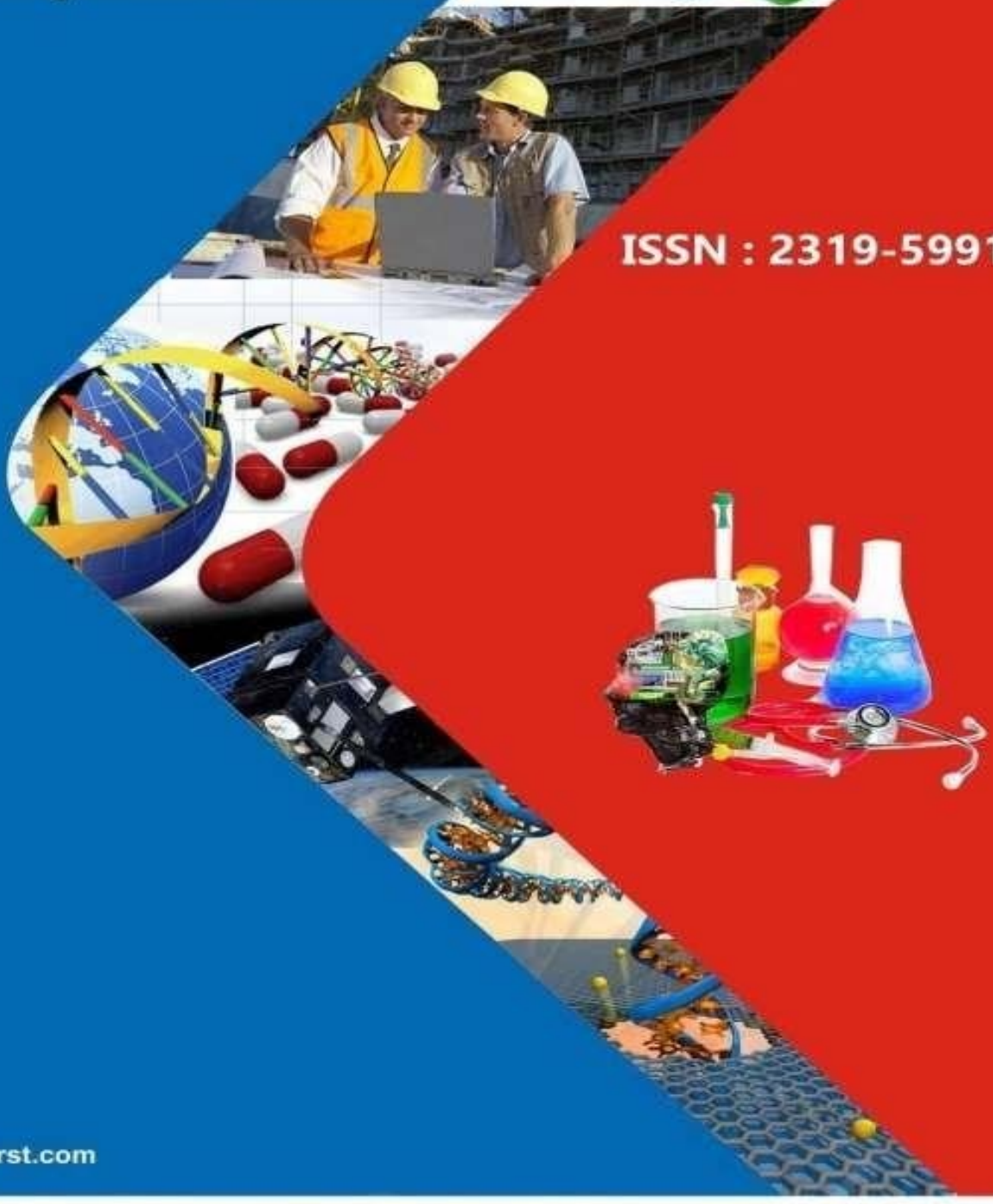
International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922

Vol. 21, Issue 2, 2025

SMART RISK, SMARTER DETECTION: A VALUE-AT-RISK APPROACH TO FINANCIAL FRAUD IN IMBALANCED DATASETS

[1]K. Chandrasekhar, MCA Student, Department of MCA

[2] H. Ateeq Ahmed, M.Tech, (Ph.D), Assistant Professor, Department of MCA

[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

As more people utilise online banking services, the large losses that banks and other financial institutions have incurred as a result of new bank account (NBA) fraud are concerning. Machine learning (ML) models have been severely challenged by the intrinsic skewness and rarity of NBA fraud occurrences. This occurs when the number of non-fraud instances is greater than the number of fraud instances, causing the ML models to ignore and mistakenly classify fraud as non-fraud instances. Customers' confidence and trust may be damaged by such mistakes. While addressing the skewness of fraud datasets, previous research has focused on fraud patterns rather than possible losses of NBA fraud risk characteristics. As a risk measure that views fraud cases as the worst-case scenario, the identification of NBA fraud is suggested in this study within the framework of value-at-risk. Value-at-risk models risk characteristics as a skewed tail distribution and estimates possible losses of risk features using historical simulation. ML was used to classify the risk-return characteristics derived from value-at-risk on the bank account fraud (BAF) dataset. In order to provide weight to the skewed NBA fraud cases, the value-at-risk manages the fraud skewness using an adjustable threshold probability range. The effectiveness of the fraud detection algorithm was assessed using a unique detection rate (DT) metric that takes risk fraud characteristics into account. A K-nearest neighbour with a true positive (TP) rate of 0.95 and a DT rate of 0.9406 is used to create an enhanced fraud detection model. Value-at-risk offers a clever way to create data-driven standards for fraud risk management within a banking industry's acceptable loss tolerance.

I. INTRODUCTION

According to a financial fraud report published by the Association of Certified Fraud Examiners (ACFE) 2022, 2,110 fraud incidents involving financial sectors in 133 countries caused losses of around $3.6 billion [1]. The intentional use of illegal methods or strategies to generate financial benefit is known as financial fraud [2]. Economic disruption, increased living costs, and weakened customer confidence are all possible outcomes of financial fraud [3]. Financial fraud may take many different forms, such as mortgage fraud, credit and debit card fraud, money laundering, insurance fraud, and new bank account fraud [4.5]. "New bank account (NBA) fraud" refers to the practice of creating an account with the intention of committing fraud at banks or other financial institutions. [6]. Fraud has wider repercussions, impacting consumers and financial systems via market volatility and fuelling more significant macroeconomic downturns, in addition to causing immediate financial losses and undermining public trust in institutions [7]. Typical characteristics of fraudulent datasets include skewness, changing trends, high dimensionality, and limited access to pertinent data. Studies have been particularly concerned by fraud skewness, which reflects the majority fraud class over the non-fraud class, as it has an impact on the effectiveness of fraud detection models. Machine learning systems, including distance-based algorithms, may be negatively impacted by skew fraud cases [8]. Rule-based expert systems, statistical techniques, machine learning, and risk-

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

based approaches have all been developed in the past to combat fraud [9], [10]. In order to address financial fraud, decision-makers choose to use statistical techniques such autoregressive models [11], [12], and [13] since rule-based solutions are ineffective and costly to maintain [10]. Because of the high dimensionality and intricate patterns of frauds, statistical approaches are less successful, which is why machine learning models were used [10], [14]. Nevertheless, a substantial False Positive (FP) rate was discovered in a few of the investigations that used machine learning approaches [15], [16], and [17]. High-dimensional data and intricate fraud case patterns may be handled by machine learning algorithms.

Jesus et al. [18] provided the first domain-specific and real-world bank account fraud (BAF) dataset in order to assess the efficacy of the machine learning model. Generative adversarial networks (GANs) were used to create the datasets, and the light gradient boosting technique (LGBM) was used to assess them.In order to optimise the LGBM model, the research [18], [19] uses 25 sets of hyperparameter configurations. Utility aware reweighing was performed to address the class skewness of the BAF dataset. In order to assess the BAF dataset and deal with the evolving fraud tendencies, the research [15] makes use of stacking in ensembled learning with majority voting. The research [20] classifies fraud cases using deep neural networks and federated learning to solve data privacy concerns of the BAF dataset. These studies do an excellent job of resolving BAF difficulties, however they don't take into account the possible losses of fraud risk characteristics. As far as we are aware, not much study has been done on the use of machine learning approaches for NBA fraud detection. This research suggests NBA fraud detection in the framework of risk management that treats skewed fraud cases as a worst-case scenario using value-at-risk. Value-at-risk was supplemented with anticipated loss and expected shortfall of frauds, which further quantifies the mean and severe loss impacts, respectively, in order to accurately evaluate the losses of fraud risks. The combination of these risk metrics will make it possible to quantify hazards in mean, worst-case, and extreme situations. Value-at-risk estimates possible losses of risk characteristics using historical simulation. Their risk exposure to fraud risk is the basis for the risk-return characteristics derived from value at risk. The NBA fraud detection model receives the risk-return characteristics as input. After training a variety of machine learning models, the K-nearest neighbour model outperformed the others. This study makes the following contributions: • Rather than modelling the fraud pattern, it employed an extreme value theorem to represent the tails (possible losses).

• To more effectively represent the skewness of fraud cases, this article used value-at-risk.

• Since this study made no assumptions about any distribution, it used historical simulation to determine value-at-risk.

• To capture the entire performance in detecting NBA fraud cases that integrate risk fraud characteristics, this article used innovative detection rate performance indicators.

This is how the rest of the paper is organised: Section II presents the study's review of the literature. Section III presents the problem definition. Section IV presents the materials and processes. Section V presents the experimental setup. Section VI presents the findings. Section VII presents the study's findings and comments.

II. LITERATURE SURVEY

"An effective fraud detection mechanism based on blockchain and machine learning,"

I. A. Hameed, T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, and S. Alsafari,

We discuss the issues of fraud and irregularities in the Bitcoin network in this article. These are typical issues with online transactions and e-banking. However, fraud and anomaly techniques also change as the financial industry does.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

Furthermore, blockchain technology is emerging as the safest approach to financial integration. But every year, a number of scams also rise in tandem with these sophisticated technology. As a result, we provide a safe fraud detection methodology that combines blockchain technology and machine learning. For transaction classification, two machine learning methods are utilised: random forest (RF) and XGboost. Machine learning approaches forecast future incoming transactions and train the dataset based on integrated and fraudulent transaction patterns. To identify fraudulent transactions in the Bitcoin network, machine learning algorithms are used with blockchain technology. The XGboost and random forest (RF) algorithms are used in the suggested model to categorise transactions and forecast transaction trends. To gauge the accuracy, we also compute the models' precision and AUC. To demonstrate the resilience of our system, a security study of the suggested smart contract is also carried out. In order to defend the suggested system from vulnerabilities and assaults, an attacker model is also suggested.

"Enhanced machine learning model for detecting credit card fraud,"

According to N. S. Alfaiz and S. M. Fati, the COVID-19 epidemic has somewhat reduced people's mobility, making it more difficult to buy products and services offline. As a result, there is now a greater reliance on internet services in society. Fraud is a significant problem in the world of online purchases and is one of the main problems with utilising credit cards. Therefore, in order to stop almost all fraudulent credit card transactions, the finest machine learning strategy must be developed. A total of 66 machine learning models based on two assessment phases are examined in this article. Each model uses stratified K-fold cross-validation and a real-world credit card fraud detection dataset of European cardholders. Nine machine learning methods are evaluated in the first phase to identify fraudulent transactions. 19 resampling strategies are used to each of the top three algorithms, which are nominated to be utilised again in the second round. The All K-Nearest Neighbours (AllKNN) undersampling approach combined with CatBoost (AllKNN-CatBoost) is regarded as the best suggested model out of 330 assessment metric values that took over a month to achieve. As a result, similar studies are compared with the AllKNN-CatBoost model. According to the findings, the suggested model performs better than the others, with an F1-Score of 87.40%, an AUC of 97.94%, and a recall of 95.91%.

An assessment system for risk-based cybersecurity compliance (RC2AS),

M. Ahmed, I. Almomani, and A. Alfaadhel,

Attacks on cybersecurity continue to pose serious risks to people and businesses, impacting almost every element of daily life. As a result, several nations attempt to address this issue by implementing cybersecurity regularity frameworks to preserve the data and digital assets of organisations. By creating the essential cybersecurity control (ECC) as a national cybersecurity regulatory reference, Saudi Arabia has taken proactive measures in this regard. The compliance assessment procedures for the various international cybersecurity standards and controls (ISO2700x, PCI, and NIST) are typically generic for all organisations with varying scopes, business functionality, and criticality levels; the security control risk is not taken into account, and the overall compliance score is absent. In order to overcome all of these drawbacks, this study builds a thorough and personalised risk-based cybersecurity compliance assessment system (RC2AS) using the ECC as a baseline. ECC was selected because of its clarity and inspiration from several international standards. The paucity of relevant research that have thoroughly examined ECC is another factor in this decision. The purpose of RC2AS is to work with the existing ECC tool. With the use of its offline self-assessment tool, the organisation may speed up the evaluation process, pinpoint existing shortcomings, and improve planning to raise its level in accordance with its priorities.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

Furthermore, RC2AS suggests four ways to determine the total ECC compliance score. To evaluate these approaches and compare their effectiveness, a number of scenarios are run. The objective is to accurately represent an organization's compliance score while taking into account its domain, requirements, resources, and security control risk level. Lastly, the assessment process's results are shown in rich dashboards that provide a thorough presentation of the company's cybersecurity maturity and provide a roadmap for raising its compliance level.

"Community detection algorithm for bank fraud detection,"

S. Hossain, D. Sarma, W. Alam, I. Saha, M. N. Alam, and M. J. Alam,

Bank fraud is a federal offence that entails dishonest efforts to defraud financial organisations in order to get financial advantages. Fraud costs banks and other financial organisations billions of dollars annually. Scammers use deception to get bankers to part with their money. Debit and credit card fraud, account fraud, insurance fraud, money laundering fraud, and other forms of bank fraud are the most prevalent. To protect the global financial system, bankers must protect both their institutional integrity and their financial assets. The dodging tactics used by scammers often get by anti-fraud defence systems. Using a community detection algorithm that finds patterns that may indicate fraud, this research suggested a method for detecting bank fraud. The web-based tool to identify the scam was designed using an agile methodology. The application served as a focal point for communication between clients and banks. The database was created and represented using Neo4j, a graph database, and the graph query language was Cypher. All of the frauds that were shown throughout the test experiment were effectively identified by the suggested approach. This article will help bankers fight fraud by identifying and preventing similar incidents.

"Research the issue of class imbalance using a modified KNN for classification,"

B. Kanisha, S. Kaliraj, and R. Sasirekha,

In the present day, identifying data imbalance is quite difficult. There would be a lot of data in a data warehouse, but in any kind of industry, maintaining the balance of data and managing it are highly challenging tasks. Data imbalance occurs when specimens are categorised according to their behaviour. In order to determine the most effective method for addressing data imbalance issues, this study analyses the imbalance condition of the data and thoroughly examines machine learning approaches. It is possible to do extensive analysis of the k-nearest neighbour (KNN) technique to maintain the categorisation of specimens in an equal group.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

Numerous studies in the literature assess financial fraud using statistical techniques. In particular, it was discovered that important research used autoregressive (AR) and ordinary least squares (OLS) regression models to assess financial fraud. The research [21] examines the relationship between auditor attributes and fraud detection in developing countries using a regression model using the Tehran Stock Exchange dataset. The authors provide helpful details to increase the results' dependability. The research [22] provides insights into the relationship between politics and fraud by examining the impact of political alignment on corporate fraud convictions using pooled OLS and panel regressions.

The evaluation of financial fraud from the standpoint of risk reduction is provided by an existing system. To determine the degree of fraud risk, the current research use a variety of risk metrics, including value-at-risk (VaR), anticipated loss, and projected deficit. By linking the motive of the fraud triangle to human characteristics that result in certain acts and the meta-model of fraud together, the research [29] provides ways for decomposing the risk of fraud, detecting possible fraudsters, and allowing more focused anti-fraud

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

efforts. In order to ascertain how control environments, risk assessments, control activities, information and communication, and monitoring contributed to fraud prevention and detection efforts in Indonesian organisations, the research [30] uses regression analysis to examine how enterprises manage risk.

Using chi-square, fisher test, and correlation, the study [35] found a positive relationship between fraud risk assessment and management and the effective use of forensic accounting; however, there is no relationship between fraud risk assessment and management and fraud-causing techniques. In addition to handling data skewness, a triage model that takes input from the ensemble model, and a risk model that assesses the financial losses, the research [9] investigates fraud utilising ensemble learners for anomaly detection. From machine learning approaches to risk assessment, the authors effectively provide an efficient fraud risk-based detection method. However, they do not examine fraud detection by taking into account the risk component before putting it via machine learning detection.

Studies that classify fraud applications using machine learning approaches are presented by an existent system. Most of the research that are given take into account detection while addressing the skewed nature of fraud cases. To counteract the skewed nature of fraud datasets, sampling techniques, hybrid techniques, and other innovative approaches are mostly used. The work [36] uses support vector machines (SVM) and quantum machine learning (QML) to solve class skewness in credit card fraud. The findings demonstrate that although QML applications may be used to time-series-based and highly skewed data, traditional machine learning approaches are still helpful for non-time series data. According to the research, quantum neural networks (QNNs) perform well in fraud detection [37]. XGBoost outperformed all other machine learning models in the research [38], which trained several models with default implementations and settings. The research [39] evaluates the efficacy of telecom fraud using a dynamic graph neural network (DGNN). The authors successfully provide a recommended approach to address the problem of telecom fraud detection in large phone social networks.

Disadvantages

- Although fraud cases are uncommon and result in significant losses when they do occur, the majority of current research does not account for possible losses of fraud risk characteristics.
- A highly skewed distribution results from the intrinsic skewness of fraud occurrences relative to non-fraud cases.
- While methods like logistics regression or regression presuppose normalcy and projections may provide an erroneous result, fraud patterns often have more irregular and high values.

PROPOSED SYSTEM

The stages and procedure involved in NBA fraud detection are described in the suggested design that was put into practice. An essential component of this study is value-at-risk, which is intended to simulate the severe and extreme fraud risk characteristics. It also concentrates on infrequent fraud cases that, when they do occur, are harmful and very expensive. Machine learning algorithms, particularly distance-based ones like KNN, might be distorted by the few but highly skewed examples. In contrast to traditional approaches that use a constant fraud probability weight that is associated to the skewed fraud cases, the value-at-risk may manage the fraud skewness by using flexible threshold probability ranges (confidence level). Value-at-risk received the preprocessed, extracted, and engineered characteristics as input for simulation. In the meanwhile, a distance-based KNN is made to be adjustable in order to discover unusual clusters with a closest neighbour distance of k, which may be used to detect fake characteristics. Especially for the KNN model with hyperparameter k, the selected confidence level treats the infrequent fraud events as higher risk characteristics that would lead to less training

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

sets. For the fraud detection model to adequately simulate the fraudulent characteristics in the uncommon cluster, k must be optimised to a lower level. By giving nearby instances a larger weight, the distance weight of KNN effectively reduces fraud skewness and makes it easier to identify skewed examples.

Advantages

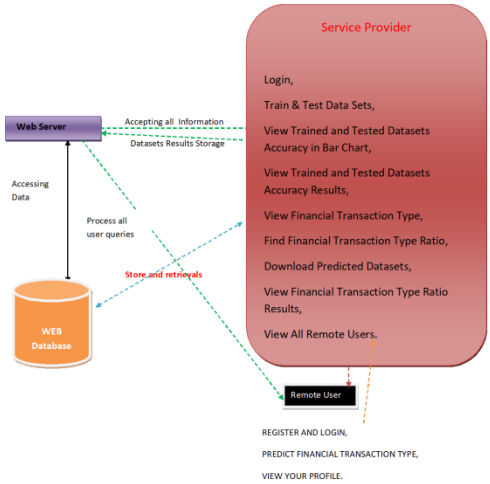
• Rather of modelling the fraud pattern, this research employed an extreme value theorem to predict the tails (possible losses).

• To more effectively represent the skewness of fraud cases, this article used value-at-risk.

• Since this study made no assumptions about any distribution, it used historical simulation to determine value-at-risk.

• To measure the total effectiveness of detecting NBA fraud cases that include risk fraud elements, this study used innovative detection rate performance indicators.

SYSTEM ARCITECTURE



IV. IMPLEMENTATION

MODULES:

Service Provider

The Service Provider must use a working user name and password to log in to this module. Following a successful login, he may do several tasks including training and testing data sets, See the Accuracy of Trained and Tested Datasets in a Bar Chart View Financial Transaction Type, Find Financial Transaction Type Ratio, Download Predicted Datasets, and View Accuracy Results of Trained and Tested Datasets View All Remote Users and Financial Transaction Type Ratio Results.

View and Authorize Users

The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

Remote User

A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user will be able to see their profile, register and log in, and choose the kind of financial transaction.

ALGORITHMS

Logistic regression Classifiers

The relationship between a collection of independent (explanatory) factors and a categorical dependent variable is examined using logistic regression analysis. When the dependent variable simply has two values, like 0 and 1 or Yes and No, the term logistic regression is used. When the dependent variable contains three or more distinct values, such as married, single, divorced, or widowed, the technique is sometimes referred to as multinomial logistic regression. While the dependent variable's data type differs from multiple regression's, the procedure's practical application is comparable.

When it comes to categorical-response variable analysis, logistic regression and discriminant analysis are competitors. Compared to discriminant analysis, many statisticians believe

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

that logistic regression is more flexible and appropriate for modelling the majority of scenarios. This is due to the fact that, unlike discriminant analysis, logistic regression does not presume that the independent variables are regularly distributed.

Both binary and multinomial logistic regression are calculated by this software for both category and numerical independent variables. Along with the regression equation, it provides information on likelihood, deviance, odds ratios, confidence limits, and quality of fit. It does a thorough residual analysis that includes diagnostic residual plots and reports. In order to find the optimal regression model with the fewest independent variables, it might conduct an independent variable subset selection search. It offers ROC curves and confidence intervals on expected values to assist in identifying the optimal classification cutoff point. By automatically identifying rows that are not utilised throughout the study, it enables you to confirm your findings.

Naïve Bayes

The supervised learning technique known as the "naive bayes approach" is predicated on the straightforward premise that the existence or lack of a certain class characteristic has no bearing on the existence or nonexistence of any other feature. However, it seems sturdy and effective in spite of this. It performs similarly to other methods of guided learning. Numerous explanations have been put forward in the literature. We emphasise a representation bias-based explanation in this lesson. Along with logistic regression, linear discriminant analysis, and linear SVM (support vector machine), the naive bayes classifier is a linear classifier. The technique used to estimate the classifier's parameters (the learning bias) makes a difference.

Although the Naive Bayes classifier is commonly used in research, practitioners who want to get findings that are useful do not utilise it as often. On the one hand, the researchers discovered that it is very simple to build and apply, that estimating its parameters is simple, that learning occurs quickly even on extremely big datasets, and that, when compared to other methods, its accuracy is rather excellent. The end users, however, do not comprehend the value of such a strategy and do not get a model that is simple to read and implement.

As a consequence, we display the learning process's outcomes in a fresh way. Both the deployment and comprehension of the classifier are simplified. We discuss several theoretical facets of the naive bayes classifier in the first section of this lesson. Next, we use Tanagra to apply the method on a dataset. We contrast the outcomes (the model's parameters) with those from other linear techniques including logistic regression, linear discriminant analysis, and linear support vector machines. We see that the outcomes are quite reliable. This helps to explain why the strategy performs well when compared to others. We employ a variety of tools (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b, and RapidMiner 4.6.0) on the same dataset in the second section. Above all, we make an effort to comprehend the outcomes.

Random Forest

Random forests, also known as random decision forests, are ensemble learning techniques that build a large number of decision trees during training for tasks like regression and classification. The class chosen by the majority of trees is the random forest's output for classification problems. The mean or average forecast of each individual tree is given back for regression tasks. The tendency of decision trees to overfit to their training set is compensated for by random decision forests. Although random forests are less accurate than gradient enhanced trees, they often perform better than choice trees.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

However, their performance may be impacted by data peculiarities.

Tin Kam Ho[1] developed the first algorithm for random decision forests in 1995 by using the random subspace technique, which in Ho's definition is a means of putting Eugene Kleinberg's "stochastic discrimination" approach to classification into practice.

Leo Breiman and Adele Cutler created an algorithm extension and filed for a trademark in 2006 for "Random Forests" (owned by Minitab, Inc. as of 2019). The extension builds a set of decision trees with controlled variance by combining Breiman's "bagging" concept with random feature selection, which was initially proposed by Ho[1] and then separately by Amit and Geman[13].

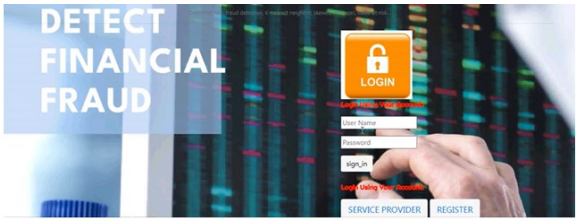
Businesses often employ random forests as "blackbox" models since they need minimal setup and provide accurate forecasts across a variety of inputs.

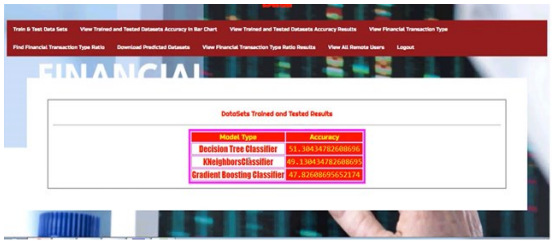
SVM

The goal of a discriminant machine learning approach in classification problems is to identify a discriminant function that can accurately predict labels for newly acquired instances based on an independent and identically distributed (iid) training dataset. A discriminant classification function takes a data point x and assigns it to one of the several classes that are part of the classification job, in contrast to generative machine learning techniques that call for calculations of conditional probability distributions. Discriminant techniques are less effective than generative approaches, which are mostly used when prediction entails the identification of outliers. However, they need less training data and processing resources, particularly when dealing with a multidimensional feature space and when just posterior probabilities are required. Finding the equation for a multidimensional surface that optimally divides the various classes in the feature space is the geometric equivalent of learning a classifier.

SVM is a discriminant approach that, unlike genetic algorithms (GAs) or perceptrons, which are both often used for classification in machine learning, always returns the same optimum hyperplane value since it solves the convex optimisation issue analytically. The initialisation and termination criteria have a significant impact on the solutions for perceptrons. While the perceptron and GA classifier models are distinct every time training is started, training yields uniquely specified SVM model parameters for a given training set for a certain kernel that converts the data from the input space to the feature space. The only goal of GAs and perceptrons is to reduce training error, which will result in several hyperplanes satisfying this criterion.

V. SCREEN SHOTS



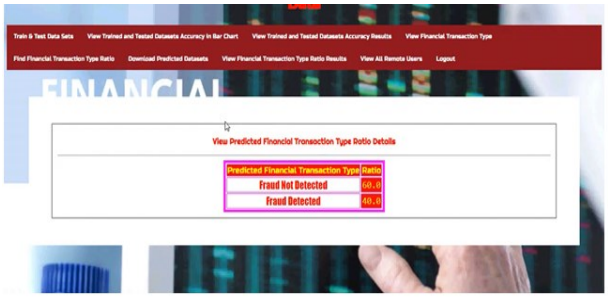


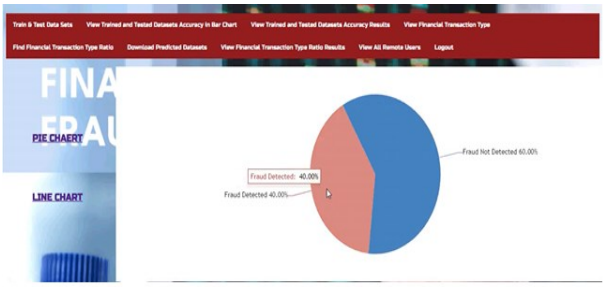
International Journal of Engineering Research and Science & Technology

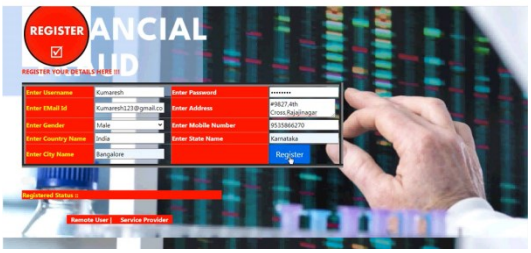
ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922







VI. CONCLUSION

The value-at-risk-based fraud detection approach described in this work makes it possible to quantify and mitigate fraud risk characteristics while also overcoming the impact of skewed fraud cases, both of which are critical in addressing financial fraud issues. The value-at-risk gives the infrequent fraud scenarios with closest neighbour distance k a confidence probability weight. By giving nearby examples a larger weight, the distance weight of KNN effectively reduces class skewness and makes it easier to identify skewed instances. By using projected shortfall and expected loss by value at risk, risk may be quantified in mean, worst-case, and extreme situations, allowing their strengths to be aggregated. As a result, a precise fraud detection system helps businesses make wise decisions and lower the total cost of fraud detection and prevention. The experiment's time frames are not taken into account in this research. The main obstacle, however, is the scarcity of data for NBA fraud detection.

REFERENCES

[1] ACFE. Association of Certified Fraud Examiners (ACFE) 2022 Report to the Nations. Accessed: 2023. [Online]. Available: https://legacy.acfe.com/report-to-the-nations/2022/

[2] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. Alsafari, and I. A. Hameed, ‘‘A machine learning and blockchain based efficient fraud detection mechanism,’’ Sensors, vol. 22, no. 19, p. 7162, Sep. 2022.

[3] N. S. Alfaiz and S. M. Fati, ‘‘Enhanced credit card fraud detection model using machine learning,’’ Electronics, vol. 11, no. 4, 662, 2022.

[4] A. Alfaadhel, I. Almomani, and M. Ahmed, ‘‘Risk-based cybersecurity compliance assessment system (RC2AS),’’ Appl. Sci., vol. 13, no. 10,p. 6145, May 2023.

[5] D. Sarma, W. Alam, I. Saha, M. N. Alam, M. J. Alam, and S. Hossain,‘‘Bank fraud detection using community detection algorithm,’’ in Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA), Jul. 2020, pp. 642–646.

[6] A. Pagano, ‘‘Digital account opening fraud on demand deposit accounts:An assessment of available technology,’’ Ph.D. thesis, Utica College, Utica,NY, USA, 2020.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp913-922
Vol. 21, Issue 2, 2025

[7] Shuftipro. New Account Fraud—A New Breed of Scams. Accessed: 2023. [Online]. Available: https://shuftipro.com/reports-whitepapers/newaccount-fraud.pdf
[8] R. Sasirekha, B. Kanisha, and S. Kaliraj, ‘‘Study on class imbalance problem with modified KNN for classification,’’ in Intelligent Data Communication Technologies and Internet of Things, vol. 101. Singapore: Springer, 2022, pp. 207–217, doi: https://doi.org/10.1007/978-981-16-7610-9_15.
[9] P. Vanini, S. Rossi, E. Zvizdic, and T. Domenig, ‘‘Online payment fraud: From anomaly detection to risk management,’’ Financial Innov., vol. 9, no. 1, p. 66, Mar. 2023, doi: 10.1186/s40854-023-00470-w.
[10] X. Zhu, X. Ao, Z. Qin, Y. Chang, Y. Liu, Q. He, and J. Li, ‘‘Intelligent financial fraud detection practices in post-pandemic era,’’ Innovation,vol. 2, no. 4, Nov. 2021, Art. no. 100176, doi: 10.1016/j.xinn.2021.100176.
[11] M. Monge, C. Poza, and S. Borgia, ‘‘A proposal of a suspicion of tax fraud indicator based on Google Trends to foresee Spanish tax revenues,’’ Int. Econ., vol. 169, pp. 1–12, May 2022, doi: 10.1016/j.inteco.2021.11.002. 12] S. Kannan and K. Somasundaram, ‘‘Autoregressive-based outlier algorithm to detect money laundering activities,’’ J. Money Laundering Control, vol. 20, no. 2, pp. 190–202, May 2017, doi: 10.1108/jmlc-07-2016-0031.
[13] B. Xiao, B. Lei,W. Lan, and B. Guo, ‘‘A blockwise network autoregressive model with application for fraud detection,’’ Ann. Inst. Stat. Math., vol. 74, no. 6, pp. 1043–1065, Dec. 2022, doi: 10.1007/s10463-022-00822-w.
[14] G. Moschini, R. Houssou, J. Bovay, and S. Robert-Nicoud, ‘‘Anomaly and fraud detection in credit card transactions using the ARIMA model,’’ in Proc. 7th Int. Conf. Time Forecasting, Jul. 2021, p. 56, doi: 10.3390/engproc2021005056.
[15] A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, and R. Effghi, ‘‘An ensemble-based fraud detection model for financial transaction cyber threat classification and countermeasures,’’ Eng., Technol.Appl. Sci. Res., vol. 13, no. 6, pp. 12433–12439, Dec. 2023, doi:10.48084/etasr.6401.
[16] R. M. Aziz, R. Mahto, K. Goel, A. Das, P. Kumar, and A. Saxena,‘‘Modified genetic algorithm with deep learning for fraud transactions of ethereum smart contract,’’ Appl. Sci., vol. 13, no. 2, p. 697, Jan. 2023, doi:10.3390/app13020697.
[17] M. Hegazy, A. Madian, and M. Ragaie, ‘‘Enhanced fraud miner: Credit card fraud detection using clustering data mining techniques,’’ Egyptian Comput. Sci. J., vol. 40, no. 3, pp. 1–10, 2016.
[18] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. Ribeiro, J. Gama, and P. Bizarro, ‘‘Turning the tables: Biased, imbalanced, dynamic tabular datasets for ML evaluation,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 35, 2022, pp. 33563–33575.
[19] J. Pombal, P. Saleiro, M. A. T. Figueiredo, and P. Bizarro, ‘‘Fairness-aware data valuation for supervised learning,’’ 2023, arXiv:2303.16963.
[20] T. Awosika, R. Mani Shukla, and B. Pranggono, ‘‘Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection,’’ 2023, arXiv:2312.13334.