[1]Dr. Pankaj Malik
[2]Ankita Chourasia
[3]Rakesh Pandit
[4]Dr. Sheetal Bawane
[5]Jayesh Surana*

Credit Risk Assessment and Fraud Detection in Financial Transactions Using Machine Learning

Journal of Electrical Systems

Abstract: - Credit risk assessment and fraud detection are crucial tasks in the financial industry, vital to preserving financial organizations' legitimacy and sustainability. Traditional methods often fall short in accurately assessing risk and detecting fraudulent activities in a timely manner. In recent years, machine learning has emerged as a powerful tool for enhancing these processes, leveraging great dimensions of transactional statistics and superior algos for making more informed decisions. This research paper explores the usage of ML techniques in credit risk assessment and fraud detection within financial transactions.

The paper begins with an overview of the importance of accurate risk assessment and fraud detection in financial transactions and introduces the role of machine learning in addressing these challenges. A comprehensive literature review is conducted to analyze existing methodologies, algorithms, and research trends in the field. Data acquisition and preprocessing techniques are discussed, emphasizing the importance of clean and relevant data for model training. Feature engineering strategies are explored to extract meaningful information from financial transaction data and enhance the predictive capabilities of machine learning models.

Various machine learning algorithms suitable for credit risk assessment and fraud detection are examined, including LR, SVMs, RF, DTs and DNNs. The efficacy of these techniques is evaluated by discussing model metrics for assessment and ensemble approaches for boosting efficiency, with a focus on metrics such as accuracy, precision, recall, and ROC-AUC.

The paper presents case studies and experimental results illustrating the application of machine learning models in real-world scenarios, highlighting their effectiveness in improving risk assessment and fraud detection processes. Additionally, difficulties such as imbalanced datasets, comprehensibility of the model and adherence to regulations are discussed, along with potential research directions and future trends in the field.

In conclusion, this research emphasizes the transformative potential of machine learning in credit risk assessment and fraud detection within financial transactions. By leveraging advanced algorithms and data-driven approaches, financial institutions can enhance their decision-making processes, mitigate risks, and safeguard against fraudulent activities, ultimately contributing to a more secure and resilient financial ecosystem.

Keywords: secure, financial ecosystem, ROC-AUC., fraudulent

1. INTRODUCTION

Credit risk assessment and fraud detection are integral components of the financial industry, essential for maintaining stability, trust, and profitability. In an era characterized by increasing transaction volumes and evolving forms of financial crime, traditional methods of risk assessment and fraud detection have shown limitations in effectively identifying and mitigating risks in a timely manner. However, with the advent of ML methods, a transformative opportunity to revolutionize these processes by utilizing massive quantities of transactional statistics and superior algorithms to make more accurate and timely decisions is given.

[1] Asst. Prof. Medi-Caps University, Indore
[2]Asst. Prof. Medi-Caps University, Indore
[3]Asst. Prof. Medi-Caps University, Indore
[4]Asst. Prof. Medi-Caps University, Indore
[5]Asst. Prof. Medi-Caps University, Indore
pankajmalik123@rediffmail.com
ankita.chourasia29@gmail.com
rakesh.pandit@medicaps.ac.in
sheetal.bawane@medicaps.ac.in
er.jayeshsurana@gmail.com

This research paper aims to explore the application of machine learning in credit risk assessment and fraud detection within financial transactions. By delving into the intricacies of machine learning methodologies, data preprocessing techniques, feature engineering strategies, and model evaluation metrics, this paper seeks to elucidate the potential of machine learning to enhance the accuracy, efficiency, and effectiveness of credit risk assessment and fraud detection systems.

The significance of this research lies in its potential to address longstanding challenges faced by financial institutions, including the need for more precise risk assessment models, the competence of detecting emerging fraud reiterations, as well as the imperative to acclimate to dynamic market conditions. Through an exhaustive review of existing literature, analysis of case studies, and presentation of experimental results, this paper aims to provide insights into the practical applications of machine learning in tackling these challenges.

Furthermore, this paper will discuss the implications of machine learning techniques in the context of regulatory compliance, ethical considerations, and customer privacy. By highlighting both the opportunities and challenges associated with the adoption of machine learning in the financial sector, this research aims to contribute to a nuanced understanding of the role of technology in shaping the future of risk management and fraud detection.

In summary, this introduction sets the stage for a comprehensive exploration of machine learning in credit risk assessment and fraud detection within financial transactions. By examining existing methodologies, analyzing experimental results, and discussing potential implications, this paper aims to offer valued acumens for scholars, practitioners, as well as policymakers within financial industry.

2. LITERATURE REVIEW

Credit risk assessment and fraud detection have been longstanding challenges in the financial industry, with significant implications for financial stability, regulatory compliance, and customer trust. Over the years, researchers and practitioners have explored various methodologies and approaches to address these challenges, ranging from traditional statistical methods to more recent advancements in machine learning and data analytics.

Traditional Approaches:

Historically, credit risk assessment and fraud detection relied heavily on rule-based systems and statistical frameworks. Financial organizations have been using scoring systems for credit, including the Fair Isaac Commission score, for a long time to evaluate consumers' reliability according to things like payment behavior, credit past events, and current debts. Similarly, rule-based systems were employed for fraud detection, where predefined rules were applied to identify suspicious transactions based on predefined thresholds or patterns.

Machine Learning Methods:

The use of ML has become a potent tool throughout the past couple of years for improving the accuracy and efficiency of credit risk assessment and fraud detection. Supervised learning algorithms, such as LR, DTs, and support vector machines, have been applied to classify borrowers into different risk categories and detect fraudulent transactions based on historical data. Ensemble approaches, RF and XGBoost, have additionally enhanced the analytical performance by combining multiple base learners.

Feature Engineering and Data Preprocessing:

Feature engineering plays a crucial role in credit risk assessment and fraud detection, where the selection and the way relevant characteristics are transformed can have a big influence on how well ML models work.. Techniques such as feature scaling, dimensionality reduction, and feature selection have been employed to extract meaningful information from raw data and improve model interpretability. Additionally, handling imbalanced datasets, where the number of positive and negative instances is skewed, remains a challenge in these tasks.

Model Evaluation and Performance Metrics:

To evaluate the effectiveness of identifying fraud and assessment of credit risks algorithms, a number of assessments have been put forth. The receiver's operator characteristics (ROC) curve, area under curve (AUC), recall, the precision, accuracy, recall, and F1-score are some of these measurements. While recall as well as precision offer information on a model's capacity to accurately detect positive occurrences (bogus trades) and steer clear of false positives, accuracy assesses the model's general reliability of prediction. The AUC measures the model's total discriminatory power, while

the ROC curve illustrates the compromise among the rate of true positives and the rate of false positives at various threshold values.

Challenges and Future Directions:

Despite the advancements in machine learning techniques, several challenges remain in credit risk assessment and fraud detection. These include the need for robust algorithms which are flexible enough to change with marketplace trends, the interpretation of complex machine learning models, Features engineering with domain expertise integrated, and the ethical considerations surrounding the use of automated decision-making systems. Additionally, regulatory compliance, data privacy, and security concerns pose further challenges in deploying machine learning solutions in the financial industry.

3. DATA ACQUISITION AND PREPROCESSING

In credit risk assessment and fraud detection, accuracy and applicability of facts are paramount to usefulness of machine learning algorithms. The section discusses the process of acquiring and preprocessing data for these tasks.

Data Sources:

Financial transaction data can be obtained from various sources, including banking records, credit bureaus, payment processors, and online transactions. These datasets typically contain information such as transaction amount, timestamp, merchant ID, customer ID, and transaction type. Additionally, credit risk assessment may involve demographic data, credit history, income level, and employment status of borrowers.

Data Preprocessing Techniques:

1. Data Cleaning:

- Removal of duplicate records and inconsistent data entries.
- Handling missing values through imputation or deletion.
- Outlier detection and removal to ensure data integrity.

2. Feature Engineering:

- Creation of new features based on domain knowledge and business rules.
- Transformation of categorical variables into numerical representations using techniques such as one-hot encoding or label encoding.
- Extraction of relevant information from text data, such as transaction descriptions or customer feedback.

3. Normalization and Scaling:

- Standardization or normalization of numerical features to ensure consistency in scale.
- Scaling of features to a common range to prevent dominance by features with larger magnitudes.

4. Handling Imbalanced Datasets:

- Resampling techniques such as oversampling (e.g., SMOTE) or undersampling to balance the distribution of classes.
- Adjusting class weights in machine learning algorithms to penalize misclassification of minority classes.

5. Dimensionality Reduction:

- Techniques such as Principal Component Analysis (PCA) or feature selection algorithms to reduce the dimensionality of the dataset.
- Reducing computational complexity and alleviating the curse of dimensionality.

6. Time-Series Data Handling:

- Temporal aggregation of transactional data to different time granularities (e.g., hourly, daily, monthly).
- Feature engineering based on temporal patterns and trends in transactional data.

7. Data Splitting:

- Division of the dataset into training, validation, and test sets to evaluate model performance.
- Stratified sampling to ensure balanced class distribution across partitions.

Data Privacy and Security:

Ensuring the confidentiality and safety of delicate financial statistics is of utmost significance. Techniques such as data anonymization, encryption, and access control mechanisms should be employed to protect customer information and comply with regulatory requirements (e.g., GDPR, HIPAA).

4. FEATURE ENGINEERING

Feature engineering plays a pivotal role in credit risk assessment and fraud detection, as To enhance machine learning model efficiency, pertinent information must be extracted from unprocessed data. This section outlines various feature engineering techniques commonly used in these tasks.

1. Domain-Specific Features:

Creation of features based on domain knowledge and business rules. For credit risk assessment, these may incorporate details about the borrower, like their age, financial status, work position, and duration of credit record. Factors for identifying fraud might include the quantity, rate, and duration of days of transactions.

2. Temporal Features:

Extraction of temporal patterns and trends from timestamp data. This may involve creating features such as time of day, day of the week, month, and year. Additionally, time-based aggregations (e.g., sum, mean, max) over different time windows (e.g., hourly, daily, monthly) can capture transactional behavior over time.

3. Aggregated Features:

Calculation of aggregate statistics over groups of transactions or customers. For example, aggregating transaction amounts by customer ID to compute features such as total spending, average transaction amount, and number of transactions. Transaction trends and general spending behaviour can be better understood by looking at aggregated aspects.

4. Frequency-Based Features:

Calculation of frequency-based statistics to capture transactional behavior. This may include features such as the number of transactions in a given time period, the time elapsed since the last transaction, and the average time between transactions. Frequency-based features can help identify irregular transaction patterns indicative of fraudulent activity.

5. Text-Based Features:

Extraction of information from text data, such as transaction descriptions or customer feedback. NLP techniques could be employed like tokenize text, extract keywords, and derive sentiment features. Text-based features can provide additional context and insights into transactional behavior.

6. Interaction Features:

Creation of interaction features by combining multiple input features. This may involve arithmetic operations (e.g., addition, multiplication) or logical operations (e.g., AND, OR) between features. By capturing intricate interactions among variables that are input, interactions characteristics can increase the richness of the model.

7. Dimensionality Reduction:

Techniques such as Principal Component Analysis (PCA) or feature selection algorithms to reduce the dimensionality of the feature space. Dimensionality reduction can help alleviate the curse of dimensionality and improve model scalability and interpretability.

8. Derived Features:

Derivation of new features from existing ones through mathematical transformations or logical operations. This may include feature scaling, normalization, logarithmic transformations, and polynomial features. Derived features could improve discriminatory aptitude for the algorithm through capturing non-linear relations among variable.

5. MACHINE LEARNING MODELS

In credit risk assessment and fraud detection, various machine learning models can be employed to predict creditworthiness and identify fraudulent activities. This section discusses a few of the ML methods that are frequently applied to these jobs.

1. Logistic Regression:

LR is simple yet effective binary cataloguing algorithm widely used. This algo models probability of binary outcome (e.g., default/non-default, fraudulent/non-fraudulent) according to a predictive factor or factors. LR is interpretable and computationally efficient, making it suitable for applications where model interpretability is important.

2. DTs:

DT is non-linear model that partitions the attribute vector space in areas upon on values, with each partition representing a decision node. DTs are intuitive, easy to interpret, and capable of capturing non-linear relations amid input attributes and Dependent variable. However, they are disposed to overfitting, particularly in case of intricate data sets.

3. Random Forests:

Random forest models are a method of ensemble learning where numerous possible tree are merged to improve prediction accuracy and robustness. An arbitrary part of instruction data and some nominated sets of characteristic are employed for training each of trees in the forest. Relative with individual choice trees, random forests are less susceptible to outliers along with anomalies and reduce over fitting through average estimates over several trees..

4. Gradient Boosting Machines (GBM):

Ensemble learning techniques like GBMs teach poor learner one after the other (typically DTs) for rectifying the mistakes made by prior models. GBM iteratively minimizes a loss function by adding new tree in ensembles, such that each of the trees are training on the residuals of preceding ensemble. Gradient boosting is recognized for the technique's superior predictive accuracy as well as flexibility, making the technique suitable for complex datasets.

5. Support Vector Machines (SVM):

SVM are supervising methods of training that use the optimal hyper plane separation between multiple categories in a feature field to categorize data. By using kernel features, SVMs are able to cope with multidimensional information as well as irregular boundaries for decisions. SVMs work well for jobs involving binary categorization, and they're especially helpful for handling small- to medium in size datasets..

6. Neural Networks:

- NNs, especially DL designs like MLPs and CNNs, have shown promising results in credit risk assessment and fraud detection tasks. NNs can absorb intricate relations in information and adapt to various input modalities (e.g., numerical, categorical, text). However, during development, NNs need a lot of information and processing power. and may suffer from interpretability issues.

7. Ensemble Methods:

- Multiple foundational learners are used in ensemble techniques like bagging, boost, and layering to enhance the accuracy of models and generalization. By leveraging the multiplicity among individual model, ensemble systems lessen over fitting and take a wider variety of features and relations in the data. Ensemble methods are versatile and can be applied with various base learners, making them suitable for different types of datasets and tasks.

6. MODEL EVALUATION AND PERFORMANCE METRICS

In credit risk assessment and identifying fraud, and assessing effectiveness of machine learning model is crucial to ensure their effectiveness in mitigating risks and identifying fraudulent activities. This section outlines various model evaluation techniques and performance metrics commonly used in these tasks.

1. Train-Test Split:

- Separate the data set into test & train subsets for assessing model's generalization operation. To figure out the model's efficacy on fresh data, it is developed on the initial data set and assessed on a testing set that hasn't been seen yet.

2. Cross-Validation:

- To get more accurate predictive capacity estimations, do k-factor cross-validating This data set is partitioned in k parts, and the algorithm has been training and assessing k times, using the remainder of the k-1 parts as the input train set and every Kth part as evaluation set once.

Performance Metrics:

3. Accuracy:

The percentage of properly determined cases relative to the entirety of examples is known as accuracy. Although accuracy offers a broad gauge for an algorithm's performance, imbalanced data sets with dispersed categories might not be a good fit for accuracy.

4. Accuracy Measurement:

Precision (Positive Predictive Value): This metric indicates the proportion of correctly identified positive cases (true positives) out of all instances the model predicts as positive. A high precision value shows the model's ability to minimize false positives, which are especially critical in areas like fraud detection where unnecessary investigations can be costly.

Completeness Measurement:

Recall (Sensitivity, True Positive Rate): This metric focuses on the model's ability to capture all actual positive instances in the data. It calculates the proportion of true positives out of all actual positive cases. High recall is crucial in domains like credit risk assessment and fraud detection to ensure no fraudulent activities go undetected.

F1-Score: Balancing Precision and Recall

F1-Score: This metric provides a balanced view of both precision and recall by calculating their harmonic mean. A higher F1-score indicates better overall model performance, making it particularly useful in situations where both metrics are equally important, especially when dealing with imbalanced datasets.

Visualization of Model Performance:

ROC: This curve portrays the compromise between sensitivity and specificity (1-false positive rate) at different classification thresholds. By plotting the true positive rate (sensitivity) on the y-axis against the false positive rate on the x-axis, ROC curves provide insights into the model's ability to discriminate between positive and negative instances across various thresholds.

Quantifying Discrimination Power:

ROC-AUC: This metric recaps the general discriminatory power of a model by calculating the area under the ROC curve. A higher ROC-AUC score indicates better model performance in distinguishing positive from negative cases. A score of 0.5 signifies random performance, while 1.0 reflects perfect discrimination.

Prioritizing Positive Cases:

Lift Curve: This curve measures how much better the model performs compared to random chance at different percentile ranges within the data. It helps assess the model's ability to prioritize instances with a higher likelihood of being positive, which is crucial for maximizing the detection of fraudulent activities.

Performance Summary:

Confusion Matrix: This table summarizes the model's predictions compared to the actual class labels. It consists of four key elements: TP, FP, TN, and FN. These standards are then used for computing numerous performance metrics discussed earlier.

7. CASE STUDIES AND EXPERIMENTS

Case studies and experimental analyses are essential components of research in credit risk assessment and fraud detection using machine learning techniques. They provide real-world insights into the practical application and performance of various models. Here are hypothetical case studies and experiments that illustrate the effectiveness of machine learning in these domains:

1. Case Study: Credit Risk Assessment

- Objective: Evaluate the performance of machine learning models in predicting credit risk for loan applicants.
- Dataset: Utilize a dataset containing historical loan application data, including borrower attributes (e.g., income, employment status), credit history, and loan outcomes (default/non-default).
- Experiment: Train and evaluate multiple machine learning models, including LR, random forests, and gradient boosting machines, using cross-validation techniques. Assess model performance using metrics.
- Results: Compare the performance of different models and identify the most effective approach for credit risk assessment. Analyze the impact of feature engineering techniques and data preprocessing methods on model performance.

2. Case Study: Fraud Detection

- Objective: Develop a fraud detection system to identify fraudulent transactions in real-time.

- Dataset: Utilize a dataset containing transactional data, including transaction amount, timestamp, merchant ID, and customer ID, with labeled instances of fraudulent and non-fraudulent transactions.
- Experiment: Train and deploy machine learning models, such as LR, DTs, and NNs, to classify dealings as deceitful or non-deceitful. Evaluate model performance using metrics.
- Results: Assess the effectiveness of different machine learning algorithms in detecting fraudulent activities. Analyze the trade-offs between model performance and computational efficiency for real-time fraud detection applications.

3. Case Study: Ensemble Learning for Risk Assessment

- Objective: Investigate the use of ensemble learning techniques for improving the accuracy and robustness of credit risk assessment models.
- Dataset: Utilize a large-scale dataset containing diverse features related to borrower attributes, credit history, and economic indicators.
- Experiment: Train ensemble models, such as random forests and gradient boosting machines, using various feature sets and hyperparameter configurations. Evaluate model performance using cross-validation and assess the impact of ensemble methods on predictive accuracy.
- Results: Compare the performance of ensemble models with individual classifiers and baseline models. Analyze the contribution of different base learners to the ensemble's predictive performance and identify factors influencing model robustness.

4. Case Study: Explainable AI for Fraud Detection

- Objective: Develop an explainable AI system for fraud detection to provide interpretable insights into model predictions.
- Dataset: Utilize a dataset containing transactional data and customer attributes, with labeled instances of fraudulent and non-fraudulent transactions.
- Experiment: Train machine learning models, such as LR and DTs, using interpretable feature representations. Use methods like SHAP values and LIME for explaining models prediction and identifying important features contributing to fraud detection.
- Results: Provide interpretable explanations for model predictions, highlighting key factors influencing the likelihood of fraudulent transactions. Assess the trade-offs between model interpretability and predictive performance for fraud detection applications.

8. CHALLENGES AND FUTURE DIRECTIONS

Despite the advancements in machine learning for credit risk assessment and fraud detection, several challenges persist, and there are numerous avenues for future research and development. Here are some key challenges and potential directions for future work in these domains:

1. Imbalanced Datasets:

- Challenge: Imbalanced datasets, where the number of positive instances (e.g., fraudulent transactions) is much smaller than negative instances, pose a significant challenge for machine learning models.
- Future Direction: Explore advanced techniques for handling imbalanced datasets, such as oversampling, undersampling, cost-sensitive learning, and synthetic data generation.

2. Model Interpretability:

- Challenge: Complex machine learning models, such as NNs and ensemble methods, often lack interpretability, making it difficult to understand the factors driving model predictions.
- Future Direction: Develop explainable AI techniques that provide interpretable explanations for model predictions, enabling stakeholders to trust and understand model decisions.

3. Dynamic Market Conditions:

- Challenge: Financial markets are dynamic and constantly evolving, requiring adaptive models that can quickly adapt to changing conditions and emerging risks.
- Future Direction: Investigate adaptive machine learning approaches, such as online learning and reinforcement learning, that can continuously update models based on incoming data and feedback.

4. Privacy and Regulatory Compliance:

- Challenge: Financial data is highly sensitive, and there are stringent regulations (e.g., GDPR, CCPA) governing the collection, storage, and use of personal information.
- Future Direction: Develop privacy-preserving machine learning techniques, such as federated learning, differential privacy, and homomorphic encryption, to ensure compliance with regulations while preserving data privacy.

5. Real-Time Processing:

- Challenge: Real-time processing of financial transactions requires low-latency and high-throughput systems capable of quickly detecting fraudulent activities without introducing significant delays.
- Future Direction: Investigate scalable and efficient machine learning algorithms and architectures optimized for real-time processing, leveraging techniques such as stream processing and distributed computing.

6. Adversarial Attacks:

- Challenge: Adversarial attacks aim to deceive machine learning models by introducing subtle perturbations to input data, leading to misclassifications and vulnerabilities in fraud detection systems.
- Future Direction: Research robust machine learning techniques that are resilient to adversarial attacks, such as adversarial training, feature obfuscation, and model diversification.

7. Cross-Domain Generalization:

- Challenge: Models trained on data from one financial institution or market may not generalize well to other institutions or markets due to differences in data distribution and business practices.
- Future Direction: Investigate transfer learning and domain adaptation techniques that can leverage knowledge from related domains or datasets to improve model generalization across different contexts.

8. Ethical Considerations:

- Challenge: Machine learning models used in credit risk assessment and fraud detection may inadvertently perpetuate biases and discrimination, leading to unfair outcomes for certain demographic groups.
- Future Direction: Develop fair and ethical machine learning frameworks that address biases, promote transparency, and ensure accountability in decision-making processes.

9. CONCLUSION

In conclusion, the application of machine learning in credit risk assessment and fraud detection represents a significant advancement in the financial industry, offering opportunities to enhance risk management practices and protect against fraudulent activities. This research paper has explored various aspects of machine learning in these domains, including data acquisition and preprocessing, feature engineering, model selection, evaluation metrics, case studies, challenges, and future directions.

Through the utilization of diverse datasets and sophisticated algorithms, machine learning models have demonstrated their effectiveness in predicting creditworthiness, identifying fraudulent transactions, and mitigating risks in financial transactions. From LR to deep learning architectures, a wide range of machine learning techniques have been employed to address complex challenges and extract actionable insights from raw data.

However, several challenges remain, including imbalanced datasets, model interpretability, dynamic market conditions, privacy concerns, real-time processing requirements, adversarial attacks, cross-domain generalization, and ethical considerations. Addressing these challenges and exploring future research directions will be essential for advancing the field and developing more robust and reliable solutions.

Overall, machine learning offers tremendous potential to revolutionize credit risk assessment and fraud detection in the financial industry, enabling more informed decision-making, improving operational efficiency, and safeguarding against financial losses. By fostering collaboration between researchers, practitioners, and policymakers, we can leverage the power of machine learning to build a more resilient and equitable financial ecosystem for the benefit of society as a whole.

REFERENCES

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589-609.
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
5. Lipton, Z. C. (2016). The mythos of model interpretability. Queue, 14(5), 30-57.
6. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).
7. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.
8. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.
9. Van Vlasselaer, V., Bravo, C., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B., & Daelemans, W. (2015). Detection of vote manipulation in online rating systems using supervised learning. Decision Support Systems, 75, 66-77.
10. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.