

Digitální data mining v managementu a ekonomii

Ondřej Jurášek

Bakalářská práce
2022



Univerzita Tomáše Bati ve Zlíně
Fakulta managementu a ekonomiky

Univerzita Tomáše Bati ve Zlíně
Fakulta managementu a ekonomiky
Ústav ekonomie

Akademický rok: 2021/2022

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Ondřej Jurásek**
Osobní číslo: **M19676**
Studijní program: **B0413A050024 Ekonomika a management**
Specializace: **Ekonomika a management podniku**
Forma studia: **Prezenční**
Téma práce: **Digitální data mining v managementu a ekonomii**

Zásady pro vypracování

Úvod

Definujte cíle práce a použité metody zpracování práce.

I. Teoretická část

- Zpracujte teoretické poznatky týkající se data miningu v managementu a ekonomii.

II. Praktická část

- Analyzujte současný stav využití data miningu v podniku.
- Vypracujte doporučení pro implementaci data miningu v podniku.
- Doporučení podrobně nákladové, časové a rizikové analýze.

Závěr

Rozsah bakalářské práce: cca 40 stran
Forma zpracování bakalářské práce: tištěná/elektronická

Seznam doporučené literatury:

AGGARWAL, Charu C. *Data mining: the textbook*. Cham: Springer, 2015, 734 s. ISBN 978-3319141411.
HAN, Jiawei, Micheline KAMBER a Jian PEI. *Data mining: concepts and techniques*. 3rd ed. Waltham: Elsevier, 2012, 703 s. ISBN 978-0123814791.
HOLUBOVÁ, Irena, Jiří KOSEK, Karel MINAŘÍK a David NOVÁK. *Big Data a NoSQL databáze*. Praha: Grada, 2015, 281 s. ISBN 978-80-2475-466-6.
LEVENTHAL, Barry. *Predictive analytics for marketers: using data mining for business advantage*. London: Kogan Page, 2018, 251 s. ISBN 978-0-7494-7993-0.

Vedoucí bakalářské práce: **Ing. Milan Damborský, Ph.D.**
Ústav ekonomie

Datum zadání bakalářské práce: **11. února 2022**
Termín odevzdání bakalářské práce: **20. května 2022**

L.S.

prof. Ing. David Tuček, Ph.D.
děkan

doc. Ing. Petr Novák, Ph.D.
garant studijního programu

Ve Zlíně dne 11. února 2022

PROHLÁŠENÍ AUTORA BAKALÁŘSKÉ/DIPLOMOVÉ PRÁCE

Prohlašuji, že

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen na elektronickém nosiči v příruční knihovně Fakulty managementu a ekonomiky Univerzity Tomáše Bati ve Zlíně;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen připouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

Prohlašuji,

1. že jsem na diplomové/bakalářské práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
2. že odevzdaná verze diplomové/bakalářské práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně 17. 5. 2022

Jméno a příjmení: Ondřej Jurášek

.....
podpis diplomanta

ABSTRAKT

Tato bakalářská práce se zabývá digitálním data miningem a jeho využitím v managementu a ekonomii. Cílem této bakalářské práce bude na základě analýzy současného stavu využití data miningu ve vybrané společnosti vytvořit návrh doporučení jeho efektivnějšího využití. Vybraná společnost nese název Martenz.

V teoretické části jsou zmíněny základní pojmy související s data miningem, metody, které data mining využívá, proces data miningu dle struktury CRISP a následné možnosti využití data miningu dle oboru. Po vypracování teoretické části a s ní obdrženými novými znalostmi jsem analyzoval současné využití data miningu ve společnosti. Společnost již určité platformy data miningu využívala, ale stále zde byl velký prostor pro zlepšení. Cíl práce byl splněn vypracováním návrhu o použití nového efektivnějšího systému CRM (řízení vztahu se zákazníky). Systémy CRM jsou jedny z nejpoužívanějších aplikací data miningu ve společnostech. Tento návrh byl poté ve společnosti i aplikován.

Společnost Martenz tak nyní může provádět efektivnější cílený marketing, má lepší přehled o svých zákaznících, a může tak zlepšovat své výsledky prodeje.

Klíčová slova: data mining, CRM, management, ekonomie, marketing, CRISP DM, měřicí kódy, business intelligence

ABSTRACT

This dissertation focuses on a digital data mining and its use in the management and economics. The aim of the dissertation is to create suggestions for more effective application of the data mining. This will be done by analysing the current state of the data mining use in the chosen company. The chosen company is called Martenz.

The theoretical part covers the basic concepts related to data mining, methods, the process of data mining according to the CRISP model, and the consequent possibilities of data mining use in the specific fields. The theoretical part is followed by the analysis of the current data mining use in the chosen business. The company had already been using certain platforms of data mining. Nevertheless, there was still room for improvement. The aim of the dissertation was accomplished by producing the proposal of the application of the new, more efficient CRM system (Customer Relationship Management). The CRM is one of the most utilized technologies of data mining amongst business. Subsequently, the suggested proposal was applied in the company.

Hence, the Martenz enterprise is now able to carry out more effective target marketing, have a better overview of their customers, and consequently be improving the sales results.

Keywords: data mining, CRM, management, economy, marketing, CRISP DM, measuring codes, business intelligence

Rád bych poděkoval panu Ing. Milanovi Damborskému Ph.D. za odborné vedení mé práce, cenné rady, inspirující podněty a připomínky.

Chtěl bych poděkovat společnosti Martenz za součinnost a poskytnutí interních informací, které byly důležité pro vypracování praktické části této bakalářské práce.

Dále bych chtěl poděkovat rodině a přátelům za jejich podporu.

Prohlašuji, že odevzdaná verze bakalářské/diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

OBSAH

ÚVOD.....	10
CÍLE A METODY ZPRACOVÁNÍ PRÁCE.....	11
I TEORETICKÁ ČÁST.....	12
1 VYMEZENÍ ZÁKLADNÍCH POJMŮ V DANÉ OBLASTI.....	13
2 TECHNIKY DATA MININGU.....	16
2.1 ASOCIACE.....	16
2.2 KLASIFIKACE.....	16
2.3 SHLUKOVÁNÍ.....	17
2.4 ROZHODOVACÍ STROMY.....	17
2.5 PREDIKCE.....	18
2.6 NEURONOVÉ SÍTĚ.....	18
3 POKROČILÉ TECHNIKY DATA MININGU.....	19
3.1 WEB MINING.....	19
3.2 TEXT MINING.....	19
3.3 MULTIMEDIA MINING.....	19
4 PRŮBĚH DATA MININGU.....	21
4.1 PŘEDCHOZÍ ZNALOSTI.....	22
4.1.1 Cíl.....	22
4.1.2 Předmětná oblast.....	22
4.2 PŘÍPRAVA DAT.....	23
4.2.1 Průzkum dat.....	23
4.2.2 Kvalita dat.....	23
4.2.3 Chybějící hodnoty.....	23
4.2.4 Datové typy.....	24
4.2.5 Mimořádné hodnoty.....	24
4.2.6 Výběr prvků.....	24
4.2.7 Vzorkování dat.....	25
4.3 MODELOVÁNÍ.....	25
4.3.2 Algoritmus nebo technika modelování.....	25
4.3.3 Souborové modelování.....	26
4.4 APLIKACE.....	26
4.4.1 Připravenost modelu.....	26
4.4.2 Technická integrace.....	27
4.4.3 Doba odezvy.....	27
4.4.4 Aktualizace modelu.....	27
4.4.5 Přizpůsobení.....	27
4.5 ZNALOST.....	28

5	VYUŽITÍ DATA MININGU DLE OBORU	29
5.1	MARKETING	29
5.1.1	Řízení vztahu se zákazníky – CRM	31
5.1.2	Analýza zákaznických hodnot.....	33
5.1.3	Určení nákupních vzorců zákazníků	33
5.1.4	Zvýšení míry odezvy v e-mailových kampaních	33
5.1.5	Udržování a získávání zákazníků.....	33
5.1.6	Analýza tržního koše.....	33
5.2	EKONOMIE	34
5.2.1	Finance a bankovníctví	34
II	PRAKTICKÁ ČÁST.....	36
6	CHARAKTERISTIKA SPOLEČNOSTI MARTENZ	37
6.1	ZÁKLADNÍ INFORMACE.....	37
6.2	CÍLE A VIZE SPOLEČNOSTI	39
7	ANALÝZA SOUČASNÉHO VYUŽITÍ DATA MININGU VE SPOLEČNOSTI.....	40
7.1	GOOGLE ANALYTICS	40
7.2	FACEBOOK PIXEL.....	46
7.3	WIX.....	47
7.4	SHOPTET.....	50
7.5	SHRNUTÍ.....	50
8	DOPORUČENÍ IMPLEMETACE DATA MININGU VE SPOLEČNOSTI PRO JEHO EFEKTIVNĚJŠÍ VYUŽITÍ	51
8.1	RAYNET CRM.....	51
8.2	IMPLEMENTACE RAYNET CRM VE SPOLEČNOSTI	52
9	NÁKLADOVÁ, ČASOVÁ A RIZIKOVÁ ANALÝZA DOPORUČENÍ	53
	ZÁVĚR	55
	SEZNAM POUŽITÉ LITERATURY.....	57
	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....	59
	SEZNAM OBRÁZKŮ	60
	SEZNAM TABULEK.....	61

ÚVOD

V souvislosti s aktuálně probíhajícím digitálním věkem vzniká velké množství dat, které si jednotlivci a společnosti ukládají do svých databází. S těmito daty lze nadále pracovat a získávat tak z nich hodnotné informace a znalosti. Metodologie zabývající se získáváním potencionálně užitečných informací a znalostí z těchto nasbíraných dat se nazývá data mining. Data mining je v současnosti využíván v každé velké úspěšné společnosti a nabízí značné množství výhod, a i přesto nebývá velice často diskutován.

Téma digitální data mining v managementu a ekonomii jsem si vybral proto, že se jedná o velice zajímavé a komplexní téma, které má v praxi široké využití. Velké společnosti často vyhledávají uchazeče právě na pracovní pozice týkající se data miningu a s ním spojenými dalšími činnostmi. Využití benefitů data miningu přináší velkou konkurenční výhodu a pro prosperující společnosti hraje využití této metodologie klíčovou roli.

Moje bakalářská práce je rozdělena na teoretickou a praktickou část. Teoretická část je rozdělena na více kapitol a budu zde čerpat z odborných knih a odborných elektronických publikací. V první kapitole vymezím základní pojmy týkající se data miningu. Druhá kapitola je zaměřena na základní techniky data miningu, které jsou v současnosti využívány. Ve třetí kapitole se zmíním o pokročilejších technikách data miningu, které jsou pro svou výkonnost v posledních letech velice oblíbené. Čtvrtá kapitola bude detailně popisovat průběh jednoho z nejpoužívanějších procesů data miningu s názvem CRISP. Pátá kapitola se pak bude zabývat možnostmi využití data miningu dle oboru.

V praktické části jsou obsaženy čtyři kapitoly. V první kapitole charakterizují společnost Martenz, její cíle a vize. Pro tuto společnost bude nadále provedena analýza současného využití data miningu.

Ve třetí části navrhu doporučení pro implementaci data miningu v podniku. Na základě analýzy v předešlé části a znalostí obdržených v teoretické části navrhu společnosti možnost efektivnějšího využití data miningu.

V poslední části podrobím doporučení nákladové, časové a rizikové analýze. Budu zde uvádět cenové náklady, rizika, které mohou společně s realizací doporučení vzniknout a jejich eliminaci. Bude zde i termín realizace a časová náročnost implementace doporučení.

CÍLE A METODY ZPRACOVÁNÍ PRÁCE

Cílem této bakalářské práce bude na základě analýzy současného stavu využití data miningu ve vybrané společnosti vytvořit návrh doporučení jeho efektivnějšího využití.

Při zpracování teoretické části aplikuji metodu literární rešerše, kde budu čerpat z odborných knih a vědeckých publikací. Na téma data mining existuje velmi omezené množství zdrojů, přičemž se většina vyskytuje v anglickém jazyce. Na tohle téma se také vyskytuje velmi malý počet knih a informace tak bude nutné z větší části čerpat z odborných článků, publikovaných na stránkách IEEE Xplore a Web of Science.

Praktická část bude zpracována formou konzultací s majitelem společnosti a přidělením přístupů do aktuálně využívaných platforem v roli diváka. Přičemž díky roli diváka získám na těchto platformách přístup k informacím potřebných k vyhotovení praktické části této bakalářské práce. Při doporučení efektivnějšího využití data miningu budu postupovat na základě obdržených znalostí z teoretické části bakalářské práce a výsledků analýzy současné situace ve společnosti. Práce byla rovněž konzultována s expertem technologického inovačního centra. Výsledné doporučení bude následně předloženo vybrané společnosti ke zvážení o jeho realizaci.

I. TEORETICKÁ ČÁST

1 VYMEZENÍ ZÁKLADNÍCH POJMŮ V DANÉ OBLASTI

V následující kapitole jsou vysvětleny jednotlivé termíny, které se v oblasti data miningu velice často využívají.

Big data

Pojem Big data, v překladu velká data, se objevil především s příchodem nových technologií, služeb a jejich kombinací. Jako příklad můžeme uvést senzorové sítě nebo vědecké přístroje zkoumající přírodní jevy, sociální sítě, mobilní technologie a související aplikace. Tyto typy technologií a aplikací, s nemalou pomocí svých uživatelů, generují každou sekundu obrovská množství dat, která potřebujeme efektivně uložit a účelně zpracovat. Myšlenku dobře vystihla například společnost IMB, která ji definuje následovně: „V závislosti na odvětví a organizaci zahrnují Big Data informace z interních a externích zdrojů, jako jsou transakce, sociální média, podniková data, senzory a mobilní zařízení. Firmy mohou tato data využívat, aby lépe přizpůsobily své výrobky a služby potřebám zákazníka, dále optimalizovaly provoz a infrastrukturu anebo našly zcela nové zdroje příjmů" (Holubová a další, 2015, s. 20).

Data mining

Data mining je proces přeměny dat na užitečné informace. Jedná se o proces objevování a interpretace vzorců v datech za účelem řešení obchodních problémů. Data mining je spojován s velkým množstvím dat, které obsahují velké množství záznamů a atributů. K identifikaci smysluplných vztahů a jejich využití je potřebný sofistikovaný software (Leventhal, 2018, s. 22-23).

Je to studium shromažďování, čištění, zpracování, analýzy a získávání užitečných informací a poznatků z dat (Aggarwal, 2015, s. 1).

Umělá inteligence

Umělá inteligence je inženýrský a studijní obor zaměřený na modelování inteligence. Je charakteristická svými schopnostmi simulovat lidské chování, jako je učení, uvažování a sebe opravování. Umělá inteligence se skládá z podkategorií, jako je strojové učení, neuronové sítě, evoluční algoritmy, robotika a zpracování jazyka (Durand a další, 2020).

Business intelligence

Business intelligence označuje procedurální a technickou infrastrukturu, která shromažďuje, ukládá a analyzuje data produkovaná činnostmi společnosti. (Frankenfield, 2021)

Jedná se o široký pojem, který zahrnuje data mining, analýzu procesů, srovnávání výkonu a popisnou analýzu. Potřeba business intelligence byla odvozena z konceptu, že manažeři s nepřesnými nebo neúplnými informacemi budou mít v průměru větší tendenci dělat horší rozhodnutí, než kdyby měli lepší informace. Business intelligence analyzuje všechna data generovaná podnikem a poskytuje společnosti která ji užívá výhody jako rychlejší a přesnější reporting, lepší kvalitu dat, lepší spokojenost zaměstnanců, nižší náklady, vyšší výnosy a schopnost činit lepší obchodní rozhodnutí (Frankenfield, 2021).

CRM

Customer relationship management, v překladu řízení vztahu se zákazníky je nedílnou součástí každého podnikání. Jde o kombinaci strategií, postupů a technologií, které společností umožňují spravovat interakce se zákazníky a jejich daty v průběhu jejich spotřebitelského životního cyklu. Ve výsledku nám tedy pomáhá udržet si stávající zákazníky a zároveň získávat nové, dochází tedy ke zvýšení prodeje (Singh, 2019).

Hlavním účelem CRM je navázat vztah se zákazníky a dalšími jednotlivci. Pro zajištění nejlepšího vztahu se zákazníky je nezbytné mít všechny správné informace a správně je seskupit. Se všemi informacemi, které systém shromažďuje, je data mining velkým pomocníkem. Data mining může pomoci analyzovat a zpracovat data, což podniku usnadní interakci se zákazníky a budoucími klienty. Kombinace CRM a data miningu nám tedy může pomoci vybrat správného potenciálního zákazníka, segmentovat publikum, nastavit optimální cenovou politiku a další (Singh, 2019).

Data Warehouse

Data warehouse, v překladu datový sklad, je sbírka integrovaných, předmětově orientovaných databází navržených tak, aby poskytovaly informace potřebné pro rozhodování (Zagan a další, 2019, s. 31).

Pro usnadnění rozhodování jsou data v datovém skladu uspořádána kolem hlavních subjektů, jako jsou zákazníci, položky, dodavatelé a činnosti. Data jsou uložena a sumarizována, aby nám poskytovala informace z historické perspektivy, například za posledních 6 až 12 měsíců (Han a další, 2012, s. 10).

Data set a data lake

Data set, v překladu datový soubor, je sbírka dat s definovanou strukturou (Kotu a další, 2015, s. 21).

Data lake, v překladu datové jezero, je moderní technologie pro ukládání dat v nezpracované podobě pro budoucí použití. Data jsou přijímána v nezpracované podobě, takže jej mohou využívat různí uživatelé, kteří poté mohou přistupovat k těmto datovým jezerům za účelem analýzy, struktury nebo dalšího jiného zpracování (Zagan a další, 2019, s. 32).

Datová jezera byla vyvinuta v reakci na omezující limity datových skladů. Zatímco datové sklady poskytují podnikům vysoce výkonnou a škálovatelnou analýzu, jsou drahé a nedokážou si poradit s moderními případy použití, která většina společností hledá. Datová jezera se používají ke konsolidaci všech dat organizace na jediném centrálním místě, kde je lze uložit v jakékoliv podobě, aniž by bylo předem nutné zavádět formální strukturu pro uspořádání dat, jako je to u datového skladu. Oproti datovým skladům jsou datová jezera schopna zpracovávat data jako jsou obrázky, videa, zvuk a dokumenty, které jsou v dnešní době strojového učení a pokročilé analýzy velice důležité (Databricks, 2022).

Atribut

Atribut, jinak řečeno vlastnost, vstup, rozměr, proměnná nebo prediktor je vlastnost datového souboru. Atributy mohou být číselné, kategorické, datové, časové, textové nebo booleovské datové typy (Kotu a další, 2015, s. 22).

2 TECHNIKY DATA MININGU

Techniky data miningu jsou procesem identifikace datových vzorů a trendů, s cílem získat užitečné informace z obrovských datových sad, abychom mohli činit lepší rozhodnutí. Existuje mnoho technik data miningu, kde každá technika má své pravidla a metody, které řeší určitý druh problému (Osman, 2019).

2.1 Asociace

Mezi velmi známé techniky data miningu patří asociace, která objevuje vzorce založené na vztahu mezi proměnnými ve stejné transakci. Je také známa jako relační technika, protože používá vztah mezi položkami a zjišťuje četnost výskytu různých položek, které se v rámci datové sady objevují s nejvyšší frekvencí. Asociační pravidla používají příkazy if – then, aby ukázala pravděpodobnost vztahů mezi daty nebo proměnnými v rámci velkých datových souborů v různých typech databází. Asociační pravidlo má řadu aplikací a je široce používáno pro pomoc při odhalování prodejních korelací v transakčních datech. Asociace je široce využívána vlastníky maloobchodů, protože pomáhá porozumět nákupnímu chování zákazníků. Na základě historických dat o prodeji mohou vlastníci maloobchodů zjistit synergetické typy produktů, jakou jsou třeba pivo a bramborové lupínky a poté je v obchodě postavit vedle sebe, aby zákazníkovi ušetřili čas a zvýšili si tržby. Vzhledem ke svému rozšířenému využití v maloobchodním prodeji se asociační pravidlo označuje jako analýza nákupního košíku (Osman, 2019).

2.2 Klasifikace

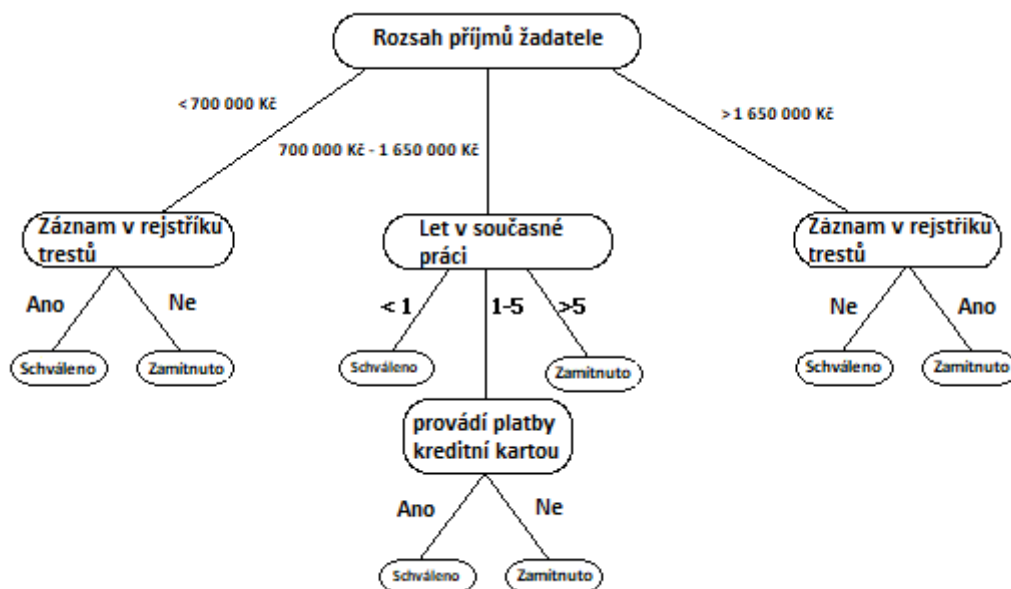
Klasifikační technika je používána ke klasifikaci nasbíraných dat do různých skupin nebo tříd za účelem získání přesné predikce a analýzy z obrovského souboru dat. Klasifikaci lze použít k vytvoření představy o kategorii zákazníka, objektu nebo předmětu v sadě dat popisem více jejich atributů za účelem identifikace konkrétní třídy. Můžeme například snadno kategorizovat budovy do různých typů na základě jejich obsazenosti, nebo je identifikovat podle atributů jako jsou výška, struktura nebo šířka. Tyto principy lze aplikovat na zákazníky a klasifikovat je podle jejich věku, pohlaví a sociálních skupin. Kromě toho můžeme použít klasifikaci jako zdroj výsledků jiných technik, jako jsou rozhodovací stromy nebo shlukování (Osman, 2019).

2.3 Shlukování

Při data miningu byla jedna z prvních používaných technik metoda shlukování. Proces techniky shlukování zahrnuje analýzu jednoho nebo více atributů potřebných k identifikaci dat, která jsou si navzájem podobná, aby bylo možné porozumět rozdílům a podobnostem které datová sada obsahuje. Procesu shlukování se někdy říká segmentace, protože rozděljuje data do kategorií, kde se nachází data se vzájemnou korelací. Techniku shlukování můžeme použít například při managementu knihovny. Tím, že dáme knihy, které pojednávají o stejných tématech blízko sebe a do jednoho regálu, je budou moci čtenáři rychleji a efektivněji najít (Osman, 2019).

2.4 Rozhodovací stromy

Rozhodovací stromy se používají k sestavení klasifikačních a regresních modelů. Používají se k vytváření datových modelů, které budou předpovídat hodnoty pro rozhodovací proces. Pomocí rozhodovacích stromů lze například snadno porozumět rozhodnutím, které zákazníci provádějí (Osman, 2019).



Obrázek 1 Rozhodovací strom – půjčka v bance (Zdroj: Plapinger, What is a decision Tree?)

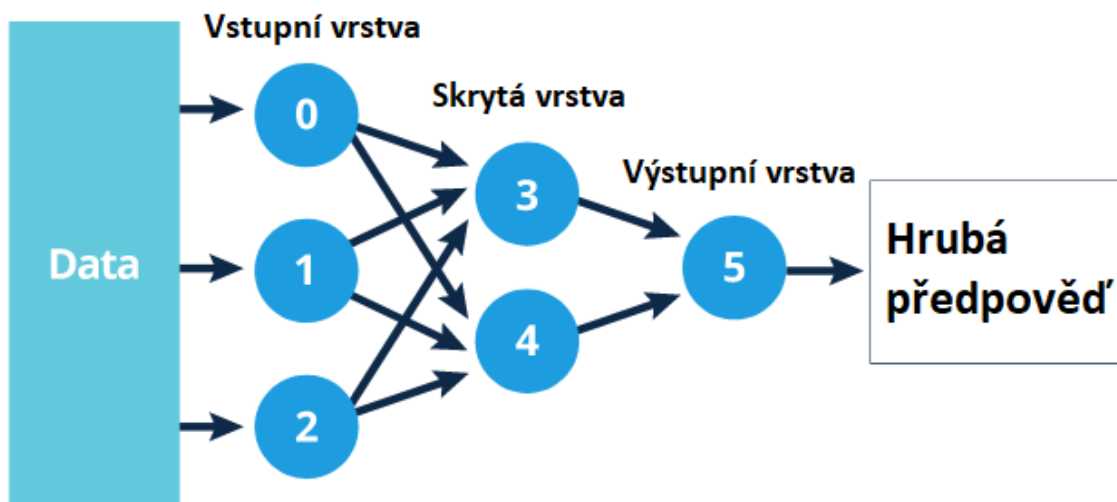
2.5 Predikce

Predikce je velice rozsáhlou technikou a sahá od předvídání podvodů k předpovídání zisků společnosti. Používá se v kombinaci s jinými technikami data miningu, predikce zahrnuje analýzu trendů, klasifikaci a porovnání modelů. Jedná se o predikci, která se provádí analýzou minulých událostí. Jako příklad lze uvést autorizaci kreditní karty, kde kombinací analýzy rozhodovacích stromů a minulých transakcí s klasifikačními historickými shodami zjišťujeme, jestli byla transakce podvodná nebo ne (Osman, 2019).

2.6 Neuronové sítě

Neuronové sítě jsou velice důležitou technikou, které jsou v dnešní době velice široce využívány. Tato technika je často používána v raných fázích data miningu. Neuronové sítě jsou velice snadno použitelné, protože jsou do určité míry automatizovány (Osman, 2019).

Jedná se o matematický model, na kterém funguje princip biologických neuronových sítí. Jinými slovy se jedná o simulaci lidského myšlení. Po spuštění programu začnou neuronové sítě hledat vztah mezi nově obdrženými daty a daty které v paměti neuronové sítě již jsou (Sinkov a další, 2016).



Obrázek 2 Neuronové sítě v data miningu (Zdroj: Ramos, Real-Life and Business Applications of Neural Networks)

3 POKROČILÉ TECHNIKY DATA MININGU

Data mining z textu, webu a multimédií jsou vzájemně propojené oblasti, které jsou v posledních letech pro svou vysokou výkonnost velice oblíbené. Data mining z textu je analýza rozsáhlých dokumentů a vytváření skrytých vzorů v textových datech. Data mining webu zahrnuje analýzu dat související s webem, včetně webového obsahu, struktur stránek a statistik webových odkazů. Multimediální data mining se používá pro extrakci zajímavých informací z multimediálních datových sad, jako je zvuk, video, obrázky, grafika, řeč a text. (Tekin a další, 2019, s. 48)

3.1 Web mining

Web mining, je dílčí oblastí data miningu. Webový data mining lze rozdělit do tří kategorií jako mining webového obsahu, mining webové struktury a mining webového využití. Data mining z webového obsahu se zabývá objevováním užitečných informací z obsahu webových stránek. Mining struktury webu je zaměřený na propojení mezi strukturou webu a webovými stránkami a mining využití webu odhaluje vstupní vzory webových uživatelů (Tekin a další, 2019, s. 48).

Cílem webového miningu je vytváření podrobných závěrů o uživateli, úprava obsahu dle tendencí uživatelů a provádění zlepšení pro lepší využitelnost webu. V posledních letech s rozmachem e-commerce a online nakupování nabízí web mining velkou konkurenční výhodu (Tekin a další, 2019, s. 48).

3.2 Text mining

Text mining je kombinací zpracování přirozeného jazyka a data miningu. Cílem je získat strukturovaná data prostřednictvím textu. K dosažení těchto cílů se používá lexikální analýza, distribuce frekvence slov, extrakce informací, a dokonce i vizualizace (Tekin a další, 2019, s. 48).

3.3 Multimedia mining

Rychlý vývoj v oblasti multimédií a technologií pro ukládání dat vedl k obrovskému růstu databází. Multimediální databázové systémy ukládají a spravují velké množství multimediálních dat, jako jsou obrázky, videa, zvuk a hypertextová data. Při práci v multimediálních dokumentech se pracuje s nestruturovanými informacemi. Nástroje

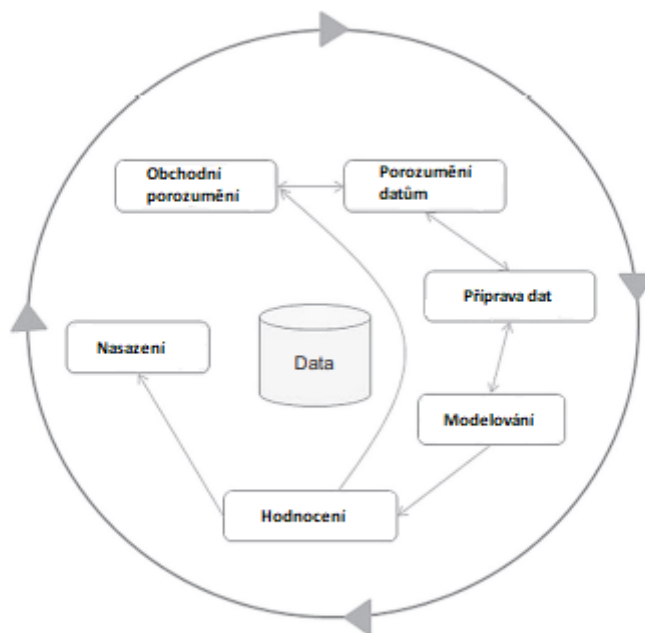
používané v této technice jsou například klasifikace obrázků na základě obsahu, kategorizace hudby a rozpoznávání objektů ve video streamech (Tekin a další, 2019, s. 49).

4 PRŮBĚH DATA MININGU

Metodologické objevování užitečných vztahů a vzorců v datech umožňuje soubor interaktivních aktivit známých jako proces data miningu. Standardní proces zahrnuje pochopení problému, přípravu vzorků dat, vývoj modelu, aplikaci modelu na sadu dat, abychom viděli, jak bude model fungovat v reálném světě a posledním krokem je produkční nasazení. Pro proces data miningu byly v průběhu let různými akademickými a komerčními institucemi navrženy různé procesní struktury (Kotu a další, 2014, s. 17).

Jedna z nejpoblárnějších procesních struktur pro data mining je CRISP-DM, což je zkratka pro Cross Industry Standard Process for Data Mining. Proces CRISP-DM je nejrozšířenější přijatou strukturou pro vývoj řešení data miningu. Další strukturou pro data mining je SEMMA, což je zkratka pro Sample, Explore, Modify, Model a Assess, vyvinutý institutem SAS (Kotu a další, 2014, s.17).

Tyto struktury průběhu data miningu vykazují společné charakteristiky, a proto bude v této kapitole použita struktura velmi podobná procesu CRISP (Kotu a další, 2014, s.17).



Obrázek 3 Data miningový rámec CRISP (zdroj: Kotu a další, Predictive Analytics and Data Mining)

4.1 Předchozí znalosti

Cíl data miningu se vždy vyvíjí na základě existujícího tématu a kontextových informací, které jsou známy. Krok předchozích znalostí v procesu data miningu pomáhá definovat jaký problém řešíme, jak se nám hodí do obchodního kontextu a jaká data potřebujeme k jeho vyřešení (Kotu a další, 2014, s. 19).

4.1.1 Cíl

Proces data miningu je potřeba začít analýzou, otázkou nebo obchodním cílem. Jedná se o nejdůležitější krok v procesu data miningu. Bez správně definovaného problému není možné získat správný soubor dat a vybrat potřebný algoritmus. I když je proces data miningu sekvenčním procesem a je běžné vracet se k předchozím krokům a revidovat předpoklady, přístup a taktiku, je nezbytně nutné definovat správný cíl celého procesu (Kotu a další, 2014, s. 20).

4.1.2 Předmětná oblast

Proces data miningu odhaluje skryté vzory v datové sadě pomocí odhalení vztahů mezi atributy. Hlavním problémem v tomto procesu jsou falešné signály, které jsou zapříčiněny odhalením velkého množství vzorců. Řešení tohoto problému je na odborníkovi pro těžbu dat. Jeho úkolem je tyto falešné signály eliminovat a použít jen ty, které jsou platné a relevantní při řešení stanoveného cíle. Proto je naprosto nezbytné znát předmět, kontext a obchodní proces, který generuje data (Kotu a další, 2014, s. 20).

4.1.3 Data

Podobně jako předchozí znalosti v oblasti předmětu existují také předchozí znalosti v datech. V typickém podniku se data obvykle shromažďují jako součást obchodních procesů. Pro proces data miningu je nezbytné pochopit, jak jsou data v podniku shromažďována, ukládána, transformována a použita. Existuje celá řada faktorů, které je třeba vzít v úvahu, jako třeba kvalita dat, množství dat, dostupnost dat, co dělat v případě, když data nejsou k dispozici a mnoho dalšího. Cílem tohoto kroku je přijít s datovým souborem, jehož těžba řeší definovaný problém. Kvalita celého modelu se poté odvíjí od kvality použitých dat (Kotu a další, 2014, s. 21).

4.2 Příprava dat

Příprava datové sady tak, aby vyhovovala stanovenému cíli data miningu, je časově nejnáročnější část procesu. Velmi zřídka jsou data dostupné ve formě, kterou vyžadují používané algoritmy data miningu. Většina algoritmů data miningu vyžaduje, aby data byla strukturována v tabulkovém formátu se záznamy v řádcích a atributy ve sloupcích (Kotu a další, 2014, s. 22).

4.2.1 Průzkum dat

Příprava dat začíná hloubkovým prozkoumáním dat a lepším porozuměním datové sady. Průzkum dat poskytuje sadu jednoduchých nástrojů pro dosažení základního porozumění. Základní přístupy průzkumu zahrnují výpočet popisné statistiky a vizualizaci dat. Základní průzkum může odhalit strukturu dat, rozložení hodnot, přítomnost extrémních hodnot a zdůrazní vzájemné vztahy v rámci souboru dat. Popisná statistika jako průměr, modus, medián, směrodatná odchylka a rozsah pro každý atribut poskytují snadno čitelný souhrn klíčových charakteristik distribuce dat (Kotu a další, 2014, s. 23).

4.2.2 Kvalita dat

Kvalita dat je všeobecným problémem všude tam, kde se data shromažďují, zpracovávají a ukládají. Organizace používají techniky čištění a transformace ke zlepšení kvality dat a poté je ukládají do celopodnikových úložišť nazývaných datové sklady. Data získaná z dobře udržovaných datových skladů mají vysokou kvalitu, protože jsou v nich zavedeny řádné kontroly, které zajišťují úroveň přesnosti jak nových, tak stávajících dat. Praktiky čištění dat zahrnují eliminaci duplicitních záznamů, odstranění nepodstatných záznamů, standardizaci hodnot atributů a nahrazení chybějících hodnot (Kotu a další, 2014, s. 24).

4.2.3 Chybějící hodnoty

Jedním z nejběžnějších problémů s kvalitou dat je, že některé záznamy postrádají hodnoty atributů. Existuje několik metod, jak tento problém zmírnit, avšak každá metoda má své klady a zápory. Prvním krokem při správě chybějících hodnot je pochopit důvod, proč hodnoty chybí. Sledováním datové linie zdroje dat může vést k identifikaci systémových problémů při sběru dat, chybám v transformaci, nebo může dojít k jevu, kterému uživatel doposud nerozumí. Znalost zdroje chybějící hodnoty často určí, jakou metodologii zmírňování problému použít. Chybějící hodnotu můžeme nahradit umělými daty, abychom problém zvládli vyřešit s minimálním dopadem na pozdější kroky v data miningu. Umělá

data bývají odvozena ze souboru dat a jedná se o střední, maximální nebo minimální hodnoty atributů. Tato metoda je užitečná, pokud se chybějící hodnoty vyskytují zcela náhodně a frekvence výskytu je poměrně vzácná. Alternativně můžeme použít metodu, kde budeme ignorovat všechny datové záznamy s chybějící hodnotou nebo nízkou kvalitou dat (Kotu a další, 2014, s. 24).

4.2.4 Datové typy

Atributy v sadě dat mohou být různých typů, například spojité numerické, celočíselné nebo kategorické. Každý algoritmus data miningu si vymezuje, jaké typy dat bude na vstupu akceptovat. Pokud se jedná například o jednoduchý lineární regresní model, vstupní atributy musí být číselné. Pokud jsou tedy dostupná data kategorická, je nutné je převést na atributy číselné. To lze provést zakódováním specifického číselného skóre pro každou hodnotu kategorie. Podobně lze číselné hodnoty převést na kategorické datové typy technikou zvanou binning, kde je pro každou kategorii určen rozsah hodnot (Kotu a další, 2014, s. 25).

4.2.5 Mimořádné hodnoty

Mimořádné hodnoty jsou anomálií v souboru dat. Mohou se vyskytovat oprávněně, například příjem v miliardách, nebo chybně, například výška člověka 1.73 centimetru. Bez ohledu na to je třeba přítomnost mimořádných hodnot pochopit a věnovat jim zvýšenou pozornost. Účelem vytvoření reprezentativního modelu je zobecnit vzor nebo vztah a přítomnost mimořádných hodnot model zkresluje. Detekce mimořádných hodnot může být primárním cílem některých aplikací pro data mining, jako je detekce podvodů (Kotu a další, 2014, s. 25).

4.2.6 Výběr prvků

Pokud má datový set stovky až tisíce atributů, nastává v praxi při data miningu problém. Ne všechny atributy jsou stejně důležité nebo užitečné při předpovídání požadované cílové hodnoty. Přítomnost vysokého počtu atributů v datové sadě výrazně zvyšuje složitost modelu a může snížit i jeho produktivitu. Snížení počtu atributů, aniž by došlo k výrazné ztrátě produktivity modelu, se nazývá výběr prvků. Pokud dojde ke snížení počtu atributů v sadě dat, bude model efektivnější (Kotu a další, 2014, s. 26).

4.2.7 Vzorkování dat

Vzorkování je proces výběru podmnožiny jako reprezentace původní datové sady pro použití při analýze dat nebo modelování. Vzorková data slouží jako zástupce původního souboru dat s podobnými vlastnostmi, jako je například průměr. Vzorkování snižuje množství dat, které je třeba zpracovat pro analýzu a modelování. K získání reprezentativních prediktivních modelů stačí ve většině případů pracovat se vzorky. Vzorkování urychluje proces modelování (Kotu a další, 2014, s. 26).

4.3 Modelování

Model je abstraktní reprezentace dat a jejich vztahů v daném souboru dat (Kotu a další, 2014, s. 27).

V dnešní době se pro data mining používají stovky algoritmů, odvozených ze statistik, strojového učení, rozpoznávání vzorů a počítačových věd (Kotu a další, 2014, s. 27).

4.3.1 Zkušební a testovací datové sady

Pro vývoj stabilního modelu je třeba využít předem připravenou datovou sadu, kde známe všechny atributy. Tato datová sada se nazývá zkušební datový soubor a používá se k vytvoření modelu. Platnost vytvořeného modelu je třeba ověřit pomocí jiné datové sady, která se nazývá testovací datová sada. Pro usnadnění tohoto procesu bude celková datová sada rozdělena na sadu zkušební a testovací. Standardním pravidlem je, že dvě třetiny dat jsou přiděleny do sady zkušební a zbývající třetina připadá sadě testovací (Kotu a další, 2014, s. 28).

4.3.2 Algoritmus nebo technika modelování

Cíl a dostupnost dat určují, jakou kategorii data miningu je třeba použít. Kategorií může být například asociace, klasifikace a regrese. Volba algoritmu data miningu je na kvalifikovaném pracovníkovi, který ho zvolí dle vybrané kategorie. V rámci klasifikace lze například zvolit algoritmy jako rozhodovací stromy, indukce pravidel nebo neuronové sítě. K řešení stanoveného cíle data miningu je běžné používat více kategorií a algoritmů najednou (Kotu a další, 2014, s.28).

4.3.3 Souborové modelování

Souborové modelování je proces, při kterém se vytváří několik různých modelů za účelem predikce výsledků, a to buď pomocí mnoha modelovacích algoritmů nebo pomocí různých zkušebních sad. Důvodem pro použití souborového modelování je snížit chybovost predikce (Kotu a další, 2014, s. 31).

Na konci modelovací fáze byl analyzován cíl, byly získány relevantní data k dosažení cíle, byla vybrána správná technika data miningu, byl vybrán správný algoritmus, data byla rozdělena na zkušební a testovací sady, byl vytvořen zobecněný model ze zkušební sady, a nakonec byl zkušební model porovnán s modelem testovacím. Nyní je třeba model nasadit (Kotu a další, 2014, s. 32).

4.4 Aplikace

Aplikace je fáze, ve které se model stává připraveným být spuštěn v provozu. Ve fázi nasazení modelu jsou řešeny klíčové aspekty jako posouzení připravenosti modelu, technická integrace, doba odezvy, údržba modelu a asimilace (Kotu a další, 2014, s. 32).

4.4.1 Připravenost modelu

Část připravenosti modelu určuje primární vlastnosti, které jsou požadovány pro cíl nasazení modelu. Lze uvést dva odlišné případy použití: určení, zda spotřebitel splňuje podmínky pro úvěrový účet u přední komerční instituce a určení skupin zákazníků pro podnik (Kotu a další, 2014, s. 32).

Proces schvalování spotřebitelského úvěru probíhá v reálném čase. Rozhodovací model potřebuje shromáždit data od zákazníka, integrovat data třetích stran jako je úvěrová historie a rozhodnout se o schválení půjčky a podmínkách během několika sekund. Primární vlastností nasazení tohoto modelu je predikce v reálném čase (Kotu a další, 2014, s. 32).

Segmentace zákazníků na základě jejich vztahu se společností je komplikovaný proces, kde se berou v úvahu signály z různých interakcí mezi odděleními ve společnosti. Pro tuto aplikaci je použito dávkové zpracování, kdy se data shromažďují přes noc z různých oddělení a zdrojů, a celkové záznamy zákazníků jsou poté segmentovány. Primární vlastností této aplikace je najít jedinečné vzory mezi zákazníky (Kotu a další, 2014, s. 33).

4.4.2 Technická integrace

K vytvoření modelů data miningu bude použit určitý druh softwarového nástroje. Nástroje data miningu šetří čas tím, že nevyžadují psaní vlastních kódů pro implementaci algoritmu. To umožňuje analytikovi zaměřit se na data, obchodní logiku a průzkum datových vzorců. Modely jako jednoduchá regrese, rozhodovací stromy a indukční pravidla lze snadno začlenit přímo do obchodních aplikací a systémů business intelligence. Vzhledem k tomu, že jsou tyto modely reprezentovány jednoduchými rovnicemi a pravidly if – then, lze je snadno přenést do většiny programovacích jazyků (Kotu a další, 2014, s. 33).

4.4.3 Doba odezvy

Některé algoritmy pro data mining lze snadno sestavit, ale jejich předpovídání cílových proměnných může být poměrně pomalé. Je třeba zvážit kompromis mezi dobou odezvy a dobou sestavení algoritmů a v případě, že doba odezvy není pro obchodní aplikaci přijatelná, je třeba přehodnotit fázi modelování. Kvalita predikce, dostupnost vstupních dat a doba odezvy predikce zůstávají nejdůležitějšími kvalitativními faktory při aplikaci data miningu v podniku (Kotu a další, 2014, s. 33).

4.4.4 Aktualizace modelu

Klíčovým kritériem pro trvalou relevanci modelu je reprezentativnost souboru dat, který zpracovává. Podmínky, pro které je model vytvořen, se mohou časem měnit. Validitu modelu lze rutinně testovat pomocí nového testovacího data setu a výpočtem chybovosti. Vytvoření plánu pro aktualizaci modelu je klíčovou součástí fáze aplikace data miningu, pouze s plánem aktualizace bude model pracovat v souladu se současnými podmínkami (Kotu a další, 2014, s. 34).

4.4.5 Přizpůsobení

V deskriptivních aplikacích data miningu nemusí být aplikace modelu hlavním cílem. Úkolem často bývá převzetí znalostí získaných díky data miningu do organizace nebo konkrétní aplikace. Cílem může být například nalezení logických shluků v zákaznické databázi, aby bylo možné každé zákaznické skupině poskytnout individuální řešení. Dalším krokem může být provedení klasifikace nových zákazníků, aby je bylo možné přiřadit již k existujícím skupinám. Asociační analýza zase poskytuje řešení problémů tržního koše, kde je úkolem zjistit, které dva produkty jsou nejčastěji nakupovány společně. Úkolem pro

odborníka data miningu je tato zjištění formulovat, zjistit jejich význam pro stanovený cíl data miningu a provést kvantifikaci rizik v modelu (Kotu a další, 2014, s. 34).

4.5 Znalost

Proces data miningu poskytuje strukturu pro extrahování netriviálních informací z dat. S příchodem masivních úložišť, zvýšeným sběrem dat a pokročilými výpočetními modely se množství dat, které máme dispozici jen zvětšuje. K získání znalostí z těchto masivních datových zásob je třeba kromě standardních časových řad nebo jednoduchého statického zpracování použít pokročilé přístupy, jako jsou data miningové algoritmy (Kotu a další, 2014, s. 34).

Proces data miningu začíná předchozími znalostmi a končí obohacením o znalosti, které byly během procesu získány. Stejně jako u každé kvantitativní analýzy může proces data miningu poukázat na irelevantní vzorce ze souboru dat. Tyto nerelevantní vzorce by měly být identifikovány odborníkem na data mining a následně odstraněny (Kotu a další, 2014, s. 35).

5 VYUŽITÍ DATA MININGU DLE OBORU

Data mining se stal strategickou analytickou metodou, která je používána v rozhodovacím procesu a dosahování organizačních cílů. Aplikace pro data mining využívají hledání skrytých vztahů mezi finančními ukazateli, určují nákupní vzorce zákazníků v marketingu a řeší také jejich rizikovost. Data mining je dnes využíván především společnostmi ve finančním, komunikačním a marketingovém sektoru. Data mining umožňuje firmám určovat interní faktory, jako je cena, plánování výroby a personální dovednosti. Ekonomické ukazatele navíc umožňují určit vnější faktory, jako je konkurence a struktura trhu. Lze tak určit pozitivní a negativní vlivy na tržby společností a spokojenost jejich zákazníků. Data mining může být aplikován na mnoho dalších oblastí. Dalšími oblastmi využití jsou medicína, bankovníctví, pojišťovnictví a zdravotní služby, astronomie, biologie a telekomunikace (Tekin a další, 2019, s. 49).

5.1 Marketing

Význam data miningu a míra jeho využití v marketingovém rozhodování výrazně vzrostla. Data mining, který pomáhá podnikům demonstrovat vývoj vztahů se zákazníky, je procesem zkoumání mnoha úrovní vztahů se zákazníky, tržních trendů a modelů chování. Informace uchovávané v marketingové databázi jsou zásadní pro strategické rozhodování společnosti. Z tohoto důvodu musí být data uspořádána stejně tak jako marketingové funkce. Podniky v dnešním neustále měnícím se a složitém tržním prostředí potřebují k udržení své existence konzistentní informace. Data mining je důležitým nástrojem, který tyto konzistentní informace podnikům poskytuje. V marketingovém pojetí data miningu se nejčastěji používají modely, které se zabývají databázovým marketingem a řízením vztahů se zákazníky (Tekin a další, 2019, s. 50).

Data mining v současnosti využívají společnosti se silnými maloobchodními, finančními, komunikačními a marketingovými organizacemi zaměřenými na spotřebitele. Data mining umožňuje podnikům porozumět skrytým vzorcům v historických transakčních datech a pomáhá tak rychle a levně plánovat a implementovat nové marketingové kampaně. Podniky používají metody data miningu k vývoji produktů a propagačních aktivit pro konkrétní segmenty zákazníků, preference a umístění produktu, dopad na prodej, spokojenost zákazníků, analýza dat v místě prodeje, optimalizace nabídky a umístění prodejen. Segmentace trhu, analýza konkurenceschopnosti, hodnocení zákazníků a analýza křížového prodeje se provádí pomocí data miningu (Tekin a další, 2019, s. 50).

Data mining lze úspěšně aplikovat v mnoha různých oblastech, jako jsou individualizované kampaně, prodejní politika, nové produkty, křížový a doplňkový prodej. Jejich podrobnější popis bude níže (Tekin a další, 2019, s. 50-51):

a) Organizace kampaně

Výběr stávajících zákaznických segmentů v různých kampaních a vývoj přístupů k charakteristikám chování těchto zákazníků. Cílem je zvýšit míru odezvy na marketingové nebo maloobchodní kampaně a snížit náklady na řízení vztahů se zákazníky.

b) Organizace speciálních kampaní

Použití speciálních kampaní na skupinu zákazníků, kde je potenciál nejvyššího zisku.

c) Organizace prodejních zásad pro zákazníky

Nalezení modelových skupin zákazníků, kteří sdílejí stejné vlastnosti (úroveň příjmů, zájmy, výdajové návyky) a poté určování prodejních podmínek a cen na základě jejich nákupních profilů.

d) Vývoj nového produktu

Určením funkcí, které různé skupiny zákazníků potřebují a odstraněním funkcí, které již nepotřebují.

e) Vývoj trhu a uspořádání regálů

Nejčastějším příkladem použití asociačních pravidel je analýza tržního koše. Tento proces analyzuje nákupní návyky zákazníků hledáním asociací mezi produkty při nákupu, které provádějí. Tento typ asociace odhaluje, jaké produkty zákazníci nakupují společně a manažeři tak mohou vyvinout efektivnější prodejní strategie nebo efektivněji uspořádat rozmístění regálů v prodejně.

f) Křížový prodej

Hledáním souvislostí mezi prodejem produktů a modely, které dokáží porozumět skupinám zákazníků dle výdajů na jejich kreditní kartě, lze vytvořit zákaznický profil. Dle zákaznického profilu lze poté pomocí vypracovaných odhadů nabízet další produkty.

g) Dodatečné prodeje

Hledání nejlepších zákazníků nebo zákaznických skupin a vývoj personalizovaných produktů a služeb na základě identifikace potřeb těchto zákazníků.

5.1.1 Řízení vztahu se zákazníky – CRM

Rozdíly mezi produkty se zmenšily, ziskové marže se snížily a životní styl zákazníků a s ním i jejich nákupní návyky se v globalizujícím se světě změnily. Tohle prostředí vyžaduje od podniků změnu přístupu k zákazníkům. Jedním z nástrojů používaných podniky v této souvislosti je řízení vztahů se zákazníky. Cílem CRM je najít zákazníky, efektivněji je oslovit, a poskytovat jim vhodné zboží a služby na základě jejich požadavků. CRM se skládá ze souborů procesů a systémů, které podporují obchodní strategii k budování dlouhodobých a ziskových vztahů s konkrétními zákazníky. Cílem řízení vztahu se zákazníky je optimalizovat hodnotu organizace pro zákazníky prostřednictvím analýzy dat a komunikace. Důležitými nástroji podniků, které CRM používají, je datový sklad a data mining. V počátečních fázích CRM byla přikládána důležitost aplikacím, jako jsou data související se zákazníky, call centrum a prodejní kampaně. V pozdějších aplikacích CRM je analyzována dlouhodobá hodnota zákazníka a za základ jsou brány aplikace, které zajišťují, aby zákazník u dané společnosti opětovně nakupoval (Tekin a další, 2019, s. 51).

Segmentace zákazníků je proces rozdělování zákazníků do homogenních podskupin založených na různých kvalitách a vlastnostech. Organizace poté lépe porozumí zákazníkům a mohou vytvářet lepší strategie. Níže budou uvedeny kategorie segmentace zákazníků získané algoritmem asociačních pravidel (Tekin a další, 2019, s. 52-53):

a) Segmentace připravenosti kupujícího

Rozdělení zákazníků do skupin odrážející různé fáze, kterými kupující procházejí během nákupního procesu. Tyto fáze zahrnují například neznalost, povědomí, znalosti, preference a přesvědčení.

b) Segmentace dle výhod

Rozdělení trhu do skupin dle různých výhod, které spotřebitelé od produktů požadují.

c) Segmentace dle chování

Rozdělení zákazníků do skupin podle přístupu nebo reakce na produkt nebo propagaci

d) Segmentace dle příležitosti

Rozdělení zákazníků do skupin, které spotřebovávají produkt nebo službu v určitých časech, situacích nebo v reakci na určité události.

e) Segmentace dle životního stylu

Rozdělení zákazníků do skupin podle životního stylu, sociálního chování, hodnot a osobních charakteristik.

f) Demografická segmentace

Rozdělení zákazníků do různých skupin na základě demografických proměnných, jako je věk, pohlaví, velikost rodiny, příjem, povolání, vzdělání, jazyk, náboženství, rasa a národnost.

g) Geografická segmentace

Rozdělení zákazníků do různých skupin na základě země, regionu, podnebí a hustoty obyvatelstva.

h) Segmentace dle loajality

Rozdělení zákazníků do různých skupin na základě stupňů loajality vůči dodavateli nebo značce.

i) Segmentace produktů

Rozdělení zákazníků do různých skupin na základě typu použití produktu nebo služby.

j) Segmentace ziskovosti

Rozdělení zákazníků do různých skupin na základě ziskovosti.

k) Interakční segmentace

Rozdělení zákazníků do různých skupin na základě jejich preferencí ohledně platby, propagací a komunikace.

l) Segmentace dle spokojenosti

Rozdělení zákazníků do různých skupin na základě zaznamenané úrovně spokojenosti, historie stížností, historie chyb a historie zlepšení. Potencionální přínos firmy pro každého klienta se nazývá celoživotní hodnota zákazníka. Celoživotní hodnota zákazníka je měřítkem potencionálu generování zisku při řízení vztahu se zákazníky. Přesné stanovení celoživotní hodnoty zákazníka určí oblasti, na které se bude firma v budoucnosti soustředit, a poskytne nesčetné výhody, zejména při nabízení personifikovaných výhod zákazníkům. Zákaznický profil definuje vlastnosti zákazníků, jako je věk, příjem a životní styl. Profilování se dosahuje shromažďováním demografických a behaviorálních informací.

5.1.2 Analýza zákaznických hodnot

Hodnota zákazníka je v dnešní době jednou z nejzákladnějších marketingových strategií. Vztahuje se k emočnímu procesu, který se vyskytuje v mysli jako výsledek srovnání peněžní hodnoty se zakoupeným produktem nebo službou. Analýza hodnot umožňuje organizacím nasměrovat vývoj produktů do oblastí s největší hodnotou pro zákazníky (Tekin a další, 2019, s. 53).

5.1.3 Určení nákupních vzorců zákazníků

Podniky ukládají mnoho dat o zákaznících, vlastnostech produktů a služeb a interakcích. Data mining může poskytnout vysoce rozvinutou analýzu zákaznického výzkumu hledáním skrytých vzorců v databázích. Pokud bude společnost vytvářet a zaznamenávat charakteristiky zákazníků, produktů a služeb, bude jim schopna pomocí aplikace data miningu porozumět a vytvářet profily zákazníků. Tyto zákaznické profily předpovídají nákupní chování a tendence zákazníků (Tekin a další, 2019, s. 54).

5.1.4 Zvýšení míry odezvy v e-mailových kampaních

Mail marketing je oblíbený nástroj marketingové komunikace, který představuje jednu z nejvíce nákladných investicí do přímého marketingu. Získat však vysokou míru odezvy a také mít vynikající nákladovou efektivitu e-mailových kampaní je pro obchodníky velice obtížné. V e-mailových kampaních se segmentace provádí za účelem cílení seznamu e-mailů na zájmy, nákupní chování a demografické údaje. V důsledku rozvoje databázových technik se stala segmentace v přímém marketingu produktivnější. Aplikace data miningu se ukazuje jako nejlepší způsob, jak rozvíjet stávající marketingové strategie segmentováním stávajících zákazníků nebo jejich nasměrováním na trh. Současně se v oblasti databázového marketingu používají i modely neuronových sítí (Tekin a další, 2019, s. 54).

5.1.5 Udržování a získávání zákazníků

Životní cyklus zákazníka označuje fáze vztahu mezi zákazníkem a podnikem. Pomocí analýzy data miningu a business intelligence lze získávat zpětnou vazbu, která pomáhá předcházet ztrátám zákazníků (Tekin a další, 2019, s. 55).

5.1.6 Analýza tržního koše

S rozvojem použití čárových kódů se při realizaci prodejního procesu všechny informace týkající se produktů přenáší na elektronické médium. Tyto informace se nazývají data

tržního koše. V záznamu tržního koše jsou obsaženy informace, jako je číslo transakce, datum a kód produktu, množství a cena zakoupených produktů. Účelem analýzy tržního koše je najít vztahy mezi produkty a vytvářet nové obchodní strategie (Tekin a další, 2019, s. 56).

5.2 Ekonomie

Počítače jsou nyní zapojeny do mnoha ekonomických transakcí. Data, které lze pomocí transakcí zachytit mohou být následně analyzována (Varian, 2014, s. 3).

Zatímco data byla tradičně shromažďována pouze pro konkrétní účel, často národní statistickou agenturou, svět se stále více kvantifikuje, kde nyní i ta nejmenší společnost shromažďuje a zaznamenává data. Tato datová lavina dramaticky zvýšila jak rozmanitost dat, tak i rychlost jakou jsou data zaznamenávána. Vznikají nové příležitosti pro vytváření nových datových souborů z dříve nestrukturovaných souborů, jako jsou text a satelitní snímky. Tento vývoj otevřel nové oblasti ekonomických dotazů. Otázky, na které bylo dříve možné odpovědět až po několika měsících nebo dokonce letech, lze nyní řešit v reálném čase. Ekonomové tak přešli od prognózování k nowcastingu (Harding a další, 2018, s. 2).

5.2.1 Finance a bankovníctví

V online transakcích dochází ke generaci velkého množství dat, proto je ve finančním sektoru velice důležitá schopnost určit správné informace ve správný čas. V současné době mnoho bank a finančních institucí nabízí širokou škálu bankovních služeb, jako jsou investice, úvěry kreditní karty a mnoho dalšího. Data shromážděná těmito organizacemi jsou obecně velmi spolehlivá, úplná a vysoce kvalitní (Tekin a další, 2019, s. 57).

Tyto organizace používají metody data miningu pro detekci podvodů, předpovídání plateb úvěrů, analýzu získávání a udržení zákazníků, křížový prodej, analýzu úvěrové politiky zákazníků a segmentaci zákazníků pro cílový trh. Jelikož je po dodání produktů v pojišťovacím a bankovním sektoru vysoký potenciál ztráty, je nutné tyto rizika řádně analyzovat. Banky se tedy snaží předpovídat finanční riziko při poskytování úvěrů zákazníkům a předpovídají rizikové modely a možnost, že dlužníci nebudou schopni splácet své půjčky. Modely jsou nastaveny způsobem spotřebitelského chování minulých období (Tekin a další, 2019, s. 57).

Riziko nežádaného chování je pro banky rovněž také důležitým problémem. Při ztrátě kreditní karty přebírají banky určitou část vzniklé ztráty. Pro tyto případy byly navrženy systémy detekce podvodů, které dokážou tyto ztráty snížit. Jedná se o jednu z metod používaných k předdefinování typických vzorců výdajů zákazníků, k detekci náhlých změn trendů ve výdajích a případnému zastavení schvalování nákupů (Tekin a další, 2019, s. 58).

Úspěšné výsledky data miningu jsou dosahovány při burzovních transakcích, jako je odhad cen akciového trhu, obecná analýza trhu a optimalizace obchodních strategií, zjišťování příčin ztrát zákazníků v pojišťovacích činnostech, předcházení nesrovnalostem, snižování hlavních nákladů a určování cen pojistek. Využití data miningu je výhodné v následujících finančních aktivitách (Tekin a další, 2019, s. 58):

- a) Shromažďování a analýza údajů o chování zákazníků a přijímání strategických rozhodnutí zvyšuje jejich loajalitu
- b) Pomáhá najít skryté vztahy mezi různými finančními ukazateli, aby bylo možné odhalit podezřelou aktivitu a identifikovat aktivity s vysokým potenciálem rizika
- c) Podporuje rozhodování v reálném čase
- d) Pomáhá definovat podvodné akce shromažďováním minulých dat a jejich přeměnou na platné a užitečné informace
- e) Data mining pomáhá předpovídat životní hodnotu každého zákazníka v bance a vhodně obsluhovat každý segment nabídkou speciálních příležitostí a slev.

II. PRAKTICKÁ ČÁST

6 CHARAKTERISTIKA SPOLEČNOSTI MARTENZ

6.1 Základní informace

Společnost Martenz byla založena v roce 2016 a zahájila prodej v roce 2017. Jedná se o obchodní sdružení dvou fyzických osob se sídlem ve Zlíně. Společnost nemá žádné zaměstnance. Hlavním produktem této výrobní společnosti je pravá kvasná pálenka z malin.



Obrázek 4 Logo společnosti Martenz
(Zdroj: Martenz.cz)

Společnost Martenz prodává své výrobky prostřednictvím svého e-shopu, ve vybraných restauracích v České republice, na stránkách Destilerka.cz a prostřednictvím aukcí na Alkohol.cz. V klasických obchodních řetězcích její výrobky nelze zakoupit.



Obrázek 5 Malinovice Gold VIP –
hlavní produkt společnosti Martenz
(Zdroj: Martenz.cz)

Ekonomické pozadí společnosti

Tabulka 1 Vývoj čistého obratu společnosti Martenz v letech 2017-2021 (zdroj: vlastní tvorba)

Vývoj čistého obratu společnosti Martenz v letech 2017–2021 (v Kč)					
Účetní období	2017	2018	2019	2020	2021
Čistý obrat	242 937	1 658 249	3 110 237	4 553 183	6 816 951



Obrázek 6 Vývoj čistého obratu společnosti Martenz v letech 2017-2021 – grafické znázornění (zdroj: vlastní tvorba)

Historie společnosti

V roce 2015 se podařilo majiteli společnosti Zdeňkovi Juráškovvi vyrobit první pálenku z malin. Výroba pálenky z malin patří k těm nejsložitějším, a i přes tuto náročnost se povedla a chutnala dokonce i těm, co pálenky moc nepijí. Tato skutečnost rozhodla o navýšení produkce a založení společnosti.

V roce 2016 vznikla společnost, jejímž hlavním produktem je řemeslně vyráběná pálenka z malin – malinovice. Jednalo se o výjimečný produkt, který v České republice dosud téměř nikdo nevyrobil. Produkt měl mezi zákazníky velký úspěch, a tak se začala společnost dále rozvíjet.

V roce 2018 rozšířila společnost svůj sortiment o další pálenky, slivovici a třešňovici.

V roce 2020 se společnost zúčastnila mezinárodní soutěže IWSC vína a destilátů v Londýně, kde se jí s malinovicí podařilo získat stříbrnou medaili v kategorii ovocných destilátů, což se dosud žádné společnosti v České republice nepodařilo.

V roce 2021 společnost oficiálně otevřela showroom ve Zlíně. Ve stejném roce proběhla premiéra pálenky malinovice black, kde společnost Martenz představila první pálenku z černých malin na světě. Na konci roku 2021 byla vydána první archivní malinovice.

V současnosti je společnost největším výrobcem kvasných pálenek z malin v České republice. Aktuálně připravuje pálenku, která bude vyrobena ze stejných malin jako tomu bylo v roce 2017 jako limitovanou kolekci k 5. letému výročí.

MartenzClub

Věrnostní program společnosti Martenz nese název MartenzClub. Členem MartenzClubu se může stát každý, kdo zakoupí alespoň jednu láhev malinovice Martenz. Členové poté mohou čerpat výhody jako přístup k archivním lahvím a pálenkám z mikrokolekcí.

6.2 Cíle a vize společnosti

Vizí společnosti je povýšení vnímání moravských pálenek, jako vysoce kvalitního alkoholu v České republice.

Cílem společnosti je nadále vyrábět pálenky v maximální kvalitě a v omezeném množství.

7 ANALÝZA SOUČASNÉHO VYUŽITÍ DATA MININGU VE SPOLEČNOSTI

Společnost Martenz v současnosti využívá základní nástroje data miningu, jejich popis bude uveden níže.

7.1 Google Analytics

Google Analytics je analytický nástroj od společnosti Google, který slouží ke sledování dění na webových stránkách. Poskytuje majitelům stránek velké množství užitečných informací, dle kterých se lze následně rozhodovat. Tento nástroj byl založen v roce 2005 a stal se ihned velmi populárním. Nástroj je pro každého a zdarma.

V dnešní době používá nástroj Google Analytics 56 % všech webových stránek na celém internetu. To je mezi analytickými nástroji 85 % stránek, která tato data sledují. (Usage statistics of traffic analysis tools for websites, 2022)

Samotné zprovoznění nástroje lze realizovat více způsoby, avšak nejjednodušším způsobem je vložení specifického kódu od společnosti Google v administrativní sekci nastavení stránek. Společnost Martenz využívá e-shop platformy Shoptet, která nabízí velice jednoduché propojení. Veškeré konkrétní výsledky data miningu analytického nástroje Google Analytics se tedy budou vztahovat na e-shop společnosti Martenz.

Služba nabízí velké množství funkcí, které budou popsány níže:

Domovská stránka

Na domovské stránce Google Analytics je zobrazen souhrn nasbíraných informací ve zvoleném časovém období. Slouží především pro rychlý přehled nad děním na webových stránkách. Pro vytváření závěrů a rozhodnutí je potřeba si dané funkce rozkliknout do plného zobrazení v nabídce.

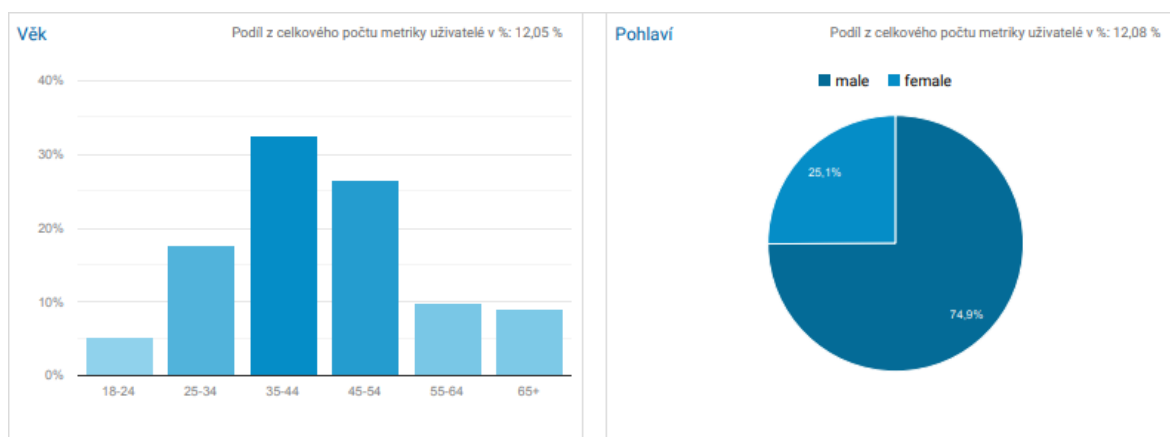
Monitorování uživatelů v reálném čase

První funkcí jsou informace, které služba monitoruje v reálném čase. Najdeme zde informace jako aktuální počet aktivních uživatelů na webu, z jaké země se aktivní uživatelé připojili, z jakého zdroje naši stránku navštívili a v jaké části stránky se momentálně nachází. Dále se dozvíme, kolik právě aktivních uživatelů uskutečnilo během jejich relace objednávku, registraci nebo postoupilo v sekci košíku ke vyúčtování.

Publikum

Sekce publikum nabízí velké množství zajímavých podstránek. První podstránka přehled uvádí celkový počet uživatelů v daném období, počty návštěv a míru okamžitého opuštění. Je zde vyobrazen také graf, který ukazuje poměr nových a vracejících se uživatelů.

Demografické údaje ukazují informace o lidech, kteří web používají. Google Analytics zde odhalí jejich věk a pohlaví. Dle pohlaví nebo věku lze poté zobrazit údaje jako návštěvnost, míra odchodu, konverzní poměr a mnoho dalšího. Díky těmto informacím lze snadno rozpoznat, která část obyvatelstva nejlépe reaguje na sdělení webu a určit tak cílovou skupinu. Nejvíce uživatelů e-shopu Martenz se pohybuje ve věku 35-44 let, další velkou zákaznickou skupinou jsou uživatelé ve věkovém rozmezí 45-54 let. Uživatelskou skupinu tvoří ze 74,9% muži a 24,1% ženy.



Obrázek 7 Demografické údaje za období 1. 2. 2022 - 8. 5. 2022 (zdroj: Google Analytics)

Další podstránkou jsou zájmy. Tato sekce rozděluje uživatele dle jejich zájmů. Vlastník webu poté vidí, které zájmové skupiny navštěvují stránky nejvíce, jak se na stránkách chovají a v jaké míře vykonávají konverzi. Z těchto dat lze poté vytvářet rozhodnutí, na kterou cílovou skupinu se při cíleném marketingu zaměřit, popřípadě kterou cílovou skupinu vyřadit. Po spuštění kampaně lze pak i sledovat její úspěšnost. Každou kategorii zájmu lze rozkliknout pro zobrazení výsledků dle věku uživatelů. Podstránka, která se nachází ihned za zájmy je rozdělení dle zaměstnání. Díky tomuto rozdělení lze rozšířit znalosti ohledně okruhu potenciálních zákazníků. Na e-shopu Martenz tvoří nejvíce konverzí uživatelé se zájmy o zprávy a politiku, jídlo a vaření, cestování a životní styl.

Kategorie zájmů (zásah)	Podíl z celkového počtu metricky uživatelé v %: 12,72 %	Segment podle zájmu	Podíl z celkového počtu metricky uživatelé v %: 12,01 %
4,65 %	News & Politics/Avid News Readers	3,42 %	Autos & Vehicles/Motor Vehicles
4,60 %	Food & Dining/Cooking Enthusiasts/30 Minute Chefs	3,26 %	Travel/Trips by Destination/Trips to Europe
3,88 %	Lifestyles & Hobbies/Outdoor Enthusiasts	3,06 %	Consumer Electronics/Mobile Phones
3,60 %	Lifestyles & Hobbies/Business Professionals	2,99 %	Real Estate/Residential Properties
3,42 %	Home & Garden/Do-It-Yourselfers	2,04 %	Financial Services/Investment Services
3,23 %	Sports & Fitness/Sports Fans	2,01 %	Home & Garden/Home Appliances/Small Appliances
3,07 %	Media & Entertainment/TV Lovers	1,74 %	Home & Garden/Home Appliances
2,91 %	Sports & Fitness/Health & Fitness Buffs	1,68 %	Consumer Electronics/Televisions
2,90 %	Travel/Travel Buffs	1,68 %	Travel/Bus & Rail Travel
2,79 %	Media & Entertainment/Book Lovers	1,64 %	Computers & Peripherals/Computers/Laptops & Notebooks

Obrázek 8 Segmentace zákazníků dle jejich zájmů za období 1. 2. 2022 - 8. 5. 2022 (zdroj: Google Analytics)

Geografické údaje rozdělují návštěvníky dle jejich lokality a používaného jazyka. Tyto údaje jsou obzvláště užitečné, pokud je web orientován mezinárodně. Společnost Martenz nabízí své produkty v České republice a na Slovensku. Po spuštění kampaně pak může společnost na základě získaných informací hodnotit její úspěšnost dle lokality. E-shop navštěvují převážně uživatelé z České republiky a Slovenska, návštěvnost z těchto zemí je pak nejvíce ovlivněna probíhající kampaní.

Přehled technologie nabízí informace o tom, jaký prohlížeč návštěvníci nejčastěji používají. Pokud bude u určitého prohlížeče zjištěna vysoká míra okamžitého opuštění, je velice pravděpodobné, že jeden z doplňků, který stránka používá, nefunguje správně a je třeba ho opravit. Mobilní přehled zobrazuje rozložení návštěvníků mezi počítači, mobily a tablety. Pokud se na některém z těchto zařízení vyskytuje podezřele malá konverze, je pravděpodobné, že web není pro dané zařízení dobře optimalizovaný. Projevy mohou být nečitelné texty, špatně umístěná nebo malá tlačítka, nepoužitelný košík. Web je tedy třeba opravit a optimalizovat. E-shop je nejvíce navštěvován z prohlížečů Android Webview, Chrome a Safari. Žádný z těchto prohlížečů nevykazuje příliš vysoké hodnoty v míře okamžitého opuštění, a tak lze posoudit, že jsou stránky dobře optimalizovány.

Podstránka se zařízeními zobrazí, jak výkonné zařízení uživatelé pro návštěvy webu používají. Pokud uživatelé s méně výkonnými zařízeními stránky rychle opouštějí, může to znamenat, že jsou stránky náročné a je třeba provést lepší optimalizaci.

Prohlížeč	Akvizice			Chování		
	Uživatelé	Noví uživatelé	Návštěvy	Míra okamžitého opuštění	Počet stránek na 1 relaci	Prům. doba trvání relace
	3 271 Podíl z celku v %: 100,00 % (3 271)	3 074 Podíl z celku v %: 100,03 % (3 073)	4 646 Podíl z celku v %: 100,00 % (4 646)	46,60 % Prům. pro výběr dat: 46,60 % (0,00 %)	3,64 Prům. pro výběr dat: 3,64 (0,00 %)	00:02:07 Prům. pro výběr dat: 00:02:07 (0,00 %)
1. Android WebView	1 131 (34,49 %)	1 060 (34,48 %)	1 291 (27,79 %)	51,82 %	3,31	00:01:35
2. Chrome	982 (29,95 %)	905 (29,44 %)	1 681 (36,18 %)	43,07 %	3,93	00:02:30
3. Safari (in-app)	392 (11,95 %)	391 (12,72 %)	426 (9,17 %)	50,94 %	3,03	00:01:17

Obrázek 9 Technologie – prohlížeč a operační systém za období 1. 2. 2022 - 8. 5. 2022 (zdroj: Google Analytics)



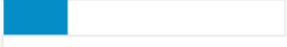



Akvizice

Přehled s názvem akvizice podrobně zobrazuje zdroje, které přivádějí návštěvníky na web. V obecném přehledu lze vidět nejlepší kanály, počty návštěv uživatelů a konverzní poměr.

Po rozkliknutí záložky veškerá návštěvnost a zvolení kategorie kanály lze vidět přehled všech kanálů, které přivedly návštěvníky na web za zvolené období. V tabulce, která se nachází pod grafem se nachází počty všech uživatelů, nových uživatelů a celkový počet návštěv. Dalšími údaji jsou míra okamžitého opuštění, počet stránek na 1 relaci a průměrná doba trvání relace. Nechybí zde ani konverze, která zobrazuje počet transakcí zákazníků z daného kanálu a tržby. Z těchto údajů lze měřit úspěšnost reklamních kampaní, zjistit, který kanál má největší tržby a nadále do něj investovat. Naopak pokud zjistíme, že nám určitá kampaň nepřináší uspokojivé výsledky, můžeme ji předčasně ukončit, aby nedocházelo ke zbytečným peněžním ztrátám. Nejvíce návštěvníků přichází na e-shop Martenz přímo z webu společnosti, kde jsou uživatelé odkázáni z marketingových kampaní.

Dále se zde nachází záložka sociální sítě, která nám podrobně ukazuje, jaký vliv mají sociální sítě na návštěvnost a konverze. Podstránka vstupní stránky ukazuje, ze kterých sociálních sítí návštěvníci přišli, podstránka konverze nám zase ukazuje, přes které sociální sítě vzniká největší počet konverzí. Díky těmto informacím máme dokonalý přehled o výkonnosti jednotlivých sociálních sítí a vidíme, které investice do reklamy se nám vyplácejí a naopak. Kampaně na sociálních sítích odkazují na web společnosti, nikoliv na její e-shop. Přes tento web se návštěvníci následně prokliknou na e-shop. V Google Analytics pro e-shop tedy nelze tyto data vyhodnocovat. Na návštěvnost má však největší vliv sociální síť Facebook,

na které běží většina marketingových kampaní. Stejně tak tomu bude i s uskutečněnými konverzemi.

		Akvizice			Chování		
		Uživatelé ↓	Noví uživatelé ↓	Návštěvy ↓	Míra okamžitého opuštění ↓	Počet stránek na 1 relaci ↓	Prům. doba trvání relace ↓
		3 271	3 073	4 646	46,60 %	3,64	00:02:07
1	Direct	2 012			41,25 %		
2	Paid Search	742			57,44 %		
3	Other Advertis	344			62,56 %		

Obrázek 10 Akvizice dle zdrojů, které přivádí zákazníky na e-shop za období 1. 2. 2022 - 8. 5. 2022, kde je zdroj Direct zkruslen prokliky ze sociálních sítí (zdroj: Google Analytics)

Poslední záložka kampaně nám zobrazí všechny probíhající kampaně a jejich úspěšnost. Je zde zobrazena i výkonnost placených klíčových slov. Analýza nákladů zobrazí kolikrát se reklama zobrazila ve vyhledávací síti, kolikrát na ni uživatelé klikli a kolik nás kampaň celkově stála. Dále jsou tu zobrazeny podrobné údaje jako procentuální podíl zobrazení reklamy, která vedla ke kliknutí, průměrná cena za proklik, tržby za proklik a návratnost investic do reklamy.

Chování

Přehled chování nám nabízí ještě podrobnější informace o dění na webových stránkách. V podstránce obsah webu všechny stránky se zobrazí veškeré stránky na webu. U těchto stránek lze zobrazit informace jako celkový počet jejich zobrazení, unikátní zobrazení stránek, průměrná doba na stránce, míra okamžitého opuštění, procento odchodů a hodnota stránky. Dále se zde nachází kategorie výstupní stránky, která nám ukazuje, ze kterých stránek nejčastěji návštěvníci odcházejí. Díky těmto informacím lze odhalit na kterých stránkách se nachází chyba a je třeba ji opravit.

Konverze

Záložka s názvem konverze slouží k měření výkonnosti elektronického obchodu. Tato funkce se nespouští automaticky, ale je ji třeba spustit v nastavení zvlášť. Po spuštění této funkce se všechny transakce, které na webu vznikly, zasílají do Google Analytics. V základním zobrazení podstránky elektronického obchodu lze vidět přehled tržeb a konverzního poměru za posledních 7 dní. Dále se zde nachází počet transakcí, průměrná hodnota objednávky a výkon jednotlivých produktů.

Chování nakupujících prozradí, jak návštěvníci na e-shopu nakupovali. Nachází se zde porovnání všechny relace vs žádná aktivita v nákupech, návštěvy se zhlédnutím produktů vs žádné přidání do košíku, návštěvy s přidáním do košíku vs opuštění košíku a návštěvy s navštívením pokladny vs opuštění pokladny. Tyto informace prozradí, kde nejčastěji dochází ke ztrátě zákazníka při nákupním procesu, případně správné načasování remarketingu.

Výkon produktů lze do detailů sledovat v samostatné záložce a nachází se zde celkové tržby za produkt, unikátní nákupy, množství, průměrná cena a průměrné množství na nákup. Dle těchto informací lze snadno odhalit ty nejlepší produkty a více se tak zaměřit na jejich propagaci.

7.2 Facebook pixel

Dalším využitím data miningu ve společnosti Martenz je Facebook pixel. Jedná se o měřicí kód, který je implementován do hlavičky webu. Shromažďuje a odesílá údaje o návštěvnosti a splněných konverzí na webu společnosti Martenz do nástroje Facebook Ads.

Facebook pixel funguje na velice podobné bázi jako Google Analytics. Pokud uživatel navštíví web, Pixel umístí do jeho prohlížeče cookies a údaje o tom, co uživatel na webu dělá a odesílá je ihned do Facebooku. Pixel poskytuje detailní přehled o návštěvnosti webu, splněných konverzích a hodnotě objednávek.

Facebook pixel umí oproti Google Analytics spárovat jednotlivé návštěvníky s konkrétními uživateli Facebooku nebo Instagramu, což mu propůjčuje obrovskou výhodu. Facebook ví úplně přesně kdo na web přišel, ale majitel webu se to kvůli GDPR nedozví. Slouží to tedy k tomu, že Facebook dokáže evidovat stejnou osobu, která na stránky přichází z více zařízení a data tak nejsou zkresleny. Google analytics tuto problematiku zatím neumí vyřešit, a tak každé zařízení které webové stránky navštíví počítá jako unikátního zákazníka, přičemž dochází ke značnému zkreslení dat.

Společnost Martenz používá placenou reklamu na Facebooku a Instagramu a Facebook pixel je tak pro ni nutností. Bez Facebook pixelu není možné provádět Facebook remarketing a dynamický marketing, který zobrazuje přesně ty produkty, které uživatelé na webu navštívili.

Facebook pixel umožňuje měření úspěšnosti konverzí. Dokáže analyzovat úspěšnost PPC (pay-per-click) kampaní a určit, jestli se investice do reklamy vrací nebo ne. Další funkcí je Facebook machine learning, díky které se Facebook učí z výsledků reklam a snaží se zlepšovat jejich výkonnost.

K implementaci Facebook pixelu je nutné mít založený Facebook Business Manager, který všechny výsledky, které pixel nasbírá dokáže zobrazit a je tak s nimi možné dál pracovat. Společnost Martenz Facebook Business Manager již používá a je jejím nástrojem pro marketing na sociálních sítích Facebook a Instagram. Jelikož společnost vlastní internetové stránky platformem, které jsou partnerem Facebook pixel je implementace velmi jednoduchá. V administraci webu se nachází plugin, do kterého stačí vložit zkopírovaný kód. Odpadá tak nutnost ručního přidání kódu do webu.

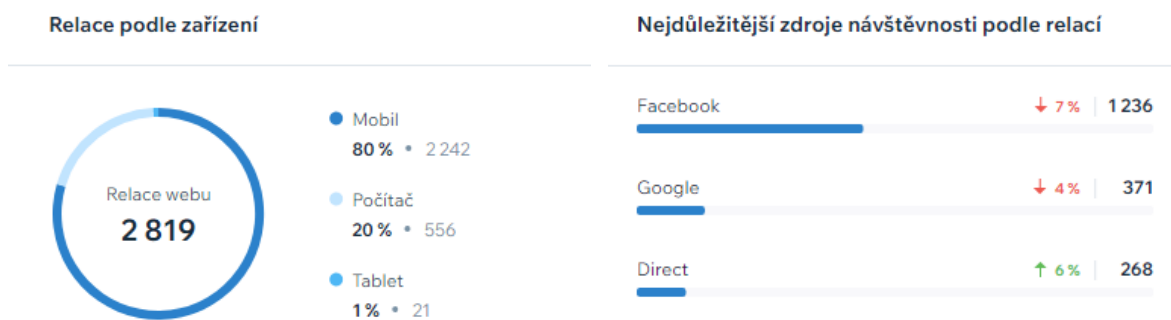
7.3 Wix

Společnost Martenz používá pro své webové stránky platformu od společnosti Wix. Jedná se o webový editor, který nabízí možnosti e-shopu, systém CRM a také stručnou analýzu návštěvnosti webu.

V počáteční fázi společnosti Martenz si zákazníci objednávali produkty společnosti Martenz prostřednictvím prostého kontaktního formuláře na webových stránkách. Nyní společnost využívá od společnosti Wix pouze službu webových stránek a e-shop byl přesunut na platformu od společnosti Shoptet. Společnost nabízela velmi malý počet produktů, a tak byl kontaktní formulář dostačující. Data, které zákazníci vyplnili v objednávkovém formuláři se poté uložily do zákaznické databáze CRM systému společnosti Wix.

V této databázi jsou uloženy informace jako jméno, příjmení, telefonní číslo a emailová adresa. Informace, které se zde naopak neukládají jsou datum posledního nákupu, historie nákupů, předmět nákupu a peněžní částka. S informacemi, které se ukládají lze tedy pracovat pouze v omezené míře. Je nutné si manuálně zakládat a třídit kontakty do skupin, dle jejich nákupního chování. Pomocí nasbíraných kontaktů lze poté využít emailový marketing nebo propagace na sociálních sítích.

Dále se zde nachází funkce analytika a přehledy. V přehledu návštěvnosti lze velice přehledně sledovat podobné údaje, které zobrazuje nástroj Google Analytics. Jsou zde opět zobrazeny nejdůležitější zdroje návštěvnosti, nejdůležitější stránky dle relací, relace dle zařízení, noví vs. vracející se zákazníci a relace dle země za vybrané období. Z těchto údajů se lze dozvědět, že za posledních 30 dní na web společnosti Martenz přichází nejvíce návštěvníků z platformy Facebook, Google a po přímém vyhledání na webu. Nejvíce uživatelů navštíví web prostřednictvím mobilního telefonu (80 %), počítače (20 %) a z tabletů pak stránky navštěvuje nejmenší počet uživatelů (necelé 1 %). Co se týče relací dle země, navštěvuje web více uživatelů z České republiky (52 %) než ze Slovenska (46 %).



Obrázek 12 Relace podle zařízení za posledních 30 dní (17.4.-16.5. 2022) (zdroj: Wix.com)

Obrázek 11 Zdroje návštěvnosti za posledních 30 dní (17.4.-16.5. 2022) ve srovnání s předchozím obdobím (18.3.-16.4. 2022) (zdroj: Wix.com)

Pokud by společnost Martenz měla na platformě Wix e-shop, mohla by si zobrazit nasbíraná data z prodejů a přehled o zákaznících, kteří u ní nakupují. E-shop je však na jiné platformě a k zobrazení výsledků zde nejsou nasbíraná data. Společnost tak tuto funkci nemůže využívat.

Oproti Google Analytics nabídne Wix navíc funkci statistiky, kde zobrazuje vyhodnocení nasbíraných dat a zobrazuje tak příležitosti, trendy a oznámení. U příležitostí by mělo být zvaženo využití relativních výhod, trendy zobrazují přehled výkonnosti webu v čase a u oznámení by měla být zvažena optimalizace webu. Příležitosti vidí platforma Wix u stránek, které jsou mezi návštěvníky nejoblíbenější, na kterých tráví uživatelé nejvíce času a u stránek, u kterých byl zaznamenán vysoký nárůst návštěvnosti. Zobrazení trendů poté identifikuje výraznou kladnou změnu oproti minulému období, což je momentálně u společnosti Martenz nejoblíbenější zdroj návštěvnosti Facebook. Oznámení zobrazuje poslední stránky, které lidé často navštíví předtím, než web opustí a stránky, které zaznamenávají výrazný pokles návštěvnosti.

Nejoblíbenější stránka mezi návštěvníky je / (Domovská stránka)	Příležitost
Lidé tráví více času na vaší stránce /single-post/2018/10/03/jak-poznat-kvalitni-ovocnou-palenku-nejzajimaveji-ze-stranek	Příležitost
Poslední stránka, kterou lidé často navštíví předtím, než web opustí, je /single-post/2018/01/08/jaka-je-vhodna-teplota-na-konzumaci-palenek	Oznámení

Obrázek 13 Funkce statistiky (zdroj: Wix.com)

Další funkcí jsou benchmarky, které zjišťují, jak se stránce majitele daří v porovnání s ostatními weby. Jsou zde zobrazena a následně porovnána nasbíraná data návštěvnosti s konkurencí se stejnou kategorií webu. Lze zde zvolit celosvětové porovnání nebo porovnání dle konkrétní země. Na srovnání s konkurencí v České republice nemá platforma Wix dostatek nasbíraných dat, a tak je nutné zvolit celosvětové srovnání. Po načtení výsledků se zobrazí radar konkurence, který vizuálně znázorňuje výsledky benchmarků. Následně platforma vypíše doporučení pro zlepšení statistik.

Srovnání klíčových statistik

Vy v porovnání s weby z odvětví [Výrobci a prodejci potravin a nápojů](#) v oblasti [Celosvětově](#). [Změnit kategorii](#)

Návštěvnost	Vaše statistiky	Lepší než	
Unikátní návštěvníci	2,430	95 %	>
Prům. doba trvání relace	2 min 7 s	74 %	>
Míra opuštění	40 %	76 %	>
Vracející se návštěvníci	19 %	85 %	>
Organické hledání	18 %	27 %	>
Sociální sítě	9 %	65 %	>

Statistiky jsou relevantní pro posledních 30 dní: 17. 4. – 16. 5.

Obrázek 14 Funkce Benchmark pro srovnání klíčových statistik (zdroj: Wix.com)

Poslední analytickou funkcí je rychlost webu a následné doporučení pro jeho zrychlení. Dle platformy Wix je rychlost webové stránky společnosti Martenz průměrná s tím, že většina návštěvníků interaguje s webem během 5,9 sekund nebo méně. V porovnání s konkurencí je web rychlejší než 47 % webů ve stejném prodejním odvětví po celém světě. Je zde možné zjistit i rychlost webu dle používaného zařízení. Kde při použití počítače web získal rychlostní skóre 71/100 a u mobilních zařízení pouze 16/100 přičemž 100 je nejlepší možný výsledek.

7.4 Shoptet

Společnost Martenz používá pro svůj e-shop platformu od společnosti Shoptet, která vlastní od roku 2016 titul jedničky na českém trhu platformem internetových obchodů.

Tato platforma v základu nabízí pouze zákaznickou databázi, seznam odběratelů newsletteru a informaci o tržbách.

V zákaznické databázi jsou základní údaje zákazníků, ale chybí zde opět podrobnější informace jako datum posledního nákupu, nákupní historie, produkty, které zákazník nejčastěji nakupuje a průměrná částka nákupu. Zákaznická databáze tak nenabízí dostatečné množství informací na jakoukoliv další analýzu.

Výhodou platformy Shoptet je možnost přidání velkého množství doplňků přímo z oficiálního katalogu. Společnost Martenz prozatím žádné doplňky nevyužívá.

7.5 Shrnutí

Společnost aktuálně využívá sběr dat a následnou analýzu od platform Google Analytics, Facebook pixel, Wix a Shoptet.

Platformy Google analytics, Facebook pixel a Wix pomáhají při vytváření a hodnocení výkonnosti marketingových kampaní a optimalizaci webových a e-shopových stránek.

Facebook pixel pomáhá pomocí nasbíraných dat optimalizovat marketing na sociálních sítích.

Platformy Wix a Shoptet vytváří zákaznické databáze. Tyto platformy však mezi sebou nejsou propojeny a emailový marketing probíhá prostřednictvím zákaznických databází v platformě Wix. Nové kontakty jsou však aktuálně automaticky vytvářeny v databázi platformy Shoptet. V současnosti je tedy nutné tyto nové kontakty manuálně importovat do zákaznické databáze Wix, což je časově velmi náročné.

Ve společnosti je tedy co se týče data miningu prostor pro zlepšení, které bude doporučeno v následující kapitole.

8 DOPORUČENÍ IMPLEMENTACE DATA MININGU VE SPOLEČNOSTI PRO JEHO EFEKTIVNĚJŠÍ VYUŽITÍ

Vhledem k nedokonalostem aktuálních zákaznických databází není možné provádět kompletní analýzu prodeje. Společnost nemá k dispozici data, která by jí umožňovala určit ty nejlepší zákazníky, vytvářet relevantní speciální nabídky nebo například odhalit zákazníky, kteří si pravidelně nevyzvednou svoji objednávku. Tyto databáze si lze vytvářet manuálně, avšak s aktuálním počtem zákazníků by se jednalo o časově velice náročnou práci. Proto je třeba tyto akce automatizovat.

Mým doporučením je tedy implementace CRM systému, pomocí kterého by všechna tato potřebná data byla zaznamenávána automaticky a na jednom místě.

V současnosti se na trhu vyskytuje velké množství CRM systémů, ale ne všechny jsou kompatibilní s využívanými platformy společnosti. Pro automatizaci celého procesu je nutné napojení CRM systému na e-shopovou platformu Shoptet. Systém, který tuto možnost nabízí je Raynet CRM.

8.1 Raynet CRM

Raynet CRM je nejoblíbenější CRM systém v České republice s více než 12000 aktivními uživateli. Jedná se o cloudový systém a přístup ke svému účtu je tak možný kdekoliv, dokonce i na mobilním zařízení. Oproti konkurenčním CRM systémům nabízí přímou integraci s platformou Shoptet a není tak nutné data přeposílat přes externí pluginy. Všechny provedené obchody a data s nimi související se tak automaticky uloží do CRM databáze Raynet. Dále systém nabízí přímou integraci s mnoho dalšími systémy třetích stran, jako jsou třeba služby Google a Microsoft. Tento systém zároveň v současnosti využívají také velké společnosti jako Liftago, REMAX a ČSFD.

Základní funkce, které systém Raynet CRM nabízí jsou adresář, obchodní případy, aktivity a analýzy.

Adresář

První a velice důležitou funkcí pro společnost Martenz je pokročilá databáze klientů. Do této databáze se ukládají veškerá data týkající se klientů a jejich objednávek. Nalezneme zde jméno a příjmení klienta, bydliště klienta, kompletní historii nákupů, rating klienta a mnoho dalšího. Je zde možnost i pokročilého filtru, pomocí kterého lze klienty roztřídit dle nejmenších detailů. Tato funkce je velice užitečná pro následný cílený marketing. Adresář

nabízí i možnost importu a exportu klientů. Společnost Martenz si tak může importovat své dosavadní kontakty z platformy Wix. Export kontaktů pak může být využit pro emailový marketing. Všechny kontakty jsou zde uchovávány v souladu s GDPR.

Obchodní případy

Další funkcí jsou obchodní případy. Společnost tak získá nadhled a kontrolu nad probíhajícími objednávkami. V této databázi jsou uchovány všechny objednávky a pomocí pokročilého filtru lze mezi nimi efektivně vyhledávat.

Aktivity

Funkce aktivity nabízí přehledný kalendář, který zobrazuje úkoly, schůzky a události. Aktivity si lze propojit s kalendářem od ostatních společností a mít tak vše na jednom místě.

Analýzy

Raynet CRM nabízí také analýzy obchodu. Nachází se zde prodejní trychtýř, výkon prodejního trychtýře, odhad prodeje, vývoj prodeje, prodej dle produktů, prodej dle obchodníků a výkon produktu. V oblasti ziskovosti najdeme analýzu vývoje zisku, ziskovost obchodních případů a ABC analýzu, která zobrazuje, jak se jednotliví klienti podílejí na celkovém obratu. Je zde i analýza obchodníků, ale tu společnost Martenz nevyužije. Poslední nabízenou analýzou je mapová analýza, která zobrazuje rozmístění klientů a aktivit na mapě. Tyto analýzy společnosti Martenz zobrazí nejhodnotnější zákazníky.

8.2 Implementace Raynet CRM ve společnosti

Vzhledem k přímé integraci platform Shoptet a Raynet CRM je implementace velice jednoduchá a efektivní. Nejdříve je nutné, aby si společnost založila účet u Raynet CRM. V oficiálním obchodě doplňků Shoptet s názvem Shoptet doplňky vyhledáme rozšíření Raynet a poté ho objednáme. Na email společnosti poté dojde aktivační kód, který slouží k propojení platformy Shoptet a Raynet CRM. V nastavení Raynet CRM se nachází sekce rozšíření, kde vyhledáme Shoptet. Pro aktivaci zadáme aktivační kód z emailu a administrátorské ID Shoptetu. Raynet CRM je nyní propojen s platformou Shoptet a přenos objednávek tak aktivován.

Společnost Martenz poté importuje své zákazníky z databáze na platformě Wix a Raynet CRM je plně připraven k použití.

9 NÁKLADOVÁ, ČASOVÁ A RIZIKOVÁ ANALÝZA DOPORUČENÍ

Analýza času a nákladů

Realizace implementace CRM systému Raynet není časově náročná, samotná aplikace trvala v řádu desítek minut. Časově náročnější tak bylo získávání potřebných informací k samotné možnosti implementace. Import kontaktů z dosud využívané platformy Wix je časově náročnější, a to z důvodu potřeby úprav importního dokumentu. Co se týče nákladů, platforma Raynet a Shoptet si účtují poplatky za využití této služby zvlášť. Platformě Raynet se platí měsíční poplatek za využití služby CRM a platformě Shoptet za využívání doplňku, pomocí kterého je možné automaticky odesílat data z e-shopu do systému CRM. Přesná kalkulace nákladů bude níže v tabulce. Jelikož se jedná o cloudový software, tak nejsou kladeny žádné speciální nároky na nákup hardwarového vybavení. Společnost Martenz tak může k využívání nového systému CRM používat dosavadní vybavení.

Tabulka 2 Explicitní náklady na implementaci systému Raynet CRM (zdroj: vlastní tvorba)

Explicitní náklady		
Platforma	Náklady měsíčně (bez DPH)	Náklady ročně (bez DPH)
Raynet CRM	650 Kč	7 800 Kč
Shoptet	150 Kč	1 800 Kč
Celkem	800 Kč	9 600 Kč

Tabulka 3 Implicitní náklady na implementaci systému Raynet CRM (zdroj: vlastní tvorba)

Implicitní náklady		
Název nákladu	Náklady měsíčně	Náklady ročně
Čas věnovaný obsluze systému	4 000 Kč	48 000 Kč

K pokrytí explicitních a implicitních nákladů stačí, aby realizace doporučení zvýšila obrat společnosti Martenz o 3 %, při předpokladu 50 % marže z výrobku, která se v tomto odvětví běžně uplatňuje.

Rizika

Při velké vytíženosti nebo aktualizací serverů Raynet CRM a Shoptetu může docházet ke zpomalení přenosu dat mezi jednotlivými platformami. Hrozí tedy zpomalený tok dat, který již byl uživateli systému v minulosti zaznamenán. Eliminace tohoto rizika však není možná,

jelikož se jedná o záležitosti, které nelze ovlivnit. Dalším rizikem může být neznalost týkající se obsluhy CRM systému pro nové uživatele. Systém je však uživatelsky velice přívětivý a toto riziko lze případně eliminovat sledováním krátkých video sérií na stránkách poskytovatele, kde je vše podrobně vysvětleno. V budoucnosti se také může stát, že poskytovatel změní cenu služby a její využití tak nebude udržitelné, tuto skutečnost však nelze nijak ovlivnit. Platformy Shoptet a Raynet CRM můžou v budoucnosti ukončit spolupráci a doplněk tak již nebude možné používat, jedná se tedy o riziko, které nelze eliminovat. Posledním shledaným rizikem je zánik platformy Raynet CRM, toto riziko však není možné eliminovat.

Tabulka 4 Rizika spojená s implementací systému Raynet CRM (zdroj: vlastní tvorba)

Riziko	Pravděpodobnost výskytu rizika (1-5)	Závažnost rizika (1-5)
Personál	2	3
Nefunkčnost systému	1	4
Zvýšení ceny	2	3
Přerušení integrace	1	5
Zánik poskytovatele	1	5
Rizika se hodnotí na stupnici 1-5, přičemž 5 je nejhorší možný výsledek		

ZÁVĚR

Cílem teoretické části této bakalářské práce bylo zpracovat poznatky týkající se digitálního data miningu a možnosti jeho využití v managementu a ekonomii. V praktické části bakalářské práce bylo cílem na základě analýzy současného stavu využití data miningu ve vybrané společnosti vytvořit návrh doporučení jeho efektivnějšího využití. Bakalářská práce tak byla rozdělena na dvě hlavní části, teoretickou a praktickou.

Teoretická část byla rozdělena na pět kapitol. První kapitola se věnovala vymezení základních pojmů v předmětné oblasti, konkrétně pojmy big data, data mining, business intelligence, umělá inteligence, CRM (řízení vztahu se zákazníky), data warehouse, data lake, atribut a data set.

Druhá kapitola byla orientována na současně využívané techniky data miningu, kde jsem popsal princip asociace, klasifikace, shlukování, rozhodovacích stromů, predikcí a neuronových sítí.

Třetí kapitola se týkala pokročilých technik data miningu, které v dnešní době rychle narůstají na oblibě a jsou nimi web mining, text mining a multimedia mining.

Čtvrtá kapitola se zabývala procesem dolování dat s názvem CRISP (mezioborový standardní proces pro dolování dat). Jedná se o nejrozšířenější analytický model, který popisuje běžné přístupy dolování dat. Tato kapitola pomáhá pochopit podstatu data miningu a vysvětluje, jak tento proces funguje.

Pátá kapitola pojednává o možnosti využití data miningu dle oborů. První obor, který byl v této práci zmíněn byl marketing. Jedná se o obor, ve kterém je uplatnění data miningu velice rozsáhlé, a to především v oblasti managementu, kde se na základě výsledků data miningu dělají důležitá marketingová rozhodnutí. Druhým oborem byla ekonomie, kde se data mining využívá k posuzování finančního chování subjektů. Konkrétně je zde řešeno využití data miningu ve finančním sektoru a bankovníctví.

V praktické části byla představena společnost Martenz, její produkty, historie, vize a cíle. Pomocí poznatků získaných v teoretické části této bakalářské práce byla provedena analýza současného využití data miningu ve společnosti. Po provedení analýzy bylo zjištěno, že by společnost mohla využívat data mining efektivněji, a že se zde nachází prostor pro zlepšení. Na základě analýzy bylo tedy vytvořeno doporučení pro efektivnější využití data miningu ve společnosti. Konkrétně doporučení nového systému CRM (systém řízení vztahů se

zákazníky), které bylo společnosti předloženo a později také implementováno. Zmíněné doporučení bylo také podrobena nákladové, časové a rizikové analýze. Tyto analýzy ukázaly náklady spojené s doporučením, jeho rizika a případnou eliminaci rizik a jeho časovou náročnost.

SEZNAM POUŽITÉ LITERATURY

Monografie:

AGGARWAL, Charu, 2015. *Data mining: the textbook*. Cham: Springer. ISBN 9783319141411.

HAN, Jiawei, Micheline KAMBER a Jian PEI, c2012. *Data mining: concepts and techniques*. 3rd ed. Waltham: Morgan Kaufmann. Morgan Kaufmann series in data management systems. ISBN 9780123814791.

HOLUBOVÁ, Irena et al., 2015. *Big Data a NoSQL databáze*. Praha: Grada. Profesionál. ISBN 9788024754666.

KOTU, Vijay a Bala DESHPANDE, 2014. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Elsevier Books. ISBN 13: 9780128014608.

LEVENTHAL, Barry, 2018. *Predictive analytics for marketers: using data mining for business advantage*. London: Kogan Page. ISBN 9780749479930.

Internetové zdroje:

DURAND, Taetse a Marie HATTINGH, 2020. Data Mining and Artificial Intelligence Techniques Used to Extract Big Data Patterns. In: *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*. IEEE, s. 1-8. ISBN 978-1-7281-9520-9. Dostupné z: doi:10.1109/IMITEC50163.2020.9334069

FRANKENFIELD, Jake, 2021. Business Intelligence (BI). *INVESTOPEDIA* [online]. [cit.2022-02-16]. Dostupné z: <https://www.investopedia.com/terms/b/business-intelligence-bi.asp>

Google Analytics [online], c2022. [cit. 2022-05-16]. Dostupné z: <https://analytics.google.com>

HARDING, Matthew a Jonathan HERSH, 2018. Big Data in economics: New Tricks for Econometrics. In: *IZA World of Labor*. 28(2), s. 3-28. ISSN 20549571. Dostupné z: doi:10.15185/izawol.451

Introduction to Data Lakes, c2022. *Databricks* [online]. [cit. 2022-05-09]. Dostupné z: <https://databricks.com/discover/data-lakes/introduction>

Martenz [online], c2017. Zlín [cit. 2022-05-09]. Dostupné z: <https://www.martenz.cz>

OSMAN, Abdullahi Sidow, 2019. Data Mining Techniques: Review. *International journal of data science research* [online]. 2(1), 3 [cit. 2022-04-07]. Dostupné z: <http://ojs.mediu.edu.my/index.php/IJDSR/article/view/18417>

PLAPINGER, Thomas, 2017. What is a Decision Tree?. *Towards data science* [online]. [cit. 2022-05-09]. Dostupné z: <https://towardsdatascience.com/what-is-a-decision-tree-22975f00f3e1>

RAMOS, Diana, 2021. Real-Life and Business Applications of Neural Networks. *Smartsheet* [online]. [cit. 2022-05-09]. Dostupné z: <https://www.smartsheet.com/neural-network-applications>
<https://www.smartsheet.com/neural-network-applications>

SINGH, Himanshu, 2019. How Data Mining in CRM Will Benefit Your Business. *Really Simple Systems* [online]. [cit. 2022-02-17]. Dostupné z: <https://www.reallysimplesystems.com/blog/data-mining-in-crm/>

SINKOV, A. et al., 2016. Neural networks in data mining. In: *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*. IEEE, s. 1-5. ISBN 978-1-5090-1322-7. Dostupné z: doi:10.1109/ICIEAM.2016.7911596

TEKIN, Mahmut et al., 2019. Data Mining in Digital Marketing: New Tricks for Econometrics. In: *Proceedings of the International Symposium for Production Research 2018*. Cham: Springer International Publishing, 28(2), s. 44-61. ISBN 978-3-319-92266-9. ISSN 20549571. Dostupné z: doi:10.1007/978-3-319-92267-6_4

Usage statistics of traffic analysis tools for websites, 2022. *W3Techs* [online]. [cit. 2022-05-09]. Dostupné z: https://w3techs.com/technologies/overview/traffic_analysis

VARIAN, Hal R., 2014. Big Data: New Tricks for Econometrics. In: *Journal of Economic Perspectives*. 28(2), s. 3-28. ISSN 0895-3309. Dostupné z: doi:10.1257/jep.28.2.3

Wix [online], c2022. [cit. 2022-05-16]. Dostupné z: <https://cs.wix.com>

ZAGAN, Elisabeta a Mirela DANUBIANU, 2019. From Data Warehouse to a New Trend in Data Architectures – Data Lake. *International Journal of Computer Science and Network Security* [online]. 19(3), 30-34 [cit. 2022-02-17]. Dostupné z: http://paper.ijcsns.org/07_book/201903/20190305.pdf

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

CRM	Customer relationship management
PPC	Pay per click
CRISP	Cross-industry standard process

SEZNAM OBRÁZKŮ

Obrázek 1 Rozhodovací strom – půjčka v bance.....	17
Obrázek 2 Neuronové sítě v data miningu.....	18
Obrázek 3 Data miningový rámec CRISP	21
Obrázek 4 Logo společnosti Martenz	37
Obrázek 5 Malinovice Gold VIP – hlavní produkt společnosti Martenz	37
Obrázek 6 Vývoj čistého obratu společnosti Martenz v letech 2017-2021 – grafické znázornění.....	38
Obrázek 7 Demografické údaje za období 1. 2. 2022 - 8. 5. 2022	41
Obrázek 8 Segmentace zákazníků dle jejich zájmů za období 1. 2. 2022 - 8. 5. 2022.....	42
Obrázek 9 Technologie – prohlížeč a operační systém za období 1. 2. 2022 - 8. 5. 2022 ..	43
Obrázek 10 Akvizice dle zdrojů, které přivádí zákazníky na e-shop za období 1. 2. 2022 - 8. 5. 2022, kde je zdroj Direct zakreslen prokliky ze sociálních sítí	44
Obrázek 11 Zdroje návštěvnosti za posledních 30 dní (17.4.-16.5. 2022) ve srovnání s předchozím obdobím (18.3-16.4. 2022).....	48
Obrázek 12 Relace podle zařízení za posledních 30 dní (17.4-16.5. 2022)	48
Obrázek 13 Funkce statistiky.....	48
Obrázek 14 Funkce Benchmark pro srovnání klíčových statistik	49

SEZNAM TABULEK

Tabulka 1 Vývoj čistého obrátu společnosti Martenz v letech 2017-2021.....	38
Tabulka 2 Explicitní náklady na implementaci systému Raynet CRM.....	53
Tabulka 3 Implicitní náklady na implementaci systému Raynet CRM.....	53
Tabulka 4 Rizika spojená s implementací systému Raynet CRM.....	54