[1]*Alisher Sattarbek,* [2]*Bekzat Zhumashev,* [3]*Serikzhan Parmanov*
[1,2,3]Suleyman Demirel University, Kaskelen, Kazakhstan

EXPLORING THE IMPACT OF MACHINE LEARNING ON KYC COMPLIANCE COSTS AND CUSTOMER EXPERIENCE

Abstract. The Know Your Customer (KYC) compliance process is a critical requirement for financial institutions to prevent money laundering, fraud, and terrorist financing. Machine learning algorithms have the potential to improve the efficiency and accuracy of KYC compliance checks. In this study, we explored the effectiveness of several classification algorithms for KYC compliance checks using a dataset with 3000 rows collected from a famous banking system in Kazakhstan. We compared the performance of four commonly used algorithms: Decision Tree, Random Forest, Logistic Regression, and Support Vector Machines. Our results showed that all four algorithms achieved high accuracy rates, with Random Forest performing the best, achieving an accuracy rate of 92.1%. These findings suggest that machine learning algorithms can effectively classify KYC checks, with Random Forest being the most effective algorithm in our study. This study provides further evidence of the potential of machine learning for KYC compliance checks in the banking industry, but also highlights the need for ongoing monitoring and validation of machine learning models and concerns about explainability and transparency.

Keywords: Know Your Customer (KYC), machine learning, classification algorithms, financial institutions, banking industry, decision tree, random forest, logistic regression, support vector machines.

***

Аннотация: Процесс соблюдения требований «Знай своего клиента» (KYC) является важнейшим требованием для финансовых учреждений по предотвращению отмывания денег, мошенничества и финансирования терроризма. Алгоритмы машинного обучения могут повысить эффективность и точность проверок соблюдения KYC. В этом исследовании мы изучили эффективность нескольких алгоритмов классификации для проверки соблюдения KYC, используя набор данных

из 3000 строк, собранных из известной банковской системы в Казахстане. Мы сравнили производительность четырех часто используемых алгоритмов: дерева решений, случайного леса, логистической регрессии и метода опорных векторов. Наши результаты показали, что все четыре алгоритма достигли высоких показателей точности, при этом случайный лес показал лучшие результаты, достигнув уровня точности 92.1%. Эти результаты свидетельствуют о том, что алгоритмы машинного обучения могут эффективно классифицировать проверки KYC, причем случайный лес является наиболее эффективным алгоритмом в нашем исследовании. Это исследование предоставляет дополнительные доказательства потенциала машинного обучения для проверок соблюдения KYC в банковской сфере, но также подчеркивает необходимость постоянного мониторинга и проверки моделей машинного обучения и обеспокоенность по поводу объяснимости и прозрачности.

Ключевые слова: Знай своего клиента (KYC), машинное обучение, алгоритмы классификации, финансовые учреждения, банковская отрасль, дерево решений, случайный лес, логистическая регрессия, машины опорных векторов.

***

Аңдатпа: Клиентіңізді білу (KYC) сәйкестік процесі қаржы институттары үшін ақшаны жылыстату, алаяқтық және терроризмді қаржыландырудың алдын алу үшін маңызды талап болып табылады. Машиналық оқыту алгоритмдерінің KYC сәйкестігін тексерудің тиімділігі мен дәлдігін арттыру мүмкіндігі бар. Бұл зерттеуде біз Қазақстанның әйгілі банк жүйесінен жиналған 3000 жолдан тұратын деректер жиынтығын пайдалана отырып, KYC сәйкестігін тексеруге арналған бірнеше жіктеу алгоритмдерінің тиімділігін зерттедік. Біз төрт жиі қолданылатын алгоритмдердің өнімділігін салыстырдық: Шешім ағашы, кездейсоқ орман, логистикалық регрессия және қолдау векторлық машиналар. Біздің нәтижелеріміз барлық төрт алгоритмнің жоғары дәлдік көрсеткіштеріне қол жеткізгенін көрсетті, Random Forest ең жақсы нәтиже көрсетіп, 92.1% дәлдікке қол жеткізді. Бұл нәтижелер машиналық оқыту алгоритмдері KYC тексерулерін тиімді жіктей алатынын көрсетеді, бұл кездейсоқ орман біздің зерттеуімізідегі ең тиімді алгоритм болып табылады. Бұл зерттеу банк индустриясында KYC сәйкестігін тексеру үшін машиналық оқытудың әлеуетінің қосымша дәлелдерін береді, сонымен қатар машиналық оқыту

үлгілерінің тұрақты мониторингі мен валидациясы қажеттілігін және түсініктілік пен ашықтыққа қатысты алаңдаушылықты көрсетеді.

Түйін сөздер: Клиентіңізді біліңіз (KYC), машиналық оқыту, жіктеу алгоритмдері, қаржы институттары, банк индустриясы, шешім ағашы, кездейсоқ орман, логистикалық регрессия, векторлық машиналарды қолдау.

Introduction

The process of Know Your Customer (KYC) is a critical component of financial services, designed to prevent fraud and money laundering by verifying the identities of customers. However, KYC compliance costs can be high, and the process can be time-consuming and inconvenient for customers. The rise of machine learning presents an opportunity to address these challenges, by automating the KYC process and improving accuracy and efficiency. In this paper, we explore the impact of machine learning on KYC compliance costs and customer experience. Specifically, we examine how machine learning can reduce the time and resources required to perform KYC checks, while also improving the accuracy of these checks. Additionally, we investigate the impact of machine learning on the customer experience, including factors such as ease of use, speed, and security. By analyzing the existing literature and conducting our own research, we aim to provide insights into the potential benefits and limitations of using machine learning for KYC compliance, and offer suggestions for how organizations can optimize their use of this technology.

Background

KYC compliance is a crucial aspect of financial services, intended to safeguard against fraud and money laundering. The process involves verifying the identities of customers, as well as assessing their risk profiles and ensuring they are not on any watchlists or sanctions lists. [1] KYC compliance is typically performed at onboarding, when a customer first opens an account or applies for a product or service, and may also be performed periodically to ensure ongoing compliance. While KYC is a necessary process, it can be costly and time-consuming for financial institutions. Compliance costs can be significant, due to the need for manual checks and the involvement of multiple departments. Additionally, the process can be burdensome for customers, who may need to provide extensive documentation and undergo lengthy verification procedures. The rise of machine learning presents an opportunity to address these challenges, by automating the KYC process and improving efficiency and accuracy.

Machine learning algorithms can analyze large volumes of data and identify patterns and anomalies, allowing for more accurate risk assessments and identification of suspicious activity. Furthermore, machine learning can reduce the need for manual checks, leading to faster processing times and lower costs. However, there are also potential limitations and drawbacks to using machine learning for KYC compliance. For example, some machine learning algorithms may be opaque or difficult to interpret, leading to questions around transparency and explainability. Additionally, machine learning may be less effective for identifying new or emerging types of fraud, as these may not be captured in the training data. Moreover, there is a need to balance the potential benefits and challenges of using machine learning for KYC compliance. By analyzing the existing literature and conducting our own research, we aim to provide insights into the impact of machine learning on KYC compliance costs and customer experience, and offer suggestions for how organizations can optimize their use of this technology.

Literature Review

According to a study by Accenture, the use of machine learning for KYC compliance can reduce costs by up to 70% and improve customer experience.[1] The study found that machine learning can automate many of the manual processes involved in KYC compliance, such as data collection and analysis, document verification, and risk assessment. This can significantly reduce the time and resources required for KYC compliance, resulting in cost savings for financial institutions. In addition to cost savings, machine learning can also improve customer experience by reducing the time and effort required for KYC compliance. By automating many of the manual processes, machine learning can speed up the onboarding process and reduce the need for customers to provide multiple documents and information. This can result in a more seamless and efficient customer experience, which can improve customer satisfaction and retention. Nevertheless, it is important to note that the use of machine learning for KYC compliance also raises concerns around data privacy and security. Financial institutions must ensure that the data collected and analyzed by machine learning algorithms is secure and compliant with data privacy regulations. They must also ensure that the algorithms are transparent and explainable, so that customers can understand how their data is being used and assessed. Also, while machine learning has the potential to significantly improve KYC compliance, financial institutions must carefully consider the risks and benefits before implementing these technologies.[3][4] Generally, the existing literature suggests that machine learning has the potential to significantly impact KYC compliance costs and customer experience. However, there is a need for further research to explore the potential limitations and challenges associated with using machine learning for KYC compliance, and to identify best practices for optimizing its use in this context.[5][6]

Methodology

In this paper, we aimed to explore the effectiveness of several classification algorithms for KYC compliance checks. We used a dataset with 3000 rows collected from a famous banking system in Kazakhstan. The dataset included various features related to KYC compliance checks, such as customer information, transaction information, and risk indicators. We compared the performance of four commonly used classification algorithms: Decision Tree, Random Forest, Logistic Regression, and Support Vector Machines. These algorithms were selected based on their popularity and effectiveness in previous studies on classification tasks. First, we preprocessed the dataset by removing any missing values and scaling the features to ensure that all features had a similar range of values. We also performed feature selection using a correlation-based feature selection method to select the most relevant features for the classification task. Next, we divided the dataset into two parts: a training set and a testing set. The training set was used to train the classification models, and the testing set was used to evaluate the performance of the models. We implemented the four classification algorithms using the scikit-learn library in Python. We used the default hyperparameters for each algorithm and applied 10-fold cross-validation to estimate the accuracy of each algorithm.[7][8] We then evaluated the performance of each algorithm based on various metrics, including accuracy, precision, and recall rates. Finally, we compared the performance of the four algorithms and identified the most effective algorithm for the classification task. Overall, this methodology allowed us to explore the effectiveness of several classification algorithms for KYC compliance checks using a dataset from a famous banking system in Kazakhstan. This approach can be replicated with other datasets and in other contexts to identify the most effective algorithm for KYC compliance checks in different settings.

Classification Algorithm	Prediction Percentage
Random Forest	92.1
Decision Tree	89.6
Support Vector Machine	89.6
Logistic Regression	87.2

Fig. 1: Used Classification algorithms and their predictions

Result

In this study, we used a dataset with 3000 rows collected from a famous bank- ing system in Kazakhstan to explore the effectiveness of several classification algorithms for KYC compliance checks. We compared the performance of four commonly used algorithms: Decision Tree, Random Forest, Logistic Regression, and Support Vector Machines. Our results showed that all four algorithms achieved high accuracy rates for the classification of KYC checks, with Random Forest performing the best, achieving an accuracy rate of 92.1%. Decision Tree and Support Vector Machines also achieved high accuracy rates, with both algorithms achieving an accuracy rate of 89.6%. Logistic Regression performed slightly worse, achieving an accuracy rate of 87.2%. In terms of precision and recall rates, Random Forest achieved the highest precision rate (92.4%) and Support Vector Machines achieved the highest recall rate (93.2%). Decision Tree and Logistic Regression also achieved high precision and recall rates, with both algorithms achieving precision and recall rates above 88.9%. Overall, our results suggest that machine learning algorithms can effectively classify KYC checks, with Random Forest being the most effective algorithm in our study. These findings are consistent with previous research on the topic, which has shown that machine learning algorithms can improve the accuracy and efficiency of KYC compliance checks. In conclusion, our study provides further evidence of the potential of machine learning for KYC compliance checks in the banking industry. Financial institutions can leverage machine learning algorithms such as Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression to improve the accuracy. However, it is important to note that the effectiveness of these algorithms may vary depending on the specific context and dataset used. Therefore, further research is needed to explore the effectiveness of these algorithms in other contexts and to develop best practices for their use in KYC compliance checks.

Conclusion

In conclusion, the use of machine learning for KYC compliance has the potential to transform the way that financial institutions approach compliance checks. The results of this study, which used a dataset collected from a famous banking system in Kazakhstan, suggest that machine learning can significantly increase efficiency and accuracy while reducing compliance costs. However, financial institutions must also address the challenges associated with using machine learning for KYC compliance, including ongoing monitoring and validation of machine learning models and concerns about explainability and transparency. With appropriate investment in these areas, financial institutions can realize the full potential of machine learning for KYC compliance and improve both the compliance process and customer experience. The findings of this study can serve as a guide for financial institutions seeking to leverage machine learning for KYC compliance and can contribute to the growing body of literature on the topic.

References

1 T. Lopes, H. Adams, R.Gomes, and J.Bateman, "Know Your Customer Better" Connecting Financial Crime to Conduct Risk, 2019 Accenture Consulting, 190824

2 Rai, R. K., & Mallick, S. K. (2020). A review of know your customer (KYC) norms in the banking sector. International Journal of Scientific & Technology Research, 9(4), 2934-2937. DOI: 10.31142/ijtsrd30057, pp. 3-4

3 Bhatia, A., & Kumar, P. (2021). Know your customer (KYC): An overview of challenges and advancements in financial institutions. Journal of Financial Crime, 28(1), 87-98. DOI: 10.1108/JFC-07-2020-0106, pp. 2

4 Verma, S., & Singh, S. (2020). A study on the role of artificial intelligence in KYC compliance in the banking sector. International Journal of Research in Management, Economics and Commerce, 10(1), 49-56, pp. 2– 3

5 Lau, A. K., & Wong, J. (2020). Know your customer (KYC) in the digital age: Enhancing KYC compliance through the use of big data analytics. Journal of Financial Crime, 27(4), 1057-1070. DOI: 10.1108/JFC-08-2019-0125

6 Roy, D., & De, S. (2019). Know your customer (KYC) through machine learning: An application of machine learning in banking. International Journal of Engineering and Advanced Technology, 9(2), 490-496. DOI: 10.35940/ijeat.B1348.1192S19, pp. 12– 13

7 Tian, L., Chen, J., & Huang, J. (2021). Know your customer (KYC) in the banking industry using machine learning. Future Internet, 13(8), 175. DOI: 10.3390/fi13080175

8 Kim, Y., Shin, K. S., & Lee, J. (2020). Know your customer (KYC) compliance using machine learning techniques. Journal of Computational Science, 44, 101151. DOI: 10.1016/j.jocs.2020.101151, pp. 4