E-ISSN: 2708-3977
P-ISSN: 2708-3969
IJEDC 2024; 5(1): 07-11

www.datacomjournal.com
Received: 07-11-2023
Accepted: 13-12-2023

Sarthak Rout
B.E. CSE AIML, Chandigarh University, Mohali, Punjab, India

Khyati Jaiswal
B.E. CSE AIML, Chandigarh University, Mohali, Punjab, India

Corresponding Author:
Sarthak Rout
B.E. CSE AIML, Chandigarh University, Mohali, Punjab, India

Fraud detection using deep learning

Sarthak Rout and Khyati Jaiswal

DOI: https://doi.org/10.22271/27083969.2024.v5.i1a.37

Abstract
Fraud detection is a critical aspect of various industries, such as finance, e-commerce, and insurance, to safeguard against fraudulent activities. Machine learning (ML) techniques have emerged as powerful tools for fraud detection, enabling the identification of patterns and anomalies that indicate fraudulent behavior. This paper explores two distinct approaches to fraud detection using ML: classical machine learning and neural networks. The classical machine learning approach utilizes K-means clustering to group similar transactions and three types of logistic regression models to predict the probability of a transaction being fraudulent. The neural network approach employs a simple neural network, Gaussian noise addition, and oversampling, scaling, and PCA to enhance model performance. A general outline for fraud detection model building is proposed, encompassing data preprocessing, feature engineering, data splitting, model selection, model training, model evaluation, imbalanced data handling, ensemble methods, threshold optimization, monitoring and updating, explainability and interpretability, and compliance and security.

Keywords: Fraud Detection, machine learning, neural networks, logistic regression, K-means clustering, PCA

Introduction
This is the start of the body text of your paper. In today's increasingly digital world, fraud has become a pervasive issue, causing significant financial losses and reputational damage to businesses and organizations across various industries. The ability to effectively detect and prevent fraud is crucial for maintaining the integrity and security of financial transactions, e-commerce platforms, and insurance systems. Traditional fraud detection methods, often relying on rule-based systems and manual reviews, have limitations in adapting to evolving fraud patterns and handling large volumes of complex data. In recent years, deep learning, a subset of artificial intelligence (AI), has emerged as a powerful tool for fraud detection due to its ability to learn intricate patterns and identify anomalies from vast amounts of data.
Deep learning algorithms, particularly neural networks, have demonstrated superior performance in various fraud detection tasks. These algorithms can extract meaningful features from complex data, including transaction details, customer profiles, and historical behavior patterns, to identify patterns that are indicative of fraudulent activities. Additionally, deep learning models can continuously adapt to new fraud patterns and emerging threats, providing a proactive approach to fraud prevention.

The application of deep learning in fraud detection has gained significant traction in various industries, including
Financial Institutions: Fraudulent transactions in the financial sector, such as credit card fraud and insurance fraud, can lead to substantial financial losses. Deep learning models are being employed to analyze vast amounts of transaction data to identify anomalies and patterns associated with fraudulent activities. E-commerce Platforms: E-commerce platforms are susceptible to various types of fraud, including account takeovers, fake reviews, and fraudulent purchases. Deep learning models are being used to analyze customer behavior, transaction patterns, and device information to identify suspicious activities and prevent fraudulent transactions.

Insurance Companies: Insurance fraud, such as false claims and exaggerated losses, poses a significant challenge to insurance companies. Deep learning models are being applied to analyze claims data, medical records, and customer behavior to detect fraudulent patterns and prevent financial losses

The adoption of deep learning for fraud detection has brought several transformative benefits:

- Enhanced Fraud Detection Accuracy: Deep learning models can extract hidden patterns and anomalies from complex data, leading to improved accuracy in identifying fraudulent activities compared to traditional methods.
- Adaptability to Evolving Fraud Patterns: Deep learning models can continuously learn and adapt to new fraud patterns, ensuring that fraud detection systems remain effective against emerging threats.
- Scalability for Large Data Volumes: Deep learning models can effectively handle large volumes of complex data, making them suitable for fraud detection in large-scale operations.
- Reduced Reliance on Manual Reviews: Deep learning models can automate fraud detection tasks, reducing the need for manual reviews and improving operational efficiency.

Related Work

Table 1: Literature survey

Paper Title	Year	Author(s)	Key Findings
A novel deep learning-based fraud detection model for imbalanced financial datasets	2021	R. A. Al-Qaheri and H. El-Khozondar	Proposed a novel deep learning-based fraud detection model for imbalanced financial datasets.
A survey on deep learning for cyber security: Applications, challenges and future directions	2020	A. Ahmed, M. Marjani, and M. Imran	Provided a comprehensive survey of deep learning applications in cybersecurity, including fraud detection.
Deep learning for fraud detection in retail transactions	2018	A. Khodayari	Discussed the use of deep learning for fraud detection in retail transactions.
Fraud detection in e-commerce using deep learning: A review	2018	S. Dua and M. Islam	Provided a comprehensive review of fraud detection in e-commerce using deep learning.
A deep learning approach to network intrusion detection	2019	I. Shen and J.-H. Chou	Proposed a deep learning approach to network intrusion detection. The proposed approach is based on a convolutional neural network (CNN) that is able to extract features from network traffic data and classify it as normal or intrusive.

Proposed System

The proposed system utilizes deep learning to effectively identify fraudulent activities. It collects and preprocesses data, trains and evaluates a deep learning model, and deploys the model for real-time fraud detection.

- Data Collection and Preprocessing- A comprehensive dataset of transactions, customer information, and historical fraud cases is collected. Data cleaning and preprocessing are performed to ensure quality and consistency. Feature engineering techniques extract meaningful features.
- Model Training and Evaluation- An appropriate deep learning algorithm is selected based on the data characteristics. The model is trained on the preprocessed data to minimize classification errors. Its performance is evaluated using a validation dataset.
- Fraud Detection and Monitoring- The trained model is deployed to classify new transactions. Real-time detection prevents financial losses and protects customers. Continuous monitoring ensures the model's effectiveness over time.
- Additional Considerations- Imbalanced data is addressed using oversampling or under sampling. Explainability and interpretability are considered for transparency and compliance. Data privacy regulations and security measures are adhered to.

The proposed system offers a comprehensive approach to safeguard against fraudulent activities. It effectively identifies anomalies, adapts to evolving fraud patterns, and provides real-time protection.

Methodology

We thought of two different approached when working with our model of fraud detection. The first one involved a very classic machine learning approach that implemented the K-means algorithm along with three different types of logistic regressions, namely, vanilla logistic regression, logistic regression with SMOTE over sampling and finally the logistic regression with balanced class weights. This is further explained in detail.

The other approach implemented the usage of Neural Networks, something we're actually after. This is important in terms of emerging technologies and moreover to consider better optimized models that are capable of handling real world datasets in comparatively much lesser time. Out of the plethora of neural networks available, we use three different customized networks. In order, we used the simplest neural network that involved ReLU and sigmoid activation functions with only 31 neurons. The second one involved adding Gaussian noise deliberately to the data in order to reduce the risk of overfitting and increase the chances of generalization. The third approach concerned itself with the incorporation of algorithms like Oversampling, Scaling and Principle Component Analysis (PCA) for 10 components to extract the important features of the data to not only increase the accuracy of the model but also reduce the time it takes to perform its tasks.

Following are the diagrammatic representations of all three approaches

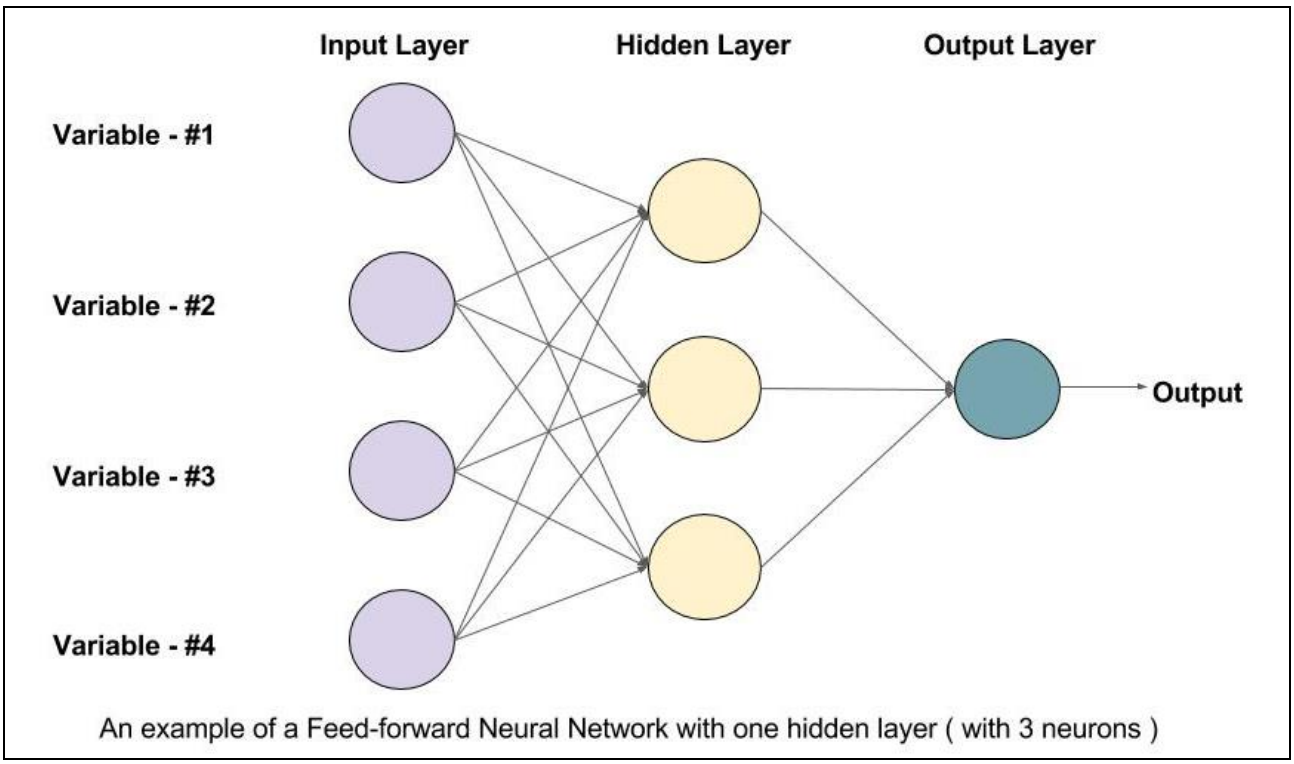

Image 1: Feed forward neural network

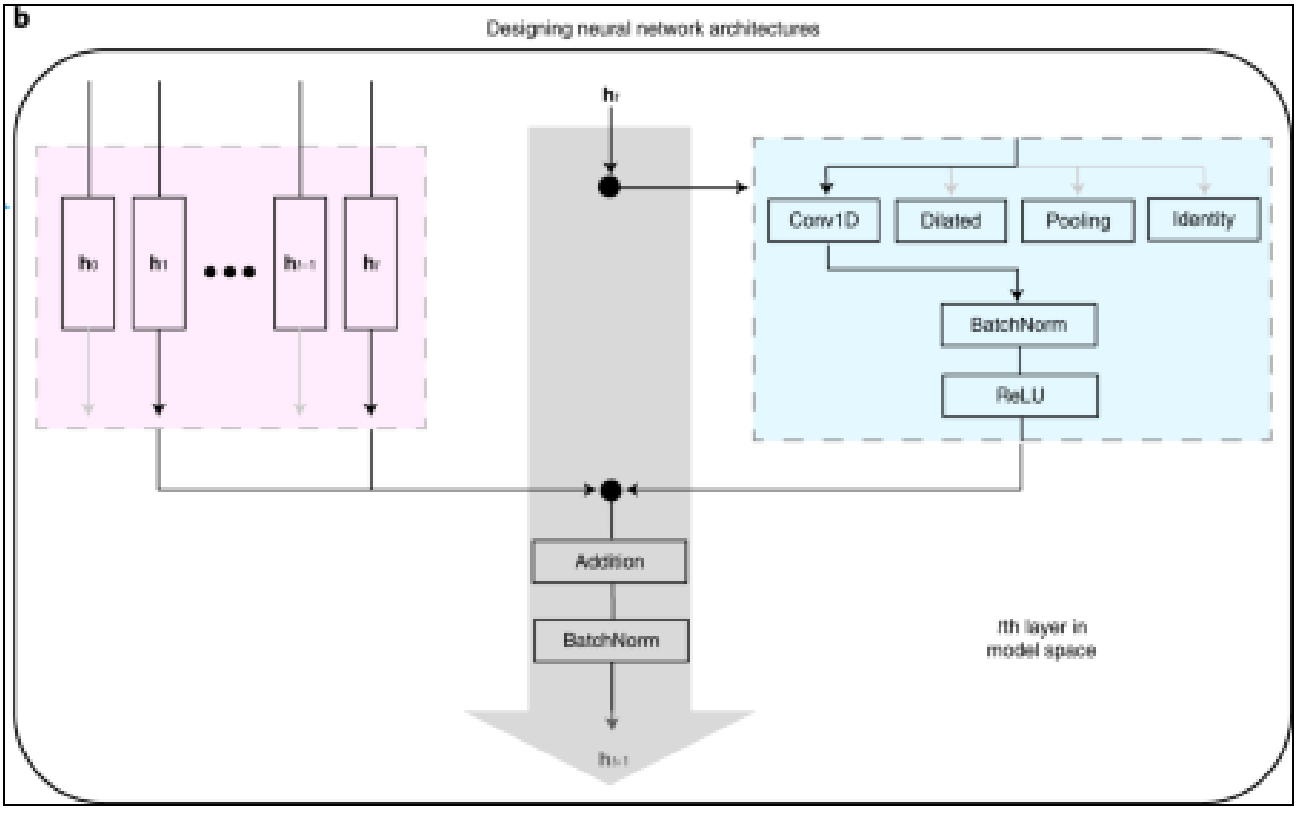

Image 2: Neural network architecture

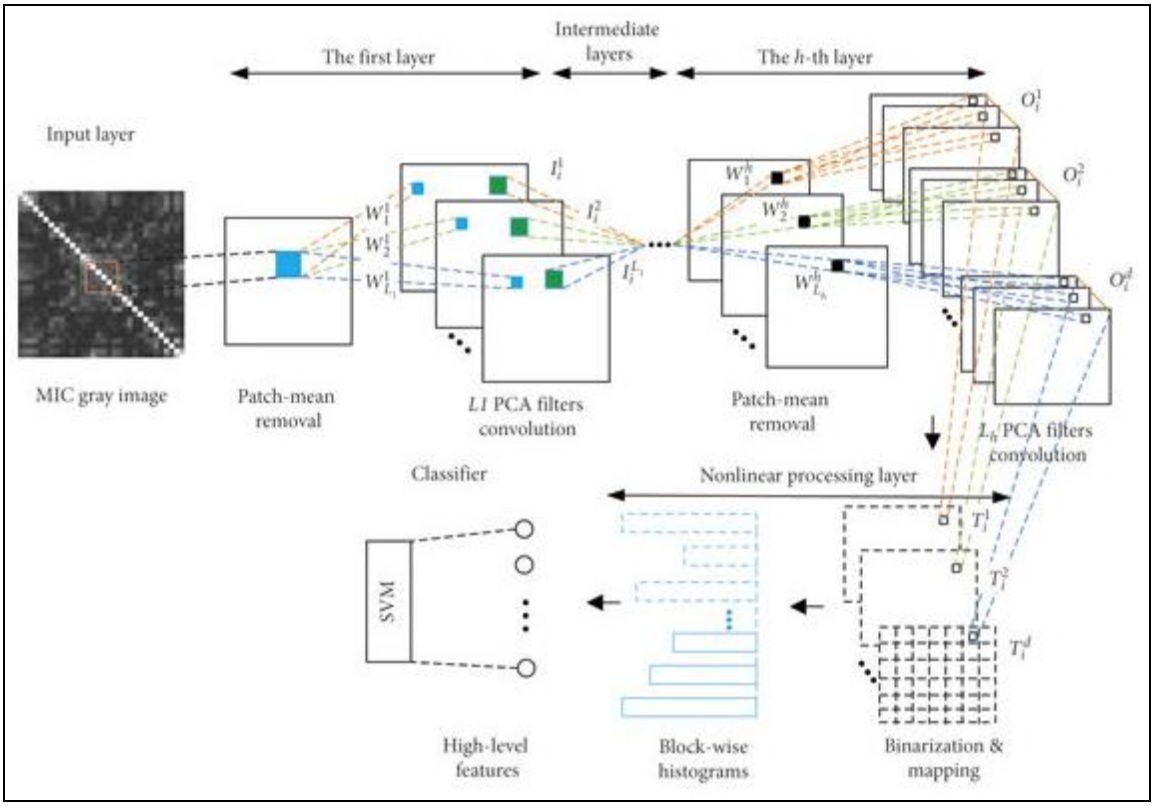

Image 3: Complete Classification Using PCA

The following is a proposed general outline for the building of the fraud detection model

- Data Preprocessing
- Feature Engineering
- Data Splitting
- Model Selection
- Model Training
- Model Evaluation
- Handling Imbalanced Data
- Ensemble Methods
- Threshold Optimization
- Monitoring and Updating
- Explainability and Interpretability
- Compliance and Security

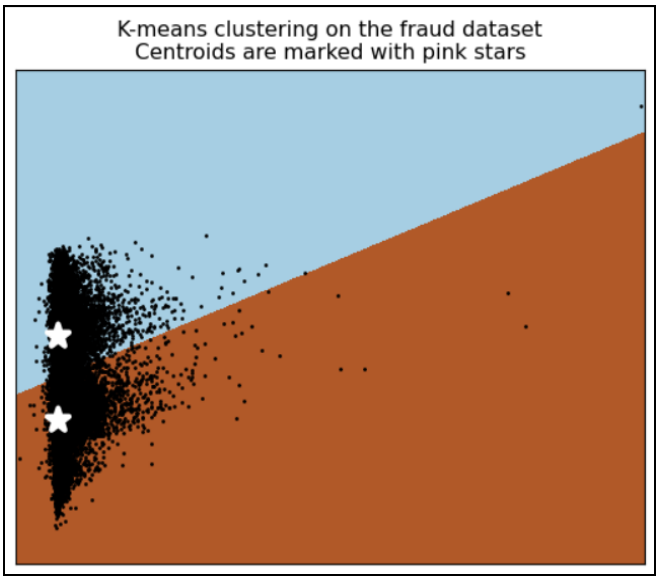

Image 4: K-means clustering

Conclusion

The proposed methodology for fraud detection using deep learning provides a robust and adaptable approach to identifying fraudulent activities in various domains. By leveraging the power of deep learning, the system can effectively extract meaningful patterns from complex data, adapt to evolving fraud patterns, and provide real-time protection against fraudulent transactions. The methodology encompasses data preprocessing, feature engineering, model training, model evaluation, imbalanced data handling, ensemble methods, threshold optimization, monitoring and updating, explainability and interpretability, and compliance and security. The implementation of this methodology can significantly enhance fraud detection capabilities and safeguard organizations from financial losses and reputational damage.

References

1. Ahmed A, Marjani M, Imran M. [Title not available]. IEEE Xplore. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9565320&casa_token=3B1aBAFFpQ0AAAAA:aAbnOiBUh3wECu5xoKARTRsCEph-Rc42hyeAUqUrSv9ulyPKahNpDxm_5YjYOzGXJACDI-Qib00cfw&tag=1
2. Khodayari A. [Title not available]. IEEE Xplore. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9755930
3. Chopra Y, Kaushik P, Rathore SPS, Kaur P. Uncovering Semantic Inconsistencies and Deceptive Language in False News Using Deep Learning and NLP Techniques for Effective Management. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):681-

692. https://doi.org/10.17762/ijritcc.v11i8s.7256
4. Shen I, Chou J-H. [Title not available]. ResearchGate. Available from: https://www.researchgate.net/profile/Pradheepan-Raghavan/publication/339411416
5. Kaushik P. Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation. International Journal for Global Academic & Scientific Research. 2022;1(3):01–11. https://doi.org/10.55938/ijgasr.v1i3.15
6. Kaushik P. Deep Learning Unveils Hidden Insights: Advancing Brain Tumor Diagnosis. International Journal for Global Academic & Scientific Research. 2023;2(2):01–22. https://doi.org/10.55938/ijgasr.v2i2.45
7. Kaushik P. Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT. International Journal for Global Academic & Scientific Research. 2023;2(2):23–45. https://doi.org/10.55938/ijgasr.v2i2.46
8. Kaushik P, Rathore SPS. Deep Learning Multi-Agent Model for Phishing Cyber-attack Detection. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(9s):680–686. https://doi.org/10.17762/ijritcc.v11i9s.7674
9. Kaushik P, Miglani S, Shandilya I, Singh A, Saini D, Singh A. HR Functions Productivity Boost by using AI. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):701–713. https://doi.org/10.17762/ijritcc.v11i8s.7672
10. Kaushik P, Singh Rathore SPS, Kaur P, Kumar H, Tyagi N. Leveraging Multiscale Adaptive Object Detection and Contrastive Feature Learning for Customer Behavior Analysis in Retail Settings. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(6s):326–343. https://doi.org/10.17762/ijritcc.v11i6s.6938
11. Kaushik P, Yadav R. Reliability design protocol and block chain locating technique for mobile agent. Journal of Advances in Science and Technology (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
12. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
13. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE
14. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
15. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
16. Pratap Singh Rathore S. Analysing the efficacy of training strategies in enhancing productivity and advancement in profession: theoretical analysis in Indian context. International Journal for Global Academic & Scientific Research. 2023;2(2):56-77. https://doi.org/10.55938/ijgasr.v2i2.49
17. Pratap Singh Rathore S. The Impact of AI on Recruitment and Selection Processes: Analysing the role of AI in automating and enhancing recruitment and selection procedures. International Journal for Global Academic & Scientific Research. 2023;2(2):78-93. https://doi.org/10.55938/ijgasr.v2i2.50
18. Al-Qaheri RA, El-Khozondar H. [Title not available]. Available from: https://pdfs.semanticscholar.org/01be/7624aa0e0251182593350a984411c2e5128a.pdf
19. Rachna Rathore. Application of Assignment Problem and Traffic Intensity in Minimization of Traffic Congestion. IJRST. 2021;11(3):25-34. DOI: http://doi.org/10.37648/ijrst.v11i03.003
20. Rathore R. A Review on Study of application of queueing models in Hospital sector. International Journal for Global Academic & Scientific Research. 2022;1(2):01–05. https://doi.org/10.55938/ijgasr.v1i2.11
21. Rathore R. A Study on Application of Stochastic Queuing Models for Control of Congestion and Crowding. International Journal for Global Academic & Scientific Research. 2022;1(1):01–07. https://doi.org/10.55938/ijgasr.v1i1.6
22. Rathore R. A Study Of Bed Occupancy Management In The Healthcare System Using The M/M/C Queue And Probability. International Journal for Global Academic & Scientific Research. 2023;2(1):01–09. https://doi.org/10.55938/ijgasr.v2i1.36
23. Dua S, Islam M. [Title not available]. Available from: https://web.archive.org/web/20210624021145id_/https://irojournals.com/aicn/V3/I2/03.pdf
24. Sharma T, Kaushik P. Leveraging Sentiment Analysis for Twitter Data to Uncover User Opinions and Emotions. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):162–169. https://doi.org/10.17762/ijritcc.v11i8s.7186
25. Sharma V. A Study on Data Scaling Methods for Machine Learning. International Journal for Global Academic & Scientific Research. 2022;1(1):23–33. https://doi.org/10.55938/ijgasr.v1i1.4
26. Yadav M, Kakkar M, Kaushik P. Harnessing Artificial Intelligence to Empower HR Processes and Drive Enhanced Efficiency in the Workplace to Boost Productivity. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):381–390. https://doi.org/10.17762/ijritcc.v11i8s.7218