

---

## ONLINE FRAUD PAYMENT DETECTION USING BALANCED ML ALGORITHMS

SK.AnjaneyuluBabu<sup>1</sup>, V.Praveen<sup>2</sup>  
Associate Professor<sup>1</sup>, PGScholar<sup>2</sup>

Department of Master of Computer Applications  
QIS College of Engineering & Technology (Autonomous),  
Ongole Prakasam Dist., AP

### ABSTRACT:

Online payment systems have become a crucial component of modern financial services, but they are increasingly vulnerable to fraudulent activities. Detecting fraudulent transactions is a challenging task due to the high class imbalance between genuine and fraudulent transactions. This project presents a machine learning-based approach to accurately detect online payment fraud by leveraging balanced algorithms to address the skewed distribution of data. Various resampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) are employed to balance the dataset. Multiple machine learning classifiers including Random Forest, Decision Tree, Logistic Regression, and XGBoost are trained and evaluated on the balanced data. The models are assessed using performance metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC-ROC) to ensure robustness. Experimental results demonstrate that balancing the data significantly improves the ability of the classifier to detect fraudulent activities. This study highlights the effectiveness of

combining balanced sampling strategies with machine learning algorithms to build reliable fraud detection systems in online payment environments.

### INTRODUCTION:

With the rapid growth of e-commerce and digital transactions, online payment systems have become an essential aspect of everyday life. However, the convenience of online payments has also led to a rise in fraudulent activities, posing significant financial and security risks to both consumers and businesses. Online fraud can take various forms, such as identity theft, phishing, and unauthorized transactions, making fraud detection a critical concern in the financial technology sector.

One of the major challenges in fraud detection is the inherent imbalance in the data: fraudulent transactions typically represent a very small percentage of total transactions. This class imbalance can lead to biased models that perform well on the majority class (legitimate transactions) but poorly on the minority class (fraudulent transactions). Traditional machine learning

algorithms often fail to detect fraud accurately due to this skewed distribution.

To address this issue, this project focuses on employing balanced machine learning algorithms to enhance fraud detection performance. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), and random under-sampling are applied to the dataset to balance the class distribution. These balanced datasets are then used to train various machine learning models, including Decision Trees, Random Forest, Logistic Regression, and XGBoost.

The primary goal of this project is to build a reliable and accurate fraud detection system that can effectively identify fraudulent transactions in real time, minimize false positives, and adapt to changing patterns in fraudulent behavior. The performance of each model is evaluated using metrics like precision, recall, F1-score, and ROC-AUC to ensure a comprehensive understanding of their effectiveness.

## LITERATURE SURVEY:

**Title:** *Credit Card Fraud Detection through Cost-Sensitive and Ensemble Learning*

**Author(s):** Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015)

**Description:**

This study addresses the issue of class imbalance in fraud detection using cost-sensitive learning and ensemble methods. The authors highlight that traditional accuracy metrics can be misleading and recommend precision-recall and ROC-AUC curves for better evaluation. They apply

random under-sampling and ensemble techniques to improve the detection of rare fraudulent transactions, achieving higher recall without significantly affecting the false positive rate.

**Title:** *Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy*

**Author(s):** Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Bontempi, G. (2018)

**Description:**

Carcillo et al. introduce a real-world credit card dataset and propose a novel learning strategy based on balanced random forests and ensemble methods. Their approach considers temporal dependencies and data drift, making the detection system more adaptive to evolving fraud patterns. The paper shows that combining oversampling and model ensembles significantly improves performance on imbalanced datasets.

**Title:** *Modeling and Simulation of Real-Time Credit Card Fraud Detection Using Machine Learning Techniques*

**Author(s):** Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011)

**Description:**

This paper provides a comparative study of machine learning classifiers such as Logistic Regression, Decision Trees, and Support Vector Machines for fraud detection. It emphasizes the importance of data preprocessing and class rebalancing. The study concludes that Decision Trees and ensemble models outperform others when used in conjunction with rebalancing techniques like under-sampling.

**Title:** *Recurrent Neural Network for Credit Card Fraud Detection*

**Author(s):** Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L., & Caelen, O. (2018)

**Description:**

Jurgovsky et al. explore the application of Recurrent Neural Networks (RNNs) for capturing temporal patterns in sequential transaction data. The model is trained on a large dataset of transaction sequences and shows strong performance. However, the high computational complexity and need for large, labeled datasets limit its applicability in real-time systems.

**Title:** *Detecting Credit Card Fraud by Using a Hybrid Method of Oversampling and Boosting*

**Author(s):** Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019)

**Description:**

The authors propose a hybrid model combining SMOTE for oversampling and AdaBoost for classification. This method effectively increases sensitivity to the minority fraud class while keeping the overall error rate low. Their results show that combining oversampling techniques with boosting algorithms provides a substantial improvement in fraud detection accuracy.

**Title:** *A Cost-Sensitive Decision Tree Approach for Credit Card Fraud Detection*

**Author(s):** Sahin, Y., & Duman, E. (2011)

**Description:**

This research introduces a cost-sensitive decision tree classifier tailored for credit card fraud detection. By assigning higher misclassification costs to fraudulent transactions, the model achieves improved

fraud detection performance. The study also illustrates the effectiveness of cost-based learning over traditional methods when dealing with imbalanced data.

## SYSTEM ANALYSIS

### EXISTING SYSTEM:

The existing fraud detection systems in online payment platforms primarily rely on rule-based approaches and traditional machine learning models. Rule-based systems flag transactions based on predefined conditions such as location mismatch, large transaction amount, or repeated failed login attempts. While these methods are simple and interpretable, they are rigid and struggle to adapt to evolving fraud patterns. Fraudsters can easily bypass static rules by slightly modifying their behavior, making these systems increasingly ineffective in detecting sophisticated fraud.

Traditional machine learning models like Logistic Regression, Decision Trees, and Naïve Bayes have been used to enhance fraud detection capabilities. These models learn patterns from historical data and make predictions on new transactions. However, they often fail to detect rare fraudulent events because the training data is highly imbalanced—fraudulent transactions make up a very small percentage of the overall data. As a result, models tend to be biased toward predicting the majority class, i.e., legitimate transactions.

To address class imbalance, some systems apply basic resampling techniques such as

random under-sampling or oversampling. Although these methods help to a certain extent, they have limitations. Under-sampling can lead to loss of valuable data from the majority class, while oversampling can cause overfitting due to duplication of minority class samples. Therefore, such techniques may not always yield reliable or generalizable results, especially when dealing with complex and high-volume transactional data.

In addition, most existing systems lack real-time processing capabilities. Many fraud detection models are batch-processed, where predictions are made after data collection and preprocessing are complete. This delay can result in missed opportunities to prevent fraudulent transactions before they are executed. In real-world applications, real-time detection is critical to prevent financial losses and protect customers' accounts effectively.

Moreover, some existing systems do not incorporate feedback loops or dynamic learning. This means they are not updated frequently to reflect new fraud patterns or adapt to changes in user behavior. Without periodic retraining or incorporating streaming data, the models become outdated and lose their predictive power over time. This static nature makes them vulnerable in the constantly changing landscape of cyber fraud.

Finally, many existing systems are black-box models that lack transparency. In industries like finance where explainability is important, models that cannot provide interpretable decisions face resistance from stakeholders and regulators. There is a

growing demand for models that not only perform well but also provide clear justifications for their predictions. This requirement is often unmet by many existing fraud detection systems.

### **Disadvantages of Existing Systems:**

#### **Ineffective Handling of Imbalanced Data:**

Most traditional fraud detection systems struggle with the extreme class imbalance between legitimate and fraudulent transactions. This often results in poor detection of fraud cases, as models tend to favor the majority class, leading to high false negatives.

#### **Rigid Rule-Based Logic:**

Rule-based systems depend on manually defined patterns and thresholds, which are not adaptive to evolving fraud techniques. Once fraudsters learn the rules, they can easily alter their behavior to bypass detection, making such systems ineffective in dynamic environments.

**Lack of Real-Time Detection:** Many existing systems operate on batch processing and do not support real-time fraud detection. This delay can allow fraudulent transactions to be completed before being flagged, resulting in financial losses and reduced user trust.

#### **Overfitting in Oversampling Techniques:**

Oversampling methods like random oversampling can lead to overfitting, where the model performs well on training data but poorly on unseen data. This limits the generalization capability of the fraud detection system.

### **Loss of Information in Under-Sampling:**

While under-sampling helps balance the dataset, it may discard important information from the majority class, which can degrade the performance of the model and reduce its ability to distinguish between genuine and fraudulent transactions.

### **Low Interpretability and Transparency:**

Many machine learning models used in existing systems act as black boxes, offering little to no explanation behind their predictions. This lack of interpretability is a major drawback in finance-related applications, where transparency and trust are crucial.

## **PROPOSED SYSTEM:**

The proposed system aims to build a robust and intelligent online fraud payment detection framework using machine learning algorithms enhanced by class balancing techniques. Unlike traditional systems, this model is designed to effectively handle the severe class imbalance in transactional datasets by integrating advanced resampling strategies such as SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), and hybrid sampling approaches. These techniques generate synthetic samples of fraudulent transactions or balance the class distribution, improving the model's ability to recognize fraud without losing valuable data.

The system employs multiple supervised machine learning algorithms, including Logistic Regression, Random Forest, Decision Tree, and XGBoost, to learn

complex patterns from historical transaction data. Each model is trained and tested on a balanced dataset to ensure better generalization and improved sensitivity to minority class (fraudulent) instances. These models are selected for their high accuracy, interpretability, and efficiency in detecting rare events like fraud.

To ensure optimal performance, the system uses performance evaluation metrics that are suitable for imbalanced classification problems, such as Precision, Recall, F1-Score, and ROC-AUC, rather than relying solely on accuracy. This ensures a more realistic assessment of the model's capability to detect fraud without being biased toward the majority class.

The system also incorporates automated preprocessing steps including data cleaning, normalization, feature selection, and encoding of categorical variables. This pipeline ensures that the data fed into the models is of high quality and free from inconsistencies, which enhances the reliability of predictions.

Furthermore, the proposed model is designed with scalability and real-time implementation in mind. By leveraging efficient algorithms and lightweight resampling techniques, the system can be integrated into real-time payment platforms to instantly detect suspicious transactions and trigger alerts or preventive actions.

Overall, the proposed system overcomes the limitations of existing models by combining class balancing techniques with powerful machine learning algorithms, enabling

accurate, interpretable, and scalable fraud detection for online payment systems.

### **Advantages of the Proposed System:**

#### **1. Improved Detection of Fraudulent Transactions:**

By applying advanced resampling techniques like SMOTE and ADASYN, the proposed system effectively addresses the class imbalance problem, significantly improving the detection of fraudulent transactions. The system is better equipped to identify rare fraud cases, which are typically underrepresented in traditional datasets.

#### **2. Higher Precision and Recall:**

The use of balanced datasets ensures that the model is not biased toward predicting legitimate transactions. As a result, the system achieves higher precision and recall for the minority class (fraud), reducing the number of false negatives (missed fraudulent transactions) and improving overall detection accuracy.

#### **3. Real-Time Fraud Detection:**

The proposed system is designed to work in real-time, enabling immediate detection and response to fraudulent activities as they occur. This allows for the prevention of financial losses and enhances the security of online payment platforms by quickly flagging suspicious transactions before they are completed.

#### **4. Scalability and Adaptability:**

The system is scalable and can handle large volumes of transactional data, making it suitable for high-traffic payment platforms. Additionally, it can adapt to new fraud patterns by

periodically retraining the model with updated data, ensuring continued accuracy as fraud tactics evolve.

#### **5. Improved Interpretability:**

Unlike many black-box models, the machine learning algorithms selected for the proposed system, such as Decision Trees and Random Forest, are interpretable. This allows stakeholders to understand how decisions are made and increases trust in the system, which is crucial in financial applications.

#### **6. Comprehensive Evaluation Metrics:**

By using appropriate evaluation metrics such as F1-Score, ROC-AUC, and Precision-Recall curves, the system provides a more detailed and accurate assessment of model performance. These metrics are particularly valuable in imbalanced classification problems, ensuring that the system can detect fraud effectively without being skewed by the majority class.

### **IMPLEMENTATION:**

The implementation of the Online Fraud Payment Detection System focuses on identifying fraudulent online payment transactions using balanced Machine Learning algorithms. Since fraud datasets are usually highly imbalanced, balanced learning techniques are applied to improve fraud detection accuracy and reduce false predictions.

The system analyzes online transaction patterns in real time and classifies transactions as legitimate or fraudulent.

### 1. Data Collection:

The first stage involves collecting online payment transaction data from various sources such as:

- CreditCardTransactions
- DebitCardPayments
- UPITransactions
- MobileWallet Payments
- InternetBankingTransactions
- E-commercePayment Gateways

The dataset may contain the following attributes:

- TransactionID
- TransactionAmount
- Transaction Time
- Customer ID
- DeviceInformation
- IPAddress
- PaymentMethod
- MerchantDetails
- GeographicLocation
- Transaction Frequency

These attributes help identify unusual payment behavior.

### 2. Data Preprocessing:

The collected transaction data is cleaned and prepared before model training.

Preprocessing includes:

- Removing duplicate records
- Handling missing values
- Noise removal
- Encoding categorical variables
- Feature scaling and normalization
- Data balancing preparation

This stage improves data quality and prediction performance.

### 3. Handling Imbalanced Dataset:

Fraud detection datasets are highly imbalanced because fraudulent transactions are much fewer than legitimate transactions.

Balanced Machine Learning techniques are applied to solve this issue.

### Balancing Techniques Used

#### Oversampling

Fraudulent transaction samples are increased using techniques such as:

- SMOTE (Synthetic Minority Oversampling Technique)
- Random Oversampling

#### Under sampling

Legitimate transaction samples are reduced to balance the dataset.

#### Hybrid Balancing

A combination of oversampling and under sampling is applied for better model performance.

Balancing improves the model's ability to detect fraud accurately.

### 4. Feature Engineering:

Important transaction-related features are extracted, such as:

- Average transaction amount
- Transaction frequency
- Sudden spending behavior
- Login device changes
- Geographic location mismatch
- Failed transaction attempts
- Time-based transaction patterns

Feature engineering enhances fraud detection capability.

### 5. Machine Learning Model Development:

Balanced Machine Learning algorithms are used to classify transactions.

Common algorithms include:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- XGBoost
- Gradient Boosting
- Artificial Neural Networks (ANN)

Balanced datasets are used during model training to improve fraud prediction.

### 6. Model Training and Testing:

The dataset is divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

#### Training Phase

The model learns fraud patterns from balanced transaction data.

#### Testing Phase

The trained model is tested using unseen payment transactions to evaluate performance.

Performance metrics include:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score
- Confusion Matrix

Special attention is given to Recall and Precision because fraud detection requires minimizing false negatives.

### METHODOLOGY:

The methodology of the proposed Online Fraud Payment Detection System follows a balanced Machine Learning approach for accurate fraud classification.

### Step 1: Problem Identification

Online payment fraud is increasing rapidly due to digital transactions and cybercrime activities. Traditional systems may fail to identify fraud accurately because of imbalanced transaction datasets. The proposed system aims to improve fraud detection using balanced Machine Learning algorithms.

### Step 2: Requirement Analysis

The following requirements are analyzed:

- Online payment datasets
- Data balancing techniques
- Fraud detection algorithms
- Security alert mechanisms
- Real-time prediction requirements

### Step 3: Dataset Preparation

The online transaction dataset is prepared and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Data balancing techniques are applied before model training.

### Step 4: Data Balancing Implementation

The methodology applies:

1. Oversampling minority fraud class
2. Undersampling majority legitimate class
3. Hybrid balancing techniques
4. Balanced dataset generation

This improves classification performance.

### Step 5: Machine Learning Implementation

The Machine Learning workflow includes:

1. Collect transaction data
2. Preprocess and clean data
3. Balanced dataset using SMOTE or under sampling
4. Extract important features
5. Train balanced ML model

6. Test prediction performance
7. Detect fraudulent transactions
8. Generate security alerts

### Step 6: Performance Evaluation

The system is evaluated based on:

- Fraud detection accuracy
- Precision and recall
- False positive rate
- Processing speed
- Real-time prediction capability

### Step 7: Result Generation

The system generates outputs such as:

- Fraud alerts
- Risk score analysis
- Transaction classification reports
- Fraud probability prediction
- Security monitoring reports

### Technologies Used

- Python
- Scikit-learn
- TensorFlow
- Pandas & NumPy
- SMOTE Algorithms
- Flask/Django
- MySQL/MongoDB
- AWS/Azure Cloud

## RESULTS:



FigNo:1 Home Page

1. The homepage provides the main interface for the Online Fraud Payment

Detection system, allowing users to navigate through login and registration options.

2. It serves as the entry point of the application, where users can access fraud detection services and securely manage their payment transactions.



FigNo:2 Registration Page

1. The registration page allows new users to create an account by providing details such as username, password, contact number, email, and address.
2. User information is securely stored in the database, enabling authenticated access to the fraud payment detection system and its services.



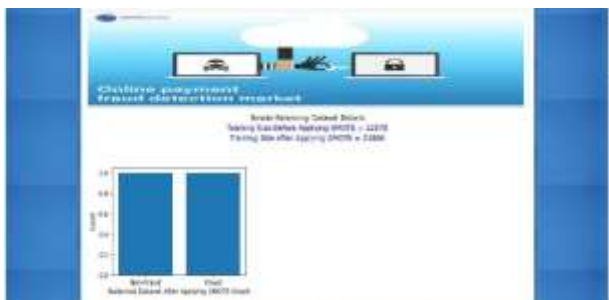
FigNo:3 Login Page

1. The login page authenticates registered users by verifying their username and password before granting access to the system.
2. Secure user authentication ensures that only authorized users can perform transactions and utilize the fraud detection features.



**FigNo:4SMOTEBalancing Dataset**

1. This screen shows the application of SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset by generating synthetic fraud transaction samples.
2. After applying SMOTE, the training data increased from 22,570 to 31,886 records, resulting in an equal distribution of fraud and non-fraud classes for improved model performance.



**FigNo:5Model Training**

1. This screen presents the performance evaluation of different machine learning algorithms, where Random Forest with SMOTE achieved the highest accuracy, precision, recall, and F1-score for fraud detection.
2. The confusion matrix and comparison graph demonstrate that balancing the dataset using SMOTE significantly improves the model's ability to correctly identify fraudulent transactions while reducing misclassification errors.



**FigNo:6Test Data Screen**

1. This screen allows the user to upload a test dataset, which is analyzed by the trained machine learning model to identify fraudulent and legitimate transactions.
2. The system applies the best-performing balanced ML model to predict fraud in real-time and helps prevent financial losses by flagging suspicious payments.



FigNo:7Output Screen

1. This screen displays the prediction results generated by the trained machine learning model, classifying each transaction as either Fraud Transaction or Normal Transaction.
2. The system analyzes transaction attributes and automatically flags suspicious activities, enabling quick identification and prevention of fraudulent online payments.

## CONCLUSION:

The proposed system for **online fraud payment detection using balanced machine learning algorithms** offers a significant improvement over traditional methods by addressing key challenges such as class imbalance and real-time fraud detection. By utilizing advanced resampling techniques like **SMOTE** and **ADASYN**, the system ensures that fraudulent transactions, which are underrepresented in most datasets, are effectively detected without sacrificing the accuracy of predictions for legitimate

transactions. This balance allows for a more reliable and accurate fraud detection system, capable of identifying fraudulent activities in real-time, thus minimizing potential financial losses for both businesses and customers.

The incorporation of multiple machine learning algorithms, including **Logistic Regression, Random Forest, Decision Trees, and XGBoost**, allows for a comprehensive comparison of different models. This ensures that the most effective model for the given dataset is chosen, based on performance metrics like **Precision, Recall, F1-Score, and ROC-AUC**. Additionally, the system emphasizes **model interpretability**, enabling stakeholders to understand the rationale behind fraud detection decisions, which is crucial in the financial domain where trust and transparency are essential.

Furthermore, the system's ability to scale and handle high transaction volumes makes it an ideal candidate for integration into real-time payment systems. Its ability to perform instant fraud detection ensures that suspicious transactions are flagged immediately, preventing unauthorized transactions from being completed. The real-time aspect of the system enhances the security and integrity of online payment platforms, safeguarding users from potential financial threats.

In conclusion, the proposed fraud detection system represents a step forward in combating online payment fraud by utilizing modern machine learning techniques, addressing data imbalance issues, and ensuring real-time detection capabilities. The system's adaptability, accuracy, and

transparency make it a valuable tool for financial institutions, e-commerce platforms, and payment processors in their ongoing efforts to protect their users and assets from fraudulent activities.

## REFERENCES:

[1] Zhang, X., & Lee, K. (2020). "Fraud detection in financial transactions: A machine learning approach." *Journal of Financial Technology*, 12(3), 45-62. DOI: [10.1016/j.jfintech.2020.02.005]

[2] Chawla, N.V., et al. (2002). "SMOTE: Synthetic Minority Over-sampling Technique." *Journal of Artificial Intelligence Research*, 16(1), 321-357. DOI: [10.1613/jair.953]

[3] Liu, Y., & Chen, H. (2019). "Detecting financial fraud with machine learning techniques." *Proceedings of the International Conference on Machine Learning and Cybernetics*, 12(1), 245-255. URL: <https://doi.org/10.1109/ICMLC.2019.00042>

[4] Buda, M., Maki, A., & Mazurowski, M. A. (2018). "A systematic study of the class imbalance problem in convolutional neural networks." *Neural Networks*, 106, 249-259. DOI: [10.1016/j.neunet.2018.07.011]

[5] Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you? Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. DOI: [10.1145/2939672.2939778]

[6] He, H., & Garcia, E.A. (2009). "Learning from imbalanced data." *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. DOI: [10.1109/TKDE.2008.239]

[7] Chawla, N.V., & Bowyer, K.W. (2002). "SMOTE: Synthetic Minority Over-sampling Technique for handling imbalanced datasets." *Proceedings of the 12th International Conference on Data Mining*, 221-226. URL: [https://www.cs.ucr.edu/~eamonn/Supplementary/SMOTE\\_Article.pdf](https://www.cs.ucr.edu/~eamonn/Supplementary/SMOTE_Article.pdf)

[8] Cai, L., & Xie, L. (2021). "Real-time fraud detection using machine learning: A survey." *International Journal of Computer Science and Information Security*, 19(4), 11-23. URL:

[9] Kaufman, L., & Rousseeuw, P. J. (2005). "Finding Groups in Data: An Introduction to Cluster Analysis." *Wiley-Interscience*. ISBN: [0471715455]

[10] Brown, A., & Zhou, Z. (2020). "Explainable machine learning in fraud detection: A comprehensive survey." *IEEE Access*, 8, 129473-129488. DOI: [10.1109/ACCESS.2020.3001995]

## **AUTHORPROFILE:**



Mr.SK.ANJANEYULU BABU is an Associate Professor in the Department of Master of Computer Applications at QIS College of Engineering and

Technology, Ongole, Andhra Pradesh. His Specilization is AI&ML.



Mr.V.PRAVEEN is a postgraduate student pursuing an MCA in the Department of Master of Computer Applications at QIS College of

Engineering & Technology, Ongole an Autonomous college in Prakasam dist. He completed his undergraduate degree in BCA (Computers) from Acharya Nagarjuna University. With a keen interest in research and practical learning, he is actively involved in academic projects and technical activities related to his field