

UNIVERZITA KARLOVA

Filozofická fakulta

Katedra psychologie



# DIPLOMOVÁ PRÁCE

Bc. Rozálie Nyirendová

**Úvod do problematiky využití pokročilých analytických postupů k  
optimalizaci personálních rozhodnutí a procesů se zaměřením na  
snižování fluktuace zaměstnanců**

**The introduction to people analytics and its usage for optimization of  
personnel decisions and processes with a focus on reduction of employee  
turnover**

Praha, 2020

Vedoucí práce: PhDr. Luděk Stehlík, Ph.D.

*Velmi děkuji PhDr. Lud'ku Stehlíkovi, Ph.D., který mi po celou dobu psaní práce poskytoval příjemnou a inspirativní zpětnou vazbu a který mě nepřestal překvapovat širokým rozsahem svých znalostí.*

*Dále bych ráda poděkovala Mgr. Matěji Kratochvílovi, který mi pomáhal s modelováním a podporoval mě při boji s chybovými hláškami v Rku.*

*Rovněž děkuji Tomáši a Radku Dohnalovi za velkou podporu a inspiraci při psaní práce a pomoc s jejím textem.*

*Současně děkuji Pavlovi, Martinovi a Kristýně za zprostředkování možnosti psát práci na dané téma a také za jejich ochotu a odborné rady.*

*Nakonec upřímně děkuji své babičce za podporu, které se mi z její strany při studiu dostalo.*

## Prohlášení

*Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.*

*V Praze dne 26. července 2020*

.....  
*Rozálie Nyirendová*

## **Abstrakt**

Cílem práce je představit možnosti využití pokročilých analytických nástrojů pro optimalizaci rozhodování v rámci personální praxe. Literárně přehledová část práce se zabývá tzv. HR analytikou, jejím vývojem, možnostmi využití a metodologickým rámcem, o který se opírá. V další části se práce věnuje konkrétnímu využití HR analytiky v oblasti retence zaměstnanců dle metodického rámce CRISP-DM. V poslední kapitole je podrobně popsán fenomén fluktuace, její důsledky a možné vysvětlující proměnné.

Empirická část práce je koncipována jako kvantitativní, aplikovaný výzkum a zabývá se dobrovolnou fluktuací zaměstnanců v konkrétní společnosti – větší české bance. V první, statisticko-inferenční části, výzkum prostřednictvím binární logistické regrese identifikuje několik statisticky signifikantních prediktorů fluktuace – míru nezaměstnanosti, počet vystřídaných týmů, dobu strávenou ve společnosti, výši platu a celkového příjmu, rychlost růstu platu, velikost týmu, výši mimořádného bonusu, platovou třídu a pohlaví. V druhé, datově-vědecké části, je sestaveno několik predikčních modelů rovněž pomocí binární logistické regrese a na základě několika technik strojového učení. Modely jsou posouzeny s ohledem na jejich přesnost a možnost využití k predikci odchodu pro jednotlivé zaměstnance. Diskuse zasazuje výsledky do kontextu předchozích studií, uvádí limity pro oba oddíly empirické části a doporučení pro společnost, které se analýza týká.

## **Klíčová slova**

HR analytika, fluktuace zaměstnanců, rozhodování založené na datech, predikční modelování

## **Abstract**

The aim of this paper is to present the possibilities of the usage of advanced analytical tools to optimize decision-making in personnel practice. The literature review part of the thesis deals with the so-called HR analytics, its development, possibilities of its usage, and the methodological framework on which it is based. The next part of the paper deals with the specific application of HR analytics in the field of employee retention according to the methodological framework of CRISP-DM. The last chapter describes in detail the phenomenon of employee turnover, its consequences, and possible explanatory variables.

The empirical part of the paper is framed as a quantitative, applied research and deals with voluntary turnover of employees in a particular company—a large Czech bank. Firstly, the statistical-inference part of the research identifies several statistically significant predictors of employee turnover through binary logistic regression—unemployment rate, number of changed teams, time spent in the company, salary and total income, salary growth rate, team size, extraordinary bonus, and gender. Secondly, in the data-science part, several prediction models are compiled, one using binary logistic regression as well and another based on several machine learning techniques. The models are assessed with regard to their accuracy and the possibility of using them to predict departure for individual employees. The discussion puts the results in the context of previous studies, outlines the limits for both sections of the empirical part, and provides the recommendations for the company to which the analysis relates.

## **Keywords**

HR analytics, employee turnover, data-driven decision making, predictive models

## Obsah

Úvod .....	4
Literárně přehledová část.....	6
1. HR analytika .....	6
1.1 Vymezení pojmu .....	6
1.2 Úvod do HR analytiky .....	7
1.3 Dovednosti potřebné k realizaci HR analytických projektů .....	8
1.4 Vývoj HR analytiky .....	9
1.5 Případové studie .....	12
2. Principy HR analytiky .....	15
2.1 Metodologie.....	15
2.2 Zdroje dat.....	17
3. Řešení problému fluktuace zaměstnanců s pomocí HR analytiky .....	19
3.1 Porozumění zkoumané oblasti a datům .....	19
3.2 Příprava dat.....	20
3.3 Analýza dat.....	21
Bodově biseriální korelace .....	21
Binární logistická regresní analýza.....	21
Analýza přežití.....	22
Analýza časových řad.....	22
Klasifikační a regresní stromy .....	23
Naivní bayesův klasifikátor ( <i>Naive bayes</i> ).....	23
Algoritmus <i>k</i> -nejbližších sousedů ( <i>k-nearest neighbors</i> ).....	24
Metoda podpurných vektorů ( <i>Support Vector Machine-SVM</i> ).....	24
Náhodné lesy ( <i>Random Forests</i> ) .....	24
<i>Extreme Gradient Boosting (XGBoost)</i> .....	25
Umělé neuronové sítě .....	25

3.3 Křížová ( <i>cross</i> ) validace .....	26
3.4 Evaluace modelu.....	26
3.5 Využití výsledků.....	28
4. Fluktuace zaměstnanců.....	30
4.1 Definice fluktuace zaměstnanců a jejích typů .....	30
4.2 Metody měření fluktuace.....	31
4.3 Důsledky fluktuace .....	31
4.3.1 Finanční dopady .....	32
4.3.2 Dopady na výkon.....	33
4.3.3 Dopady na další zaměstnance.....	34
4.4 Strategie prevence fluktuace.....	34
4.4.1 Faktory související s fluktuací.....	35
4.3.2 Předcházení fluktuace.....	43
Výzkumná část .....	45
5. Výzkumný problém, cíle výzkumu a hypotézy nebo výzkumné otázky.....	47
5.1 Statisticko-inferenční část.....	47
5.2 Datově-vědecká část.....	52
6. Design výzkumného projektu.....	54
6.1 Typ výzkumu.....	54
6.2 Metody získávání dat.....	54
6.3 Metody zpracování a analýzy dat .....	55
6.3.1 Statisticko-inferenční část.....	56
6.3.2 Datově-vědecká část .....	59
6.4 Etika výzkumu .....	60
7. Výzkumný soubor .....	62
8. Výsledky.....	63
8.1 Statisticko-inferenční část.....	63

8.2 Datově-vědecká část .....	66
9. Diskuse .....	73
9.1 Statisticko-inferenční část .....	73
9.2 Datově-vědecká část .....	76
Závěr .....	81
Seznam použité literatury .....	82
Seznam grafů .....	90
Seznam obrázků .....	91
Seznam tabulek .....	92



## Úvod

Rozhodování založené na datech nahrazuje rozhodování odvíjející se od intuice či dílčích pozorování jednotlivce. Již delší dobu je součástí mnoha organizací a souvisí s rozvojem technologií a snahou využívat data, která organizace mnohdy sbírají. S rozhodováním založeným na datech se běžně setkáváme v např. v bankovníctví, marketingu nebo v rámci optimalizace obchodního rozhodování. Subjektem, kterého se rozhodování týká, jsou často lidé. Není tedy divu, že tento trend našel v nedávné době své místo i v oblasti řízení lidských zdrojů. V rámci personální praxe organizacím umožňuje optimalizovat rozhodování týkající se zaměstnanců, a tím, v ideálním případě, maximalizovat zisky. Jednou z možných oblastí aplikace HR analytiky v rámci personální praxe je oblast predikce a prevence odchodovosti zaměstnanců, jelikož nadměrná fluktuace zaměstnanců má pro společnost mnoho negativních důsledků. Tato oblast je spolu s úvodem do HR analytiky stěžejním tématem této diplomové práce.

Práce svým tématem spadá do oblasti psychologie práce a organizace a jejím cílem je seznámit odborníky z psychologie s možností využití dat k rozhodování v personální praxi a demonstrovat tuto aplikaci na konkrétním příkladě.

První kapitola se zabývá obecným úvodem do oblasti HR analytiky, tedy seznamuje čtenáře s popisem disciplíny, terminologickými odlišnostmi, jejím vznikem a možnostmi využití v personální praxi.

V druhé kapitole jsou uvedeny technické aspekty HR analytiky, konkrétně metodologické rámce pro zvýšení úspěšnosti HR analytických projektů, a možné zdroje dat.

Třetí kapitola podrobně čtenáře seznamuje s problematikou řešení fluktuace zaměstnanců s pomocí HR analytiky.

V další části práce na ni navazuje čtvrtá kapitola, která především prohlubuje tematiku vysvětlujících proměnných fluktuace. Vedle toho také čtenáři představuje obecný úvod do fluktuace zaměstnanců spolu s jejími následky.

Nakonec v kapitolách pět až devět je popsán výzkum, jehož předmětem je predikce odchodovosti zaměstnanců a identifikace faktorů, které s odchodovostí souvisí, na jedné specifické pozici v konkrétní společnosti.

Co se týče výběru literatury, v práci jsou zařazeny vedle akademicky uznávaných zdrojů (např. metaanalytické studie) také internetové zdroje (např. blogy). Důvodem je zprvée snaha práce propojit akademický výzkum s praxí a také to, že HR analytika, jakožto nový obor, není zatím akademicky zcela upevněna. Citační norma použitá v práci je APA (2010).

# Literárně přehledová část

## 1. HR analytika

### 1.1 Vymezení pojmu

Pojem HR analytiky označuje na důkazech založený přístup k rozhodování, které se týká zaměstnanců, se snahou provádět jej co nejefektivněji ve vztahu k obchodním výsledkům společnosti (Marler & Boudreau, 2017). Vzhledem k tomu, že lidský kapitál představuje jedno z nejcennějších aktiv podniku, společnost optimálními rozhodováními ve vztahu k zaměstnancům získává kompetitivní výhodu na trhu (DiClaudio, 2019). Cílem HR analytiky je především maximalizovat zisky (resp. minimalizovat náklady) společnosti (Marler & Boudreau, 2017).

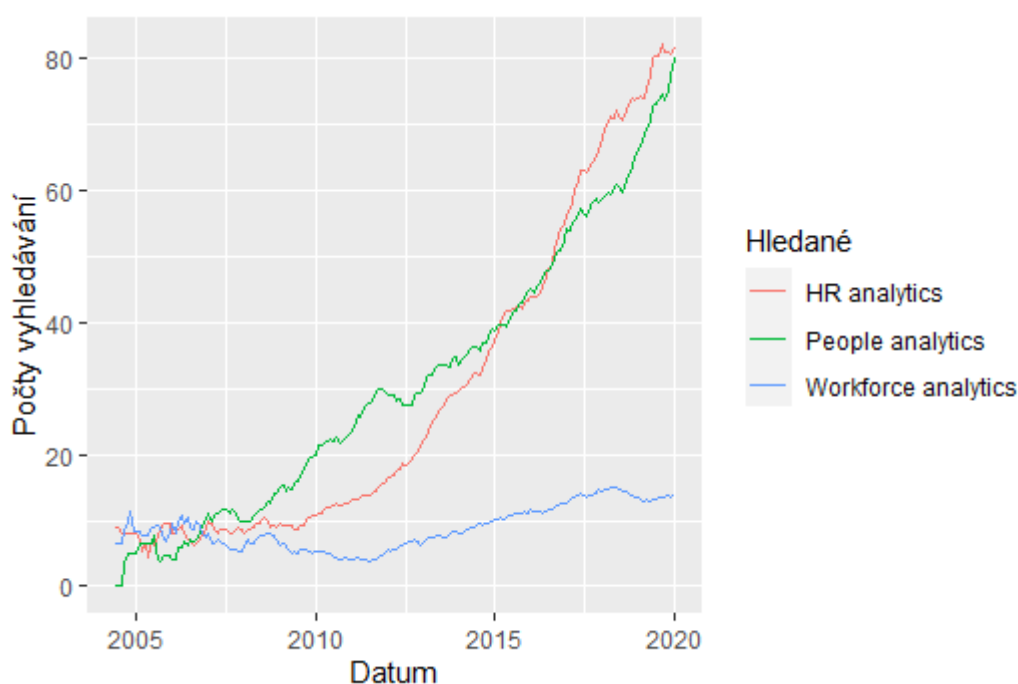
HR analytika se v anglickém jazyce skrývá pod vícero názvy – jedná se jak o doslovný překlad tzv. *HR analytics*, tak např. o tzv. *people analytics* nebo tzv. *workforce analytics*. Co se týče preferencí v užívání pojmů dle geografie, v Evropě a Indii se s vyšší frekvencí vyskytuje pojem *HR analytics*, celosvětově se těší větší oblibě pojem *people analytics* (Google Trends, 2020).

Nástrojem HR analytiky jsou nejrůznější statistické a vizualizační metody, prostřednictvím kterých data týkající se zaměstnanců a obchodních výsledků propojujeme (Marler & Boudreau, 2017; Mondore, Douthitt, & Carson, 2011). Vzhledem k tomu, že je HR analytika podložena daty, stojí v opozici vůči rozhodování, které je založeno na intuici, domněnkách o tom, co funguje, a osobních preferencích (Rousseau & Barends, 2011).

Pojem HR analytiky můžeme odlišit od termínu HR metrika. HR metrika je méně komplexní, slouží k popisu daných jevů a procesů uvnitř společnosti – např. míra fluktuace či spokojenosti zaměstnanců. Oproti tomu HR analytika zahrnuje jak využívání HR metrik, tak sofistikovanější metodiku a propojuje data z více oblastí včetně externích dat (Marler & Boudreau, 2017). Díky HR analytice společnosti mohou potenciální změny dostatečně brzy anticipovat, což jim dává dostatek času a prostoru pro naplánování potřebné strategie, aby situaci např. předešly, minimalizovaly její pravděpodobnost nebo aby se jí přizpůsobily (DiClaudio, 2019).

## 1.2 Úvod do HR analytiky

HR analytika je relativně nová a inovativní disciplína, která se v současné době nachází ve fázi počátečních osvojitelů. V současnosti si zatím disciplínu osvojilo méně než 20 % organizací, ale zájem o ni rychle roste (Marler & Boudreau, 2017). Pro ilustraci rostoucího zájmu o oblast HR analytiky slouží Graf 1 znázorňující počet vyhledání jednotlivých názvů pro HR analytiku v anglickém jazyce mezi lety 2004 a 2020 prostřednictvím internetového vyhledávače *Google* (Google Trends, 2020).



**Graf 1:** Počty vyhledávání jednotlivých pojmů, zpracováno na základě dat z Google Trends (červen, 2020)

Dle některých odpůrců HR analytiky se jedná pouze o módní záležitost a zájem o ni brzy opadne (Angrave, Charlwood, Kirkpatrick, Lawrence, & Stuart, 2016). Důvodem pro kritiku HR analytiky je především nedostatek relevantní a akademické literatury. Dále to, že některé organizace preferují kvantitu dat před jejich kvalitou a relevancí, namísto kvalitního statistického modelování v datech zachycují náhodné jevy, což je způsobeno jejich nesprávným přístupem k analýze dat (Angrave, Charlwood, Kirkpatrick, Lawrence, & Stuart, 2016).

HR analytika nemůže splnit svůj účel, budou-li data analyzovaná jen proto, aby byla analyzovaná. Tomu lze předejít, pokud projekt začneme otázkou, proč dané analýzy chceme provádět (viz kapitola 2.1). Stejně tak nebezpečí hrozí při špatné interpretaci získaných výsledků. Zde je navíc nutné jednat v souladu s etickými principy a právními předpisy (viz kapitola 3.5; Guenole et al., 2017). Aby HR analytika přinesla společnostem kýžené výsledky, je třeba správně propojit akademický výzkum s praxí. Zatímco akademici se zabývají teorií, vyznačují se precizností a mají tendenci na problémy pohlížet izolovaně, tak aby je bylo možné testovat pomocí hypotéz, zaměstnanci a zaměstnavatelé řídicí podniky čelí složitým problémům, a proto potřebují komplexní řešení, která budou alespoň na přijatelné úrovni v praxi fungovat. Vzhledem ke komplexitě HR analytiky, která propojuje podnikání, psychologii a pokročilé statistické metody, hrozí její selhání tam, kde bude nedostatek takto všestranných odborníků (Guenole et al., 2017; Marler & Boudreau, 2017).

### **1.3 Dovednosti potřebné k realizaci HR analytických projektů**

Jak je zmíněno v předchozí kapitole, má-li HR analytický projekt uspět, je třeba kombinace komplexních dovedností, kterou je třeba brát v potaz např. při sestavování HR analytického týmu. Guenole a kol. (2017) uvádí šest dovedností, které vedou k úspěchu v rámci HR analytických projektů.

- Zprvce se jedná o znalosti týkající se podnikání – finanční gramotnost, organizační politika, povědomí o interním a externím pracovním prostředí. Tato skupina znalostí umožňuje danému týmu porozumět situacím, které v rámci podnikání nastávají, a efektivně se s nimi vypořádat.
- Zadruhé, důležitou roli hrají poradenské dovednosti, které zahrnují definici problému a hypotéz, projektový management, vývoj řešení, řízení změn a vedení zúčastněných stran. Tedy jedná se o schopnost poskytnout organizaci expertní posudek ohledně toho, jak vylepšit některé aspekty v rámci podnikání.
- Další skupina dovedností se týká lidských zdrojů. V týmu by měli být obsaženi jedinci s profesionální znalostí oblasti lidských zdrojů, kteří mají znalosti z nábory, odměňování a vzdělávání, kteří ví, jak spolu jednotlivé funkce v rámci lidských zdrojů souvisí, a mají znalost předpisů o ochraně soukromí.
- Čtvrtá skupina dovedností spadá do oblasti pracovní psychologie – týká se poznatků z oblastí pracovní motivace, postoje a výkonnosti zaměstnanců, pracovní diagnostiky, výběru prediktorů apod.

- Zapáté, v týmu by měli být odborníci na datovou vědu, kteří hrají klíčovou roli při hledání kauzálních mechanismů a možných řešení. Tato oblast kombinuje znalosti matematiky a statistiky, programování, databází a povědomí o možných datových zdrojích.
- Posledním předpokladem jsou komunikační dovednosti. Ty jsou stěžejní pro předání výsledků pověřeným osobám. Do této skupiny dovedností patří např. vizualizace zjištěných poznatků, prezentace výsledků a um přesvědčit posluchače.

## 1.4 Vývoj HR analytiky

Cílem této kapitoly je podchytit faktory související se vznikem HR analytiky. Dle dostupné, níže uvedené literatury např. (Rousseau & Barends, 2011; Ulrich & Dulebohn, 2015; Van der Lupen, 2018), se zdá, že mezi hlavní incentivy pro její vznik patří transformace účelu HR oddělení ve společnostech, rozvoj tzv. na důkazech založeného rozhodování v rámci HR a zrod datové vědy.

### Transformace HR oddělení

HR oddělení původně zastávala v organizacích spíše administrativní funkci, a to až do 80. let, kdy se dočkala transformace a začala v organizacích obsazovat stěžejní, strategickou roli (Ulrich & Dulebohn, 2015). Od té doby akademický výzkum přináší poznatky týkající se vlivu HR oddělení na obchodní výsledky společností. Dle dostupné literatury HR oddělení mohou pozitivně působit na obchodní výsledky organizace intervencemi, které zahrnují např. důkladný nábor a výběr zaměstnanců, odměňování, sledování výkonu a poskytování školení (Crook et al., 2011; Huselid, 1995; Jiang et al., 2012). Huselid (1995) jako jeden z průkopníků publikoval ve své práci, že prostřednictvím investic do HR aktivit je možné snížit odchodovost zaměstnanců a pozitivně působit na jejich produktivitu a také na obchodní výsledky společnosti. Další metaanalytická studie se zabývala činností HR oddělení zaměřenou na rozvoj zaměstnanců a snahu motivovat zaměstnance. Tyto dvě oblasti přímo souvisely s obchodními výsledky společnosti (Jiang et al., 2012). Aktivity HR oddělení mohou přinášet odlišné výsledky v závislosti na kontextu organizace a sektoru, ve kterém se daná společnost nachází (Paauwe & Farndale, 2017; Ulrich & Dulebohn, 2015). Stejně tak se liší jejich vliv u jednotlivých zaměstnanců v závislosti na jejich pozici – velký vliv má např. na zaměstnance, jichž práce

zahrnuje strategické plánování (Aguinis et al., 2010). Tyto rozdíly by HR oddělení při plánování své činnosti měla brát na zřetel.

S transformací HR oddělení vznikla potřeba propojit aktivity HR oddělení s obchodními výsledky společnosti. Zde vznikl prostor právě pro HR analytika. Dle van Vulpena (2018) jsou vedle HR aktivit (např. odměňování či nábor) důležité především jejich měřitelné výsledky (např. odchodovost nebo spokojenost zaměstnanců). Právě výsledky HR aktivit se dají propojit s obchodními výsledky společnosti (výnosy, tržní hodnota společnosti, kvalita produktů apod.) a HR analytika nám toto propojení umožňuje díky svým nástrojům, které umí mezi těmito oblastmi najít implicitní souvislosti.

### **Na důkazech založený přístup k rozhodování v rámci HR**

I přes existenci mnoha literárních zdrojů, které mluví ve prospěch některých praktik v rámci oddělení lidských zdrojů, v mnoha společnostech dochází k výzkumem a daty nepodloženému rozhodování, které může vyústit ke zbytečným finančním ztrátám (Rousseau & Barends, 2011). Příkladem jsou pohovory v rámci náboru zaměstnanců – ty nabízejí velkou flexibilitu, ale zároveň při nesprávné strukturovanosti mohou vyústit ve špatná a nákladná rozhodnutí. Přitom existují mnohé studie, které se zabývají např. vztahem mezi hodnocením při pohovoru a pozdějším pracovním výkonem a které by mohly být při náboru pro zaměstnavatele dobrým podkladem (Stevens, 2009). Rousseau a Barends (2011) uvádí příklad, kdy rozhodování při náboru kandidátů založené na testech s dobrými psychometrickými vlastnostmi přináší lepší výsledky než ty samé výsledky doplněné o úsudek manažera. Tento příklad má demonstrovat, že subjektivní, intuitivní úsudky mohou být nesprávné a že je nutné praktikovat v rámci HR na důkazech založené rozhodování. Při praktikování na důkazech založeného rozhodování v rámci HR by společnosti měly kombinovat čtyři oblasti: zaprvé výše zmíněnou kvalitní vědeckou literaturu s vhodným výzkumným designem, nejlépe pak metaanalytické studie; zadruhé validní HR metriky – např. míru fluktuace; zatřetí kritický úsudek jedince, který provádí daný projekt; a nakonec zhodnocení vlivu rozhodování na všechny zainteresované osoby (Rousseau & Barends, 2011).

HR analytika umožňuje poznatky z existujících výzkumů převést do praxe prostřednictvím analýzy dat získaných speciálně pro danou společnost. Díky tomu je možné identifikovat problémy, s kterými se daný podnik potýká, spolu s jejich souvislostmi a najít pro ně adekvátní řešení (Van der Laken, 2018).

## Datová věda

V současné digitální době s rostoucím množstvím dat se nejrůznější instituce, mezi kterými dominují obchodní společnosti, snaží data zužitkovat ve svůj prospěch, aby získaly na trhu kompetitivní výhodu (Provost & Fawcett, 2013). Tyto tendence stojí za vznikem nové disciplíny – datové vědy. Cílem datové vědy je získat z dat takové informace, které se dají zevšeobecnit a využít pro predikci do budoucna. Získané informace je třeba integrovat a interpretovat tak, aby měla smysl. To z datové vědy činí komplexní disciplínu kombinující např. nástroje statistiky, počítačové vědy, sociologie, ekonometrie nebo lingvistiky (Dhar, 2013). Příkladem oblastí, v rámci kterých se společnosti snaží svá rozhodování optimalizovat, je marketing (např. segmentace zákazníků, chování zákazníků), obchodní rozhodování (tzv. *business intelligence*), modelování finančního rizika, úvěrového rizika či detekce podvodů. Je přirozené, že se postupně principy datové vědy dostaly i do oddělení lidských zdrojů, kde se dá k optimalizaci procesů přistupovat analogicky jako v jiných, výše zmíněných oblastech (Provost & Fawcett, 2013).

Vznik HR analytiky tedy pravděpodobně nasedá na vývoj snah činit ve společnostech rozhodování týkající se zaměstnanců na základě důkazů, ale vedle toho také souvisí s rozvojem datové vědy a snahou vnímat lidský kapitál jako oblast, která silně souvisí se strategickým plánováním podniku. Tedy ačkoliv HR analytika představuje novou disciplínu, její vznik není nic nepředvídatelného. I přes zjevné souvislosti v literatuře vztah HR analytiky a výše zmíněnými oblastmi není přesně definován. Např. vztah mezi HR analytikou a na důkazech založeným přístupem k rozhodování v HR může být matoucí (Ulrich & Dulebohn, 2015). V praxi někdy mohou být obě oblasti obsaženy pod pojmem HR analytiky (nebo některou z jeho alternativ), čehož jsou důkazem případové studie uvedeny dále v této práci (viz kapitola 2.2.1), kde jsou oba přístupy k rozhodování kombinovány. Van der Laken (2018) uvádí, že dle jeho názoru se HR analytika od na důkazech založeného rozhodování v rámci HR odlišuje dvěma aspekty: HR analytika slouží rozdílnému účelu a využívá jiných statistických metod pro modelování. Zatímco cílem samotné statistiky je snaha vysvětlit minulost, datové predikce nám poskytují scénáře do budoucnosti (Dhar, 2013). Ukážeme-li si tento rozdíl na příkladu retence zaměstnanců, běžný akademický výzkum zkoumá faktory, které s odchodovostí zaměstnanců souvisí, zatímco HR analytika nám poskytuje aparát k predikci odchodovostí zaměstnanců do budoucna. A tedy, zatímco v akademickém výzkumu může dobře posloužit bodově biseriální korelace, v rámci HR analytiky bychom se přiklonili spíše k predikčním



statistickým technikám, tedy např. logistické regresi, analýze přežití nebo klasifikačním stromům (Speer, Dutta, Chen, & Trussell, 2019).

## 1.5 Případové studie

HR analytika nachází své využití ve všech fázích životního cyklu zaměstnance – upoutání pozornosti pracovníků (např. segmentace zaměstnanců dle zdroje nábory), jejich výběr (predikce úspěšných zaměstnanců nebo zaměstnanců, kteří dobře zapadnou do firemní kultury), nástup a adaptace, rozvoj a vzdělávání (identifikace klíčových dovedností na dané pozici), retence (predikce odchodovosti zaměstnanců, analýza faktorů, které s odchodovostí souvisí) a separace, přičemž umožňuje optimalizovat procesy, které jsou pro jednotlivé fáze charakteristické (van Vulpen, 2018). V následující části práce jsou uvedeny případové studie, které demonstrují využití HR analytiky v několika fázích životního cyklu zaměstnance.

### Upoutání pozornosti zaměstnanců

Prvním příkladem společnosti, která svá rozhodnutí týkající se zaměstnanců optimalizuje pomocí HR analytiky, je největší kanadská banka tzv. *The Royal Bank of Canada* (RBC), která pravidelně monitoruje postoje svých zaměstnanců prostřednictvím dotazníkových šetření a data dále propojuje s dalšími metrikami (Boudreau & Jesuthasan, 2011).

Jedna z jejich analýz zahrnovala segmentaci zaměstnanců dle jimi vyznávaných hodnot ve vztahu ke kanálu jejich nábory. Výsledkem bylo zjištění, že zaměstnanci oslovení v rámci univerzitních akcí si nejvíce zakládají na typu lidí, se kterými spolupracují, zaměstnanci získaní v rámci akvizice jiné společnosti byli zaměřeni především na možnosti kariérního postupu, nakonec ti zaměstnanci, které společnost původně oslovila, když pracovali v jiné společnosti, byli nejvíce zaměřeni na finanční ohodnocení. Tomuto zjištění RBC přizpůsobila své jednání při snaze inzerovat zaměstnání prostřednictvím jednotlivých kanálů – např. při snaze oslovit zaměstnance skrze univerzitní akce kladla důraz na možnost navázání interakcí se současnými zaměstnanci společnosti, oproti tomu při snaze nabrat zaměstnance pracující v jiných společnostech kladla důraz na finance a bonusy (Boudreau & Jesuthasan, 2011).

## **Nábor zaměstnanců**

Jednou ze společností, která se rozhoduje na základě HR analytiky je *Google*. Jedná se o technologickou společnost proslulou především svým eponymním internetovým prohlížečem („Google“, 2020). *Google* je známý pro své inovativní pracovní prostředí a účinné HR praktiky, což z něj dělá jednoho z nejžádanějších zaměstnavatelů na světě. *Google* prostřednictvím rozsáhlého využívání HR analytiky, tedy upřednostněním dat před pocity a intuicí, dosáhl znatelných obchodních výsledků (Shrivastava et al., 2018).

Příkladem oblasti, kde *Google* využívá HR analytiky, je oblast nábory zaměstnanců. Algoritmus, který *Google* používá k nábory kandidátů, předpovídá na základě osobnostních charakteristik (např. biografických údajů, postojů či vlastností) pravděpodobnost, že se jednotlivým kandidátům v zaměstnání dobře povede s ohledem na jejich výkon i soulad s firemní kulturou (Hansell, 2007; Shrivastava et al., 2018).

## **Rozvoj a vzdělávání**

Další z oblastí aplikace HR analytiky v *Googlu* je efektivní leadership. Cílem projektu týkajícího se leadershipu, který *Google* prováděl v rámci HR analytiky, byla identifikace a zhodnocení klíčových vlastností manažerů prostřednictvím kvalitativní analýzy komentářů od zaměstnancům v průzkumech s hodnocením, aby bylo možné zaměřit se na tyto vlastnosti v rámci dalšího rozvoje. Výsledkem byly následující praktiky: zmocňování týmu, a nikoliv praktikování detailní kontroly podřízených (tzv. *micromanagement*); projevování zájmů o jednotlivé členy týmu a jejich osobní pohodu; dobré komunikační schopnosti jak pro naslouchání, tak sdílení informací; dobré koučovací dovednosti; orientace na výsledky; snaha napomáhat jednotlivým členům týmu v kariérním rozvoji; jasná vize pro tým a klíčové technické dovednosti k vedení týmu (Shrivastava et al., 2018).

## **Rozvoj a vzdělávání, retence**

Jiná společnost, jejíž cílem bylo zaměřit se prostřednictvím HR analytiky na rozvoj a vzdělávání zaměstnanců, byl řetězec restaurací s rychlým občerstvením. Cílem společnosti bylo prostřednictvím svých zaměstnanců v první linii zvýšit spokojenost zákazníků. Vedení společnosti se domnívalo, že možným nástrojem pro zvýšení spokojenosti zákazníků by mohlo být větší porozumění zaměstnancům (Arellano et al., 2017).

Projekt byl zaměřen na tři oblasti. Zaprvé na osobní charakteristiky zaměstnanců, které podle vedení mohly stát za variabilitou v jejich výkonu (data o osobních charakteristikách byla dosbírána pomocí gamifikačních technik). Zadruhé na kvalitu managementu, jelikož cílem bylo odhalit vliv leadershipu na pracovníky v první linii. Nakonec bylo zaznamenáváno chování a interakce zaměstnanců – senzory zachycovaly míru jejich pohybu po restauraci, intonaci jejich mluvy a podíl mezi aktivitou a pasivitou v konverzaci se zákazníky (Arellano et al., 2017).

Výsledky odhalily souvislost výkonu zaprvé s osobnostními rysy (nejlépe si vedli zaměstnanci, kteří se dokázali plně soustředit na práci; ti si vedli lépe než komunikativní a přátelští zaměstnanci) a zadruhé se vzdáleností bydliště od zaměstnání. Dále u manažerů výkon souvisel s délkou jejich směny (čím déle byli na směně, tím horší výkon měli jejich podřízení), tréninkem dovedností a chováním (nejlépe si vedli manažeři, kteří zmocňovali své zaměstnance, působili na ně inspirativně, upozorňovali na jejich úspěchy atd.). Po implementaci změn zaměřených na rozvoj a vzdělávání zaměstnanců se spokojenost zákazníků zvýšila o více než 100 %, zvýšila se celková rychlost obsluhy, snížila se odchodovost zaměstnanců a tržby vzrostly o 5 % (Arellano et al., 2017).

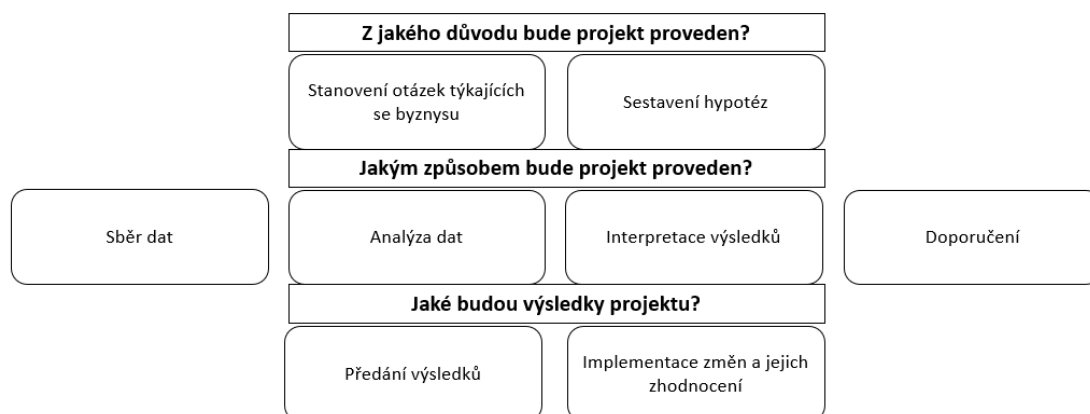
## 2. Principy HR analytiky

### 2.1 Metodologie

Pro HR analytiku vzhledem k tomu, že se jedná o obor relativně nový a dosud se vyvíjející, není charakteristické užívání standardní metodologie. Přesto se v současné době objevují snahy formulovat metodologický rámec, který HR analytikům pomáhá zvyšovat pravděpodobnost úspěchu projektů, které realizují (Guenole et al., 2017). Jedním z nich je tzv. osmikrokový model smysluplné analytiky (*The Eight Step Model for Purposeful Analytics*), jehož autoři jsou Guenole a kol. (2017).

#### Osmikrokový model smysluplné analytiky

Tento postup sestává ze tří částí, které lze dále dělit celkově na dalších osm (viz Obrázek 1).



**Obrázek 1:** Osmikrokový model smysluplné analytiky, zpracováno dle Guenole a kol. (2017)

V první části je stěžejní stanovit si, z jakého důvodu je projekt prováděn. Tedy je třeba stanovit si otázky týkající se problému, kterému společnost čelí, a formulovat příslušné hypotézy. DiClaudio (2019) uvádí příklad, kdy je cílem společnosti zvýšit prodej obchodníků. Pak by si měla klást např. otázky ohledně toho, kteří zaměstnanci jsou nejlepší obchodníci a jakým způsobem pracují. Bez toho, aniž bychom znali účel analýz, bude projekt jen těžko přínosný. V této části je důležité promyslet si, jak HR procesy mohou souviset s obchodními výsledky. Dále zvážit to, zda je projekt pro společnost opravdu potřebný. Pokud je znám účel a přínos projektu, dají se pak mnohem snadněji získat potřebné finance od vedení společnosti. Kladení správných otázek a dobrá formulace hypotéz je

důležitým předpokladem pro sběr relevantních dat, a navíc zabraňuje tomu, abychom brali vážně vztahy, které byly v danech zachyceny jen v důsledku náhodného šumu. Jako v každém výzkumu musí hypotézy být v testovatelné podobě a podloženy relevantní a kvalitní literaturou z posledních let (Guenole et al., 2017).

Druhá část zahrnuje sběr dat, jejich čištění, samotné testování hypotéz, interpretaci výsledků a formulaci doporučení, která na výsledky nasedají. V rámci sběru dat na základě stanovených hypotéz je možné využít již existujících dat nebo dosbírat data nová. Při práci s daty je nezbytné jednat v souladu s obecným nařízením o ochraně osobních údajů (GDPR<sup>1</sup>). Veškerá data by měla obsahovat jedinečný klíč, na základě kterého je možné je vzájemně propojit. V další části je nutné provést samotnou analýzu dat, samozřejmě za využití vhodných statistických postupů. V opačném případě bychom ohrozili platnost výsledků celého projektu. Předposledním krokem této části je verbalizace a vizualizace závěrů plynoucích z výsledků. Díky tomu je možné zajistit plné pochopení a předejít možnému zkreslení výsledků na straně společnosti, která projekt poptala. V poslední fázi této části je nezbytné společnosti předat doporučení, protože právě doporučení jsou cestou k lepšímu výkonu společnosti, a tedy jádrem celého projektu (Guenole et al., 2017).

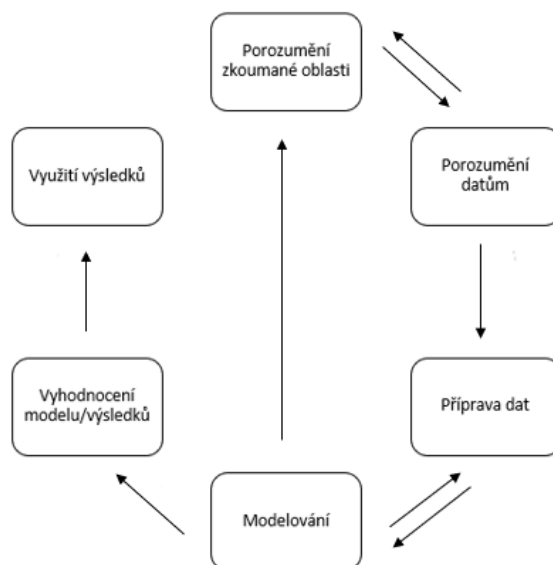
Třetí etapa zahrnuje změny a následnou evaluaci implementovaných změn v organizaci. Organizace by měla přijmout doporučení, na základě kterých učiní rozhodnutí. V případě komplexních rozhodnutí, která ovlivní velké množství zaměstnanců, není od věci zapojit konzultanty z oblasti řízení změn (Guenole et al., 2017).

## **CRISP-DM**

Historicky starším metodologickým rámcem je tzv. CRISP-DM (Wirth & Hipp, 2000). CRISP-DM označuje hierarchický postup při procesu modelování, který sestává ze šesti částí, přičemž gradient postupu mezi jednotlivými částmi je od obecného ke konkrétnímu. Tento metodologický rámec se dodnes v rámci datových projektů v různých modifikacích používá (Wirth & Hipp, 2000). Jednotlivé fáze seřazené dle jejich logické návaznosti jsou uvedeny níže spolu s jejich stručným popisem podle autorů Wirtha a Hippa (2000). Obrázek 2 znázorňuje jakými způsoby lze mezi jednotlivými fázemi přecházet:

---

<sup>1</sup> „Nařízení Evropského parlamentu a Rady (EU) č. 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES“ (Wikipedie, 2020)



**Obrázek 2:** CRISP-DM, zpracováno dle Wirth a Hipp (2000)

- Porozumění zkoumané oblasti: snaha porozumět cílům projektu a převedení problému do testovatelné podoby
- Porozumění datům: sběr/získání dat, zhodnocení jejich kvality
- Příprava dat: příprava finálního datového souboru zahrnující např. čištění dat nebo transformaci proměnných
- Modelování: výběr a aplikace samotných modelovacích technik, kalibrace parametrů
- Vyhodnocení modelu/výsledků: vyhodnocení získaných výsledků, zhodnocení modelu/modelů
- Využití výsledků: definice způsobů, jakými koncový uživatel může využít získané výsledky

## 2.2 Zdroje dat

Výhodou aplikovaného výzkumu je oproti akademickému výzkumu větší dostupnost dat. Mezi běžné datové zdroje pro analytické projekty týkající se zaměstnanců patří nejrůznější zaměstnanecké průzkumy (hodnocení z 360° zpětné vazby, průzkumy spokojenosti a angažovanosti, výstupní pohovory), objektivní hodnocení výkonu (např. počet prodaných položek, chybovost, tržby), psychodiagnostické testy (osobnostní rysy), data nasbíraná během pohovoru (hodnocení, motivační dopis, CV), systémy poskytující informace o charakteristikách pozice (O\*NET), personální informační systémy (obsahující

demografické údaje, dobu v zaměstnání, informace o povýšení, výše platu, změny platu, benefity, školení, absence, pozdní příchody, vzdálenost zaměstnání od bydliště; Speer, Dutta, Chen, & Trussell, 2019).

Kromě toho je možné využít i dat nad rámec běžných datových zdrojů jako jsou průzkumy spokojenosti nebo hodnocení výkonu. Hodí se jakákoliv data, která nám mohou přiblížit sledovanou oblast, tedy např. informace z kalendáře, čerpání dovolené nebo informace týkající se služebních cest (DiClaudio, 2019). Existují i kontroverznější zdroje dat jako zaměstnancovy stopy činností jako elektronická komunikace nebo např. klíčová slova vyhledávaná na Googlu. Ač tato data mohou být bohatým zdrojem informací, vždy je třeba brát v potaz právní předpisy a etickou stránku věci (Bodie et al., 2016), více viz kapitola 3.5. Získání všech dat tak v důsledku ochrany osobnostních údajů a interní politiky dané společnosti může představovat nemalé komplikace (Speer, Dutta, Chen, & Trussell, 2019).

Co se týče podoby dat, dají se zužtkovat jak strukturovaná data (např. v podobě demografických údajů), tak data nestrukturovaná (např. v podobě komentářů v hodnocení zaměstnance nebo údajů z píchacích hodin v docházkovém systému; DiClaudio, 2019).

Samozřejmě ne vždy jsou k dispozici všechna potřebná data. V takovém případě se nabízí možnost si některé proměnné ze stávajících dat vytvořit, případně potřebná data dosbírat (Speer, Dutta, Chen, & Trussell, 2019).

Přestože společnosti sbírají velká množství dat, jedná se o citlivé informace, které nemohou být volně distribuovány po společnosti (resp. mimo společnost). Tato skutečnost do jisté míry omezuje HR analytické projekty, jelikož si HR oddělení přístup k datům musí vyjednat a po celou dobu jsou povinna zacházet s daty v souladu s právní legislativou (Guenole et al., 2017).

### **3. Řešení problému fluktuace zaměstnanců s pomocí HR analytiky**

Tato kapitola je zaměřena na podrobný popis procesu aplikace HR analytické metodologie na vybraný personální problém. Vzhledem k zaměření práce je řešeným problémem fluktuace zaměstnanců. Téma fluktuace je dále v následující kapitole (viz kapitola 4) podrobně popsáno z hlediska pracovní a organizační psychologie.

Jak již bylo zmíněno v předchozí kapitole (viz kapitola 2), HR analytické projekty by se měly opírat o některý z metodologických rámců, za účelem zvýšení pravděpodobnosti jejich úspěchu. Pro popsání jednotlivých fází v rámci modelování fluktuace byl vybrán CRISP-DM, který sestává z následujících kroků: porozumění zkoumané oblasti, porozumění datům, příprava dat, modelování, vyhodnocení výsledků a využití výsledků (Wirth & Hipp, 2000).

#### **3.1 Porozumění zkoumané oblasti a datům**

Prvním krokem v rámci procesu modelování odchodovosti zaměstnanců je nadefinování problému, kterému daná společnost čelí. Respektive, k analýze fluktuace zaměstnanců by mělo dojít ve chvíli, kdy společnost vyslovila např., že se potýká s nadměrnou fluktuací či že by ráda zabránila odchodu talentů. Kromě toho je vzhledem k dalšímu postupu analýzy důležité, aby společnost společně s daným analytikem (resp. analytiky) nadefinovali, co by mělo být výsledkem projektu (Guenole et al., 2017).

Dalším krokem je určení typu fluktuace, kterým se v rámci analýzy chceme zabývat. Tento krok ovlivňuje jak volbu prediktorů, tak jedinců, kteří budou do modelu zahrnuti (např. při modelování dobrovolné fluktuace můžeme z modelu vyloučit osoby, které odchází do důchodu; Speer, Dutta, Chen, & Trussell, 2019).

Dále je důležité stanovit si časový rámeček, tedy období, po které je nutné sbírat data týkající se odchodovosti zaměstnanců, stejně tak jako časový úsek, na který do budoucna budeme odchodovost zaměstnanců předpovídat. Zde je důležité zvážit, k jaké vysoké míře fluktuace v organizaci dochází, čím více fluktuace zaměstnanců organizace zakouší, tím kratší stačí perioda, po kterou jsme data týkající se minulosti sbírali. Kromě toho by perioda, na kterou fluktuaci predikujeme, neměla být delší než doba, po kterou jsme data sbírali. Nedodržení těchto podmínek by mohlo vést např. k tomu, že v organizaci, kde téměř



nedochází k fluktuaci, budeme jen pouhou náhodou poměrně přesně predikovat nízkou odchodovost zaměstnanců (Speer, Dutta, Chen, & Trussell, 2019).

Čtvrtým krokem je zvážení toho, zda výstupem budou individuální či skupinové výstupy. Při modelování skupinové odchodovosti se výstupy mohou týkat např. jednotlivých týmů, oddělení či poboček. Jednotlivé skóry jsou zprůměrovány a výstupem je pouze skór skupinový. Při výběru individuální či skupinových výstupů je důležité zvážit výhody a nevýhody, které jednotlivé přístupy přináší. Reportování individuálních např. pravděpodobností odchodu může být riskantní tam, kde model vykazuje menší zlepšení oproti nulovému (vysvětluje malé procento variability) a dále jsou-li v modelu obsaženy citlivé proměnné týkající se např. pohlaví nebo věku. Pokud se společnost rozhodne pro individuální výstupy, mělo by být velmi dobře ošetřeno, jak bude s výsledky zacházeno. Skupinové výstupy mohou být z výše uvedených důvodů bezpečnější variantou. Ovšem u skupinových výstupů je třeba počítat s chybou způsobenou v důsledku průměrování, což ještě více zkresluje získané výsledky (Speer, Dutta, Chen, & Trussell, 2019).

V neposlední řadě je nutné identifikovat proměnné, na základě kterých budeme odchodovost zaměstnanců předpovídat. Výběr proměnných se odvíjí od hypotéz, které se týkají možných příčin odchodovosti a které jsme si předem formulovali (Guenole et al., 2017; Speer, Dutta, Chen, & Trussell, 2019). Výčet těchto prediktorů je velmi podrobně popsán níže (viz kapitola 4.4.1).

### **3.2 Příprava dat**

Stejně jako v jiných analýzách, i v rámci modelování fluktuace zaměstnanců je třeba podstatnou část času strávit přípravou dat. Čištění dat je nezbytným krokem pro zajištění validity a zobecnitelnosti výsledků (Osborne, 2013). V rámci této etapy se výzkumník seznamuje s daty, upravuje jejich strukturu, aby byla připravena pro pozdější manipulaci (pro modelování i vizualizaci), vypořádává se s odlehlými a chybějícími hodnotami nebo hodnotami, které v daném kontextu nedávají smysl (Wickham, 2014). Co se týče chybějících a extrémních dat, výzkumník by měl zvážit proč se tomu tak stalo (např. prostřednictvím srovnání skupiny jedinců, kteří odpověděli, a jedinců, kteří neodpověděli) a podle toho se se situací vhodně vypořádat (např. použitím robustnější statistické metody, rozdělením datové matice na menší podsoubory, useknutím průměru, vyřazením chybějících jedinců či odhadnutím chybějících dat). Podrobnější postup uvádí Osborne (2013).

### 3.3 Analýza dat

#### Bodově biseriální korelace

Mnoho studií, které se zaměřují na vztah mezi fluktuací a dalšími faktory, využívají bodově biseriální korelace (Allen, Hancock, Vardaman, & Mckee, 2014). Prostřednictvím příslušného korelačního koeficientů je možné zachytit vztah mezi fluktuací znázorněnou pomocí dichotomické proměnné (např. 1 – odešel/odešla, 0 – neodešel/neodešla) a jiných kvantitativních proměnných (např. věk, svědomitost apod.). Při použití této metody je důležité stanovit si pro dichotomickou proměnnou časový úsek, pro který je daný zaměstnanec považován za „fluktuujícího“ a pro který již nikoliv (např. 1 – odešel/odešla do 2 let, 0 – neodešel/neodešla do 2 let). Bodově biseriální korelace ovšem není vhodnou statistickou metodou pro predikci fluktuace do budoucnosti (Speer, Dutta, Chen, & Trussell, 2019).

#### Binární logistická regresní analýza

Logistická regrese bývá častou metodou pro modelování fluktuace a její predikci do budoucnosti (Allen, Hancock, Vardaman, & Mckee, 2014). Z tohoto důvodu a stejně tak proto, že je využita v rámci zpracování dat v praktické části této diplomové práce, jí bude v dalších odstavcích věnováno více pozornosti než jiným metodám.

Prostřednictvím regresních modelů je možné zachytit vztah mezi jednou závislou proměnnou a vysvětlujícími proměnnými. Binární logistická regrese je speciálním případem zobecněných lineárních modelů. V jejím případě je závislá proměnná (stejně jako jedna z proměnných v případě bodově biseriální korelace) opět binární, nabývající hodnot 1, pokud výsledek nastane, a 0, pokud výsledek nenastane. Binární logistická regrese nám umožňuje vypočítat pravděpodobnost, s jakou daný jev nastane nebo nenastane a udává váhy jednotlivých prediktorů (Hosmer, Lemeshow, & Sturdivant, 2013). Z tohoto důvodu se jedná o metodu vhodnou pro modelování fluktuace.

Na rozdíl od lineární regrese, která je založena na metodě nejmenších čtverců, logistická regrese se opírá o metodu maximální věrohodnosti. Pro zhodnocení kvality celkové modelu využívá např. několika typů indexu determinace např. McFaddenovo pseudo  $R^2$ , které se stejně tak jako u lineární regrese pohybuje v intervalu  $<0,1>$ . Pro zhodnocení důležitosti jednotlivých prediktorů logistická regrese využívá Z-skóřů (na rozdíl od  $t$ -testu u

lineární regrese) nebo tzv. věrohodnostního poměru (tzv. *likelihood ratio*). Pro výběr prediktorů volíme např. krokové metody, tedy metodu vzestupného nebo sestupného výběru (Field, Miles, & Field, 2012). Kvalitu modelu bychom dále měli hodnotit pomocí křížové validace a specifity, senzitivity, přesnosti, preciznosti a ROC křivky (viz dále).

Při modelování fluktuace prostřednictvím logistické regrese je, stejně tak jako u bodově biseriální korelace, třeba stanovit si pro dichotomickou proměnnou časový úsek, po který je zaměstnanec ještě považován za „fluuktujícího“, a kdy již nikoliv. Máme-li informace o příčině odchodu daných zaměstnanců, je možné datovou matici podle typů fluktuace rozdělit na menší vzorky a pro každý z nich udělat samostatnou logistickou regresi. Tento postup by nám měl zajistit přesnější výsledky (Speer, Dutta, Chen, & Trussell, 2019).

### **Analýza přežití**

Analýza přežití zahrnuje soubor technik, které umožňují odhadovat, kolik uběhne času, než dojde k nějaké události – např. relapsu onemocnění, úmrtí či rozvodu (Lee & Wang, 2003). V případě modelování fluktuace zaměstnanců odhadujeme, kdy dojde k odchodu daného zaměstnance. Tedy zatímco logistická regrese uvádí pravděpodobnosti odchodovosti pro jednotlivé zaměstnance, analýza přežití poskytuje časové odhady, kdy k odchodu dojde (Speer, Dutta, Chen, & Trussell, 2019). Analýza přežití pro predikce dané skutečnosti může zahrnout vedle času i další vysvětlující proměnné a stejně jako u logistické regrese poskytuje váhy k jednotlivým prediktorům (Lee & Wang, 2003). Výhodou analýzy přežití je to, že se, na rozdíl od logistické regrese, lépe vyrovná s cenzorovanými daty. Analýza přežití zahrnuje více přístupů k modelování. Jedním z nich je tzv. *Coxův model proporcionálního hazardu*, který se k modelování fluktuace často využívá (Speer, Dutta, Chen, & Trussell, 2019).

### **Analýza časových řad**

Časová řada je pojem označující soubor pozorování, která jsou seřazena z hlediska času (Hamilton, 1994). Tato technika se hodí v případě, že nemáme k dispozici model, který by vysvětloval vztahy mezi prediktory a predikující proměnnou. Analýza časových řad umožňuje modelovat fluktuaci bez externích prediktorů za užití trendové a sezónní složky (univariační metody bez externích prediktorů), s externími prediktory (např. pomocí dynamické regrese a dekompozice) či za užití multivariačních a nelineárních metod (Zhu et al., 2016). Prediktivní síla výsledného modelu by měla být zhodnocena pomocí ukazatelů

jako  $R^2$  (procento vysvětlené variance),  $MSE$  (střední kvadratická chyba odhadu) a dalších ukazatelů, které se týkají středních chyb –  $ME, MAE, MPE$  apod. (Hamilton, 1994).

### **Klasifikační a regresní stromy**

Klasifikační a regresní stromy klasifikují osoby na základě proměnných. Zatímco klasifikační stromy se hodí pro situace, kde má závislá proměnná konečně mnoho neseřazených hodnot, regresní stromy pracují se spojitou nebo diskrétní (seřaditelnou) závislou proměnnou (Loh, 2011). Proměnné na sebe hierarchicky navazují, resp. dělí jednotlivé případy v datové matici na menší a menší podskupiny. Např. v případě fluktuace bychom zaměstnance mohli nejdříve rozdělit podle výše platu, dále podle spokojenosti, posléze podle záměrů odejít atd. až se dostaneme k výslednému uzlu, kde jsou zaměstnanci dělení např. podle rizika odchodu. Výsledná klasifikace může být jak do kategorií (např. zaměstnanec odejde/neodejde), tak do numerických intervalů (např. zaměstnanec odejde za 1-12/13-24 měsíců; Speer, Dutta, Chen, & Trussell, 2019).

Výhodou této metody je, že se vypořádá s nelineárními vztahy bez toho, abychom nelineární vztah předem specifikovali. Dále je robustní vůči chybějícím a odlehlým hodnotám. Model také relativně rychle natrénujeme, což nám umožňuje pracovat s velkými datovými soubory. Kromě toho se jedná o intuitivní statistickou metodu, která se dá snadno vizualizovat a interpretovat, díky čemuž je pro laiky dobře srozumitelná, což může zvýšit efektivitu projektů v rámci HR analytiky (Speer, Dutta, Chen, & Trussell, 2019). Naopak kontraindikace klasifikačních a regresních stromů jsou velké stromy, které se dají těžko interpretovat. Hrozí riziko zachycení náhodného šumu (tzv. *overfittingu*) a vysoké variance, které negativně působí na možnost zobecnění modelu.

### **Naivní bayesův klasifikátor (*Naive bayes*)**

Naivní bayesův klasifikátor označuje klasifikační techniku založenou na bayesově větě o podmíněné pravděpodobnosti. V případě predikování fluktuace zaměstnanců nám algoritmus poskytuje pravděpodobnosti zařazení zaměstnance do skupin (např. odejde/neodejde do jednoho roku) podobně jako tomu je u logistické regrese (Ajit, 2016; Devi & Umadevi, 2018).

Předpokladem pro použití tohoto algoritmu je nezávislost mezi jednotlivými prediktory. Při dodržení tohoto předpokladu modelu naivní bayesův klasifikátor produkuje lepší výsledky než jiné modely jako logistická regrese. Kromě toho je algoritmus rychlý a

k natrénování modelu stačí menší množství pozorování. Bohužel předpoklad nezávislosti mezi prediktory je v běžné praxi těžko splnitelný (Ajit, 2016; Devi & Umadevi, 2018).

### **Algoritmus *k*-nejbližších sousedů (*k-nearest neighbors*)**

Další klasifikační a regresní technikou je algoritmus *k*-nejbližších sousedů (*k-nearest neighbor*), která se opírá o základní myšlenku podobnosti mezi elementy, které jsou si blízko. V případě klasifikace, která je relevantní pro predikci fluktuace, algoritmus klasifikuje data do skupin (tzv. *clusterů*) na základě třídy jejich nejbližšího souseda. Tedy výsledkem predikce může být zařazení do skupiny zaměstnanců, který např. jsou/nejsou ohroženi odchodem (Ajit, 2016; Hastie et al., 2009).

Algoritmus sestává ze dvou částí: nejprve jsou určeny sousední body a posléze dochází k identifikaci třídy sousedů. Sousedé mohou být určeni např. na základě euklidovské vzdálenosti. *K* v názvu označuje množství skupin a může být určeno pomocí několika technik – jednou z nich je sutinový graf (tzv. *elbow method*). Výhodou je srozumitelnost a tvárnost algoritmu – může být použit jak pro regresi, tak klasifikaci. Naopak nevýhodná je jeho nižší rychlost při větším množství prediktorů (Ajit, 2016, Hastie et al., 2009).

### **Metoda podpůrných vektorů (*Support Vector Machine-SVM*)**

Metoda podpůrných vektorů je algoritmus, který nám umožňuje řešit lineární a nelineární klasifikační problémy. Jedním z principů SVM je převedení vstupního problému do nelineárního, kde je možné lineárně rozdělit pozorování do skupin prostřednictvím nadroviny, která maximalizuje vzdálenosti bodů mezi jednotlivými skupinami. Tedy prvky náležící do rozdílných skupin jsou co nejoptimálněji rozděleny (Harmady, 2015). Cílem predikce při užití SVM je opět klasifikace zaměstnanců do skupin podle toho, zda u daného jedince předpovídáme či nepředpovídáme odchod z organizace (např. v určitém časovém horizontu). Indikace pro použití SVM jsou jasná hranice mezi skupinami, naopak při překrytí skupin náhodný šum zkresluje výsledky. Další nevýhodou je, že algoritmus přímo neposkytuje pravděpodobnosti zařazení do jednotlivých skupin (Hastie et al., 2009).

### **Náhodné lesy (*Random Forests*)**

Náhodné lesy jsou koncepčně podobné klasifikačním a regresním stromům. Ovšem díky kombinaci mnoha stromů jsou robustnější vůči zachycení náhodného šumu, který je zmíněn výše jakožto primární kontraindikace rozhodovacích a regresních stromů. Další

výhodou je, že náhodné lesy poskytují kvalitní odhady o síle jednotlivých prediktorů. Nevýhodou oproti stromům je náročnější interpretace. Co se týče mechanismu, který stojí za kombinováním jednotlivých stromů do lesů, vybíráme náhodné vzorky s vrácením, tzv. *bagging* nebo *bootstrapping* (*bootstrapping* navíc přiděluje váhy tak, aby vzorky s dříve nepřesnými odhady měly větší váhu), na kterých modelujeme jednotlivé stromy a dále u regrese zprůměrujeme výsledky nebo u klasifikačních stromů vybíráme modus predikované třídy (Hastie et al., 2009).

### ***Extreme Gradient Boosting (XGBoost)***

*XGBoost* je další technikou strojového učení, která umožňuje řešit regresní a klasifikační problémy a která je založena na kombinaci více modelů. Jedná se o pokročilou verzi *gradient boostingu*, která si klade za cíl zvýšit rychlost a efektivitu modelu. Principem těchto algoritmů je *boosting*. Individuální modely, které v rámci *gradient boostingu* (resp. *XGBoostingu*) kombinujeme, mohou opět být rozhodovací a regresní stromy, tedy jedná se o paralelu náhodných lesů. Díky *XGBoostingu* obvykle dosahujeme vyšší efektivity a přesnosti modelu (Elsinghorst, 2018).

### **Umělé neuronové sítě**

Další technikou pro prediktivní modelování jsou neuronové sítě. Tato metoda je s to zachytit komplexní, nelineární vztahy mezi pozorovanými proměnnými. Neuronové sítě ze své podstaty nejsou zcela transparentní. Při tvorbě prediktivního modelu, je důležitým faktorem hledání kompromisu mezi zkreslením a rozptylem (tzv. *bias-variance trade off*). Jak název napovídá, při modelování bychom mezi zkreslením a rozptylem měli najít rovnováhu (Christian & Griffiths, 2017). Zatímco pod zkreslením chápeme rozdíl mezi skutečnou a predikovanou hodnotou, rozptyl označuje, jak moc se liší predikce mezi daty, na kterých byl model natrénován, a daty, na kterých je model testován. Jednodušší modely (jako např. lineární regrese) mohou vést k vyššímu zkreslení (tzv. *underfitting modelu*), ale dávají v čase konzistentní odhady. Oproti tomu komplexnější modely (např. neuronové sítě) bývají přesnější (s nižším zkreslením), ale v důsledku zachycení náhodného šumu (tzv. *overfitting modelu*) nemusí být konzistentní při generalizaci modelu na testovací datové soubory. Vzhledem k netransparentnosti neuronových sítí a nižší generalizaci modelu na nová data tato metoda není příliš vhodným nástrojem pro predikci v rámci HR analytiky (Briscoe & Feldman, 2011).

### 3.3 Křížová (*cross*) validace

Z důvodu, abychom zabránili zachycení souvislostí v datech jen v důsledku náhody, je nutností model křížově validovat. Postupujeme tak, že původní vzorek rozdělíme na tři menší (např. v poměru 6:2:2) – trénovací, validační a testovací. Trénovací a validační vzorek slouží k výběru prediktorů a odhadu parametrů, na testovacím vzorku ověřuje, jak bude model fungovat na neznámých datech. Speciálním typem křížové validace je tzv. *k-násobná křížová validace*, v rámci které jsou data rozdělena více způsoby (konkrétně *k*-krát) a tolikrát se proces opakuje (Speer, Dutta, Chen, & Trussell, 2019).

### 3.4 Evaluace modelu

Primární metodou pro zhodnocení klasifikačního modelu je tzv. matice záměn nebo konfusní matice (*confusion matrix*) v psychologii také známá pod názvem kontingenční tabulka. Matice záměn porovnává skutečnou klasifikaci pozorování s predikovanou klasifikací. Správně zařazená pozorování jsou tzv. skutečně pozitivní/negativní (*true positive/negative*) a špatně klasifikovaných pozorování tzv. falešně pozitivní/negativní (*false positive/negative*), přičemž správně zařazená pozorování jsou v matici umístěny na diagonále (Santra & Josephine Christy, 2012). Pro ilustraci matice záměn slouží Tabulka 1.

Predikce/Skutečnost	Odešel	Neodešel
Odešel	345	12
Neodešel	12	345

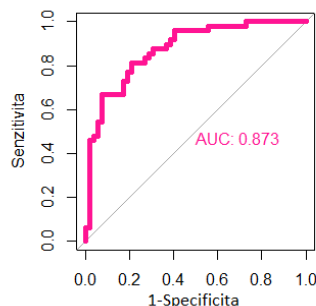
**Tabulka 1:** Matice záměn (vlastní zpracování)

Pro další zhodnocení kvality modelu můžeme využít několika statistik, které jsou od matice záměn odvozeny, např. senzitivita, specifická, přesnost a preciznost. K výpočtu těchto charakteristik můžeme přistoupit potom, co byl model po nějaký časový úsek používán nebo je můžeme spočítat na testovacím vzorku. Pokud tyto charakteristiky aplikujeme přímo na model fluktuace, hodnotíme, zda zaměstnanci, které model označit za potenciálně ohrožené odchodem, skutečně odešli, a naopak zda pracovníci, kteří dle modelu vykazují nízké riziko odchodu, neodešli. Senzitivitu a specifitu můžeme nadefinovat pomocí podmíněné pravděpodobnosti. Senzitivita testu v tomto případě označuje pravděpodobnost, že zaměstnanec byl označen za rizikového, pokud skutečně odešel [TP/(TP+FN)]. Naopak specifitu chápeme jako pravděpodobnost, že byl označen jako

nerizikový, pokud ze zaměstnání neodešel  $[TN/(FP+TN)]$ . Vysoká senzitivita je důležitá tam, kde je prioritní zachytit většinu skutečně pozitivních případů (příkladem je letištní kontrola, kde je důležité identifikovat veškeré předměty, jejichž přeprava by mohla představovat potenciální riziko, proto budou zaznamenány i předměty, které nebezpečné nejsou – např. lžička). Oproti tomu specifita je důležitá tam, kde chceme správně identifikovat co nejvíce skutečně negativních případů – příkladem může být situace, kdy je důležité správně identifikovat jedince bez nemoci – např. kvůli drastické léčbě. Co se týče přesnosti a preciznosti, přesnost určuje, kolik predikcí bylo správně zařazeno  $[(TP+TN)/(TP+FP+TN+TP)]$ , proti tomu preciznost určuje, kolik z našich správných predikcí bylo skutečně pozitivních – tedy kolik ze správně zařazených zaměstnanců odešlo  $[TP/(TP+FP)]$ . Tento ukazatel je v praxi důležitý např. tam, kde dochází k nízké míře fluktuace, abychom zbytečně nealokovali finanční zdroje směrem k zaměstnancům, kteří stejně odejít nechtěli (Speer, Dutta, Chen, & Trussell, 2019).

Ze specifity a senzitivity vychází ROC (tzv. *Receiver Operating Characteristic*) křivka a plocha pod ní zvaná AUC (tzv. *Area Under Curve*), které jsou dalším nástrojem pro hodnocení kvality modelu. Pomáhají nám určit pravděpodobnostní hranici, od které budou zaměstnanci označeni že odejdou nebo neodejdou. Graficky shrnují informace o poměru senzitivity a specifity při různých hranicích. Kvalitní binární klasifikátor skóruje dobře jak pro skutečně pozitivní, tak pro skutečně negativní. ROC graf nám pomáhá určit optimální hranici podle toho, jakou míru senzitivity a specifity od modelu vyžadujeme (Vuk & Curk, 2006). Hosmer a kol. (2013) uvádí, jakým způsobem je možné AUC interpretovat – hodnoty okolo čísla 0,5 naznačují, že dělení je zcela náhodné (v takovém případě by se ROC pohybovala okolo diagonály v grafu), čísla mezi 0,7 a 0,8 poukazují na akceptovatelnou kvalitu diskriminace modelu, jakákoliv hodnota nad 0,8 je považovaná za excelentní. AUC nám rovněž umožňuje porovnat jednotlivé modely (např. budeme-li modelovat fluktuaci jak pomocí logistické regrese, tak klasifikačního stromu), abychom mohli zvolit ten přesnější. Volíme model s vyšší AUC (Speer, Dutta, Chen, & Trussell, 2019). Příklad křivky ROC a plochy pod ní je znázorněn níže (viz Graf 2).





**Graf 2:** ROC křivka (vlastní zpracování)

Podobným grafem pro zhodnocení výkonnosti klasifikačního modelu jsou tzv. *gain* a *lift chart*, které spočítáme jako poměr mezi výsledky získanými s modelem a bez modelu. Metodu více popisují např. autoři Vuk a Curk (2006).

### 3.5 Využití výsledků

Samotnou analýzou dat a získáním výsledků proces nekončí. Výsledek je třeba prezentovat zadavateli projektu (např. pracovníkům na různých úrovních managementu) ve srozumitelné podobě (Wirth & Hipp, 2000). Cílem HR analytiky totiž není pouze potvrzení existence vztahu mezi sledovanými proměnnými, ale také získání informací o tom, jak se ve vztahu k obchodním výsledkům v budoucnu rozhodovat. Proto by prezentace výsledků měla být doplněna o relevantní doporučení (Rasmussen & Ulrich, 2015).

Prezentace výsledků může probíhat ve formě příběhu (tzv. *storytelling*). Neměla by obsahovat zbytečné množství technických pojmů, ale spíše prvky, které působí na emoce posluchače (např. zjištění vztáhnout k fiktivní postavě, se kterou se posluchač může identifikovat). Prezentace by měla posluchače zasvětit do situace a seznámit s výsledky hypotéz. Dále by měla být zaměřena na to, co je možné do budoucna udělat, aby se situace zlepšila. Nakonec je důležité zmínit důležitost vizuálních prvků v podobě obrázků a grafů, které mohou přilákat pozornost k některým informacím, což zvyšuje efektivitu prezentace (Guenole et al., 2017).

### Etická stránka

HR analytika, stejně tak jako výzkum obecně, může být snadno zneužita. Např. chceme-li potvrdit účinnost intervence, můžeme si v datech snadno dohledat informace, které našemu tvrzení nějakým způsobem odpovídají (Rasmussen & Ulrich, 2015). V rámci

retence zaměstnanců je důležité upozornit především na riziko poškození zaměstnanců, bude-li s výsledky špatně naloženo. Jak již bylo výše zmíněno, při modelování fluktuace zaměstnanců je možné prezentovat buď individuální či skupinové pravděpodobnosti odchodu. Při individuální prezentaci výsledků je nutné zvážit, jak vedoucí pracovníci se získanými informacemi naloží (tj. zda nebudou protežovat zaměstnance s nižší rizikovostí a zanedbávat ty s vyšší pravděpodobností odchodu apod.). Tím spíše, pokud není k dispozici dostatečné množství proměnných a model vysvětluje menší procento variance odchodovosti. Vždy je nutné mít na paměti, že se jedná pouze o model přibližující skutečnost, nikoliv o skutečnost, která nutně nastane (Speer, Dutta, Chen, & Trussell, 2019).

## 4. Fluktuace zaměstnanců

### 4.1 Definice fluktuace zaměstnanců a jejích typů

V oblasti lidských zdrojů pojem fluktuace označuje odchodovost zaměstnanců, kteří jsou většinou nahrazeni novými pracovníky. Konkrétně může být fluktuace pojímána např. jako subjektivně chtěný odchod zaměstnance z organizace, který zahrnuje nejen odchod pracovníka, ale zároveň i nástup nového pracovníka na jeho místo (Milkovich & Boudreau, 1993) či jako veškeré příchody a odchody zaměstnanců (Armstrong, 2007). Fluktuaci je možné dále klasifikovat dle nejrůznějších kritérií. Jedním z příkladů je dělení fluktuace zaměstnanců podle tří dimenzí (Allen, Bryant, & Vardaman, 2010). Zprv je lze dělit na dobrovolnou a nedobrovolnou podle toho, zda iniciativa k odchodu je na straně zaměstnavatele či zaměstnance. Snaha o retenci zaměstnanců v rámci HR analytiky je většinou zaměřena na zaměstnance, kteří odchází dobrovolně. Oproti tomu nedobrovolná odchodovost zaměstnanců (např. pro nedostatečný výkon či v rámci restrukturalizace organizace) bývá pro zaměstnavatele žádoucí. Dalším kritériem dělení je míra schopnosti zaměstnanci v odchodu zabránit – zde mluvíme o odvratitelné (resp. neodvratitelné) fluktuaci, kde je (resp. není) v silách zaměstnavatele odchodovost eliminovat. Eliminace nadbytečné fluktuace zaměstnanců je možná např. tam, kde zaměstnanci organizaci opouští pro nespokojenost či pocity nedocení. Odchodu není možné zabránit u osob odcházejících na mateřskou (resp. otcovskou dovolenou), do důchodu či ze zdravotních důvodů. Pro organizace je strategické svůj čas a investice směřovat k spíše odvratitelnému typu fluktuace. Poslední dimenze odlišuje funkční a dysfunkční fluktuaci. Dysfunkční fluktuace má na organizace neblahý vliv. Zahrnuje takové odchody zaměstnanců, kdy společnost přichází o vysoce kvalifikované a jiné těžce nahraditelné pracovníky. Oproti tomu s funkční fluktuací, kde jsou zaměstnanci snadno nahraditelní, se společnost lehce vypořádá a někdy může být dokonce prospěšná např. tam, kde noví pracovníci jsou levnější nebo přinášejí do společnosti nový pohled na věc nebo jiné inovace. Proto je pro společnost opět strategické cílit na dysfunkční fluktuaci. Speciální kategorií jsou zaměstnanci, kteří neodešli z organizace, nýbrž pouze přešli na jinou pozici (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012; Allen, Bryant, & Vardaman, 2010).

Správné, užší nadefinování fluktuace je důležité, jelikož přispívá k přesnosti analytických postupů v rámci HR analytiky – např. při tvorbě predikčních modelů. Při dostatečně velkém vzorku je např. možné odlišit některé typy fluktuace, a tak utvořit dva

specifičtější modely. Dělení také umožňuje vyřazení některých jedinců (např. těch, jejichž odchod byl neodvratitelný či takových, kteří odešli na jiné místo v rámci jedné organizace). Obecně je důležité na některé typy fluktuace nahlížet zvlášť (Speer, Dutta, Chen, & Trussell, 2019).

## 4.2 Metody měření fluktuace

K numerickému vyčíslení fluktuace poslouží několik způsobů. Tím pravděpodobně nejužívanějším je míra fluktuace (tzv. *turnover rate*), ke které dospějeme, vydělíme-li počet odešlých zaměstnanců průměrným počtem daných zaměstnanců za dané období a výsledek vynásobíme stem. Dalším je tzv. index stability (tzv. *stability index*), který vypočítáme podílem počtu pracovníků, kteří jsou ve společnosti déle než rok, a počtem zaměstnanců před jedním rokem, to celé vynásobeno stem. Třetím způsobem je míra „úmrtnosti“ (tzv. *survival rate*), ke které dospějeme, vezmeme-li kohortu zaměstnanců, kteří nastoupili v určité době, a podíváme-li se na počet zaměstnanců, kteří z této skupiny zůstali. Proporce „přeživších“ zaměstnanců se dá vyjádřit např. v procentech. Dalším možným způsobem vyčíslení fluktuace je index poločasu (tzv. *half-life index*). Tato metrika vyjadřuje, jak dlouhá je perioda, než se daná kohorta pracovníků zmenší na polovinu (Armstrong, 2007).

## 4.3 Důsledky fluktuace

Fluktuace má pro společnost, ve které k ní dochází, mnoho negativních důsledků. Ovšem jejich závažnost závisí právě na typu fluktuace. U vysoce kvalifikovaných a výkonných zaměstnanců stejně tak jako u zaměstnanců se specializovanými dovednostmi bývá fluktuace často dobrovolná. Jejich odchody tak bývají často nečekané a hůře řešitelné, proto pro organizaci představují závažnější důsledky (Park & Shaw, 2013).

Někteří autoři upozorňují na kvadratický vztah mezi mírou fluktuace a jejími důsledky. Podle této hypotézy škodí jak příliš vysoká, tak nízká fluktuace. Zatímco při vysoké fluktuaci je negativní dopad zjevný (např. vysoké finanční náklady), negativní dopad příliš nízké míry fluktuace bývá vysvětlen nedostatečným přísunem inovativních myšlenek od potenciálních zaměstnanců a vyššími náklady vynaloženými na seniorní zaměstnance. Avšak tento vztah se nejeví jako příliš silný (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012).

### 4.3.1 Finanční dopady

Odchod pracovníka je spojen s náklady, které souvisí s výběrem a náborem nového zaměstnance, ale také s poklesem produktivity v případě, že pracovník není okamžitě nahrazen a místo je volné či vyplněné jinými zaměstnanci (Milkovich a Boudreau, 1993). Výdaje za odchod zaměstnance snadno vyšplhají na 150 % ročních nákladů na daného zaměstnance. Ovšem náklady jsou výrazně vyšší u obchodních zástupců a na manažerských pozicích, kde mohou činit až 250 % ročních výdajů za danou pozici. Výraznější jsou také u pracovníků se specializovanými znalostmi a dovednostmi (Armstrong 2007; Bliss, 2004).

Podle Blisse (2004) a Armstronga (2007) je při vyčíslování nákladů vynaložených při odchodu zaměstnance nutné započítat následující faktory:

- Náklady na zaměstnance, který zaplní danou pozici v době, dokud nebude zaměstnán a zaučen nový jedinec.
- Přibližně 50% ztráta produktivity na dané pozici v případě, že bude prázdná pozice zaplněna jiným zaměstnancem, 100% ztráta produktivity, nebude-li pozice zaplněna vůbec.
- Náklady za tzv. výstupní pohovory – dotazníky či rozhovory určené pro odcházející zaměstnance, které zaměstnavateli pomáhají odhalit problémové oblasti a faktory, které v organizaci přispívají k odchodovosti zaměstnanců.
- Výdaje za školení, která odcházející zaměstnanec absolvoval, spolu s náklady obětované příležitosti zaučujících osob.
- Snížený zisk společnosti způsobený ztrátou vědomostí, schopností a dovedností zaměstnance, který odchází.
- Např. u obchodních zástupců finanční ztráty spojené se ztrátou zákazníků, o které společnost přijde spolu s odchodem zaměstnance.
- Výdaje spojené s vyvěšením inzerátu či s případným najmutím personální agentury.
- Výdaje za čas interního personalisty, který musí porozumět požadavkům pozice, vyvinout vhodnou strategii nábory, vybrat vhodné kandidáty k pohovorům, zorganizovat pohovory a připravit se na ně, provést pohovory, vytvořit nabídku pro vhodného kandidáta a dále kontaktovat neúspěšné kandidáty. A dále náklady obětované příležitosti spojené s časem, který interní personalista věnoval odcházejícímu zaměstnanci.

- Výdaje za další části přijímacího řízení – např. psychodiagnostika, assessment centra, příprava domácího úkolu apod., které zaměstnavateli pomáhají zhodnotit charakteristiky, schopnosti a dovednosti potenciálních kandidátů.
- Výdaje za čas dalších zaměstnanců, kteří jsou zapojeni do výběrového řízení (např. přítomnost nadřízeného na pohovoru či při konání assessment centra).
- Náklady vynaložené za zaučení nového pracovníka – výdaje za čas supervizora (instruování, kontrola práce a náklady obětované příležitosti), cena za materiály (manuály a jejich příprava), odpisy za technické vybavení potřebné k zaučení, čas za případná školení.
- Ztráty spojené s dočasně sníženou produktivitou nového pracovníka, dokud není plně zaučen.
- Administrativní náklady spojené s odchodem starého a příchodem nového zaměstnance (výdaje spojené s účetnictvím, zánikem a zřízením identifikačních karet, pracovních e-mailů, debetních karet, bezpečnostních hesel apod.).
- Časová investice vedoucího pracovníka potřebná k vzájemnému vybudování důvěry s novým zaměstnancem.

Naopak společnost ušetří za:

- Výdaje spojené s platem a benefity, které by vynaložila za původního zaměstnance.

Ovšem prostředky ušetřené za platy a benefity odešlých zaměstnanců jsou výrazně nižší než výše uvedené náklady způsobené jejich odchodem (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012).

#### **4.3.2 Dopady na výkon**

Další oblastí, která je ovlivněna fluktuací zaměstnanců, je výkonnost společnosti. Podle některých autorů působí dobrovolná fluktuace oproti fluktuaci nedobrovolné na výkon škodlivěji, jelikož nedobrovolná fluktuace může pro společnost představovat i pozitivní výsledky – např. pokud společnost opustí neproduktivní zaměstnanci (Park & Shaw, 2013). Jiní autoři se domnívají, že negativní důsledky fluktuace ve formě finančních ztrát vždy převažují nad pozitivy nově příchozí pracovní síly, která může být levnější nebo má potenciál nahlížet na problémy z jiné perspektivy (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012).

Dopad fluktuace na výkon se liší v závislosti na charakteristikách daných společností a pozic. Zprv má fluktuace výrazněji negativní vliv na výkon ve společnostech, na jejichž pozicích jsou třeba specializované dovednosti. Dále má silnější vliv ve středně velkých firmách více než v malých či velkých a stejně tak na vyšších (např. manažerských) pozicích. Silnější vztah mezi fluktuací a obchodními výsledky ve středně velkých společnostech není v literatuře objasněn. Negativní vliv fluktuace na vyšších pozicích by mohl být způsoben tím, že je náročné nahrazovat školení a rozvojové programy pro pracovníky na těchto pozicích (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012). Kromě toho jsou rozdílně ovlivněny jednotlivé oblasti společnosti. Vysoká fluktuace negativně působí v odděleních zákaznických služeb a odděleních kvality a bezpečnosti než v jiných oblastech (Hancock, Allen, Bosco, McDaniel, & Pierce, 2012). Dále míra fluktuace negativně koreluje se spokojeností zákazníků, ziskovou marží, efektivitou produkce, efektivitou prodeje a pozitivně s celkovou chybovostí společnosti (Heavey, Holwerda, & Hausknecht, 2013).

### **4.3.3 Dopady na další zaměstnance**

#### **Sociální nákaza fluktuace**

Sociální nákaza jako tendence napodobovat emoce, myšlení i chování druhých (Burt & Janicik, 1996) je psychologický fenomén, který můžeme pozorovat v mnoha oblastech zaměstnání např. u neetického či nežádoucího chování v organizaci (Zuber, 2015). Stejně tak se s nákazou setkáváme i u chování souvisejícím s odchodem z organizace. Podle některých autorů dobrovolná odchodovost zaměstnanců z organizace pozitivně koreluje s dobrovolnou odchodovostí kolegů, kteří jsou pro ně důležití (Wang, Newman, & Dipboye, 2016). V jiné studii tzv. zakotvenost pracovníků v zaměstnání, tedy množství a síla faktorů, které zaměstnance udržují v zaměstnání, negativně korelovala s chováním jejich kolegů, které souviselo s hledáním jiného zaměstnání, a stejně tak se samotnou odchodovostí ze současného zaměstnání (Felps et al., 2009).

### **4.4 Strategie prevence fluktuace**

Vzhledem k negativním dopadům, které má fluktuace zaměstnanců pro společnosti, je na místě zabývat se faktory, které s fluktuací souvisí. Díky tomu je možné fluktuaci do budoucna predikovat a také jí předcházet (Speer, Dutta, Chen, & Trussell, 2019).

#### 4.4.1 Faktory související s fluktuací

Současná literatura poskytuje výčet faktorů, které s odchodovostí zaměstnanců souvisí. Tyto charakteristiky jsou níže rozepsány jakožto potenciální prediktory fluktuace, u některých jsou uvedeny odhady síly populačního parametru. Rovněž se jedná o oblasti, na které je dobré se zaměřit v rámci prevence fluktuace (Speer, Dutta, Chen, & Trussell, 2019).

#### Schopnosti a dovednosti pracovníků

Někteří autoři uvádí, že v některých případech může s dobrovolnou fluktuací souviset kognitivní kapacita zaměstnance. U zaměstnání s vysokými požadavky na kognitivní kapacitu zaměstnance byl nalezen kvadratický vztah mezi kognitivní kapacitou a odchodovostí zaměstnanců. Pracovníci s nižší a vyšší kognitivní kapacitou měli větší tendence k brzkému odchodu z organizace. Oproti tomu pracovníci na pozicích, které vyžadovaly nižší kognitivní kapacitu, vykazovali nepřímo úměrný lineární vztah mezi kognitivní kapacitou a fluktuací. Zaměstnanci, kteří vykazovali vyšší skóre v inteligenčních testech, odcházeli ze zaměstnání po delší době než zaměstnanci skórující méně. V prvním uvedeném případě u zaměstnanců s nižší kognitivní kapacitou je možné vyšší míru odchodovosti vysvětlit skrze inkongruenci mezi jejich možnostmi a příliš vysokými nároky, které na ně pozice klade. V obou případech u zaměstnanců s vyšší kognitivní kapacitou se nabízí teorie, že je zaměstnanec k odchodu ovlivněn možnostmi, které mu pracovní trh nabízí (Maltarich, Nyberg, & Reilly, 2010). Faktor, který souvisí se schopnostmi a dovednostmi pracovníků, jsou školení. V metaanalýze týkající se prediktorů pro kolektivní odchodovost, byla nalezena pozitivní souvislost mezi pro firmu specifickým školením ( $\hat{\rho} = -0,4$ ). Tato spojitost by mohla být způsobena skutečností, že školení zaměřená na dovednosti specifické pro zaměstnancovu současnou pozici zvyšují zaměstnancův závazek vůči společnosti. Oproti tomu u obecných školení žádný vztah nalezen nebyl (Heavey, Holwerda, & Hausknecht, 2013).

#### Osobní charakteristiky

Co se týče osobnostních charakteristik studie poukazují na možnou souvislost mezi fluktuací a následujícími osobnostními rysy: emoční stabilita ( $\hat{\rho} = -0,19$ ), svědomitost ( $\hat{\rho} = -0,16$ ), otevřenost ke zkušenosti, místo řízení (*locus of control*,  $\hat{\rho} = 0,1$ ) a iniciativnost. Zdrojem těchto dat jsou především dotazníky (Speer, Dutta, Chen, & Trussell, 2019).



Rubenstein a kol. (2018) provedli rozsáhlou metaanalýzu faktorů, které souvisely s odchodovostí zaměstnanců. Z osobnostních charakteristik svědomitost, emoční stabilita a interní místo řízení (*locus of control*) nepřímo úměrně korelovaly s odchodovostí. Jedním z možných vysvětlení je, že tyto charakteristiky jedinci umožňují lépe překonávat stres a jiné překážky, které se v zaměstnání objeví. Zimmerman (2008) ve své metaanalýze uvádí, že osobnost zaměstnance souvisí jak se záměry odejít, tak i s jejich skutečným odchodem. V jeho studii se záměry odejít negativně korelovala emoční stabilita a svědomitost, se skutečnou odchodovostí pak negativně přívětivost, svědomitost a emoční stabilita, pozitivně naopak otevřenost ke zkušenosti.

Liu a kol. (2011) uvádí, že zaměstnanci méně odchází ze zaměstnání, kde cítí podporu a možnost autonomie, které vedou k jejich zmocnění (*empowerment*). Proto menší tendenci k fluktuaci pravděpodobně mají zaměstnanci, kteří jsou na pracovišti iniciativní a díky tomu ke své práci více vnitřně motivovaní (Liu, Zhang, Wang, & Lee, 2011).

### **Demografické charakteristiky**

Další skupinou charakteristik, které souvisí s odchodovostí zaměstnanců, jsou demografické údaje. Zde je však nutné mít na paměti, že není etické k charakteristikám jako věk, pohlaví či rasa přihlížet při optimalizaci rozhodování v rámci některých oblastí HR. Jelikož např. zaměstnávání osob na základě věku či pohlaví je diskriminující, je tedy nepřijatelné (Rubenstein, Eberly, Lee, & Mitchell, 2018).

Z této skupiny s odchodovostí negativně koreluje věk ( $\hat{\rho} = -0,21$ ) a počet dětí ( $\hat{\rho} = -0,2$ ; Rubenstein et al., 2018). Možných vysvětlení pro negativní souvislost mezi fluktuací a věkem je větší flexibilita a méně realistická očekávání mladších lidí (Ng & Feldman, 2009). Naopak nebyl zaznamenán vztah mezi odchodovostí, rasou a pohlavím (Rubenstein, Eberly, Lee, & Mitchell, 2018; Griffeth, Hom, & Gaertner, 2000). Podle Rubensteina a kol. (2018) také s fluktuací negativně koreluje doba, po kterou je zaměstnanec doposud v dané společnosti ( $\hat{\rho} = -0,2$ ). Důvodem pro tuto souvislost je pravděpodobně závazek, který zaměstnanec vůči společnosti pocítuje.

### **Informace z náboru**

Informace získané během nástupu zaměstnance mohou také předpovídat pravděpodobnost odchodu. Zaměstnanci, kteří se opakovaně hlásí na pozice v rámci stejné organizace, své zaměstnání opouští v méně případech než ti, kteří jsou přijati napoprvé. Dále

méně fluktuují uchazeči, kteří o sobě při náboru uvádějí více osobních informací, a uchazeči, kteří před nástupem do zaměstnání byli zaměstnání jinde, tedy nebyli nezaměstnaní. Jako možné vysvětlení se jeví vyšší motivace kandidáta do zaměstnání nastoupit, která posléze také sníží pravděpodobnost jeho odchodu (Breugh, 2014). Roli také může hrát zdroj náboru. Společnosti, které přednostně nabírají kandidáty z jiných pozic uvnitř společnosti, reportují nižší odchodovost zaměstnanců (Haines, Jalette, & Larose, 2010). Stejně tak méně ohroženi odchodem jsou zaměstnanci, kteří byli přijati na základě doporučení jiného zaměstnance, což může být vysvětleno realističtějšími očekáváním kandidáta (Pieper, 2015).

### **Charakteristiky zaměstnání**

Zaměstnanci jsou méně ohroženi odchodem, dostává-li se jim dostatečné autonomie a podpory, díky čemuž se cítí dostatečně zmocnění (Liu, Zhang, Wang, & Lee, 2011). Naopak negativně na odchodovost zaměstnanců působí, nemají-li vliv na dění ve společnosti, nemohou-li komunikovat své potřeby a názory na dění ve společnosti a také, nemohou-li se podílet na rozhodování (Allen, Shore, & Griffeth, 2016; Heavey, Holwerda, & Hausknecht, 2013). Další faktor, který negativně koreluje s odchodovostí zaměstnanců, jsou elektronické monitorování zaměstnanců a rutinně koncipované pracovní činnosti (Heavey, Holwerda, & Hausknecht, 2013). Vysvětlením pro tyto souvislosti je pozitivní vliv na zaměstnancovu motivaci, pokud mají pocit, že jim jejich nadřízený důvěřuje.

### **Plat a benefity**

Co se týče platu a benefítů, podle níže uvedených studií se záměry odejít či se skutečnou odchodovostí nějakým způsobem koreluje: rychlost růstu platu, absolutní výše platu ( $\hat{\rho} = -0,17$ ), relativní výše platu vůči ostatním, spokojenost s platem ( $\hat{\rho} = -0,07$ ), příležitosti k rozvoji schopností a dovedností a vnímaná férovost při rozdělování odměn ( $\hat{\rho} = -0,17$ ; Heavey et al., 2013; Nyberg, 2010; Rubenstein et al., 2018). Podle Rubenstein a kol. (2018) také benefity (zahrnující školení, možnost povýšení, bonusy a nefinanční benefity;  $\hat{\rho} = -0,28$ ).

Nyberg (2010) ve své metaanalýze uvádí, že pravděpodobnost dobrovolného odchodu zaměstnanců souvisí s rychlostí růstu platu. Růst platu negativně korelovala s odchodovostí zaměstnanců, u vysoce výkonných pracovníků rychleji než u méně výkonných. Avšak je důležité zmínit, že mediátorem tohoto vztahu nebyla spokojenost zaměstnanců, jak by se mohlo nabízet. Jedním z možných vysvětlení pro tu skutečnost je to,

že zaměstnanci s nižším růstem platu mohou být náchylnější k odchodu na jiné, lépe placené, pracovní pozice, přestože jsou v současném zaměstnání spokojeni (Nyberg, 2010). To je však v opozici s výsledky studie autorů Tekleaba a kol. (2005), podle kterých spokojenost s růstem platů negativně koreluje jak se záměrem, tak se skutečným odchodem z organizace. Dalším faktorem týkajícím se platu, který souvisí s odchodovostí zaměstnanců, je jeho výše v porovnání s ostatními pracovníky na podobných pozicích v dané oblasti. Čím vyšší je plat v poměru s platy ostatních, tím menší je pravděpodobnost odchodu daného zaměstnance. To je v souladu s teorií spravedlnosti, podle které na motivaci zaměstnanců pozitivně působí, je-li s nimi férově zacházeno (Armstrong, 2007; Heavey, Holwerda, & Hausknecht, 2013).

Kromě platu při odchodovosti zaměstnanců hrají roli i další benefity nabízené zaměstnavatelem. Zaprvé někteří zaměstnanci spíše zůstanou v organizaci, budou-li mít možnost rozvoje svých schopností a dovedností spolu s možností kariérního růstu. Odchod naopak hrozí tam, kde se zaměstnancům dostává příležitostí k rozvoji, ale chybí možnosti kariérního růstu. Ti totiž mohou mít větší tendence odejít ze zaměstnání, aby svůj potenciál mohli využít jinde (Kraimer, Seibert, Wayne, Liden, & Bravo, 2011). Zadruhé samotná existence benefitů negativně koreluje s odchodovostí zaměstnanců (Heavey, Holwerda, & Hausknecht, 2013). Nakonec je to pocíťovaná férovost při rozdělování odměn, která taktéž negativně koreluje s fluktuací zaměstnanců (Allen, Shore, & Griffeth, 2016).

### **Postoje k práci**

S odchodovostí zaměstnanců a jejich záměry odejít pravděpodobně také souvisí angažovanost ( $\hat{\rho} = -0,20$ ) a spokojenost pracovníků ( $\hat{\rho} = -0,28$ ), míra zakoušeného stresu ( $\hat{\rho} = -0,21$ ) a schopnost vypořádat se s ním ( $\hat{\rho} = -0,39$ ), vzájemný respekt mezi zaměstnanci a pocity naplnění. Souvislost mezi postoji zaměstnanců k práci a odchodovostí je pravděpodobně způsobena dalšími faktory, se kterými jsou tyto postoje propojeny. Např. u angažovanosti byla nalezena souvislost mezi lepším zdravím, vyšší spokojeností a lepším pracovním prostředím, které přímo souvisí s nižší tendencí pracovníků odcházet. Podobně je tomu u míry zakoušeného stresu a schopností se se stresem vypořádat. Celkově by se dalo říci, že angažovaní a spokojení pracovníci ve vztahu ke svému zaměstnání pocíťují pozitivní emoce, což může být dalším faktorem, který je v dané společnosti drží. U stresu tomu je naopak (Collini, Guidroz, & Perez, 2015).

Někteří autoři našli interakci mezi vzájemným respektem mezi zaměstnanci, pocitu naplnění pramenící z pracovní činnosti a angažovaností, která predikovala fluktuaci u pracovníků zdravotních služeb (Collini, Guidroz, & Perez, 2015). Celková spokojenost v zaměstnání je dalším faktorem, který souvisí s odchodovostí zaměstnanců. Konkrétně byl nalezena slabá negativní korelace mezi těmito dvěma proměnnými (Griffeth, Hom, & Gaertner, 2000).

V jiné studii byly měřeny zaměstnancovy úmysly odejít z organizace. Administrativní pracovníci, kteří byli vystaveni vysoké míře stresu a vykazovali nižší míru vnímané podpory za strany zaměstnavatele, ve vyšší míře plánovali odchod ze zaměstnání (Thorsteinsson, Brown, & Richards, 2014). Rubenstein a kol. (2018) ve své metaanalýze uvádí pozitivní vztah mezi mírou pocitovaného vyčerpání (resp. stresu) a odchodovostí zaměstnanců. Vztah je možné vysvětlit několika způsoby. Zaměstnanci např. odchází ve snaze vyhnout se nepohodě, kvůli sníženému výkonu, který je důsledkem stresu, či např. pro nemoc a následné absentérství (Zimmerman, Swider, Woo, & Allen, 2016). Se stresem také souvisí schopnost se s ním vypořádat, ta podle Rubensteina a kol. (2018) středně silně negativně koreluje s odchodovostí zaměstnanců.

### **Sociální vztahy**

Na důležitost sociálních vztahů na pracovišti upozorňuje hojné množství literatury, které se souvislostí mezi kvalitou sociálních vazeb a odchodovostí zaměstnanců zabývá. Zdrojem dat, které v zaměstnání vypovídají o sociálních vztahů, mohou být např. 360° zpětné vazby či četnost e-mailové nebo jiné elektronické komunikace (Speer, Dutta, Chen, & Trussell, 2019). Z této oblasti autoři uvádí kvalitu vztahů na pracovišti ( $\hat{p} = -0,14$ ), vzájemnou soudržnost ( $\hat{p} = -0,11$ ), podporu ( $\hat{p} = -0,19$ ) a podobnost, zakotvenost v zaměstnání ( $\hat{p} = -0,14$ ), vztahy s nadřízeným, styl vedení nadřízeného ( $\hat{p} = -0,63/\hat{p} = -0,26$ ), chování kolegů, networking, výměna vedoucího nebo kolegy ( $\hat{p} = -0,23$ ), poměr žen ( $\hat{p} = 0,17$ ), diverzita zaměstnanců ( $\hat{p} = 0,19$ ), průměrný věk zaměstnanců v týmu ( $\hat{p} = -0,26$ ) spolu s průměrnou dobou strávenou v zaměstnání ( $\hat{p} = -0,25$ ) a názory partnera zaměstnance.

Pojmem souvisejícím s odchodovostí pracovníků je tzv. zakotvenost v zaměstnání. Tento pojem zahrnuje soubor sil, které zaměstnance drží v zaměstnání. Jednou z těchto sil jsou právě vztahy na pracovišti a jejich kvalita. Jelikož zakotvenost v zaměstnání negativně koreluje s odchodovostí zaměstnanců, dá se předpokládat, že kvalitní vztahy vedou k silnější zakotvenosti a tím pádem i nižší odchodovosti zaměstnanců (Yang, Ma, & Hu, 2011).

K tomuto názoru se přiklání i Liu a kol. (2011), kteří uvádí, že je na pracovišti důležitá podpora ze strany ostatních pracovníků. Zaměstnanci totiž v zaměstnání spíše zůstanou, mají-li kolegy, na které se mohou spolehnout. Dle studie provedené Ellingson a kol. (2016) u pracovní síly s nízkou pracovní kvalifikací fluktuace souvisí s tím, zda zaměstnanec navázal sociální vztahy se svými kolegy. Avšak tato souvislost byla nalezena pouze u zaměstnanců v životní fázi postupně se vynořující dospělosti. Dalším faktorem, který se pozitivně odráží na odchodovosti zaměstnanců, je vzájemná soudržnost (Heavey, Holwerda, & Hausknecht, 2013) a podobnost mezi sebou a svými kolegy. Zaměstnanci, kteří se považují za odlišné od svých kolegů, mají větší pravděpodobnost odchodu z organizace (Liao, Chuang, & Joshi, 2008). Pozitivně působí jak vnitřní, psychická podobnost, tak podobnost v oblasti věku a doby strávené v zaměstnání. V podobnějším kolektivu je jedinec pravděpodobně více zakotven, což ho drží v zaměstnání delší dobu. Důvodem pravděpodobně je, že podobnost posiluje citové pouto k ostatním zaměstnancům a identifikaci s nimi (Liao, Chuang, & Joshi, 2008; Heavey, Holwerda, & Hausknecht, 2013). Zakotvenost zaměstnance je ovlivněna i demografickými charakteristiky dalších členů týmu. Tedy pokud ti mají nižší či vyšší pravděpodobnost odchodu, pak je ohrožen vyšší pravděpodobností odchodu i každý jednotlivec týmu. Vzhledem k tomu, že častěji odchází ženy, které častěji odchází např. kvůli částečným úvazkům, je poměr žen v týmu pozitivním signifikantním prediktorem fluktuace. Dále protože častěji odchází mladší lidé, věk jedinců v týmu je negativním prediktorem fluktuace. Stejně tak protože častěji odchází lidé, kteří jsou ve společnosti kratší dobu, doba, po kterou jsou členové týmu v dané společnosti, je negativním prediktorem fluktuace (Heavey, Holwerda, & Hausknecht, 2013).

Nesmíme opomenout ani vztah zaměstnance a nadřízeného, který tvoří významnou součást pracovní zkušenosti. V této oblasti s odchodovostí zaměstnanců souvisí např. typ vedení či spokojenost s nadřízeným (Wells & Welty Peachey, 2011; Speer, Dutta, Chen, & Trussell, 2019; Heavey, Holwerda, & Hausknecht, 2013). Konkrétně na retenci zaměstnanců pozitivně působí např. snahy zaměstnance vnitřně motivovat, oproti vedení, kdy nadřízeného vyvíjí nátlak (Reina, Rogers, Peterson, Byron, & Hom, 2017). Podle další studie transformační leadership souvisí se záměry ze společnosti odejít jak přímo, tak nepřímo, jelikož manažeři používající transformační leadership budují kulturu založenou na takových hodnotách (např. spolupráci), která zaměstnance lépe udrží v zaměstnání (Sun & Wang, 2016). Riziko fluktuace je také vyšší u zaměstnanců, kterým byl změněn manažer nebo kolega v týmu (Griffeth, Hom, & Gaertner, 2000).

Zaměstnanec může být jak při zvažování alternativ k současnému zaměstnání, tak při skutečném odchodu ovlivněn i chováním svých kolegů. Autoři studie Felps a kol. (2009) naznačují, že zakotvenost zaměstnanců v zaměstnání negativně korelovala se snahou kolegů poohlédnout se po jiném zaměstnání a také byla signifikantním, negativním prediktorem pro jejich odchodovost. Tento fakt je ve studii vysvětlen vzájemnou nakažlivostí chování, postojů a prožívání mezi kolegy.

Networking, tedy dobrovolné navazování pracovních kontaktů, patří mezi další faktor související s fluktuací. Zahrnuje činnosti jako získávání pracovních kontaktů na firemních akcích či vzájemné sdílení pracovní náplně s kolegy z jiných oddělení. Networking provozovaný v rámci organizace snižuje pravděpodobnost odchodu zaměstnance, jelikož zaměstnanci, kteří jej provozují, bývají v zaměstnání spokojenější a více zakotveni. Oproti tomu externí networking koreluje s odchodovostí pozitivně, jelikož zaměstnanci tak získávají více alternativních pracovních příležitostí (Porter, Woo, & Campion, 2016).

Na pravděpodobnost odchodu zaměstnance má vliv nejen sociální podpora vnímaná na pracovišti, ale také podpora, kterou zaměstnanec získává od svého partnera. Je-li partner zaměstnance s jeho zaměstnáním spokojený, snižuje se pravděpodobnost odchodu daného zaměstnance. Autoři studie tento fenomén vysvětlují tak, že kariérní rozhodnutí zaměstnance jsou ovlivněny nejen jím samotným, ale postoji, přesvědčením a chováním jeho partnera (Huffman, Casper, & Payne, 2014).

## **Hodnocení výkonu**

Dalším prediktorem fluktuace může být hodnocení výkonu zaměstnance ( $\hat{p} = -0,08$ ) nebo objektivní výkon ( $\hat{p} = -0,21$ ) a také kontextuální výkon ( $\hat{p} = -0,22$ ). Informaci o výkonu nám mohou poskytnout jak objektivní data – např. počet uzavřených smluv, tak subjektivní data – např. v podobně hodnocení nadřazeného (Speer, Dutta, Chen, & Trussell, 2019).

Morrow a kol. (1999) ve své studii zaznamenal, že zaměstnanci, jejichž výkon byl hůře hodnocen, měli větší tendenci k brzkému odchodu z organizace. Možným vysvětlením pro tento vztah může být vnímání negativního hodnocení jakožto hrozby. Jiní autoři popisují vztah mezi fluktuací a tzv. kontextovým výkonem (Harrison, Newman, & Roth, 2006), tedy zaměstnancovým výkonem nad rámec běžných pracovních povinností zahrnujícím nápomocné chování vůči svým kolegům a organizaci jako celku (Goodman & Svyantek,

1999). Konkrétně uvádí mírnou velikost efektu pro negativní lineární vztah pro tyto dvě charakteristiky. Vysvětlením tohoto vztahu může být větší ochota investovat do svého zaměstnání, pokud v něm zaměstnanec plánuje setrvat. Vedle toho také nápomocné chování vůči kolegům může posílit sociální vazby mezi zaměstnanci, což zaměstnance drží v zaměstnání (Harrison, Newman, & Roth, 2006).

### **Náznaky stažení se**

Varovným signálem mohou být i absence ( $\hat{p} = 0,23$ ) a pozdní příchody zaměstnanců ( $\hat{p} = 0,14$ ), stejně tak jako vědomé záměry z organizace odejít ( $\hat{p} = 0,56$ ) a jednání směřující k nalezení jiného zaměstnání ( $\hat{p} = 0,4$ ). Tato data se dají získat jak objektivně, tak prostřednictvím dotazníků (Rubenstein, Eberly, Lee, & Mitchell, 2018; Speer, Dutta, Chen, & Trussell, 2019). Možným vysvětlením tohoto vztahu je, že stažení se slouží jako copingový mechanismus pro vyrovnání se např. se stresem, nepohodou nebo nespokojeností (Morrow et al., 1999).

Morrow a kol. (1999) upozorňuje na souvislost mezi mírou absentérství a dobrovolnou fluktuací. Zaměstnanci, kteří v práci více chyběli, měli větší tendenci k brzkému odchodu z organizace (Morrow, McElroy, Laczniak, & Fenton, 1999). Stejně tak pravděpodobnost s odchodovostí zaměstnanců pozitivně koreluje pozdní příchody (Rubenstein, Eberly, Lee, & Mitchell, 2018). Záměr odejít z organizace je dalším z prediktorů skutečného odchodu. Zajímavým poznatkem je, že se v literatuře setkáváme s velkými rozdíly v míře variability skutečné odchodovosti vysvětlené záměry odejít – např. 5 % (Cohen, Blake, & Goodman, 2016) či 11 % (Heavey, Holwerda, & Hausknecht, 2013).

### **Charakteristiky organizace**

Důležitým faktorem je soulad mezi zaměstnancem a firemní kulturou. Míra tohoto souladu byla měřena pomocí souhlasu s výroky, které se vztahovaly např. k pocitu sounáležitosti s kolegy, vnímání manažera a jeho snahy podporovat spolupráci nebo řešit konflikty, snahám společnosti oceňovat úsilí zaměstnanců apod., a signifikantně predikovala dobu, kterou zaměstnanec strávil v dané společnosti (Thaden, Jacobs-Priebe, & Evans, 2010). Možné vysvětlení pro tuto souvislost je, že firemní kultury se vyznačují různou mírou fluktuace – některé se vyznačují nízkou, jiné vysokou, což láká určitý typ zaměstnanců, ale zároveň je zaměstnanec firemní kulturou zpětně ovlivněn. Dále také samotný rozpor mezi hodnotami vyznávanými v rámci firemní kultury a hodnotami vyznávanými zaměstnancem

může přímo působit na pravděpodobnost zaměstnancova odchodu (Goldstein et al., 2017). Na odchodovost zaměstnanců také pozitivně působí, nachází-li se organizace v dobré lokaci – např. v místě, ze kterého mají zaměstnanci snadnou dostupnost ke klientům (Heavey, Holwerda, & Hausknecht, 2013).

## **Pracovní trh**

Míra fluktuace může také souviset s mírou nezaměstnanosti a celkovou situací na pracovním trhu ( $\hat{\rho} = 0,23$ ). Při nižší míře nezaměstnanosti se pro větší množství možností mezi zaměstnanci objevuje vyšší míra fluktuace (Rubenstein, Eberly, Lee, & Mitchell, 2018; Speer, Dutta, Chen, & Trussell, 2019). Vysoce výkonní zaměstnanci při vyšší relevantní míře nezaměstnanosti jsou paradoxně více ohroženi dobrovolným odchodem z organizace než jejich méně výkonní kolegové, jelikož se jim na pracovním trhu dostává více možností (Nyberg, 2010).

## **Výzkumný design studií**

Co se týče výzkumného designu, většina studií má spíše explanační charakter, a proto často využívá strukturálního a víceúrovňového hierarchického modelování, které umožňuje dobře zachytit vztahy mezi proměnnými a zároveň se hodí tam, kde jsou jednotlivá pozorování na sobě závislá (Soukup, 2006), jako je tomu v případě zaměstnanců (Collini, Guidroz, & Perez, 2015; Felts et al., 2009; Kraimer, Seibert, Wayne, Liden, & Bravo, 2011; Liu, Zhang, Wang, & Lee, 2011; Thorsteinsson, Brown, & Richards, 2014). Někteří autoři studií popisující vztahy mezi odchodovostí a dalšími faktory pomocí bodově biseriální korelace (Huffman, Casper, & Payne, 2014). Studie, které se zaměřují na predikci fluktuace do budoucna nejčastěji volí analýzu přežití (Liao, Chuang, & Joshi, 2008; Maltarich, Nyberg, & Reilly, 2010; Pieper, 2015) nebo také binární logistickou metodu (Morrow et al., 1999). Obě tyto metody jsou vhodným nástrojem pro modelování fluktuace (více viz kapitola 3.3).

### **4.3.2 Předcházení fluktuace**

Allen a kol. (2010) spolu s dalšími autory (viz níže) udávají výčet rad, které zaměstnavateli mohou pomoci předcházet fluktuaci:



- Poskytnout při náboru potenciálním zaměstnancům realistický popis práce; zaměstnanci, kteří mají při náboru k dispozici realistický popis práce zůstávají v zaměstnání déle než ti, co jej nemají (Breaugh & Starke, 2016).
- Zaměřit se na to, zda kandidát zapadne do stávajícího kolektivu a firemní kultury; další pomocnou metodou je na základě biografických dat potenciálního zaměstnance odhadnout, zda zapadne do dané firemní kultury (Thaden, Jacobs-Priebe, & Evans, 2010).
- Řádně socializovat nového zaměstnance; zaměstnancům prospívá mají-li své mentory a školitele, dále informace o tom, jak bude probíhat proces socializace, a pozitivní zpětnou vazbu ohledně jejich výkonu již během zaučování (Allen, Bryant, & Vardaman, 2010; Porter, Woo, & Campion, 2016; Yang, Ma, & Hu, 2011)
- Poskytovat zaměstnancům příležitosti dalšího vzdělávání; školení a další příležitost ke vzdělávání snižují pravděpodobnost odchodu zaměstnanců. Děje se tak zvláště na pozicích, kde je od zaměstnanců očekáváno, že budou své schopnosti a dovednosti neustále přizpůsobovat současným požadavkům trhu. Příležitosti k dalšímu vzdělávání by ovšem měly být v souladu s možným kariérním postupem v zaměstnání. Pokud tomu tak není, hrozí naopak zaměstnancův odchod do jiné organizace (Kraimer, Seibert, Wayne, Liden, & Bravo, 2011).
- Poskytnout zaměstnancům odměny a jiné benefity; zaměstnanci oceňují benefity odpovídající jejich preferencím a potrpí si na spravedlnost při jejich udělování (Allen, Shore, & Griffeth, 2000; Heavey, Holwerda, & Hausknecht, 2013).
- Školit zaměstnance na manažerských pozicích; zaměstnanci na manažerských pozicích zůstávají déle v organizacích, kde mají možnost rozvíjet své manažerské schopnosti a dovednosti, tedy např. navazování dobrých vztahů s podřízenými, stejně tak pozitivně působí schopnosti a dovednosti manažerů na retenci jejich podřízených (Wells & Welty Peachey, 2011).
- Zaměřit se na angažovanost pracovníků: poskytovat zaměstnancům pozitivní zpětnou vazbu za jejich přínosy pro společnost, nabízet zaměstnancům možnost rozvoje, poskytovat zaměstnancům úkoly, které jsou pro ně výzvou, poskytovat zaměstnancům smysluplné úkoly podporující jejich autonomii, spolupráci s kolegy, sdělovat zaměstnancům, jak prostřednictvím své práce přispívají k dosahování cílů společnosti (Collini, Guidroz, & Perez, 2015)

## Výzkumná část

Výzkumná část navazuje na část literární a to tak, že je zaměřena na retenci zaměstnanců konkrétní společnosti, velké české banky. Konkrétně obsahuje predikční model, jehož cílem je identifikovat zaměstnance ohrožené odchodem a prediktory, které s odchodovostí souvisí. Banka v práci nebude jmenována, což je v souladu s jejími interními předpisy.

Vzhledem k cíli výzkumné části je pro větší přehlednost v následujících odstavcích uvedeno shrnutí vysvětlujících proměnných odchodovosti zaměstnanců. Tyto prediktory byly identifikovány v literárně přehledové části práce na základě předchozích výzkumů a jsou relevantní pro následující výzkum.

Z demografických charakteristik se jako nejsilnější prediktory jeví věk zaměstnance ( $\hat{\rho} = -0,21$ ), počet dětí ( $\hat{\rho} = -0,2$ ) a doba ( $\hat{\rho} = -0,2$ ), kterou jedinec již strávil na svém zaměstnání (Rubenstein et al., 2018). Pro všechny tři proměnné předpokládáme negativní vztah s fluktuací zaměstnanců. S fluktuací také korelují postoje zaměstnanců. Relevantním postojem je angažovanost ( $\hat{\rho} = -0,2$ ), která, mimo jiné, poukazuje na zaměstnancův závazek vůči zaměstnavateli, jak se se společností identifikuje a zda má důvěru v to, jaké služby společnost poskytuje (Kular et al., 2008). Vzhledem k tomu, že pro analýzu nebyly k dispozici přímo výsledky průzkumu angažovanosti, jsou ve výzkumu použity informace o čerpání jednotlivých produktů, které banka svým zaměstnancům nabízí jak zvýhodněně, tak za plnou cenu. Předpokladem je, že zaměstnanci, kteří firemní produkty hojněji využívají, budou s menší mírou ze zaměstnání odcházet. Užívání firemních bankovních produktů nemusí souviset pouze s angažovaností, ale je pro zaměstnance také velkým benefitem např. v případě nižší úrokové sazby. Rubenstein a kol. (2018) ve své metaanalýze uvádí jako prediktory fluktuace zaměstnanců nejen benefity ( $\hat{\rho} = -0,28$ ), ale také bonusy ( $\hat{\rho} = -0,28$ ) a vyšší platu ( $\hat{\rho} = -0,17$ ), Nyberg (2010) přidává rychlost růstu platu. Na základě těchto informací jsou do analýzy zahrnuty veškeré příjmy a benefity zaměstnanců – plat, finanční bonusy a benefity a rychlost jejich růstu, přičemž je předpokládán negativní vztah mezi těmito proměnnými a fluktuací zaměstnanců. Variabilní složka příjmu ve formě bonusů je pro pozice spjaté s prodejem, tedy i na pozici osobního bankéře, velmi důležitá. Vedle finančních benefitů s odchodovostí zaměstnanců souvisí také nefinanční benefity. Jedním z nich je možnost účastnit se školení zaměřená na rozvoj schopností a dovedností ( $\hat{\rho} = -0,4$ ), které jedinec uplatní na své současné pozici (Heavey, Holwerda, & Hausknecht, 2013). U

proměnných týkajících se absolvovaných školení je očekáván negativní vztah s fluktuací. Z tohoto důvodu jsou do analýzy zařazeny informace o absolvovaných školeních, konkrétně o počtu kurzů, kterých se zaměstnanec zúčastnil, a celkové doby, kterou zaměstnanec na kurzech strávil. Liao a kol. (2008) upozorňují na důležitost vzájemné podobnosti zaměstnanci (týkající se psychických i demografických charakteristik), která zřejmě působí jako jedna z mnoha sil, která zaměstnance udržuje v zaměstnání. Z tohoto důvodu jsou jako potenciální prediktory použity proměnné věková podobnost a podobnost ve vztahu k době strávené na dané pozici. Vedle toho zaměstnancova rozhodnutí ohledně odchodu ovlivňují i demografické charakteristiky členů týmů; konkrétně se jedná o poměr mužů a žen ( $\hat{\rho} = 0,17$ ), diverzitu ( $\hat{\rho} = 0,19$ ), průměrný věk ( $\hat{\rho} = -0,26$ ) a průměrnou dobu strávenou v zaměstnání ( $\hat{\rho} = -0,25$ ; Heavey, Holwerda, & Hausknecht, 2013). Pro podobnost s ostatními zaměstnanci, průměrný věk a průměrnou dobu v zaměstnání je očekáván negativní vztah s fluktuací, u diverzity naopak pozitivní. Dalším potenciálním prediktorem je výměna některého z členů týmu ( $\hat{\rho} = -0,23$ ), u které očekáváme pozitivní vztah s fluktuací. Další prediktory odchodovosti se mohou týkat výkonu zaměstnance (Rubenstein et al., 2018); z tohoto důvodu byly do analýzy zařazeny i proměnné týkající se hodnocení zaměstnance ( $\hat{\rho} = -0,08$ ) spolu s objektivními ukazateli jeho výkonu ( $\hat{\rho} = -0,21$ ). Vedle charakteristik zaměstnance nebo zaměstnavatele s odchodovostí zaměstnanců korelují i externí faktory, konkrétně situace na pracovním trhu ( $\hat{\rho} = 0,23$ ), kdy při větším množství pracovních nabídek jsou zaměstnance více ohroženi odchodem (Rubenstein, Eberly, Lee, & Mitchell, 2018). Pro prověření tohoto vztahu je ve výzkumu proměnná týkající se míry nezaměstnanosti a očekáván je negativní vztah s fluktuací zaměstnanců. Speer a kol. (2019) navrhují také zabývat se při analýze fluktuace proměnnými, které se týkají výše úvazku, pohlaví, vzdělání a zkušeností zaměstnance, vzdáleností zaměstnání od bydliště zaměstnance, velikosti jeho týmu a zkušeností jeho nadřízeného.

## 5. Výzkumný problém, cíle výzkumu a hypotézy nebo výzkumné otázky

Banka se potýká s vysokou mírou fluktuace na pozici *osobní bankéř*. Jedná se o pozici, kdy je pracovník v přímém kontaktu s klienty, a to buď osobní nebo telefonickou formou. Hlavní náplní pozice je prodej finančních produktů zákazníkům. Data týkající se zaměstnanců, která byla bankou poskytnuta pro účely této diplomové práce, obsahují historické personální záznamy sbírané do konce roku 2017.

Cílem bylo zaměřit se na dobrovolnou fluktuaci zaměstnanců, protože ta má pro společnost negativní důsledky, na rozdíl od fluktuace zaměstnanců iniciovaná ze strany zaměstnavatele, která naopak může mít dopady pozitivní (Speer, Dutta, Chen, & Trussell, 2019).

Výzkumná část práce je rozdělena na dvě části. První část, statisticko-inferenční, je zaměřena na testování stanovených hypotéz a určení velikostí efektů jednotlivých prediktorů. Druhá, datově-vědecká část, je zaměřena na tvorbu predikčního modelu, evaluaci jeho kvality a validaci. Každá z částí plní v HR analytickém projektu jinou funkci. Statisticko-inferenční část umožňuje lépe rozpoznat strukturu a sílu vztahů mezi proměnnými a fluktuací. Díky tomu může společnost získat konkrétní doporučení, která jí umožní fluktuaci zaměstnanců snížit. Oproti tomu cílem datově-vědecké části je zkonstruovat model, který bude co nejpřesněji klasifikovat zaměstnance, kteří odejdou nebo neodejdou. Vzájemná struktura vztahů zde již není tolik důležitá.

### 5.1 Statisticko-inferenční část

V rámci statisticko-inferenční částí bylo cílem zodpovědět výzkumnou otázku, jaké faktory a jakou silou s dobrovolnou odchodovostí osobních bankéřů souvisí.

Hypotézy byly formulovány v souladu s cíli práce, na základě dostupné literatury, která se zabývá prediktory fluktuace (podrobnější přehled viz výše v kapitole 4.4.1), a také dat, které bylo možné získat z dané společnosti.

**H1:** Celkový model pro odchodovost zaměstnanců je signifikantní

**H2:** Průměrný věk v týmu je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H3:** Variabilita věku v týmu je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H4:** Průměrná doba strávených let ve společnosti u členů týmu je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H5:** Variabilita doby strávených let ve společnosti u členů v týmu je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H6:** Poměr mužů v týmu je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H7:** Množství absolvovaných kurzu typu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H8:** Množství absolvovaných kurzu typu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H9:** Množství absolvovaných kurzu typu 3 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H10:** Množství absolvovaných kurzu typu 4 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H11:** Množství absolvovaných kurzu typu 5 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H12:** Množství absolvovaných kurzu typu 6 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H13:** Množství absolvovaných kurzu typu za poslední rok je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H14:** Výše bonusu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H15:** Rychlost růstu bonusu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H16:** Rychlost růstu bonusu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H17:** Výše bonusu 3 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H18:** Rychlost růstu bonusu 3 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H19:** Výše benefitu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H20:** Výše benefitu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H21:** Výše mimořádného bonusu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H22:** Výše mimořádného bonusu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H23:** Platová třída je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H24:** Sestoupení v platové třídě je pozitivním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H25:** Postoupení v platové třídě je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H26:** Čerpání produktu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H27:** Čerpání produktu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H28:** Čerpání produktu 3 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H29:** Čerpání produktu 4 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H30:** Čerpání produktu 5 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H31:** Čerpání produktu 6 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H32:** Čerpání produktu 7 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H33:** Čerpání produktu 8 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H34:** Čerpání produktu 9 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H35:** Čerpání produktu 10 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H36:** Čerpání produktu 11 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H37:** Intenzita čerpání produktu 1 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H38:** Intenzita čerpání produktu 2 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H39:** Intenzita čerpání produktu 3 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H40:** Intenzita čerpání produktu 4 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H41:** Intenzita čerpání produktu 5 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H42:** Intenzita čerpání produktu 6 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H43:** Intenzita čerpání produktu 7 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H44:** Intenzita čerpání produktu 8 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H45:** Intenzita čerpání produktu 9 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H46:** Intenzita čerpání produktu 10 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H47:** Intenzita čerpání produktu 11 je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H48:** Míra nezaměstnanosti je signifikantním, pozitivním prediktorem dobrovolné fluktuace zaměstnanců

**H49:** Počet týmů, ve kterých zaměstnanec byl, je signifikantním, pozitivním prediktorem dobrovolné fluktuace zaměstnanců

**H50:** Počet manažerů, pod kterými zaměstnanec byl, je signifikantním, pozitivním prediktorem dobrovolné fluktuace zaměstnanců

**H51:** Objektivní hodnocení zaměstnance je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H52:** Věk zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H53:** Doba, po kterou zaměstnanec pracuje ve společnosti, je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H54:** Počet dětí zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců



**H55:** Závazek zaměstnance vůči společnosti ve formě spoludlužnictví je negativním, signifikantním prediktorem dobrovolné fluktuace zaměstnanců

**H56:** Výše platu zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H57:** Výše celkového příjmu zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H58:** Subjektivní hodnocení zaměstnance od nadřízeného ve formě výtky je pozitivním prediktorem dobrovolné fluktuace zaměstnanců

**H59:** Věková podobnost je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H60:** Podobnost ohledně doby strávené v zaměstnání je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H61:** Rychlost růstu celkového příjmu zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H62:** Rychlost růstu platu zaměstnance je signifikantním, negativním prediktorem dobrovolné fluktuace zaměstnanců

**H63:** Interakce mezi pohlavím zaměstnance a poměrem mužů v týmu je signifikantním prediktorem dobrovolné fluktuace zaměstnanců

Všechny hypotézy byly testovány na hladině významnosti  $\alpha=0,05$ . Avšak pro velké množství hypotéz byla hladina významnosti upravena pomocí Bonferroniho korekce jako podíl 0,05 a počtu hypotéz náležícím jednotlivým prediktorům ( $0,05/62$ ). Výsledná hladina významnosti tedy činila přibližně 0,0008.

## **5.2 Datově-vědecká část**

V rámci datově-vědecké části práce bylo cílem zodpovědět následující výzkumné otázky:

**Výzkumná otázka 1:** S jakou přesností dokáží vytvořené modely predikovat odchody jednotlivých zaměstnanců?

**Výzkumná otázka 2:** Může být implementace vytvořených modelů pro klienta užitečná v rámci jeho personální praxe?

## **6. Design výzkumného projektu**

### **6.1 Typ výzkumu**

Pro účely výzkumu byla zvolena kvantitativní strategie, tedy cílem bylo ověřit a kvantifikovat souvislost prediktorů a fluktuace zaměstnanců v konkrétní společnosti. Jedná se o aplikovaný výzkum. Výzkumným designem je kvaziexperiment, jelikož v rámci výzkumu jsou porovnávány dvě skupiny zaměstnanců (ti co odešli a ti, co tak neučinili), přičemž rozřazení do skupin nebylo možné ovlivnit výzkumníkem.

### **6.2 Metody získávání dat**

HR analytické projekty, tedy i výzkum prováděný v rámci této diplomové práce, spadají do aplikovaného výzkumu. Tato skutečnost ovlivňuje i volbu způsobu získávání dat. Na rozdíl od akademického výzkumu je v rámci HR analytiky možné získat z HR systémů velká množství personálních dat, která společnosti pro své potřeby dlouhodobě sbírají (Speer, Dutta, Chen, & Trussell, 2019). Tímto způsobem byla získána i data využita v této diplomové práci, což je plně v souladu s metodologickými postupy běžnými pro HR analytické projekty.

Data potřebná pro výzkum prováděný v rámci této diplomové práce byla získána od banky, jejíž zaměstnanci jsou předmětem zkoumání. Vzhledem k tomu, že se jedná o historické personální záznamy, data byla vyňata z HR systému specifického pro danou banku. Vzhledem k výzkumné otázce bylo stěžejní, aby získaná data obsahovala informaci o datu nástupu a odchodu zaměstnance (resp. informaci o tom, že zaměstnanec zatím neodešel), informaci o straně, která odchod zaměstnance iniciovala (zaměstnanec či zaměstnavatel) a dále záznamy týkající se potenciálních prediktorů (např. demografické záznamy, záznamy o absolvovaných školeních, o čerpaných produktech apod.). Data banka poskytla ve velmi čisté podobě většinou bez chybějících a odlehlých hodnot. Chybějící hodnoty, které se týkaly proměnných vztahujících se k objektivnímu hodnocení zaměstnanců, byly eliminovány v rámci transformace proměnných viz dále. Dále chyběly hodnoty u proměnných, které se vztahovaly ke změně příjmu (88), bonusu (3) a platu (1), avšak zde pro nízký počet (pod 5 %) byla pozorování s danými hodnotami buď zcela eliminována nebo nahrazena střední hodnotou.

Na základě shody mezi možnými prediktory fluktuace, které byly nalezeny na základě dohledané literatury (viz kapitola 4.4.1), a proměnnými, které bylo možno od banky získat, byly určeny výzkumné hypotézy pro tuto práci. Některé informace, zejména psychologického charakteru, jako např. informace o osobních charakteristikách, stylu vedení nadřízeného, vztazích na pracovišti, nebylo možné získat, jelikož banka takové údaje nesbírá. Z dat byly následně vybrány ty proměnné, které byly potřebné pro otestování stanovených hypotéz. Data týkající se financí byla z důvodu interních předpisů banky převedena na jinou škálu.

### **Transformace proměnných**

Získaná data obsahovala většinu potřebných proměnných. Přesto byly přidány některé další proměnné za pomoci transformace původních. Některé proměnné byly vytvořeny na základě jiných proměnných – týmová podobnost týkající se věku a doby strávené v zaměstnání. Dále byla vytvořena proměnná rychlosti nárůstu platu, příjmu a bonusu jakožto podíl procentuálního nárůstu platu, příjmu, bonusu a doby strávené v zaměstnání.

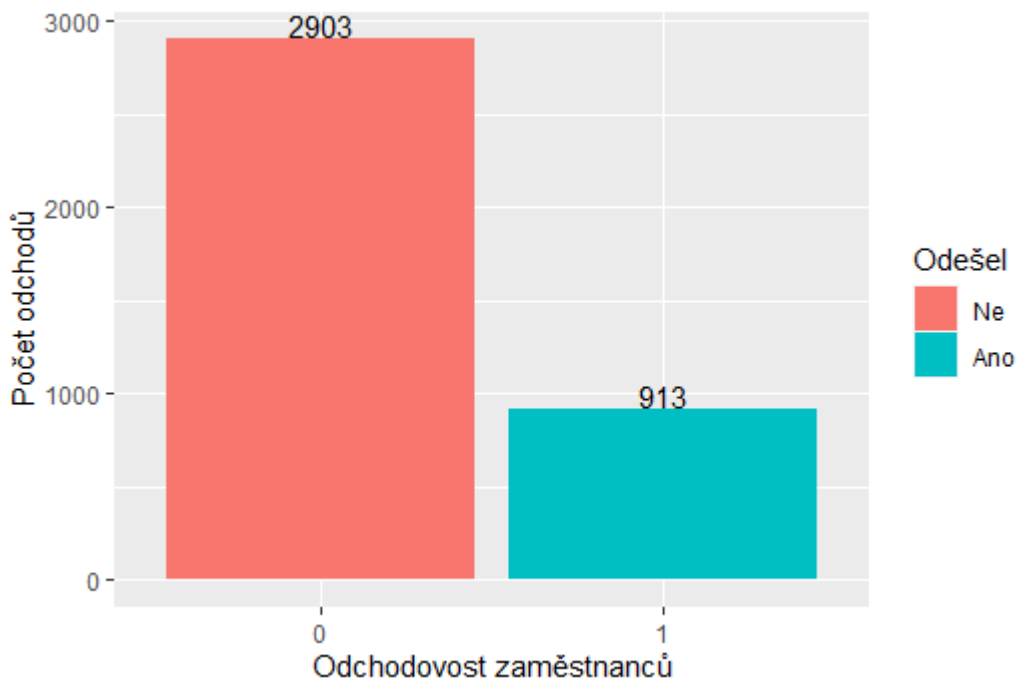
Vedle toho bylo potřeba vyrovnat se s chybějícími hodnotami u proměnných, které se vztahovaly k objektivnímu hodnocení zaměstnanců, jelikož pozorování týkající se zaměstnanců, kteří ve společnosti jsou nebo byli příliš krátkou dobu tyto hodnoty pochopitelně postrádají. Toho bylo docíleno převodem na kategorické proměnné, kdy políčka s chybějícími hodnotami tvořila jeden stupeň kategorické proměnné. Transformace spojitých proměnných na kategorické bylo docíleno pomocí kvantilů na osm rovnoměrných částí.

Na závěr bylo třeba do finální datové matice zahrnout pouze zaměstnance, kteří svůj odchod dobrovolně sami iniciovali (nikoliv ty jejichž odchod inicioval zaměstnavatel), spolu se zaměstnanci, kteří na své pozici dále setrvávají. Z datové matice byli vyjmuti ti zaměstnanci, kteří odešli z důvodu nástupu do starobního důchodu.

### **6.3 Metody zpracování a analýzy dat**

Nejdříve bylo cílem reportovat základní popisné ukazatele týkající se fluktuace zaměstnanců. V roce 2016 a 2017, tedy v období, které trvalo dva roky, dobrovolně celkově odešlo 913 (cca 24 %) zaměstnanců, zatímco 2903 (cca 76 %) zůstalo. Přibližná míra dobrovolné fluktuace je tedy 24 %, což je důležitým údajem, protože víme, že pro relativně

vysokou míru odchodovosti nám data poskytují dostatečnou variabilitu pro její predikci do budoucnosti.



**Graf 3:** Výzkumný soubor dle fluktuace zaměstnanců

### 6.3.1 Statisticko-inferenční část

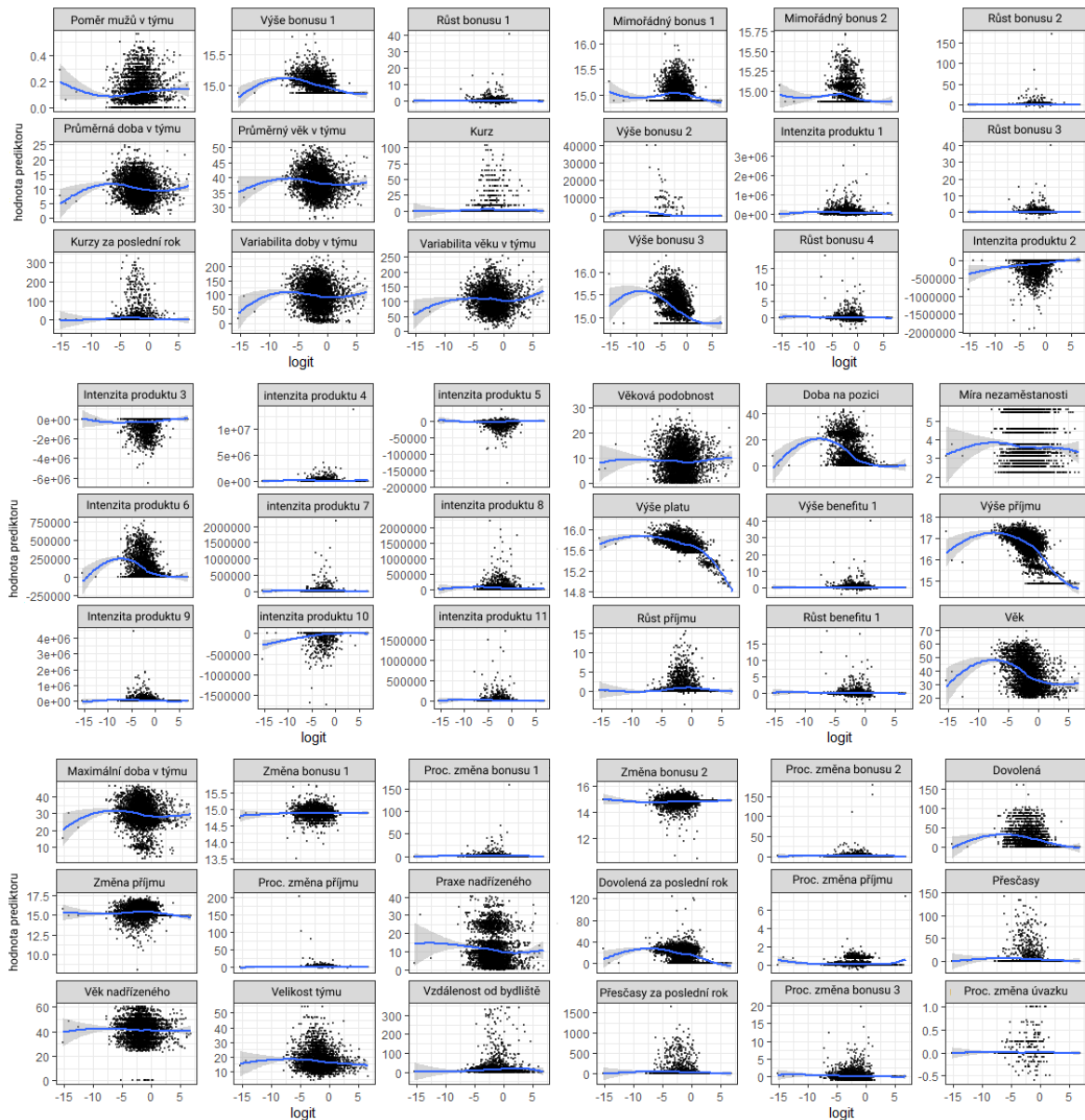
K modelování byla využita binární logistická regrese. Vzhledem k tomu, že pracovníci, kteří byli předmětem této práce, se nachází v různých pobočkách, bylo zváženo víceúrovňové modelování. To by bylo vhodné pro použití za předpokladu, že by se míra odchodovosti lišila v rámci jednotlivých poboček. Proto byly porovnány relativní četnosti týkající se odchodovosti napříč jednotlivými kraji. Model nebyl signifikantní na všech rozumných hladinách významnosti ( $\chi^2(13) = 14,4, p = 0,35$ ), z toho důvodu víceúrovňové modelování nebylo využito.

Pro použití binární logistické modelování bylo nutné ověřit některé předpoklady. Nutným předpokladem, který vyplývá z definice binární logistické regrese je binomická závislá proměnná, což bylo splněno (Field, Miles, & Field, 2012), jelikož predikovanou proměnnou byla informace o tom, zda zaměstnanec již ze zaměstnání odešel (1) či nikoliv (0).

Druhým nutným předpokladem pro použití logistické regrese je dostatečně velký vzorek. Jedním z hrubých předpokladů je 15 pozorování na prediktor (Field, Miles, & Field, 2012). Vzhledem k vysokému počtu pozorování v této práci ( $N = 3816$ ) byl tento předpoklad s rezervou dodržen. Kromě toho by nezávislé proměnné neměly obsahovat odlehlé hodnoty, což bylo opět dodrženo. Předpoklad byl testován pomocí tzv. *Cook's distance* a standardních residuí (Field, Miles, & Field, 2012).

Dalším předpokladem je absence multikolinearity, tedy prediktory by mezi sebou neměly příliš korelovat. Tento předpoklad se dá ověřit např. pomocí tzv. VIF (*variance inflation factor*) statistiky, která by rozhodně neměla být vyšší než 10 (hodnoty mezi 5 a 10 jsou diskutabilní; Field, Miles, & Field, 2012). Předpoklad absence multikolinearity byl zajištěn během přidávání proměnných do modelu – žádná z přidaných proměnných hodnotu 10 nepřesáhla.

U logistické regrese by dále měl být dodržen předpoklad o lineárním vztahu mezi spojitými prediktory a logitem – přirozeným logaritmem podílu šanci predikované proměnné (Field, Miles, & Field, 2012). Tento předpoklad byl pro spojité prediktory ověřen na finálním modelu pomocí vizualizace, tedy grafů mezi hodnotami daných proměnných a logaritmech podílů šanci predikované proměnné. U všech proměnných byl zaznamenán lineární vztah. Grafy těchto proměnných jsou uvedeny níže (viz Graf 4).



**Graf 4:** Ověření linearity mezi spojitými proměnnými a logaritmy podílů šancí predikované proměnné

Vzhledem k povaze práce byl zvolen hierarchický přístup k selekci vysvětlujících proměnných. Tato technika se vyznačuje přidáváním proměnných do modelu ve více krocích, které vycházejí z poznatků získaných na základě předchozích výzkumů (Field, Miles, & Field, 2012). V prvním kroku jsou do modelu vloženy proměnné, u nichž je známý směr vztahu s fluktuací a odhad velikosti efektu z předchozích výzkumů, v druhém kroku jsou do modelu vloženy proměnné, které sice mají oporu v literatuře, avšak bez znalosti směru jejich vztahu s fluktuací a bez znalosti velikosti efektu, v posledním kroku jsou do modelu vloženy zatím neznámé proměnné, které však v kontextu dané analýzy dávají

výzkumníkovi smysl. Pozorování, která obsahovala chybějící hodnoty, byla pro malý počet odstraněna.

### 6.3.2 Datově-vědecká část

#### Binární logistická regrese

Pro účely tvorby predikčního modelu byla zvolena kroková metoda sestupného výběru. Výchozí model obsahoval všechny proměnné, které do modelu byly vloženy během třetího kroku v rámci hierarchické selekce prediktorů. Dalším krokem bylo data rozdělit v poměru 8:2 na trénovací a testovací datový soubor, přičemž testovací byl odložen. K validaci modelu byla použita  $k$ -násobná křížová validace o deseti rozděleních.

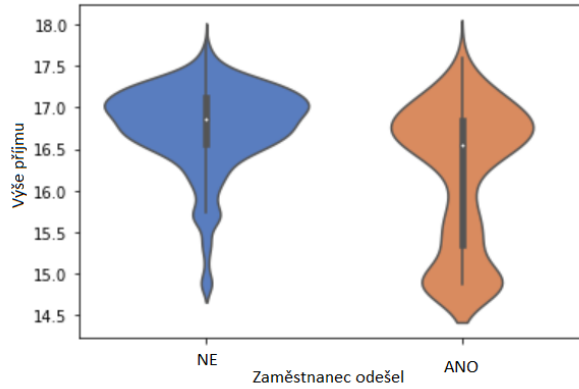
Předpoklady pro použití této metody již nebylo nutné testovat, jelikož byly splněny u výchozího modelu definovaného v rámci statisticko-inferenční části výzkumu.

#### Další modely

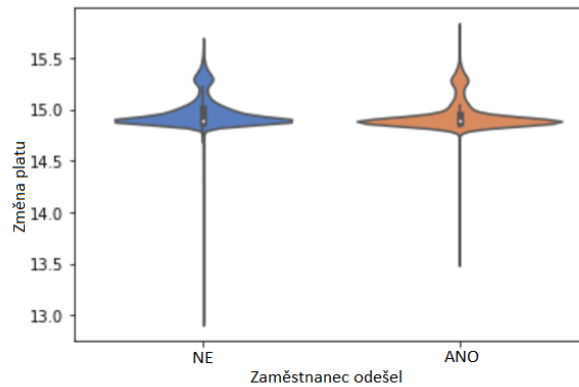
Vedle logistické regrese byly využity i modely strojového učení – náhodné lesy (*Random Forests*), tzv. *Gradient Boosting*, tzv. *Extreme Gradient Boosting* a metoda podpůrných vektorů (*Support Vector Machine*). Pro tento účel byla data opět rozdělena v poměru 8:2 na trénovací a testovací vzorek. K validaci bylo využito  $k$ -násobné křížové validace o pěti rozděleních. Další analýzy probíhaly pouze na trénovacím vzorku. Nejdříve byla zjištěna prediktivní síla vysvětlujících proměnných (pomocí Giniho koeficientu a entropie), následně byly vyřazeny slabé a vysoce korelující prediktory. K prediktorům byly také získány základní statistické charakteristiky (míry polohy a variability) a jejich typ (kategorické či numerické). Nakonec bylo třeba proměnné převést na stejnou škálu a kromě toho byly chybějící hodnoty nahrazeny mediánem dané proměnné.

Také bylo před samotným modelováním třeba seznámit se s daty a vizualizovat vztahy mezi predikovanou proměnnou a prediktory. Takové vizualizace mohou přiblížit pozorované vztahy. Grafy byly vytvořeny pro všechny proměnné, ovšem v této práci jsou uvedeny pouze dva příklady, na kterých je demonstrováno, jaké vzorce můžeme pozorovat již na začátku analýzy. Např. Graf 5 poukazuje na zřetelný rozdíl v rozložení výše příjmu napříč jednotlivými skupinami predikované proměnné. Oproti tomu Graf 6 ukazuje podobné rozložení změny platu v obou skupinách prediktoru.





**Graf 5:** Vztah výše příjmu a predikované proměnné



**Graf 6:** Vztah změny platu a predikované proměnné

Na datech byly natrénovány následující modely: *Gradient Boosting*, *XGBoosting*, náhodné lesy (*Random Forests*) a metoda podpůrných vektorů (*Support Vector Machine*). Vzhledem k povaze modelů proměnné, na základě kterých modely rozhodují, nejsou v práci uvedeny.

## 6.4 Etika výzkumu

Dodržení etických principů je důležitou součástí HR analytických projektu, tedy i výzkumné části této diplomové práce. V rámci této práce byla zpracovávána citlivá data týkající se pracovní historie zaměstnanců, tedy např. jejich hodnocení a odměn. Proto byla data pro účely práce poskytnuta v anonymní podobě a některé údaje byly zamaskovány prostřednictvím převodu na jinou škálu. Kromě toho data nebyla poskytnuta k nahlédnutí žádné další osobě. Samotný proces byl samozřejmě ošetřen smlouvou o mlčenlivosti a ochraně informací.

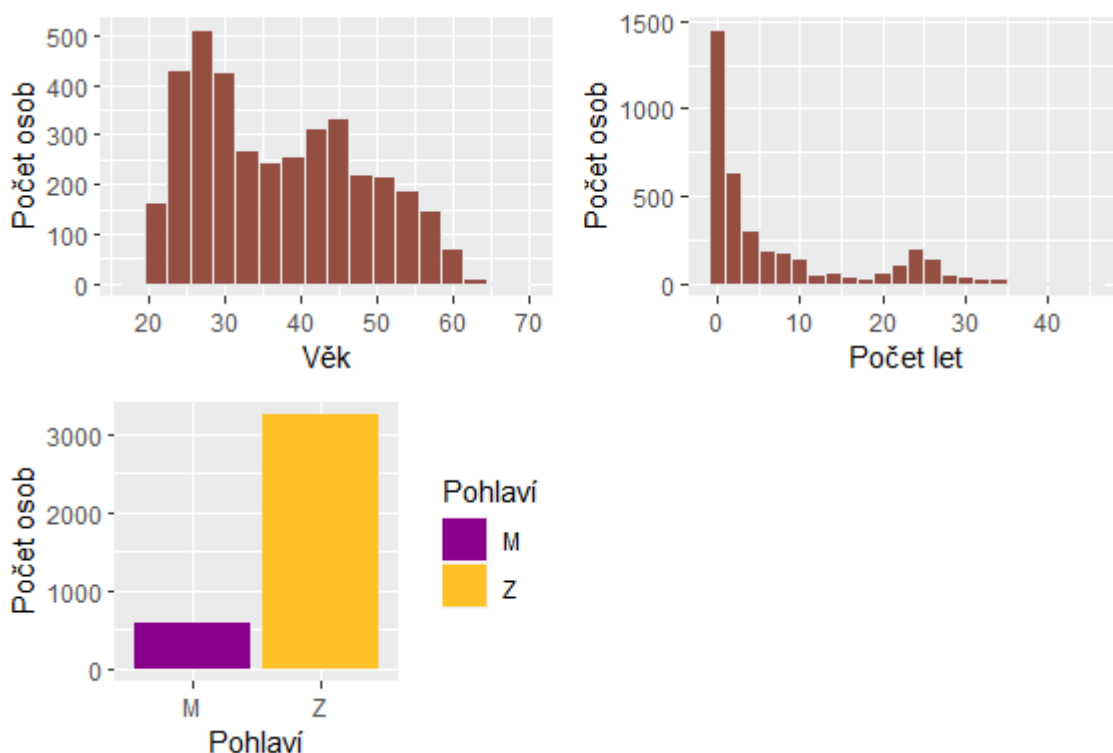
Vzhledem k tomu, že je práce zaměřena na predikci odchodovosti zaměstnanců, bylo třeba zvážit, zda budou reportovány individuální nebo kolektivní pravděpodobnosti odchodu. Nakonec byly reportovány individuální pravděpodobnosti, ovšem bylo třeba tuto skutečnost ošetřit v rámci doporučení v diskusi, aby nedošlo k poškození zaměstnanců. Stručně řečeno, je vhodné zaměřit se v rámci intervencí především na oblasti, které souvisí s jednotlivými prediktory fluktuace. Informace také není vhodné předávat v surové podobě přímým nadřízeným, jelikož by znalost individuálních pravděpodobností mohla negativně ovlivnit jejich chování vůči jednotlivým podřízeným.

Další etickou otázkou bylo zařazení demografických údajů do modelu. Vzhledem k tomu, že do modelu byly zařazeny informace o pohlaví a věku zaměstnanců, model nesmí být využíván k náboru ani jinému rozhodování, které by mohlo diskriminovat některá pohlaví nebo věkové skupiny.

## 7. Výzkumný soubor

Výběr výzkumného souboru byl ovlivněn povahou studie, tedy byli do něj zahrnuti všichni zaměstnanci banky na pozici *osobního bankéře*, kteří se z pozice dobrovolně odešli nebo na ní dále zůstávají. Z tohoto důvodu se jedná o výběr vyčerpávající. Data byla bankou poskytnuta v anonymní podobě.

Výzkumný soubor celkově tvořilo 3816 pozorování ( $N = 3816$ ), přičemž větší procento tvořily ženy  $N = 3241$  (cca 84,93 %) a menší část muži  $N = 575$  (15,07 %). Toto rozložení zcela neodpovídá rozložení mužů a žen v populaci, které je přibližně rovnoměrné (Český statistický úřad, 2017). Co se týče věku respondentů, věk ve výzkumném souboru se pohyboval od 19 do 69 let při průměru 37,21 let a směrodatné odchylce cca 10,85 let. Vzhledem k účelům práce se jeví jako důležité popsat vzorek také z hlediska doby strávené na dané pozici – ta se pohybuje od 0 (např. jedinci, kteří odešli ve zkušební době) do 45,5 let, s průměrem 7,42 let a směrodatnou odchylkou 9,82 let. Vzhledem k tomu, že je doba ve společnosti log-normálně rozdělena, vyplatí se reportovat také medián jakožto přesnější charakteristiku středu, který činí 2,12 let.



**Graf 7:** Rozložení výzkumného souboru dle věku, počtu let strávených na dané pozici a pohlaví

## 8. Výsledky

### 8.1 Statisticko-inferenční část

Proměnné do modelu byly vkládány hierarchicky ve třech krocích. Nejdříve proměnné, které mají oporu v předchozí literatuře včetně síly a směru vztahu, dále proměnné, které mají alespoň nějakou oporu v literatuře, ačkoliv bez směru a síly vztahu, a v posledním kroku byly do modelu vloženy proměnné dávající v rámci této analýzy smysl, ačkoliv jsou bez jakékoliv opory v předchozí literatuře.

V prvním kroku bylo do modelu vloženo 61 proměnných, které se vztahovaly ke stanoveným hypotézám (viz kapitola 5) – týkající se platové třídy, počtu dětí, podobnosti doby strávené ve společnosti, věkové podobnosti, doby strávené ve společnosti, čerpání finančních produktů, objektivního a subjektivního hodnocení výkonu, délky a počtu absolvovaných kurzů a školení, počtu vystřídaných manažerů, čerpání benefitů, míry nezaměstnanosti, platu a celkového příjmu, poměru mužů a žen v týmu, průměrného věku v týmu, průměrné doby ve společnosti v týmu, počtu vystřídaných týmů, variability doby strávené ve společnosti v týmu a věku. Model byl signifikantní  $\chi^2(3754) = 966,795$ ,  $p < 0,001$  a vysvětlená variance činila 23,8 % dle McFaddenova  $R^2$  (0,24). Signifikantními prediktory<sup>2</sup> při korekci požadované hladiny významnosti byly poměr mužů v týmu zaměstnance ( $p < 0,001$ ;  $z=3,77$ ;  $e^b=6,32$ ), zda zaměstnanec obdržel mimořádný bonus ( $p < 0,001$ ;  $z=5,42$ ;  $e^b=8,62$ ), platová třída ( $p < 0,001$ ;  $z=7,16$ ;  $e^b=3,36$ ), míra nezaměstnanosti ( $p < 0,001$ ;  $z=-4,39$ ;  $e^b=0,8$ ), počet týmů, které zaměstnanec vystřídal ( $p < 0,001$ ;  $z=7,48$ ;  $e^b=1,83$ ), doba, po kterou je zaměstnanec ve společnosti ( $p < 0,001$ ;  $z=-3,47$ ;  $e^b=0,96$ ), výše platu ( $p < 0,001$ ;  $z=-8,86$ ;  $e^b=0,0009$ ), výše celkového příjmu ( $p < 0,001$ ;  $z=-10,56$ ;  $e^b=0,3$ ) a rychlost růstu platu ( $p < 0,001$ ;  $z=-4,02$ ;  $e^b=0,84$ ).

V druhém kroku bylo přidáno celkem 7 proměnných, které se týkaly pohlaví, vzdělání, zkušeností, výše úvazku, velikosti týmu, zkušeností nadřízeného a vzdálenosti do zaměstnání, které vysvětlily další 2,2 % variance dle McFaddenova  $R^2$  (0,022), přičemž zlepšení modelu oproti původnímu bylo signifikantní  $\chi^2(3747) = 65,23$ ,  $p < 0,001$  a signifikantní byly dvě proměnné – velikost týmu ( $p < 0,001$ ;  $z=-5,18$ ;  $e^b=0,96$ ) a pohlaví zaměstnance ( $p < 0,001$ ;  $z=-4,86$ ;  $e^b=0,55$ ). V tomto kroku poměr mužů přestal být signifikantní ( $p = 0,7$ ;  $z=1,81$ ;  $e^b=2,54$ ), dále byly stále signifikantní proměnné: zaměstnanec

---

<sup>2</sup> Váhy prediktorů jsou kvůli lepší interpretovatelnosti uváděny ve formě poměrů šancí.

obdržel mimořádný bonus ( $p < 0,001$ ;  $z=5,25$ ;  $e^b=8,35$ ), platová třída ( $p < 0,001$ ;  $z=6,98$ ;  $e^b=3,27$ ), míra nezaměstnanosti ( $p < 0,001$ ;  $z=-4,05$ ;  $e^b=0,806$ ), počet týmů, které zaměstnanec vystřídal ( $p < 0,001$ ;  $z=7,23$ ;  $e^b=1,8$ ), doba, po kterou je zaměstnanec ve společnosti ( $p < 0,001$ ;  $z=-3,51$ ;  $e^b=0,96$ ), výše platu ( $p < 0,001$ ;  $z=-8,58$ ;  $e^b=0,001$ ), výše celkového příjmu ( $p < 0,001$ ;  $z=-10,5$ ;  $e^b=0,3$ ) a rychlost růstu platu ( $p < 0,001$ ;  $z=-4,12$ ;  $e^b=0,83$ ).

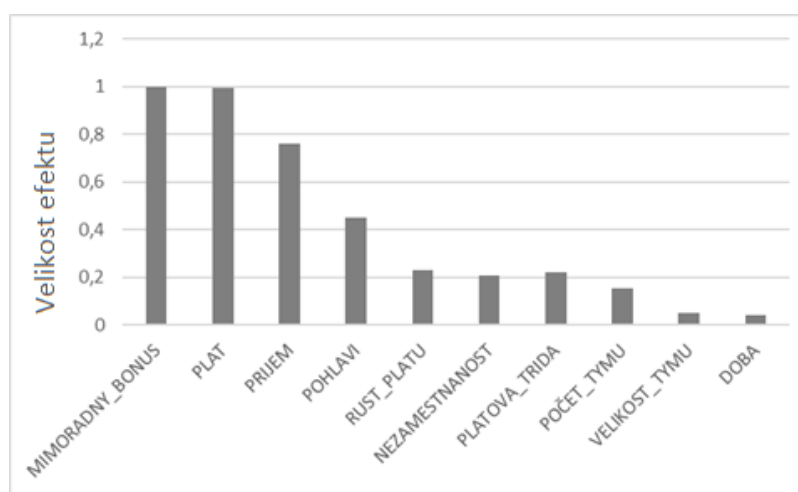
V posledním, třetím, kroku, bylo do modelu přidáno 33 proměnných týkajících se maximální doby ve společnosti v týmu, existence partnera, pohlaví a věku nadřízeného, změny adresy, typu obsluhovaných klientů, změny příjmu, změny bonusů, dovolené, přesčasů, změny platu, výše úvazku a změněných poboček a pozic, které vysvětlily další 3 % variance dle McFaddenova  $R^2$  (0,03), přičemž zlepšení modelu bylo signifikantní  $\chi^2(3624) = 86,68$ ,  $p < 0,001$ . Vzhledem ke korekci hladiny významnosti již nepřibýly žádné signifikantní proměnné. Stále byly signifikantní následující proměnné: zaměstnanec obdržel mimořádný bonus ( $p < 0,001$ ;  $z=10,54$ ;  $e^b=13,63$ ), platová třída ( $p < 0,001$ ;  $z=6$ ;  $e^b=3$ ), míra nezaměstnanosti ( $p < 0,001$ ;  $z=-3,98$ ;  $e^b=0,79$ ), počet týmů, které zaměstnanec vystřídal ( $p < 0,001$ ;  $z=7,81$ ;  $e^b=2,1$ ), doba, po kterou je zaměstnanec ve společnosti ( $p < 0,001$ ;  $z=-3,74$ ;  $e^b=0,96$ ), výše platu ( $p < 0,001$ ;  $z=-7,49$ ;  $e^b=0,003$ ), výše celkového příjmu ( $p < 0,001$ ;  $z=-9,81$ ;  $e^b=0,24$ ), rychlost růstu platu ( $p < 0,001$ ;  $z=-3,82$ ;  $e^b=0,77$ ), velikost týmu ( $p < 0,001$ ;  $z=-5,11$ ;  $e^b=0,95$ ) a pohlaví zaměstnance označené 0 pro muže a 1 pro ženy ( $p < 0,001$ ;  $z=-4,53$ ;  $e^b=0,55$ ).

Na základě výše uvedených výsledků byla v datech nalezena opora pro hypotézy 1 (hypotéza o modelu) a dále 48 (o míře nezaměstnanosti), 49 (o počtu týmů), 53 (o době ve společnosti), 56 (o výši platu), 57 (o výši příjmu) a 62 (o rychlosti růstu platu). Ostatní hypotézy nebyly v souladu s našimi daty. Hypotézy 21 a 23 týkající se výše mimořádného bonusu a platové třídy, nebyly v souladu s našimi daty, přestože prediktory vyšly signifikantní. Důvodem byl předpoklad opačného vztahu mezi prediktorem a dobrovolnou fluktuací, než byl nalezen v datech.

Co se týče interpretace výsledků signifikantních proměnných (interpretovány jsou kvůli větší srozumitelnosti poměry šancí), mezi získáním mimořádného bonusu a fluktuací byl nalezen pozitivní vztah, tedy s rostoucí výší mimořádného bonusu, vzrostla pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení bonusu o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec

dobrovolně odejde, oproti tomu, že neodejde, změní 13,63krát; mezi platovou třídou a fluktuací byl nalezen pozitivní vztah, tedy s postupem v platové třídě, vzrostla pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při změně platové třídy ze 3. na 4. při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec dobrovolně odejde, oproti tomu, že neodejde, změní 3krát; mezi mírou nezaměstnanosti a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s vyšší mírou nezaměstnanosti klesá pravděpodobnost dobrovolného odchodu zaměstnance, konkrétně při zvýšení míry nezaměstnanosti o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,79krát; dále mezi počtem vystřídaných týmů a fluktuací byl nalezen pozitivní vztah, tedy s rostoucím množstvím týmů, které zaměstnanec vystřídal, vzrostla pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení počtu vystřídaných týmů o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec dobrovolně odejde, oproti tomu, že neodejde, změní 2,1krát; mezi dobou, po kterou je zaměstnanec ve společnosti, a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s rostoucí dobou klesá pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení doby strávené ve společnosti o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,96krát; mezi výší platu a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s rostoucím platem klesá pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení platu o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,003krát (tento údaj neodpovídá skutečnosti, jelikož plat byl převeden na jinou škálu); mezi výší celkového příjmu a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s rostoucím příjmem klesá pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení příjmu o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,24krát (tento údaj neodpovídá skutečnosti, jelikož příjem byl převeden na jinou škálu); mezi rychlostí růstu platu a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s rychlejším růstem platu klesá pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvýšení platu o jednu jednotku při zafixování ostatních se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,77krát (tento údaj neodpovídá skutečnosti, jelikož plat byl převeden na jinou škálu).

Nakonec nad rámec hypotéz byly nalezeny souvislosti mezi dobrovolnou odchodovostí zaměstnanců a velikostí týmu s pohlavím. Mezi velikostí týmu a dobrovolnou fluktuací byl nalezen negativní vztah, tedy s větším týmem klesá pravděpodobnost zaměstnancova dobrovolného odchodu, konkrétně při zvětšení týmu o jednu jednotku při zafixování ostatních proměnných se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,95krát. U pohlaví změna z muže na ženu snižuje pravděpodobnost odchodu, konkrétně se očekávaný poměr šancí, že zaměstnanec odejde, oproti tomu, že neodejde, změní 0,54krát.



**Graf 8:** Seřazení signifikantních prediktorů dle velikostí efektů

## 8.2 Datově-vědecká část

### Binární logistická regrese

Výsledný model obsahoval následující proměnné: výše mimořádného bonusu ( $p < 0,001$ ;  $z=4,8$ ;  $e^b=6,85$ ), platová třída ( $p < 0,001$ ;  $z=7,43$ ;  $e^b=2,97$ ), míra nezaměstnanosti ( $p < 0,001$ ;  $z=-4,1$ ;  $e^b=0,8$ ), počet vystřídaných týmů zaměstnancem ( $p < 0,001$ ;  $z=8,92$ ;  $e^b=1,96$ ), doba strávená ve společnosti ( $p < 0,001$ ;  $z=-3,38$ ;  $e^b=0,97$ ), výše platu ( $p < 0,001$ ;  $z=-8,42$ ;  $e^b=0,002$ ), výše příjmu ( $p < 0,001$ ;  $z=-12,31$ ;  $e^b=0,26$ ), rychlost růstu platu ( $p < 0,001$ ;  $z=-3,66$ ;  $e^b=0,82$ ), pohlaví ( $p < 0,001$ ;  $z=-5,77$ ;  $e^b=0,48$ ), velikost týmu ( $p < 0,001$ ;  $z=-5,09$ ;  $e^b=0,96$ ), zkušenosti zaměstnance ( $p < 0,001$ ;  $z=3,56$ ;  $e^b=1,66$ ) pro jednotkovou změnu mezi nezkušenými proti referenční skupině zaměstnanců, co nedodali informaci o vzdělání, navýšení bonusu ( $p < 0,001$ ;  $z=-3,47$ ;  $e^b=0,71$ ), množství poboček, které zaměstnanec vystřídal ( $p < 0,001$ ;  $z=-2,72$ ;  $e^b=0,61$ ) pro změnu mezi dvěma pobočkami oproti jedné referenční pobočce. Tento model vysvětlil 24,3 % variance dle McFaddenova R.

Výsledný model lze interpretovat podobně jako výsledky statisticko-inferenční části (váhy prediktorů jsou kvůli větší srozumitelnosti opět uvedeny jako poměry šancí). Tento model pouze obsahuje navíc zkušenosti zaměstnance, navýšení bonusu a množství poboček, které zaměstnanec vystřídal. U zkušeností je signifikantní úroveň nezkušených oproti referenční skupině zaměstnanců, kteří informaci o zkušenosti nedodali; poměr šancí zaměstnancova odchodu vůči tomu, že neodejde, se zvýší 1,66krát. U navýšení bonusu o jednotku (zde je opět třeba hledět na to, že jsou hodnoty převedeny na jinou škálu), se poměr šancí sníží 0,71krát a se změnou pobočky mezi jednou a dvěma se sníží 0,68krát.

Výchozí hodnota tvořící hranici mezi zaměstnanci, u kterých je predikováno, že odejdou, a mezi zaměstnanci, u kterých je predikováno, že neodejdou činí 0,5 ( $P=0,5$ ). Při této hranici dle plochy pod křivkou ROC (AUC) model vykazuje excelentní kvalitu diskriminace (0,82). Při dané hranici model charakterizuje relativně nízká senzitivita (0,35) – tedy model neidentifikuje dobře zaměstnance, kteří odejdou; střední preciznost (0,69) – model predikuje u relativně velkého počtu zaměstnanců, že odejdou, přestože tito lidé ve skutečnosti neodejdou; výborná specificita (0,95) – model dobře identifikuje zaměstnance kteří neodejdou; a relativně dobrá přesnost (0,81) – model celkově identifikuje zaměstnance relativně dobře.

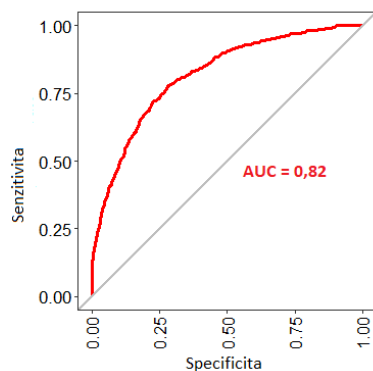
Predikce/Skutečnost	Neodešel	Odešel
Neodešel	2233	459
Odešel	113	248

**Tabulka 2:** Matice záměn pro logistickou regresi

Statistika	Hodnota
Senzitivita	0,35
Specificita	0,95
Přesnost	0,81
Preciznost	0,69

**Tabulka 3:** Charakteristiky logistické regrese

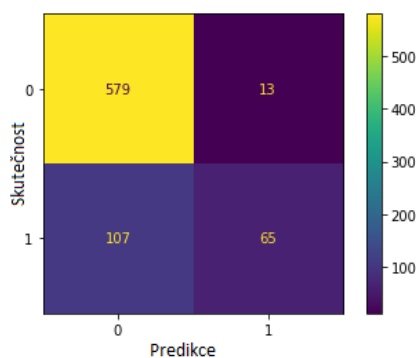




**Graf 9:** ROC a AUC pro logistickou regresi

### ***Gradient Boosting***

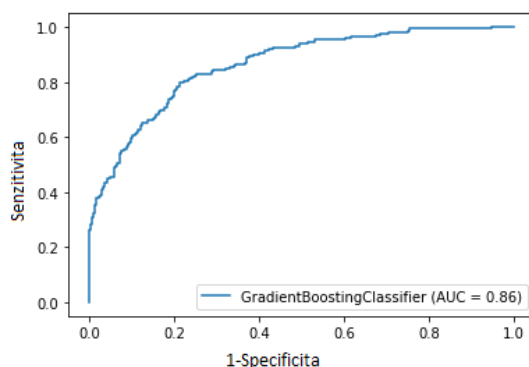
Dle plochy pod křivkou ROC (AUC) model vykazuje excelentní kvalitu diskriminace, přičemž výše AUC (0,86) je nejlepší ze všech modelů vytvořených v rámci této práce. Při hranici 0,5 ( $P = 0,5$ ) model charakterizuje relativně vysoká senzitivita (0,83) – tedy model relativně dobře identifikuje zaměstnance, kteří odejdou; nízká preciznost (0,38) – model predikuje u velkého počtu zaměstnanců, že odejdou, přestože tito lidé ve skutečnosti neodejdou; relativně vysoká specifická (0,84) – model relativně dobře identifikuje zaměstnance kteří neodejdou; a relativně vysoká přesnost (0,84) – model celkově identifikuje zaměstnance relativně dobře.



**Tabulka 4:** Matice záměn pro *Gradient Boosting*

Statistika	Hodnota
Senzitivita	0,83
Specificita	0,84
Přesnost	0,84
Preciznost	0,38

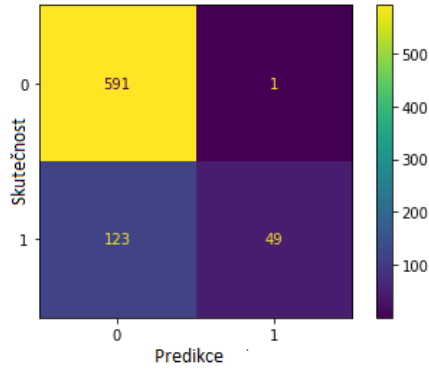
**Tabulka 5:** Charakteristiky *Gradient Boostingu*



**Graf 10:** ROC a AUC pro *Gradient Boosting*

### ***XGBoosting***

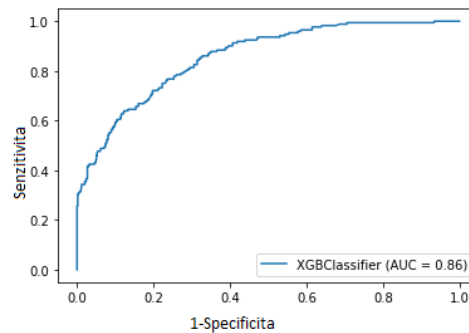
Dle plochy pod křivkou ROC (AUC) model vykazuje excelentní kvalitu diskriminace (0,86), která je stejně tak jako u modelu vytvořeného na základě *Gradient Boostingu* nejlepší ze všech modelů. Hranice mezi skupinami, do kterých jsou zaměstnanci klasifikováni, činí 0,5 ( $P=0,5$ ). Při dané hranici model charakterizuje relativně velmi vysoká senzitivita (0,98) – model výborně identifikuje zaměstnance, kteří odejdou; nízká preciznost (0,28) – model predikuje u vysokého počtu zaměstnanců, že odejdou, přestože tito lidé ve skutečnosti neodejdou; relativně vysoká specifická (0,82) – model dobře identifikuje zaměstnance kteří neodejdou; a relativně dobrá přesnost (0,84) – model celkově klasifikuje zaměstnance relativně dobře.



**Tabulka 6:** Matice záměn pro *XGBoosting*

Statistika	Hodnota
Senzitivita	0,98
Specifická	0,82
Přesnost	0,84
Preciznost	0,28

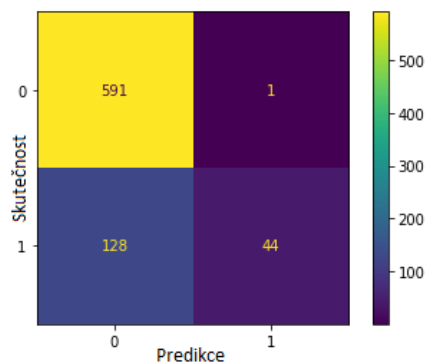
**Tabulka 7:** Charakteristiky *XGBoostingu*



**Graf 11:** ROC a AUC pro *XGBoosting*

## Náhodné lesy

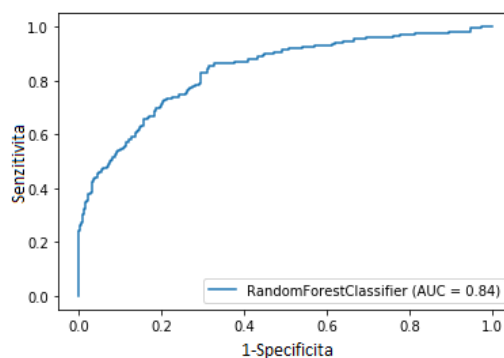
Dle plochy pod křivkou ROC (AUC) model vykazuje excelentní kvalitu diskriminace (0,84). Hranice mezi skupinami činí 0,5 ( $P=0,5$ ). Při dané hranici model charakterizuje relativně velmi vysoká senzitivita (0,98) – model výborně identifikuje zaměstnance, kteří odejdou; nízká preciznost (0,26) – model predikuje u vysokého počtu zaměstnanců, že odejdou, přestože tito lidé ve skutečnosti neodejdou; relativně vysoká specifická (0,82) – model dobře identifikuje zaměstnance kteří neodejdou; a relativně dobrá přesnost (0,83) – model celkově klasifikuje zaměstnance relativně dobře.



**Tabulka 8:** Matice záměn pro náhodné lesy

Statistika	Hodnota
Senzitivita	0,98
Specifická	0,82
Přesnost	0,83
Preciznost	0,26

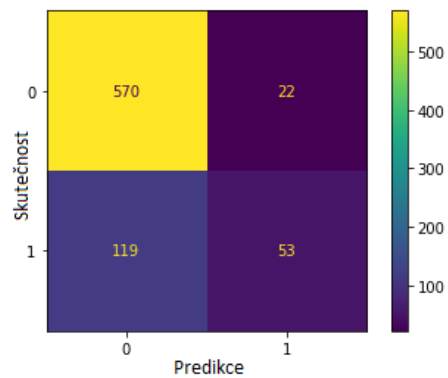
**Tabulka 9:** Charakteristiky náhodných lesů



**Graf 12:** ROC a AUC pro náhodné lesy

## Metoda podpůrných vektorů

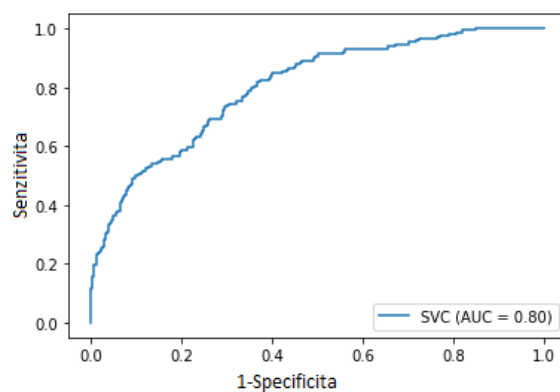
Dle plochy pod křivkou ROC (AUC) model vykazuje akceptovatelnou až excelentní kvalitu diskriminace (AUC = 0,8). Avšak ze všech modelů se jedná o nejslabší model dle AUC. Při hranici 0,5 (P=0,5) model charakterizuje střední senzitivita (0,71) – model přijatelně identifikuje zaměstnance, kteří odejdou; nízká preciznost (0,31) – model predikuje u vysokého počtu zaměstnanců, že odejdou, přestože tito lidé ve skutečnosti neodejdou; relativně vysoká specifická (0,83) – model dobře identifikuje zaměstnance kteří neodejdou; a relativně dobrá přesnost (0,82) – model celkově klasifikuje zaměstnance relativně dobře.



**Tabulka 10:** Matice záměn pro metodu podpůrných vektorů

Statistika	Hodnota
Senzitivita	0,71
Specifická	0,83
Přesnost	0,82
Preciznost	0,31

**Tabulka 11:** Charakteristiky metody podpůrných vektorů



**Graf 13:** ROC a AUC pro metodu podpůrných vektorů

## 9. Diskuse

Významná část této práce měla dva cíle – prověřit vztahy mezi vysvětlujícími proměnnými a fluktuací v konkrétní společnosti a sestrojil predikční model, který bude s to identifikovat zaměstnance ohrožené odchodem. V rámci diskuse bude zhodnoceno, zda a do jaké míry se podařilo tyto cíle naplnit, jaké jsou možnosti aplikace zjištěných informací, a nakonec, jaké měl výzkum limity.

### 9.1 Statisticko-inferenční část

V rámci statisticko-inferenční části bylo testováno celkem 62 hypotéz, přičemž jedna byla vztažena k modelu, zbylých 61 odpovídalo jednotlivým prediktorům. Z těchto hypotéz byly v souladu s našimi daty hypotéza o modelu a dále hypotézy 48 (o míře nezaměstnanosti), 49 (o počtu týmů), 53 (o době ve společnosti), 56 (o výši platu), 57 (o výši příjmu) a 62 (o rychlosti růstu platu). Pro hypotézu 21 (o výši mimořádného bonusu) a 23 (o platové třídě) nebyla nalezena opora v datech, jelikož ve výsledném modelu byl detekován opačný vztah, než bylo původně předpokládáno. Nad rámec hypotéz byly v datech zaznamenány další dva signifikantní prediktory, a to velikost týmu a pohlaví zaměstnance.

Hypotéza 48 předpokládala snížení dobrovolné pravděpodobnosti odchodu zaměstnanců při růstu míry nezaměstnanosti. Tento vztah byl v datech nalezen, což je v souladu s předchozími zjištěními, dle kterých mají zaměstnanci při nižší míře nezaměstnanosti více možností, a proto jsou více ohroženi dobrovolným odchodem (Rubenstein, Eberly, Lee, & Mitchell, 2018; Speer, Dutta, Chen, & Trussell, 2019).

Hypotéza 49 se vztahovala k počtu týmů, které jedinec vystřídal. Předpokladem byla vyšší pravděpodobnost odchodu u zaměstnanců, kteří častěji mění týmy. Podkladem pro tuto hypotézu byla předchozí zjištění, dle kterých zaměstnanci, kterým se střídají kolegové, jsou v zaměstnání méně zakotveni, což souvisí s vyšší pravděpodobností jejich dobrovolného odchodu. Zjištění v této práci jsou v souladu s těmito dřívějšími zjištěními (Heavey, Holwerda, & Hausknecht, 2013).

Hypotéza 53 předpokládala, že zaměstnanci, kteří jsou delší dobu ve společnosti, budou méně ohroženi odchodem. Možným kauzálním vztahem pro tuto souvislost je vyšší závazek vůči společnosti u zaměstnanců, kteří jsou ve společnosti delší dobu. Zjištění v této práci je v souladu s dřívější literaturou (Rubenstein, Eberly, Lee, & Mitchell, 2018).

Hypotézy 56, 62 a 67 se vztahovaly k výši příjmu zaměstnanců. Pro všechny tři hypotézy, tedy pro výši platu a příjmu a pro rychlost růstu platu byl předpokládán negativní vztah s dobrovolnou fluktuací zaměstnanců, což je v souladu s předchozí literaturou (Rubenstein, Eberly, Lee, & Mitchell, 2018). Zaměstnanci s vyšším platem by měli být méně ohroženi změnou práce pro jinou, lépe placenou pozici. Naopak v rozporu s těmito zjištěními je zamítnutí hypotéz 21 a 23, které se také vztahovaly k finančnímu ohodnocení zaměstnanců, ale v opačném vztahu, než bylo předpokládáno. Toto zjištění by mohlo poukazovat např. na kvadratický vztah mezi financemi a dobrovolnou fluktuací zaměstnanců.

U proměnné pohlaví zaměstnanců v rámci této práce vyšla vyšší pravděpodobnost dobrovolného odchodu u mužů. Tento poznatek se liší např. oproti Rubensteinovi a kol. (2018), který ve své metaanalýze nezaznamenal signifikantní vztah pro výše zmíněnou proměnnou. Navíc je zjištění v rozporu s Heavey a kol. (2013), který se domnívá, že ženy mají vyšší pravděpodobnost odchodu kvůli přetrvávajícím tendencím žen odcházet na částečné úvazky kvůli povinnostem souvisejícím s rodinou. Z tohoto důvodu se jeví jako důležité se v budoucnu tímto vztahem více zabývat v souvislosti s jinými faktory jako např. v interakci s poměrem mužů a žen v týmu zaměstnance, v souvislosti s osobnostními charakteristikami nebo v souvislosti se změnou vnímání ženské a mužské role.

Nakonec pro proměnnou velikost týmu vyšla nižší pravděpodobnost odchodu pro větší týmy. Toto zjištění je mimořádně zajímavé, jelikož zcela neodpovídá teoriím, na základě kterých je optimální mít v týmu asi 5-7 lidí, jelikož ve větších týmech dochází k většímu množství komplikací, např. k těžší koordinaci (Armstrong, 2007), což jsou faktory, které by mohly způsobovat nepohodu a stres, a tedy naopak souviset s vyšší odchodovostí zaměstnanců (Zimmerman, Swider, Woo, & Allen, 2016).

Co se týče hypotéz, pro které v datech nebyla nalezena opora, překvapivé zjištění je, že mezi statisticky nesignifikantní prediktory spadal objektivní výkon, který je považován za relativně silný prediktor fluktuace, a dále také prediktory týkající se podobnosti zaměstnanců ve vztahu k demografickým charakteristikám.

Vzhledem k velkému počtu hypotéz a tomu uzpůsobené korekci hladiny významnosti se ve výsledcích objevilo velké množství prediktorů, které nesplňovaly požadovanou hranici pro  $p$ -hodnotu, ale dosahovaly  $p$ -hodnoty pod 0,05. Tyto proměnné jsou zde pro zajímavost uvedeny – jedná se o počet vystřídaných pozic ( $p=0,01$ ;  $z=-2,57$ ;  $e^b=0,573$ ) a poboček ( $p$

=0,001;  $z=-3,76$ ;  $e^b=0,66$ ), délka odpracovaných přesčasů ( $p =0,01$ ;  $z=-2,47$ ;  $e^b=0,98$ ), změna výše bonusu ( $p <0,001$ ;  $z=3,77$ ;  $e^b=6,32$ ); navýšení bonusu ( $p =0,04$ ;  $z=-2,05$ ;  $e^b=0,8$ ); existence partnera ( $p =0,028$ ;  $z=-2,2$ ;  $e^b=0,76$ ); nejvyšší kategorie pro hodnocení objektivního výkonu oproti referenční nejnižší kategorii ( $p =0,047$ ;  $z=-1,98$ ;  $e^b=0,68$ ); čerpání jednoho z produktů ( $p =0,02$ ;  $z=-2,29$ ;  $e^b=0,81$ ), intenzita čerpání jednoho z produktů ( $p =0,03$ ;  $z=2,15$ ;  $e^b=1,000002$ ); sestoupení v platové třídě ( $p =0,04$ ;  $z=2,02$ ;  $e^b=2,5$ ), délka dobrovolných absolvovaných školení za poslední rok ( $p =0,01$ ;  $z=-2,49$ ;  $e^b=0,01$ ).

## Limity práce

Velkým limitem práce je absence dat, která by se více zaměřovala na psychologické aspekty zaměstnání. Na základě dřívější literatury je patrné, že psychologické aspekty jsou pro retenci zaměstnanců velmi důležité. Např. sociálním vztahům v zaměstnání je v této práci mezi prediktory věnována nejdelší podkapitola, což reflektuje množství studií, které se na toto téma zaměřují. Z této oblasti stojí za zmínku např. důležitost vztahu s nadřízeným, vzájemná podpora mezi zaměstnanci nebo celková kvalita vztahů na pracovišti, které se jeví jako důležité prediktory odchodovosti (Wells & Welty Peachey, 2011; Speer, Dutta, Chen, & Trussell, 2019; Heavey, Holwerda, & Hausknecht, 2013). Další důležité psychologické aspekty, které v datech chyběly jsou charakteristiky práce, mezi které lze řadit např. míru autonomie, zmocnění nebo vlivu na dění ve společnosti, které souvisí s motivací zaměstnanců a taktéž ovlivňují pravděpodobnosti zaměstnancova odchodu (Allen, Shore, & Griffeth ;Heavey, Holwerda, & Hausknecht, 2013; Liu, Zhang, Wang, & Lee, 2011). To samé platí pro postoje zaměstnanců, ať už se jedná o jejich spokojenost nebo angažovanost (Rubenstein, Eberly, Lee, & Mitchell, 2018), zde je ovšem těžší data propojit s odchodovostí, jelikož většinou bývají anonymní.

Dalším limitem práce je zaměření pouze na přímé vztahy mezi prediktory a závislou proměnnou, jelikož jednotlivé prediktory mohou ovlivňovat vztahy jiných prediktorů. Např. plat jakožto prediktor dobrovolné odchodovosti zaměstnanců může být ovlivněn tím, do jaké míry jsou zaměstnanci s jeho výší spokojeni, nebo tím, zda jeho výší zaměstnanec považuje za spravedlivou (Heavey et al., 2013; Nyberg, 2010; Rubenstein et al., 2018). Práci by tedy jistě obohatilo větší zaměření na možné moderující a mediační proměnné a také na interakce.

Diskutabilní je také možnost generalizovat získané výsledky na jiné populace. Důvodem je, že tento výzkum byl prováděn na populaci osobních bankéřů, kterým jistě odpovídají



specifické charakteristiky, a nejsou tedy reprezentativním vzorkem. Výsledky jsou tedy uplatnitelné především pro společnost, která data poskytl.

Dalším možným limitem je velké množství prediktorů. Množství prediktorů sice odpovídalo velikosti vzorku, ovšem vzhledem k nutnosti korigovat  $p$ -hodnotu možná nebyly zachyceny některé vztahy, které by byly bývaly zachyceny při menším množství prediktorů.

V rámci výzkumu bylo také zváženo víceúrovňového modelování na základě porovnání míry fluktuace v jednotlivých regionech. Zajímavé by ovšem bylo při porovnávání prověřit míru fluktuace ve vztahu k jednotlivým pobočkám nebo týmům, jelikož na takové úrovni by se mohla míra fluktuace lišit, ačkoliv na úrovni regionů žádný rozdíl nalezen nebyl. Kromě toho potřeba využití víceúrovňového modelování nebyla adekvátně ověřena a vyžaduje hlubší analýzu.

## 9.2 Datově-vědecká část

V rámci datově vědecké části bylo natrénováno 5 modelů (binární logistická regrese, *Gradient Boosting*, *XGBoosting*, náhodné lesy a metoda podpůrných vektorů), které dosahovaly podobných charakteristik kvality. Modely ve všech případech vykazovaly excelentní diskriminační schopnost ( $AUC \geq 0,8$ ). Co se týče dalších charakteristik, pro zhodnocení kvality modelu zaměřeného na predikci odchodů zaměstnanců je stěžejní především preciznost a senzitivita. Všechny modely při výchozí hranici mezi skupinami (0,5) vykazovaly nízké hodnoty preciznosti – nejlepší hodnota charakterizovala model logistické regrese (0,69) a nejhorší model náhodných lesů (0,26). U senzitivity modely dosahovaly lepších výsledků – nejlepších pro model náhodných lesů a *XGBoosting* (0,98) a nejhorší pro model logistické regrese (0,35). Z tohoto důvodu modely, i přes excelentní diskriminační schopnost dle AUC, nejsou považovány za extrémně kvalitní. Vzhledem ke vztahu mezi precizností a senzitivitou bychom zvýšením hranice dosáhli lepších výsledků pro preciznost na úkor senzitivity. Takový kompromis se nabízí např. u náhodných lesů a *XGBoostingu*, kde bylo dosaženo výborných hodnot u senzitivity, ale nízkých u preciznosti. Vyvážení těchto dvou charakteristik bychom mohli docílit pomocí optimalizačních úloh, což je rozepsáno níže. Po vyvážení by modely mohly vykazovat přijatelnou kvalitu. Z tohoto důvodu lze konstatovat, že implementace modelu je pro klienta užitečná v rámci jeho personální praxe.

## **Limity**

Limitem této části je netransparentnost většiny modelů strojového učení, kdy je obtížné zjistit, které vysvětlující proměnné jsou zodpovědné za predikce. Při rozsáhlejší analýze v budoucnu by mohlo být užitečné rozkrýt strukturu vztahů v modelu, aby byly modely lépe interpretovatelné. Pro snazší interpretaci je v práci podrobněji rozepsán model logistické regrese, na základě kterého je možno podat konkrétní doporučení a který nedostatky strojového učení do jisté míry kompenzuje. Dále výkonnost modelů snižovala jejich přetrénovanost, jelikož modely vykazovaly výrazně lepších výsledků na trénovacím datovém souboru. Co se týče chybějících hodnot nahrazených střední hodnotou proměnné a proměnných převedených kvůli chybějícím hodnotám na kategorické, tento problém by v budoucnu mohl být řešen i jinými, přesnějšími způsoby imputace.

Co se týče modelu, který byl vytvořen pomocí binární logistické regrese, je zjevným limitem kvalita modelu ve vztahu k vysvětlené varianci (McFaddenovo  $R = 0,24$ ). Výsledný model oproti nulovému vysvětluje dobrovolnou fluktuaci zaměstnanců pouze částečně, výsledné pravděpodobnosti tedy nelze brát doslova, jelikož v modelu chybí další neznámé faktory, které mohou dobrovolnou fluktuaci zaměstnanců ovlivňovat.

Kvalita modelů by mohla být vylepšena pomocí optimalizace hodnoty hranice mezi zaměstnanci, které model klasifikuje jako ty, co odejdou, a zaměstnanci, které model klasifikuje jako ty, co neodejdou. Cílem je dosažení takové hodnoty hranice, aby společnost vynaložila co nejnižší náklady za retenční program a dosáhla co největších úspor. Tento problém by mohl být řešen např. pomocí lineárního programování.

## **Doporučení pro společnost**

Vzhledem k tomu, že byly reportovány individuální pravděpodobnosti odchodu, je důležité zmínit, jakým způsobem by s výsledky mělo být nakládáno. V těchto případech je doporučeno zaměřit se v rámci intervence spíše na prediktory fluktuace, nikoliv přímo na samotné zaměstnance, jelikož z takových intervencí budou těžit nejen zaměstnanci ohrožení odchodem, ale též zaměstnanci, kteří ohrožení nejsou. Kromě toho, by se individuální pravděpodobnosti neměli dostat do rukou přímých nadřízených zaměstnanců, jelikož znalost individuálních pravděpodobností odchodu by se mohla negativně odrazit na jejich chování vůči zaměstnancům.

Důležitou informací je, že vzhledem k tomu, že do modelu byly zahrnuty demografické proměnné, není možno tento model využívat např. k náboru zaměstnanců nebo k rozhodování ohledně osudu zaměstnance v rámci zkušební doby, jelikož by mohlo dojít k diskriminaci některých kandidátů a zaměstnanců.

Další důležitou částí diskuse jsou doporučení, která se odvíjí od výsledků modelů. Vzhledem k tomu, že transparentní je pouze model logistické regrese, vztahují se poskytnutí doporučení pouze k němu.

Co se týče konkrétních výsledků ve finálním modelu, vyšlo mnoho proměnných, které negativně souvisí s dobrovolnou fluktuací a které se vztahují k příjmům zaměstnance, konkrétně výše celkového příjmu, rychlost růstu platu, výše mimořádného bonusu a navýšení bonusu. Ovšem růst příjmu nelze navyšovat do nekonečna, navíc izolovaný vliv peněz je velmi diskutabilní. Z toho důvodu by mohlo být pro společnost zajímavé zaměřit se na nefinanční benefity jako např. uznání zaměstnanců, zodpovědnost, autonomie nebo osobní růst. Dle Armstronga (2017) může být uznání, které vyvěrá ze základních potřeb člověka, pro zaměstnance stejně hodnotné jako finanční odměny. U rychlosti růstu platu, což je proměnná, která vznikla prostřednictvím změny platu ku době, kterou zaměstnanec ve společnosti strávil, by mohlo být zajímavé zaměřit se na to, zda jsou zaměstnanci odměňováni adekvátně k době, kterou v zaměstnání strávili a zda jejich platy jsou spravedlivě dorovnány vůči zaměstnancům, kteří nastoupili později.

Další proměnná, která se v modelu ukázala jako signifikantní ve vztahu k fluktuaci, je míra nezaměstnanosti, s jejíž růstem se pravděpodobnost fluktuace zaměstnanců snižuje. Situace na trhu je bohužel interně těžko ovlivnitelná, přesto tato informace může být společností využita v rámci plánování pracovní síly (tzv. *Workforce Planning*), kdy společnost optimalizuje např. kapacity zaměstnanců nebo náklady na ně na základě hospodářského cyklu.

Vzhledem k tomu, že v modelu vyšla doba, po kterou jsou zaměstnanci ve společnosti, jakožto negativní prediktor fluktuace, nabízí se věnovat zaměstnancům větší podporu v počáteční fázi jejich pobytu ve společnosti. Jak již bylo zmíněno dříve v práci, možným pojítkem mezi dobou strávenou ve společnosti a dobrovolnou fluktuací je závazek, kteří zaměstnanci pocítují vůči společnosti. Konkrétně se tedy nabízí zaměřit se u nových zaměstnanců na budování tohoto závazku, případně je možné jej monitorovat pomocí

průzkumů zaměřených na postoje zaměstnanců (angažovanost, spokojenost, identifikace apod.).

Další proměnné, které v modelu vyšly signifikantní je počet týmů a poboček, které zaměstnanec vystřídal. Vzhledem k tomu, že s vyšším počtem vystřídaných týmů roste pravděpodobnost dobrovolného odchodu zaměstnance, mohlo by pomoci, kdyby zaměstnancům, kteří z nějakého důvodu mění tým, byla dopřána podpora, aby se mohli na změny lépe adaptovat a zvýšit svou zakotvenost v týmu. Zajímavým kontrastem je, že změna pobočky naopak pravděpodobnost odchodu snižuje, ale pouze při změně mezi jednou a dvěma pobočkami; tedy neznamená to, že čím více poboček zaměstnanec vystřídá, tím nižší bude pravděpodobnost jeho odchodu.

Vzhledem k zajímavému zjištění, kdy zaměstnanci bez zkušenosti jsou méně ohroženi odchodem než zaměstnanci, kteří informaci o zkušenostech nedodali, by bylo zajímavé zaměřit se více na důvody, které vedou druhou skupinu zaměstnanců k tomuto počínání. Na základě těchto informací by dále bylo možné usuzovat na možné kauzální mechanismy mezi touto skutečností a dobrovolnou fluktuací zaměstnanců.

Co se týče rozdílů v odchodovosti mezi muži a ženami, zatím není možné poskytnout konkrétní doporučení. V budoucnu by mohlo být zajímavé zkoumat možné kauzální mechanismy za touto skutečností prostřednictvím rozhovorů se zaměstnanci (např. prostřednictvím tzv. fokusních skupin). Na základě toho by mohly být poodhaleny nové informace, které by teprve formovaly budoucí doporučení.

Podobné doporučení platí pro výši mimořádného bonusu a platovou třídu. Tyto dvě proměnné vyšly statisticky signifikantní, avšak v opačném vztahu k fluktuaci, než bylo na základě předchozí literatury předpokládáno. Z toho důvodu by bylo cenné získat informaci týkající se okolností, od jakých se výše mimořádného bonusu odvíjí (např. zda je mimořádný bonus určen na základě výkonu zaměstnance). Podobně u platové třídy je třeba získat další informace ohledně toho, na základě čeho jsou zaměstnanci řazeni. Je možné, že vyšší odchodovost zaměstnanců, kteří se nachází ve vyšší platové třídě, nesouvisí s finančním ohodnocením, ale např. se zkušenostmi nebo dovednostmi zaměstnance, který pak má na pracovním trhu více možností. Tato hypotéza již je v souladu s předchozími studiemi (Maltarich, Nyberg, & Reilly, 2010).

Jak již bylo zmíněno v části o limitech, zajímavé do dalších analýz by mohlo být pro společnost zaměřit se na sběr dat týkajících se psychologických aspektů. Model logistické regrese má velký prostor pro zlepšení to jak v části vztahující se k predikci, tak s ohledem na jím vysvětlenou variabilitu oproti nulovému modelu. Jak bylo uvedeno v kapitole 4.4.1, existuje mnoho psychologických charakteristik zaměstnanců nebo zaměstnání, které vykazují relativně silné souvislosti s odchodovostí zaměstnanců. Z osobních charakteristik se jeví jako zajímavé ve vztahu k fluktuaci emoční stabilita ( $\hat{p} = -0,19$ ), svědomitost ( $\hat{p} = -0,16$ ) a schopnost vypořádávat se se stresem ( $\hat{p} = -0,39$ ), dále z oblasti sociálních vztahu např. kvalita sociálních vztahů ( $\hat{p} = -0,14$ ), podpora, které se zaměstnancovi od kolegů dostává ( $\hat{p} = -0,19$ ) a vzájemná soudržnost ( $\hat{p} = -0,11$ ); z oblasti postojů zaměstnance se ukazují jako důležité angažovanost ( $\hat{p} = -0,20$ ) a spokojenost ( $\hat{p} = -0,28$ ) zaměstnance, dále z charakteristik zaměstnání např. míra zakoušeného stresu ( $\hat{p} = -0,21$ ) a nakonec signálem mohou být i náznaky ze strany zaměstnance např. ve formě absence ( $\hat{p} = 0,23$ ) nebo pozdních příchodů ( $\hat{p} = 0,14$ ).

## Závěr

Tato práce byla psána s nadějí, že v teoretické části čtenáře seznámí s nově vznikající disciplínou HR analytiky, a to se zaměřením na retenci zaměstnanců. Cílem výzkumné části bylo aplikovat tyto poznatky u konkrétní společnosti. Pro tuto společnost byly poodhaleny vztahy mezi dobrovolnou fluktuací zaměstnanců na pozici osobního bankéře a jejími prediktory. Vedle toho byl sestrojen model, který pomohl odhalit rizikovost odchodů na úrovni jednotlivců. Závěrem lze říci, že vytyčené cíle byly i přes své limity naplněny.

Literárně přehledová část uvedla čtenáře do problematiky HR analytiky jakožto nově vznikající disciplíny, do kontextu jejího vývoje a možností její aplikace. Dále zde byly rozepsány možné metodologické rámce spolu se zdroji dat. První část práce pak vyústila v podrobný popis HR analytického projektu se zaměřením na retenci zaměstnanců dle modelu CRISP-DM. Nakonec v poslední části byly rozepsány negativní důsledky fluktuace zaměstnanců a její možné prediktory.

Empirická část v návaznosti na literárně přehledovou část představila HR analytický projekt v konkrétní společnosti, větší české bance. Tento projekt byl zaměřen na retenci zaměstnanců, tedy na podchycení signifikantních prediktorů jejich dobrovolné fluktuace. Kromě toho bylo utvořeno několik modelů, které pro zaměstnance predikují individuální pravděpodobnosti odchodu. V rámci statisticko-inferenční části se podařilo najít oporu v datech pro 6 hypotéz spolu s hypotézou o signifikanci modelu, ty se týkaly míry nezaměstnanosti, počtu vystřídaných týmů, doby strávené ve společnosti, výše platu a celkového příjmu a rychlosti růstu platu, ostatních 55 bylo zamítnuto. Mezi zamítnutými hypotézami byla i platová třída a výše mimořádného bonusu, jejichž signifikance vyšla v opačném směru, než bylo očekáváno. Nad rámec hypotéz byla zaznamenána významnost u pohlaví a velikosti týmu. V rámci datově-vědecké části bylo sestaveno 5 predikčních modelů, které předpovídají odchody jednotlivých zaměstnanců, všechny modely vyjma jednoho vykazovaly excelentní diskriminační vlastnosti dle metriky AUC.

Jak teoretická, tak empirická část práce zvyšuje povědomí o možných způsobech využití pokročilých analytických nástrojů ke zpracování dat v rámci personální praxe. Práce tak může být inspirací a návodem pro případné zájemce o tuto oblast.

## Seznam použité literatury

- Aguinis, H., Boyd, B. K., Pierce, C. A., Short, J. C., Huselid, M. A., & Becker, B. E. (2010). Bridging Micro and Macro Domains: Workforce Differentiation and Strategic Human Resource Management. *Journal Of Management*, 37(2), 421-428.
- Ajit, P. (2016). Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. *International Journal Of Advanced Research In Artificial Intelligence*, 2016(9), 22-26.
- Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining Talent: Replacing Misconceptions With Evidence-Based Strategies. *Academy Of Management Perspectives*, 24(2), 48-64.
- Allen, D. G., Hancock, J. I., Vardaman, J. M., & Mckee, D. 'lisa N. (2014). Analytical mindsets in turnover research. *Journal Of Organizational Behavior*, 35(1), 61-86.
- Allen, D. G., Shore, L. M., & Griffeth, R. W. (2016). The Role of Perceived Organizational Support and Supportive Human Resource Practices in the Turnover Process. *Journal Of Management*, 29(1), 99-118.
- Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1-11.
- Arellano, C., DiLeonardo, A., & Felix, I. (2017). Using people analytics to drive business performance: A case study. *Mckinsey Quarterly*, 6, 1-6.
- Armstrong, M. (2007). *Řízení lidských zdrojů: nejnovější trendy a postupy: 10. vydání* (10. ed.). Praha: Grada.
- Bliss, W. G. (2004). Cost of employee turnover. *The Advisor*.
- Bodie, M. .T., Cherry, M. A., McCormick, M. L., & Tang, J. (2016). The Law and Policy of People Analytics. *University Of Colorado Law Review*, 6, 1-80.
- Boudreau, J. W., & Jesuthasan, R. (2011). *Transformative HR: How Great Companies Use Evidence-Based Change for Sustainable Advantage*. Jossey-Bass.
- Breugh, J. A. (2014). Predicting Voluntary Turnover from Job Applicant Biodata and Other Applicant Information. *International Journal Of Selection And Assessment*, 22(3), 321-332.
- Breugh, J. A., & Starke, M. (2016). Research on Employee Recruitment: So Many Studies, So Many Remaining Questions. *Journal Of Management*, 26(3), 405-434.

- Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, 118(1), 2-16.
- Burt, R., & Janicik, G. (1996). Social contagion and social structure. *Networks In Marketing*, 32-49.
- Cohen, G., Blake, R. S., & Goodman, D. (2016). Does Turnover Intention Matter? Evaluating the Usefulness of Turnover Intention Rate as a Predictor of Actual Turnover Rate. *Review Of Public Personnel Administration*, 36(3), 240-263.
- Collini, S. A., Guidroz, A. M., & Perez, L. M. (2015). Turnover in health care: the mediating effects of employee engagement. *Journal Of Nursing Management*, 23(2), 169-178.
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal Of Applied Psychology*, 96(3), 443-456.
- Devi, P. S., & Umadevi, D. B. (2018). A Novel Approach to Control the Employee's Attrition Rate of an Organization. *International Journal Of Computer Science And Mobile Applications*, 6(7), 43-52.
- Dhar, V. (2013). Data science and prediction. *Communications Of The Acm*, 56(12), 64-73.
- DiClaudio, M. (2019). People analytics and the rise of HR: how data, analytics and emerging technology can transform human resources (HR) into a profit center. *Strategic Hr Review*, 18(2), 42-46.
- Ellingson, J. E., Tews, M. J., & Dachner, A. M. (2016). Constituent attachment and voluntary turnover in low-wage/low-skill service work. *Journal Of Applied Psychology*, 101(1), 129-140.
- Elsinghorst, S. (2018). Machine Learning Basics - Gradient Boosting & XGBoost. Shirin's Playground [Blog post]. Retrieved June 29, 2020, from [https://www.shiringer.de/2018/11/ml\\_basics\\_gbm/](https://www.shiringer.de/2018/11/ml_basics_gbm/)
- Felts, W., Mitchell, T. R., Hekman, D. R., Lee, T. W., Holtom, B. C., & Harman, W. S. (2009). Turnover Contagion: How Coworkers' Job Embeddedness and Job Search Behaviors Influence Quitting. *Academy Of Management Journal*, 52(3), 545-561.
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, California: Sage.



- Goldstein, H. W., Pulakos, E. D., Passmore, J., & Carrilho Lopes Santarém Semedo, C. S. (2017). *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention*.
- Goodman, S., & Svyantek, D. (1999). Person–Organization Fit and Contextual Performance: Do Shared Values Matter. *Journal Of Vocational Behavior*, 55(2), 254-275.
- *Google Trends*. (2020). Retrieved June 29, 2020, from trends.google.com
- Guenole, N., Ferrar, J., & Feinzig, S. (2017). *The power of people: Learn how successful organizations use workforce analytics to improve business performance*. United States: Pearson Education.
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium. *Journal Of Management*, 26(3), 463-488.
- Haines, V. Y., Jalette, P., & Larose, K. (2010). The Influence of Human Resource Management Practices on Employee Voluntary Turnover Rates in the Canadian Non Governmental Sector. *Industrial And Labor Relations Review*, 63(2), 228-246.
- Hamilton, J. (1994). *Time series analysis* (1st Edition). New Jersey: Princeton.
- Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2012). Meta-Analytic Review of Employee Turnover as a Predictor of Firm Performance. *Journal Of Management*, 39(3), 573-603.
- Hansell, S. (2007). Google answer to filling jobs is an algorithm [Blog post], Retrieved June 29, 2020, from <https://www.nytimes.com/2007/01/03/technology/03google.html>
- Harmady, J. (2015). *Hodnocení spotřebitelských úvěrů pomocí SVM techniky s využitím sw nástroje Mathematica* [Diplomová práce]. Západočeská Univerzita v Plzni.
- Harris, J. G., Craig, E., & Light, D. A. (2011). Talent and analytics: new approaches, higher ROI. *Journal Of Business Strategy*, 32(6), 4-13.
- Harrison, D., Newman, D., & Roth, P. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioral outcomes and time sequences. *Academy Of Management Journal*, 49(2), 305-325.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd edition). Springer.

- Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal Of Applied Psychology, 98*(3), 412-453.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd edition). Hoboken, New Jersey: Wiley.
- Huffman, A. H., Casper, W. J., & Payne, S. C. (2014). How does spouse career support relate to employee turnover? Work interfering with family and job satisfaction as mediators. *Journal Of Organizational Behavior, 35*(2), 194-212.
- Huselid, M. (1995). The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. *Academy Of Management Journal, 38*(3), 635-672.
- Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How Does Human Resource Management Influence Organizational Outcomes? A Meta-analytic Investigation of Mediating Mechanisms. *Academy Of Management Journal, 55*(6), 1264-1294.
- Kraimer, M. L., Seibert, S. E., Wayne, S. J., Liden, R. C., & Bravo, J. (2011). Antecedents and outcomes of organizational support for development: The critical role of career opportunities. *Journal Of Applied Psychology, 96*(3), 485-500.
- Kular, S., Gatenby, M., Soane, E., Rees, C., & Truss, K. (2008). *Employee engagement: A literature review*. Kingston University.
- Lee, E., & Wang, J. (2003). *Statistical methods for survival data analysis* (3rd edition). New York: John Wiley.
- Liao, H., Chuang, A., & Joshi, A. (2008). Perceived deep-level dissimilarity: Personality antecedents and impact on overall job attitude, helping, work withdrawal, and turnover. *Organizational Behavior And Human Decision Processes, 106*(2), 106-124.
- Liu, D., Zhang, S., Wang, L., & Lee, T. W. (2011). The effects of autonomy and empowerment on employee turnover: Test of a multilevel model in teams. *Journal Of Applied Psychology, 96*(6), 1305-1316.
- Loh, W. Y. (2011). Classification and regression trees. *Wires Data Mining And Knowledge Discovery, 1*(1), 14-23.
- Maltarich, M. A., Nyberg, A. J., & Reilly, G. (2010). A conceptual and empirical analysis of the cognitive ability–voluntary turnover relationship. *Journal Of Applied Psychology, 95*(6), 1058-1070.

- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal Of Human Resource Management*, 28(1), 3-26.
- Mondore, S., Douthitt, S., & Carson, M. (2011). Maximizing the impact and effectiveness of HR Analytics to drive business outcomes. *People And Strategy*, 34(2), 20-27.
- Morrow, P. C., McElroy, J. C., Laczniak, K. S., & Fenton, J. B. (1999). Using Absenteeism and Performance to Predict Employee Turnover: Early Detection through Company Records. *Journal Of Vocational Behavior*, 55(3), 358-374.
- Ng, T. W. H., & Feldman, D. C. (2009). Re-examining the relationship between age and voluntary turnover. *Journal Of Vocational Behavior*, 74(3), 283-294.
- Nyberg, A. (2010). Retaining your high performers: Moderators of the performance–job satisfaction–voluntary turnover relationship. *Journal Of Applied Psychology*, 95(3), 440-453.
- Obecné nařízení o ochraně osobních údajů. (18. 03. 2020). Wikipedie: Otevřená encyklopedie. Retrieved June 29, 2020, from <https://cs.wikipedia.org/>
- Osborne, J. W. (2013). *Best practices in data cleaning: a complete guide to everything you need to do before and after collecting your data*. Thousand Oaks, California: SAGE.
- Paauwe, J., & Farndale, E. (2017). *Strategy, HRM, and performance: a contextual approach* (2nd edition). Oxford University Press.
- Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal Of Applied Psychology*, 98(2), 268-309.
- Pieper, J. R. (2015). Uncovering the Nuances of Referral Hiring: How Referrer Characteristics Affect Referral Hires' Performance and Likelihood of Voluntary Turnover. *Personnel Psychology*, 68(4), 811-858.
- Porter, C. M., Woo, S. E., & Campion, M. A. (2016). Internal and External Networking Differentially Predict Turnover Through Job Embeddedness and Job Offers. *Personnel Psychology*, 69(3), 635-672.
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.
- Reina, C. S., Rogers, K. M., Peterson, S. J., Byron, K., & Hom, P. W. (2017). Quitting the Boss? The Role of Manager Influence Tactics and Employee Emotional

Engagement in Voluntary Turnover. *Journal Of Leadership & Organizational Studies*, 25(1), 5-18.

- Rousseau, D. M., & Barends, E. G. R. (2011). Becoming an evidence-based HR practitioner. *Human Resource Management Journal*, 21(3), 221-235.
- Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, 71(1), 23-65.
- Santra, A. K., & Josephine Christy, C. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering. *International Journal Of Computer Science Issues*, 9(2), 322-328.
- Shrivastava, S., Nagdev, K., & Rajesh, A. (2018). Redefining HR using people analytics: the case of Google. *Human Resource Management International Digest*, 26(2), 3-6.
- Speer, A. B., Dutta, S., Chen, M., & Trussell, G. (2019). Here to stay or go? Connecting turnover research to applied attrition modeling. *Industrial And Organizational Psychology*, 12(3), 277-301.
- Soukup, P. (2006). Proč užívat hierarchické lineární modely? *Sociologický Časopis*, 42(5), 987–1012.
- Stevens, C. K. (2009). *Handbook of Principles of Organizational Behavior: Indispensable Knowledge for Evidence-Based Management*. Wiley.
- Sun, R., & Wang, W. (2016). Transformational leadership, employee turnover intention, and actual voluntary turnover in public organizations. *Public Management Review*, 19(8), 1124-1141.
- Tekleab, A. G., Bartol, K. M., & Liu, W. (2005). Is it pay levels or pay raises that matter to fairness and turnover? *Journal Of Organizational Behavior*, 26(8), 899-921.
- Thaden, E., Jacobs-Priebe, L., & Evans, S. (2010). Understanding attrition and predicting employment durations of former staff in a public social service organization. *Journal Of Social Work*, 10(4), 407-435.
- Thorsteinsson, E. B., Brown, R. F., & Richards, C. (2014). The Relationship between Work-Stress, Psychological Stress and Staff Health and Work Outcomes in Office Workers. *Psychology*, 5(10), 1301-1311.

- Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review*, 25(2), 188-204.
- van der Laken, P. A. (2018). Data-Driven Human Resource Management: The rise of people analytics and its application to expatriate management. Ridderkerk: Ridderprint BV.
- van Vulpen, E. (2016). 5 Human Resources Models Every HR Practitioner Should Know. Aihr Academy [Blog post]. Retrieved June 30, 2020, from <https://www.digitalhrtech.com/human-resources-models/>
- Vuk, M., & Curk, T. (2006). ROC Curve, Lift Chart and Calibration Plot. *Metodoloski Zvezki*, 3(1), 89-108.
- Wang, W., Newman, D. A., & Dipboye, R. L. (2016). Social Network Contagion in The Job Satisfaction-Intention-Turnover Model. *Academy Of Management Proceedings*, (1)
- Wells, J. E., & Welty Peachey, J. (2011). Turnover Intentions: Do leadership behaviors and satisfaction with the leader matter? *Team Performance Management: An International Journal*, 17(1/2), 23-40.
- Wickham, H. (2014). Tidy Data. *Journal Of Statistical Software*, 59(10), 1-23.
- Wikipedia contributors. (2020). Google. In Wikipedia, The Free Encyclopedia.
- Wikipedia contributors. (2020). McKinsey & Company. In Wikipedia, The Free Encyclopedia.
- Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a Standard Process Model for Data Mining*. London: Springer-Verlag.
- Yang, C., Ma, Q., & Hu, L. (2011). Job embeddedness: a new perspective to predict voluntary turnover. *Nankai Business Review International*, 2(4), 418-446.
- Zhu, X., Seaver, W., Sawhney, R., Ji, S., Holt, B., Sanil, G. B., & Upreti, G. (2016). Employee turnover forecasting for human resource management based on time series analysis. *Journal Of Applied Statistics*, 44(8), 1421-1440.
- Zimmerman, R. D. (2008). Understanding The Impact Of Personality Traits On Individuals' Turnover Decisions: Meta-Analytic Path Model. *Personnel Psychology*, 61(2), 309-348.
- Zuber, F. (2015). Spread of Unethical Behavior in Organizations: A Dynamic Social Network Perspective. *Journal Of Business Ethics*, 131(1), 151-172.

- Český statistický úřad (2017). *Ženy a muži v datech*. Dostupné z: <https://www.czso.cz/documents/10180/45709986/30000417.pdf/1fa799cb-c008-4271-a09c-9035e22923cc?version=1.2>

## Seznam grafů

Graf 1: Počty vyhledávání jednotlivých pojmů, zpracováno na základě dat z Google Trends (červen, 2020).....	7
Graf 2: ROC křivka (vlastní zpracování) .....	28
Graf 3: Výzkumný soubor dle fluktuace zaměstnanců6.3.1 Statisticko-inferenční část ....	56
Graf 4: Ověření linearity mezi spojitými proměnnými a logaritmy podílů šancí predikované proměnné .....	58
Graf 5: Vztah výše příjmu a predikované proměnné.....	60
Graf 6: Vztah změny platu a predikované proměnné .....	60
Graf 7: Rozložení výzkumného souboru dle věku, počtu let strávených na dané pozici a pohlaví .....	62
Graf 8: Seřazení signifikantních prediktorů dle velikostí efektů.....	66
Graf 9: ROC a AUC pro logistickou regresi .....	68
Graf 10: ROC a AUC pro Gradient boosting .....	69
Graf 11: ROC a AUC pro XGBoosting.....	70
Graf 12: ROC a AUC pro náhodné lesy .....	71
Graf 13: ROC a AUC pro metodu podpůrných vektorů.....	72

## Seznam obrázků

Obrázek 1: Osmikrokový model smysluplné analytiky, zpracováno dle Guenole a kol. (2017) .....	15
Obrázek 2: CRISP-DM, zpracováno dle Wirth a Hipp (2000) .....	17



## Seznam tabulek

Tabulka 1: Matice záměn (vlastní zpracování).....	26
Tabulka 2: Matice záměn pro logistickou regresi .....	67
Tabulka 3: Charakteristiky logistické regrese .....	67
Tabulka 4: Matice záměn pro Gradient boosting .....	68
Tabulka 5: Charakteristiky Gradient boostingu.....	69
Tabulka 6: Matice záměn pro XGBoosting .....	70
Tabulka 7: Charakteristiky XGBoostingu .....	70
Tabulka 8: Matice záměn pro náhodné lesy .....	71
Tabulka 9: Charakteristiky náhodných lesů .....	71
Tabulka 10: Matice záměn pro metodu podpůrných vektorů.....	72
Tabulka 11: Charakteristiky metody podpůrných vektorů .....	72