Project Report Format

Team ID : Team-592699

1. INTRODUCTION

1.1 Project Overview

The rapid growth of internet usage and e-commerce has led to a significant increase in online credit/debit card transactions. While this has made transactions more convenient, it has also given rise to a parallel surge in fraudulent activities. Fraud detection is crucial to ensure the security and integrity of online transactions. This project aims to implement machine learning (ML) techniques for the detection of credit/debit card fraud by employing classification algorithms. The chosen algorithms for this project include Decision Tree, Random Forest, Support Vector Machine (SVM), Extra Tree Classifier, and XGBoost Classifier.

Objective:

The primary objective of this project is to develop an effective fraud detection system for online payments using ML algorithms. By leveraging the power of classification models, the project seeks to identify and prevent fraudulent transactions in real-time. The focus is on achieving high accuracy in fraud detection while minimizing false positives, ensuring a robust and reliable system for online payment security.

Methodology:

The proposed method involves training and testing the selected classification algorithms on a dataset containing a large amount of transaction data. The algorithms will learn patterns and features indicative of fraudulent behavior, allowing them to make predictions on new transactions. The iterative training process aims to enhance the accuracy and efficiency of fraud detection.

Algorithm Selection:

The project will utilize Decision Tree, Random Forest, SVM, Extra Tree Classifier, and XGBoost Classifier for their proven effectiveness in classification tasks. Each algorithm will be evaluated based on its performance metrics, and the best-performing model will be selected for further deployment.

Flask Integration:

Once the best model is identified, the project will integrate it into a Flask web application. Flask, a micro web framework for Python, will facilitate the creation of a user-friendly interface for interacting with the fraud detection system. Users can input transaction details, and the system will provide real-time predictions regarding the likelihood of fraud.

1.2 Purpose

The project aims to contribute to the ongoing efforts in enhancing online payment security by leveraging machine learning algorithms. By combining robust classification models with user-friendly interfaces, the system seeks to provide a reliable solution for identifying and preventing fraudulent transactions in the dynamic landscape of online payments.

2. LITERATURE SURVEY

2.1 Existing problem

The existing landscape of online payment systems is marred by a significant and persistent challenge—fraudulent transactions. As the popularity of internet-based transactions grows, so

does the sophistication of fraud attempts. Traditional methods of fraud detection often fall short in accurately identifying and preventing these illicit activities. The limitations in existing approaches underscore the critical need for advanced solutions that can adapt to evolving fraud patterns, necessitating the exploration of machine learning algorithms to bolster the security of online payments.

2.2 References

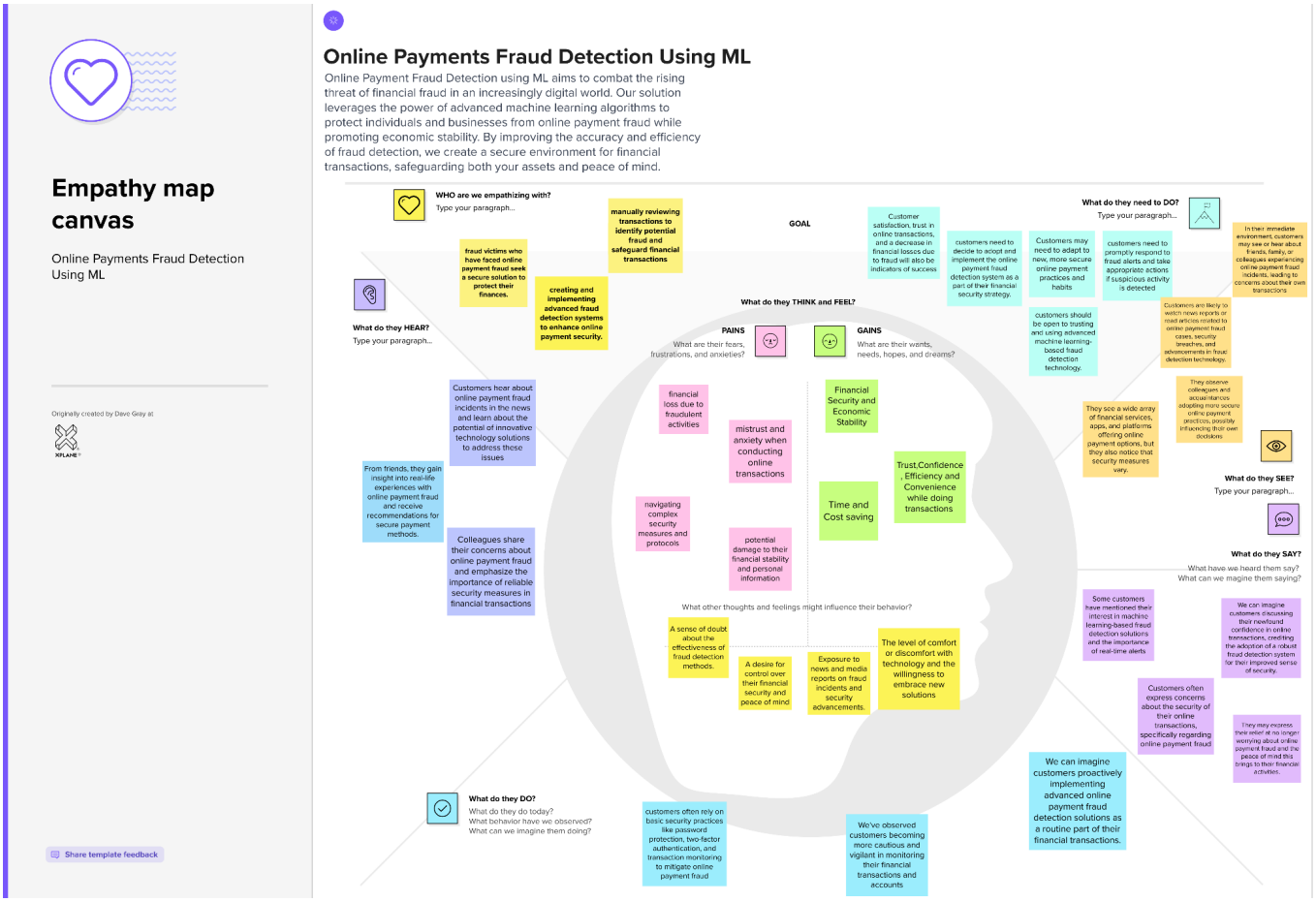
The literature survey draws upon a comprehensive set of references from both academic and industry sources. Key scholarly works include research papers such as "A Survey of Machine Learning Techniques in Credit Card Fraud Detection" by Gupta et al. (2016), providing insights into the application of machine learning in fraud detection. Additionally, industry reports from leading financial institutions and cybersecurity experts contribute to a holistic understanding of the current challenges and potential solutions in the domain of online payment fraud detection.

2.3 Problem Statement Definition

The problem at hand is the inadequacy of existing fraud detection mechanisms in addressing the growing threat of online payment fraud. Traditional rule-based systems often fail to keep pace with the dynamic nature of fraudulent activities, leading to false positives or overlooking subtle patterns indicative of fraud. The aim of this project is to define and implement an efficient fraud detection system using machine learning algorithms. The specific problem statement revolves around developing a system that can accurately and swiftly identify fraudulent transactions in real-time, ensuring the integrity and security of online payment processes.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- **10 minutes** to prepare
- **1 hour** to collaborate
- **2-8 people** recommended

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.

Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.

Open article →

1 Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

The surge in online credit/debit card transactions, driven by the growth of e-commerce, has also led to an increase in fraudulent activities. Current fraud detection methods suffer from accuracy and operational limitations. Our solution, employing classification algorithms such as Decision Trees, Random Forest, SVM, Extra Tree Classifier, and XGBoost, aims to address these issues. By training and testing these models on substantial datasets, we select the most effective one, saving it in a pkl format for convenience. We'll further enhance user access by integrating this solution with Flask and deploying it on IBM Cloud, ultimately offering a precise and user-friendly remedy for the credit/debit card fraud detection problem in the ever-evolving digital landscape.

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

2 Brainstorm

Write down any ideas that come to mind that address your problem statement.

⏱ 10 minutes

TIP

You can double click a sticky note and hit the pencil switch to sketch! Can't start drawing!

Jyoshitha

Multi-Layered Authentication by otp,password,biometrics,two factor authentication etc

geolocation data for real-time user location verification to identify and prevent potentially fraudulent transactions originating from remote or unusual locations

Real-Time Transaction Monitoring by analyzing transaction patterns and triggering alerts for any unusual or suspicious activities

Develop educational content to inform users about the risks of online fraud and guide them on best practices for secure transactions

Jyothika

Implement machine learning algorithms to spot unusual user actions like big or irregular transactions, prompting extra security actions for protection

Design user-friendly interfaces that make it easy for customers to monitor and control their transaction security preferences

Establish a dedicated customer support team to handle fraud-related inquiries and assist users with concerns or suspicious activities

Partner with banks and financial institutions to share data and collaborate on enhancing fraud detection and prevention

3 Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

⏱ 20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as the work progresses.

Authentication and Verification:
1. Multi-Layered Authentication: Combining knowledge, possession, and biometric factors for enhanced security.
2. Geographic Authentication: Using geolocation for user location verification.

Real-Time Monitoring and Alerts:
Real-Time Transaction Monitoring: Continuous analysis of transactions and alerting for anomalies

Advanced Security Measures:
Machine Learning Anomalies: Detecting unusual behavior, especially large transactions, and triggering additional safeguards.
1. Behavioral Biometrics: Continuous user authentication and fraud detection through behavior analysis.

User Involvement and Feedback:
Community Feedback Loop: Allowing users to report potentially fraudulent transactions for system improvement.
1. User Education: Developing educational content for user awareness and secure practices.

Collaboration and Support:
Collaboration with Financial Institutions: Partnering with banks to enhance fraud detection and prevention.
1. Enhanced AI Chatbots: Implementing advanced chatbots for real-time user support.
2. User-Friendly Interfaces: Designing intuitive interfaces for user control over security preferences.

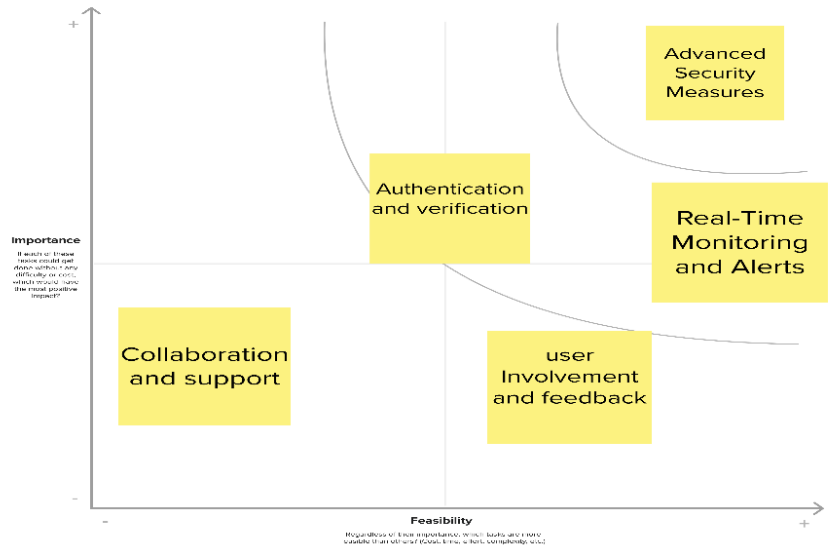
4 Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

⏱ 20 minutes

TIP

Participants can now their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the H key on the keyboard.



After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template →

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

Share template feedback

4. **REQUIREMENT ANALYSIS**

4.1 Functional requirement

The functional requirements for the online payment fraud detection system encompass the capabilities of ingesting, preprocessing, and analyzing large volumes of transactional data, training machine learning models such as Decision Tree, Random Forest, SVM, Extra Tree Classifier, and XGBoost Classifier, offering real-time predictions, conducting model evaluations, integrating with Flask for user-friendly interactions, and deploying the finalized model and application on the IBM Cloud platform

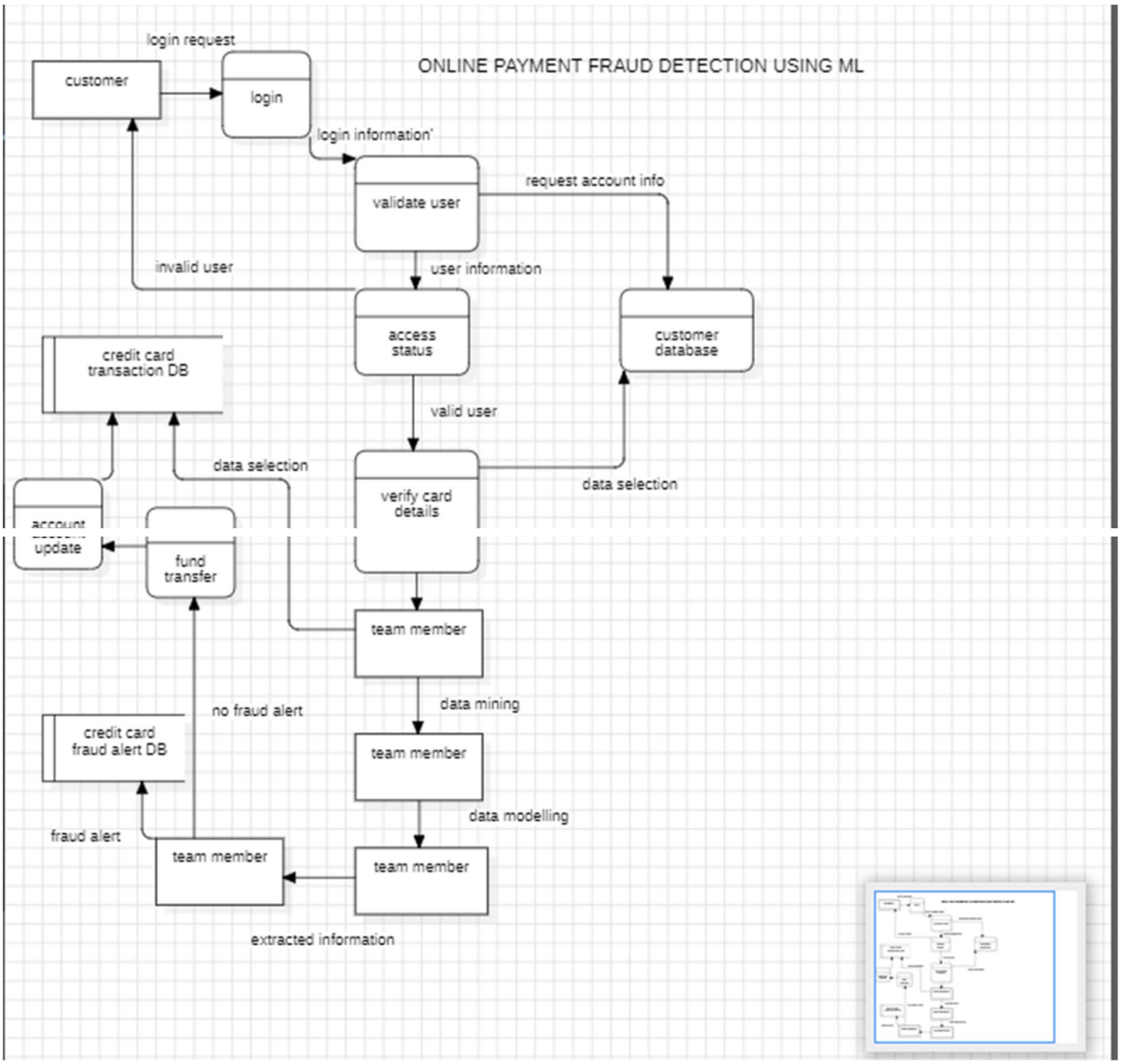
4.2 Non-Functional requirements

Non-functional requirements include ensuring high performance with minimal latency, maintaining a high level of accuracy in fraud detection, scalability to accommodate increasing transaction volumes, adherence to security standards, usability through an intuitive interface, reliability with minimal downtime, compatibility across various devices and browsers, and maintainability for seamless updates and enhancements over time.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

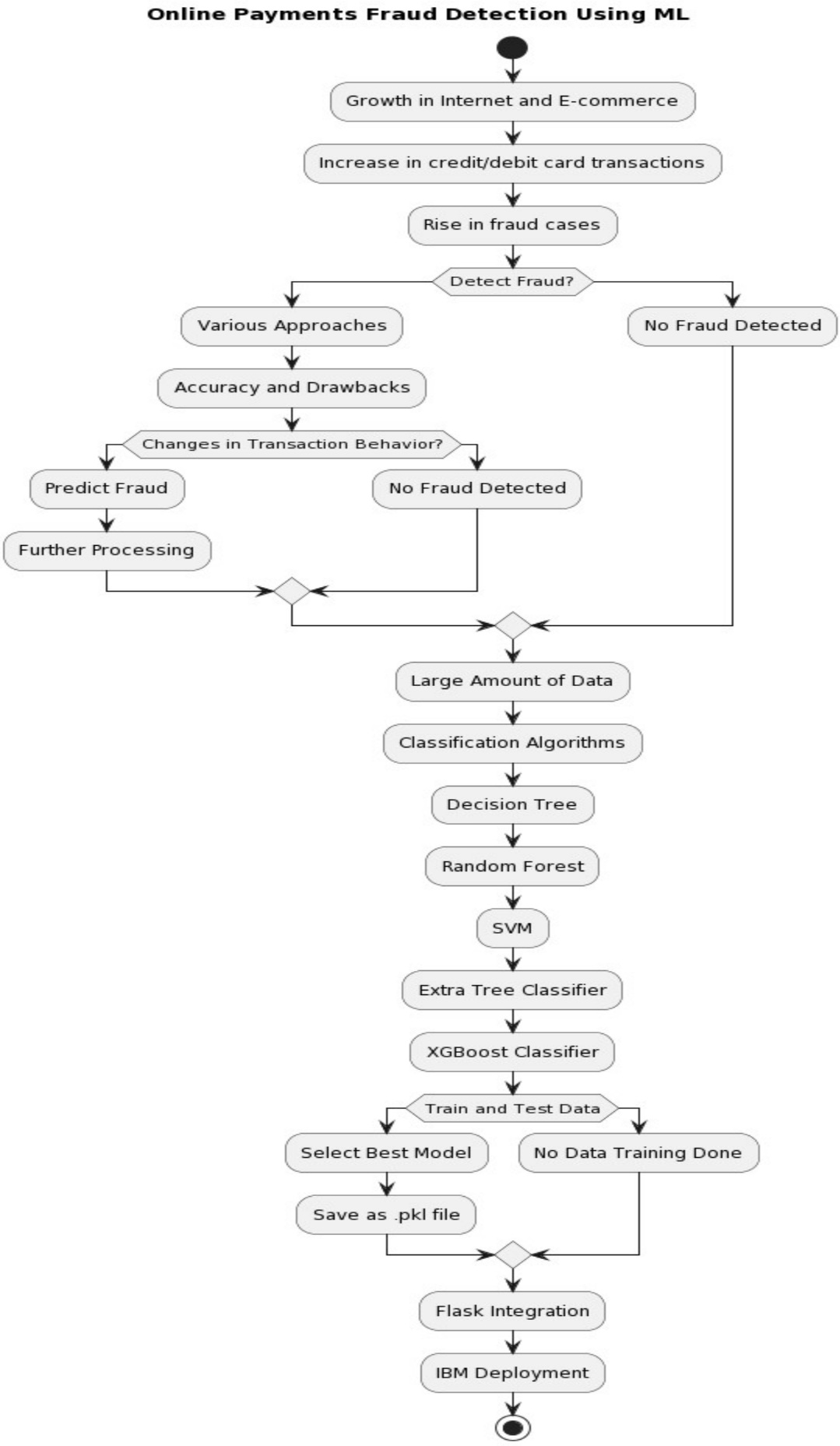
DFD :


ONLINE PAYMENT FRAUD DETECTION USING ML
login request
customer
login
login information'
validate user
request account info
invalid user
user information
access status
customer database
credit card transaction DB
valid user
data selection
verify card details
data selection
account update
fund transfer
team member
data mining
no fraud alert
credit card fraud alert DB
team member
data modelling
fraud alert
team member
team member
extracted information

User Stories :

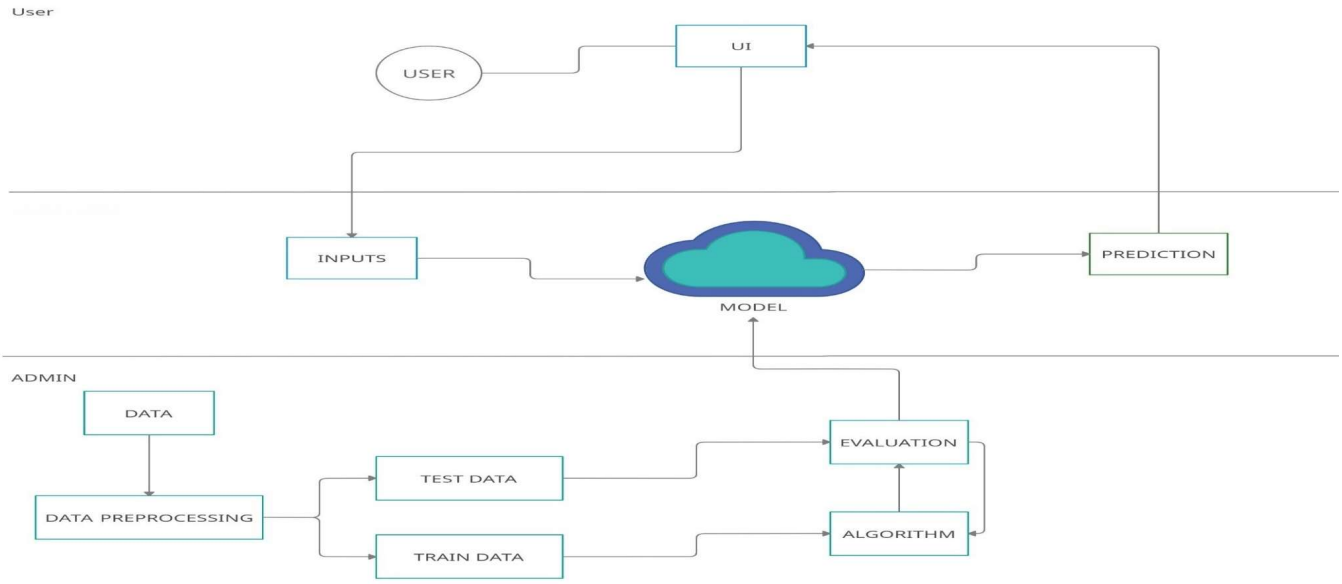
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer	Online payment fraud detection	USN-1	As an customer,I want the online payment system to have fraud detection capabilities using machine learning, so my financial transactions are secure.	The system can detect potential fraudulent activities based on transaction history and behavior patterns.	High	Sprint-1
		USN-2	As a customer, I want the system to notify me in real-time if any suspicious or fraudulent activity is detected during my online payment, so I can take immediate action.	Receive real-time notifications for suspicious transactions, along with details and recommended actions.	High	Sprint-1
		USN-3	As a customer, I want to have the option to review and verify potentially fraudulent transactions, so I can confirm or report them to the payment provider.	The system allows customers to review and verify transactions through the user interface.	High	Sprint-2
		USN-4	As a customer, I want the system to provide guidance on securing my online payment accounts in case of a security breach or fraud, so I can protect my financial information.	The system provides security recommendations and steps to follow in case of a security incident.	Medium	Sprint-2
Administrator	Online payment fraud detection	USN-5	As an administrator, I want to access a dashboard that provides insights into online payment fraud detection, so I can monitor and manage the fraud prevention system.	The dashboard displays real-time fraud detection metrics, historical data, and trends..	High	Sprint-3
		USN-6	As an administrator, I want to be able to configure and customize fraud detection rules and thresholds, so I can adapt the system to changing fraud patterns.	The system allows administrators to adjust fraud detection parameters through a user interface.	High	Sprint-3
Data Analyst	Online payment fraud detection	USN-7	As a data analyst, I want access to the system's historical data and logs for fraud detection, so I can perform data analysis and improve the machine learning models.	Access to a database of historical payment transaction data and fraud-related logs.	Medium	Sprint-5
		USN-8	As a data analyst, I want to receive regular reports on the performance and effectiveness of the fraud detection system, so I can make data-driven recommendations for improvements.	Scheduled reports on fraud detection accuracy, false positives, and other relevant metrics.	Medium	Sprint-5
		USN-9	As a data analyst, I want to collaborate with data scientists to refine and enhance the machine learning models used for fraud detection.	Access to data and tools for collaboration with data scientists.	Medium	Sprint-6

5.2 **Solution Architecture**


Online Payments Fraud Detection Using ML
Growth in Internet and E-commerce
Increase in credit/debit card transactions
Rise in fraud cases
Detect Fraud?
Various Approaches
No Fraud Detected
Accuracy and Drawbacks
Changes in Transaction Behavior?
Predict Fraud
No Fraud Detected
Further Processing
Large Amount of Data
Classification Algorithms
Decision Tree
Random Forest
SVM
Extra Tree Classifier
XGBoost Classifier
Train and Test Data
Select Best Model
No Data Training Done
Save as .pkl file
Flask Integration
IBM Deployment

6. PROJECT PLANNING & SCHEDULING

6.1 **Technical Architecture**



6.2 **Sprint Planning & Estimation**

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Project Initiation	USN-1	Define project scope,and objectives.	8	Medium	Jyoshitha, Jyothika
Sprint-1	Risk Assessment	USN-2	Identify potential risks and challenges.	6	Medium	Jyoshitha, Jyothika
Sprint-1	Data Collection	USN-3	Identifying and reviewing dataset for analysis.	6	Medium	Jyoshitha, Jyothika
Sprint-2	Data Preprocessing	USN-4	Clean,transform and prepare data for modeling.	20	High	Jyoshitha, Jyothika
Sprint-3	Model Selection	USN-5	Choose classification algorithms(Decision tree,random forest,svm,Extra tree classifier,and xg boost classifier).	10	High	Jyoshitha, Jyothika
Sprint-3	Model training and Evaluation	USN-6	Train the selected models with data and evaluate model performance(accuracy,precision).	10	High	Jyoshitha, Jyothika
Sprint-4	Best Model selection and Model Serialization	USN-7	Identify the best performing model and save the best model in .pkl format.	20	High	Jyoshitha, Jyothika
Sprint-5	Flask Integration	USN-8	Develop a web application using flask for model deployment.	20	High	Jyoshitha, Jyothika
Sprint-6	Project Review and closure	USN-9	Review project progress and make necessary adjustments and document project outcomes.	20	Medium	Jyoshitha, Jyothika

6.3 **Sprint Delivery Schedule**

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)
Sprint-1	20	2 Days	28 Oct 2023	29 Oct 2023	20
Sprint-2	20	2 Days	30 Oct 2023	31 oct 2023	20
Sprint-3	20	3 Days	01 Nov 2023	03 Nov 2023	20
Sprint-4	20	3 Days	01 Nov 2023	03 Nov 2023	20
Sprint-5	20	2 Days	04 Nov 2023	05 Nov 2023	20
Sprint-6	20	1 Day	06 Nov 2023	06 Nov 2023	20

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1: **Real-time Fraud Prediction using Flask Interface**

Overview: The first feature involves integrating a Flask web application to provide users with a real-time interface for predicting fraudulent transactions. The Flask app interacts with the trained machine learning model to process user-inputted transaction details and promptly returns the likelihood of fraud.

7.2 Feature 2 :**IBM Cloud Deployment**

Overview: The second feature involves deploying the finalized machine learning model and the Flask web application on the IBM Cloud platform for accessibility and scalability. This enables users to access the fraud detection system remotely.

8. PERFORMANCE TESTING

8.1 Performace Metrics

Accuracy:

Definition: The ratio of correctly predicted transactions (both fraud and non-fraud) to the total transactions.

Importance: Accuracy is a fundamental metric to evaluate the overall effectiveness of the machine learning model in correctly identifying fraudulent transactions.

Precision:

Definition: Precision measures the accuracy of the positive predictions, indicating the proportion of actual fraudulent transactions among those predicted as fraudulent.

Importance: High precision is essential to minimize false positives, ensuring that flagged transactions are indeed likely to be fraudulent.

Recall (Sensitivity):

Definition: Recall measures the ability of the model to correctly identify all actual fraudulent transactions, indicating the proportion of actual fraud detected.

Importance: High recall is crucial to minimize false negatives, ensuring that the system captures a significant portion of actual fraudulent activities.

F1 Score:

Definition: The F1 Score is the harmonic mean of precision and recall, providing a balanced measure of a model's overall performance.

Importance: F1 Score is particularly useful when there is an uneven class distribution, providing a comprehensive evaluation of the model's effectiveness.

Execution Time:

Definition: The time taken for the system to process a transaction and provide a fraud prediction.

Importance: Low execution time is crucial for real-time fraud detection, ensuring swift responses to incoming transactions without causing delays.

Resource Utilization:

Definition: Measures the utilization of system resources, including CPU, memory, and network bandwidth, during the execution of the fraud detection system.

Importance: Efficient resource utilization ensures that the system can handle a high volume of transactions without straining hardware resources.

Scalability:

Definition: The system's ability to handle an increasing number of transactions while maintaining performance.

Importance: Scalability is crucial to accommodate the growth in transaction volume over time without compromising the system's responsiveness.

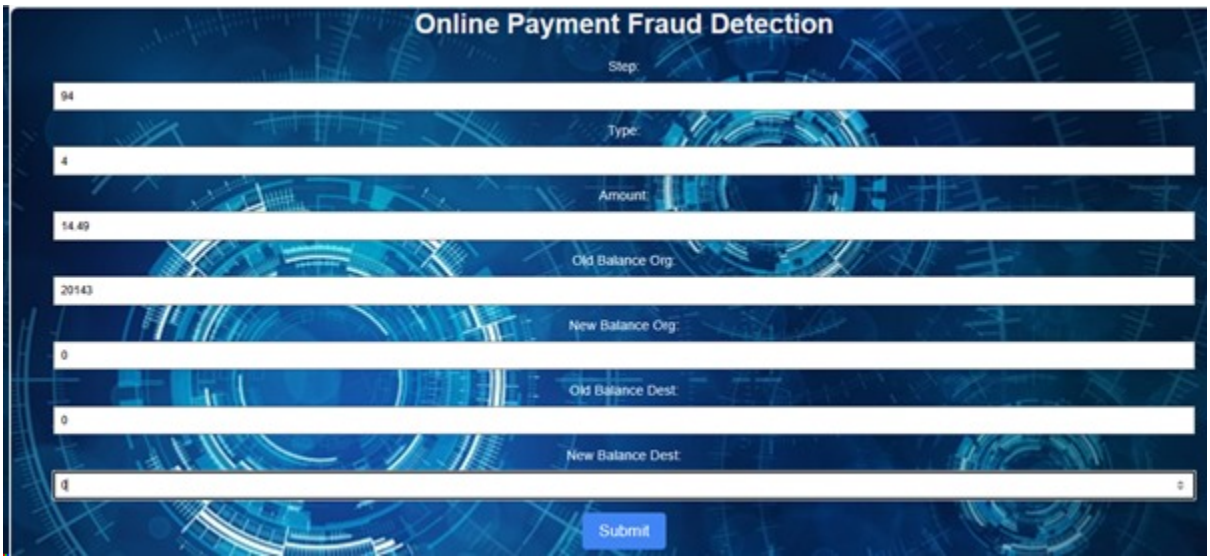
User Satisfaction:

Definition: User feedback and satisfaction with the usability and responsiveness of the Flask web interface.

Importance: Positive user experiences contribute to the overall success of the fraud detection system.

9. RESULTS

9.1 Output Screenshots



10. ADVANTAGES

- Improved Fraud Detection:

The utilization of machine learning algorithms enhances the accuracy and efficiency of fraud detection, reducing false positives and negatives.

- Real-time Prediction:

The integration of Flask enables users to receive real-time predictions on the likelihood of fraud, allowing for prompt action.

- Scalability:

Deployment on the IBM Cloud platform ensures scalability, accommodating a growing number of transactions without compromising performance.

- User-friendly Interface:

The Flask interface provides an intuitive and accessible platform for users to interact with the fraud detection system.

- Cloud Deployment Benefits:

Utilizing IBM Cloud ensures accessibility, reliability, and efficient resource management, contributing to overall system robustness.

Disadvantages:

- Dependency on Quality of Data:

The effectiveness of the machine learning models is highly dependent on the quality and

representativeness of the training data. Biases in the data can impact the accuracy of predictions.

- Initial Setup Complexity:

Implementing the system may require expertise in machine learning, Flask, and cloud deployment, making the initial setup complex for some users.

- Resource Intensiveness:

Machine learning algorithms, especially when dealing with large datasets, can be resource-intensive, requiring significant computational power.

- Continuous Monitoring Required:

The system requires continuous monitoring and updates to adapt to evolving fraud patterns and maintain optimal performance.

11. CONCLUSION

In conclusion, the online payment fraud detection system employing machine learning algorithms, Flask integration, and IBM Cloud deployment offers an effective solution to the escalating challenge of online fraud. The advantages, such as improved accuracy, real-time predictions, scalability, and user-friendly interface, outweigh the disadvantages associated with data dependency, initial setup complexity, resource intensiveness, and the need for continuous monitoring. The system represents a significant step forward in enhancing the security of online transactions, providing users and businesses with a reliable and efficient means of identifying and preventing fraudulent activities.

12. FUTURE SCOPE

The future scope of this project includes :

- Enhancing Model Robustness - Continuous refinement of machine learning models to adapt to emerging fraud tactics and improve overall robustness.
- Integration of Advanced ML Techniques
- Incident Analysis and Reporting
- User Authentication Enhancements
- Global Collaboration
- Integration with Blockchain