ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Conference Proceedings

Open Access

Advanced Machine Learning Models for Detecting and Classifying Financial Fraud in Big Data-Driven (Approved by ICITET 2024)

Achuthananda Reddy Polu[1*], Bhumeka Narra[2], Dheeraj Varun Kumar Reddy Buddula[3], Hari Hara Sudheer Patchipulusu[4], Navya Vattikonda[5] and Anuj Kumar Gupta[6]

[1]Senior SDE, Cloudhub IT Solutions, USA

[2]Sr Software Developer, Statefarm, USA

[3]Software Engineer, Elevance Health Inc, USA

[4]Senior Software Engineer, Walmart, USA

[5]Business Intelligence Engineer, International Medical Group Inc, USA

[6]Oracle ERP Senior Business Analyst ,Genesis Alkali, USA

ABSTRACT

The banking sector faces a major challenge in identifying credit card fraud, especially as online transactions increase. This study employs the Kaggle Credit Card Fraud Detection dataset to present a machine learning (ML)-based method to credit card fraud detection. The collection contains de-identified transaction information from European cardholders. With 284,807 transactions, only 492 included fraud, suggesting a significant disparity in class. Therefore, data balancing techniques were used to improve model training. Among the data pretreatment procedures were label encoding for categorical conversion and standardization to normalize feature scales. Using Euclidean distance, for the purpose of identifying the majority-belonging k-nearest neighbor class, A classifier called K-Nearest Neighbors (KNN) was created. The model was assessed using ROC-AUC, F1-score, accuracy, precision, recall, and K-fold cross-validation, among other important performance measures. The KNN model outperformed benchmark models like MLP and Naïve Bayes, obtaining 98.56% accuracy and an AUC of 96.07, according to experimental data, demonstrating great classification efficacy. The promise of KNN in creating reliable and accurate fraud detection systems for cybersecurity applications is confirmed by these findings.

***Corresponding author**

Achuthananda Reddy Polu, Senior SDE, Cloudhub IT Solutions, USA.

Received: June 06, 2024; **Accepted:** June 10, 2024; **Published:** June 10, 2024

Keywords: Financial Fraud detection, Credit Card Fraud, big data, Machine Learning K-Nearest Neighbor (KNN)

Introduction

The banking industry has been reminded again in recent years that preserving customer confidence is the most crucial element in ensuring its continued existence. Following the second-largest shock to the U.S. banking industry since the 2008 crisis, which was the fall of Silicon Valley Bank (SVB), the significance of banks maintaining client safety and confidence was underlined [1]. Although banking crises may destabilise financial markets and impede global economic progress, the growing danger of the financial industry is battling the formidable problem of fraud. organizations in the age of constantly changing technology and artificial intelligence [2,3]. Risk assessment is essential for financial firms to control risk and avoid mistakes [4]. Credit card fraud can result from a credit card being compromised by cybercriminals. By obtaining unauthorised access to credit card information, the fraudster commits fraudulent behaviour, which results in a financial loss for both the client and the business [5]. The complexity and frequency of fraudulent acts, coupled with the proliferation of digital transactions, have become growing concerns. The issue of credit card theft in online banking may be addressed with the use of trustworthy, scalable, and real-time systems [6]. Fraud detection now heavily relies on ML models as they can recognize intricate patterns and adjust to changing behavior [7]. However, the vast volume of transaction data and the inherent class imbalance in fraud detection activities require distributed systems capable of efficiently addressing these computational issues [8].

The goal of this study was to identify fraudulent financial transactions using ML models. The study's objective was to develop algorithms that could consistently recognise these types

of transactions [9]. ML algorithms and pre-processing approaches were among the strategies employed [10]. The potential of the suggested approach. This work is important because it can help identify fraudulent bank transactions more effectively, particularly during the epidemic, when a lot of transactions have moved online, and during wartime, when a lot of organisations and activities are raising money [11].

Motivation and Contribution of Study

A major rise in credit card theft has resulted from the quick expansion of digital transactions, presenting major financial risks to financial organisations and the people that use them. Fraud detection in real time is still difficult since fraud datasets are very unbalanced and fraudsters' strategies are always changing. Because they make too many false predictions about positive or negative outcomes, current fraud detection systems provide misleading findings regarding fraudulent actions. This research drives toward creating an efficient ML model with accurate results for detecting fraudulent transactions, especially when datasets exhibit heavy non-fraudulent class dominance.

Contributions

- **Data Balancing Strategy:** The study uses a balancing technique to address class imbalance problems, which both improves model learning capability and enhances accuracy in fraud detection.
- **Pre-processing Enhancement:** A combination of labelling techniques, together with normalization methods, is applied to data for machine learning enhancement purposes.
- **Application of KNN Classifier:** The KNN classification method uses Euclidean distance to perform simple yet dependable fraud detection so that it can be implemented effectively in these scenarios.
- **Comprehensive Evaluation:** The examination assesses six parameters used to measure the model's performance that provide a comprehensive assessment system: precision, accuracy, recall, confusion matrix, F1-score, and ROC-AUC.
- **Comparative Analysis:** Real-world deployment of fraud detection becomes possible because the suggested KNN model outperforms the MLP and NB approaches.

Structure of Paper

The paper is organized as follows: Section II, literature review, and Section III, methodology, are followed by Section IV, results and discussion, and Section V, conclusion and future work.

Literature Review

The credit card has garnered significant attention since its introduction as a device that combines Internet technology with financial operations. Due to the growing prevalence of credit card theft, several academics have conducted extensive study on the issue in recent years.

Kumar et al. the RFA is used in the best way to assess the veracity of offered methods for detecting fraudulent transactions. The basis of this approach is supervised learning, which uses DT to classify datasets. After classifying the dataset, a confusion matrix is generated. The confusion matrix is used to assess how well the RF Algorithm works. The accuracy of the dataset processing results is around 90% [12].

Pillai et al. suggest employing DL techniques to create a very effective approach for identifying fraudulent credit card transactions. Their findings demonstrate the equivalent effectiveness of logistic and hyperbolic tangent activation functions in detecting authentic credit card transactions. The logistic activation function works better with 100 nodes (sensitivity: 83%) and 10 nodes (82%), according to the three-hidden-layer model. At 1000 nodes, the hyperbolic tangent activation function performs better than the other two choices, achieving an 82% sensitivity for all three hidden layer numbers. This research will help us determine which model is appropriate for DL in order to achieve the greatest outcomes at the lowest possible cost [13].

Popat and Chaudhary main objective is to protect credit card transactions so that users may conveniently and safely utilise e-banking. There are several methods for identifying credit card fraud, including DL, LR, NB, SVM, NN, Artificial Immune System, KNN, Data Mining, DT, Fuzzy Logic-based System, Genetic Algorithm, and more [14].

Financial transaction fraud has the potential to seriously harm a company's reputation among clients. Focusing on a range of fraud detection techniques as well as innovative approaches to address and prevent them is therefore becoming more and more crucial. This article uses clustering based on autoencoders. A three-layered hidden-layer k-means clustering autoencoder has been used and assessed on 284807 bank transactions across Europe. The outcomes showed that this strategy outperformed the others, with a TPR of 81% and an accuracy of 98.9% [15].

Awoyemi et al. collected credit card transactions from 284,807 cardholders in Europe was collected. Oversampling and undersampling are used in a hybrid technique that distorts the data. The three methods are used to preprocess raw data. Python is employed to finish the assignment. The approaches' performance is measured using the following metrics: balanced classification rate, accuracy, sensitivity, specificity, precision, and Matthew's correlation coefficient. At 97.92%, 97.69%, and 54.86%, respectively, the data revealed that NB, KNN, and LR were the classifiers with the greatest accuracy rates. According to the comparing findings, KNN outperforms logistic regression and NB [16].

Mahmud examines and contrasts a few popular classifier algorithms that are most commonly employed to detect credit card abuse. Additionally, the metric utilised to rank such algorithms and evaluate their classification performance is the main emphasis on the dataset from the 2009 Tests of performance conducted for the UCSD-FICO Data Mining Competition. The results of the experiment show that (1) the fraud detection success rate was less than 50%, despite the classification accuracy rate being 98.25% [17].

Kho and Vea suggest that in order to catch any potentially unusual transactions, a detection model should be present as a backup in case Technology malfunctions. The highest accuracy values, 94.32% and 93.50%, were obtained by the Random Tree and J48, respectively, out of a number of classifiers that were assessed during the model-building process. Examining these two classifiers in detail reveals that the J48 is more suited to comprehending the data from transaction logs [18].

Table 1 summarizes the research on financial fraud categorization and detection using ML, organized by methodology, data sources, key findings, limitations, and suggested directions for future research.

Table 1: Comparative Analysis of Machine Learning Techniques for Credit Card Fraud Detection

Author	Methodology	Data	Key Findings	Limitation	Future Work
Kumar et al.	Confusion matrix analysis combined with the Random Forest Algorithm (RFA)	Credit card transaction dataset (unspecified, assumed public dataset)	Achieved about 90% accuracy	Dataset size/type not specified; only Random Forest evaluated	Compare with other classifiers; optimize hyperparameters
Pillai et al.	Deep learning models with logistic and tanh activation functions; varying nodes/layers	Credit card transactions (dataset unspecified)	Logistic function: Sensitivity 82%-83% (10-100 nodes); Tanh: Sensitivity 82% (1-3 hidden layers, 1000 nodes)	Dataset details missing; comparison limited to activation functions	Explore more activation functions, optimizers, and model architectures
Popat and Chaudhary	Review study on ML/DL techniques for fraud detection	General survey (no specific dataset)	Listed multiple techniques (SVM, Neural Networks, KNN, Decision Trees, Fuzzy Logic, etc.)	No experiments; only theoretical review	Practical implementation and evaluation on real-world datasets
Zamini and Montazer	Autoencoder-based unsupervised clustering with K-Means	European credit card transactions dataset (284,807 transactions)	Accuracy: 98.9%; True Positive Rate: 81%	Risk of overfitting; unsupervised approach may not generalize well	Combine supervised and unsupervised approaches for better results
Awoyemi, Adetunmbi, and Oluwadare	hybrid strategy for under- and over-sampling; classifiers like KNN, Naïve Bayes, and Logistic Regression	284,807 European credit card transactions in one dataset	KNN outperformed Naïve Bayes (97.92%), and the accuracy rate of logistic regression (54.86%)	Limited to basic classifiers; did not explore deep learning	Implement more complex ML models; optimize sampling techniques
Mahmud	Comparative study of multiple popular classifiers	Information from the 2009 UCSD-FICO Data Mining Contest	Although the fraud detection rate was less than 50%, the classification accuracy was 98.25%.	High overall accuracy masks poor fraud detection	Focus on improving sensitivity/recall for fraud class
Kho and Vea	Classifier evaluation (Random Tree and J48)	Transaction logs dataset (unspecified size)	J48 achieved 93.50% accuracy, Random Tree 94.32%; J48 better in understanding patterns	Dataset size not specified; limited classifier variety	Test ensemble and boosting methods; validate on larger datasets

Methodology

This study made use of an anonymized fraud detection model created with the Kaggle dataset for detecting credit card fraud, shown in Figure 1, which contained European cardholder transactions. Data preprocessing included standardization to normalize feature distributions to Label encoding is used to translate category data into numerical form, together with a unit variance and zero mean. To ensure accurate model evaluation, K-fold cross-validation was used to separate the dataset into training and testing sets. A KNN classifier was proposed, which used the k nearest neighbors' majority class for classification, and the distance between points was determined using the Euclidean distance formula. The model's performance was assessed using metrics including accuracy, precision, recall, and F1-score to determine how successfully it recognised fraudulent transactions; a confusion matrix is also crucial.

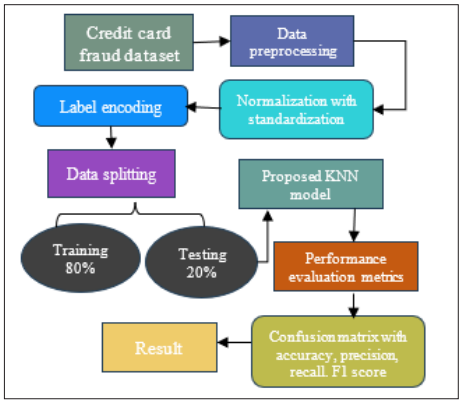

Figure 1: Flow Chart of the Credit Card Fraud Detection for Cybersecurity Farmwork

The following sections provide each step description that also shows in methodology and proposed flowchart:

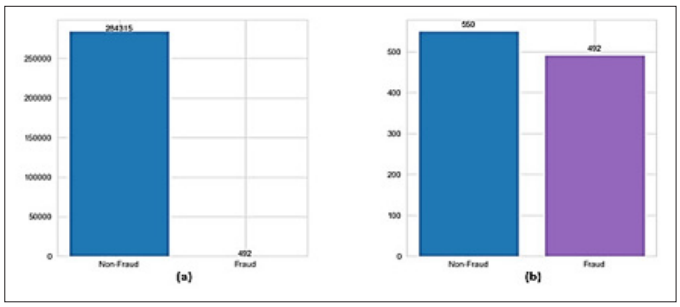

Figure 2: Data Balancing of the Credit Card Fraud Dataset

Figure 2 shows two bar charts that show the ratio of fraud to non-fraud cases, both prior to and during dataset balancing. There is a significant disparity between the original dataset and figure (a), which shows that there are only 284,315 non-fraud instances and 492 fraud cases. In contrast, chart (b) displays the balanced dataset after applying a balancing technique, where there are far fewer incidents of fraud and non-fraud, 550 non-fraud cases compared to 492 fraud cases, thereby addressing the class imbalance problem and creating a more suitable dataset for training machine learning models.

Data Collection

The anonymized According to MLG-ULB (2013), The Credit Card Fraud Detection dataset, which includes credit card transactions made by European cardholders during two days in September 2013, is available to the public on Kaggle. The data collection comprises 284,807 transactions, 492 of which are fake. There are 28 characteristics in the dataset, 27 of which are numerical characteristics produced via PCA transformation for secrecy reasons. The last feature shows whether or not a transaction was fraudulent. There is a noticeable class imbalance in the database since the majority of transactions do not fit into the fraud category. The construction of an accurate fraud transaction classifier becomes complex because it requires minimizing incorrect positive results. The accessible dataset has aided several studies and contests that evaluated ML models to detect credit card fraud. The provided dataset exists for fraud detection algorithm development research, while researchers must obtain permission to use it commercially. This dataset is accessible to the public and permits use either for academic pursuits or research goals. The dataset can only be employed for commercial purposes when the dataset owners provide written permission before usage.

Data Pre-Processing

This study's primary objective is to investigate both standardization methods and label encoding techniques which will be implemented on fraudulent transaction data.

Normalization with Standardization

The data normalization and standardization, and label encoding methods created appropriate conditions for ML models to handle the data. Standardization standardized the data through mean normalization to zero and variance normalization to 1, and label encoding transformed categorical data into numerical values. The normalization processes standardized the data to create a unified scale that made it suitable for the successful training and assessing ML models.

Standardization

Data is transformed into a zero mean by standardization distributions that become centered on 0 with unit variance, which results in a standard deviation value of 1. Standardization takes the form of the calculation shown in Equation (1):

$$x_{standardized} = \frac{x - mean(x)}{standard\ deviation(x)} \qquad (1)$$

The formula combines x with the mean(x) and the standard deviation(x) to calculate statistical results. x indicates the initial data value, while mean(x) reflects data averaging, and standard deviation(x) shows data spread.

Label Encoding

Encoding labels in order to do mathematical calculations and analysis, ML models acquire the capability of converting categorical data into numerical values. Giving each distinct category in the data an integer label is the most basic type of label encoding. The label encoding formula is computing Equation (2).

$$x_{encoded} = lable(x) \qquad (2)$$

In this case, Label(x), where x is the original categorical variable, is the numerical label assigned to each unique category value. For example, the label encoding transforms the regular and fraudulent labels into 0 and 1, respectively.

Data Splitting

The dataset was split in half using the k-fold cross-validation approach. By using 80% of the data to construct the model and 20% to evaluate the model's prediction ability, this method made it possible to create distinct training and test sets.

Proposed KNN Model

A supervised technique that every data analyst should be aware with KNN. And the KNN techniques select an integer k that divides the data from its nearest neighbors once more. It is mostly used in the classification process. The categorization of a new data point is influenced by how similar it is to previously categorized data. To separate the data from its closest neighbors once more, KNN algorithms use an integer k in Equation (3):

$$d(p,q) = \sqrt{\sum_{1=1}^{n}(P_1} - q_1)^2 \qquad (3)$$

It uses a specific norm to compute the distance between locations. For the new observation, the class with the greatest number of K nearest points is the one that was selected.

Performance Evaluation Metrics

The performance assessment measures are used in this section to evaluate how well their suggested technique detects fraudulent transactions. The following metrics are used:

Confusion Matrix: This matrix provides a comprehensive understanding of the model's classification performance by dividing predictions into TP, TN, FP, and FN. Table II shows the confusion matrix.

Accuracy

The model's total performance in properly identifying transactions that are either fraudulent or not is measured by its accuracy. It

may be expressed as the proportion of instances in Equation (4) that are properly classified to all cases.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

Precision

The proportion of fraudulent transactions that show the model's ability to avoid false positives is the definition of accuracy that the model accurately detects out of all those that it has detected as fraudulent. Inside Equation (5).

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall

In Equation (6), recall also known as sensitivity or true positive rate is the proportion of actual fraudulent transactions that the model correctly identifies.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1 Score

Table II below examines the results of the proposed models are evaluated equitably by taking into consideration both FP and FN, and performance measures such as accuracy, precision, recall, F1-score, and ROC graphs are used to identify credit card fraud in Equation (7).

$$F1-score = 2 * \frac{(precision*recall}{(precision+recall)} \quad (7)$$

Results and Discussion

In order to obtain precise dataset filing and model training results, experiments were carried out utilising contemporary hardware. An Intel Core i9-13900K CPU operating at 3.0 GHz, an NVIDIA RTX4090 GPU with 24GB VRAM, and 64 GB DDR5 RAM made up the system. Windows 11 Pro was the operating system that was used. In order to identify credit card fraud, Table II below analyses the results of the proposed models using performance measures such as accuracy, precision, recall, F1-score, and ROC graphs.

Table 2: Performance Parameters of KNN Model

Measures	Performance
Accuracy	98.56
Precision	98.62
Recall	98.54
F1-score	98.56
AUC	96.07

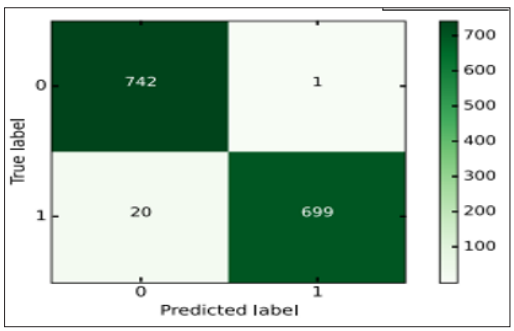

Figure 3: Confusion Matrix for KNN Model

The binary classification model performance appears through the confusion matrix illustrated in Figure 3. The matrix shows how many accurate and inaccurate predictions the model produced in relation to the actual results. Specifically, there were 699 cases of class 1 (TP), while 742 cases of class 0 (TN) were correctly predicted by the model. There were twenty instances of class 1 being mistakenly classed as class 0 (FN) and one instance of class 0 being mistakenly classified as class 1 (FP). Darker hues denote greater values in the color gradient, which shows the density of predictions.

The binary classification model's ROC curve appears in Figure 4 which demonstrates the relationship between sensitivity (or FP rate) and TP rate at various threshold values. Excellent model performance is shown by the blue-plotted curve's AUC value of 0.962. A reference line representing a random classifier is shown in orange, running diagonally from (0,0) to (1,1). Additionally, a black dot marks the curve's "Best" point, which denotes the ideal threshold with the optimum ratio of false positives to real positives.

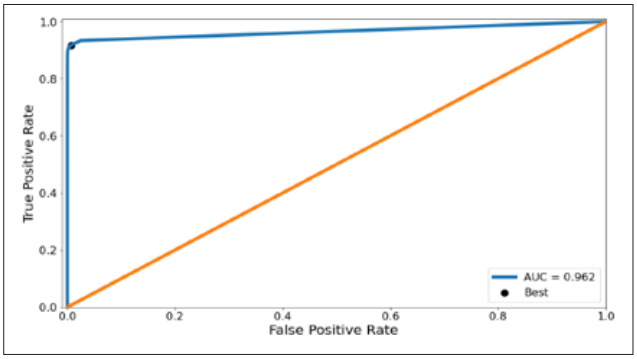

Figure 4: ROC-AUC curve of KNN Model

Table 3: Comparison between Base and Proposed Model Performance for Credit Card Fraud Detection

Matrix	Accuracy
KNN	98.56
MLP [13].	83
NB [16].	97.69

The evaluated KNN model achieved better results in credit card fraud detection than all current baseline models, according to Table III. The KNN model demonstrated enhanced accuracy of 98.56% while surpassing the accuracy levels of the MLP model, developed at 83%. The performance of KNN surpasses NB, as the NB model reached only 97.69% accuracy. The combination of a KNN classifier with proper data balancing and preprocessing makes it highly effective for finding fraudulent transactions within unbalanced datasets while achieving excellent performance results.

The KNN delivers various organizational advantages when used for financial fraud detection. KNN demonstrates effectiveness in detecting financial fraud due to its ability to operate without distribution assumptions while processing instance-based and non-parametric learning. The model performance metrics reveal excellent values with 98.56% accuracy and 98.62% precision and 98.54% recall, and 98.56% F1-score alongside 96.07% AUC. The model's ability to differentiate between actual payments is impressive and fraudulent transactions using data, which provides a well-balanced trade-off between detection and misidentification of either type of transaction. KNN functions as an effective solution for fraud detection because its basic design approach brings together powerful performance capabilities.

Conclusion and Future Scope

Financial services experience serious difficulties from credit card fraud activities. A significant amount of money, equivalent to billions of dollars, is stolen through credit card fraud annually. The researchers built an effective KNN-based credit card fraud detection algorithm that achieved 98.56% accuracy by implementing proper data preprocessing alongside data balancing and Euclidean distance for classification. The KNN A model outperformed the MLP and NB algorithms, especially when dealing with unbalanced data, to spot credit card fraud. The model demonstrates reliable fraud transaction recognition capabilities through its high precision values and recall levels, which establish it as an effective tool for operational financial institutions.

The system would benefit from future improvements that comprise ensemble methods and DL framework implementation to achieve enhanced prediction capabilities. To test its performance in dynamic conditions, the model requires assessment using real-time streaming data. Model transparency and stakeholder trust find improvement when explainable AI techniques are integrated into the system. The system's detection ability can improve by investigating different transaction features that depend on time and context, since this will enable it to handle fraud patterns that change in complex financial systems [19-41].

References

1. Varmedja M Karanovic, Sladojevic S, Arsenovic M, Anderla A (2019) Credit Card Fraud Detection - Machine Learning methods. 18th International Symposium INFOTEH-JAHORINA (INFOTEH), IEEE 1-5.
2. Herland M, Bauder RA, Khoshgoftaar TM (2019) The effects of class rarity on the evaluation of supervised healthcare fraud detection models. J Big Data https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0181-8.
3. Kolluri V (2016) An Innovative Study Exploring Revolutionizing Healthcare with AI: Personalized Medicine: Predictive Diagnostic Techniques and Individualized Treatment. J Emerg Technol Innov Res 3.
4. Zareapoor M, Shamsolmoali P (2015) Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier. Procedia Comput Sci 48: 679-685.
5. Chimonaki C, Papadakis S, Vergos K, Shahgholian A (2019) Identification of Financial Statement Fraud in Greece by Using Computational Intelligence Techniques. Business Information Processing 39-51.
6. Thennakoon A, Bhagyani C, Premadasa S, Mihiranga S, Kuruwitaarachchi N (2019) Real-time credit card fraud detection using machine learning. Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 10.1109/CONFLUENCE.2019.8776942.
7. Kolluri V (2015) A Comprehensive Analysis on Explainable and Ethical Machine: Demystifying Advances in Artificial Intelligence. Int Res J 2.
8. Patil S, Nemade V, Soni PK (2018) Predictive Modelling for Credit Card Fraud Detection Using Data Analytics. Procedia Computer Science https://www.sciencedirect.com/science/article/pii/S1877050918309347.
9. Khare N, Yunus Sait S (2018) Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models. Int J Pure Appl Math 118: 825-838.
10. Masters O, Hunt H, Steffinlongo E, Crawford J, Bergamaschi F (2019) Towards a Homomorphic Machine Learning Big Data Pipeline for the Financial Services Sector. IACR Cryptol. ePrint Arch.
11. Rao DD (2009) Multimedia based intelligent content networking for future internet. EMS 2009 - UKSim 3rd Eur Model Symp Comput Model Simul 55-59.
12. Kumar MS, Soundarya V, Kavitha S, Keerthika ES, Aswini E (2019) Credit Card Fraud Detection Using Random Forest Algorithm. 2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT.
13. Pillai TR, Hashem IAT, Brohi SN, Kaur S, Marjani M (2018) Credit Card Fraud Detection Using Deep Learning Technique. 4th International Conference on Advances in Computing, Communication and Automation, ICACCA.
14. Popat RR, Chaudhary J (2018) A Survey on Credit Card Fraud Detection Using Machine Learning. Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI.
15. Zamini Mand G. Montazer, "Credit Card Fraud Detection using autoencoder based clustering," in 9th International Symposium on Telecommunication: With Emphasis on Information and Communication Technology, IST 2018, 2018. doi: 10.1109/ISTEL.2018.8661129.
16. Awoyemi JO, Adetunmbi AO, Oluwadare SA (2017) Credit card fraud detection using machine learning techniques: A comparative analysis. International Conference on Computing Networking and Informatics (ICCNI), IEEE 1-9.
17. Mahmud MS (2017) An evaluation of computational intelligence in credit card fraud detection," in 20th International Computer Science and Engineering Conference: Smart Ubiquitos Computing and Knowledge, ICSEC.
18. Kho JRD, Vea LA (2017) Credit card fraud detection based on transaction behavior. TENCON 2017 - 2017 IEEE Region 10 Conference, IEEE 1880-1884.
19. Katnapally N, Chinta PCR, Routhu KK, Velaga V, Bodepudi V, et al. (2021) Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. American Journal of Computing and Engineering 4: 35-51.
20. Karaka LM (2021) Optimising Product Enhancements Strategic Approaches to Managing Complexity. SSRN 5147875.
21. Chint PCR, Karaka LM (2010) Agentic AI and Reinforcement Learning: Towards More Autonomous and Adaptive AI Systems. Journal for Educators, Teachers and Trainers https://jett.labosfor.com/index.php/jett/article/view/2700.
22. Boppana SB, Moore CS, Bodepudi V, Jha KM, Maka SR, et al. (2021) AI and ML Applications In Big Data Analytics: Transforming ERP Security Models For Modern Enterprises. SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5118085.
23. Chinta PCR, Katnapally N, Ja K, Bodepudi V, Babu S, et al. (2022) Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Studies.
24. Chinta PCR (2022) Enhancing Supply Chain Efficiency and Performance Through ERP Optimisation Strategies. Journal of Artificial Intelligence & Cloud Computing 1: 10-47363.
25. Sadaram G, Sakuru M, Karaka LM, Reddy MS, Bodepudi V, et al. (2022) Internet of Things (IoT) Cybersecurity Enhancement through Artificial Intelligence: A Study on Intrusion Detection Systems. Universal Library of Engineering Technology.
26. Moore C (2023) AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnology Perceptions 19: 46-64.

27. Chinta PCR (2023) The Art of Business Analysis in Information Management Projects: Best Practices and Insights. SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5103197.
28. Chinta PCR (2023) Leveraging Machine Learning Techniques for Predictive Analysis in Merger and Acquisition (M&A). Journal of Artificial Intelligence and Big Data 3: 10-31586.
29. Krishna Madhav J, Varun B, Niharika K, Srinivasa Rao M, Laxmana Murthy K (2023) Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI-104.
30. Maka SR (2023) Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. SSRN 5116707.
31. Routhu KishanKumar, Katnapally Niharika, Sakuru Manikanth (2023) Machine Learning for Cyber Defense: A Comparative Analysis of Supervised and Unsupervised Learning Approaches. Journal for ReAttach Therapy and Developmental Diversities https://jrtdd.com/index.php/journal/article/view/3481.
32. Chinta Purna Chandra Rao, Moore Chethan Sriharsha (2023) Cloud-Based AI and Big Data Analytics for Real-Time Business Decision-Making 36: 96-123.
33. Krishna Madhav J, Varun B, Niharika K, Srinivasa Rao M, Laxmana Murthy K (2023) Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI-104.
34. Bodepudi V (2023) Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. Journal of Artificial Intelligence and Big Data 3: 10-31586.
35. Jha KM, Bodepudi V, Boppana SB, Katnapally N, Maka SR, et al. & Sakuru, M. Deep Learning-Enabled Big Data Analytics for Cybersecurity Threat Detection in ERP Ecosystems. ResearchGate https://www.researchgate.net/publication/389397103_Deep_Learning-Enabled_Big_Data_Analytics_for_Cybersecurity_Threat_Detection_in_ERP_Ecosystems_1.
36. Kuraku S, Kalla D, Samaah F, Smith N (2023) Cultivating proactive cybersecurity culture among IT professional to combat evolving threats. International Journal of Electrical, Electronics and Computers 8.
37. Kalla D, Smith N, Samaah F, Polimetla K (2022) Enhancing Early Diagnosis: Machine Learning Applications in Diabetes Prediction. Journal of Artificial Intelligence & Cloud Computing 2-7.
38. Kuraku DS, Kalla D (2023) Impact of phishing on users with different online browsing hours and spending habits. International Journal of Advanced Research in Computer and Communication Engineering 12.
39. Kalla D, Kuraku S (2023) Phishing website url's detection using nlp and machine learning techniques. Journal of Artificial Intelligence 5: 145.
40. Kuraku DS, Kalla D, Samaah F (2022) Navigating the link between internet user attitudes and cybersecurity awareness in the era of phishing challenges. International Advanced Research Journal in Science, Engineering and Technology 9.
41. Kuraku DS, Kalla D, Smith N, Samaah F (2023) Exploring How User Behavior Shapes Cybersecurity Awareness in the Face of Phishing Attacks. International Journal of Computer Trends and Technology 71.