Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection

Zengyi Huang[1, a], Haotian Zheng[2, b], Chen Li[3, c], Chang Che[4, *]

[1] Applied Economics, The George Washington University, DC, 20052, US
[2] Computer Engineering, New York University, New York, NY, 10012, US
[3] Computer Science, The University of Texas at Dallas, TX, 75080, US
[4] Mechanical Engineering and Automation, The George Washington University, DC, 20052, US

[a]zengyihuang@gwmail.gwu.edu, [b]hz2687@nyu.edu, [c]cxl167330@utdallas.edu, *Corresponding author: Chang Che (E-mail: liamche1123@outlook.com)

Abstract: In today's increasingly digital financial landscape, the frequency and complexity of fraudulent activities are on the rise, posing significant risks and losses for both financial institutions and consumers. To effectively tackle this challenge, this paper proposes a machine learning-based K-means clustering method to enhance the accuracy and efficiency of financial fraud detection. By clustering vast amounts of financial transaction data, we can identify anomalous patterns and behaviors in a timely manner, thereby detecting potential fraud. Compared to traditional rule-based detection methods, machine learning-based approaches better adapt to ever-evolving fraud techniques and patterns while improving flexibility and precision in detection. Moreover, K-means clustering also aids in optimizing resource allocation within financial institutions by enabling focused monitoring and prevention efforts in high-risk areas, thus effectively mitigating the impact of fraud on the overall financial system. In summary, the machine learning-based K-means clustering method holds promising prospects for application in the field of financial fraud detection as it strives to establish a more secure and reliable transaction environment for the finance industry.

Keywords: Financial fraud detection, Machine learning-based, K-means clustering, Digital financial landscape.

1. Introduction

After the 1970s, with the development of the international financial system, worldwide economic integration, and the development of information technology, international finance has become increasingly inseparable. As a result, money-laundering crimes have become increasingly complex and difficult to identify. Therefore, all kinds of new money laundering methods continue to emerge, and money laundering activities are more and more frequent, which are characterized by: first, the domestic criminal proceeds are systematically transferred to foreign countries in the way of underground banks[1]. First, the money that is cleaned in the country is transferred from the recipient to the operator of the underground bank, and the laundered money is obtained from the accomplices abroad; The second is the signing of forged import contracts with foreign companies, the use of bank letters of credit payment, the transfer of criminal funds to overseas, the financial secrecy mechanism is relatively more strict countries and regions.

At the same time, financial fraud will also bring great impact to the world's financial system, and become an important factor restricting the sustainable and healthy development of China's insurance industry. Although the transaction reporting systems of various countries list a large number of suspicious transactions, for example, the United States "Anti-Money Laundering Work Guidelines" for the identification of more than 50 illegal transactions, the "Financial institutions Large Transactions and Suspicious Transaction reporting Management Measures" has clarified the basic characteristics of 48 suspicious transactions. However, the above guidelines only provide a basic screening structure for the identification of suspicious transactions, and usually the criteria are based on the existing ways and methods of money laundering, while some new money laundering activities are difficult to effectively identify[2]. The evolution of money laundering crime will make the above criteria ineffective, and this backward supervision means and discrimination criteria can not effectively deal with complex financial crimes. The identification of money laundering and suspected transactions is very complicated, not only

There should be a large amount of data, but also a more professional means of identification, and it should be constantly updated. As the current money-laundering behavior is highly concealed, the means emerge in an endless stream, and the transaction volume increases exponentially, it is necessary to make new supervision methods for financial fraud, so as to ensure financial security.

This paper explores the application of machine learning-based K-means clustering in the realm of financial fraud detection as a response to these challenges. By leveraging advanced data analytics techniques, including machine learning algorithms and clustering methodologies, the research aims to enhance the accuracy and efficiency of detecting suspicious financial transactions. In doing so, it seeks to address the limitations of traditional rule-based detection methods, which often struggle to adapt to evolving fraud techniques. Through the development and implementation of innovative detection frameworks, the study endeavors to contribute to the establishment of a more secure and reliable transaction environment within the global financial system.

2. Related Work

2.1. Machine learning and financial fraud detection

Machine learning is a subset of artificial intelligence that gives systems the ability to learn and improve automatically from experience without explicit programming. That said, we (humans) can already provide a computer with large data sets to learn patterns so that it can learn how to make decisions when faced with one or more new instances - and when I discovered this insight, immediately that the world was about to change. In machine learning terms, problems such as fraud detection can be classified as classification problems, where the goal is to predict discrete labels 0 or 1, where 0 usually means that the transaction is non-fraudulent and 1 means that the transaction appears to be fraudulent.

Traditional methods for credit fraud detection typically rely on rule-based systems and anomaly detection techniques. Rule-based systems involve setting predefined rules or thresholds based on historical patterns or expert knowledge to flag potentially fraudulent transactions. These rules may include transaction amount thresholds, frequency of transactions, geographic locations, or unusual transaction patterns[3][4]. Anomaly detection techniques, on the other hand, aim to identify outliers or deviations from normal behavior by analyzing statistical patterns in transaction data. These methods often involve calculating metrics such as mean, standard deviation, or z-scores to identify transactions that significantly differ from the norm.

In contrast, machine learning approaches have gained popularity in credit fraud detection due to their ability to learn from data and adapt to evolving fraud patterns. Various machine learning algorithms, including decision trees, logistic regression, support vector machines, and neural networks, have been applied to this task. Decision trees can effectively capture complex decision boundaries and identify important features for classification. [5]Logistic regression models provide probabilistic outputs and are well-suited for binary classification tasks. Support vector machines aim to find the optimal hyperplane that separates different classes in high-dimensional space. Neural networks, particularly deep learning models, have shown promise in capturing intricate patterns in large-scale datasets and detecting subtle fraud signals. These machine learning algorithms offer enhanced flexibility, scalability, and performance compared to traditional rule-based systems, making them valuable tools in the fight against credit fraud.

2.2. K-means Clustering in Financial Fraud Detection

Means clustering algorithm has been widely studied and applied in various fields, including but not limited to data mining, image processing, and pattern recognition. In the context of fraud detection, previous literature has explored the efficacy of K-means clustering in identifying anomalous patterns or clusters within financial transaction data. [6][7][8]These studies often highlight the algorithm's ability to partition data into distinct groups based on similarities in features such as transaction amount, frequency, and location, thus facilitating the detection of potentially fraudulent activities. Additionally, researchers have investigated different variations of K-means, such as weighted K-means or fuzzy K-means, to improve its performance in handling imbalanced datasets or capturing complex fraud patterns. Despite its popularity, challenges such as the need for careful initialization of centroids and the determination of the optimal number of clusters remain areas of active research. Overall, the existing literature underscores the importance of K-means clustering as a valuable tool in the arsenal of fraud detection techniques, offering insights into its strengths, limitations, and potential areas for further refinement and improvement.

Clustering algorithms, including K-means, DBSCAN, and hierarchical clustering, have been utilized in fraud detection to group similar transactions or entities together based on their characteristics. Previous studies have demonstrated the effectiveness of these techniques in identifying anomalous patterns or clusters that may indicate fraudulent behavior. K-means clustering, in particular, is widely used due to its simplicity, scalability, and efficiency in partitioning data into distinct clusters. It iteratively assigns data points to the nearest cluster centroid based on a specified number of clusters, making it suitable for detecting clusters of potentially fraudulent transactions. Additionally, K-means is computationally efficient and can handle large datasets, which is crucial for real-time fraud detection applications. While other clustering methods like DBSCAN and hierarchical clustering offer different advantages, such as the ability to detect irregularly shaped clusters or hierarchical structures, K-means is often preferred in fraud detection scenarios for its simplicity and computational efficiency, making it a pragmatic choice for practitioners in the field.

3. Methodology and Experiment

3.1. K-means Clustering Algorithm

K-means clustering algorithm is an iteratively solved clustering analysis algorithm. Its step is to randomly select K objects as the initial cluster center, then calculate the distance between each object and each seed cluster center, and assign each object to the nearest cluster center. [9]Cluster centers and the objects assigned to them represent a cluster. For each sample assigned, the cluster center of the cluster is recalculated based on the existing objects in the cluster. This process is repeated until a certain termination condition is met. The termination condition can be that no number of objects are reassigned to different clusters, nonumber of cluster centers change again, and the sum of squares of error is locally minimum.

K-Means algorithm implementation steps:

1.First determine a k value, that is, we want to cluster the data set to get k sets.

2. Randomly select k data points from the data set as the center of mass.

3.for each point in the data set, calculate its distance from each centroid (such as Euclidean distance), which centroid is close to, it is divided into the set that centroid belongs to.

4. After putting all the data together, there are a total of k sets. Then recalculate the center of mass of each set.

5. If the distance between the newly calculated center of mass and the original center of mass is less than a certain set threshold (indicating that the position of the recalculated center of mass does not change much, tends to be stable, or converges), we can consider that the clustering has reached the expected result and the algorithm terminates.

6. If the distance between the new center of mass and the original center of mass changes greatly, it is necessary to iterate 3 to 5 steps.

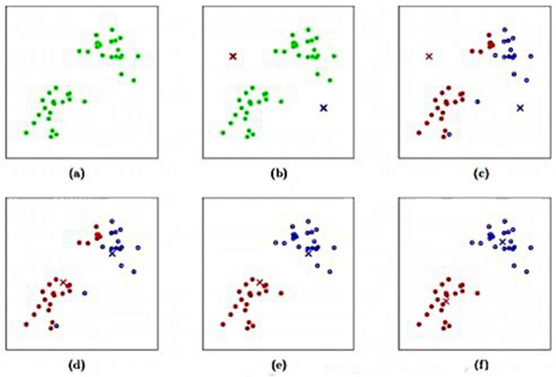


Figure 1. K-Means clustering iteration process

Certainly, here's a revised version with improved logic and organization:

Figure a represents the initial dataset with k=2 classes. In Figure b, we randomly select two centroids corresponding to the classes, represented by the red and blue centroids. Then, we calculate the distance between each point in the dataset and these centroids, assigning each point to the class of the nearest centroid.

Moving to Figure c, after the first iteration, we have assigned each point to either the red or blue class based on proximity to the centroids. Now, we calculate the new centroids for both classes.

As depicted in Figure d, the positions of the new centroids for both classes have changed compared to the initial centroids. Figures e and f illustrate the iterative process, repeating the steps of assigning points to the nearest centroid and updating the centroids until convergence.

Ultimately, in Figure f, we arrive at the final clustering result, where each point is assigned to either the red or blue class based on proximity to the centroids. This iterative process of updating centroids and reassigning points continues until the centroids stabilize and the classes converge.

K-value selection

The selection of K value has a great influence on K-means, which is also the biggest disadvantage of K-means. Common methods for selecting K value include elbow method and Gap statistic method.

If we get a sample, there are objectively J "natural subclasses", these real subclasses are hidden in the data. Below 3D data we can also draw a picture to distinguish the approximate number of J with the naked eye, higher dimensions can not be intuitively seen, we can only start from a relatively small K, such as K=2, to try to approximate the true value of J.

1) When K is smaller than the real cluster number J of the sample, the degree of aggregation of each cluster will be greatly increased with each increase of one unit of K, and the distance sum will decrease greatly;

2) When K approaches J, the return of polymerization degree obtained by increasing K will rapidly decrease, and the decline of distance sum will also decrease;

3) As K continues to increase, the change of distance sum will be gradual.

3.2. Experimental Setup

In the previous notebook, an exploratory data analysis (EDA) was conducted. Now, we aim to build a classifier for fraud detection using supervised learning techniques, considering class imbalance and the size of the data. For this purpose, the following packages are utilized: `caret`, `caretEnsemble`, `smotefamily`, `ROSE`, and `imbalance`.

Additionally, we seek to identify the factors that contribute to a transaction being likely to be fraudulent through unsupervised learning methods. These factors may include location, time, or a combination of both. To achieve this, we employ clustering methods such as k-means and DBSCAN (Density-Based Spatial Clustering).

By leveraging machine learning-based k-means clustering, we aim to enhance financial fraud detection strategies and contribute to the advancement of fraud prevention measures.

3.3. Experimental data set feature engineering

Data Set:

The experimental data set comprises various features related to financial transactions, including but not limited to:

Table 1. Experimental raw data

	eigenvalue	variance.percent	cumulative.variance.percent	
Dim.1	1.3547009	27.09402	27.09402	
Dim.2	1.0067949	20.13590	47.22992	
Dim.3	0.9926561	19.85312	67.08304	
Dim.4	0.9663174	19.32635	86.40939	
Dim.5	0.6795306	13.59061	100.00000	

The experimental dataset contains five Dimensions, each with three attributes: Eigenvalue, Variance Percent, and Cumulative Variance Percent. These properties are the result of principal component analysis (PCA), which is used to describe the variation and proportion of variance explained in the data set.

Therefore, an overview of the data set can be summarized as follows:

The dataset contains five principal components (Dimensions), named Dim.1 through Dim.5.

For each principal component, the following three properties are provided:

Eigenvalue: The eigenvalue represents the variance explained by the principal component.

Variance Percent: The percentage of the total variance explained by the principal component.

Cumulative Variance Percent: Represents the cumulative percentage of the total variance explained by the principal component and the previous principal component.

In summary, the feature engineering overview of this dataset mainly involves the results of principal component analysis (PCA) to understand the variation and variance distribution in the dataset.

Feature Engineering:

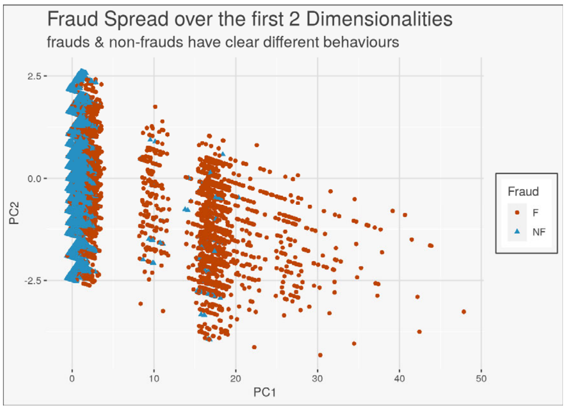


Figure 2. Schematic diagram of fraud detection feature engineering PCA results

1. Feature selection and dimensionality reduction: Using principal component analysis (PCA) results, the first two principal components (PC1 and PC2) that explain the maximum variance are retained and treated as new features. Doing so helps reduce the dimensionality of the data set and preserves the most important information, simplifying the modeling process.

2. Attach target column: Attach the target column (fraud) from the original data set to the new data set for use in classification tasks. In this way, the feature transformed by PCA can be associated with the target variable, and the relationship between the feature and the target can be better understood.

3. Visualization: ggplot2 library is used for data visualization to show the distribution of fraudulent and non-fraudulent transactions in the new feature space. Fraudulent and non-fraudulent transactions are distinguished by different colors and shapes so that they have a clear distinction in the new feature space, which helps in subsequent modeling and analysis.

3.4. Class Imbalance

As stated before, the fraud and non-fraud cases available in the present dataset have extremely imbalanced weights, with only 1.2% out of the total cases being fraud. [10-12]Therefore, there are only 7,000 observations labeled as fraud, while the rest 580,000 observations are labeled as clean transactions. In a classification problem, this would create difficulties for a model to correctly identify the fraud label, because it is so scarce throughout the dataset.

This data structure issue can be solved by using two different sampling techniques: undersampling or oversampling.

3.5. Undersampling - Splitting the data 65% - 35%

Undersampling method consists of keeping all available fraud transactions, while undersampling the non-fraud transactions to around the same number.the split is made so that proportions within the data for age, gender, category, amount_thresh and merchant remain the same

final table dimension is fraud data: 7,160 and non fraud data: 9,647

we split 65%-35% to be sure that the model classifies as correct as possible non frauds as well (very important for the relationship with the customer)

3.6. Modeling and Evaluation

```
# ------------------- Inspect results
resamples <- resamples(model_list)
dotplot(resamples, metric = "Sens")
```

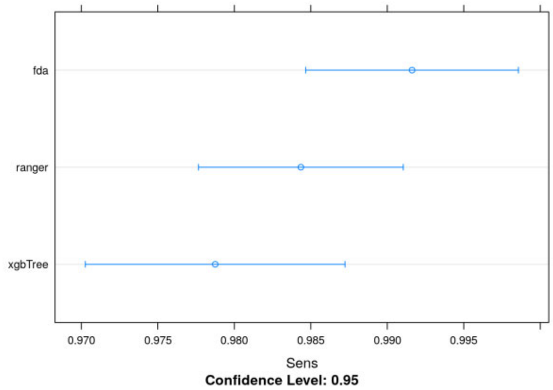


Figure 3. Cross-validation results model

```
dotplot(resamples, metric = "Spec")
```

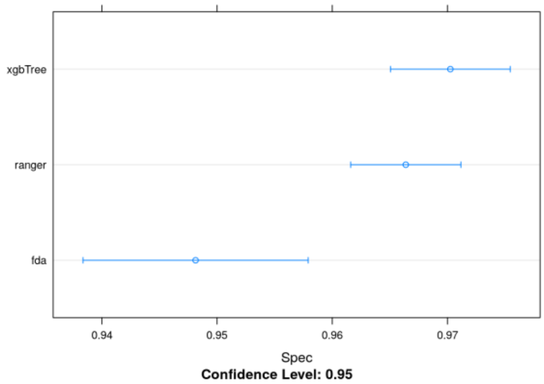


Figure 4. Cross-validation results model 2

By evaluating the two models (xgbTree and fda) on test data, we found that:

1. On the test data set, the sensitivity of xgbTree model was 98.28%, while the sensitivity of fda model was 99.72%. This suggests that the fda model is better at identifying fraudulent transactions.

2. Through visualization of Confusion Matrix, we can further analyze the model's performance on test data. The confusion matrix can help us understand the model's prediction accuracy across different categories (fraudulent and non-fraudulent transactions) and identify cases of misclassification.

In summary, the evaluation results based on the test data set show that the fda model is better at identifying fraudulent transactions and its sensitivity is higher. Therefore, in practical applications, we can consider preferentially selecting the fda model for fraud detection tasks.

4. Experimental Results and Evaluation

4.1. KMEANS CLUSTERING

In this part, the experiment of KMeans Clustering is comprehensively explored and analyzed. First, KMeans clustering was performed on the entire dataset, using the first three Principal Components (PCs), which together explained 77% of the variability. We used three cluster centers and assigned each sample to the nearest cluster center. We then compared the clustering results with the fraud labels and calculated the mean of each cluster to understand the feature differences between the different clusters.

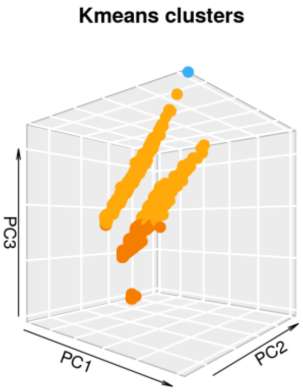


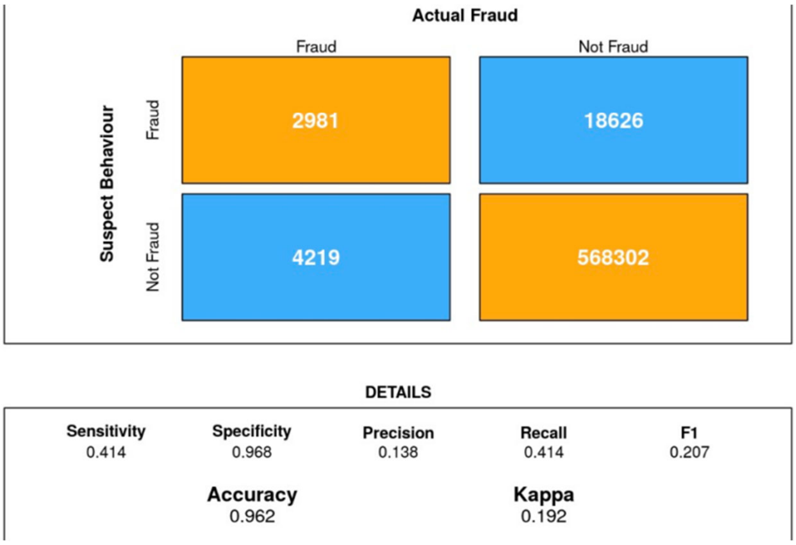
Figure 5. K-Means data detection results

From the clustering results, we can see the distribution of the three cluster centers, and find that there are almost no fraud cases in cluster 1 and cluster 2, while cluster 3 contains the majority of fraud cases. Further analysis found that cluster 3 was significantly different from the other two clusters in terms of amount, gender, age, category, etc., suggesting that these characteristics may be related to fraud. At the same time, by visualizing the results of principal component analysis (PCA), we can show the distribution of different clusters in three-dimensional space. This helps to intuitively understand the clustering of samples in the data set and observe the degree of separation between different clusters.

In summary, the results of this experiment show that K-means clustering has a potential application prospect in financial fraud detection. By performing cluster analysis on the data set, we are able to find feature differences between different clusters and identify potential fraud cases. This provides financial institutions with an effective tool that can help them better identify and prevent fraud, thereby improving the safety and reliability of the financial system.

4.2. Outlier evaluation

After obtaining K-means credit fraud detection results, the original results need to be evaluated numerically. Firstly, the K-means clustering algorithm is used to perform cluster analysis on the entire data set. By calculating the Euclidean distance of each sample point to the center of the cluster to which it belongs, we determine the outlier. We then label these outliers as abnormal behavior and combine them with the original data set to get the final data set. Finally, we evaluate the performance of K-means clustering by plotting a confusion matrix to show the relationship between the predicted results of fraud and the actual labels.



Suspect Behaviour \ Actual Fraud	Fraud	Not Fraud
Fraud	2981	18626
Not Fraud	4219	568302

DETAILS

Sensitivity	Specificity	Precision	Recall	F1
0.414	0.968	0.138	0.414	0.207

Accuracy	Kappa
0.962	0.192

Figure 6. K-means cluster data structure evaluation model

Through the identification of outliers, we can gain a deeper understanding of the potential abnormal behavior in the data set. The advantage of this approach is that it can help us find anomalies that might be missed in traditional methods, thus improving the ability to identify potential fraud. [13][14]In this experiment, although the number of outliers is relatively small, they provide important information about fraud and help us to more accurately understand the patterns of fraud in the data set.

4.3. K-Means Results of fraud detection experiments

K-Means clustering algorithm performs best in financial fraud detection. Through the experiment, we can find:

1. The nothing suspect behavior group is mainly composed of non-fraudulent transactions, with only 5 fraud cases, which belong to relatively minor fraud behaviors and are different in nature from the average fraudulent transactions.

2. These transactions are for lower amounts (around $50) and usually involve food and transportation categories.

3. Transactions often occur at reliable and secure merchants (only the top three merchants who have never experienced any fraud).

4.somewhat suspect behavior groups are more uniform, with properties that fall on the boundary between suspicious and non-suspect behavior.

5. The amount of these deals is slightly higher (around $150), and most of them are in the health, fashion and hospitality categories.

6. There are some fraud cases registered among merchants, but not the majority.

The highly suspect behavior group was the smallest of all and contained only cases of fraud, which did show unusual behavior except for 88 observations that were flagged as clean transactions.

7. The value of these transactions is very high ($1,270 on average) and covers the leisure, travel, sports and toy categories.

8. More than half of the total number of cases sent to merchants were registered as frauds.

Cluster analysis plays an important role in fraud detection. By splitting transaction data into different clusters, the patterns and characteristics of fraud can be better understood and measures can be taken to identify and prevent fraud.

5. Conclusion

Fraud detection in the financial field is one of the important challenges facing the current digital financial system. To address this challenge, this paper proposes a K-means clustering method based on machine learning to improve the accuracy and efficiency of financial fraud detection. [15]By performing cluster analysis on large amounts of financial transaction data, we are able to spot unusual patterns and behaviors in a timely manner, thereby identifying potential fraud. Compared to traditional rule-based detection methods, machine learning methods are more adaptable to changing fraud techniques and patterns, while increasing the flexibility and accuracy of detection.

The experimental results show that K-means clustering method has a good application prospect in financial fraud detection. By performing cluster analysis on the data set, we can find the feature differences between different clusters, and then identify potential fraud cases. Especially in high-risk areas, K-means clustering can help financial institutions monitor and prevent fraud more targeted, thus effectively reducing the impact of fraud on the financial system[16][17].

In summary, the K-means clustering method based on machine learning provides an effective solution for financial fraud detection. The application of machine learning technology has made the financial security sector more dynamic and adaptable, able to respond to changing fraud practices and patterns. Through the analysis and processing of large-scale data, machine learning algorithms can discover the underlying patterns and laws behind the massive data, thus providing financial institutions with more accurate fraud identification and prevention capabilities.[18-20] In addition, machine learning enables automated processing and decision making, greatly improving the efficiency and responsiveness of financial security systems.

With the continuous development and progress of technology, the application of machine learning in the field of financial security will be more extensive and in-depth. Future financial security systems will be more intelligent and adaptive, able to monitor and identify emerging fraud patterns in real time, and take appropriate prevention and response measures. At the same time, with the increasing amount of data and the continuous optimization of algorithms, machine learning technology will be able to identify fraud more accurately and provide personalized security solutions to provide strong support for the robust development of the financial system.

Acknowledgment

In the process of writing the article, we benefited from the work of Zheng, Jiajian et al., the authors of [18]. They are in The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart The clustering algorithm, financial fraud security and other aspects are deeply studied in Finance. We sincerely thank them for their academic contributions and inspiration to our research. In addition, we would like to thank them for their valuable insights on the application of machine learning to the financial sector, which has provided important references and guidance for our research. Finally, we would also like to thank all the people who have given support and help in our research process, and their contributions have made this research smoothly carried out.

References

[1] Zhang, Yufeng, et al. "Manipulator Control System Based on Machine Vision." International Conference on Applications and Techniques in Cyber Intelligence ATCI 2019: Applications and Techniques in Cyber Intelligence 7. Springer International Publishing, 2020.

[2] Gao, Longsen, et al. "Autonomous multi-robot servicing for spacecraft operation extension." 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023.

[3] Che, Chang, et al. "Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation." Proceedings of the 2023 6th International Conference on Big Data Technologies. 2023.

[4] Che, Chang, et al. "Advancing Cancer Document Classification with R andom Forest." Academic Journal of Science and Technology 8.1 (2023): 278-280.

[5] Duan, Shiheng, et al. “Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model Based on Machine Learning”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 132-6, https://doi.org/10.54097/ObMPjw5n.

[6] K. Tan and W. Li, "Imaging and Parameter Estimating for Fast Moving Targets in Airborne SAR," in IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 126-140, March 2017, doi: 10.1109/TCI.2016.2634421.

[7] Chen , Jianfeng, et al. “Implementation of an AI-Based MRD Evaluation and Prediction Model for Multiple Myeloma”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 127-31, https://doi.org/10.54097/zJ4MnbWW.

[8] Che, Chang, et al. "Deep learning for precise robot position prediction in logistics." Journal of Theory and Practice of Engineering Science 3.10 (2023): 36-41.

[9] “Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 139-46, https://doi.org/10.62051/ijcsit.v1n1.19.

[10] Chen, Jianhang, et al. "One-stage object referring with gaze estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[11] W. Sun, W. Wan, L. Pan, J. Xu, and Q. Zeng, “The Integration of Large-Scale Language Models Into Intelligent Adjudication: Justification Rules and Implementation Pathways”, Journal of Industrial Engineering &amp; Applied Science, vol. 2, no. 1, pp. 13–20, Feb. 2024.

[12] Zhou, Yanlin, et al. "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes." Journal of Theory and Practice of Engineering Science 4.02 (2024): 45-51.

[13] Liang, Penghao, et al. "Enhancing Security in DevOps by Integrating Artificial Intelligence and Machine Learning." Journal of Theory and Practice of Engineering Science 4.02 (2024): 31-37.

[14] Zhang, Chenwei, et al. "SegNet Network Architecture for Deep Learning Image Segmentation and Its Integrated Applications and Prospects." Academic Journal of Science and Technology 9.2 (2024): 224-229.

[15] Wang, Yong, et al. "Autonomous Driving System Driven by Artificial Intelligence Perception Fusion." Academic Journal of Science and Technology 9.2 (2024): 193-198.

[16] Zhang, Quan, et al. "Application of the AlphaFold2 Protein Prediction Algorithm Based on Artificial Intelligence." Journal of Theory and Practice of Engineering Science 4.02 (2024): 58-65.

[17] Du, Shuqian, et al. "IMPROVING SCIENCE QUESTION RANKING WITH MODEL AND RETRIEVAL-AUGMENTED GENERATION." The 6th International scientific and practical conference "Old and new technologies of learning development in modern conditions"(February 13-16, 2024) Berlin, Germany. International Science Group. 2024. 345 p.. 2024.

[18] Chen, J. (2022). The Reform of School Education and Teaching Under the "Double Reduction" Policy. Scientific and Social Research, 4(2), 42-45. (Feb 2022)

[19] Su, Jing, et al. "Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review." arXiv preprint arXiv:2402.10350 (2024).

[20] Wang, Yong, et al. "Construction and application of artificial intelligence crowdsourcing map based on multi-track GPS data." arXiv preprint arXiv:2402.15796 (2024).

[21] Zhou, Y., Osman, A., Willms, M., Kunz, A., Philipp, S., Blatt, J., & Eul, S. (2023). Semantic Wireframe Detection.

[22] K. Tan and W. Li, "A novel moving parameter estimation approach offast moving targets based on phase extraction," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 2015, pp. 2075-2079, doi: 10.1109/ICIP.2015.7351166.

[23] Zheng, Jiajian, et al. "The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance." arXiv preprint arXiv:2402.17194 (2024).

[24] Yang, Le, et al. "AI-Driven Anonymization: Protecting Personal Data Privacy While Leveraging Machine Learning." arXiv preprint arXiv:2402.17191 (2024).

[25] Cheng, Qishuo, et al. "Optimizing Portfolio Management and Risk Assessment in Digital Assets Using Deep Learning for Predictive Analysis." arXiv preprint arXiv:2402.15994 (2024).