

EKONOMICKÁ UNIVERZITA V BRATISLAVE
NÁRODOHOSPODÁRSKA FAKULTA

Evidenčné číslo: 101006/B/2024/36122163607853572

MODELOVANIE ZLYHANIA FIRIEM – POROVNANIE
ŠTATISTICKÝCH METÓD
Bakalárska práca

2024

Samuel Horňan

EKONOMICKÁ UNIVERZITA V BRATISLAVE
NÁRODOHOSPODÁRSKA FAKULTA

MODELOVANIE ZLYHANIA FIRIEM – POROVNANIE
ŠTATISTICKÝCH METÓD

Bakalárska práca

Študijný program: Financie, bankovníctvo a poisťovníctvo

Študijný odbor: Ekonómia a manažment

Školiace pracovisko: Katedra financií

Vedúci záverečnej práce: Mgr. Marek Káčer, PhD.

Čestné vyhlásenie

Čestne vyhlasujem, že záverečnú prácu som vypracoval samostatne a že som uviedol všetku použitú literatúru.

Dátum:

.....

Podpis

Pod'akovanie

V tejto časti by som chcel poďakovať vedúcemu mojej záverečnej práce Mgr. Marekovi Káčerovi, PhD. za cenné rady, pripomienky a čas, ktorý mi venoval pri vypracovaní tejto záverečnej práce.

ABSTRAKT

HORŇAN, Samuel: *Modelovanie zlyhania firiem – porovnanie štatistických metód* – Ekonomická univerzita v Bratislave, Národohospodárska fakulta; Katedra financií. – Vedúci záverečnej práce: Mgr. Marek Káčer, PhD. – Bratislava: NHF EU, 2024, 53s.

Táto bakalárska práca sa zameriava na vytvorenie viacerých predikčných modelov zlyhania firiem s využitím rôznych štatistických metód. Ako štatistické metódy boli použité: lineárna diskriminačná analýza, logistická regresia a náhodný les rozhodovacích stromov. Tieto metódy boli natrénované a otestované na vzorke dát vybraných slovenských firiem z rokov 2015 – 2017.

Práca je rozdelená na dve časti. V teoretickej časti je charakterizovaný podnik, zlyhanie firiem, finančná analýza podnikov, pomerové ukazovatele, Altmanove diskriminačné ukazovatele a tiež je popísané použitie strojového učenia na zostavenie predikčných modelov. V praktickej časti sa pracovalo s poskytnutou vzorkou dát. Najprv sa pracovalo s nespracovanými údajmi, z ktorých sa vypočítali vybrané pomerové ukazovatele. Potom boli zostrojené a natrénované predikčné modely s použitím vyššie zmienených štatistických metód. Ich účinnosť a výkon bol otestovaný pomocou matice zámen a ROC krivky.

KĽÚČOVÉ SLOVÁ

Zlyhanie firmy, lineárna diskriminačná analýza, logistická regresia, náhodný les rozhodovacích stromov, strojové učenie

ABSTRACT

This bachelor thesis focuses on the development of corporate failure prediction using a variety of statistical methods. The statistical methods that were used in this study are: linear discriminant analysis, logistic regression and random forest decision trees. These methods were trained and tested on a sample of data from selected Slovak firms from 2015 – 2017.

The thesis is divided into two parts. The theoretical part includes the characterization of firms, corporate failure prediction, financial analysis of enterprises, financial ratio indicators, Altman discriminant indicators, and the use of machine learning to build prediction models. The practical part includes the analysis of the provided sample of data. First, the raw data was analyzed, from which selected ratio indicators were calculated. Then the prediction models were built, trained and tested using the statistical methods mentioned above. Their efficiency and performance on a sample of test data was tested using a confusion matrix and ROC curve.

KEY WORDS

Corporate failure, linear discriminant analysis, logistic regression, random forest decision trees, machine learning

OBSAH

1	Súčasný stav problematiky doma a v zahraničí	11
1.1	Vymedzenie pojmu podnik a firma	11
1.1.1	Riziko podnikania	13
1.1.2	Kríza a zlyhanie firmy	13
1.2	Finančná analýza podniku	15
1.2.1	Zdroje pre finančnú analýzu	15
1.2.2	Metódy finančnej analýzy.....	16
1.3	Metódy prognózovania finančnej situácie podniku	17
1.3.1	Výber ukazovateľov	17
1.3.2	Metódy integrácie vybraných ukazovateľov do diskriminantu	18
1.4	Metódy bodového hodnotenia.....	18
1.5	Diskriminačná analýza	18
1.5.1	Beaverov jednozmerný model diskriminačnej analýzy.....	19
1.5.2	Altmanova viacrozmerná diskriminačná analýza	19
1.6	Strojové učenie (Machine learning)	21
1.7	Logistická regresia	23
1.8	Rozhodovacie stromy	25
1.8.1.1	Náhodné lesy rozhodovacích stromov	26
1.9	Evaluácia predikčného modelu	27
1.9.1	Matica zámen	27
1.9.2	ROC krivka	29
2	Cieľ práce	31
3	Metodika práce a metódy skúmania.....	32
3.1	Dáta a ich spracovanie.....	34

3.2	Vytvorenie predikčných modelov	35
3.3	Analýza a zhodnotenie predikčných modelov	37
4	Výsledky a diskusia.....	38
4.1	Charakteristika zlyhaných a nezlyhaných spoločností	38
4.2	Použité pomerové ukazovatele.....	41
4.2.1	Výpočet a deskriptívna štatistika vybraných pomerových ukazovateľov	43
4.3	Evalúácia zostavených predikčných modelov	44
4.3.1	Matica zámen	45
4.3.2	ROC krivka	48
Záver		51
Zoznam použitej literatúry		52

ÚVOD

Zlyhanie firmy je problémom, ktorý ovplyvňuje nielen jednotlivé podniky, ale aj celé ekonomiky. Preto je potrebné už vopred sledovať faktory, ktoré by mohli predikovať budúce nepriaznivé podmienky. Na Slovensku, kde je kapitálový trh menej rozvinutý a len veľmi malé množstvo spoločností má verejne obchodované cenné papiere, sa väčšina podnikov spolieha na účtovné výkazy. Tieto výkazy poskytujú bohaté informácie o minulých a súčasných stavoch podnikov, ktoré sa môžu použiť pri zostavovaní predikčných modelov. Na ich zostavenie sa využíva rozsiahle množstvo štatistických metód. Zostavené modely nachádzajú široké uplatnenie v mnohých ekonomických odvetviach, v ktorých samotným firmám pomáhajú identifikovať riziká a optimalizovať ich finančné rozhodnutia, čím prispievajú k zlepšeniu celkovej finančnej stability a riadenia rizika v ekonomike.

Strojové učenie patrí medzi pomerne novú technológiu, ktorá pomaly transformuje spôsob, akým firmy a ekonomické odvetvia fungujú. V kontexte modelovania zlyhania firiem, strojové učenie prináša nový rozmer analýzy. Na rozdiel od tradičných štatistických metód, ktoré sa často spoliehajú na preddefinované predpoklady a ručne vybrané ukazovatele, strojové učenie dokáže automaticky identifikovať a využiť komplexné vzťahy v dátach, ktoré nie sú hneď identifikovateľné. To umožňuje vytvárať predikčné modely s vyššou presnosťou a robustnosťou. Strojové učenie navyše umožňuje neustále vylepšovanie modelov prostredníctvom učenia sa z nových dát, v dôsledku čoho modely zostávajú relevantné a presné aj v dynamicky sa meniacom podnikateľskom prostredí.

1 Súčasný stav problematiky doma a v zahraničí

1.1 Vymedzenie pojmu podnik a firma

Podnik je komplexný pojem, ktorý má rôzne definície. Vo všeobecnosti sa charakterizuje ako subjekt, ktorý systematicky vykonáva ekonomické aktivity za účelom dosiahnutia zisku.¹ Tento pojem je mimo toho vymedzený tiež v právnom systéme SR – v Obchodnom zákonníku, § 5, nasledovne²: „*Podnik je súbor hmotných, ako aj osobných a nehmotných zložiek podnikania.*“

Firma je obdobou slova podnik. V ekonomike sa oba pojmy považujú za synonymá. V praxi je podnik menšia podjednotka vlastnená firmou. V zásade je firma právna entita, ktorá môže prevádzkovať jeden alebo viacero podnikov, zatiaľ čo podnik je len samotná organizačná jednotka, ktorá spadá pod firmu.

Podniky sa z hľadiska veľkosti delia na malé, stredné a veľké. Medzi základné kritéria, na základe ktorých sa podniky rozdeľujú podľa veľkosti, sú:³

- počet zamestnancov
- obrat
- hodnota majetku
- zisk

Podľa odporúčania komisie č. 2003/361/ES sa podniky delia do 3 skupín. Jednotlivé druhy podnikov, vrátane faktorov podľa ktorých ich rozlišujeme, sú uvedené v Tabuľke 1.

¹ MAJDÚCHOVÁ, H. a kol. *Podnikové hospodárstvo*. 2. Wolters Kluwer, 2020. 14 s. ISBN 978-80-571-0271-7.

² Zákon č. 513/1991 Zb. *Obchodný zákonník*, § 5

³ MAJDÚCHOVÁ, H. a kol. *Podnikové hospodárstvo*. 2. Wolters Kluwer, 2020. 123 s. ISBN 978-80-571-0271-7.

Tabuľka 1 Rozdelenie podnikov (Zdroj: vlastné spracovanie)

Rozdelenie podniku	Počet zamestnancov	Ročný obrat/celková súvaha
Mikropodnik	1 – 9	<2 mil EUR
Malý	10 – 49	<10 mil EUR
Stredný	50 – 249	<50 mil EUR
Veľký	250+	>50 mil EUR

Z vyššie vymedzených podnikov, majú pre hospodársku stabilitu a rast najväčší význam malé a stredné podniky (MSP). V EÚ tvoria viac ako 99 % všetkých podnikov a viac ako 2/3 všetkých zamestnaných osôb pracuje v MSP.⁴

Od slova podnik je ďalej odvodené slovo podnikanie, ktoré priamo nadväzuje na definíciu podniku. Podnikaním sa rozumie sústavne vykonávaná ekonomická aktivita za účelom dosiahnutia zisku. Podobne ako pojem podnik, aj slovo podnikanie je zadefinované v Obchodnom zákonníku § 2 odst. 1, kde sa podnikaním rozumie: „*sústavná činnosť vykonávaná samostatne podnikateľom, vo vlastnom mene a na vlastnú zodpovednosť za účelom dosiahnutia zisku.*“

Dosahovanie zisku pre podniky je hlavnou motiváciou ich vzniku a pôsobenia. Zisk sa dosahuje rôznymi spôsobmi, a to predovšetkým predajom produktov alebo poskytovaním služieb spotrebiteľom. Popri tom to môžu byť aj rôzne opatrenia, stratégie alebo iné činnosti, ktoré súvisia priamo alebo nepriamo s dosahovaním ekonomickej efektívnosti daného podniku.⁵

⁴ : Di Bella, L., Katsinis, A., Lagüera-González, J., Odenthal, L., Hell, M., Lozar, B. Annual Report on European SMEs 2022/2023, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/028705, JRC134336.

⁵ MAJDÚCHOVÁ, H. a kol. *Podnikové hospodárstvo*. 2. Wolters Kluwer, 2020. 32 s. ISBN 978-80-571-0271-7.

1.1.1 Riziko podnikania

Riziko v podnikateľskej sfére je chápané ako podnikateľský neúspech, ktorého najčastejším dôsledkom je strata. Každé podnikanie je už od svojho vzniku vystavené riziku. Riziko je to stále a nie je možné, aby úplne zaniklo. Medzi základné riziká podnikania patria ekonomické, finančné, organizačné, technologické, právne a environmentálne faktory. Uplatnením rôznych postupov a opatrení je možné toto riziko znížiť, prípadne eliminovať na minimum.

Nižšie sú vyčlenené niektoré spôsoby, ktoré podniky môžu aplikovať za účelom zníženia rizika:⁶

- diverzifikácia
- transfer rizika
- poistenie rizika

1.1.2 Kríza a zlyhanie firmy

Situácia, v ktorej sa firma výrazne negatívne odchyli od rovnovážneho stavu pôsobenia sa označuje ako krízový stav, prípadne iba kríza. Krízové stavy sa rozlišujú závažné a menej závažné. V prípade závažnej krízy je ohrozená celá existencia firmy. Menej závažné krízy narušujú nejaké čiastkové ciele firmy, no nie je pritom ohrozená celá existencia firmy, ba firma je schopná tieto ohrozenia časom prekonať.

Z časového horizontu, rozlišujeme krízu buď na pomalú alebo rýchlu. Pomalá kríza sa vyvíja postupne a môže byť spojená s rôznymi dlhodobými trendmi, ktoré sa postupne zhoršujú. Príkladom môže byť ekonomická recesia, zánik trhov, postupný rast nákladov, alebo aj fluktuácia pracovnej sily. Rýchla kríza je spôsobená náhlou a nečakanou udalosťou, ktorá vyvoláva okamžitú reakciu. Ako príklad možno uviesť prírodné katastrofy, výpadky dodávok, ale aj štátne zásahy.⁷

⁶ MAJDÚCHOVÁ, H. a kol. *Podnikové hospodárstvo*. 2. Wolters Kluwer, 2020. 16-18 s. ISBN 978-80-571-0271-7.

⁷ SMEJKAL, Vladimír a RAIS, Karel. *Řízení rizik ve firmách a jiných organizacích*. 4., aktualiz. a rozš. vyd. Expert (Grada). Praha: Grada, 2013. 25-26 s. ISBN 978-80-247-4644-9.

Ako už bolo naznačené, závažná kríza môže viesť k zlyhaniu podniku. V prípade zlyhania podniku neexistuje žiadna všeobecne prijatá definícia, ktorá by charakterizovala túto udalosť. Autori, ktorí sa týmto problémom zaoberali vo svojich štúdiách, považovali pojem zlyhanie za binárny jav, pričom sa snažili predovšetkým vymedziť hranicu, kedy je podnik považovaný za zlyhávajúci a kedy ešte nie.⁸

Všeobecne by sme mohli povedať, že zlyhanie je úzko späté s neúspechom v podnikaní, ktorý má za následok prerušenie, prípadne ukončenie činnosti s cieľom zabrániť ďalším stratám. Zlyhanie tiež môže byť definované ako spôsobenie strát veriteľom, ktoré dlžník nie je schopný splatiť.

Okrem toho sa na opísanie zlyhania využívajú aj pojmy ako napr. insolvenca alebo bankrot.⁹ Obe synonymá môžu opisovať istú formu zlyhania, ale pri ich detailnejšej charakteristike oboch pojmov, sa stáva zrejmým rozdiel medzi nimi. Hlavný rozdiel je ten, že insolvenca je širší stav finančnej neschopnosti, zatiaľ čo bankrot je konkrétny právny proces, keď dlžník formálne vyhlasuje, že nie je schopný splatiť svoje dlhy. Insolvenca môže viesť k bankrotu, no nie všetky prípady insolvencie nutne skončia bankrotom.

Insolvenca sa ďalej rozdeľuje na dočasnú alebo trvalú a môže viesť k rôznym právnym postupom, ako sú reštrukturalizácia, konkurz alebo likvidácia.

Ani v právnom rámci Slovenskej republiky nie je presne definované zlyhanie podniku. Namiesto toho sa v Zákone č. 7/2005 Z. z. o konkurze a reštrukturalizácii vymedzuje pojem platobná neschopnosť, ktorá túto udalosť popisuje ako formu úpadku: „Právnická osoba sa podľa zákona považuje za platobne neschopnú, ak nie je schopná plniť 30 dní po lehote splatnosti aspoň dva peňažné záväzky viac ako jednému veriteľovi. Fyzická osoba je platobne neschopná, ak nie je schopná plniť 180 dní po lehote splatnosti aspoň jeden peňažný záväzok.“¹⁰

⁸ VEGANZONES, David a SEVERIN, Eric. Corporate failure prediction models in the twenty-first century: a review. Online. *European Business Review*. 2020, roč. 33, č. 2, s. 204-226. ISSN 0955-534X.

⁹Kücher, A., Mayr, S., Mitter, C. *et al.* Firm age dynamics and causes of corporate bankruptcy: age dependent explanations for business failure. *Rev Manag Sci* **14**, 633–661 (2020).

¹⁰ Zákon č. 7/2005 Z. z. o konkurze a reštrukturalizácii a o zmene a doplnení niektorých zákonov v znení neskorších predpisov.

1.2 Finančná analýza podniku

Prvotným a základným indikátorom zdravia podnikov nám poskytuje finančná analýza. Tá na základe zozbieraných ukazovateľov a údajov poskytuje a odpovedá na otázky nielen ohľadom súčasnej finančnej situácie podniku, ale aj budúcej. Finančná situácia podniku, označovaná tiež ako „viditeľná časť“ podnikovej ekonomiky, odráža hlavný znak schopnosti, alebo neschopnosti plniť záväzky podnikov.

1.2.1 Zdroje pre finančnú analýzu

Existuje nespočetné množstvo zdrojov informácií, ktoré môžu analytici využiť na finančnú analýzu. Medzi najdôležitejší zdroj, patrí účtovná závierka podniku. V nej je zahrnutý celý transformačný proces podniku od jeho začiatkov až po aktuálnu súčasnosť. Z účtovnej závierky sa dozvedáme o realizácii, priebehu a výsledkoch podniku. Popri tom nám údaje z nej poskytujú objektívny pohľad na činnosť podniku v čase a formálne dokumentujú jeho výkony. Údaje musia byť dôkladne zosúladené s príslušnými finančnými štandardmi a právnymi normami a často bývajú podrobené dôkladnej kontrole a verifikácii auditormi.

Na Slovensku je účtovná závierka vymedzená v *Zákone o účtovníctve č. 431/2002 Z.Z.* Zákon nariaďuje účtovným jednotkám účtovať v sústave podvojného účtovníctva, pričom v zákone sú uvedené aj výnimky, na ktoré sa táto povinnosť nevzťahuje. Účtovná závierka má obsahovať:

- súvahu
- výkaz ziskov a strát
- poznámky¹¹

Ako alternatíva k účtovnej závierke môžu analytikom ešte poslúžiť ďalšie iné externé informácie z ekonomického okolia, ktorými môžu byť napr. Obchodný register, Register účtovných závierok, prípadne sa môžu využiť rôzne verejne dostupné ekonomicko-štatistické štúdie.

¹¹ *Zákone o účtovníctve č. 431/2002 Z.Z.*

1.2.2 Metódy finančnej analýzy

Pri finančnej analýze podnikov možno využiť rôzne komplexné metódy, ktoré sa odvíjajú od viacerých ukazovateľov. My sa zameriavame na metódu, ktorá sa pri analýze opiera o súbor relatívnych finančných ukazovateľov. Táto metóda patrí medzi najpoužívanejšie metódy pri hodnotení finančnej situácie podniku. Základom tejto metódy je výpočet relatívnych hodnôt, ktoré sú odvodené z finančných ukazovateľov obsiahnutých v účtovných závierkach. Tie sa delia podľa faktorov uvedených nižšie:¹²

- a) *Schopnosť hradiť záväzky* – závisí od finančnej situácie podniku. Na jej analýzu používame ukazovatele likvidity. Tie vyjadrujú, či má spoločnosť dostatočnú hotovosť na hradenie svojich záväzkov. Obecne platí, že vyššia hodnota týchto ukazovateľov je charakteristická pre spoločnosti, ktoré nie sú v úpadku. Patrí tu napr. bežná likvidita (*current ratio*) alebo okamžitá likvidita (*cash ratio*).
- b) *Úroveň využitia majetku* – zisťuje, či je podnik schopný efektívne investovať svoje prostriedky do aktív a či dosahuje primerané výnosy z týchto investícií. O tom ako efektívne podnik využíva svoje aktíva, hovoria ukazovatele aktivity. Vyššie hodnoty týchto ukazovateľov sú charakteristické pre podniky, ktoré prosperujú. Medzi dosť výrazný ukazovateľ z tejto kategórie, ktorý silne diskriminuje s bankrotom, patrí obrat aktív. Ten popisuje efektívnosť využívania aktív na generovanie tržieb, pričom podniky ktoré neskrachovali, produkujú vyššie tržby na celkové aktíva. Ďalšie významné ukazovatele, ktoré stoja za zmienku, sú obrat obežných aktív, obrat pracovného kapitálu atď.
- c) *Štruktúra finančných zdrojov* – vyjadruje pomer medzi dlhovými a vlastnými zdrojmi financovania, ako aj pomer medzi dlhodobým a krátkodobým financovaním. Tento pomer poskytuje informácie o finančnej stabilite podniku. Popisujú ho ukazovatele zadlženosti. Očakáva sa, že zdravé a prosperujúce spoločnosti budú mať nižšie ukazovatele zadlženosti (finančnej páky) na rozdiel od spoločností v úpadku. Ukazovateľ dlhového zaťaženia (*debt ratio*) je jedným z kľúčových faktorov, ktoré môžu významne ovplyvniť finančnú stabilitu spoločnosti. Jeho vysoká hodnota môže

¹² ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšír. vyd. Bratislava: Sprint 2, 2010. 84-85 s. ISBN 978-80-89393-15-2.

naznačovať potenciálne riziko bankrotu. Medzi ďalšie pomerové ukazovatele patrí pomer dlhu k celkovému majetku, kapitálu k aktívam a iné.

- d) *Celková výnosnosť podniku* – zahŕňa všetky tri vyššie spomenuté ukazovatele do jedného celku. Popisujú ju ukazovatele rentability. Obecne platí, že tieto ukazovatele negatívne korelujú s finančným zlyhaním podniku. Najdôležitejší ukazovateľ, ktorý predpovedá zlyhanie podnikov, je rentabilita aktív (ROA). Ten meria, ako efektívne spoločnosť využíva svoje aktíva na generovanie zisku. Medzi ďalšie dôležité ukazovatele patrí rentabilita vlastného kapitálu (ROE) alebo zisk pred odpočítaním úrokov a zdanením a odpismi (EBITDA).¹³
- e) *Trhová hodnota podniku* – poskytuje pohľad na hodnotu podniku z vonku, či je z hľadiska investícií vhodný pre investorov.¹⁴

1.3 Metódy prognózovania finančnej situácie podniku

Metódy prognózovania finančnej situácie podniku sú dôležitým nástrojom na predpovedanie výsledkov finančnej situácie spoločnosti. Táto hodnota je založená na porovnaní a interpretovaní dosiahnutých aktuálnych hodnôt. Na základe toho vieme s istou pravdepodobnosťou odhadnúť kategóriu prosperujúcich a neprosperujúcich firiem. To sa robí pomocou prevedenia štatisticky významných ukazovateľov do jednočíselného vyjadrovania označovaného ako tzv. diskriminátor, ktorý umožňuje ich zhrnutie. Ako údaje sú použité informácie z účtovných závierok, ktoré sú dostatočné pre väčšinu použitých metodík.

1.3.1 Výber ukazovateľov

Na správny výber ukazovateľov sa využíva niekoľko rozlišovacích metód, z ktorých sa štatistickou analýzou vyberajú štatisticky významné ukazovatele. Na to slúžia buď jednoduché testy deskriptívnej štatistiky, medzi ktoré patrí napr. mediánový test alebo rozličné grafické porovnania. Nie je vylúčené použitie aj komplikovanejších štatistických

¹³ Vuković, Bojana & Milutinović, Sunčica & Milicevic, Nikola & Jaksic, Dejan. (2020). Corporate Bankruptcy Prediction: Evidence from Wholesale Companies in the Western European Countries. Ekonomický časopis. 68. 477-498.

¹⁴ ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšír. vyd. Bratislava: Sprint 2, 2010. 84-85 s. ISBN 978-80-89393-15-2.

metód ako napr. Mann-Whitneyho test, Kolmogorov-Samirnov test ktoré môžu poskytovať lepšiu rozlišovaciu schopnosť vybraných ukazovateľov.¹⁵

1.3.2 Metódy integrácie vybraných ukazovateľov do diskriminantu

Na integráciu vybraných ukazovateľov do diskriminantu sa najčastejšie využívajú metódy vymedzené nižšie:

1. Metódy bodového hodnotenia
2. Matematicko štatistické metódy
 - a. Diskriminačná analýza
 - b. Logistická regresia
3. Neurónové siete

Pre jasnejšie vysvetlenie si popíšeme niektoré z nich.

1.4 Metódy bodového hodnotenia

Metódy bodového hodnotenia sú jednoduché metódy, ktoré hodnotia finančné zdravie podniku. Sú založené na tom, že k hodnotám vybraných finančných ukazovateľov sa priradujú body pridelené na stupnici. Zo súčtu priradených bodov sa predikuje finančný vývoj firmy, pričom v niektorých metodikách sa považuje za pozitívny najväčší počet bodov, kým v iných naopak. Ako príklad bodového hodnotenia možno uviesť napr. Tamariho rizikový index alebo Rýchly test.

1.5 Diskriminačná analýza

Diskriminačná analýza patrí medzi matematicko-štatistickú metódu, ktorá sa používa na klasifikáciu pozorovaní do preddefinovaných skupín na základe jednej alebo viacerých prediktorov. Jej hlavným cieľom je určiť, ktoré premenné najviac prispievajú k separácii skupín. Metóda vypočíta diskriminačné funkcie, ktoré maximalizujú rozdiely medzi priemermi skupín a minimalizujú variabilitu vo vnútri skupín.

¹⁵ ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšír. vyd. Bratislava: Sprint 2, 2010. 84-85 s. ISBN 978-80-89393-15-2.

Na základe počtu zvolených sledovaných ukazovateľom sa rozlišuje:

1. Jednorozmerná diskriminačná analýza:
2. Dvojrozmerná diskriminačná analýza
3. Viacrozmerná diskriminačná analýza

1.5.1 Beaverov jednozmerný model diskriminačnej analýzy

Beaver bol jeden z prvých ekonómov, ktorý zostrojil model s využitím jednorozmernej diskriminačnej analýzy na predpoveď úpadku podniku. Jeho súbor obsahoval 138 podnikov, pričom vzorky zlyhaných a nezlyhaných podnikov boli v pomere 1:1. Pri zostrojení Beaver využil dichotomický model hodnotenia pre jednoduchý t-test v jednorozmernom poli.¹⁶

Analýzou zistil, že významné štatistické rozdiely boli preukázané v týchto ukazovateľoch:

- cash flow/cudzí kapitál
- čistý zisk/celkový kapitál
- cudzí kapitál/celkový kapitál
- čistý prevádzkový kapitál/celkový kapitál
- likvidita
- no credit interval

Po zistení vyššie uvedených ukazovateľov skúmal hodnotu determinantu pre jednotlivé ukazovatele. Súbor rozdelil na dve podskupiny – odhadovaciu a verifikačnú. U oboch skúmal ich spoľahlivosť na základe chybovosti. Zistil, že najnižšiu hodnotu chybovosti mali ukazovatele cash flow/cudzí kapitál, a preto im pridelil najvyššiu diskriminačnú silu.¹⁷

1.5.2 Altmanova viacrozmerná diskriminačná analýza

Altman vo svojej publikácii v roku 1968 ako prvý predstavil model, ktorý dokázal do istej miery predpovedať pravdepodobnosť úpadku podnikov. Na zostavenie tohto modelu využil

¹⁶ HASSAN, Ehsan ul; ZAINUDDIN, Zaemah a NORDIN, Sabariah. A Review of Financial Distress Prediction Models: Logistic Regression and Multivariate Discriminant Analysis. Online. *Indian-Pacific Journal of Accounting and Finance*. 2017, roč. 1, č. 3, s. 13-23. ISSN 2590-406X.

¹⁷ ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšir. vyd. Bratislava: Sprint 2, 2010. 135-136 s. ISBN 978-80-89393-15-2.

súbor dát 66 podnikov, ktoré rozdelil do dvoch skupín po 33, pričom polovicu z nich tvorili podniky, ktoré dosiahli bankrot. Na jeho zostavenie spočiatku pripravil 22 potenciálnych ukazovateľov, ktoré by mohli mať signifikantný význam pri zlyhaní podnikov. Analýzou zistil, že týchto päť ukazovateľov najlepšie vypovedá o budúcej finančnej situácii podnikov:

- X_1 – čistý prevádzkový kapitál/celkový kapitál
- X_2 – nerozdelený zisk/celkový kapitál
- X_3 – zisk pred zdanením + dane/celkový kapitál
- X_4 – tržová hodnota vlastného kapitálu /cudzí kapitál
- X_5 – tržby/celkový kapitál

S využitím viacrozmernej diskriminačnej analýzy Altman zostavil rovnicu, pričom závislú premennú označil ako tzv. Z-skóre.¹⁸ Rovnica mala tvar funkcie pre firmy s akciami obchodovanými na trhu nasledovný:

$$Z = 1,2X_1 + 1,4X_2 + 3,3X_3 + 0,6X_4 + 1,0X_5$$

Výsledná hodnota Z-skóre pritom indikovala:

$Z > 2,99$ – podnik je v dobrej finančnej situácii

$1,81 \leq Z \leq 2,99$ – šedá zóna – nevieme isto predpovedať situáciu podniku

$Z < 1,81$ – podnik sa nachádza v zlej finančnej situácii

Rovnako bola zostavená rovnica pre firmy, ktoré neemitujú akcie na trhu, s tvarom:

$$Z' = 0,717X_1 + 0,847X_2 + 3,107X_3 + 0,420X_4 + 0,998X_5$$

Výsledná hodnota Z-skóre pritom indikovala:

$Z' > 2,90$ – podnik je v dobrej finančnej situácii

$1,20 \leq Z' \leq 2,90$ – šedá zóna – nevieme isto predpovedať situáciu podniku

$Z' < 1,20$ – podnik sa nachádza v zlej finančnej situácii

¹⁸ ALTMAN, Edward I.; ZAINUDDIN, Zaemah a NORDIN, Sabariah. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy: Logistic Regression and Multivariate Discriminant Analysis. Online. *The Journal of Finance*. 1968, roč. 23, č. 4, s. 13-23. ISSN 00221082. Dostupné na: <https://doi.org/10.2307/2978933>. [cit. 2024-02-03].

Pre nevýrobné podniky, bola zostavená rovnica, ktorá mala tvar funkcie:

$$Z'' = 6,56X_1 + 3,26X_2 + 6,72X_3 + 1,05X_4$$

Výsledná hodnota Z-skóre pritom indikovala¹⁹:

$Z'' > 2,60$ – podnik je v dobrej finančnej situácii

$1,10 \leq Z'' \leq 2,60$ – šedá zóna – nevieme isto predpovedať situáciu²⁰ podniku

$Z'' < 1,10$ – podnik sa nachádza v zlej finančnej situácii

1.6 Strojové učenie (Machine learning)

Strojové učenie, angl. *machine learning*, je termín, ktorý bol prvýkrát použitý v roku 1959 americkým informatikom Arthurom Samuelom. Ten ako prvý predstavil návrh, že je možné (programovaním) naučiť zariadenia, aby sami od seba vykonávali úlohy. V súčasnosti *machine learning* patrí do oblasti umelej inteligencie a jeho fundamentálne využitie spočíva v efektívnejšom spracovaní dát. To je založené na tom, že niekedy pri spracovaní dát nevieme nájsť vzťahy medzi premennými a vyextrahovať dôležité informácie.

V súvislosti s analýzou dát sa strojové učenie opiera o širokú škálu algoritmov, ktoré sú aplikované na príslušné dáta za účelom naučenia a riešenia príslušných problémov. Najčastejšie sa aplikujú klasifikačné, regresné alebo zhlukovacie algoritmy.

Klasifikácia v strojovom učení je dvojstupňový proces, ktorý sa skladá z dvoch krokov:

1. vývoj modelu
2. vyhodnotenie modelu

Vývoj modelu strojového učenia je proces, ktorý je primárne založený na rozdelení datasetu na trénovaciu a testovaciu skupinu. V prvotnej fáze sa príslušný algoritmus trénuje na súbore trénovacích dát, aby vytvoril predikčný model. Následne sa potom zo zostavených modelov

¹⁹ ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšír. vyd. Bratislava: Sprint 2, 2010. 142-143 s. ISBN 978-80-89393-15-2.

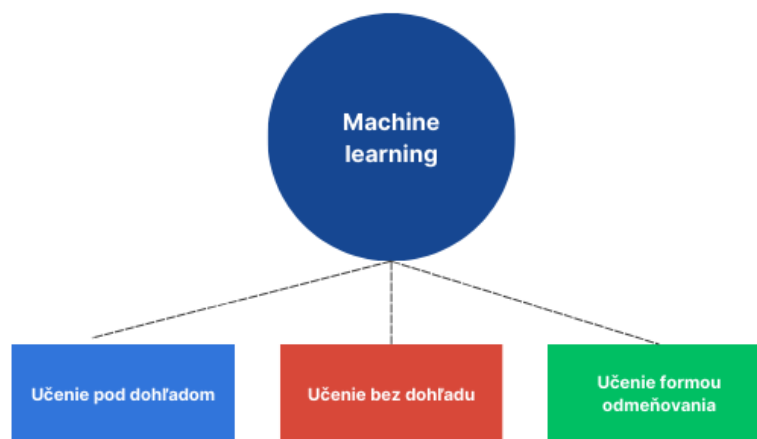
generujú predpovede alebo klasifikácie založené na naučených vzorcoch. Keď sú do tohto modelu následne vložené nové, testovacie dáta, model sa znova snaží predpovedať odozvu.

To sa následne porovnáva z hľadiska presnosti. Celkový úspech modelu sa meria nie len jeho presnosťou na tréningových dátach, ale aj jeho schopnosťou správne klasifikovať nové, neoznačené príklady. V prípade nedostatočnej presnosti, je potrebné model upraviť alebo trénovať na rozsiahlejšom alebo rôznorodejšom súbore dát. Tento proces sa opakuje, kým model nedosiahne požadovanú úroveň presnosti.²¹

Pri výbere algoritmu je tiež dôležité poznamenať, že neexistuje jediný univerzálny algoritmus ktorý by bol aplikovaný na všetky problémy. Z tohto dôvodu sa zostavujú rôzne modely aj s využitím kombinácie rôznych algoritmov, ktoré by mohli produkovať lepšie výsledky. V prípade že sa dosiahne požadovaná presnosť, je možné model s príslušne zvoleným algoritmom aplikovať na ďalšie použitie.²²

V závislosti od celého procesu sa rozlišujú tri základné typy strojového učenia. Tie sú znázornené na Obrázku 1:

Obrázok 1 Základné typy učenia learning (Zdroj: vlastné spracovanie)



²¹ PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.

²² Mahesh, Batta. (2019). Machine Learning Algorithms - A Review. *International Journal of Science and Research* 10.21275/ART20203995.

1. ***Učenie pod dohľadom*** je typ strojového učenia, kde sú nami zvolené algoritmy pod dohľadom a zostavený model trénovaný na vymedzenom datasete. Model je rozdelený na trénovacie dáta a testovacie dáta, pričom sa najprv výstupná premenná natrénuje na a potom sa predikuje alebo klasifikuje. Medzi najznámejšie algoritmy ktoré tu patria sú Logistická regresia, Rozhodovacie stromy alebo Support Vector Machine (SVM). Tento typ učenia bude aplikovaný aj v praktickej časti tejto záverečnej práce.
2. ***Učenie bez dohľadu*** je typ strojového učenia, kde je model trénovaný na neoznačenom datasete a na rozdiel od učenia pod dohľadom tu neexistujú správne odpovede a nieje tu žiaden „učiteľ“. Model sa musí učiť identifikovať vzory a štruktúry v dátach bez akejkoľvek predchádzajúcej znalosti správnej odpovede. V prípade natrénovaného modelu na nové dáta, model pracuje z predošlých naučených vlastností. Medzi najčastejšie používané algoritmy patria Zhhlukovanie (K-means), analýza hlavných komponentov (PCA) a neurónové siete.
3. ***Učenie formou odmeňovania*** je typ strojového učenia, kedy sa softvérový agent sa učí robiť rozhodnutia vykonávaním akcií v prostredí. Ten dostáva odmeny alebo pokuty na základe toho, či splní alebo nesplní úlohu²³. Tento typ učenia sa dá najbližšie prirovnať ľudskému učeniu, pričom v praxi je dosť rozšírený v robotike.²⁴

1.7 Logistická regresia

Logistická regresia je významnou metódou v oblasti štatistiky, ktorá umožňuje modelovať pravdepodobnosť výskytu udalostí s diskretnými výsledkami. Taktiež je ju možné použiť aj z dôvodu nelineárneho vzťahu, pričom nie je nevyhnutné splniť podmienky ako u lineárnych regresných vzťahoch. Vyznačuje sa schopnosťou analyzovať vzťah medzi jednou alebo viacerými nezávislými premennými a binárnou (dichotomickou) závislou

²³ Mahesh, Batta. (2019). Machine Learning Algorithms - A Review. *International Journal of Science and Research* 10.21275/ART20203995.

²⁴ PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.

premennou. Metóda je vhodná na predikciu určitého javu, ktorý môže mať iba 2 možné výsledky – v našom prípade zlyhania firiem – úspech/neúspech.

Logistická regresia odhaduje pravdepodobnosť príslušnosti príkladu k príslušej triede pomocou sigmoidálnej funkcie – tiež označovaná ako S-krivka. Nadobúda tu hodnoty rozsahu (0; 1), čo zodpovedá pravdepodobnosti výsledku.

Funkcia je definovaná ako: ²⁵

$$p\left(y = \frac{1}{X}\right) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$$

Kde:

$p\left(y = \frac{1}{X}\right)$ - pravdepodobnosť, že sa daný jav sa stane za podmienky konkrétnych hodnôt vysvetľujúcich nezávislých premenných x

$X, \dots, X_n$ – vstupné premenné

$\beta_0, \beta_1, \dots, \beta_n$ – konštanty lineárneho regresného výsledku

Logitovou transformáciou je možné transformovať vzťah medzi logaritmom šanci a vektorom vysvetľujúcich premenných z nelineárneho na lineárny. Tým sa hodnoty pravdepodobností prevedú na reálnu os. Táto transformácia je definovaná ako prirodzený logaritmus z pomeru pravdepodobnosti udalosti. Vyjadrujeme ju nasledovne: ²⁶

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln(e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

V súvislosti so strojovým učením spadá algoritmus logistickej regresie pod techniku učenia pod dohľadom. Zostavený model vychádza z princípov logistickej regresie a má schopnosť

²⁵ POPOOLA, Oluwatoyin Muse Johnson. Preface to the Third Issue of Indian-Pacific Journal of Accounting and Finance. Online. Indian-Pacific Journal of Accounting and Finance. 2017, roč. 1, č. 3, s. 1-3. ISSN 2590-406X.

²⁶ ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšir. vyd. Bratislava: Sprint 2, 2010. 152 s. ISBN 978-80-89393-15-2.

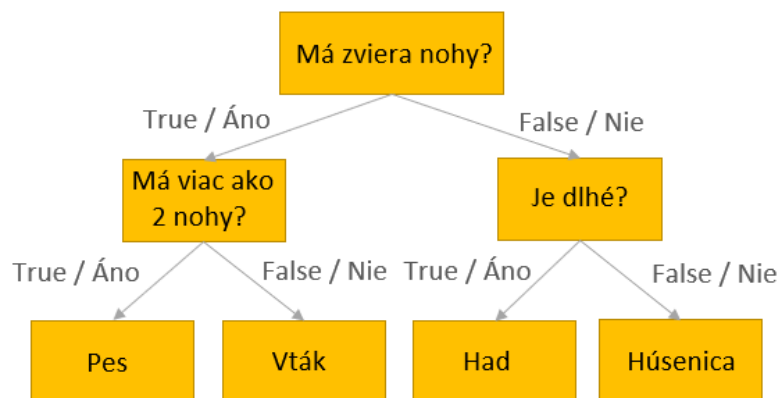
predikovať pravdepodobnosti a klasifikovať nové údaje pomocou spojitých a diskretných súborov údajov.²⁷

1.8 Rozhodovacie stromy

Rozhodovacie stromy (angl. *decision trees*) patria medzi klasifikačno-regresné algoritmy v oblasti strojového učenia. Už z názvu vyplýva, že rozhodovacie stromy sú založené na hierarchickej štruktúre podobnej stromu. Rozhodovacie stromy sa začínajú od koreňa a každý uzol v strome predstavuje rozhodnutie alebo test na určitý atribút, zatiaľ čo vetvy predstavujú výsledky týchto testov. Listové uzly, ktoré sú koncovými bodmi stromu, predstavujú konečné rozhodnutia alebo predpovede.

Užitočnosť rozhodovacích stromov spočíva v tom, že výsledná klasifikácia poskytuje informácie v ľahko čitateľnej a zrozumiteľnej forme, ktorá často pripomína ľudské rozhodovanie. Štruktúra rozhodovacieho stromu s jednoduchým príkladom je znázornená na Obrázku 2.²⁸

Obrázok 2 Štruktúra rozhodovacieho stromu vysvetlená na príklade²⁹



²⁷ MUSA, Abdallah Bashir. Comparative study on classification performance between support vector machine and logistic regression. Online. *International Journal of Machine Learning and Cybernetics*. 2013, roč. 4, č. 1, s. 13-24. ISSN 1868-8071. Dostupné na: <https://doi.org/10.1007/s13042-012-0068-x>.

²⁸ PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.

²⁹ ING. MICHAL CHMULÍK, PhD. Náhľad do algoritmu rozhodovací strom. *Umelá Inteligencia.sk* [online]. Dostupné na: <https://umelainteligencia.sk/nahlad-do-algoritmu-rozhodovaci-strom/>

Existujú rôzne typy klasifikačných algoritmov rozhodovacích stromov, ktoré sa používajú v strojovom učení. Medzi najznámejšie patrí ID3 (Iterative Dichotomizer 3), C4.5 (Classification 4.5) a CART (Classification And Regression Trees). Každý z týchto algoritmov má svoju vlastnú metódu pre konštrukciu rozhodovacieho stromu a vytváranie predpovedí.

Pri ID3 a C.5 sú vetvy stromu tvorené z informácií, ktoré sa merajú pomocou entropie. Entropia vyjadruje, do akej miery je systém informácií usporiadaný pomocou vzniknutých vetví.

Informačný zisk pre atribút A a množinu S je možné vyjadriť týmto vzorcom:³⁰

$$IG(A, S) = H(S) - \sum_{t \in T} p(t)H(t)$$

Vysvetlivky:

$H(S)$ – entropia množiny S

$p(t)$ – pravdepodobnosť výskytu hodnoty t atribútu A v množine S

$H(t)$ – entropia množiny príkladov S , ktoré majú hodnotu t pre atribút A

$\sum$ – suma pre všetky možné hodnoty t , ktoré môže atribút A nadobúdať

Algoritmus končí, keď sa vyčerpajú atribúty alebo sú všetky príklady klasifikované.

1.8.1.1 Náhodné lesy rozhodovacích stromov

Náhodné lesy sú algoritmom strojového učenia, ktorý trénuje viacero rozhodovacích stromov súčasne na vzorkách údajov a náhodných podmnožinách prvkov, čím zabraňuje nadmernému prispôsobeniu a stráca schopnosť generalizácie na nových dátach. Tieto lesy sa dajú použiť na klasifikačné aj regresné problémy, pričom výsledky sa buď vyberú podľa väčšinovej kategórie alebo sa spriemerujú. Vďaka technike nazývanej bootstrapping sú efektívne najmä

³⁰ PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.

pri klasifikačných problémoch. Ich významnou výhodou je schopnosť spracovať chýbajúce hodnoty.³¹

1.9 Evaluácia predikčného modelu

Evaluácia predikčných modelov je proces, pri ktorých sa hodnotí efektívnosť a presnosť zostrojených predikčných modelov. Tento proces nám umožňuje identifikovať, kde model exceluje a kde bude potrebovať zmeny. Existuje mnoho možností, ako si túto výkonnosť overiť. Tie najčastejšie si popíšeme v nasledujúcich kapitolách.

1.9.1 Matica zámen

Matica zámen (angl. *confusion matrix*) je matica s rozmermi $N \times N$, ktorá sa používa na vyhodnotenie výkonnosti algoritmov v zostavenom predikčnom modeli. Jedna dimenzia matice predstavuje skutočnú triedu, zatiaľ čo druhá predpovedané hodnoty modelu. Špeciálny prípad tejto matice je dvojrozmerná matica s rozmermi 2×2 , pričom triedy v tejto matici sú označované ako pozitívne alebo negatívne. Bunky v matici sú označené ako skutočne negatívne (TN), falošne pozitívne (FP), falošne negatívne (FN) alebo skutočne pozitívne (TP). Pre lepšiu predstavu je to znázornené v Tabuľke 2:³²

Tabuľka 2 Matica zámen s rozmermi 2×2

Trieda	Negatívne (Predpovedané)	Pozitívne (Predpovedané)
Negatívne (Skutočné)	Skutočne Negatívne (TN)	Falošne Pozitívne (FP)
Pozitívne (Skutočné)	Falošne Negatívne (FN)	Skutočne Pozitívne (TP)

³¹ PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.

³² TING, Kai Ming. Confusion Matrix. Online. *Encyclopedia of Machine Learning*. 2010, s. 209-209. ISBN 978-0-387-30768-8.

Vysvetlivky:

Skutočne Negatívne (TN) – počet správne klasifikovaných negatívnych prípadov.

Falošne Pozitívne (FP) – počet negatívnych prípadov, ktoré boli nesprávne klasifikované ako pozitívne.

Falošne Negatívne (FN) – počet pozitívnych prípadov, ktoré boli nesprávne klasifikované ako negatívne.

Skutočne pozitívne (TP) – počet správne klasifikovaných pozitívnych prípadov.

Z týchto výsledkov je následne možné vypočítať množstvo metrík, medzi ktoré patria:³³

1. Presnosť – vypočíta, koľko správnych predpovedí urobil model. Vypočítame ju pomocou vzorca:

$$\text{Presnosť} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Senzitivita – vypovedá, koľko skutočných pozitívnych prípadov sme vedeli predpovedať v našom modeli. Vypočítame ju pomocou vzorca:

$$\text{Senzitivita} = \frac{TP}{TP + FN}$$

3. Precíznosť – vypovedá, koľko správne predpovedaných prípadov sa ukázalo ako pozitívnych. Vypočítame ju pomocou vzorca:

$$\text{Precíznosť} = \frac{TP}{TP + FN}$$

4. Špecifita – vypovedá o tom, ako dobre sú identifikované negatívne príklady. Vypočítame ju pomocou vzorca:

$$\text{Špecifita} = \frac{TN}{TN + FP}$$

³³ TING, Kai Ming. Confusion Matrix. Online. *Encyclopedia of Machine Learning*. 2010, s. 209-209. ISBN 978-0-387-30768-8.

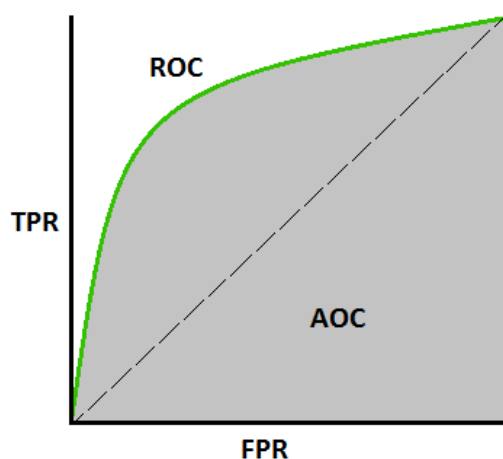
1.9.2 ROC krivka

ROC krivka (angl. *Receiver operating Characteristic*) je graf, ktorý popisuje výkonnosť binárneho klasifikačného modelu. Na osi x sa nachádza relatívna hodnota falošných pozitívnych predpovedí (FPR), ktorá vyjadruje pravdepodobnosť nesprávnej klasifikácie negatívneho prípadu. Na osi y sa nachádza relatívna hodnota skutočne pozitívne predpovedanej (TPR), ktorá vyjadruje pravdepodobnosť správne zaradeného prípadu.

Pomocou ROC krivky je možné vizualizovať výkonnosť modelu pri rôznych zmenách prahovej hodnoty.

Táto krivka, vrátane plochy pod krivkou (AUC), je znázornená na Obrázku 3.

Obrázok 3 ROC krivka³⁴



Z ROC krivky je následne odvodená tzv. plocha pod krivkou (AUC), ktorá je empirickým meradlom klasifikačnej výkonnosti na základe plochy pod krivkou ROC. Táto plocha pod krivkou (AUC) je robustným ukazovateľom schopnosti modelu správne klasifikovať prípady bez ohľadu na konkrétnu prahovú hodnotu.

³⁴NARKHEDE, Sarang. Understanding AUC - ROC Curve. *Medium* [online]. Dostupné na: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

Hodnoty AUC sa pohybujú v rozmedzí od 0 do 1, pričom, maximálna hodnota $AUC = 1$ znamená, že model vie dokonale rozlíšiť rozdiely a každá skupina je klasifikovaná správne. V opačnom prípade, ak by hodnota $AUC = 0$, tak by to signalizovalo, že model vždy predpovedá opačný výsledok oproti skutočnosti a všetky hodnoty sú klasifikované nesprávne. Avšak toto je extrémne nepravdepodobné. V praxi sa hodnoty AUC zvyčajne pohybujú v rozmedzí od 0,5 do 0,9. Hodnota 0,5 naznačuje, že model nemá žiadnu rozlišovaciu schopnosť medzi pozitívnymi a negatívnymi prípadmi, ktorá by prekračovala očakávanie čistej náhody.³⁵

³⁵ FAWCETT, Tom. An introduction to ROC analysis. Online. *Pattern Recognition Letters*. 2006, roč. 27, č. 8, s. 861-874. ISSN 01678655. Dostupné na: <https://doi.org/10.1016/j.patrec.2005.10.010>.

2 Cieľ práce

Cieľom tejto bakalárskej práce je zostrojenie predikčných modelov zlyhania firiem s využitím rôznych štatistických metód. Tie budú aplikované na dátach slovenských firiem, ktoré ešte neboli testované.

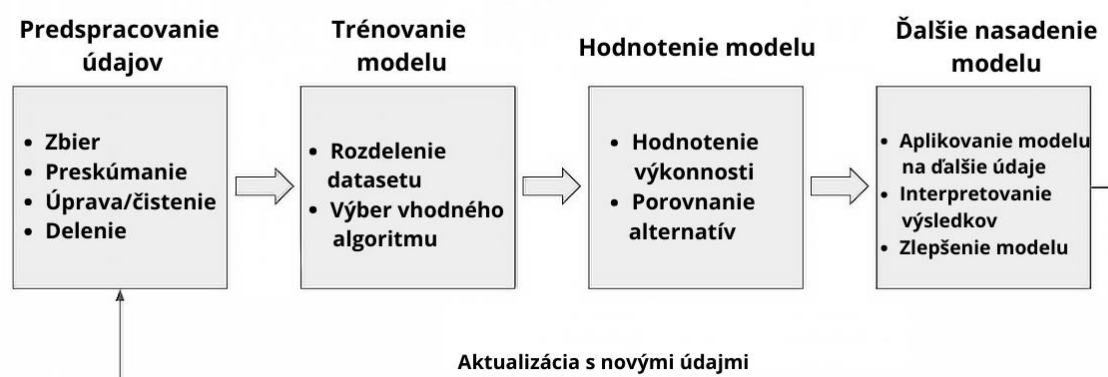
V rámci dosiahnutia tohto cieľa bolo zvolených viac čiastkových cieľov:

- Poskytnúť základné teoretické poznatky o danej problematike.
- Spraviť základnú charakteristiku poskytnutého datasetu a vykonať deskriptívnu analýzu štatistických údajov.
- Vytvoriť modely zlyhania firiem s využitím algoritmov – lineárna diskriminačná analýza, logistická regresia a náhodný les rozhodovacích stromov.
- Vyhodnotiť a porovnať zostavené predikčné modely s využitím príslušných evaluačných metód.

3 Metodika práce a metódy skúmania

Na zostrojenie predikčných modelov zlyhania firiem sme vychádzali zo všeobecnej schémy zostrojenia predikčného modelu s využitím príslušných klasifikačných a regresných algoritmov strojového učenia. Všeobecná schéma predbežného spracovania údajov vrátane tréningu modelov, ich hodnotenia a nasadenia optimalizovaného modelu je znázornená na Obrázku 4.

Obrázok 4 Schéma zostavenia modelu (Zdroj: vlastné spracovanie)



Ako údaje nám boli poskytnuté finančné vybrané ukazovatele, ktoré obsahovali takmer 600 000 riadkov pripravených na ďalšie spracovanie. Údaje boli získané z databázy Finstat. Finančné výkazy poskytovali informácie o malých a stredných podnikoch, z obdobia rokov 2015 – 2017. Konkrétne šlo o podniky, ktoré tvoria značnú časť ekonomiky na Slovensku a v EÚ. Okrem finančných výkazov náš dataset obsahoval aj ďalšie nefinančné informácie o týchto podnikoch, ktoré zahŕňali odvetvie v ktorom podnik pôsobí, počet zamestnancov, právnu formu a dátum vzniku.

Náš dataset tiež obsahoval binárnu premennú nazvanú „default“, ktorá indikovala, či daný podnik zlyhal alebo nie. V prípade, že podnik nezlyhal, táto premenná obsahovala hodnotu 0, a v prípade zlyhania hodnotu 1.

Zhrnutie všetkých ukazovateľov, ktoré boli v nespracovanom datasete je v Tabuľke 3.

Tabuľka 3 Všetky poskytnuté ukazovatele v nespracovanom datasete

Demografické informácie	Súvaha	Výkaz ziskov a strát
Rok	Spolu majetok	Výroba
Default	Tržby z predaja tovaru	Tržby z predaja vlastných výrobkov a služieb
Odvetvie	Neobežný majetok	Výrobná spotreba
Zamestnanci_kategoria	Dlhodobý nehmotný majetok súčet	Mzdové náklady
Zamestnanci_prese	Dlhodobý hmotný majetok súčet	Výsledok hospodárenia z hospodárskej činnosti
Právnaforma	Dlhodobý finančný majetok súčet	Výnosy z dlhodobého finančného majetku
Dátum vzniku	Pôžičky s dobou splatnosti najviac jeden rok	Nákladové úroky
	Obežný majetok	Výsledok hospodárenia z finančnej činnosti
	Zásoby súčet	Výsledok hospodárenia z bežnej činnosti pred zdanením
	Dlhodobé pohľadávky súčet	Výsledok hospodárenia z bežnej činnosti po zdanení
	Krátkodobé pohľadávky súčet	Výsledok hospodárenia z mimoriadnej činnosti pred zdanením
	Finančné účty súčet	Výsledok hospodárenia z mimoriadnej činnosti po zdanení
	Spolu vlastné imanie a záväzky	Výsledok hospodárenia za účtovné obdobie pred zdanením
	Vlastné imanie	Výsledok hospodárenia za účtovné obdobie po zdanení
	Základné imanie súčet	
	Kapitálové fondy súčet	
	Fondy zo zisku súčet	
	Výsledok hospodárenia minulých rokov	
	Výsledok hospodárenia za účtovné obdobie po zdanení	
	Záväzky	
	Rezervy súčet	
	Dlhodobé záväzky súčet	
	Krátkodobé záväzky súčet	
	Krátkodobé finančné výpomoci	
	Bankové úvery	
	Časové rozlíšenie súčet	

Spracovanie, úprava dát a zostrojenie predikčných modelov prebehlo v cloudovom výpočtovom prostredí – Google Colab. Ako programovací jazyk bol zvolený programovací jazyk Python, ktorý je široko využívaný v oblastiach strojového učenia a analýzy dát. V rámci nášho projektu sme využili niekoľko knižníc Pythonu, ktoré sú špecificky navrhnuté pre analýzu dát a strojové učenie. Tieto knižnice nám umožnili efektívne pracovať s dátami, vytvárať a testovať modely a poskytli nám tiež príslušné výsledky popísané v ďalšej kapitole.

V tejto práci boli na prácu s dátami použité a importované tieto knižnice v Pythone:

- *Pandas* – knižnica použitá na analýzu a manipuláciu s dátami
- *NumPy* – knižnica použitá na numerické operácie
- *Scikit-learn* – knižnica použitá na predikciu a analýzu dát, strojové učenie
- *Matplotlib a Seaborn* – knižnica použitá na vizualizáciu dát

3.1 Dáta a ich spracovanie

Správna príprava dát je nevyhnutný proces v správnom zostrojení robustného modelu. Keďže v mnohých prípadoch sa pracuje s veľkým množstvom dát, dataset je často neusporiadaný a obsahuje chýbajúce údaje. Preto je vhodné buď vynechať alebo nahradiť – tzv. imputovať chýbajúce hodnoty pre každý stĺpec vo vstupných údajoch pred ďalším pokračovaním. V prípade imputácie sa najčastejšie aplikuje buď nahradenie priemerom, mediánom, modusom, nie je však vylúčené aj použitie konštantnej hodnoty.

V súvislosti s našim datasetom, sme sa rozhodli chýbajúce ukazovatele vynechať. Rozhodli sme sa tak preto, lebo náš dataset obsahoval aj po vynechaní dostatok údajov na zostrojenie predikčných modelov. Vynechaním chýbajúcich hodnôt sme sa rozhodli predísť skresleniu výsledkov v našom modeli.

Z poskytnutých ukazovateľov v nespracovanom datasete sme vypočítali päť Altmanových finančných pomerových ukazovateľov, ktoré boli použité na zostrojenie predikčných modelov. Z týchto premenných bola tiež spravená základná deskriptívna štatistika, ktorá nám poskytla základné sumárne informácie o našich dátach. Názvy a popisy vypočítaných pomerových ukazovateľov sú znázornené v Tabuľke 4.

Tabuľka 4 Názvy a popis vypočítaných pomerových finančných ukazovateľov

Názov	Popis
X1	Čistý prevádzkový kapitál/celkový kapitál
X2	Nerozdelený zisk/celkový kapitál
X3	Zisk pred zdanením + dane/celkový kapitál
X4	Trhová hodnota vlastného kapitálu /cudzí kapitál
X5	Tržby/celkový kapitál

3.2 Vytvorenie predikčných modelov

Vypočítané pomerové ukazovatele boli ďalej použité ako vstupné premenné pre zostrojenie modelov. Každý zostrojený model bol nastavený tak, aby sa spočiatku „trénoval“ na týchto dátach s cieľom naučiť sa vzťahy medzi nezávislými premennými a závislou – či podnik zlyhal alebo nie (default). Dataset bol rozdelený v pomere 70/30, pričom 70 % dát bolo použitých na tréningovanie a zvyšných 30 % bolo použitých na testovanie.

Po rozdelení dát na tréningovú a testovaciu vzorku sme vykonali škálovanie údajov (angl. *feature scalling*), a to z toho dôvodu, že mnoho algoritmov pri zostavení modelu počíta vzdialenosť za pomoci euklidovskej metriky alebo iných metrík vzdialenosti, ktoré sú citlivé na vysoký rozsah odľahlých hodnôt. Ak sú funkcie na rôznych vzdialenostných škálach, môže to skresliť vzdialenosti a viesť k nesprávnym predpovediam, resp. k vzniku úplne iných modelov.³⁶³⁷

V našej práci sme použili funkciu *StandardScaler*, importovanú z knižnice *sklearn*, ktorý implementuje štandardizáciu, t. j. odstráni priemernú hodnotu a škáluje jednotlivé funkcie na jednotkovú varianciu. Dáta teda boli normalizované na škále [0;1].

Okrem toho bola využitá technika *SMOTE* (*Synthetic Minority Over-sampling Technique*) z knižnice *imblearn* na vyváženie dát v tréningovej triede. Keďže v našom datasete bol niekoľkonásobne nižší počet firiem, ktoré skrachovali, aplikovali sme túto techniku za

³⁶ BEHRENS, T.; SCHMIDT, K.; VISCARRA ROSSEL, R. A.; GRIES, P.; SCHOLTEN, T. et al. Spatial modelling with Euclidean distance fields and machine learning. Online. *European Journal of Soil Science*. 2018, roč. 69, č. 5, s. 757-770. ISSN 1351-0754.

³⁷ *Scikit-learn Machine Learning in Python*. Online. Dostupné na: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html.

účelom zvýšenia modelu pri správnom klasifikovaní menšej triedy. Táto technika nám umožnila synteticky generovať nové príklady v menšej triede, čím sme dosiahli lepšiu rovnováhu medzi triedami a zlepšili sme schopnosť modelu správne klasifikovať príklady z menšej triedy. Z percentuálneho hľadiska sme mali v našom dátovom súbore 99,8368 % nezlyhaných a 0,1632 % zlyhaných prípadov. V prípade ignorovania práce s nerovnovážnym datasetom, dochádza k nedostatočnej reprezentácii menšej triedy a model by sa vôbec nemusel naučiť správne rozpoznať hranicu medzi týmito dvomi triedami. Naopak, príliš prehnané zvýšenie počtu minoritných vzoriek v tréningovom sete môže negatívne ovplyvniť citlivosť modelu a model sa môže stať viac zaujatým pri predikcii nižšej triedy. V našom prípade sme si nastavili *sampling_strategy=0.9*, čiže v tréningovom sete malo 90 % počet vzoriek väčšej triedy. Za účelom dosiahnutia rovnakých opakovateľných výsledkov bol nastavený *random_state* na hodnotu 0.

Po natrénovaní modelu (na tréningových dátach) a realizácii vyššie zmienených opatrení bol model s využitím nami zvolených algoritmov testovaný, ako dobre dokáže predpovedať výsledky na netestovaných dátach. Za účelom dosiahnutia rovnakých opakovateľných výsledkov bol nastavený *random_state* na hodnotu 0.

Na testovanie boli zvolené tieto tieto príslušné algoritmy, ktoré boli importované z knižnice *sklearn*:

1. Lineárna diskriminačná analýza
2. Logistická regresia
3. Náhodný les rozhodovacích stromov

U náhodného lesa rozhodovacích stromov boli nastavené tieto hyperparametre:

- Počet stromov ktoré sa majú vytvoriť bol nastavený na 50 (*n_estimators = 50*)
- Maximálna hĺbka stromov bola nastavená na 10 (*max_depth = 10*)
- Počet vzoriek potrebných na rozdelenie interného uzla (*min_samples_split = 5*)
- Min počet vzoriek na uzol – 2 (*min_samples_leaf = 2*)
- Ako kritérium na meranie kvality rozdelenia bol použitý Gini index (*criterion = gini*)

- Počet funkcií, ktoré treba zvážiť pri hľadaní najlepšieho rozdelenia – (*max_features* = 2)

3.3 Analýza a zhodnotenie predikčných modelov

Po implementácii a trénoch týchto modelov, boli výsledky validované niekoľkými rôznymi spôsobmi. Tak sme zistili, ako dokáže nami navrhnutý model predpovedať výsledky.

Na porovnanie výkonnosti týchto klasifikačných modelov boli zvolené tieto vyhodnocovacie techniky, ktoré boli importované z knižnice *sklearn*:

1. Matica zámen (*Confusion Matrix*)
2. ROC krivka

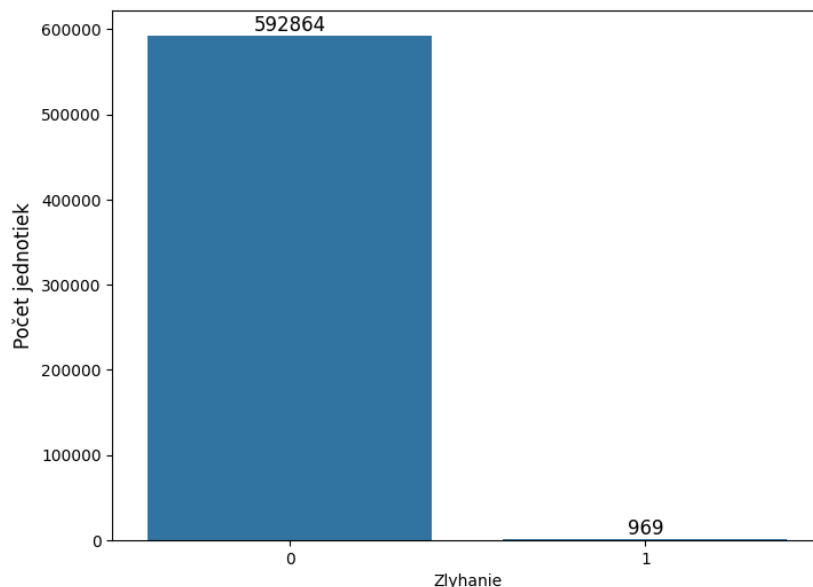
4 Výsledky a diskusia

V tejto kapitole popíšeme výsledky, ktoré sme získali analýzou poskytnutého datasetu. Ako už bolo zmienené, na analýzu a zostrojenie modelov nám bol poskytnutý dataset finančných výkazov firiem, ktorý obsahoval 593 833 záznamov z 231 268 podnikateľských subjektov z rokov 2015 – 2017. Na získanie a zistenie odpovedí sme šli postupne. Najprv začiatok sme si vytvorili charakteristiku zlyhaných a nezlyhaných spoločností, ktorú sme spracovali aj vizuálne. Z poskytnutých výkazov boli vypočítané vybrané finančné pomerové ukazovatele, ktoré boli použité na zostrojenie modelu s využitím strojového učenia. Pri výbere týchto ukazovateľov sme sa rozhodli použiť Altmanove ukazovatele, u ktorých bolo zistené, že majú signifikantný význam pri predpovedaní zlyhania spoločností. Sem patria pomerové ukazovatele týkajúce sa likvidity, ziskovosti, finančnej páky a aktivity spoločností.³⁸ Z vybraných ukazovateľov boli zostrojené modely s využitím strojového učenia za použitia algoritmov: lineárna diskriminačná analýza, logistická regresia a náhodné lesy rozhodovacích stromov (*Decision Trees*), ktoré budú zhodnotené v tejto kapitole pomocou vybraných evaluačných metód.

4.1 Charakteristika zlyhaných a nezlyhaných spoločností

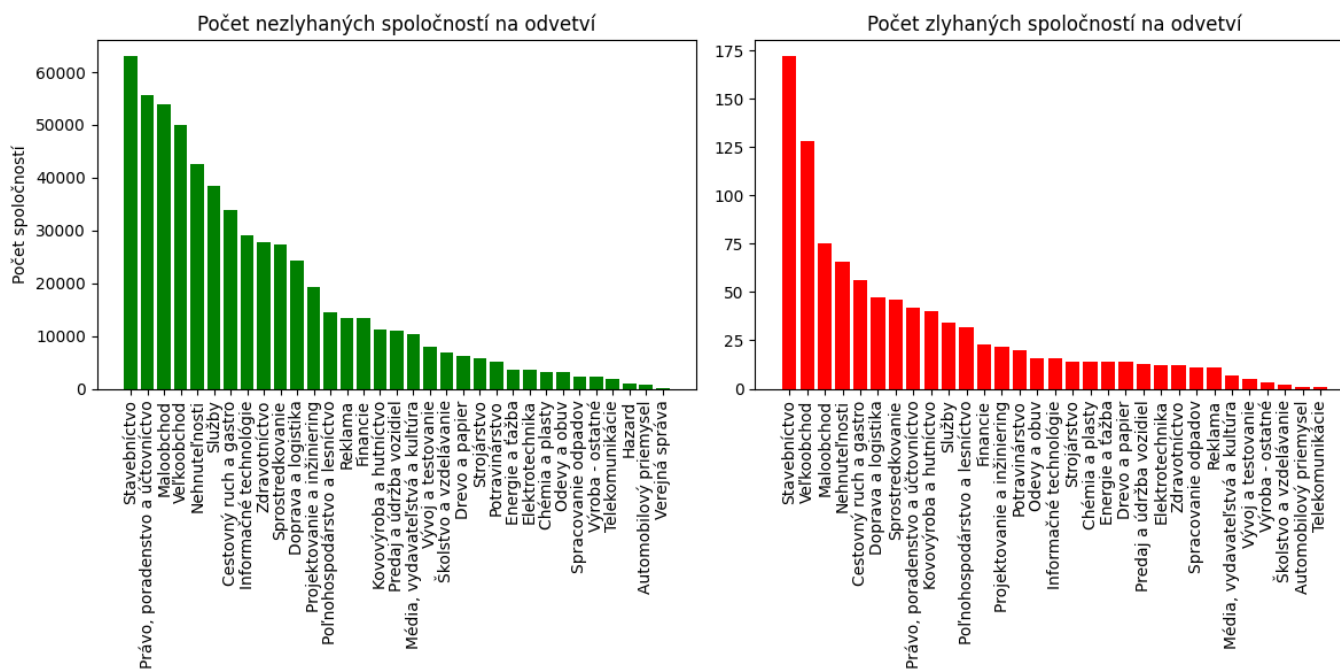
Na začiatok bol zostrojený graf závislosti všetkých údajov, pomocou ktorého bolo vizualizované celkové rozloženie zlyhaných a nezlyhaných jednotiek. To je znázornené na stĺpcovom Grafe 1. Z grafu je zrejmé, že počet zlyhaných údajov o jednotlivých spoločnostiach bol 592 864 a nezlyhaných bol 969. Z percentuálneho hľadiska to predstavuje 99,8368 % nezlyhaných a 0,1632 % zlyhaných spoločností. Na základe týchto použitých údajov vzniká problém nevyvážených dát medzi týmito dvomi kategóriami. To môže spôsobiť ďalšie problémy pri zostavení nášho modelu, ktoré však boli vyriešené oversampling metódou, popísanej v predošlej kapitole.

³⁸ ALTMAN, Edward I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Online. *The Journal of Finance*. 1968, roč. 23, č. 4. ISSN 00221082.



Graf 1 Pomer počtu zlyhaných a nezlyhaných jednotiek

Následne boli zostrojené grafy 2 a 3, ktoré vizuálne zobrazovali rozloženie závislosti zlyhaných a nezlyhaných spoločností na odvetví. Oba grafy sú vyobrazené na dole:



Graf 2 a 3 Počet zlyhaných a nezlyhaných spoločností v závislosti na odvetví

Na základe údajov vyobrazených v stĺpcových grafoch možno pozorovať, že najväčší počet zlyhaných spoločností z poskytnutých dát pochádza z odvetvia stavebníctvo, veľkoobchod, maloobchod a nehnuteľnosti. Možno to zdôvodniť aj tým, že odvetvia ako stavebníctvo a nehnuteľnosti sú silno ovplyvnené ekonomickými cyklami. V čase recesie môže dopyt po nových stavbách klesnúť, čo môže viesť k finančným ťažkostiam pre spoločnosti v týchto odvetviach. V odvetviach ako je veľkoobchod a maloobchod môže byť vysoká úroveň konkurencie, čo môže viesť k cenovému tlaku a tenkým maržám. To môže byť náročné pre spoločnosti, ktoré nemajú dostatočné finančné rezervy alebo efektívne operácie. Odvetvia ako cestovný ruch a gastro sú taktiež často ovplyvnené sezónnosťou a individuálnymi zmenami v spotrebiteľskom správaní.

Na druhej strane vysoký počet zlyhaných prípadov u niektorých z týchto spoločností môže byť spôsobený aj vysokým množstvom záznamov u odvetví ako napr. stavebníctvo, maloobchod a veľkoobchod. To nasvedčuje väčšiemu zastúpeniu týchto spoločností, čo následne zvyšuje pravdepodobnosť že nejaká spoločnosť zlyhá. Presné percentuálne zastúpenie všetkých spoločností z poskytnutého datasetu je uvedené v Tabuľke 5.

Tabuľka 5 Percentuálne zastúpenie spoločností

Odvetvie	%	Odvetvie	%
Stavebníctvo	10,64	Média, vydavateľstvá a kultúra	1,76
Právo, poradenstvo a účtovníctvo	9,37	Vývoj a testovanie	1,32
Maloobchod	9,08	Školstvo a vzdelávanie	1,17
Veľkoobchod	8,44	Drevo a papier	1,04
Nehnuteľnosti	7,18	Strojárstvo	0,98
Služby	6,49	Potravinárstvo	0,87
Cestovný ruch a gastro	5,71	Energie a ťažba	0,63
Informačné technológie	4,91	Elektrotechnika	0,60
Zdravotníctvo	4,68	Chémia a plasty	0,53
Sprostredkovanie	4,60	Odevy a obuv	0,52
Doprava a logistika	4,11	Spracovanie odpadov	0,39
Projektovanie a inžiniering	3,25	Výroba - ostatné	0,38
Poľnohospodárstvo a lesníctvo	2,45	Telekomunikácie	0,33
Reklama	2,27	Hazard	0,16
Financie	2,26	Automobilový priemysel	0,14
Kovovýroba a hutníctvo	1,91	Verejná správa	0,01
Predaj a údržba vozidiel	1,84		

Z hľadiska právnej formy je najväčšie zastúpenie spoločností typu s. r. o. a to konkrétne 95,75 %. Po nej nasleduje akciová spoločnosť, ktorá tvorí 2,75 % všetkých spoločností. Ešte menej bežné sú družstvá – 0,72 % , kom. spoločnosť – 0,46 %, prípadne ďalšie iné – 0,32 %, ktorých zastúpenie je minoritné.

4.2 Použité pomerové ukazovatele

Na zostrojenie predikčných modelov s využitím príslušných algoritmov strojového učenia boli použité Altmanove pomerové ukazovatele, ktoré boli vypočítané nasledovne:

X₁ – čistý prevádzkový kapitál/celkový kapitál

Pomer pracovného kapitálu k celkovým aktívam je dôležitým ukazovateľom likvidity a veľkosti podniku. Čistý prevádzkový kapitál definovaný ako rozdiel medzi obežnými aktívami a krátkodobými záväzkami je meraním schopnosti spoločnosti plniť svoje bežné záväzky. Väčšia hodnota tohto ukazovateľa je lepšia, pretože to znamená, že podnik má k dispozícii viac likvidných aktív na pokrytie svojich záväzkov. V prípade nižšej hodnoty tohto ukazovateľa, podnik zaznamenáva sústavné prevádzkové straty a bude mať menšiu tendenciu likvidných aktív. Vypočítali sme ho z našich ukazovateľov takto:

$$x_1 = \frac{\text{Obežný majetok (aktiva30)} - \text{krátkodobé záväzky (pasiva106)}}{\text{Spolu majetok (aktiva1)}}$$

X₂ – nerozdelený zisk/celkový kapitál

Tento pomer je dôležitým ukazovateľom finančnej stability a rentability spoločnosti. Obecne platí, že tento ukazovateľ popisuje schopnosť nadobúdať profit. Vyššia hodnota tohto pomeru naznačuje, že spoločnosť má viac zisku, ktorý môže použiť na svoj rast alebo na vyplatenie dividend svojim akcionárom. Ak má staršia firma nízky objem nerozdeleného zisku, môže to signalizovať jej ťažkosti. Mladé firmy majú nižšiu hodnotu tohto ukazovateľa, ktorý by mal s postupom času rásť. Obvykle sa v čitateli používa výsledok hospodárenia minulých rokov. My sme sa rozhodli v našom prípade rozhodli modifikovať premennú X₂ a použili sme výsledok hospodárenia za účtovné obdobie po zdanení. Rozhodli sme sa použiť tento ukazovateľ z dôvodu nedostupnosti potrebných údajov. Okrem toho tento ukazovateľ môže poskytovať aktuálnejší pohľad na finančnú situáciu firmy, čo môže byť v niektorých prípadoch dokonca výhodnejšie. Vypočítali sme ho z našich ukazovateľov takto:

$$x_2 = \frac{\text{Výsledok hospodárenia za účtovné obdobie po zdanení (vzas61)}}{\text{Spolu majetok (aktiva1)}}$$

X₃ – zisk pred zdanením + dane/celkový kapitál

Ukazovateľ zisku pred úrokmi a zdanením je meraním skutočnej produktivity firmy oslobodených od úrokov a daní. Narozdiel od zisku po zdanení sa tento ukazovateľ hodí viac, pretože sa berie do úvahy skutočnejšia výnosová sila jej aktív. Opäť platí, že čím je hodnota vyššia, tým má spoločnosť vyššiu výnosovú silu jej aktív a je schopnejšia generovať zisk. Čím nižšia je hodnota tohto ukazovateľa, tým vyššia je očakávaná náchylnosť k úpadku. Vypočítali sme ju z našich ukazovateľov takto:

$$x_3 = \frac{\text{Výsledok hospodárenia z hospodárskej činnosti (vzas26)}}{\text{Spolu majetok (aktiva1)}}$$

X₄ – trhovú hodnotu vlastného kapitálu /cudzí kapitál

Pomer trhovej hodnoty vlastného a cudzieho kapitálu vyjadruje mieru finančnej stability a rentability spoločnosti. Medzi ukazovateľ trhovej hodnoty akcií patrí hodnota všetkých akcií, prioritných aj kmeňových. Celkové dlhy zahŕňajú krátkodobé aj dlhodobé záväzky spoločnosti. Tento pomer vyjadruje, o koľko môže klesnúť hodnota aktív firmy predtým, než sa jej finančná stabilita a rentabilita stanú neudržateľnými. Tiež je dosť rozšírený podobný a bežne používaný pomer čistá hodnota/celkový dlh. V našom prípade sme tento ukazovateľ vypočítali takto:

$$x_4 = \frac{\text{Vlastné imanie (pasiva67)}}{\text{Dlhodobé záväzky + Krátkodobé záväzky (pasiva94 + pasiva106)}}$$

X₅ – tržby/celkový kapitál

Pomer tržieb s celkovým kapitálom vyjadruje schopnosť aktív generovať tržby. Tiež môže vyjadrovať schopnosť manažmentu vyrovnávať sa s podmienkami hospodárskej súťaže. Vypočítali sme ho takto:

$$x_5 = \frac{\text{Tržby z predaja tovaru + Tržby z predaja vl. výrobkov a služieb (vzas1 + vzas5)}}{\text{Spolu majetok (aktiva1)}}$$

4.2.1 Výpočet a deskriptívna štatistika vybraných pomerových ukazovateľov

V ďalšom kroku bola vygenerovaná deskriptívna štatistika z vybraných pomerových ukazovateľov pre zlyhané a nezlyhané podniky. Deskriptívna štatistika nám poskytla základné informácie o distribúcii hodnôt vypočítaných pomerových ukazovateľov v zlyhávajúcich a nezlyhávajúcich firmách. Jej výstup je znázornený v Tabuľkách 5 a 6:

Tabuľka 5 Deskriptívna štatistika pre zlyhané firmy jednotlivých pomerových ukazovateľov

	X1	X2	X3	X4	X5
Počet	889				
Priemer	-648,366	-17,188	-14,810	-0,891	15,385
SD	15592,572	394,869	397,874	49,784	274,586
Min	-462229,516	-10880,097	-10849,129	-972,187	-8,667
25%	-1,519	-0,607	-0,464	-0,816	0,004
50%	-0,197	-0,078	-0,039	-0,171	0,541
75%	0,141	0,002	0,013	0,173	2,247
Max	254,600	3432,865	3831,378	624,000	7781,364

Tabuľka 6 Deskriptívna štatistika pre nezlyhané firmy jednotlivých pomerových ukazovateľov

	X1	X2	X3	X4	X5
Počet	561532				
Priemer	-20,940	-3,261	-2,339	34,608	3,375
SD	3543,491	1018,570	914,747	9017,930	304,683
Min	-2143989,000	-630417,230	-632037,539	-432847,000	-7828,667
25%	-0,105	-0,066	-0,013	-0,049	0,227
50%	0,296	0,021	0,047	0,744	0,981
75%	0,740	0,133	0,186	4,214	2,067
Max	89541,000	101546,000	102026,000	6634174,000	132257,000

Na základe predbežného porovnania priemerných hodnôt vypočítaných ukazovateľov, možno pozorovať a tvrdiť, že u zlyhávajúcich firiem sa nachádza vyššia negatívna hodnota týchto ukazovateľov.

U pomerového ukazovateľa X₁ vykazujeme z priemernej hodnoty údajov vyššiu negatívnu hodnotu pre zlyhávajúce podniky. To naznačuje, že tieto podniky majú k dispozícii menej likvidných aktív na pokrytie svojich záväzkov, čiže tieto podniky zaznamenávali sústavné prevádzkové straty, čoho dôsledkom bol zánik.

U pomerových ukazovateľov X_2 , X_3 tiež možno pozorovať vyššie záporné hodnoty u zlyhaných podnikov. Nižšie hodnoty týchto ukazovateľov signalizujú, že firma utrpela straty, ktoré prevýšili jej celkový zisk. To naznačuje, že firma má problémy s generovaním zisku. Môže to byť dôsledok rôznych faktorov, ako napr. vysoké prevádzkové náklady, nízka tržná hodnota jej produktov alebo služieb, alebo neefektívne riadenie.

U pomerového ukazovateľa X_4 , sme zistili, že zlyhávajúce firmy mali tiež mať tendenciu zápornej hodnoty. Toto zistenie naznačuje, že tieto firmy mali relatívne vyššie dlhy (cudzí kapitál) v porovnaní s ich trhovou hodnotou vlastného kapitálu. Môže to byť dôsledok nadmerného zadlženia alebo poklesu trhovej hodnoty ich akcií.

U pomerového ukazovateľa X_5 pri porovnaní hodnôt vyšlo, že zlyhávajúce firmy mali väčšiu hodnotu tohto ukazovateľa. To naznačuje, že tieto firmy mali relatívne vyššie tržby v porovnaní s celkovým kapitálom, avšak mohli mať však vysoké prevádzkové náklady, nadmerný dlh alebo iné problémy, ktoré napomohli k zlyhaniu.

Z tabuliek možno tiež sledovať prítomnosť odľahlých hodnôt. Tieto hodnoty majú potenciál skresliť výsledky štatistických meraní, ako sú priemery a rozptyl, a môžu mať nepriaznivý vplyv na odhady našich modelov. Často sa tieto problémy riešia winsorizáciou dát pred ich analýzou. Avšak, v kontexte našej štúdie sme sa rozhodli nevykonať winsorizáciu. Naším cieľom bolo skúmať výkonnosť modelov na neupravených dátach, ktoré môžu obsahovať odľahlé hodnoty, aby sme získali realistický obraz o ich robustnosti. Je však dôležité zdôrazniť, že v praxi je potrebné riešiť prítomnosť odľahlých hodnôt.

4.3 Evaluácia zostavených predikčných modelov

Po natrénovaní, ktoré bolo popísané v predošlej kapitole, bola vykonaná evaluácia modelov strojového učenia. Konkrétne boli vytvorené model s využitím algoritmov lineárna diskriminačná analýza, logistická regresia a náhodný les rozhodovacích stromov. Na porovnanie výkonnosti modelov boli zvolené tieto vyhodnocovacie techniky:

1. Matica zámen (*confusion matrix*)
2. ROC krivka

4.3.1 Matica zámen

Zhodnotenie výkonnosti modelu s algoritmom logistickej regresie s využitím strojového učenia pomocou matice zámen je uvedené v Tabuľkách 7, 8, 9:

Tabuľka 7 Matica zámen modelu s využitím lineárnej diskriminačnej analýzy

Trieda	Negatívne (Predpovedané)	Pozitívne (Predpovedané)
Negatívne (Skutočné)	168460	9
Pozitívne (Skutočné)	258	0

Z matice zámen možno diagonálne vyčítať, že predikčný model s využitím lineárnej diskriminačnej analýzy v testovacom súbore predpovedal správne 168460 negatívnych prípadov, keď spoločnosť nezlyhala a model ju zaradil správne. Model nevedel zaradiť žiadny pozitívny prípad správne, keď spoločnosť zlyhala a klasifikoval ju ako zlyhanú. Na druhú stranu model nesprávne predpovedal 9 prípadov ako zlyhaných avšak patrili medzi nezlyhané. Taktiež model nevedel zaradiť ani 1 prípad zlyhaných spoločností pričom zaradil 258 spoločností ako nezlyhané keď naozaj zlyhali.

Tabuľka 8 Matica zámen modelu s využitím logistickej regresie

Trieda	Negatívne (Predpovedané)	Pozitívne (Predpovedané)
Negatívne (Skutočné)	147164	21305
Pozitívne (Skutočné)	147	111

V prípade zostrojenia predikčného modelu s algoritmom logistická regresia z matice zámen možno vyčítať, že model správne predpovedal 147164 negatívnych prípadov, keď spoločnosť nezlyhala a model ju zaradil správne. Taktiež model správne zaradil 111 prípadov, keď spoločnosť zlyhala a model ju zaradil ako zlyhanú. Na druhej strane model nesprávne predpovedal 21305 prípadov ako zlyhaných, no patrili medzi nezlyhané. Taktiež model nesprávne zaradil 147 prípadov, keď spoločnosti zlyhali a model ich zaradil ako nezlyhané.

Tabuľka 9 Matica zámen modelu s využitím algoritmu náhodného lesu rozhodovacích stromov

Trieda	Negatívne (Predpovedané)	Pozitívne (Predpovedané)
Negatívne (Skutočné)	137792	30677
Pozitívne (Skutočné)	141	117

V prípade zostrojenia predikčného modelu s využitím algoritmu náhodný les rozhodovacích stromov možno vyčítať, že model správne predpovedal 137792 negatívnych prípadov, keď spoločnosť nezlyhala a model ju zaradil správne. Model tiež zaradil správne 117 prípadov, keď spoločnosť zlyhala a model ju zaradil ako zlyhanú. Na druhej strane model nesprávne predpovedal 30677 prípadov ako zlyhaných, ktoré však patrili medzi nezlyhané. Model tiež nesprávne zaradil 141 prípadov, keď spoločnosť zlyhala a zaradil ich ako nezlyhané.

Z matice zámen boli tiež vypočítané nasledovné ukazovatele, ktoré sú uvedené v nasledujúcej Tabuľke 10:

$$\text{Presnosť} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Špecifita} = \frac{TN}{TN + FP}$$

$$\text{Senzitivita} = \frac{TP}{TP + FN}$$

$$\text{Precíznosť} = \frac{TP}{TP + TN}$$

Tabuľka 10 Vypočítané príslušné ukazovatele vybraných modelov z matice zámen

Model	Presnosť (%)	Špecifita (%)	Senzitivita (%)	Precíznosť (%)
Lineárna diskriminačná analýza	99,84	99,99	–	–
Logistická regresia	87,29	87,35	43,02	0,08
Náhodný les rozhodovacích stromov	81,73	81,79	45,35	0,08

Na základe porovnania týchto ukazovateľov možno uzavrieť, že najvyššiu presnosť mal model s lineárnou diskriminačnou analýzou. Presnosť modelu činila 99,84 %. No aj napriek vysokej hodnote presnosti možno tvrdiť, že tento model bol z celkovej evaluácie slabý, pretože nebol schopný nájsť a správne rozlíšiť zlyhané a nezlyhané triedy z nášho datasetu.

V prípade predikčného modelu s využitím logistickej regresie s možno pozorovať o niečo nižšiu celkovú presnosť 87,29 %. Model bol však schopný do istej miery klasifikovať a zaradiť správne 147164 nezlyhaných a 111 zlyhaných prípadov.

Presnosť predikčného modelu s využitím algoritmu náhodný les rozhodovacích stromov bola 81,73 %. Model bol schopný zaradiť správne 137792 nezlyhaných a 117 zlyhaných prípadov. Pri porovnaní tohto modelu s logistickou regresiou si tento model viedol o niečo horšie. Je však možné, že ladením a optimalizáciou týchto hyperparametrov by výsledky mohli prekonať aj logistickú regresiu.

Na základe pozorovania parametrov možno ďalej vidieť, že predikčné modely majú vysoké hodnoty špecifity a nízke hodnoty senzitivity a presnosti. Na základe tohto môžeme tvrdiť, že nami zostavené predikčné modely vedia dobre predpovedať negatívne príklady (nezlyhané prípady), avšak majú problém s predikciou pozitívnych (zlyhané prípady). Nasvedčuje tomu aj fakt, že pomer zlyhaných a nezlyhaných spoločností je v nerovnováhe. Z percentuálneho hľadiska to predstavuje 99,8368 % nezlyhaných a 0,1632 % zlyhaných spoločností.

Pri zostavení predičných modelov bola na tréningový set aplikovaná technika *SMOTE* oversampling nastavená na 90%, kedy synteticky generovanie nových príkladov v menšej triede, dopomohlo k zlepšeniu schopnosti predikčných modelov s algoritmi logistická regresia a náhodný les rozhodovacích stromov, kedy boli tieto modely schopné lepšie detekovať zlyhané firmy. Avšak ďalším testovaním bolo zistené že nadmerný oversampling síce zvýši predikciu zlyhaných firiem, avšak sa rapídne zvýšia hodnoty falošne negatívnych a falošne pozitívnych tried.

4.3.2 ROC krivka

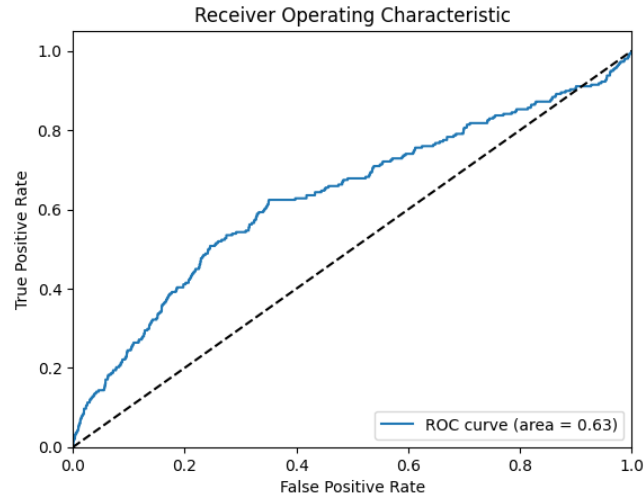
Na základe získanej hodnoty AUC v ROC krivke bola tiež zistená výkonnosť nášho klasifikačného modelu. AUC skóre je užitočným ukazovateľom, ktorý nám pomáha posúdiť, ako dobre model rozlišuje medzi pozitívnymi a negatívnymi príkladmi. Výkonnosť našich predikčných modelov boli ohodnotené pomocou hodnôt AUC v Tabuľke 11:

Tabuľka 11 Hodnoty AUC popisujúce výkonnosť modelu ³⁹

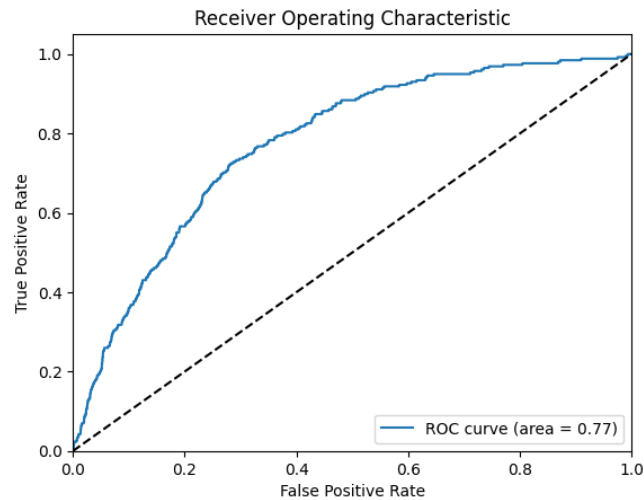
Hodnota AUC	Popis
$0,9 \leq \text{AUC}$	Vynikajúci
$0,8 \leq \text{AUC} < 0,9$	Dobrý
$0,7 \leq \text{AUC} < 0,8$	Priemerný
$0,6 \leq \text{AUC} < 0,7$	Slabý

U modelu s využitím lineárnej diskriminačnej analýzy bola hodnota AUC stanovená na 63,39 %. Tá podľa Tabuľky 11, spadá do oblasti slabej schopnosti klasifikovať správne pozitívne a negatívne prípady. Táto hodnota sa nachádza nad priemernou hodnotou 0,5, vďaka čomu môžeme tvrdiť, že náš zostavený model je lepší ako náhodné zaradovanie, avšak nie o veľa.

³⁹ NARKHEDE, Sarang. Understanding AUC - ROC Curve. *Medium* [online]. Dostupné na: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

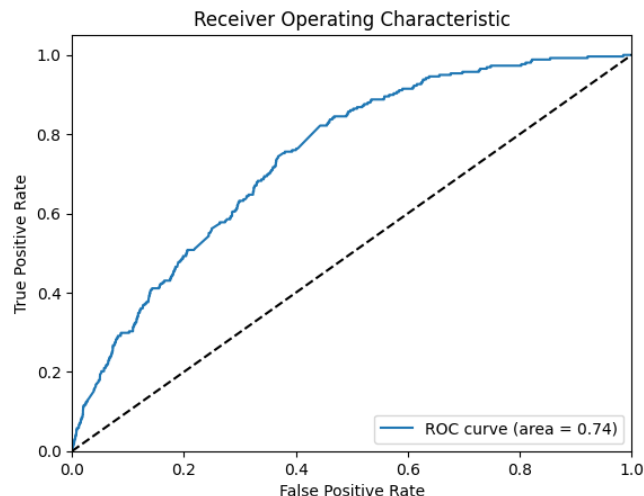


Graf 4 ROC krivka a AUC skóre pre model lin. diskriminačnej analýzy



Graf 5 ROC krivka a AUC skóre pre model s využitím algoritmu logistická regresia

Pri zostavenom predikčnom modeli s využitím logistickej regresie bola hodnota AUC stanovená na 77,46 %. Z hľadiska výkonnosti spadá náš zostavený model s logistickou regresiou do oblasti priemernej schopnosti klasifikovať správne pozitívne a negatívne prípady. Táto hodnota je vyššia ako u predošlého modelu, avšak, stále je tu priestor na vylepšenie nášho modelu za účelom dosiahnutia lepšej hodnoty.



Graf 6 ROC krivka a AUC skóre pre model s využitím algoritmu náhodný les rozhodovacích stromov

Pri zostavenom predikčnom modeli s využitím algoritmu náhodného lesu rozhodovacích stromov bola hodnota AUC stanovená na 74,39 %. Z hľadiska výkonnosti spadá náš zostavený model s logistickou regresiou do oblasti priemernej schopnosti klasifikovať správne pozitívne a negatívne prípady. Hodnota je väčšia ako 0,5, vďaka čomu môžeme tvrdiť, že náš zostavený model je lepší ako náhodné zaradovanie. Taktiež je možné tvrdiť, že ďalším ladením a optimalizáciou hyperparametrov by bolo možné dosiahnuť ešte lepšie výsledky a zlepšiť aj túto hodnotu.

Záver

V tejto bakalárskej práci sme sa zaoberali modelovaním zlyhania firiem. Keďže zlyhanie je problém, ktorý je kritický pre finančnú stabilitu a udržateľnosť spoločností, našim cieľom bolo vytvoriť predikčné modely, ktoré by pomohli identifikovať zaradiť zlyhanie spoločností z dostupných finančných výkazov SR spoločností.

V tejto práci boli zostrojené predikčné modely s použitím Altmanových pomerových ukazovateľov a s využitím algoritmov: lineárna diskriminačná analýza, logistická regresia a náhodný les rozhodovacích stromov. Tieto modely boli spočiatku trénované a následne testované na datasete slovenských firiem v pomere 70/30. Na záver boli vyhodnotené zostavené predikčné modely pomocou príslušných evaluačných metód.

Pri porovnaní týchto zostrojených predikčných modelov sme zaznamenali najlepšie výsledky u modelu s využitím algoritmu logistická regresia. Za ním sa umiestnil model s náhodným lesom rozhodovacích stromov. Tento model síce nedosiahol najlepšie výsledky, avšak je dosť pravdepodobné, že ďalším testovaním a ladením hyperparametrov by bolo možné ešte dosiahnuť optimálnejšie výsledky. Aj tu prichádza do popredia využitie strojového učenia na zostrojenie predikčných modelov, ktoré dokážu automaticky identifikovať a využiť komplexné vzťahy v dátach, ktoré nie sú hneď inými metódami identifikovateľné. To umožňuje vytvárať predikčné modely s vyššou presnosťou a robustnosťou. Navyše neustále učenie sa z nových dát zabezpečuje, že modely zostávajú relevantné a presné aj v dynamicky sa meniacom podnikateľskom prostredí.

Zoznam použitej literatúry

1. ALTMAN, Edward I.; ZAINUDDIN, Zaemah a NORDIN, Sabariah. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy: Logistic Regression and Multivariate Discriminant Analysis. Online. *The Journal of Finance*. 1968, roč. 23, č. 4, s. 13-23. ISSN 00221082. Dostupné na: <https://doi.org/10.2307/2978933>.
2. BEHRENS, T.; SCHMIDT, K.; VISCARRA ROSSEL, R. A.; GRIES, P.; SCHOLTEN, T. et al. Spatial modelling with Euclidean distance fields and machine learning. Online. *European Journal of Soil Science*. 2018, roč. 69, č. 5, s. 757-770. ISSN 1351-0754
3. Di Bella, L., Katsinis, A., Lagüera-González, J., Odenthal, L., Hell, M., Lozar, B. *Annual Report on European SMEs 2022/2023*, Publications Office of the European Union, Luxemburg, 2023, doi:10.2760/028705, JRC134336.
4. FAWCETT, Tom. An introduction to ROC analysis. Online. *Pattern Recognition Letters*. 2006, roč. 27, č. 8, s. 861-874. ISSN 01678655. Dostupné na: <https://doi.org/10.1016/j.patrec.2005.10.010>.
5. HASSAN, Ehsan ul; ZAINUDDIN, Zaemah a NORDIN, Sabariah. A Review of Financial Distress Prediction Models: Logistic Regression and Multivariate Discriminant Analysis. Online. *Indian-Pacific Journal of Accounting and Finance*. 2017, roč. 1, č. 3, s. 13-23. ISSN 2590-406X. Dostupné na: <https://doi.org/10.52962/ipjaf.2017.1.3.15>.
6. ING. MICHAL CHMULÍK, PhD. Náhľad do algoritmu rozhodovací strom. *Umelá Inteligencia.sk* [online]. Available at: <https://umelainteligencia.sk/nahlad-do-algoritmu-rozhodovaci-strom/>
7. Kücher, A., Mayr, S., Mitter, C. et al. Firm age dynamics and causes of corporate bankruptcy: age dependent explanations for business failure. *Rev Manag Sci* **14**, 633–661 (2020). <https://doi.org/10.1007/s11846-018-0303-2>
8. MAJDÚCHOVÁ, H. a kol. *Podnikové hospodárstvo*. 2. Wolters Kluwer, 2020. 424 s. ISBN 978-80-571-0271-7.
9. Mahesh, Batta. (2019). Machine Learning Algorithms -A Review. *International Journal of Science and Research* <https://doi.org/10.21275/ART20203995>.
10. MUSA, Abdallah Bashir. Comparative study on classification performance between support vector machine and logistic regression. Online. *International Journal of Machine Learning and Cybernetics*. 2013, roč. 4, č. 1, s. 13-24. ISSN 1868-8071. Dostupné na: <https://doi.org/10.1007/s13042-012-0068-x>.

11. NARKHEDE, Sarang. Understanding AUC - ROC Curve. *Medium* [online]. Dostupné na: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
12. PANESAR, Arjun. Machine Learning and AI for Healthcare. Online. Dostupné na: <https://doi.org/10.1007/978-1-4842-3799-1>.
13. POPOOLA, Oluwatoyin Muse Johnson. Preface to the Third Issue of Indian-Pacific Journal of Accounting and Finance. Online. *Indian-Pacific Journal of Accounting and Finance*. 2017, roč. 1, č. 3, s. 1-3. ISSN 2590-406X. Dostupné na: <https://doi.org/10.52962/ipjaf.2017.1.3.20>.
14. *Scikit-learn Machine Learning in Python*. Online. Dostupné na: https://scikit-learn.org/stable/modules/cross_validation.html.
15. *Scikit-learn Machine Learning in Python*. Online. Dostupné na: https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html.
16. SMEJKAL, Vladimír a RAIS, Karel. *Řízení rizik ve firmách a jiných organizacích*. 4., aktualiz. a rozš. vyd. Expert (Grada). Praha: Grada, 2013. ISBN 978-80-247-4644-9.
17. TING, Kai Ming. Confusion Matrix. Online. *Encyclopedia of Machine Learning*. 2010, s. 209-209. ISBN 978-0-387-30768-8
18. VEGANZONES, David a SEVERIN, Eric. Corporate failure prediction models in the twenty-first century: a review. Online. *European Business Review*. 2020, roč. 33, č. 2, s. 204-226. ISSN 0955-534X. Dostupné na: <https://doi.org/10.1108/EBR-12-2018-0209>.
19. Vuković, Bojana & Milutinović, Sunčica & Milicevic, Nikola & Jaksic, Dejan. (2020). Corporate Bankruptcy Prediction: Evidence from Wholesale Companies in the Western European Countries. *Ekonomický časopis*. 68. 477-498.
20. ZALAI, K. a kol. *Finančno-ekonomická analýza podniku*. 7. preprac. a rozšir. vyd. Bratislava: Sprint 2, 2010. 487s. ISBN 978-80-89393-15-2.
21. *Zákon č. 513/1991 Zb. Obchodný zákonník, § 5*
22. *Zákon č. 7/2005 Z. z. o konkurze a reštrukturalizácii a o zmene a doplnení niektorých zákonov v znení neskorších predpisov*
23. *Zákone o účtovníctve č.431/2002 Z.Z.*