



UNIVERZITA KOMENSKÉHO V BRATISLAVE

Fakulta matematiky, fyziky a informatiky

Mlynská dolina F1, 842 48 Bratislava



**Témy dizertačných prác vypísané na prijímacie konanie  
so začiatkom štúdia v akad. roku 2023/2024**

Pracovisko FMFI:

Pracovisko SAV<sup>1</sup>: Ekonomický ústav SAV, v.v.i.

Téma <sup>2</sup>	Detection of tax manipulations using machine learning and artificial intelligence methods Odhaľovanie daňových manipulácií pomocou metód strojového učenia a umelej inteligencie
Školiteľ/ka	doc. Ing. Eduard Baumöhl, PhD.
Konzultant/ka <sup>3</sup>	

<sup>1</sup> Ekonomický ústav SAV

<sup>2</sup> V jazyku/jazykoch práce

<sup>3</sup> Vyplniť ak vhodné



Anotácia <sup>4</sup>	<p>ENG:</p> <p>Juggling with the corporate accounting and financial statements is an integral part of corporate finance, particularly in Slovakia. The main goal of the thesis is to develop a new model based on the recent advances in neural networks and machine learning, suited for the conditions of Slovak business environment. The thesis has a strong applied research focus, and as such, a close cooperation with the official authority in Slovakia – Financial Directorate of SR (FD SR) – will be maintained. There is a submitted APVV project on this topic, in which the student will be involved and from the FD SR part, a short-term contract might be provided.</p> <p>The most commonly used data mining techniques at the beginning of 21st century to detect financial manipulations includes neural networks, Bayesian analysis, and decision trees (Ngai et al., 2011; Ravisankar et al., 2011; Feroz et al., 2000; Lin et al., 2003). Later on, in addition to these methods, the Support Vector Machine (SVM) was also used (Perols, 2011; Albashrawi, 2016). All these techniques, led by neural networks, are also a standard part of current research, in which several techniques are used to compare their performance simultaneously (Lin et al., 2015). Some works extended the techniques to KNN (K-Nearest Neighbor), but mainly to hybrid systems used to identify factors predicting manipulative behaviour (Kirkos et al., 2007). A less frequently used method in the analysis of financial manipulations is the Random Forest machine learning technique, which, however, appears to have significantly better results compared to other methods (Whiting et al., 2012; Patel et al., 2019; An and Suh, 2020; Wyrobek, 2020). Due to its accuracy, this technique is thus coming to the forefront in analysing manipulative behaviours in current research.</p> <p>A typical problem of this type of analysis is the class imbalance in the disproportion of subjects identified as manipulators. Many methodological approaches have been developed in recent years to cope with the issue of imbalanced learning. In general, these can be divided into several categories: sampling methods, cost-sensitive methods for imbalanced learning, ensemble methods and various hybrid methods (He and Garcia, 2009). From the methodological point of view, tackling the problem of high data imbalance will be one of the major concerns in the thesis.</p>
-----------------------	--

<sup>4</sup> Stručný popis práce v jazyku/jazykoch práce



SK:

Žonglovanie s podnikovým účtovníctvom a účtovnou závierkou je neoddeliteľnou súčasťou podnikových financií, obzvlášť na Slovensku. Hlavným cieľom tejto práce je vyvinúť nový model založený na najnovších pokrokoch v oblasti neurónových sietí a strojového učenia, ktorý by bol vhodný pre podmienky slovenského podnikateľského prostredia. Práca má silné zameranie na aplikovaný výskum, a preto sa bude udržiavať úzka spolupráca s oficiálnym orgánom na Slovensku – Finančným riaditeľstvom SR (FR SR). Na túto tému je podaný projekt APVV, do ktorého bude študent zapojený a zo strany FR SR môže byť poskytnutá krátkodobá zmluva.

Medzi najčastejšie používané techniky dolovania údajov na začiatku 21. storočia na odhaľovanie finančných manipulácií patria neurónové siete, Bayesovská analýza a rozhodovacie stromy (Ngai et al., 2011; Ravisankar et al., 2011; Feroz et al., 2000; Lin et al., 2003). Neskôr sa okrem týchto metód začala používať aj metóda SVM (Support Vector Machine; Perols, 2011; Albashrawi, 2016). Všetky tieto techniky na čele s neurónovými sieťami sú štandardnou súčasťou aj existujúceho výskumu v tejto oblasti, v ktorých sa používa viacero techník na porovnanie ich výkonnosti (Lin et al., 2015). Niektoré práce rozšírili techniky na KNN (K-Nearest Neighbor), ale hlavne na hybridné systémy používané na identifikáciu faktorov predikujúcich manipulatívne správanie (Kirkos et al., 2007). Menej často používanou metódou pri analýze finančných manipulácií je technika strojového učenia Random Forest, ktorá však môže poskytovať výrazne lepšie výsledky v porovnaní s inými metódami (Whiting et al., 2012; Patel et al., 2019; An a Suh, 2020; Wyrobek, 2020). Táto technika sa tak vďaka svojej presnosti dostáva do popredia pri analýze manipulatívneho správania v súčasnom výskume.

Typickým problémom tohto typu analýzy je nevyváženosť v dátach, t.j. disproporcía subjektov identifikovaných ako manipulátori. V posledných rokoch sa vyvinulo mnoho metodických prístupov na zvládnutie problému nevyváženého učenia. Vo všeobecnosti ich možno rozdeliť do niekoľkých kategórií: metódy vzorkovania, metódy citlivé na náklady pre nevyvážené učenie, súhrnné metódy a rôzne hybridné metódy (He a Garcia, 2009). Z metodologického hľadiska bude jednou z hlavných tém práce riešenie problému vysokej nevyváženosti údajov.



Forma štúdia	Denná/Externá
Jazyk práce	AJ/SJ

### References:

- Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, 14(3), 553-569.
- An, B., Suh, Y. (2020). Identifying financial statement fraud with decision rules obtained from Modified Random Forest. *Data Technologies and Applications*, 54(2), 235-255.
- Feroz, E. H., Kwon, T. M., Pastena, V. S., Park, K. (2000). The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach. *Intelligent Systems in Accounting, Finance & Management*, 9(3), 145-157.
- He, H., Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9), 1263-1284.
- Kirkos, E., Spathis, C., Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995-1003.
- Lin, C. C., Chiu, A. A., Huang, S. Y., Yen, D. C. (2015). Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, 89, 459-470.
- Lin, J. W., Hwang, M. I., Becker, J. D. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*, 18(8), 657-665.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3), 559-569.
- Patel, H., Parikh, S., Patel, A., Parikh, A. (2019). An application of ensemble random forest classifier for detecting financial statement manipulation of Indian listed companies. In: *Recent developments in machine learning and data analytics* (pp. 349-360). Springer, Singapore.
- Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2), 19-50.
- Ravisankar, P., Ravi, V., Rao, G. R., Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision support systems*, 50(2), 491-500.
- Whiting, D. G., Hansen, J. V., McDonald, J. B., Albrecht, C., Albrecht, W. S. (2012). Machine learning methods for detecting patterns of management fraud. *Computational Intelligence*, 28(4), 505-527.
- Wyrobek, J. (2020). Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture. *Procedia Computer Science*, 176, 3037-3046.