: 978-1-449-36132-7.

REŠL, Š.: Využití prediktivní analytiky pro finanční plánování firmy, BP, VSE, 2020.

ŘEPA, V.: Podnikové procesy. Praha, Grada 2007.

SEGAL, T.: How Prescriptive Analytics Can Help Businesses. In: Investopedia [online]. 2019-07-05 [cit. 2020-03-16]. Dostupné z: https://www.investopedia.com/terms/p/prescriptive-analytics.asp

SHARPE, P. K. and SOLLY, R. J.: Dealing with missing values in neural network-based diagnostic systems. *NEURAL COMPUTING & APPLICATIONS.* 1995, pp. 73-77.

SCHIESSER, R.: IT Systems Management. New York, Prentice Hall 2010. ISBN 978-0-13-702506-0.

SIEGEL, E: Predictive Analytics. New York, John Wiley & Sons, 2016. ISBN 978-1-119-14567-7.

SLÁNSKÝ, D.: Data and Analytics for the 21st Century: Architecture and Governance, Professional Publishing, 2018. ISBN 978-80-88260-16-5.

ŠOLJAKOVÁ, L. FIBÍROVÁ, J.: Reporting. Praha, Grada, 2010. ISBN 978-80-247-2759-2.

SYNEK, M. a kol.: Manažerská ekonomika. Praha, Grada 2011. ISBN 978-80-247-3494-1.

SYNEK, M., KISLINGEROVÁ, E. a kol.: Podniková ekonomika. Praha, C H Beck 2015. ISBN 978-80-7400-274-8.

TOGAF: Architecture Development Method – (The_Open_Group).

TOMEK, G., VÁVROVÁ, V.: Průmysl 4.0 aneb nikdo sám nevyhraje. Praha, Professional Publishing, 2017. ISBN 978-80-906594-4-5.

UČEŇ, P.: Zvyšování výkonnosti firmy na bázi potenciálu zlepšení. Praha, Grada 2008. ISBN: 978-80-247-2472-0.

VAN DECKER, J, CHANDLER, N.: Top Processes for Corporate Performance Management, Gartner, 2011.

VOMLELOVÁ, MARTA: Rozhodovací stromy. *Výuka - Marta Vomlelová.* [Online] 10 20, 2009. [Cited: 11 10, 2010.] http://kti.mff.cuni.cz/~marta/slistromy.pdf.

VOŘÍŠEK, J., POUR, J. a kol.: Management podnikové informatiky, Professional Publishing, 2012, ISBN 978-80-7431-102-4.

VOŘÍŠEK, J. a kol.: Principy a modely řízení podnikové informatiky. Praha, Oeconomia 2008. ISBN: 978-80-245-1440-6.

WEST, M.: *People Analytics For Dummies.* John Wiley and Sons, 2019. ISBN: 978-1-119-43476-4.

WILSON, J., E.: Predictive Analytics for Business Forecasting and Planning. Graceway Publishing Company, 2021. ISBN 978-0-9839413-8-5.

ZHONG, M. 2007. An Analysis of Misclassification Rates for Decision Trees. Florida, 2007. PhD Thesis. University of Central Florida Orlando.

ŽŮRKOVÁ, Hana, 2007. Plánování a kontrola - klíč k úspěchu. 1. vydání. Praha: Grada Publishing. ISBN 80-247-1844-8.

BORTLÍČEK, Z. ROC Křivky. Brno, 2009. Diplomová práce. Masarykova univerzita

v Brně, Přírodovědecká fakulta.

ELDER IV, J. 2003. The generalization paradox of ensembles. Journal of Computational and

Graphical Statistics, 2003. 853-864.

FAUSKE, MAGNE K. 2006. Example: Neural network. texample.net. [Online] 2006. [cit. 2013-11-04]. Dostupné z: http://www.texample.net/tikz/examples/neural-network/

HAND, D, MANNILA, H a SMYTH, P. 2001. Principles of data mining (adaptive computation and machine learning). Cambridge, MIT Press, 2001. ISBN 978-0262082907.

CHALUPNÍK, V. 2012. Biologické algoritmy (5) - Neuronové sítě. Root.cz. [Online] 2012. [Cit. 2013-8-14]. Dostupné z: http://www.root.cz/clanky/biologicke-algoritmy-5-neuronovesite/#ic=serial-box&icc=text-title

MATIGNON, R. 2005. Neural Network Modeling Using Sas Enterprise Miner. Bloomington : AuthorHouse, 2005. ISBN 978-1-4184-2341-4.

MATINGON, R. 2007. Data Mining Using SAS Enterprise Miner. Bloomington : Wiley, 2007. ISBN 978-0-470-14901-0.

Prescriptive Analytics. In: Gurobi [online]. [cit. 2020-03-15]. Dostupné z: https://www.gurobi.com/company/about-gurobi/prescriptive-analytics/

Descriptive, Predictive and Prescriptive Analytics Explained. In: Logility Blog [online]. 2020 [cit. 2020-03-15]. Dostupné z: https://www.logility.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/

ROOSEVELT, M. 2005. The use of predictive modeling in the insurance industry. místo neznámé, PINNACLE Actuarial Resources, INC, 2005.

78

SMITH, A. 2004. Branch Prediction with Neural Networks: Hidden layers and Recurrent Connections. Department of Computer Science University of California : San Diego, 2004.

STATSOFT. 2013. Electronic Statistics Textbook. [Online] 2013. [Cit. 2013-7-15]. Dostupné z: http://www.statsoft.com/Textbook

ŠEVČÍK, O. Současné trendy v business intelligence. BP, VŠE, 2020.

TEASLEY, B. 2005. Using Predictive Models, Part 1. Clickz.com. [Online] 2005. [Cit. 2013-9-4]. Dostupné z: http://www.clickz.com/clickz/column/1707212/using-predictive-modelspart

W3SCHOOLS. 2013. SQL JOIN. W3schools.com. [Online] 2013. [Cit: 2013-10-3]. Dostupné z: http://www.w3schools.com/sql/sql_join.asp