Capstone Project – Final: Review – 1 on

Online Payment Fraud Detection

Presented by the Batch No. : CSBS-3

G.AKASH	**BU21CSEN0200131(TL)**
V.JATHIN	**BU21CSEN0200046**
A.SHARAN	**BU21CSEN0200079**
G.MANINDAR	**BU21CSEN0200168**

Under the Guidance of

Mrs.Chandra Prabha

Assistant Professor

Department of Artificial Intelligence & Data Science

GITAM School of Technology, GITAM (Deemed to be University),
Bengaluru, Karnataka, INDIA.

CONTENTS

Abstract

The system first preprocesses transaction data to extract relevant features, followed by the implementation of XGBoost for initial fraud prediction, leveraging its efficiency and robustness in handling imbalanced datasets. Subsequently, the predictions are refined using a DNN, which captures complex patterns in the data. To optimize model parameters and improve accuracy, WOA is employed, inspired by the hunting behavior of whales, to efficiently navigate the solution space.

CONTENTS

Introduction

Online payment fraud detection using XGBoost and DNN with WOA is a method for identifying fraudulent transactions in online payments. It combines XGBoost, a powerful machine learning algorithm, and Deep Neural Networks (DNN) to analyze transaction data. The Whale Optimization Algorithm (WOA) is used to improve the accuracy of these models by optimizing their parameters. Together, this approach enhances the ability to detect and prevent fraudulent activities, ensuring safer online transactions.

Challenges

Complexity: Combining XGBoost, DNN, and WOA into a single model adds complexity in terms of both architecture and computational resources. Computational Cost: Training both XGBoost and DNN models, along with running the Whale Optimization Algorithm, may require significant computational resources, especially if the dataset is large Tuning WOA: WOA itself has parameters (e.g., population size, number of iterations) that need tuning for optimal performance. These must be carefully adjusted to get the best results from the model.

Introduction..contd

- **XGBoost** (Extreme Gradient Boosting) is a highly efficient and scalable implementation of the gradient boosting algorithm. It's widely used in machine learning for tasks such as classification and regression due to its high performance, flexibility, and ability to handle a wide range of data types.

- A **DNN (Deep Neural Network)** is a type of artificial neural network with multiple layers between the input and output layers. It is called "deep" because it has more than one hidden layer, allowing it to learn complex patterns and features from the data.

- The **Whale Optimization Algorithm (WOA)** is a nature-inspired optimization algorithm that mimics the hunting behavior of humpback whales. It is used to find optimal solutions to complex problems by simulating how whales encircle and trap prey in a bubble-net formation.

CONTENTS

Literature Survey

Table 1: Literature Survey

S.No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
1.	Personal loan default prediction and impact analysis of debt/DR raj Gopal April 2023	It highlights how debt levels influence default risks, emphasizing the importance of early intervention strategies to mitigate financial loss..	Logistics Regression, SVM	Expand dataset for better generalization

Literature Survey…contd

Table 1: Literature Survey Contd…

S.No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
2.	Research on credut risk models via machine learning algorithm and logestic regression for prediction/kera	It compares these models in terms of accuracy and efficiency, highlighting their strengths in identifying potential defaulters and managing financial risks effectively.	Decision trees, XG Boost	Simplifying interface for users

Literature Survey…contd

Table 1: Literature Survey Contd…

S.No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
3 .	Credit default via interpretable ensemble tranfer/p.oladejo juhe 2024	By incorporating data from multiple domains and adjusting the weights of different datasets, TrLightGBM enhances prediction accuracy and transparency, making it more reliable for credit risk assessment.	SVM, Logistic regression	Challenges remain in handling imbalanced datasets and improving model interpretability

Literature Survey…contd

Table 1: Literature Survey Contd…

S. No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
4.	Advanced fraud detection method for dnn .sep.23	The study leverages the ability of DNNs to learn complex patterns from large datasets, making them particularly effective in the dynamic landscape of financial fraud.	Feature Selection and Engineering, Data Preprocessing	It highlights the potential of advance deep neural networks for effective fraud detection.

Literature Survey…contd

Table 1: Literature Survey Contd…

S. No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
5.	Advanced fraud detection method for dnn .sep.23	The study leverages the ability of DNNs to learn complex patterns from large datasets, making them particularly effective in the dynamic landscape of financial fraud.	Feature Selection and Engineering, Data Preprocessing	It highlights the potential of advance deep neural networks for effective fraud detection.

Literature Survey…contd

Table 1: Literature Survey Contd…

6.	Comparision of machine learning clsscification models for credit card default data.	studies comparing machine learning classification models for credit card default prediction. and classify the data models.	SVM , XG Boost	Challenges remain in handling imbalanced datasets and improving model interpretability
7.	Identification of Potential Future Credit Card Defaulters from non Defaulter using Self Oraganizing maps	SOM is a type of unsupervised neural network that helps in clustering and visualizing high-dimensional data, making it suitable for identifying patterns that may indicate the likelihood of default.	Clustering analysis, Self organization maps	Addressing the identified research gaps could lead to more robust models that enhance prediction.

Literature Survey…contd

Table 1: Literature Survey Contd…

S. No.	Paper Details	Summary	Techniques/Algorithms Used	Research Gap
8.	Default credit card prediction via machine learninhg/Sofianita Multlib Shuzlina Abdul Rahman 2023	various machine learning techniques were used for predicting credit card defaults.	Random Forest , XGBoost, Deep Neura Networks	Better model interpretability
9.	Credit default via interpretable ensemble tranfer/p.oladejo juhe 2024	Transfer Learning helps models generalize across different datasets, while LightGBM ensures efficient, scalable predictions.	RF, DNN	Enhancement of cross-domain adaptability and refinement of interpretability

CONTENTS

Problem Statement

- As online transactions continue to grow in prevalence and complexity, the incidence of payment fraud has become a significant concern for businesses and consumers alike. Traditional fraud detection methods often struggle to keep pace with the evolving tactics employed by fraudsters, leading to increased financial losses and a decline in consumer trust.
- This project aims to address the urgent need for more robust and effective online payment fraud detection systems. By leveraging advanced machine learning techniques, specifically XGBoost and Deep Neural Networks (DNN), combined with the Whale Optimization Algorithm (WOA) for feature selection and parameter tuning, this project seeks to enhance the accuracy and efficiency of fraud detection in real-time transactions.

CONTENTS

Objectives

- To Collect the dataset and preprocess the data.
- Train the model XG Boost, DNN, along with WOA.
- Evaluate the performance of the developed model (Confusion Matrix, Precision, Recall, F1-score, Accuracy, Error.
- Compare the model Accuracy with Efficiency Sample result.

CONTENTS

Requirement Analysis

- **1. Hardware Requirements**
- To run machine learning models like XGBoost and Deep Neural Networks (DNN), especially when optimizing with algorithms like Whale Optimization Algorithm (WOA), you need powerful hardware for efficient computation. Here's the breakdown:
- **Processor (CPU):**
  - Minimum: Intel i5 (8th Gen or higher) / AMD Ryzen 5
  - Recommended: Intel i7/i9 (10th Gen or higher) / AMD Ryzen 7/9 or above
  - Rationale: Multi-core processors ensure faster data processing and model training.
- **Graphics Processing Unit (GPU):**
  - Minimum: NVIDIA GTX 1050 or equivalent
  - Recommended: NVIDIA RTX 3060/3080 or equivalent

Requirement Analysis

•**Memory (RAM):**
•Minimum: 8 GB
•Recommended: 16 GB or more
•Rationale: More RAM allows handling larger datasets in memory, which is crucial for fraud detection systems with large transactional datasets.

•**Storage:**
•Minimum: 256 GB SSD
•Recommended: 512 GB SSD or more
•Rationale: Faster read/write speeds during data processing and model execution.

•**Other Peripherals:**
•Monitor with full HD resolution
•Stable Internet connection (for cloud services or data retrieval)

Requirement Analysis

2. Software Requirements

•Operating System:

- Windows 10 or 11 / Linux Ubuntu 20.04 or later / macOS (if applicable)

•Development Environment:

- **IDE/Code Editor:**
  - Jupyter Notebook, Google Colab

.

•Programming Languages:

- **Python 3.7 or higher**
  - Primary language for machine learning libraries, easy integration with optimization algorithms, and data preprocessing.

Requirement Analysis

•**Libraries/Frameworks:**

•**XGBoost:**

- •For building the gradient-boosted decision tree models..

- •**Pandas & NumPy:**
- •For handling and processing large datasets efficiently.

•**Matplotlib & Seaborn:**

- •For data visualization and plotting performance metrics.

•**Imbalanced-learn (imbalanced-learn):**

- •To address the class imbalance problem in fraud detection (e.g., SMOTE, ADASYN).

•**Optimization Algorithms:**

•**WOA (Whale Optimization Algorithm):**

- •A custom library or Python implementation for integrating WOA for hyperparameter optimization or feature selection.

CONTENTS

Identification of tools/technology/algorithms

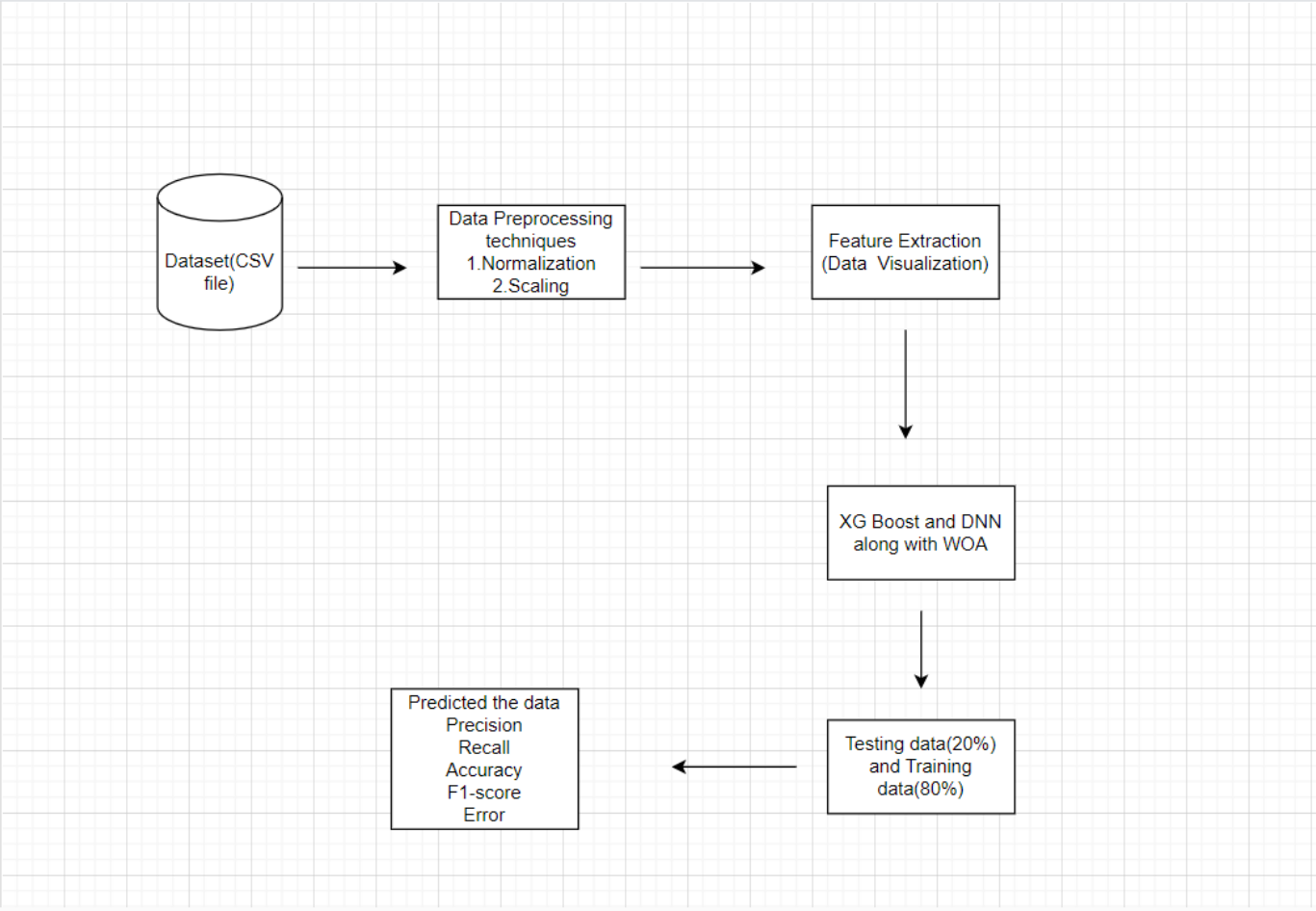
- XGBoost:
- A powerful gradient boosting algorithm, known for its accuracy and efficiency in handling large datasets.

- Deep Neural Networks (DNN)
- DNNs excel at learning complex patterns from data, enhancing fraud detection capabilities.

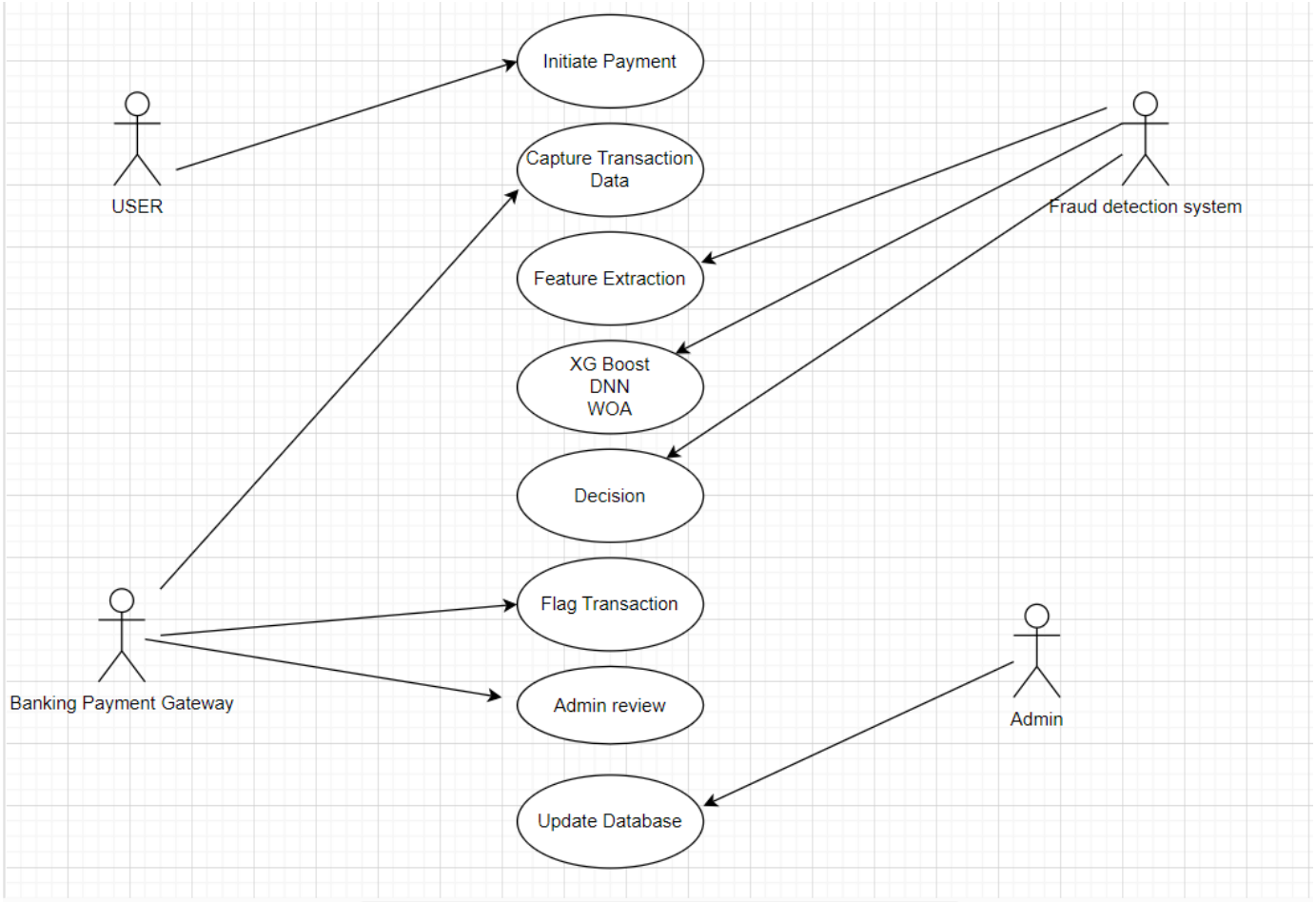
- Whale Optimization Algorithm (WOA)
- WOA is metaheuristic optimization algorithm inspired by humpback whales, used to optimize hyperparameters for the DNN.

CONTENTS

Final Design Strategy



usecase diagram



CONTENTS

Status of Implementation

Choose Dataset and implement Whale optimization,DNN(deep neural networks) and XG Boost

- Packages : pandas,numpy.
- Libraries : sk learn,tensorflow,scikit learn,matplot lib.
- Classes : train_test_split,classification_report,accuracy_score.
- Methods : Whale Optimization, XG boost, DNN.

- **Objective 1: *Collect and Preprocess the Dataset**

  1. *Collect Dataset*: - Identify and source a reliable dataset for online payment fraud detection.

  Common datasets include:

  - Publicly available datasets like those on Kaggle (e.g., the Credit Card Fraud Detection dataset).
  - Private datasets from finance institutions or payment systems.

2.Explore and Understand the Data:

- Load the dataset and examine its structure: -

-Check the number of rows and columns. - --

Understand feature descriptions (categorical vs. numerical).

- Identify the target variable (fraudulent or not).

- **Features:**

1. **Transaction ID**: Unique identifier for each transaction (categorical, will be excluded during modeling).
2. **Amount**: Transaction amount (numeric).
3. **Time Since Last Transaction**: Time since the last transaction (numeric).
4. **Location Match**: Binary indicator if the transaction's location matches expected (0 or 1).
5. **Device Match**: Binary indicator if the device matches the user's usual device (0 or 1).
6. **Transaction Channel**: The channel used for the transaction (e.g., Mobile, Web).
7. **Merchant Category**: Type of merchant (categorical).
8. **User Age**: Age of the user (numeric).
9. **Account Tenure**: Length of time the account has been active (numeric).
10. **Fraud Label**: Target variable indicating whether the transaction was fraudulent (0 or 1)

DATASET

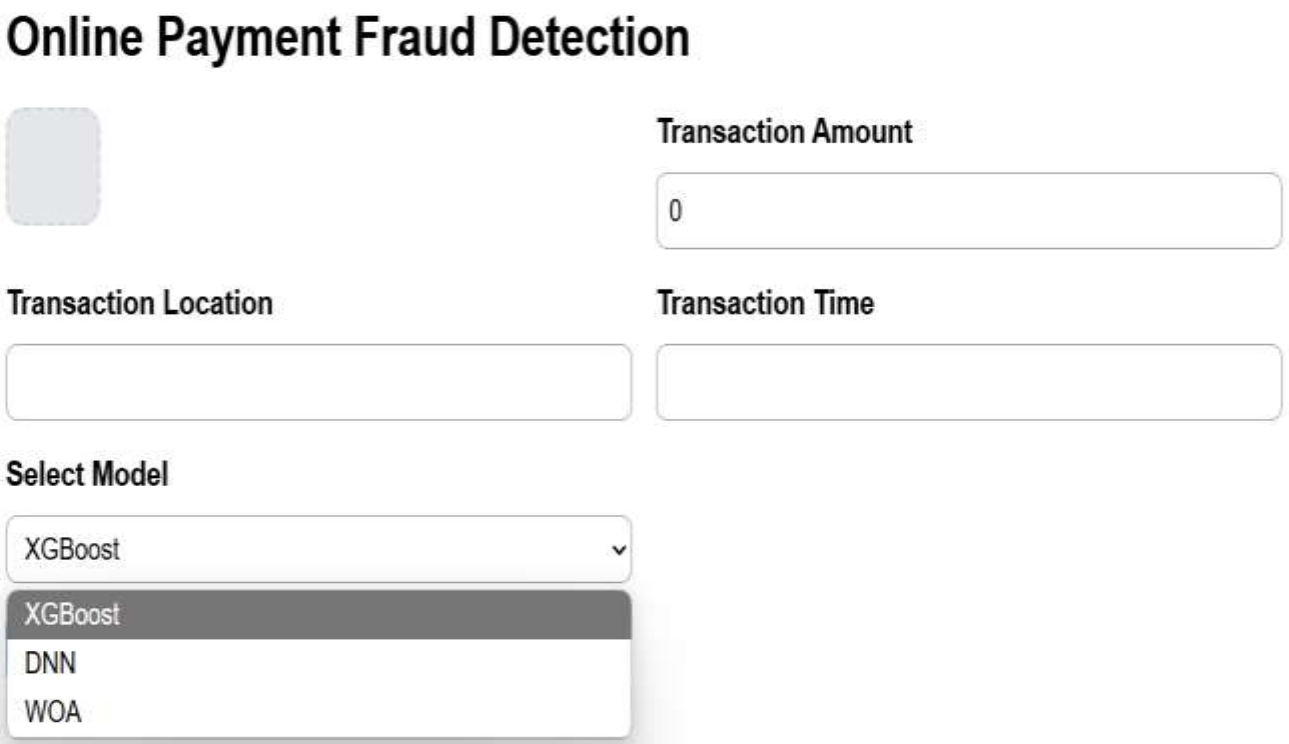
Q1

	A	B	C	D	E	F	G	H	I	J	K
1	Transaction	Amount	Time_Since_Last_Transaction	Location_Match	Device_Match	Transaction	Merchant_Category	User_Age	Account_Tenure	Fraud_Label	
2	TXN006252	338.139568	3.057399295288973	1	1	Mobile	Grocery	29	3.50078023550573	0	
3	TXN004684	775.729791	46.730939302493	1	1	Web	Retail	38	0.41947911809454	0	
4	TXN001731	624.271940	7.0275387007844605	0	0	Web	Dining	62	9.56013991946032	0	
5	TXN004742	36.5546405	49.18232647736028	1	1	Mobile	Retail	25	1.66762727615239	0	
6	TXN004521	577.539965	85.3731337937525	1	1	Mobile	Electronics	32	4.28241890737776	0	
7	TXN006340	248.215518	0.0428023681550465	0	0	POS	Electronics	49	3.48403055651821	0	
8	TXN000576	147.926407	112.50066596009373	1	1	Web	Dining	53	3.41237587409923	0	
9	TXN005202	643.134636	17.078885459036044	0	1	Mobile	Grocery	47	5.71753858011714	0	
10	TXN006363	470.334517	14.083461896282676	0	0	POS	Grocery	40	5.25289005713900	0	
11	TXN000439	82.2674318	12.667416374772888	0	1	POS	Dining	36	7.74265954692481	1	
12	TXN002750	388.777540	0.8607796078894258	0	1	POS	Clothing	46	3.29043203808983	0	
13	TXN007487	961.456160	22.405475502896696	1	0	POS	Clothing	43	5.32163788299509	0	
14	TXN005272	298.561538	13.48950677341387	1	1	Web	Electronics	43	8.43233129452143	0	
15	TXN005653	353.128565	8.234621939548969	0	1	Mobile	Retail	66	3.29816575861504	0	
16	TXN003999	162.205991	41.635538425604246	0	1	Web	Dining	25	5.31280118499556	0	
17	TXN006033	234.456758	16.22920649968193	0	0	POS	Clothing	23	9.67682001754241	0	
18	TXN000582	493.125475	19.17573798952754	0	1	Web	Clothing	36	1.75205888312075	0	
19	TXN009930	823.403132	9.406560614703375	0	1	Web	Retail	30	4.82802179800394	0	
20	TXN007051	527.727894	2.873120838805981	1	1	Mobile	Dining	64	3.89063945538701	0	
21	TXN008158	402.892383	11.384509310200936	1	0	POS	Grocery	42	6.90172277110148	0	
22	TXN009896	718.042784	9.46331029697384	1	0	Mobile	Grocery	47	9.31933715861329	0	
23	TXN002249	115.818064	105.27345383746449	1	0	Mobile	Electronics	29	2.77167080441769	0	
24	TXN004640	945.803952	22.793667608667025	0	1	Mobile	Electronics	61	2.98379379699239	0	
25	TXN009485	624.129630	9.837252555947607	0	1	Mobile	Dining	34	4.40014947959159	0	
26	TXN004947	741.885660	17.455881749940755	0	1	Web	Grocery	26	4.27116812660835	0	
27	TXN009920	278.555066	12.4187196263278	0	0	Mobile	Retail	21	9.67905790813587	0	

- **Objective 2: Train the Model (XG Boost, DNN, and WOA)**

1. **Train XG Boost**:
    1. Import XG Boost libraries.
    2. Set up the model:
        1. Define hyperparameters (learning rate, max depth, n-estimators, etc.).
    3. Train the model using the training dataset.
    4. Evaluate the model on the test set and record the performance metrics.
2. **Develop DNN Model**:
    1. Import libraries like TensorFlow or PYtorch.
    2. Build the network:
        1. Define input, hidden, and output layers

SAMPLE USER INTERFACE



Status of Implementation

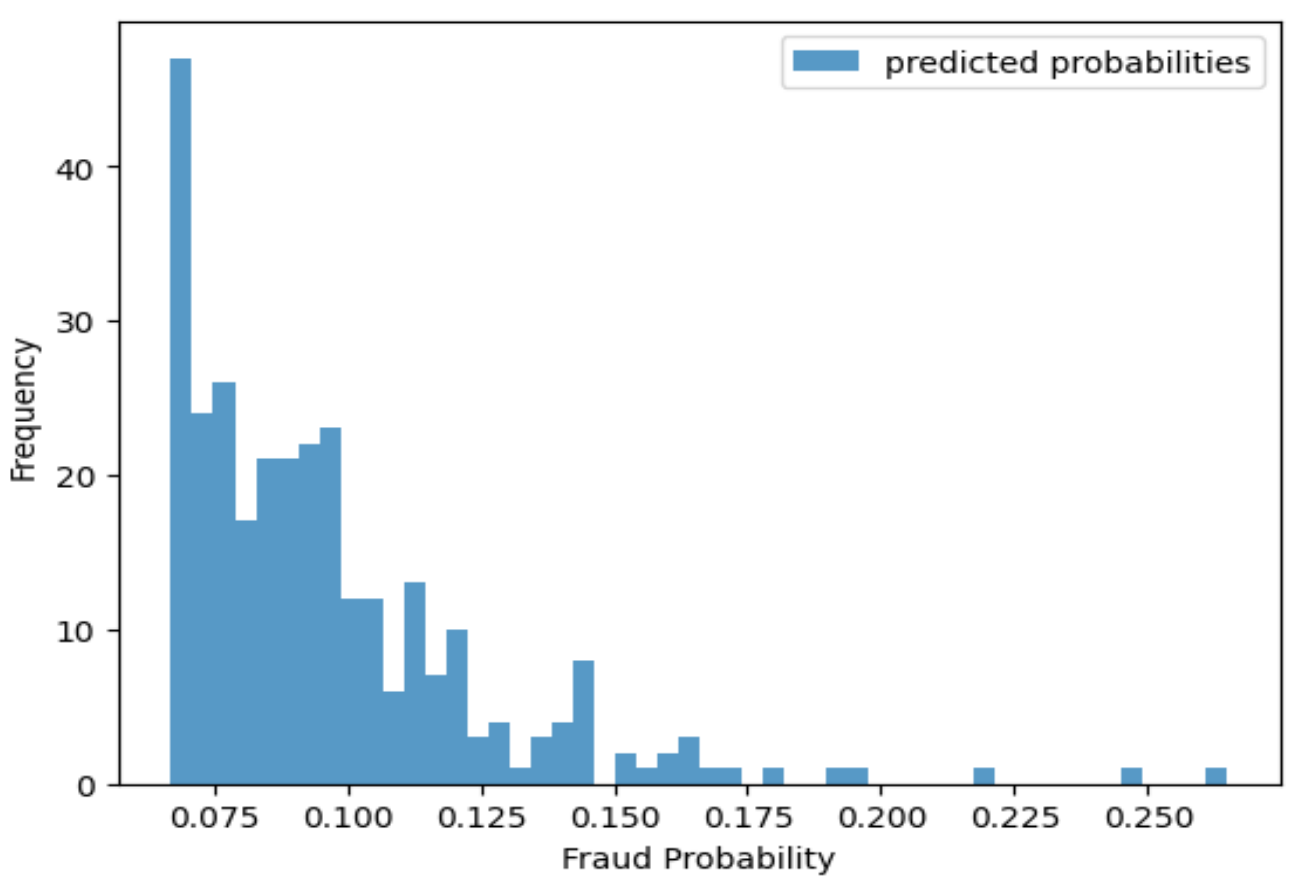
```
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.8966666666666666

Classification Report:
               precision    recall  f1-score   support

           0       0.90      1.00      0.95       269
           1       0.00      0.00      0.00        31

    accuracy                           0.90       300
   macro avg       0.45      0.50      0.47       300
weighted avg       0.80      0.90      0.85       300

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

Accuracy and Classification report

Status of Implementation

```
plt.show()
```



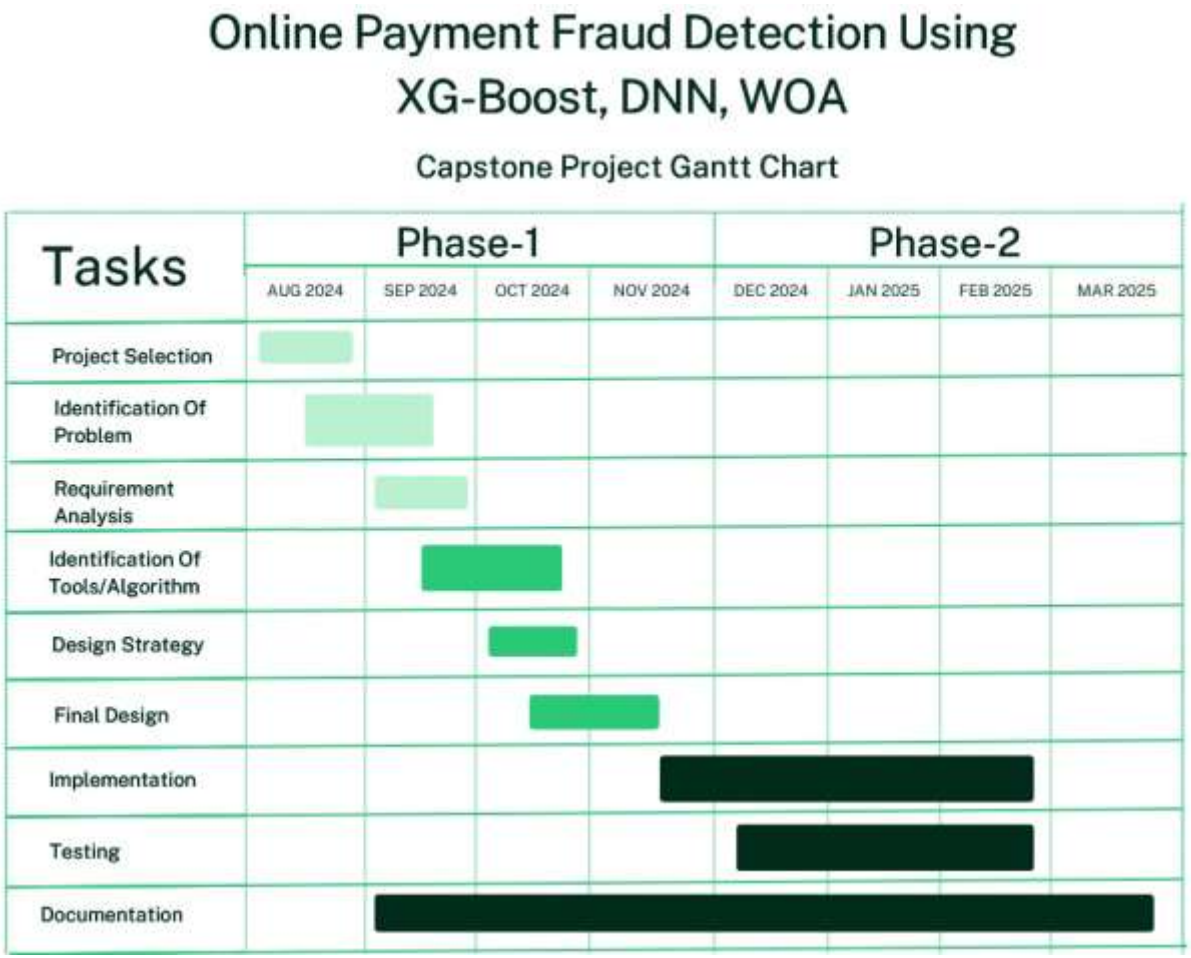
Fraud Probability Plotting

CONTENTS

Conclusion

In this project, we developed a hybrid model for fraud detection in UPI (Unified Payments Interface) transactions using a combination of XG Boost, Deep Neural Networks (DNN), and the Whale Optimization Algorithm (WOA). The goal was to improve the accuracy and efficiency of fraud detection in online payments, particularly focusing on UPI, which has become one of the most widely used digital payment systems in India.

Gantt Chart


Online Payment Fraud Detection Using
XG-Boost, DNN, WOA
Capstone Project Gantt Chart
Tasks
Phase-1
Phase-2
AUG 2024
SEP 2024
OCT 2024
NOV 2024
DEC 2024
JAN 2025
FEB 2025
MAR 2025
Project Selection
Identification Of Problem
Requirement Analysis
Identification Of Tools/Algorithm
Design Strategy
Final Design
Implementation
Testing
Documentation

CONTENTS

Journals

References

[1]l Hamoud, A., Hoenig, A. and Roy, K., 2022. Sentence subjectivity analysis of a political and ideological debate dataset using LSTM and BiLSTM with attention and GRU models. *Journal of King Saud University-Computer and Information Sciences*, 34(10), pp.7974-7987, 2022.

[2] Bouaziz, F., Oulhadj, H., Boutana, D. and Siarry, P., 2019. Automatic ECG arrhythmias classification scheme based on the conjoint use of the multi-layer perceptron neural network and a new improved metaheuristic approach. *IET Signal Processing*, 13(8), pp.726-735.

[3] A. Varshney, R. Kolhe, S. Gatne and V. V. Ingale, "Arrhythmia Classification of ECG Signals Using Undecimated Discrete Wavelet Transform," *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, Mumbai, India, 2022, pp. 1-5.

[4] Babu, D.V., Karthikeyan, C. and Kumar, A., 2020, December. Performance analysis of cost and accuracy for whale swarm and RMSprop optimizer. In *IOP Conference Series: Materials Science and Engineering* (Vol. 993, No. 1, p. 012080). IOP Publishing.