Application of supervised machine learning models for card payment workflow optimization

Oskars Rodiņš[1]

[1] Riga Technical University, Ķīpsalas iela 6a, Riga, LV-1048, Latvia

Abstract

This paper explores the topic of the application of machine learning models in the context of workflow optimization in the financial industry. Within this study a review of existing literature is conducted to determine possible applications of machine learning in the financial sector and a practical experiment is performed: a machine learning classification model is developed to predict the result of a card deposit workflow based on historical transaction data, gathered in a financial institution during April 2024. The experiment evaluates machine learning as one of the available tools for workflow optimization. This paper briefly describes the process of building a classification model: including data selection, and the choice of the classification model and elaborates on the results of the developed model's performance.

Keywords

workflow, optimization, machine learning, classification model, finance, card payments

1. Introduction

Financial organizations leverage the use of workflows to ensure a controlled and effective execution of various business processes, such as card payments, wire transfers and account-related activities. As observed in the author's previous work [1], various optimization approaches and technologies can be applied to make the execution of workflows more efficient and bring meaningful improvements to an organization's business process performance. A survey conducted within the 2023 study [1], targeting 60 employees from 2 financial companies, has shown that the majority (83%) of the respondents find workflow optimization topic important or highly important. Workflow optimization techniques can be classified into the following categories:

- Input data optimization,
- Workflow execution step sequence changes,
- Manual processing reduction.

Machine learning (ML) is one of the promising technologies to address the concerns related to the workflow optimization topic, for example, a 2023 study [2] suggests that the integration of machine learning in the field of business processes is bringing a transformative wave in the field of automation and also significantly elevating the operational efficiency, so a sufficient alignment of machine learning model development with business goals of workflow optimization is crucial for the identification of effective optimization methods. This research aims to explore the application of machine learning as an integral part of the workflow optimization.

This study specifically explores the category of input data optimization by exploring the capabilities of machine learning applications with the goal of improving the quality of the observed workflows. The quality improvement, in the scope of this work, implies increasing the workflow success rate by leveraging the historical data of similar transactions.

BIR-WS 2024: BIR 2024 Workshops and Doctoral Consortium, 23rd International Conference on Perspectives in Business Informatics Research (BIR 2024), September 11-13, 2024, Prague, Czech Rep.

Oskars.Rodins@edu.rtu.lv (O.Rodiņš)

Studies suggest that machine learning models are being employed by the industry at a large scale to effectively manage the risk of financial damage [3] and the use of artificial intelligence and machine-learning applications has increased significantly in the banking, financial services and insurance sector (BFSI) due to its potential to automate processes, enhance decision-making, improve customer experience and detect fraud [4]. With the rise of popularity of large language models (LLM), such as ChatGPT, new opportunities and use-cases for this technology are being researched, for example, a 2023 study [5] describes the application of ChatGPT for various conceptual modelling tasks, such as generation of ER, Business Process, UML and other models, based on textual input. The experiments conducted within the study [5] showed the enormous potential of large language models such as ChatGPT for supporting conceptual modeling tasks, indicating the potential of AI application in the workflow optimization topic.

There is a wide array of possibilities for the use of AI for workflow optimization. This paper explores a specific application: the use of classification models to predict the result of a card deposit workflow based on the historical payment data. This application was selected due to its practical relevance. The machine learning model developed within this study can be used, for example, to adjust the input parameters of a card payment workflow to maximize the likelihood of a positive outcome. Thus, the paper focuses on two research questions:

- RQ1 - What are ML applications in finance?
- RQ2 - Is machine learning an effective tool for card payment workflow input data optimization?

The objective of this study is to perform preliminary assessment of the efficiency of machine learning for workflow optimization in the financial sector. The contribution of the paper is the literature review on the topic of ML application in the finance sector and the machine learning models built during the research, their comparison and the conclusions drawn based on the results achieved.

The rest of the paper is organized as follows. The overview of research related to the topic of ML applications in the financial industry is conducted in Section 2. Section 3 describes an experiment: development of a ML model aimed to predict card deposit workflow results, based on the historical transaction data. Section 4 concludes.

2. Literature on ML applications for workflow optimization

To answer the RQ1 and gain a better understanding of the existing research on the topic of ML applications in the finance sector, a review of existing literature was conducted. A query was requested in the Scopus abstract and citation database with the following search parameters:

- Search target: article title, abstract, keywords.
- Search terms: workflow, optimization, machine learning.
- Range: 2015 –2023.

The resulting data suggests that the research interest for the topic of machine learning and workflow optimization has been gradually increasing in recent years. (Figure 1):

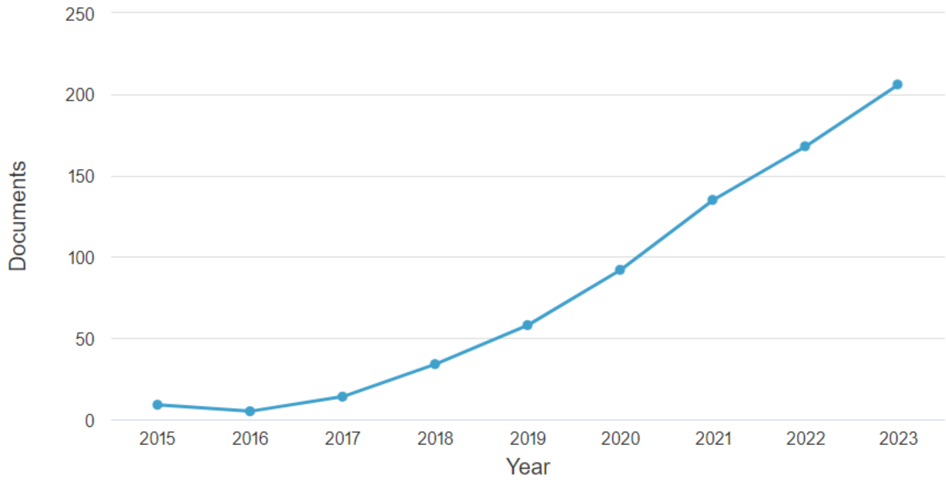

Figure 1: Scopus database - yearly number of publications search by keywords: workflow, optimization, machine learning (2015 - 2023). Number of publications increased by 58% yearly on average.

The topics of workflows and machine learning cover a wide range of disciplines, which makes the optimization approaches a universal tool that can be applied to workflows in different areas. Based on the data from the Scopus database, the studies related to machine learning application in the context of workflow optimization mostly fall into the following subject areas:

- Computer science: 16.5%
- Other (physics and astronomy, environmental science, finance and other): 16.4%
- Engineering: 13.3%

A more detailed division into subject areas is displayed below (Figure 2):

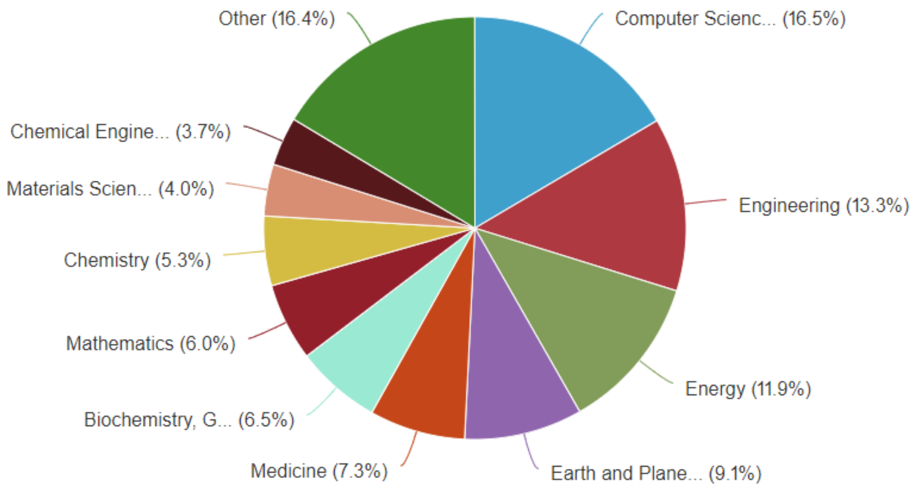

Figure 2: Scopus database - yearly number of publications search by keywords: workflow, optimization, machine learning divided into subject areas.

2.1. ML Taxonomy

Machine learning, a subset of artificial intelligence, encompasses a broad range of algorithms and techniques that enable computers to learn from data and make predictions or decisions without being explicitly programmed [6]. The rapid advancement of machine learning in recent years has led to a proliferation of algorithms, each with their unique characteristics and applications. To understand of this vast landscape of machine learning techniques, this section provides a brief overview of commonly used algorithm types [7] is describe below.

1. **Supervised learning**: a subset of machine learning algorithms that learn from labeled training data, where each input is paired with the correct output. Supervised learning models can be further divided into classification and regression models. Classification models aim to predict if the input data is affiliated to a certain target class and regression models aim to determine the specific target value with the given the input data.
2. **Unsupervised learning**: a subset of machine learning algorithms that learn from unlabeled data, finding patterns and relationships in the data, typically used for clustering tasks, to find patterns and groups data into clusters.
3. **Reinforcement learning**: algorithms learn to make sequences of decisions by interacting with an environment and receiving rewards or penalties. Reinforcement learning problems are related to understanding, which is the best action to perform, situation-by-situation, to maximize the aggregated reward. RL agent has to learn a policy by trying actions without any domain expert has told it, as in other forms of machine learning [8].

2.2. ML applications in the financial sector

As this study focuses on the application of ML as means of improving a workflow, specific to the financial sector (payment card deposit), this section provides a brief overview of various ML use-cases and examples found in other publications. According to a 2024 study [9], the financial industry is being significantly impacted by the rising influence of artificial intelligence and machine learning, especially in the risk management sectors of credit, liquidity, market and operational risk. According to the publication in the International Journal of Accounting, Finance, Auditing, Management and Economic [10] ML applications in financial sector can be classified into the following 7 categories described below.

2.2.1. Portfolio management and robo-advisory

AI and ML are being employed to improve the customer experience, increase the efficiency and accuracy of operational workflows, and enhance performance by supporting multiple aspects of the investment process [11]. Companies leverage ML for adjusting investment portfolios for the needs of specific customers, an example of this is the launch of the robo-advisor in the online bank Revolut in February 2024 [12], the company states that the technology can automatically rebalance customer portfolios based on the performance of the assets within the portfolio and perform periodic reviews to maintain customer risk tolerances and target portfolio allocations.

A study published in the Journal of Physics Conference Series [13] observed five algorithms: ARIMA, SVR, LSTM, and Facebook Prophet to predict the price movements of various stocks. The result suggests that supervised and unsupervised ML algorithms can be effectively used to recognize the trends that exist in the financial markets.

2.2.2. Risk management and financial distress prediction

The study by the International Journal of Accounting, Finance, Auditing, Management and Economic [10] concludes that many authors and scholars used various statistical and ML methods to develop financial prediction models and tested their effectiveness in the two main topics of financial distress prevention: credit scoring and bankruptcy prediction. Supervised ML algorithms like SVM and ANN have been used to assess credit risk with promising performance accuracy. A 2022 study [14] on the topic of ML applications for bankruptcy prediction has performed experiments with various supervised machine-learning algorithms: AdaBoost, Decision tree, J48, Bagging and Random Forest on a data from developing business sectors across the globe for a 5-year period and concludes that the highest accuracy for bankruptcy prediction: 97%, was obtained by the Bagging algorithm.

2.2.3. Financial fraud detection and AML

A common application for ML application in the financial sector is fraud detection and AML. As the regulators often require the financial institutions to rapidly detect and react to any suspicious financial activity, and the volumes of financial transactions are constantly increasing worldwide. Data from statista.com [15] suggests that total digital payment value is expected to show an annual growth rate (2024-2028) of 9.52% resulting in a projected total amount of US$16.59tn by 2028 (Figure 3).

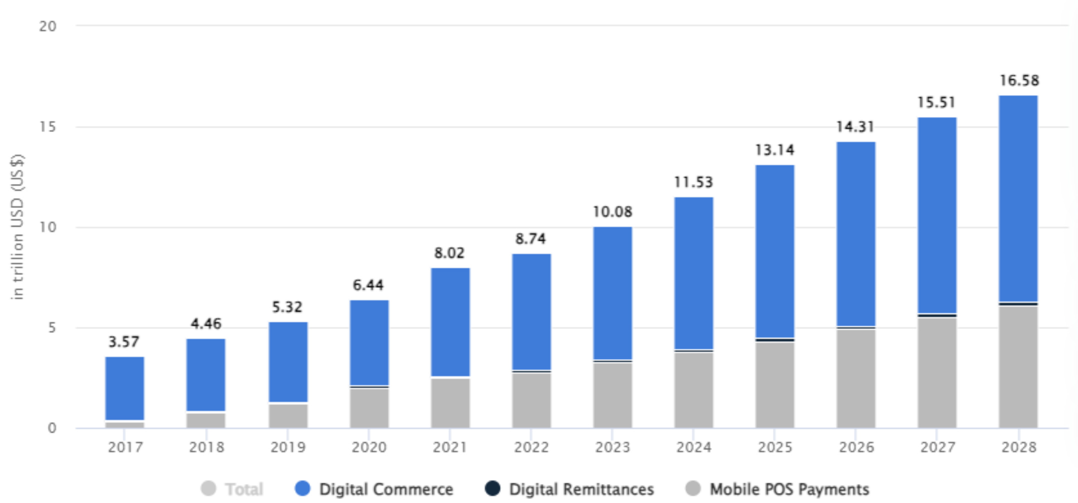

Figure 3: Statista.com - total transaction value in the Digital Payments market 2017-2028, historical and projected values. (https://www.statista.com/outlook/fmo/digital-payments/worldwide#transaction-value).

The increasing digital payment volumes suggest that the requirements for the quality and efficiency of the fraud-detection and AML systems are also increasing, making the potential benefits of ML applications valuable for the industry. Research has proven the effectiveness of the ML models in detecting fraudulent companies and many other researchers have focused their analysis on structured data using data mining models [10]. In a 2011 study [16], a group of researchers have been able to identify and track organizations that commit financial statement fraud using ML algorithms such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN), with PNN outperforming all the techniques without feature selection.

There are many studies focusing on the application of ML for AML processes, artificial intelligence is used to detect fraudulent transactions, determine risk scores for companies and individuals and predict money-laundering activities and detecting credit card fraud. For example, a 2022 study [17] has conducted a comparative analysis of the literature review considering the ML techniques for credit card fraud detection by observing various supervised ML algorithms: Random Forest, Artificial Neural Network and Support Vector Machine proposed a hybrid solution, using the neural network (ANN) in a federated learning framework as the most effective solution.

2.2.4. Sentiment analysis and investor behavior

Sentiment analysis is a great example of ML in finance and is critical for all firms in today's workplace [18]. Companies apply ML to analyze large quantities of data to make predictions about costumer behavior and market trends and use it to make data-driven business decisions. The most prevalent use of sentiment analysis in the financial field is the study of financial news, specifically anticipating market behavior and potential trends [10], as an example, a 2014 study [19] analyzed 32 million Yahoo! Finance messages to evaluate the potential of ML to forecast stock returns, trading volume, and volatility and found evidence indicating that past stock price performance positively influenced investor sentiment. The 2022 study [10] on the use of ML in finance suggests that exploring social media, entertainment platforms, and other relevant data sources to predict client sentiments will play a significant role in the future applications of machine learning.

2.2.5. Stock market prediction

As the ML presumes the potential to make predictions about future values based on historical data, one of its possible uses in the financial sector is the stock market price movement prediction. The advantages of AI compared to traditional econometric models have ignited significant academic interest in applying machine learning to algorithmic trading. Research [10] on the use of ML in finance points out that machine learning algorithms such ANN, SVM and reinforcement learning have been used to:

- make stock market price predictions
- recognize triggers for market irregularities
- make decisions related to purchasing, holding, or selling a stock
- acquiring unique insights into market movements

2.2.6. Data protection and cyber security

The introduction of ML systems has reduced the costs of data engineering and pre-processing, leading many banks and financial institutions to adopt AI for enhancing customer experience. This shift has automated operations like account creation, money transfers, and bill payments through mobile apps, reducing the need for personnel. AI chatbots, using natural language processing, efficiently address customer inquiries. However, AI systems face data protection and cybersecurity challenges, with vulnerabilities to new security threats [20]. In cybersecurity, AI and ML are crucial for malware detection, identifying hacker attacks, and ensuring network security. Organizations require robust anti-malware, anti-spyware, firewalls, and intrusion prevention systems. A 2016 study [21] on the applications of text mining in the financial domain categorized financial service cybersecurity ML applications into the following categories:

- phishing detection
- spam detection
- malware detection
- intrusion detection
- fraud detection

2.2.7. Big data analytics

Financial institutions often obtain large amounts of historical data, which encourages the application of ML to use the data to find patterns and make predictions, allowing the institutions to make data-driven decisions. A 2023 study suggests that ML models and big data are increasingly being used by banks and financial institutions to assess the creditworthiness of prospective borrowers and make underwriting decisions [22]. The companies can use the data to their advantage, for example: to train and test ML models for various purposes:

- solving classification problems, by predicting the outcome of various business processes
- solving regression problems, by predicting the value of a certain target value
- clustering problems, by finding patterns and grouping data into clusters

3. Experiment: ML for card payment workflow optimization

In this section, the author seeks to answer the RQ2 - is machine-learning an effective tool for card payment workflow input data optimization? To answer the question, the author conducts a practical experiment of ML application for workflow optimization: leveraging historical transaction data to train and test a ML model. The trained model then can be applied to the workflow input data, to select the best possible set of parameters.

3.1. Experiment goal

The goal of the experiment is to develop a machine learning model that would predict the outcome of a card deposit workflow based on historical data about previously executed card deposit workflows. Depending on the results, the author draws conclusions whether applying a ML model for workflow input data optimization is an effective optimization technique.

3.2. Experiment context

To conduct the experiment, the author collected historical data on card deposit workflows (transactions made during a 7-day period in April 2024) that were conducted in an online bank, which services about 300’000 client accounts worldwide and provides internet-banking services such as online personal and business account opening, wire payments, card operations, currency exchange, investments etc. As the data used originates from real-life transactions in an online-bank, the experiment results can be useful for similar organizations that utilize workflows where one of the input parameters can have various categorical values.

3.3. Optimization purpose

In the context of the observed card deposit workflow, a specific workflow parameter “provider” is chosen randomly by the system. The “provider” parameter in this case is a variable that indicates which payment provider will perform the card charging operation. The optimization purpose is to leverage the historical transaction data and, with the given input parameters, make predictions with the ML model about which payment provider would show a higher probability of a successful workflow completion and using the best result to initiate the workflow. The developed ML model can be applied to optimize the input data of the workflow by performing predictions with various values of “provider” parameter to select the best value for workflow execution. A graphical representation of the two scenarios is displayed below (Figure 4).

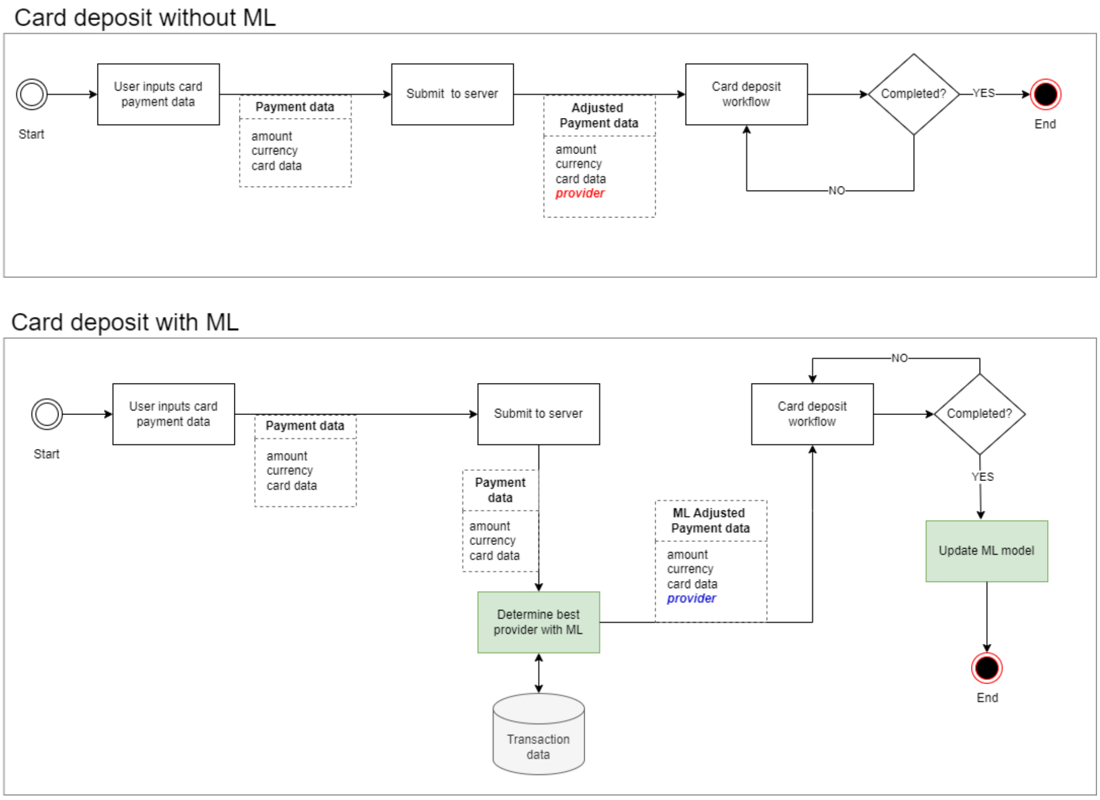

Figure 4: Diagram for the card deposit workflow for 2 scenarios, without and with assistance of ML to determine provider.

3.4. Dataset

3.4.1. Initial dataset

To build the ML model, historical data about card deposit workflows (transactions) has been collected for the period of 7 days: from 01.04.2024 to 07.04.2024. Total of 1826 have been collected in the initial dataset The structure of the dataset is displayed below (Table 1)

Table 1
Card deposit transaction dataset structure (with examples)

amount	currency	card_bin	card_issuer_country	service_provider_id	sub_account_type_id	status
314.68	USD	535825	SA	4	21	1

The target attribute, that the model will attempt to predict is status, it can hold two values:

- Status: 1, meaning the deposit was completed successfully
- Status: 2, meaning the deposit was completed non-successfully

The number of transactions with status 1 and 2 is displayed below (Figure 5):

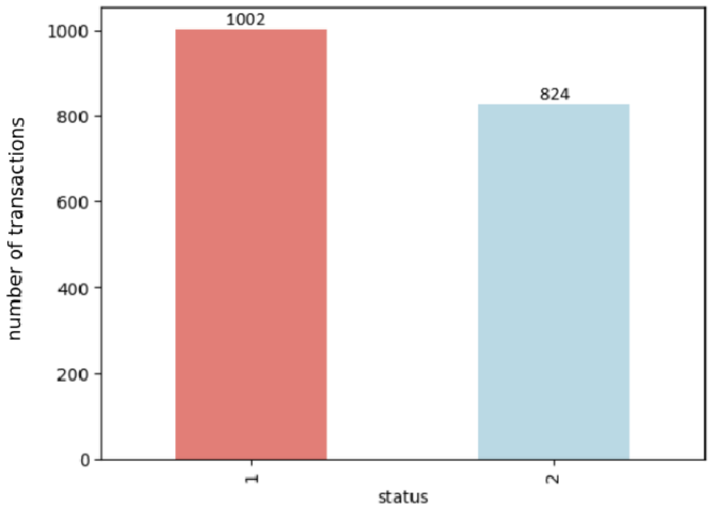

Figure 5: The number of transactions with status 1 and 2 in the initial dataset. The proportion of transactions is not completely balanced, with 54% of transactions being completed successfully.

3.4.2. Data preparation for ML

Before using the gathered data to train and test the model, various activities had to be performed:

- Removal of missing values: the dataset contained missing values in 21 row (1.15% of all entries), to the missing values were removed from the dataset.
- Outliers: the parameter “amount” of the initial dataset contained values far beyond the third quartile. To remove the outliers, interquartile range technique [21] was applied (removing data that is more than 1.5 * IQR above the 3rd quartile and below the 1st quartile).

After the adjustments to the initial dataset, the resulting number of entries has been reduced to 1644.

3.5. ML model selection

The objective of the model is to predict the value of the attribute “status”, which has two possible values: 1 or 2 (successful or non-successful), hence a classification problem. As the model was developed with the tools offered by the open-source Scikit-learn machine learning library, the author chose to select models offered by this library. The model Scikit-learn model selection diagram [23] is displayed below (Figure 6). From the various classification models, offered by the Scikit-learn library, the author selected the following models:

- KNN (K-Nearest Neighbor classifier)
- Logistic Regression
- Random Forest classifier

To train and test the 3 models, the dataset has been divided into two splits:

- Training set: 80% of the dataset, used to train the models
- Test set: 20% of the dataset, used to test the models and evaluate performance

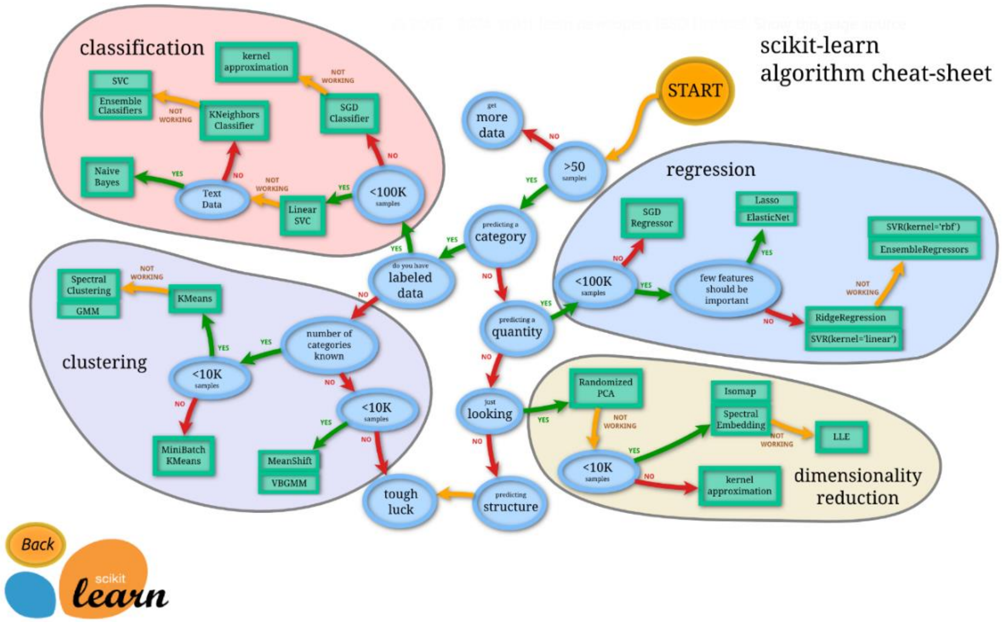

Figure 6: ML model algorithm selection diagram [Public domain], via Scikit Learn. (https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html).

3.6. Experiment results

After training and testing the chosen models, the following accuracy was achieved (Figure 7) for the ability to predict a successful or unsuccessful card deposit workflow result. The accuracy is measured as the proportion between the model's correct predictions and total predictions for the test set, expressed as a percentage:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

As the dataset was not completely balanced (18% difference), the accuracy of the observed models has been complemented with the F1 metric [24], which can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The formula for the F1 score is:

$$F1 = \frac{2 \times \text{True Positives}}{2 \times \text{True Positives} + \text{False Positives} + \text{False Negatives}}$$

The best performance was shown by the Random Forest Classifier (accuracy: 0.74, F1: 0.74), followed by the Logistic Regression (0.73, F1: 0.73) and KNN model (0.60, F1: 0.61).

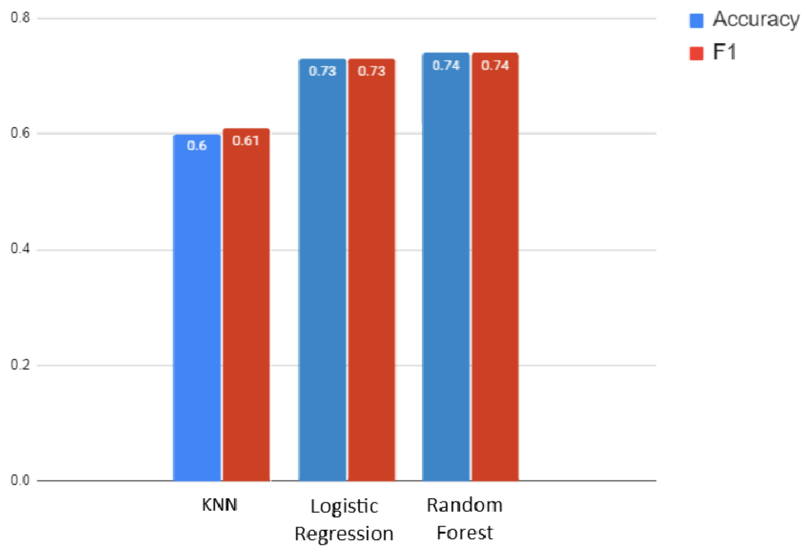

Figure 7: Accuracy and F1-score achieved by ML models for predicting an un-successful card deposit workflow result.

To determine which model could generate the highest accuracy, additional experiments were performed with various model training parameters. First, the KNN model, which displayed the worst result (0.60) was re-trained with an increased counts of "n_estimators", however the accuracy did not increase over 0.65, so the KNN model could not outperform the other models. As a result, the KNN model was discontinued.

The experiments continued with the two other models: Random Forest Classifier and Logistic Regression. Attempts to increase the model's accuracy were made with the help of Scikit Learn library RandomizedSearchCV [25], which randomly applies various model creation parameters and performs cross-validation to find a set of parameters for highest accuracy. The highest accuracy was still achieved by the Random Forest Classifier, reaching the accuracy of 0.78 and F1-score of 0.76. The final set of parameters used to train the model were the following:

- "n_estimators": 300
- "max_depth": 5
- "min_samples_split": 5
- "max_samples_leaf": 5
- "cv": 2

To find the attributes, that have the highest impact on the un-successful completion of the card deposit workflow, a function of Scikit Learn library "feature_importances_" was used [26], which computes the importance as the mean and standard deviation of accumulation of the impurity decrease within each tree. As a result, the attributes that most significantly contribute to unsuccessful completion of the card deposit workflow are displayed below (Figure 8):

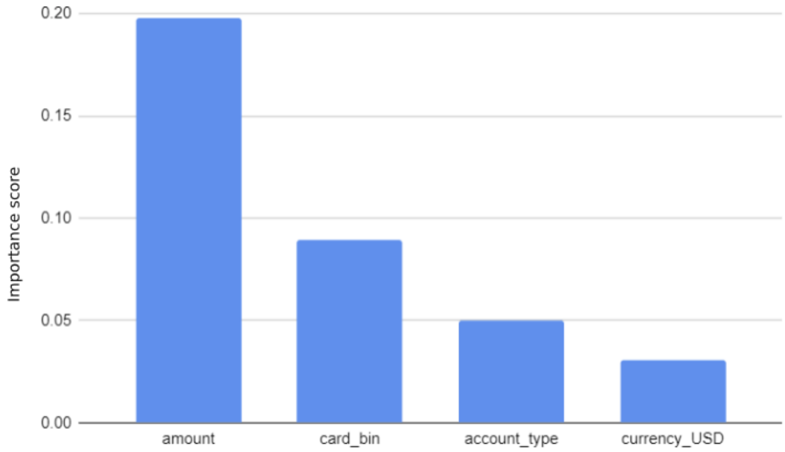

Figure 8: Attributes that have the highest impact on unsuccessful completion of card deposit workflow.

4. Conclusion

This study explores the topic of ML application in the financial sector and focuses on two research questions:

- RQ1 - What are ML applications in finance?
- RQ2 - Is machine-learning an effective tool for card payment workflow input data optimization?

To answer the RQ1, a review of existing literature was conducted. The data gathered from the Scopus database suggests that the interest for the topic of machine learning and workflow optimization has been gradually increasing during 2015 -2023 (number of articles increases by 58% yearly on average). The literature review performed within this study suggests, that the applications of machine learning in the financial sector can be categorized into the following groups:

- Portfolio management and robo-advisory
- Risk management and financial distress prediction
- Financial fraud detection and AML
- Sentiment analysis and investor behavior prediction
- Stock market prediction
- Data protection and cyber security
- Big data analytics

To answer the RQ2, an experiment was conducted in this study, 3 machine learning classification models were built with the aim to predict the outcome of a card deposit workflow result, and the best classification metrics (accuracy: 0.78, F1: 0.76) were achieved by the Random Forest Classifier.

The experiment confirms that a machine-learning model can be used as a tool for workflow input data optimization, for cases when one of the input variables can have various values, as it can predict the best possible value for the variable and thus increase the probability of a desired outcome, however RQ2 is not fully answered, as it requires further evaluation of the performance of the developed model on optimized card deposit workflows. The research displays a possible use case for the alignment of machine learning model development with business goals of workflow optimization. Overall, the author can make the following recommendations for future research on the topic:

- to verify the effectiveness of this optimization approach, further research is needed to compare the results of optimized and non-optimized workflows
- to provide higher accuracy, the dataset used to build the ML model should cover a larger period, to increase the variety of data
- experiments with various types of ML algorithms (depending of the goal) should be performed to further evaluate the effectiveness of ML as a workflow optimization tool
- although the outliers were removed in the dataset used to build the ML model, the dataset was not normalized, which could lead to worse accuracy

Acknowledgements

The author expresses gratitude Dr. eng. sc. Jānis Grabis for guiding the research process and providing mentorship during the academic activities. The author also would like to thank his colleagues for their support during the research process.

References

[1] O.Rodiņš, Workflow optimization at financial institutions: survey and case study, Joint Proceedings of the BIR 2023 Workshops and Doctoral Consortium co-located with 22nd International Conference on Perspectives in Business Informatics Research (2023).

[2] J. K. Chaudhary, R. Sobti, et al., Machine Learning Strategies for Business Process Optimization, 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (2023). doi: 10.1109/UPCON59197.2023.10434915

[3] V. Siddharth, J. Manoj, Applying Machine Learning Methods for Credit Card Payment Default Prediction With Cost Savings, Biomedical and Business Applications Using Artificial Neural Networks and Machine Learning, IGI Global, pp. 285-305. (2022).

[4] D. Pattnaik, S. Ray, R. Raman, Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review, Journal of Behavioral and Experimental Finance. doi: https://doi.org/10.1016/j.jbef.2021.100577

[5] H. G. Fill, P. Fettke, J. Köpke, Conceptual Modeling and Large Language Models: Impressions From First Experiments With ChatGPT, Enterprise Modelling and Information Systems Architecture, Vol. 18, No. 3 (2023). doi: https://doi.org/10.18417/emisa.18.3

[6] S. Belkacem, Machine learning approaches to rank news feed updates on social media, Machine Learning, Ph. D. thesis, University of Science and Technology HOUARI BOUMEDIENE, Algeria (2021). URL: https://www.researchgate.net/publication/350993856_Machine_learning_approaches_to_rank_news_feed_updates_on_social_media

[7] T. O. Ayodele, Types of Machine Learning Algorithms, New Advances in Machine Learning, InTech (2010). doi: http://dx.doi.org/10.5772/9385

[8] M. Naeem, S. T. H. Rizvi, A. Corontano, A Gentle Introduction to Reinforcement Learning and its Application in Different Fields, IEEE Access (2020). doi: 10.1109/ACCESS.2020.3038605

[9] T. Muhammad, S. Ness, Exploring AI and Machine Learning Applications in Banking: A Comprehensive Review of Literature, International Journal of Innovative Science and Research Technology (2024). doi: 10.5281/zenodo.10707332

[10] W. Lakhchini, R. Wahabi, M. El Kabbouri, Artificial Intelligence & Machine Learning in Finance: A literature review, International Journal of Accounting, Finance, Auditing, Management and Economic (2022). doi: 10.5281/zenodo.7454232

[11] Blackrock, Artificial intelligence and machine and learning in asset management, 2019. URL: https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-artificial-intelligence-machine-learning-asset-management-october-2019.pdf

[12] Revolut, Revolut launches Robo-Advisor in EEA to automate investing, 2024. URL: https://www.revolut.com/en-LV/news/revolut_launches_robo_advisor_in_eea_to_automate_investing/
[13] A. Potdar, M. B. Pande, Comprehensive Analysis of Machine Learning Algorithms Used in Robo-Advisory Services, Journal of Physics Conference Series (2021). doi: 10.1088/1742-6596/1964/6/062105
[14] M. Keya, H. Akter, A. Rahman et al., Comparison of Different Machine Learning Algorithms for Detecting Bankruptcy, 6th International Conference on Inventive Computation Technologies (2021). doi: 10.1109/ICICT50816.2021.9358587
[15] Statista, Digital Payments - Worldwide, 2024. URL: https://www.statista.com/outlook/fmo/digital-payments/worldwide
[16] P. Ravisankar, V. Ravi, G. Raghava Rao, I. Bose, Detection of financial statement fraud and feature selection using data mining techniques, Decision Support Systems, Volume 50, Issue 2, pp. 491-500 (2011). doi: 10.1016/j.dss.2010.11.006
[17] R. Bin Sulaiman, V. Schetinin, P. Sant, Review of Machine Learning Approach on Credit Card Fraud Detection, Human-Centric Intelligent Systems (2022). doi: 10.1007/s44230-022-00004-0
[18] H. Tatsat, S. Puri, B. Lookabaugh, Machine Learning and Data Science Blueprints for Finance, O'Reilly Media, Inc., Sebastopol, CA (2020)
[19] S. Kim, D. Kim, Investor sentiment from internet message postings and the predictability of stock returns, Journal of Economic Behavior & Organization Volume 107, pp. 708-729 (2014). doi: doi.org/10.1016/j.jebo.2014.04.015
[20] B. C. Stahl, Conclusion. Artificial Intelligence for a Better Future, SpringerBriefs in Research and Innovation Governance, Springer, pp. 117-122 (2021). doi: doi.org/10.1016/j.jebo.2014.04.015
[21] H. P. Vintuha, B. Poornima, B. M. Sagar, Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset, Information and Decision Sciences, pp. 511-518 (2018). doi: 10.1007/978-981-10-7563-6_53
[22] M. Belhaj, Y. Hachaïchi, Artificial Intelligence, Machine Learning and Big Data in Finance Opportunities, Challenges, and Implications for Policy Makers (2023). doi: 10.13140/RG.2.2.27950.18248
[23] Scikit-learn, Choosing the right estimator, 2024. URL: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html
[24] Scikit-learn, f1_score, 2024. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html
[25] Scikit-learn, RandomizedSearchCV, 2024. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html
[26] Scikit-learn, Feature importance based on mean decrease in impurity 2024. URL: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

A. Online Resources

Scopus database: https://www.elsevier.com/products/scopus
Scikit learn open-source machine-learning library: https://scikit-learn.org/stable/