
ONLINE
PAYMENT
FRAUD
DETECTION

TEAM MEMBERS:

- SUBIN SIBY

  MUT23MCA-2058
- SAHALA MOL V S

  MUT23MCA-2055
- NAEEMA KUNHIMON

  MUT23MCA-2047
- SHERIN MARIYA ROBERT

  MUT23MCA-2056

GUIDE:DR.GEETHU S

ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER APPLICATIONS

INTRODUCTION

- The primary goal of this project is to implement supervised machine learning models for fraud detection, with the goal of analyzing prior transaction information.
- The goal is to predict whether a transaction is a legal transaction or a fraudulent transaction

LITERATURE REVIEW

S.NO	TITLE	AUTHOR & YEAR	KEY FINDINGS
1	Online payment fraud: from anomaly detection to risk management.	Paulo Vanini, Sebastiano Rossi, Ermin and Thomas Domenig. Published online on 13 march 2023.	Integrated approach of anomaly detection and risk modeling and improve fraud detection capability.
2	Fraud_Detection_ML: Machine Learning based on online payment Fraud Detection.	Maged Farouk, Nashwa S Ragaba, Diaa Salama,Omnia Elrashidy Journal of Computing and Communication Vol.3 , No.1 , PP. 116-131 , 2024 Published on: 1 January 2024	This paper explores an effective framework for detecting online payment fraud. Gradient Boosting is identified as the optimal solution due to its high accuracy and robustness across different datasets.

S.NO	TITLE	AUTHOR & YEAR	KEY FINDINGS
3	Leveraging Machine Learning Algorithms for Real-Time Fraud Detection in Digital Payment Systems.	Nakra, V., Pandian, P. K. G., Paripati, L., Choppadandi, A., & Chanchela, P. (2024). International Journal of Multidisciplinary Innovation and Research Methodology (IJMIRM), 3(2), 165-171.	The study proposes an ensemble approach combining logistic regression and deep learning to improve fraud detection in digital payment systems, emphasizing the role of feature engineering and the potential for significant financial savings.
4	Financial Fraud Detection Model: Based on Random Forest	Kazmi, S. H. A., Liu, C., Chan, Y., & Fu, H. (2015). Financial Fraud Detection Model: Based on Random Forest. International Journal of Economics and Finance, 7(7), 178-188. doi:10.5539/ijef.v7n7p178.	This study presents a financial fraud detection model using Random Forest (RF), demonstrating its high accuracy in detecting fraud in large datasets. The research, based on data from Chinese listed companies, highlights the effectiveness of RF, especially when analyzing key financial variables.

S.NO	TITLE	AUTHOR & YEAR	KEY FINDINGS
5	Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review	Abdulalem Ali, Shukor Abd Razak, Siti Hajar Othman, Taiseer Abdalla Elfadil Eisa, Arafat Al-Dhaqm, Maged Nasser, Tusneem Elhassan, et al. Published on: 26 Sept 2022	This review paper comprehensively analyzes machine learning techniques used in financial fraud detection, highlighting key trends, challenges, and advancements in the field.
6	Predictive-Analysis-based Machine Learning Model for Fraud Detection with Boosting Classifiers	M. Valavan, S. Rita. Department of Statistics, Periyar University, Salem, Tamil Nadu, India. Published on: 6 April 2022	This paper explores the effectiveness of various machine learning algorithms, such as Decision Tree, Random Forest, Linear Regression, and Gradient Boosting, in detecting and predicting fraud, particularly in loan cases.

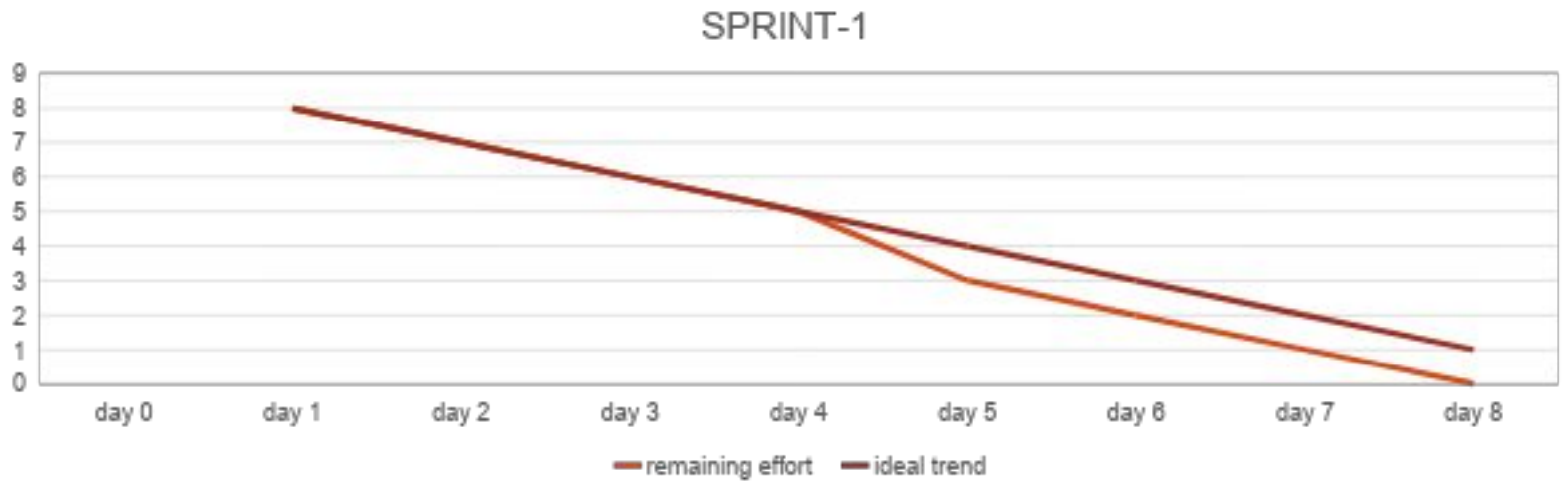
S.NO	TITLE	AUTHOR & YEAR	KEY FINDINGS
7	An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection	Minastireanu, E. A., & Mesnita, G. (2019). An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection. *Informatica Economica*, *23*(1).	The paper reviews online fraud detection techniques, highlighting that supervised learning algorithms-Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT)are most effective for detecting fraud, particularly in credit card transactions. It identifies key factors like accuracy, coverage, and cost efficiency as crucial for evaluating these algorithms.
8	Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System	Choi, D., & Lee, K. (2017). Machine learning-based approach to financial fraud detection process in mobile payment system. *IT Convergence Practice (INPRA)*, *5*(4), 12-24.	The research explores a dual approach to mobile payment fraud detection, combining supervised and unsupervised machine learning methods to manage large datasets and address data imbalance.

PRODUCT BACKLOG

BACKLOG ID	USER STORIES	PRODUCT BACKLOG
111	As a user, I want to perform a literature review to understand the project objectives and context	• Review academic papers, industry reports, and existing solutions related to fraud detection. • exploring the dataset • clean the dataset • preprocess the dataset
112	As a user, I want to preprocess data and perform feature engineering for fraud detection.	• Scale and normalize numerical features • Perform feature selection • Split data into training, validation, and test sets.
113	As a user, I want to evaluate different models and analyze their performance for fraud detection.	• Choose candidate algorithms based on literature review findings • Implement initial models and compare performance. • Fine-tune hyperparameters for the best-performing model. • Cross-validate model performance to check for overfitting

BACKLOG ID	USER STORIES	PRODUCT BACKLOG
114	As a user, I want to evaluate the model & analysing the metrics	• Evaluate model accuracy, precision, recall, F1-score. • Generate confusion matrix and classification report • Assess model performance using AUC-ROC curve. • Interpret model results for clinical relevance.
115	As a user, I want to improve the model & test it	• Adjust settings to improve the model's accuracy • Check where the model is making mistakes and try to fix them. • Try combining multiple models to get better results. • Compare how different models work and pick the best one.
116	As a user, I want to do Final checks and Make a report	• Make sure the model works well with different types of data. • Explain in simple terms how the model makes its decisions. • Write down a short guide on how to use the model. • Double-check everything

Burndown Chart										
Backlog ID	USER STORIES	Initial Estimate	Jul-22	Jul-25	Jul-29	Aug-02	Aug-05	Aug-06	Aug-13	Aug-19
		day 0	day 1	day 2	day 3	day 4	day 5	day 6	day 7	day 8
111	literature review	2	1			1				
111	gather dataset	2		1					1	
111	clean dataset	2			1		1			
111	preprocessing	2				1		1		
remaining effort		8	7	6	5	3	2	1	0	0
ideal trend		8	7	6	5	4	3	2	1	0



DATA PROCESSING

- **Data Collection:**

  Gather Data from significant sources

- **Data Cleaning :**

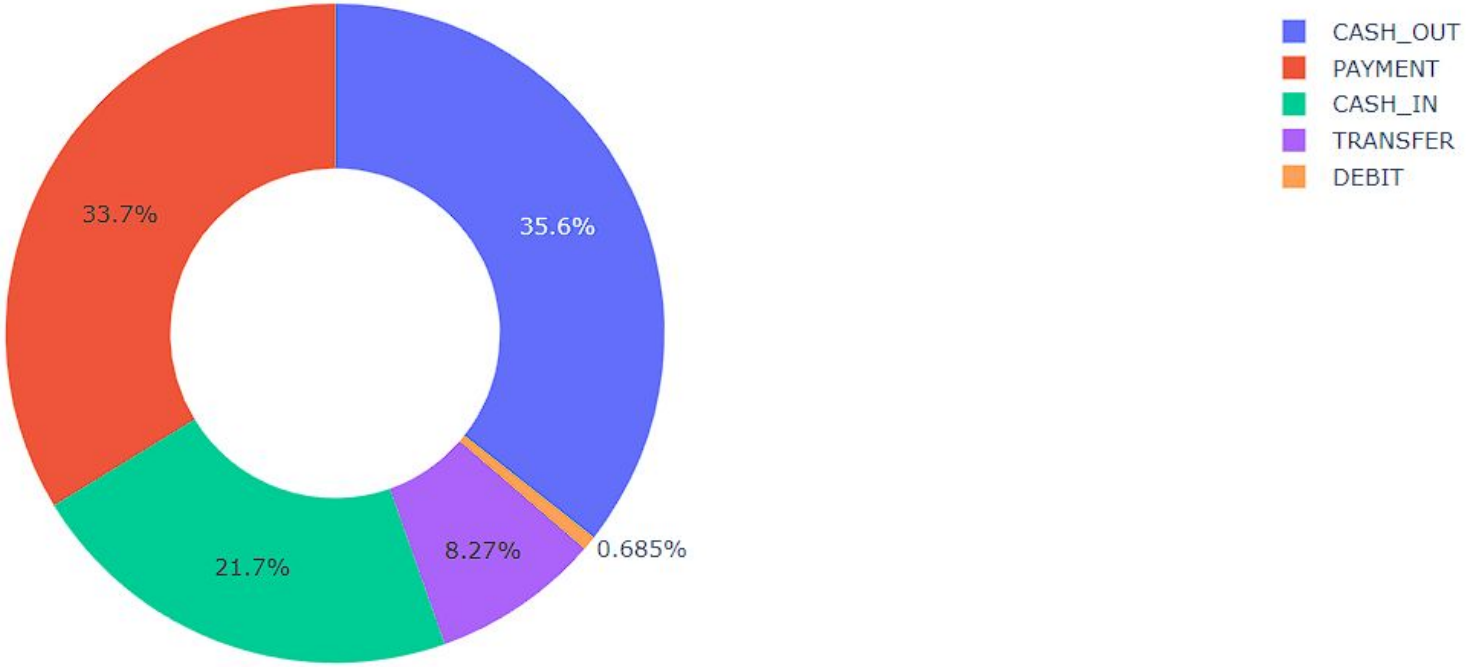
  Remove or handle irrelevant or duplicate data

- **Encoding categorical Data:**

  Convert categorical features into numerical values

DATA PREPROCESSING STEPS

- Understand the Dataset
- Importing libraries
- Handle Missing Values
- Outlier Detection
- Data Splitting



MODEL SELECTION

MODEL SELECTION

We are planning to use 4 models in this project, and they are:

- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Decision_Tree
- Random Forest

K-NEAREST NEIGHBORS (KNN)

Overview:

- **KNN Approach:** Compares transactions to nearest neighbors.
- **Legitimate:** Close to legitimate neighbors, likely legitimate.
- **Fraud Detection:** Close to fraudulent neighbors, likely fraud.

Why These Models Were Used:

- Simple and effective for spotting fraud by comparing new transactions to past cases.

Advantages of These Models:

- **Simple to Implement**.
- **Effective for Similar Patterns**

SUPPORT VECTOR MACHINE (SVM)

Overview:

- **SVM Approach:** Separates fraud from legitimate.wade
- **Hyperplane:** Maximizes margin between classes.
- **Effective:** Works well in complex, high-dimensional spaces.

Why These Models Were Used:

- Great at handling complex data and finding clear boundaries between fraudulent and legitimate transactions..

Advantages of These Models:

- **Handles Complex Data**.
- **Clear Separation.**

DECISION TREE

Overview:

- **Decision Tree:** Splits transactions into branches.
- **Features:** Uses transaction details like amount, frequency, location.
- **Classification:** Each decision point leads to fraud or legitimate outcome.

Why These Models Were Used:

- Easy to understand and explain, providing clear rules for identifying fraud.

Advantages of These Models:

- **Handles Complex Data**.
- **Clear Separation.**

RANDOM FOREST

Overview:

- **Random Forest:** Creates multiple decision trees.
- **Ensemble Method:** Combines results from random subsets.
- **Enhanced Detection:** Averages tree results to reduce errors.

Why These Models Were Used:

- Combines multiple decision trees for higher accuracy, especially useful for large datasets with rare fraud cases.

Advantages of These Models:

- **High Accuracy**.
- **Effective for Imbalanced Data.**
- **Resistant to Overfitting**

CONCLUSION

- Effective fraud detection and prevention are crucial in today's digital landscape, driven by rapid technological advancements.
- Adopting best practices such as continuous monitoring and multi-layered security measures, like two-factor authentication, is essential to safeguarding against fraud.
- Looking ahead, future innovations will continue to enhance these defenses, underscoring the importance of a steadfast commitment to security.
- Staying proactive and adaptive in our approach will ensure robust protection against evolving threats.

REFERENCES

- https://www.kaggle.com
- https://link.springer.com/article/10.1186/s40854-023-00470-w
- https://journals.ekb.eg/article_339929.html
- https://www.revistaie.ase.ro/content/89/01%20-%20minastireanu,%20mesnita.pdf
- https://isyou.info/inpra/papers/inpra-v5n4-02.pdf
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2625215
- https://ijmirm.com/index.php/ijmirm/article/view/97
- https://premiumedutech.com/wp-content/uploads/2023/12/paper-3.pdf
- https://www.mdpi.com/2076-3417/12/19/9637

THANK YOU