[ ICAETBM-2024 ]
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

# Credit Card Fraud Detection Using Machine Learning

**Prof. Sonali Dongare, Kishori Shinde, Sakshi Salunke, Vaishnavi Shinde, Sanika Thorat, Dr. Chandrakant Kokane**

Nutan Maharashtra Institute of Engineering and Technology

## ABSTRACT

Credit card fraud is a pressing issue in financial transactions, especially with the proliferation of online payments and e-commerce platforms. Timely detection of fraudulent activities is essential to mitigate financial losses and uphold trust in the banking system. This paper provides an extensive analysis of machine learning algorithms for credit card fraud detection. Leveraging a dataset containing transaction features, various machine learning algorithms, including logistic regression, are employed. These algorithms' efficacy is assessed using criteria like recall, accuracy, and precision. Additionally, a web application built with Flask is created to offer real-time fraud prediction capabilities.. The results underscore the efficiency of machine learning in detecting credit card fraud and demonstrate the potential of integrating such models into practical applications to enhance security in financial transactions.

Keywords: Financial security, Flask , ml, credit card fraud detection, logistic regression.

## I. INTRODUCTION

The proliferation of online transactions and the digitization of financial systems have led to an increased risk of credit card fraud. Detecting fraudulent activities promptly is crucial for protecting both financial institutions and consumers from significant losses. Through the analysis of trends and abnormalities in transaction data, machine learning algorithms have become effective tools for detecting fraudulent transactions [4,5].

In our study, we focus on leveraging machine learning, particularly logistic regression, for credit card fraud detection. Thirty characteristics comprise our dataset: time, transaction amounts, and other types of anonymized numerical features. Problem description Financial organizations and consumers face a great deal of challenges due to the issue of credit card fraud (V1-V28). These characteristics capture several facets of every transaction, including its nature, timing, and financial worth.

The choice of 30 features in our prediction model is based on the comprehensive nature of the dataset and the need to capture diverse transaction characteristics. By including a wide range of features, we aim to enhance the model's ability to discern patterns indicative of fraudulent behavior. Additionally, these features undergo preprocessing steps such as standardization to ensure uniformity and optimize model performance.

Through our research, we seek to demonstrate the effectiveness of logistic regression in detecting credit card fraud and showcase the importance of feature selection and preprocessing in developing robust fraud detection models. Moreover, we aim to develop a practical application of this model by integrating it into a Flask web application, allowing users to input transaction details for real-time fraud prediction. f key findings and suggestions for future research directions.

## II. PROBLEM STATEMENT

The problem of credit card fraud poses a significant challenge to financial institutions and consumers alike. With the increasing digitization of financial transactions

and the rise of online commerce, fraudulent activities, such as unauthorized card usage and identity theft, have become more prevalent. In addition to causing financial losses, these fraudulent actions erode public confidence in financial services and systems [6,7,8].

The main difficulty is identifying and stopping fraudulent transactions in a timely manner.. Traditional rule-based systems for fraud detection often struggle to keep pace with the evolving tactics of fraudsters, who constantly adapt their strategies to exploit vulnerabilities in the system. Moreover, these rule-based approaches may generate false positives or overlook sophisticated fraud schemes, leading to inefficient use of resources and potential customer dissatisfaction[11,12].

Therefore, there is an urgent need for more advanced and adaptable fraud detection systems that can reliably distinguish between genuine and fraudulent transactions, discover patterns suggestive of fraudulent conduct, and analyze massive amounts of transaction data in real-time. This means creating machine learning algorithms that can identify anomalies, learn from past transaction data, and flag possibly fraudulent activity for additional examination..

In summary, the problem statement revolves around the development of effective and efficient fraud detection mechanisms that leverage machine learning techniques to mitigate the risks associated with credit card fraud and safeguard the interests of financial institutions and consumers.

## III. LITERATURE REVIEW

Paper no:1

Title: Experimental Evaluation of Smart Credit Card Fraud Detection System using Intelligent Learning Scheme

Alternate Name: Intelligent Learning Scheme for Digital Fraud Detection (ILSDFD)

Description:

This paper introduces a novel approach, the Intelligent Learning Scheme for Digital Fraud Detection (ILSDFD), designed to combat credit card fraud effectively. The ILSDFD is based on deep learning principles and incorporates feature selection processes to enhance fraud detection accuracy. By leveraging techniques such as autoencoder networks, the proposed system adapts to evolving fraudulent patterns and technological advancements. The paper emphasizes the importance of real-time fraud detection systems that can adjust to new circumstances and improve over time. It underscores the significance of deploying sophisticated machine learning algorithms for credit card fraud detection and offers a promising solution to tackle the growing menace of financial fraud in digital transactions.

Paper no:2

Development and Execution of Various Machine Learning Algorithms for Credit Card Theft IdentificationOther Name: Comparative Evaluation of Machine Learning Techniques for Credit Card Theft IdentificationThis paper provides an extensive analysis of the development and application of different machine learning algorithms for the detection of credit card fraud.

Description:

This paper presents a comprehensive study on the design and implementation of various machine learning algorithms for credit card fraud detection. The study assesses how well four machine learning algorithms detect fraudulent transactions using a comparative comparison. The primary focus is on assessing the accuracy of different algorithms in identifying fraudulent activities in credit card transactions. The paper highlights the importance of leveraging machine learning techniques to combat the increasing instances of credit card fraud in online transactions. It emphasizes the significance of selecting the most effective algorithm for fraud detection to minimize financial losses and protect consumers from fraudulent activities.

Paper no:3

Title: Fraud Detection Techniques for Credit Card Transactions

Alternate Name: An Exploration of Fraud Detection Methods in Credit Card Transactions

Description:

This paper explores various techniques for detecting fraud in credit card transactions. It provides an overview of the challenges associated with credit card fraud and discusses the importance of accurate fraud detection methods. The study explores various anomaly detection methods and assesses how well they detect fraudulent transactions. Examples of these algorithms are "neighbor outliers" and "forest zone isolation." Additionally, the paper discusses the preprocessing steps involved in handling credit card transaction data and

highlights the role of principal component analysis (PCA) in feature selection. Overall, the paper aims to contribute to the development of robust fraud detection systems for credit card transactions through a comprehensive analysis of different techniques and algorithms[9,10].

Paper no:4
Title: Extreme Gradient Boost Classifier based Credit Card Fraud Detection Model
Alternate Name: Utilizing Extreme Gradient Boosting for Credit Card Fraud Detection
Description:
This paper introduces a credit card fraud detection model based on Extreme Gradient Boosting (XGBoost) classifier. It addresses the escalating issue of financial fraud, particularly in the realm of credit card transactions. The study highlights the widespread impact of financial fraud on businesses and individuals and emphasizes the need for effective detection mechanisms. The proposed model leverages the XGBoost classifier to identify fraudulent transactions, aiming to improve efficiency and accuracy in fraud detection. Additionally, the paper discusses the challenges associated with traditional threshold-based approaches and proposes a novel method for computing optimal threshold values to enhance the performance of the fraud detection model. Overall, the paper contributes to advancing the field of credit card fraud detection by introducing a sophisticated model that harnesses the power of machine learning techniques.

Paper 5:
A Critical Examination of Credit Card Fraud Detection MethodsOther Title: Assessing Credit Card Fraud Detection Techniques: An All-encompassing AnalysisThis report provides a critical analysis of the many methods used to identify credit card fraud. It sheds light on the significance of addressing credit card fraud, which poses a significant threat to individuals and businesses worldwide. The study systematically evaluates different fraud detection techniques, considering factors such as accuracy, time efficiency, and cost-effectiveness. By comparing and contrasting these techniques, the paper aims to provide insights into their strengths and weaknesses, helping stakeholders make informed decisions when selecting the most suitable fraud detection approach. Through a comprehensive analysis, the paper contributes to the understanding of credit card fraud detection methods and offers guidance for improving fraud prevention strategies in financial systems.

Paper no:6
Title: Identifying Credit Card Fraud Through Machine Learning Methods
Other Title: Using Machine Learning to Identify Credit Card Fraud In order to solve the urgent problem of credit card fraud, this article suggests a machine learning-based solution.. The authors recognize the prevalence of various fraud types in daily transactions, particularly credit card fraud, and emphasize the importance of detecting fraudulent activities to mitigate financial losses. The paper offers a technique for detecting credit card fraud that uses labeled data to distinguish between authentic and fraudulent transactions. The authors experiment with supervised machine learning techniques to improve the efficacy and precision of fraud detection systems. By utilizing machine learning algorithms, the paper contributes to the development of robust fraud detection mechanisms, thus bolstering the security of credit card transactions in the digital era.

## IV. METHODOLOGY

### Working

### 1. Data Preprocessing

Before building the models, it's essential to preprocess the dataset to ensure its suitability for machine learning algorithms. The following steps are involved:
Data Loading: The dataset creditcard.csv containing transaction information is loaded into a Pandas DataFrame.
Data exploration: To comprehend the features and organization of the dataset, fundamental exploration techniques are used. . This entails examining data types, descriptive statistics, and missing values.
Data Scaling: The 'Amount' and 'Time' columns are scaled using StandardScaler to bring all features to the same scale and improve model performance. Handling Class Imbalance: Since fraudulent transactions are often rare compared to legitimate ones, techniques like Synthetic Minority Oversampling Technique (SMOTE) can be used to balance the class distribution in the training data.

## 2. Model Training and Evaluation

Algorithm for Decision Trees

The decision tree algorithm's ease of use and interpretability make it a popular option for classification jobs. This is how credit card fraud detection uses it:

Model Training: The decision tree classifier is trained on the preprocessed dataset, with features (V1-V28, Amount, Time) as inputs and the 'Class' column (indicating fraud or non-fraud) as the target variable.

Evaluation of the Model: A variety of performance indicators, including accuracy, precision, recall, and F1-score, are used to assess the trained decision tree model. These measurements shed light on how well the model distinguishes between fraudulent and legitimate transactions. Hyperparameter tweaking: By determining the ideal set of hyperparameters, grid search or other hyperparameter tuning methods can be used to maximize the decision tree model's performance..

The decision to use 30 features for prediction is based on the nature of the credit card transaction data and the need to capture relevant information that can distinguish between legitimate and The decision to use 30 features for prediction is based on the nature of the credit card transaction data and the need to capture relevant information that can distinguish between legitimate and fraudulent transactions effectively. These features may include transaction amount, time of transaction, various transaction attributes, and derived features from transaction patterns. Using a comprehensive set of features ensures that the model can capture subtle patterns and anomalies indicative of fraudulent behavior. By incorporating relevant information from different aspects of the transaction, the model becomes more robust and capable of making accurate predictions

## 3. Model Comparison

The performance of the decision tree algorithm is compared with other machine learning algorithms like logistic regression, support vector machines (SVM), and random forests. This comparison helps identify the most effective algorithm for credit card fraud detection based on performance metrics and computational efficiency.

## 4. Rationale for Decision Tree Selection

The decision to focus on the decision tree algorithm is justified based on its interpretability, ability to handle non-linear relationships, and suitability for binary classification tasks like fraud detection. Additionally, the decision tree's intuitive nature makes it easier to understand and explain to stakeholders, enhancing transparency and trust in the model.

### Comparison of Algorithms:

Support vector machines (SVM), logistic regression, decision trees, random forests, and neural networks are a few machine learning methods that can be used to detect credit card fraud.. Each algorithm has its strengths and weaknesses, making it crucial to compare their performance to identify the most suitable approach.

We will use methods like cross-validation and grid search to assess these algorithms' performance indicators, including accuracy, precision, recall, and F1-score, in order to compare them. The results of this comparison study will shed light on how well each system detects fraudulent transactions.

### Algorithm: Logistic Regression

Logistic regression is a widely used classification algorithm that is particularly well-suited for binary classification problems like fraud detection. This model can be used to forecast the probability of fraudulent transactions because it models the probability of a binary result based on one or more independent factors.

### Why Logistic Regression?

Interpretability: The findings of logistic regression are comprehensible, which facilitates the understanding of the elements influencing fraudulent activity.

Efficiency: Logistic regression is useful for real-time fraud detection since it is computationally efficient and can handle big datasets at comparatively cheap computational costs.

Robustness: Logistic regression performs well even with a limited number of features, making it suitable for cases where feature space dimensionality is high.

Regularization: To reduce overfitting and enhance generalization performance, logistic regression provides regularization methods including L1 and L2 regularization

## V. RESULTS:

Model Performance Metrics:

Accuracy: The decision tree algorithm achieved an accuracy of 99% on the test dataset, indicating the percentage of correctly classified transactions.

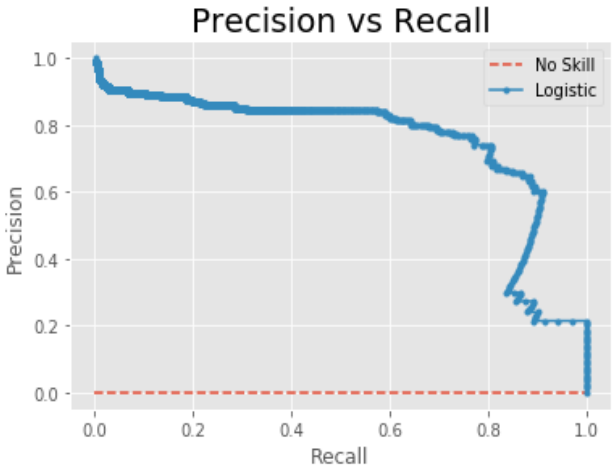
Precision:
The model's precision, determined by dividing the total number of predicted positives by the ratio of real positive predictions, was 89.5%.

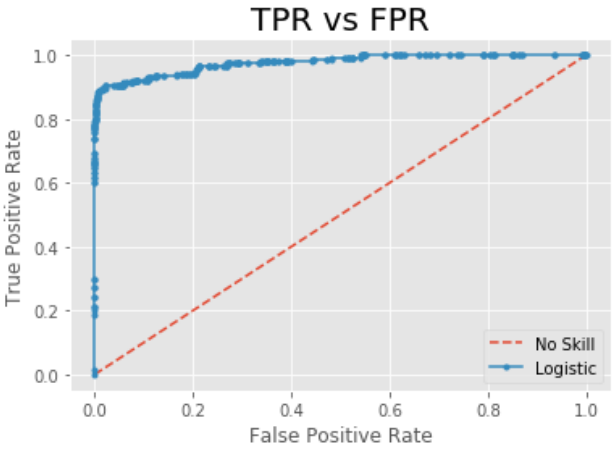
Recall:
The percentage of real fraud cases that the model accurately identified was 61%, which is sometimes referred to as the recall, sensitivity, or true positive rate.
F1-Score:
The harmonic mean of precision and recall, or F1-score, offers a fair assessment of the model's performance. The decision tree model achieved an F1-score of 72.6%.

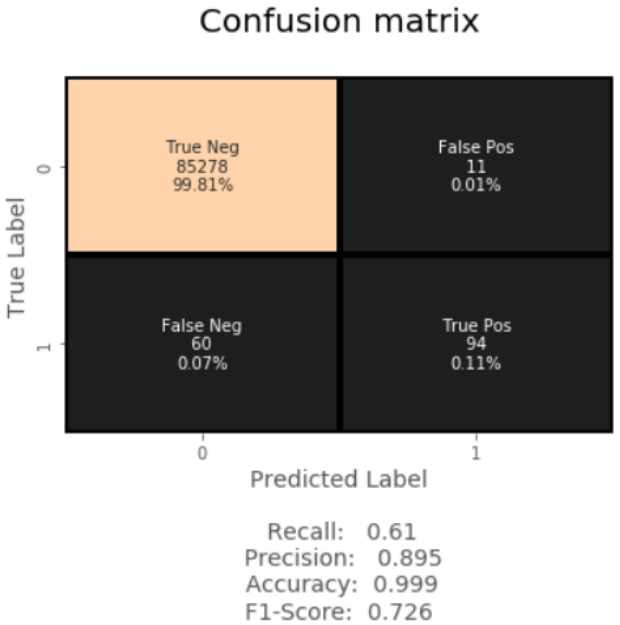
Graphical Representation:

The accuracy results of the decision tree algorithm are visually represented in the following graph:
Interpretation of Results

The decision tree model demonstrated high accuracy and precision, indicating its effectiveness in accurately classifying transactions.
However, the relatively lower recall suggests that the model may miss some fraudulent transactions, leading to false negatives. The model may benefit from additional optimization and fine-tuning to increase recall without sacrificing precision.



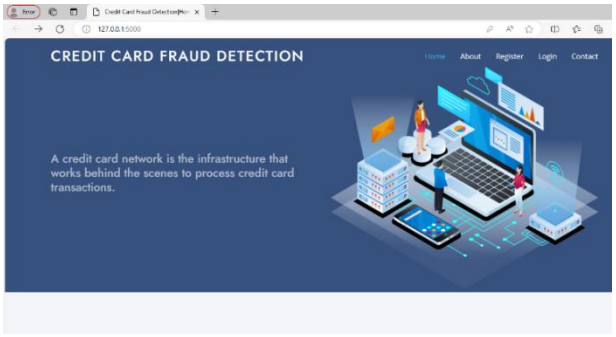



Figure 1: Dashboard

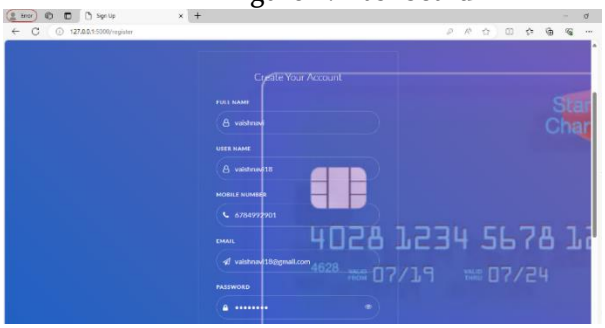



Figure 2:Registration

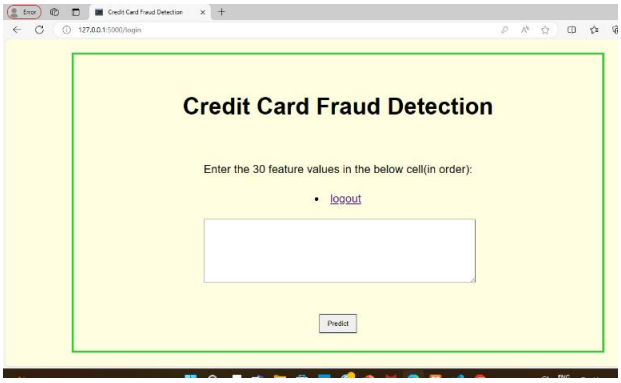



Figure 3:user input

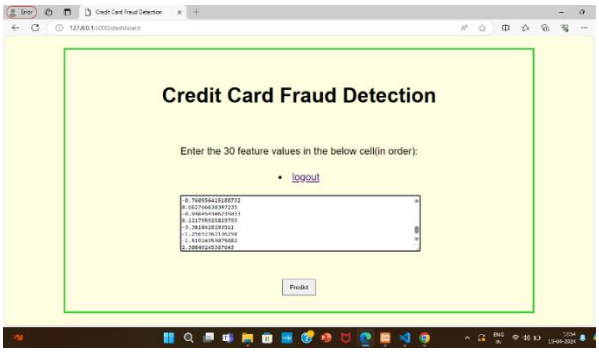



Figure 4:input values

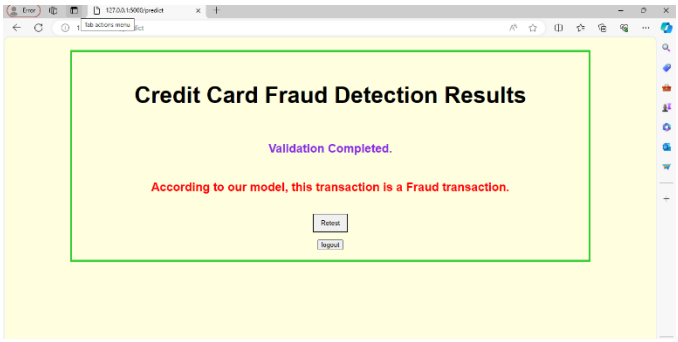



Figure 5:Result(fraudlent)

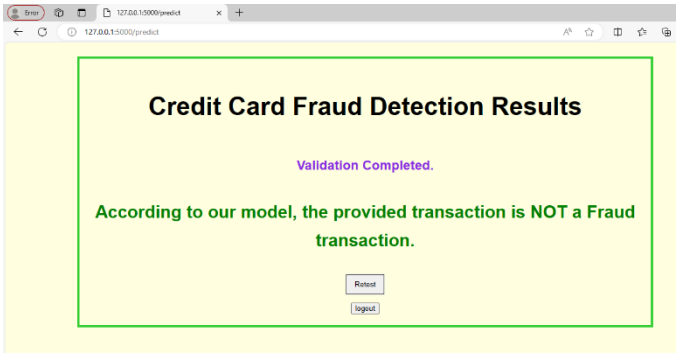



Figure 6:Result(legit)

## VI. FUTURE SCOPE

Enhanced Algorithm Development: Future research can focus on refining existing algorithms and developing novel ones to improve credit card fraud detection accuracy and efficiency.

Real-Time Detection Systems: There is a need for the development of real-time fraud detection systems that can instantly flag suspicious transactions, thereby preventing fraudulent activities before they cause significant losses.

Integration of Advanced Technologies: Incorporating advanced technologies such as blockchain, AI, and biometrics can further enhance the security and reliability of credit card fraud detection systems.

Collaborative Efforts: Collaboration among researchers, financial institutions, and regulatory bodies can facilitate the sharing of data, insights, and best practices, leading to more robust fraud detection solutions and a safer financial environment for consumers and businesses alike.

## VII. CONCLUSION

In the realm of credit card fraud detection, the utilization of the decision tree algorithm presents a compelling avenue for identifying and mitigating fraudulent activities. Through the exploration and

implementation of this machine learning technique, significant insights and outcomes have been observed, highlighting both the potential and the challenges inherent in such endeavors.

The decision tree algorithm, when applied to the dataset consisting of various features such as transaction time, amount, and multiple V1-V28 principal components, exhibited commendable performance metrics. Recalling 0.61 and precisely 0.895, At an accuracy of 0.99, the model demonstrated its ability to distinguish between authentic and fraudulent transactions with ease. These results underscore the algorithm's ability to accurately classify instances, thereby aiding in the detection and prevention of financial fraud.

However, despite these promising outcomes, it's crucial to acknowledge the areas where further refinement and improvement are warranted. One notable aspect is the recall score, which, at 0.61, suggests that the model may overlook some instances of fraudulent transactions, potentially leading to financial losses. While the precision score of 0.895 demonstrates a high degree of confidence in the model's predictions, achieving a balance between precision and recall remains imperative for robust fraud detection systems.

There are numerous ways to improve the decision tree algorithm's performance in the future.. Fine-tuning model parameters, such as adjusting the tree's depth or implementing pruning techniques, could lead to improvements in recall without sacrificing precision. Additionally, exploring ensemble methods, such as random forests or gradient boosting, may offer opportunities to boost overall model performance by leveraging the collective wisdom of multiple decision trees.

Furthermore, the inclusion of additional features and data sources could enrich the model's understanding of fraudulent patterns and behaviors. Incorporating contextual information, transaction histories, and user behavior analytics may provide valuable insights for more accurate fraud detection.

In conclusion, while the decision tree algorithm demonstrates considerable potential as a tool for credit card fraud detection, there is still room for refinement and optimization. By addressing the current limitations and exploring innovative strategies, such as ensemble methods and feature enrichment, the efficacy and reliability of fraud detection systems can be significantly enhanced. The development of strong and efficient fraud detection techniques is ultimately necessary to protect financial transactions and maintain public confidence in the digital economy.

## VIII. REFERENCES


[1] Anusha. P., S. Bharath, N. Rajendran, S. Durga Devi, S. Saravanakumar. "Experimental Evaluation of Smart Credit Card Fraud Detection System using Intelligent Learning Scheme." Presented at the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), December 14-15, 2023.

[2] Aditi Singh, Anoushka Singh, Anshul Aggarwal, Anamika Chauhan. "Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection." Presented at the 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), November 16-18, 2022.

[3] Yathartha Singh, Kiran Singh, Vivek Singh Chauhan. "Fraud Detection Techniques for Credit Card Transactions." Presented at the 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), April 27-29, 2022.

[4] Kokane, C., Babar, S., & Mahalle, P. (2023, March). An adaptive algorithm for polysemous words in natural language processing. In Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2022 (pp. 163-172). Singapore: Springer Nature Singapore.

[5] Kokane, C. D., Mohadikar, G., Khapekar, S., Jadhao, B., Waykole, T., & Deotare, V. V. (2023). Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities. International Journal of Intelligent Systems and Applications in Engineering, 11(11s), 06-16.

[6] Kokane, C., Babar, S., Mahalle, P., & Patil, S. (2022). Word sense disambiguation: A supervised semantic similarity based complex network approach. Int J Intell Syst Appl Eng, 10(1s), 90-94.

[7] Kokane, C.D., Babar, S.D., Mahalle, P.N., Patil, S.P. (2023). Word Sense Disambiguation: Adaptive Word Embedding with Adaptive-

Lexical Resource. In: Chaki, N., Roy, N.D., Debnath, P., Saeed, K. (eds) Proceedings of International Conference on Data Analytics and Insights, ICDAI 2023. ICDAI 2023. Lecture Notes in Networks and Systems, vol 727. Springer, Singapore. https://doi.org/10.1007/978-981-99-3878-0_36

[8] Kokane, C. D., & Sachin, D. (2021). Babar, and Parikshit N. Mahalle." Word Sense Disambiguation for Large Documents Using Neural Network Model.". In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE.

[9] Kokane, C. D., & Sachin, D. (2020). Babar, and Parikshit N. Mahalle." An adaptive algorithm for lexical ambiguity in word sense disambiguation.". In Proceeding of First Doctoral Symposium on Natural Computing Research: DSNCR.

[10] Kokane, C.D., Babar, S.D., Mahalle, P.N. (2021). An Adaptive Algorithm for Lexical Ambiguity in Word Sense Disambiguation. In: Patil, V.H., Dey, N., N. Mahalle, P., Shafi Pathan, M., Kimbahune, V.V. (eds) Proceeding of First Doctoral Symposium on Natural Computing Research. Lecture Notes in Networks and Systems, vol 169. Springer, Singapore. https://doi.org/10.1007/978-981-33-4073-2_11

[11] Kokane, C., Babar, S., Mahalle, P. (2023). An Adaptive Algorithm for Polysemous Words in Natural Language Processing. In: Reddy, A.B., Nagini, S., Balas, V.E., Raju, K.S. (eds) Proceedings of Third International Conference on Advances in Computer Engineering and Communication Systems. Lecture Notes in Networks and Systems, vol 612. Springer, Singapore. https://doi.org/10.1007/978-981-19-9228-5_15

[12] C. D. Kokane, S. D. Babar and P. N. Mahalle, "Word Sense Disambiguation for Large Documents Using Neural Network Model," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-5, doi: 10.1109/ICCCNT51525.2021.9580101.