COMPARISON OF VARIOUS FRAUD DETECTION USING MACHINE LEARNING

MOTIVATION BEHIND THE PROJECT

In today's digital world, financial fraud is a growing concern, causing global losses over $1 trillion annually. Our project aims to combat this threat with a real-time, high-precision machine learning model for fraud detection. Data breaches often take 197 days to detect, giving fraudsters ample time. We're driven to change this by providing a swift defense against fraudulent transactions, protecting assets and ensuring financial stability. Our goal: reduce detection time from months to minutes, minimizing losses and rebuilding trust in digital transactions for a safer financial landscape.

TYPE OF PROJECT

Research cum development project

CRITICAL ANALYSIS OF RESEARCH PAPER READ

Research papers	Abstract
A State of the Art Survey of Fraud Detection Technology by Ulrich Flegel, Julien Vayssiere, and Gunter Bitz	The survey highlights key challenges in fraud detection,such as: ● Developing effective IT-based access control. ● Ensuring quality Separation of Duties (SoD) model and managing risks.
Fraud Detection in Online Payments using Machine Learning Techniques Pasala Anjaneyulu, Y. Haritha, Mande Ramesh	● This research study introduces a machine learning model with XG Boost for fraud detection. ● XG Boost excels in speed and accuracy compared to other algorithms.
Fraud Transaction Detection Approach Using Machine Learning Hybrid Techniques-International Journal of Scientific Research in Science and Technology	● Various algorithms(Linear SVC, SVC with RBF Kernel, Logistic Regression, Random Forest, Decision Tree) were tested on a mobile money dataset for fraud detection. ● Evaluated based on accuracy, precision, recall, and F1-score. ● Identified while minimizing false positives.

Survey of Fraud Detection Techniques	• Methods like data mining, statistics, and artificial intelligence focus on detecting anomalies and patterns in data. • Credit card fraud detection uses outlier detection to identify significant deviation

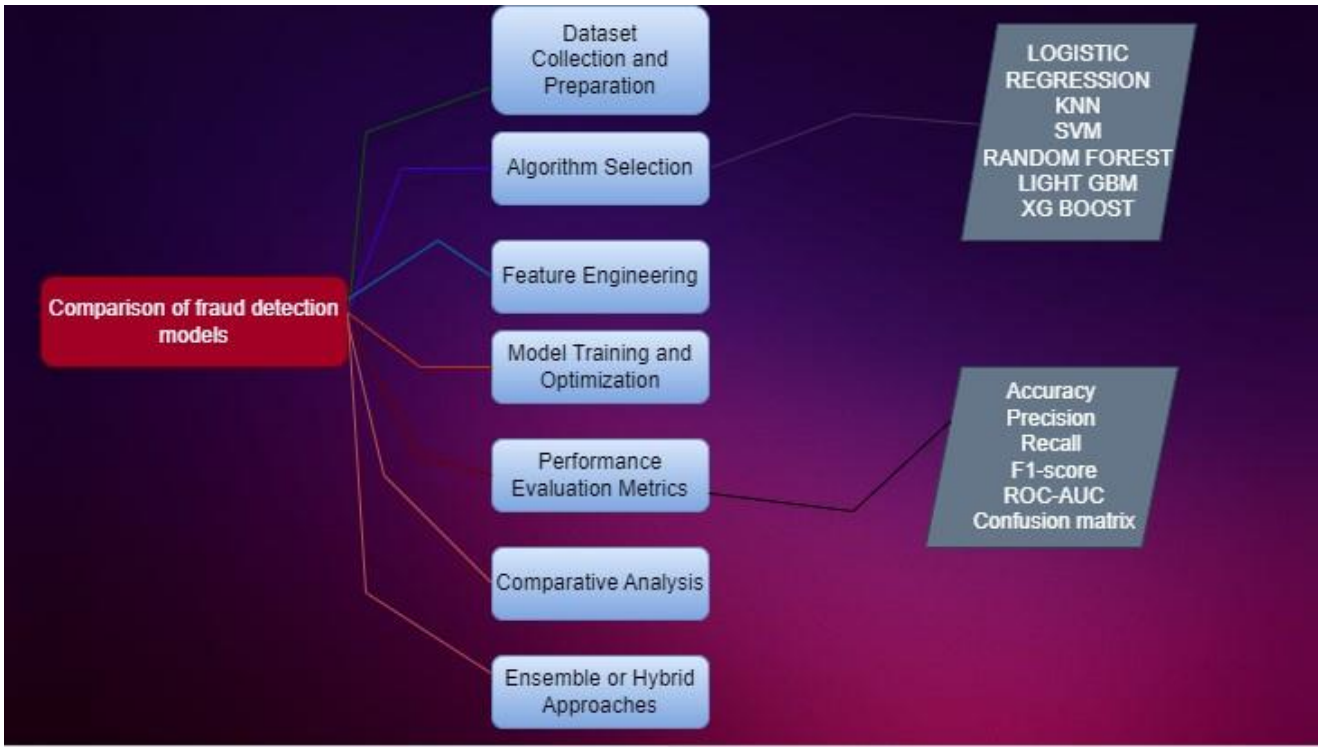
In parallel with the project's development, the team has proactively pursued additional learning opportunities through online courses.

TOOLS OR SOFTWARE USED

Programming language involved: Python

Libraries used: numpy, pandas, matplotlib, seaborn, scipy, sklearn

DESIGN OF THE PROJECT



FEATURES

- Dataset Collection and Preparation: Gathered diverse real-world fraud data and preprocessed it for analysis, ensuring data quality and consistency across models.

- Algorithm Selection: Defined selection criteria based on literature review insights, compares 6 models: XG Boost,Light GBM,Random Forest,SVM,KNN,Logistic Regression And one Dummy Classifier.
- Feature Engineering: Engineered Features like Step,Diff_new_balance,Diff_new_Destiny and OldBalanceorg, distinctive features from the fraud data to enhance model detection capabilities and feature relevance.

- Model Training and Optimization: Trained multiple machine learning models using historical data, employing cross-validation techniques.

- Performance Evaluation Metrics: Evaluated models using standard metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices to compare their accuracy.

- Comparative Analysis: Conducted a comparative analysis of different models' performances to identify strengths and weaknesses across various fraud detection approaches.

- Ensemble or Hybrid Approaches: Explored the potential of ensemble or hybrid methods to combine the strengths of different algorithms for improved fraud detection outcomes.

PROPOSED METHODOLOGY

- Data Preparation: Collected real-world fraud data and preprocessed it, including data cleaning, transformation, and feature extraction.
- Algorithm Selection: Based on the findings from the literature review, selected appropriate machine learning algorithms for fraud detection, considering factors like sensitivity, specificity, and adaptability to different data scenarios.
- Feature Engineering: Engineered relevant features from the fraud data to improve all the models detection capabilities.
- Model Training: Trained all the selected machine learning models using historical data, optimizing hyperparameters, and employing cross-validation techniques.
- Model Evaluation: Evaluated the models performances using standard metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices.

DESCRIPTION OF THE PROJECT

The project involves a systematic investigation of machine learning algorithms for fraud detection in financial transactions. It spans multiple phases: collecting and preprocessing real-world fraud data, selecting algorithms based on literature insights, and engineering features to enhance model capabilities. Training models using historical data involves optimization techniques and rigorous evaluation using standard metrics. The comparative analysis aims to highlight strengths and weaknesses, exploring ensemble methods for potential enhancements. A continuous feedback loop ensures ongoing refinement, while documented findings will guide future research in this domain.

DIVISION OF WORK

Although we did all our work together but here are the divisions of work :-

- Mayank Johri :-Data visualization and Preprocessing And Logistic Regression

- Kaanayan Sukhija :- Model comparison(Dummy Classifier, SVM,KNN,Random Forest)

- Devansh Agarwal :- Model Comparison(XG Boost,Light GBM) And Evaluation(Accuracy,Precision,Recall,F1 Score,Confusion Matrix,ROC,Precision Recall Curve)

RESULT

After evaluating diverse machine learning methods for fraud detection, notable performance discrepancies were observed. Each model showcased strengths in various areas, highlighting the importance of tailored selection based on dataset and detection requirements. Our analysis emphasizes the potential for enhanced fraud detection by combining algorithms strategically, underlining the value of customized solutions in e ective fraud prevention strategies.

CONCLUSION

The project's comprehensive exploration of various machine learning algorithms for fraud detection in financial transactions has yielded valuable insights.This research project will significantly contribute to the domain of fraud detection by providing a thorough examination of existing techniques and introducing a novel machine learning model tailored to unsupervised data. The technical specifications, design, test plan, and implementation details ensure a rigorous and scientific approach to this problem. Furthermore, the originality of the problem statement underscores the significance of this research endeavor in addressing contemporary challenges in fraud detection