**Ime i prezime: Sarah Djouab**

Adresa: Ulica Emanuela Vidovića 3f

Telefon: 0912588882

Adresa elektroničke pošte: sdjouab@efzg.hr

JMBAG: 0067506763

**Naziv studija: Elektroničko poslovanje u privatnom i javnom sektoru**

Godina upisa: 2020

**VIJEĆU SMJERA EKONOMSKOG FAKULTETA SVEUČILIŠTA U ZAGREBU**

**PREDMET: PRIJAVA TEME DIPLOMSKOG RADA**

**Tema/naslov diplomskog rada**: **Analiza odlazaka klijenata banke konkurentima pomoću stabla odlučivanja**

**Mentor**: prof.dr.sc. Mirjana Pejić Bach

**Prijedlog povjerenstva** (ispunjava potencijalni mentor):

1. Prof.dr.sc. Mirjana Pejić Bach
2. Prof.dr.sc. Katarina Ćurko
3. Izv.prof.dr.sc. Ljubica Milanović Glavan

**Potpis studenta:**

**Potpis mentora:**

U Zagrebu, 09. Rujan 2022

**Suglasnost Vijeća smjera:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Prilozi**:

* Sadržaj rada
* Obrazloženje teme
* Popis literature
* Potpisana Izjava o akademskoj čestitosti

**Sadržaj rada**

**Tema/naslov diplomskog rada**: **Analiza odlazaka klijenata banke konkurentima pomoću stabla odlučivanja**

**1. Uvod**

1.1. Predmet i cilj rada

1.2. Izvori podataka i metode prikupljanja

1.3. Struktura rada

**2. Podatkovna znanost**

2.1. Pojam podatkovne znanosti

2.2. Razvoj podatkovne znanosti

2.3. Područja primjene podatkovne znanosti u poslovanju

2.4. Specifičnosti primjene podatkovne znanosti u bankarstvu

**3. Metode modeliranja u podatkovnoj znanosti**

3.1. Metode klasifikacije

3.2. Metode predviđanja

3.3. Metode grupiranja i povezivanja

3.4. Stablo odlučivanja kao metoda podatkovne znanosti

**4. Istraživanje odlazaka klijenata banke pomoću stabla odlučivanja**

4.1. Metodologija istrživanja

4.2. Rezultati istraživanja

4.2.1. Deskriptivna analiza varijabli

4.2.2. Izrada modela predviđanja odlazaka klijenata banke pomoću stabla odlučivanja

4.3. Predviđanja budućih trendova primjene podatkovne znanosti u bankarstvu

**5. Zaključak**

**Literatura**

**Popis slika**

**Popis tablica**

**Sažetak i ključne riječi**

**Životopis kandidata**

**Obrazloženje teme:**

**Tema/naslov diplomskog rada**: **Analiza odlazaka klijenata banke konkurentima pomoću stabla odlučivanja**

Znanost o podacima je interdisciplinarno polje koje koristi znanstvene metode, procese, algoritme i sustave za ekstrakciju znanja i uvida iz strukturiranih i nestrukturiranih podataka. Znanost o podacima koristi tehnike i teorije iz različitih disciplina, uključujući matematiku, statistiku, informatiku, i znanje o specifičnom području kao što je bankarstvo.

Svaki radno sposobni čovjek, koji ima aktivan radni status koristi banku kako bi barem primao plaću putem tekućeg računa, ako već nije korisnik kredita, kreditne kartice ili štednje. U Hrvatskoj posluje tridesetak banaka, klijentima je omogućen veliki izbor kako bi mogao odlučiti za svoj najisplativiji izbor među Bankama. Banke nude lepezu proizvoda, i klijentima, s obzirom na potrebe i interese, nude različite usluge. Cilj istraživanja je identificirati karakteristike podataka klijenata koji mijenjaju banku, tj. odlaze od svoje matične banke kod konkurenata.

Za ovaj rad koristit će se izvor podataka koji sadrži informacije o klijentima banaka u Njemačkoj, Francuskoj i Španjolskoj, među ostalim i informacije o odlascima ovisno o njihovim gubitcima, kreditnoj sposobnosti i ostalim atributima. Koristi se jedna od najčešće korištenih metoda podatkovne znanosti, a to je stablo odlučivanja, specifično algoritam C.45 (J.48 unutar Weka platforme za rudarenje podataka). Stablo odlučivanja koristi se za klasifikaciju, predikciju, grupiranje, procjenu vrijednosti, vizualizaciju pomoću histograma i stabala i opisivanje podataka. Metoda je izabrana zbog jednostavnosti korištenja i razumljivosti modela.

C4.5 je algoritam za generiranje stabla odlučivanja koje se može koristiti za klasifikaciju, zbog čega se ponekad naziva i statistički klasifikator. Ovaj algoritam je jedan od najčešće korištenih algoritama za klasifikaciju podataka, zbog mogućnosti korištenja nominalnih varijabli kao ciljnih varijabla sa dvije i više kategorije, te zbog mogućnosti korištenja numeričkih i nominalnih varijabli kao nezavisnih varijabli.

Cilj ovoga rada je koristeći bazu podataka o klijentima banaka u raznim zemaljam i alate za podatkovnu analizu istražiti dostupne informacije i doći do vrijednih zaključaka vezano uz odlazak klijenata iz banke istovremeno zadržavši visoku stopu preciznosti korištenog modela.

**Popis literature**

1. Ali, Ö. G., & Arıtürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. Expert Systems with Applications, 41(17), 7889-7903.
2. Arasu, B. S., Seelan, B. J. B., & Thamaraiselvan, N. (2020). A machine learning-based approach to enhancing social media marketing. Computers & Electrical Engineering, 86, 106723.
3. Bilal Zorić, A. (2016). Predicting customer churn in banking industry using neural networks. Interdisciplinary Description of Complex Systems: INDECS, 14(2), 116-124.
4. COŞER, A., Aldea, A., Maer-Matei, M. M., & BEŞIR, L. (2020). Propensity to churn in banking: what makes customers close the relationship with a bank?. Economic Computation & Economic Cybernetics Studies & Research, 54(2).
5. Dias, J., Godinho, P., & Torres, P. (2020, July). Machine learning for customer churn prediction in retail banking. In International Conference on Computational Science and Its Applications (pp. 576-589). Springer, Cham.
6. Ivanišević, F. (2016). KOMPARATIVNA ANALIZA PREDIKTIVNIH TEHNIKA RUDARENJA PODATAKA (Doctoral dissertation, University of Split. Faculty of economics Split).
7. Jain, H., Yadav, G., & Manoov, R. (2021). Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In Advances in Machine Learning and Computational Intelligence (pp. 137-156). Springer, Singapore.
8. Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation, 2(1), 1-13.
9. Kumar, D. A., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. Int. J. Data Anal. Tech. Strateg., 1(1), 4-28.
10. Macanga, A. (2017). Rudarenje podataka kao alat za upravljanje odnosima s klijentima (Doctoral dissertation, University of Zagreb. Faculty of Organization and Informatics. Department of Information Systems Development).
11. Mukhopadhyay, D., Malusare, A., Nandanwar, A., & Sakshi, S. (2021). An approach to mitigate the risk of customer churn using machine learning algorithms. In Machine Learning for Predictive Analysis (pp. 133-142). Springer, Singapore.
12. Mutanen, T., Nousiainen, S., & Ahola, J. (2010). Customer churn prediction–a case study in retail banking. In Data Mining for Business Applications (pp. 77-83). IOS Press.
13. Pakšec, M. (2019). Tehnike rudarenja podataka za procjenu kreditnog rizika (Doctoral dissertation, University of Zagreb. Faculty of Organization and Informatics. Department of Information Systems Development).
14. Pejić Bach, M. (2005). Rudarenje podataka u bankarstvu. Zbornik ekonomskog fakulteta u Zagrebu, 3(1), 181-193.
15. Pejić Bach, M., Zoroja, J., Jaković, B., & Šarlija, N. (2017, May). Selection of variables for credit risk data mining models: preliminary research. In 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 1367-1372). IEEE.
16. Popović, D., & Bašić, B. D. (2009). Churn prediction model in retail banking using fuzzy C-means algorithm. Informatica, 33(2).
17. Rahman, M., & Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1196-1201). IEEE.
18. Ullah, I., Hussain, H., Ali, I., & Liaquat, A. (2019, July). Churn prediction in banking system using K-means, LOF, and CBLOF. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-6). IEEE.
19. Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. Expert Systems with Applications, 36(3), 5445-5449.
20. Živolić, S. (2021). RUDARENJE PODATAKA I NAPREDNE ANALITIČKE TEORIJE I METODE. Zbornik radova Fakulteta tehničkih nauka u Novom Sadu, 36(03), 448-451.

Potpisana izjava o akademskoj čestitosti

IZJAVA O AKADEMSKOJ ČESTITOSTI

*Izjavljujem i svojim potpisom potvrđujem da je diplomski rad / prijava teme diplomskog rada isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija.*

*Izjavljujem da nijedan dio rada / prijave teme nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog izvora te da nijedan dio rada / prijave teme ne krši bilo čija autorska prava.*

*Izjavljujem, također, da nijedan dio rada / prijave teme nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.*

 

 (vlastoručni potpis studenta)

Zagreb, 09.09.2022.
 (mjesto i datum)

STATEMENT ON THE ACADEMIC INTEGRITY

*I hereby declare and confirm by my signature that the final thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography.*

*I declare that no part of the thesis has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the thesis infringes any of the copyrights.*

*I also declare that no part of the thesis has been used for any other work in any other higher education, scientific or educational institution.*

**

 (personal signature of the student)

Zagreb, 09.09.2022.
 (place & date)