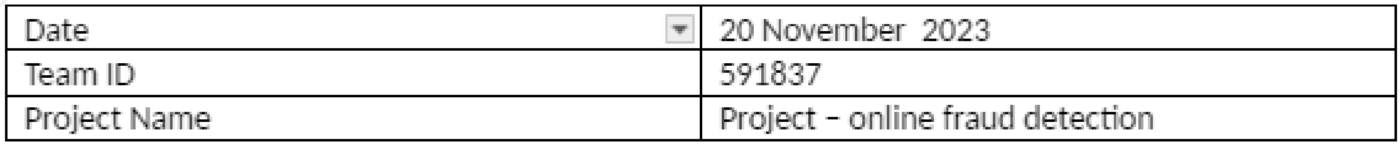
Project report



Date	20 November 2023
Team ID	591837
Project Name	Project – online fraud detection

Online Payments Fraud Detection using ML

1 .INTRODUCTION

1.1 Project Overview:

The primary objective of this project is to develop a robust and efficient online fraud detection system using machine learning (ML) techniques. The system aims to proactively identify and prevent unauthorized access and fraudulent activities within an online platform.

1. **User Authentication Analysis:**
   . Develop algorithms to analyze user authentication patterns.
   . Implement multi-factor authentication for enhanced security.
   . Monitor login attempts and detect anomalies or suspicious patterns.
2. **Behavioral Analysis:**
   . Utilize machine learning to analyze user behavior patterns.
   . Establish a baseline for normal user behavior.
   . Detect deviations from the baseline that may indicate fraudulent activities.
3. **Transaction Monitoring:**
   . Implement real-time monitoring of transactions.
   . Analyze transaction patterns and identify unusual activities.
   . Set thresholds for transaction amounts and frequency to trigger alerts.
4. **Geographical Analysis:**
   . Incorporate geolocation data to identify unusual login locations.
   . Implement rules to flag logins from unexpected or high-risk locations.
   . Geo-fencing to restrict access from certain geographical areas.
5. **Device Fingerprinting:**
   . Implement device fingerprinting techniques to uniquely identify devices.
   . Track and analyze changes in device characteristics.
   . Identify and flag multiple account logins from the same device
6. **Machine Learning Models:**
   . Train supervised machine learning models using historical data.
   . Explore anomaly detection algorithms for identifying unusual patterns.
   . Regularly update and retrain models to adapt to evolving fraud tactics.
7. **Alerts and Notifications:**
   . Set up a notification system for real-time alerts on suspicious activities.
   . Classify alerts based on the level of risk.
   . Provide detailed information on flagged activities for further investigation.
8. **User Education and Communication:**
   . Develop features to educate users about online security best practices.
   . Communicate with users when suspicious activities are detected.
   . Provide guidance on securing their accounts and reporting potential issues.
9. **Regulatory Compliance:**
   . Ensure that the system complies with relevant data protection and privacy regulations.

. Implement features for auditing and reporting to meet compliance requirements.

10. **Continuous Improvement:**

. Establish a feedback loop for continuous improvement.
. Collect feedback from flagged activities and use it to enhance the system.
. Regularly review and update the system to adapt to new fraud patterns.

11. **Integration with Other Systems:**

. Integrate the fraud detection system with other security systems.
. Share information with threat intelligence databases.
. Collaborate with law enforcement when necessary.

12. **Scalability and Performance:**

. Design the system to handle a growing user base and increasing data volume.
. Optimize algorithms and infrastructure for performance.
. Implement load balancing and scalability measures.

Timeline:

Define a project timeline with key milestones, including data collection, model development, system integration, testing, and deployment.

Risks and Mitigations:

Identify potential risks such as data privacy concerns, model performance degradation, or system vulnerabilities. Develop mitigation strategies for each identified risk.

Release Plan:

Specify the release schedule for different components of the project, including any planned incremental releases.

Success Criteria:

Define success criteria based on the achievement of key performance indicators (KPIs) and the system's ability to effectively detect and prevent fraudulent activities.

1.2 Purpose:

The purpose of implementing Online Fraud Detection Using Machine Learning (ML) is to enhance online platform security by proactively identifying and preventing fraudulent activities. This improves user trust, operational efficiency, and compliance with regulations, while also providing data-driven insights and adaptability to evolving threats. Ultimately, it aims to reduce financial losses, minimize disruptions for genuine users, and streamline fraud prevention processes.

2 LITERATURE SURVEY

Existing problem:

Online fraud detection using machine learning (ML) faces several challenges, including inaccuracies in identifying fraud (false positives and negatives), data imbalances affecting model accuracy, struggles in adapting to evolving fraud patterns, opacity in model decision-making, resource-intensive computations, concerns about regulatory compliance, potential drift in model performance over time, and difficulties in integrating with existing platforms. To tackle these issues, a comprehensive strategy is needed. This involves refining models, ensuring diverse and up-to-date datasets, transparent decision-making processes, and regular system updates to effectively combat emerging fraud threats while adhering to compliance standards and facilitating seamless integration with existing platforms.

2.1 References:

https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90

https://c4model.com/

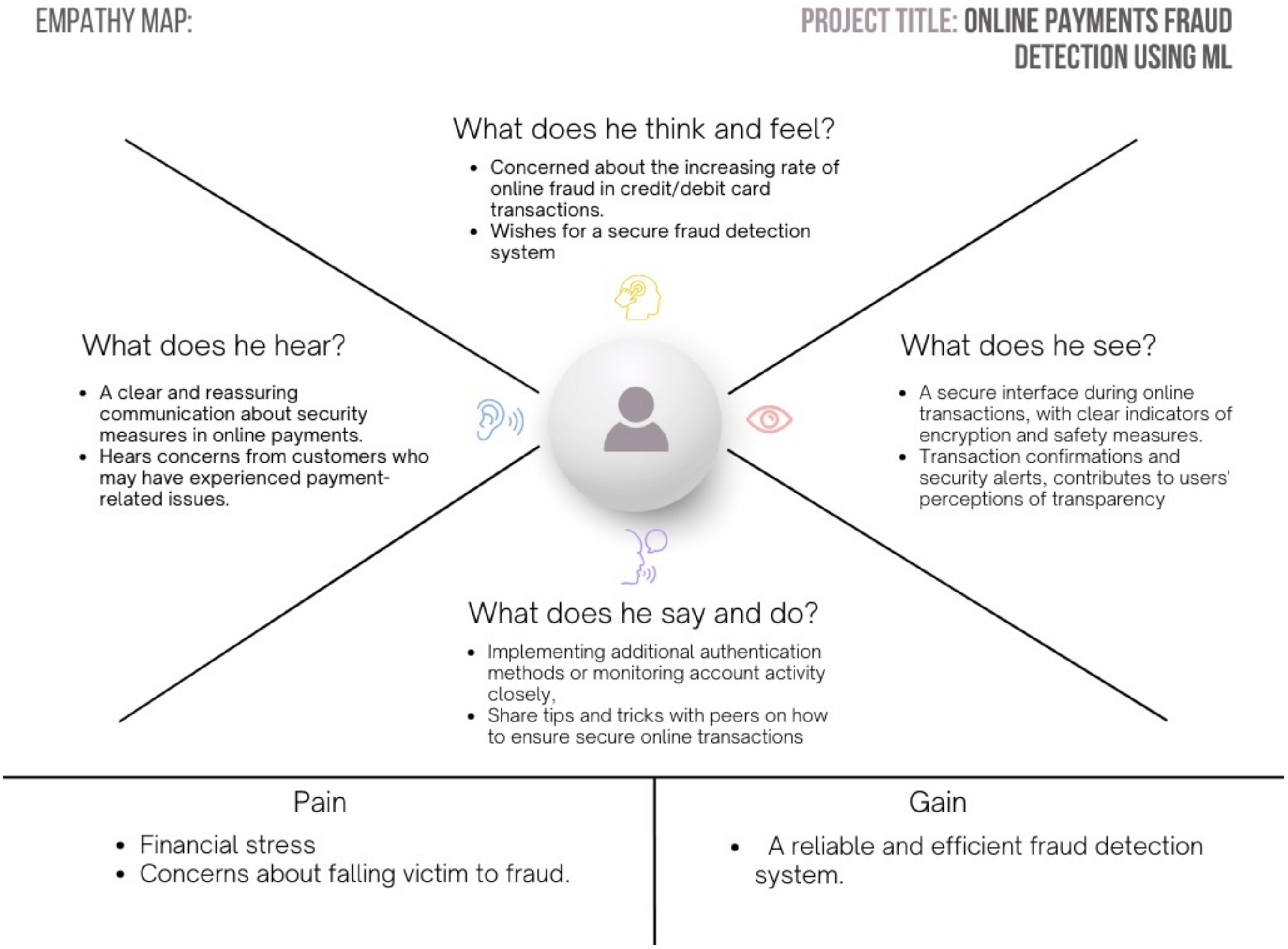
https://developer.ibm.com/patterns/online-order-processing-system-during-pandemic/

2.2 Problem Statement Definition:

The problem statement for Online Fraud Detection Using Machine Learning (ML) encompasses various challenges that need to be addressed for an effective fraud detection system. These challenges include the occurrence of false positives and negatives, imbalanced datasets impacting model accuracy, the need for adaptability to emerging fraud patterns, transparency in model decision-making, computational resource intensity, compliance with regulations, potential model drift over time, and integration challenges with existing platforms. The problem is to develop a fraud detection solution that minimizes false positives and negatives, ensures the model's adaptability to evolving threats, provides transparent and explainable decisions, optimizes computational efficiency, complies with regulatory standards, mitigates model drift, and seamlessly integrates with online platforms. Addressing these aspects is crucial for building a reliable and efficient Online Fraud Detection system using ML.

3.IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas:



EMPATHY MAP:

PROJECT TITLE: ONLINE PAYMENTS FRAUD DETECTION USING ML

What does he think and feel?

- Concerned about the increasing rate of online fraud in credit/debit card transactions.
- Wishes for a secure fraud detection system

What does he hear?

- A clear and reassuring communication about security measures in online payments.
- Hears concerns from customers who may have experienced payment-related issues.

What does he see?

- A secure interface during online transactions, with clear indicators of encryption and safety measures.
- Transaction confirmations and security alerts, contributes to users' perceptions of transparency

What does he say and do?

- Implementing additional authentication methods or monitoring account activity closely,
- Share tips and tricks with peers on how to ensure secure online transactions

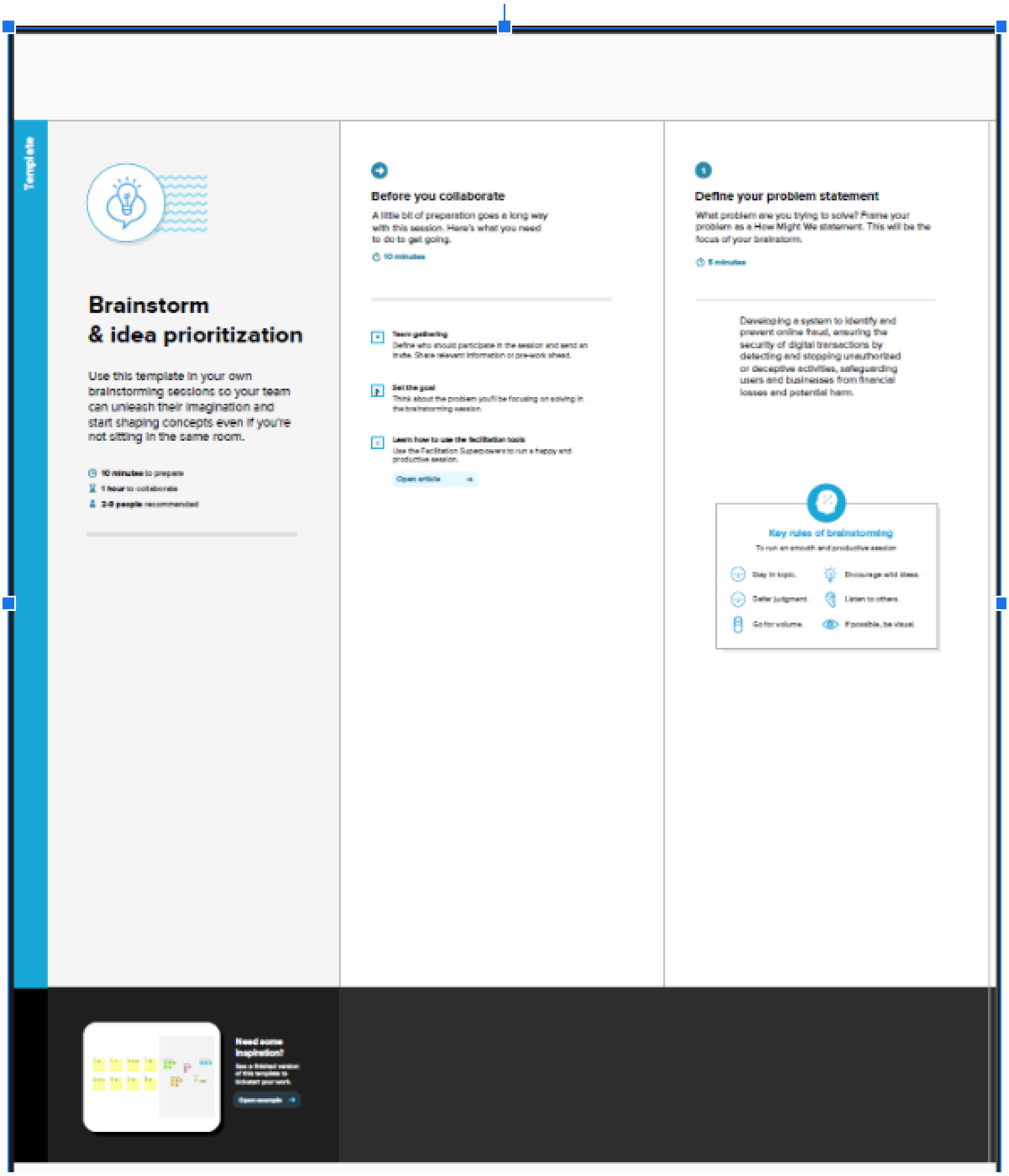
Pain

- Financial stress
- Concerns about falling victim to fraud.

Gain

- A reliable and efficient fraud detection system.

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
- **3-8 people** recommended

Need some inspiration?

See a finished version of this template to kickstart your work.

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.

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

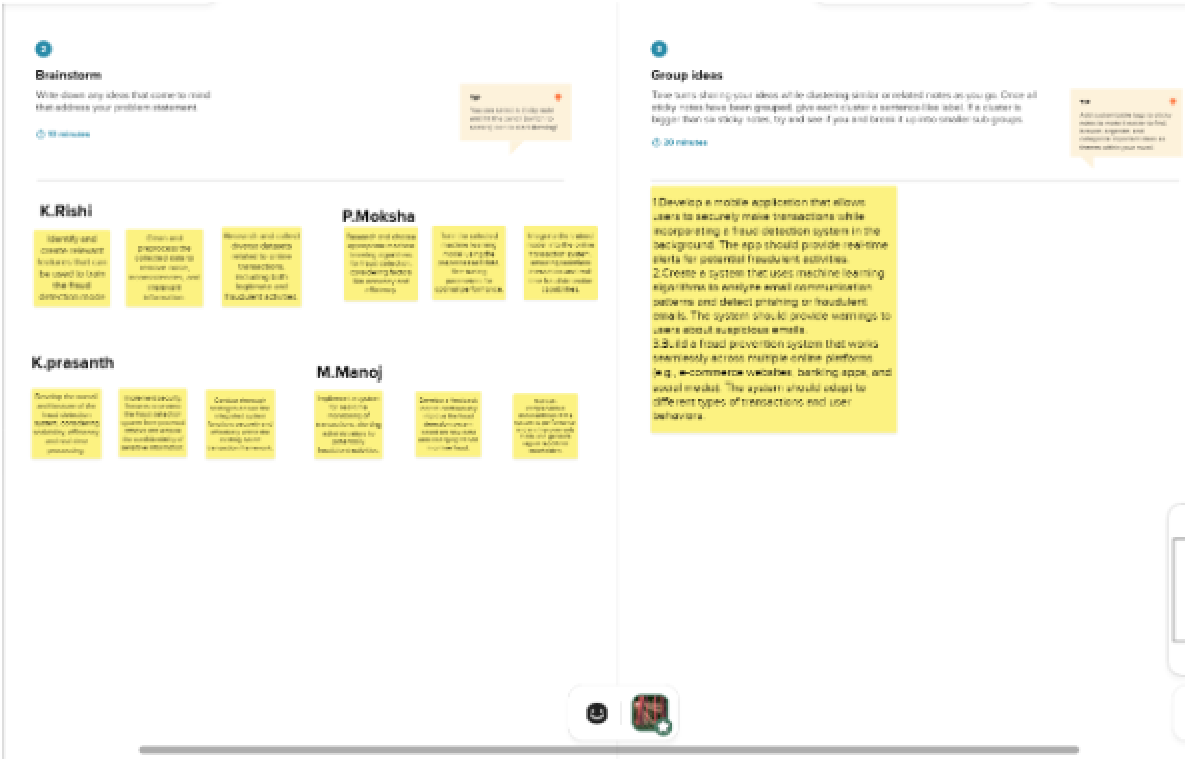
Developing a system to identify and prevent online fraud, ensuring the security of digital transactions by detecting and stopping unauthorized or deceptive activities, safeguarding users and businesses from financial losses and potential harm.

Key rules of brainstorming

To run an smooth and productive session

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

Step-2: Brainstorm, Idea Listing and Grouping



Brainstorm

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

10 minutes

TIP

[illegible]

K.Rishi

Identify and create relevant features that can be used to train the fraud detection model

Clean and preprocess the collected data to remove noise, inconsistencies, and irrelevant information.

Research and collect diverse datasets related to online transactions, including both legitimate and fraudulent activities.

P.Moksha

[illegible]

K.prasanth

[illegible]

M.Manoj

[illegible]

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.

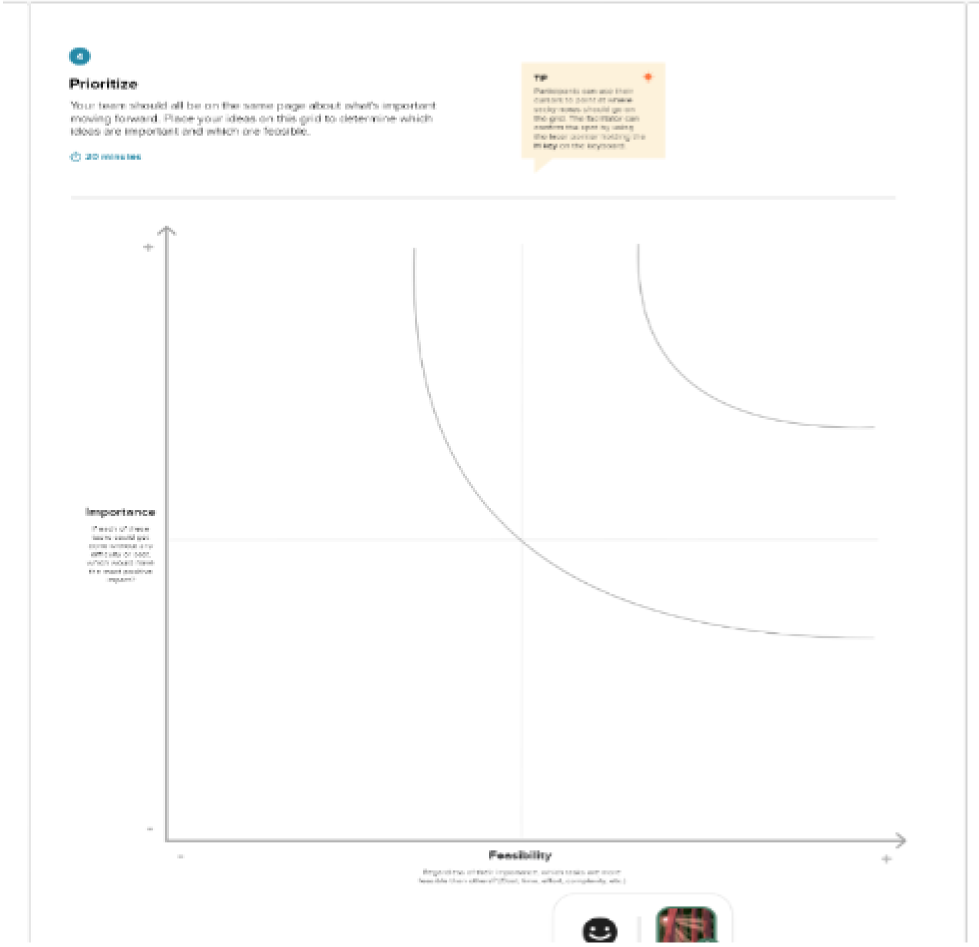
20 minutes

TIP

[illegible]

1.Develop a mobile application that allows users to securely make transactions while incorporating a fraud detection system in the background. The app should provide real-time alerts for potential fraudulent activities.
2.Create a system that uses machine learning algorithms to analyze email communication patterns and detect phishing or fraudulent emails. The system should provide warnings to users about suspicious emails.
3.Build a fraud prevention system that works seamlessly across multiple online platforms (e.g., e-commerce websites, banking apps, and social media). The system should adapt to different types of transactions and user behaviors.

Step-3: Idea Prioritization



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 use 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.

+
Importance
If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?
-
-
Feasibility
Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)
+

4 REQUIREMENT ANALYSIS

4.1 Functional requirement:

Real-time Transaction Monitoring

DESCRIPTION: The system must continuously monitor online transactions in real-time to identify potentially fraudulent activities.

ACCEPTANCE CRITERIA:

The system should process transactions as they occur, providing immediate feedback.

DESCRIPTION: Real-time monitoring should cover various transaction types and channels. Anomaly Detection

Implement machine learning algorithms for anomaly detection to identify deviations from normal transaction patterns.

ACCEPTANCE CRITERIA:

The system should define thresholds for normal behavior based on historical data.

DESCRIPTION: Anomalies should trigger alerts for further investigation. Model Training and Adaptation

The system must undergo regular model training using updated datasets to adapt to changing fraud patterns.

ACCEPTANCE CRITERIA:

The model should be retrained at defined intervals, incorporating the latest transaction data.

Adaptive learning mechanisms should be in place to dynamically adjust to emerging fraud tactics.

User Interface for Administrators

DESCRIPTION:

Develop an intuitive user interface for administrators to monitor and manage fraud detection activities.

ACCEPTANCE CRITERIA:

The interface should provide real-time visualization of flagged transactions and their status.

Include features for administrators to drill down into transaction details and apply manual interventions.

Alerting and Notification System

DESCRIPTION:

Implement an alerting system to notify administrators of potentially fraudulent transactions.

ACCEPTANCE CRITERIA:

Alerts should be generated in real-time when anomalies surpass predefined thresholds.

Notifications should include relevant transaction details for quick decision-making.

Performance Metrics Tracking

DESCRIPTION:

Define and track key performance metrics to evaluate the effectiveness of the fraud detection system.

ACCEPTANCE CRITERIA:

Metrics such as precision, recall, false positive rate, and F1 score should be monitored regularly.

Establish a reporting mechanism for administrators to review performance.

4.2 Non-Functional requirements:

Performance

Responce Time:

The system should provide real-time responses to flagged transactions, with a response time of no more than 2 seconds.

SCALABILITY:

The fraud detection system must scale horizontally to accommodate increasing transaction volumes without significant performance degradation.

Reliability

Availability:

The system should be available 99.9% of the time to ensure continuous fraud monitoring.

FAULT TOLERENCE:

Implement mechanisms to handle system failures gracefully, ensuring minimal impact on fraud detection capabilities.
Security

DATA ENCRYPTION:

All sensitive transaction and user data must be encrypted during transmission and storage to ensure confidentiality.

ACCESS CONTROL:

Implement robust access controls to restrict system access based on roles and responsibilities, preventing unauthorized manipulation of data.

Adaptability

MODEL APTABILITY:

The ML model should be adaptable to changes in transaction patterns and emerging fraud tactics without requiring extensive manual intervention.

CONFIGURABILITY:

Provide configuration options for anomaly detection thresholds, allowing administrators to fine-tune the system according to specific needs.

Usability

USER INTERFACE INTUITIVENESS:

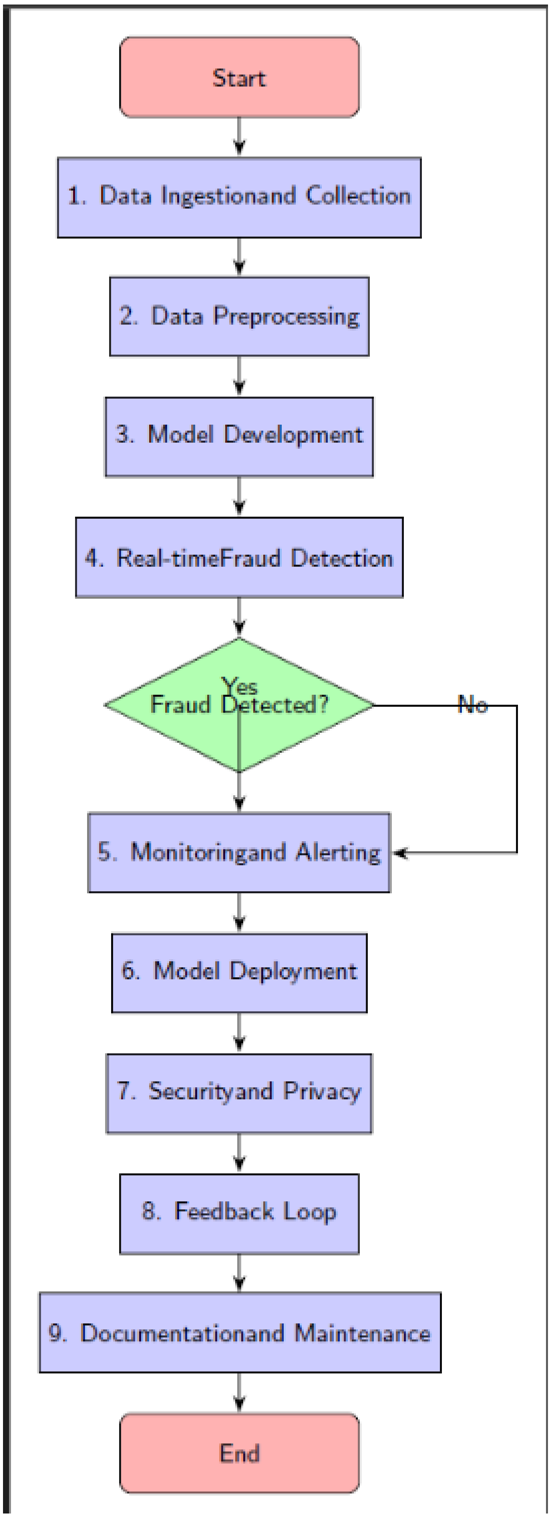
The administrator interface should be intuitive and user-friendly, requiring minimal training for effective use.

DOCUMENTATION:

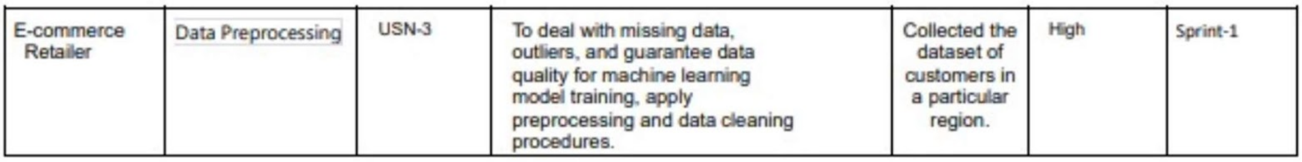
Comprehensive documentation should be provided for system administrators, detailing system functionalities and troubleshooting procedures.

5 PROJECT DESIGN

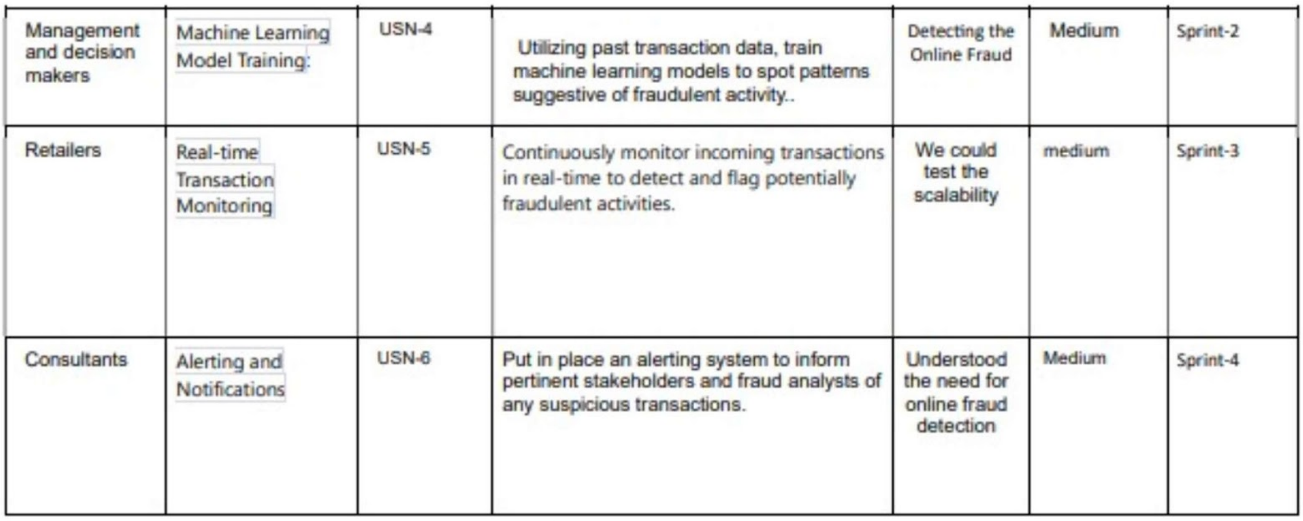
5.1 Data Flow Diagrams & User Stories:


Start
1. Data Ingestionand Collection
2. Data Preprocessing
3. Model Development
4. Real-timeFraud Detection
Fraud Detected?
Yes
No
5. Monitoringand Alerting
6. Model Deployment
7. Securityand Privacy
8. Feedback Loop
9. Documentationand Maintenance
End

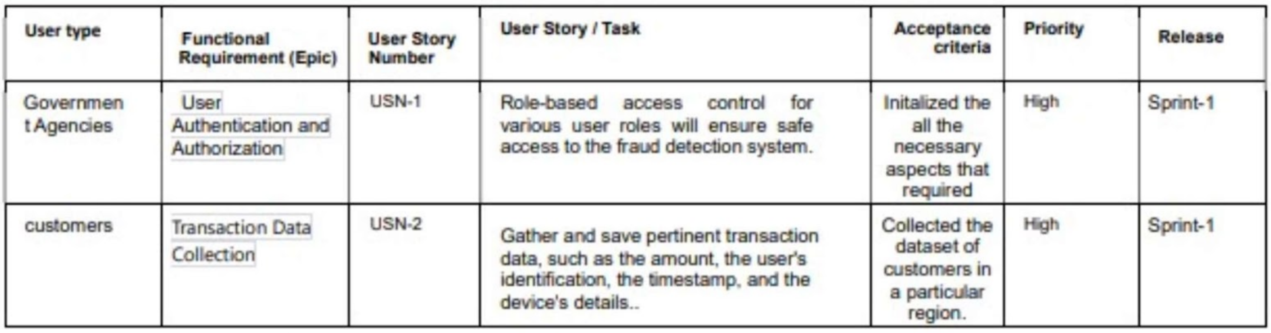
User story:



E-commerce Retailer	Data Preprocessing	USN-3	To deal with missing data, outliers, and guarantee data quality for machine learning model training, apply preprocessing and data cleaning procedures.	Collected the dataset of customers in a particular region.	High	Sprint-1

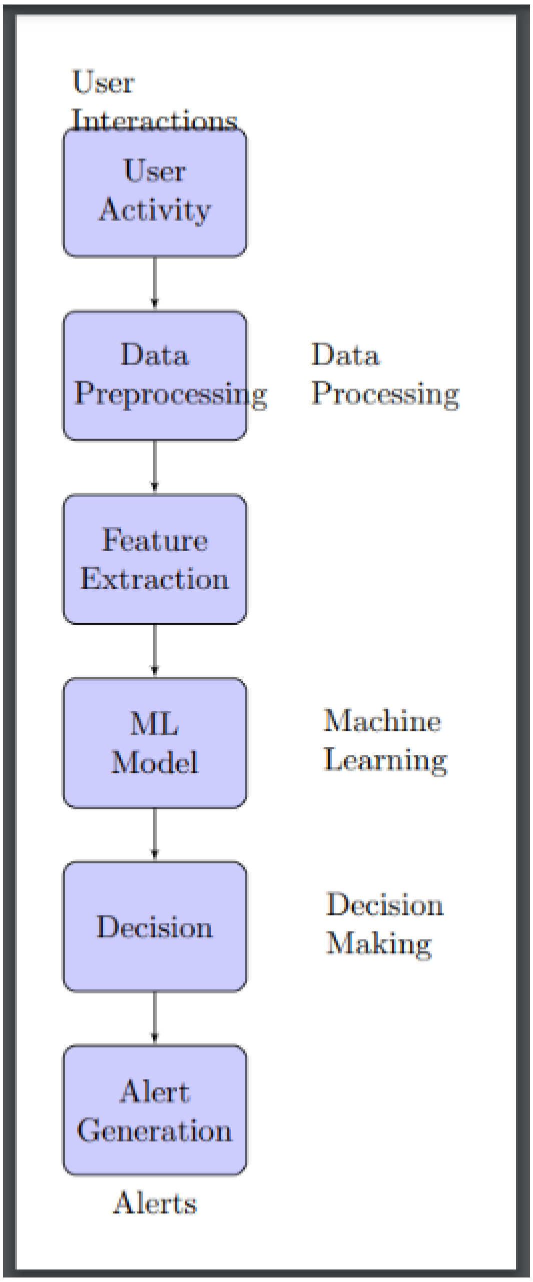


Management and decision makers	Machine Learning Model Training:	USN-4	Utilizing past transaction data, train machine learning models to spot patterns suggestive of fraudulent activity..	Detecting the Online Fraud	Medium	Sprint-2
Retailers	Real-time Transaction Monitoring	USN-5	Continuously monitor incoming transactions in real-time to detect and flag potentially fraudulent activities.	We could test the scalability	medium	Sprint-3
Consultants	Alerting and Notifications	USN-6	Put in place an alerting system to inform pertinent stakeholders and fraud analysts of any suspicious transactions.	Understood the need for online fraud detection	Medium	Sprint-4



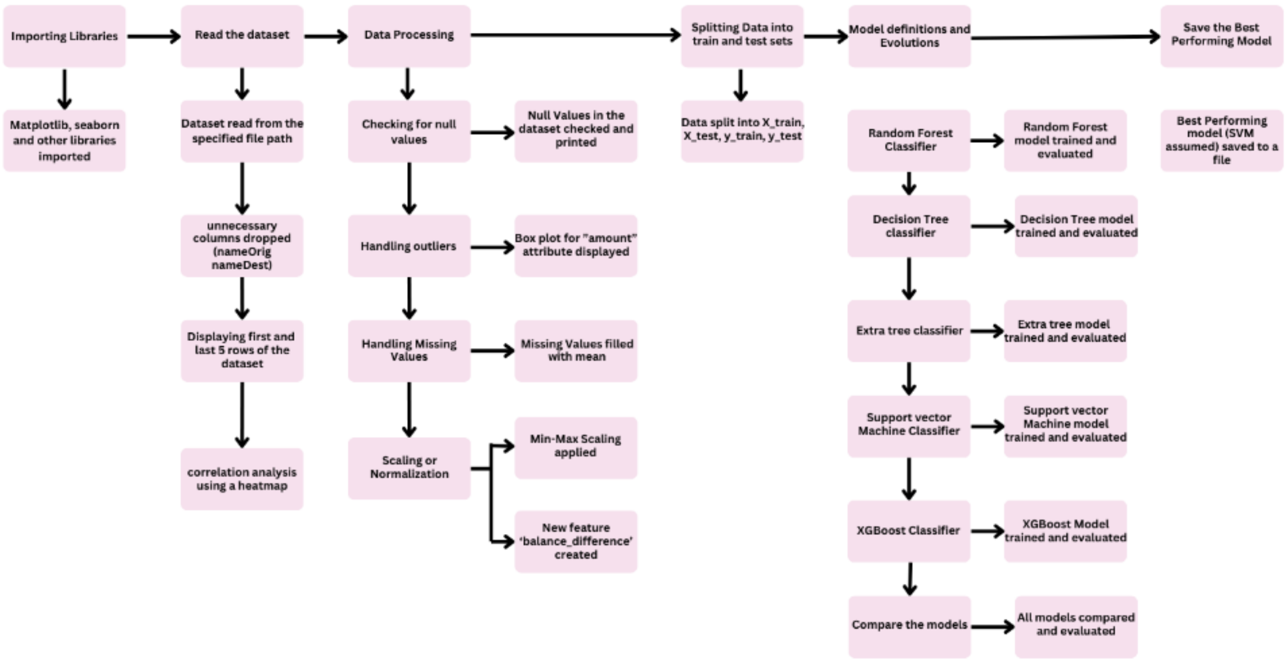
User type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Governmen t Agencies	User Authentication and Authorization	USN-1	Role-based access control for various user roles will ensure safe access to the fraud detection system.	Initalized the all the necessary aspects that required	High	Sprint-1
customers	Transaction Data Collection	USN-2	Gather and save pertinent transaction data, such as the amount, the user's identification, the timestamp, and the device's details..	Collected the dataset of customers in a particular region.	High	Sprint-1

5.2 Solution Architecture:



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture:

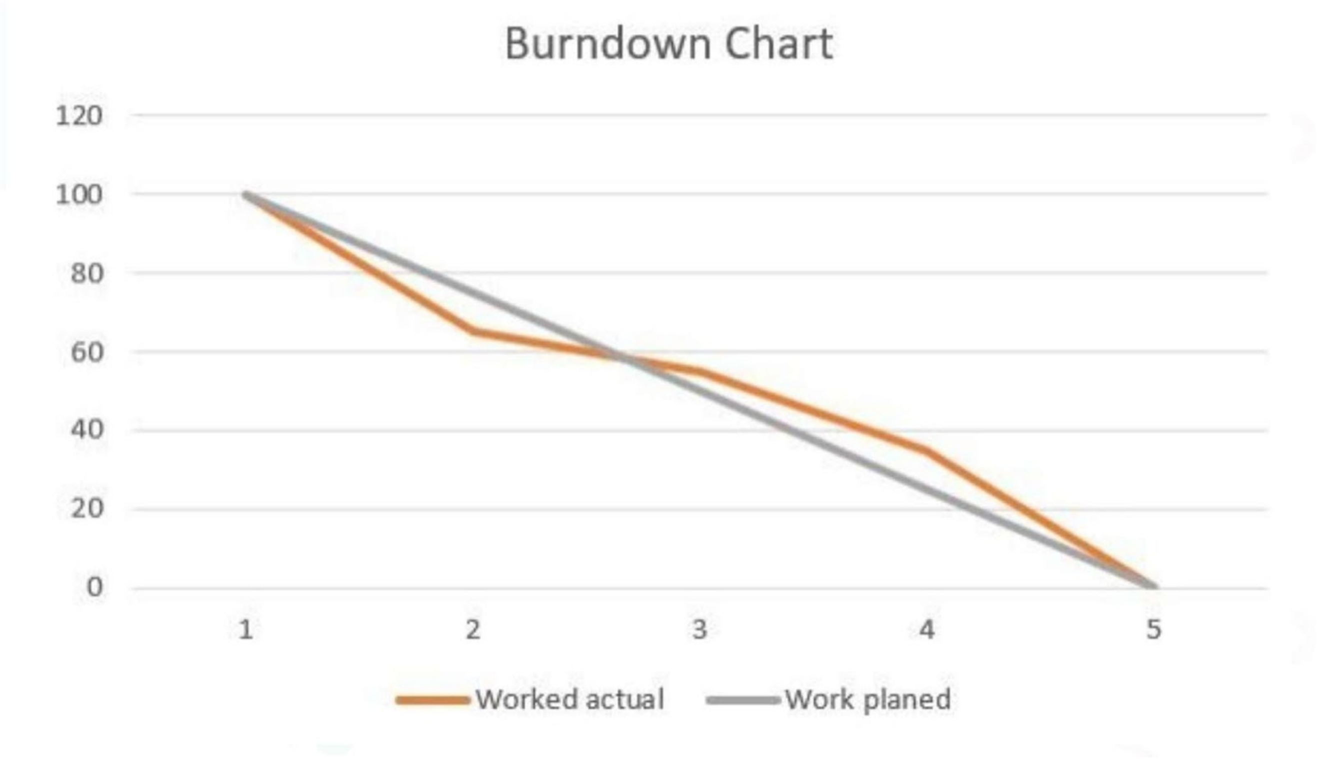

Importing Libraries
Matplotlib, seaborn and other libraries imported
Read the dataset
Dataset read from the specified file path
unnecessary columns dropped (nameOrig nameDest)
Displaying first and last 5 rows of the dataset
correlation analysis using a heatmap
Data Processing
Checking for null values
Null Values in the dataset checked and printed
Handling outliers
Box plot for "amount" attribute displayed
Handling Missing Values
Missing Values filled with mean
Scaling or Normalization
Min-Max Scaling applied
New feature 'balance_difference' created
Splitting Data into train and test sets
Data split into X_train, X_test, y_train, y_test
Model definitions and Evolutions
Random Forest Classifier
Random Forest model trained and evaluated
Decision Tree classifier
Decision Tree model trained and evaluated
Extra tree classifier
Extra tree model trained and evaluated
Support vector Machine Classifier
Support vector Machine model trained and evaluated
XGBoost Classifier
XGBoost Model trained and evaluated
Compare the models
All models compared and evaluated
Save the Best Performing Model
Best Performing model (SVM assumed) saved to a file

6.2 Sprint Planning & Estimation:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	User Authentication and Authorization	USN-1	Role-based access control for various user roles will ensure safe access to the fraud detection system.	1	High	K.Rishi
Sprint-1	**Transaction Data Collection**	USN-2	Gather and save pertinent transaction data, such as the amount, the user's identification, the timestamp, and the device's details..	2	High	P.Moksha

Sprint-1	**Data Preprocessing**	USN-3	To deal with missing data, outliers, and guarantee data quality for machine learning model training, apply preprocessing and data cleaning procedures.	2	High	K.prasanth
Sprint-2	**Feature Extraction and Engineering**	USN-4	To improve the effectiveness of the fraud detection algorithms, extract pertinent features from transaction data and create new features.	3	Medium	M.Manoj
Sprint-3	**Machine Learning Model Training**:	USN-5	Utilizing past transaction data, train machine learning models to spot patterns suggestive of fraudulent activity..	4	Medium	P.Moksha
Sprint-3		USN-6	**Continuously monitor incoming transactions in real-time to detect and flag potentially fraudulent activities.**	6	High	K.prasanth
Sprint-3	**Alerting and Notifications**	USN-7	Put in place an alerting system to inform pertinent stakeholders and fraud analysts of any suspicious transactions.	1	Medium	K.rishi
Sprint-4	**Scalability**	USN-8	**Ensure that the system is scalable to handle a growing volume of transactions as the business expands.**	1	Low	M.Manoj

6.3 Sprint Delivery Schedule:

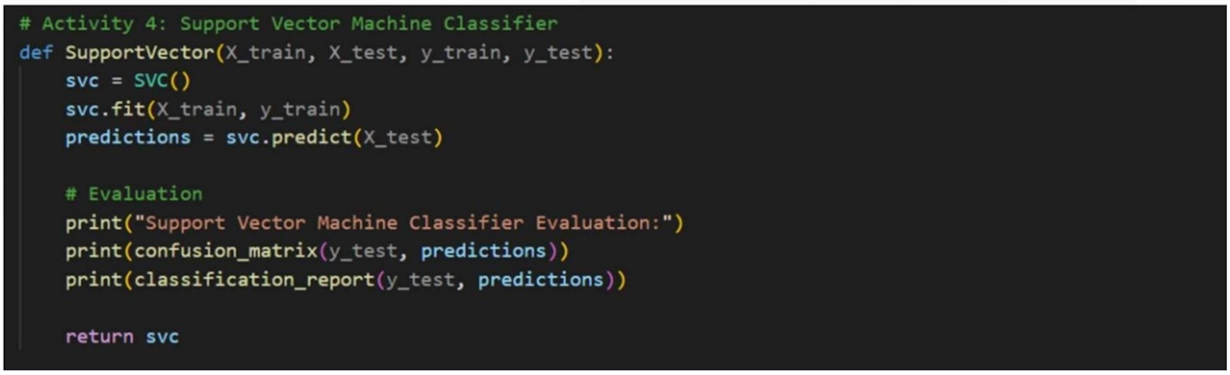


7 CODING & SOLUTIONING

7.2 Feature 1:

We train our data. I can train our data on different algorithms. For this project we are applying Three classification algorithms, SVM, XGboost and Random Forest Classifier. The best model is saved based on its performance.

SVM:

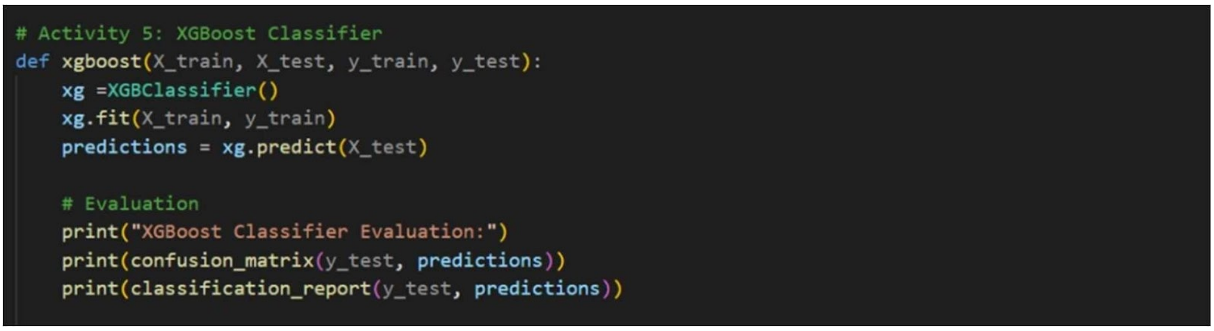


```
# Activity 4: Support Vector Machine Classifier
def SupportVector(X_train, X_test, y_train, y_test):
    svc = SVC()
    svc.fit(X_train, y_train)
    predictions = svc.predict(X_test)

    # Evaluation
    print("Support Vector Machine Classifier Evaluation:")
    print(confusion_matrix(y_test, predictions))
    print(classification_report(y_test, predictions))

    return svc
```

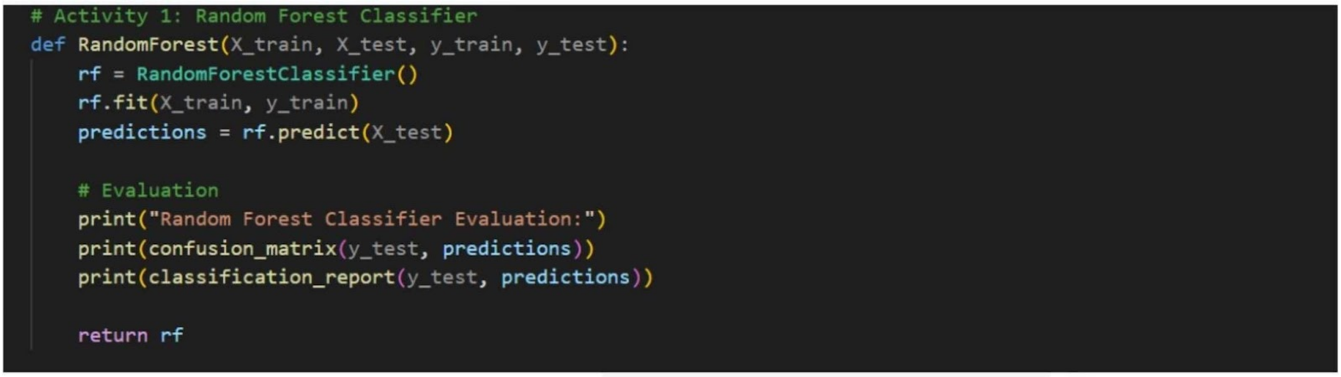
XGboost:



```
# Activity 5: XGBoost Classifier
def xgboost(X_train, X_test, y_train, y_test):
    xg =XGBClassifier()
    xg.fit(X_train, y_train)
    predictions = xg.predict(X_test)

    # Evaluation
    print("XGBoost Classifier Evaluation:")
    print(confusion_matrix(y_test, predictions))
    print(classification_report(y_test, predictions))
```

RANDOM FOREST CLASSIFIER

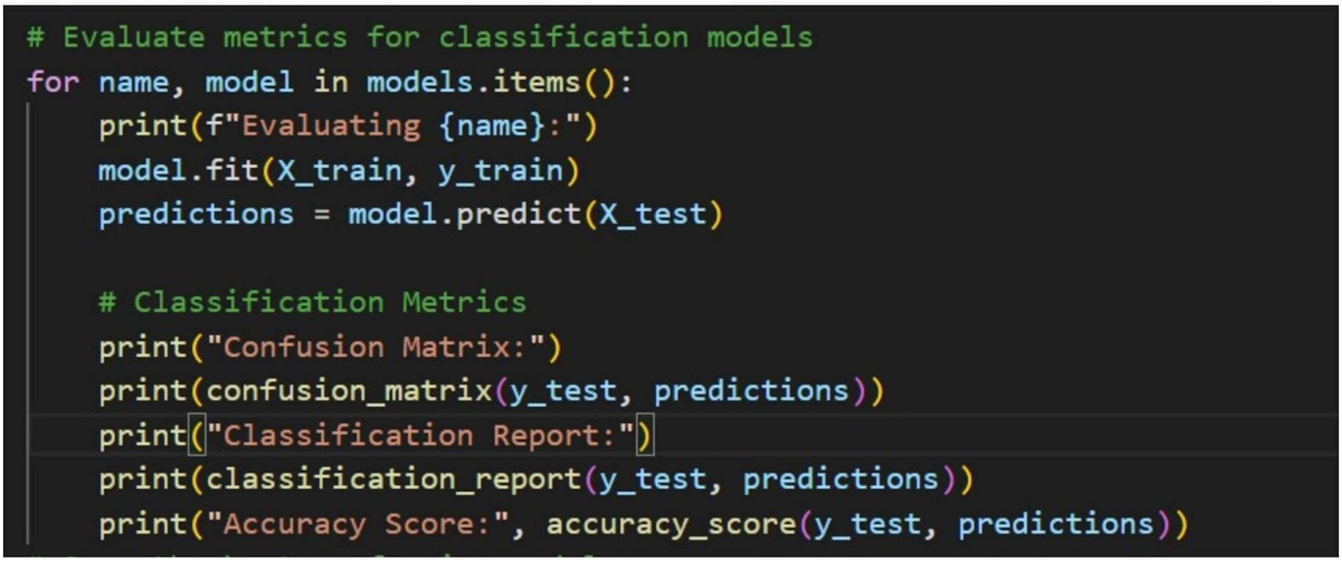


```
# Activity 1: Random Forest Classifier
def RandomForest(X_train, X_test, y_train, y_test):
    rf = RandomForestClassifier()
    rf.fit(X_train, y_train)
    predictions = rf.predict(X_test)

    # Evaluation
    print("Random Forest Classifier Evaluation:")
    print(confusion_matrix(y_test, predictions))
    print(classification_report(y_test, predictions))

    return rf
```

7.3 Feature 2: confusion matrix



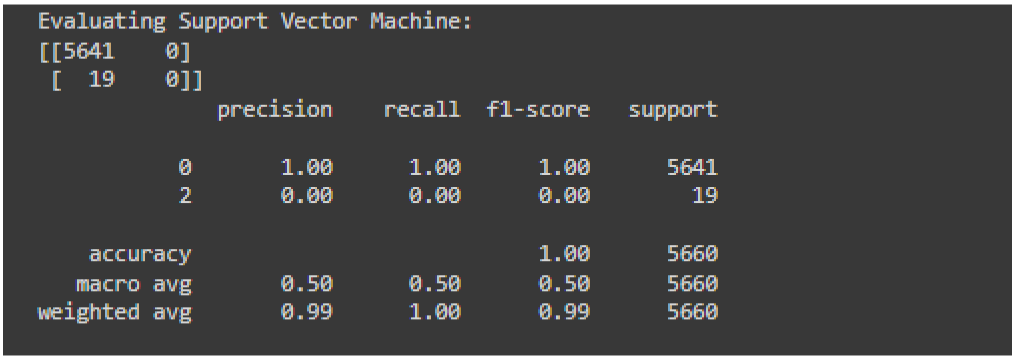
```
# Evaluate metrics for classification models
for name, model in models.items():
    print(f"Evaluating {name}:")
    model.fit(X_train, y_train)
    predictions = model.predict(X_test)

    # Classification Metrics
    print("Confusion Matrix:")
    print(confusion_matrix(y_test, predictions))
    print("Classification Report:")
    print(classification_report(y_test, predictions))
    print("Accuracy Score:", accuracy_score(y_test, predictions))
```

8 RESULTS

8.2 Output Screenshots

SVM:

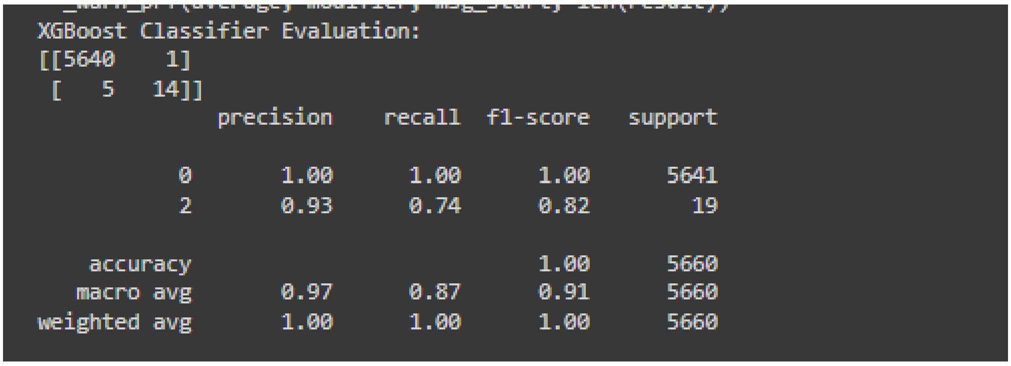


```
Evaluating Support Vector Machine:
[[5641    0]
 [  19    0]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5641
           2       0.00      0.00      0.00        19

    accuracy                           1.00      5660
   macro avg       0.50      0.50      0.50      5660
weighted avg       0.99      1.00      0.99      5660
```

XGBOOST:

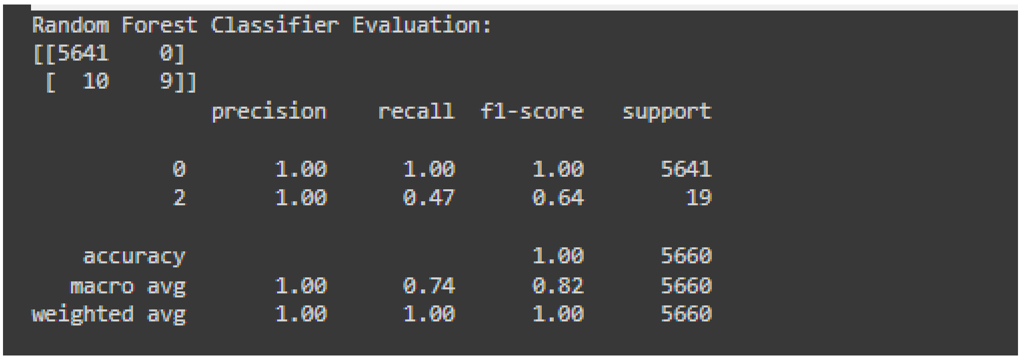


```
  _warn_prf(average, modifier, msg_start, len(result))
XGBoost Classifier Evaluation:
[[5640    1]
 [   5   14]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5641
           2       0.93      0.74      0.82        19

    accuracy                           1.00      5660
   macro avg       0.97      0.87      0.91      5660
weighted avg       1.00      1.00      1.00      5660
```

Random Forest Classifier:



```
Random Forest Classifier Evaluation:
[[5641    0]
 [  10    9]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5641
           2       1.00      0.47      0.64        19

    accuracy                           1.00      5660
   macro avg       1.00      0.74      0.82      5660
weighted avg       1.00      1.00      1.00      5660
```

9 .ADVANTAGES & DISADVANTAGES

Advantages of Online Fraud Detection:

Early detection of potential threats: Recognizes suspicious activities in their initial phases, preventing financial losses and safeguarding users.

Instantaneous notifications: Offers real-time alerts, facilitating prompt responses and the mitigation of potential dangers.

Augmented security: Reinforces the overall security of online transactions, fostering confidence among both users and businesses.

Dynamic learning capabilities: Adjusts to changing fraud patterns through machine learning, staying ahead of emerging threats.

Worldwide accessibility: Empowers users to engage in secure transactions from any location, fostering the global expansion of online commerce.

Disadvantages of Online Fraud Detection:

False Positives: False alarms may be triggered, causing inconvenience for users engaged in legitimate transactions and potentially impacting the overall user experience.

Challenging Implementation: The development and upkeep of an efficient fraud detection system can pose technical difficulties and demand significant resources.

Privacy Issues: Examining user behavior for fraud detection raises privacy concerns, emphasizing the need for cautious handling of sensitive information.

Resource Demands: The continuous monitoring and analysis of extensive datasets can strain resources, necessitating a robust infrastructure.

Adaptation to Fraud Tactics: Fraudsters evolve, and some may discover ways to evade detection methods, requiring ongoing system updates to stay effective.

10. CONCLUSION

online fraud detection is a vital component of ensuring the security and trustworthiness of digital transactions. Its advantages, such as early threat detection, real-time alerts, and adaptive learning through machine learning, contribute significantly to safeguarding users and businesses. However, challenges like false positives, complex implementation, and privacy concerns necessitate a thoughtful and balanced approach. As technology evolves, continuous refinement and adaptation of fraud detection systems are imperative to stay ahead of emerging threats. Ultimately, the benefits of enhancing online security and user trust outweigh the challenges, making ongoing advancements in fraud detection crucial for the sustainable growth of digital commerce.

11. FUTURE SCOPE

The future scope of online fraud detection holds promising developments and opportunities for further advancement. Key areas of future focus include:

Advanced Machine Learning Techniques:

Continued exploration and integration of advanced machine learning algorithms to enhance the accuracy and adaptability of fraud detection systems, particularly in the face of increasingly sophisticated fraud tactics.

Behavioral Biometrics:

Emphasis on leveraging behavioral biometrics, such as keystroke dynamics and mouse movements, to add an additional layer of user verification and enhance the overall security posture.

AI-Powered Predictive Analytics:

Integration of predictive analytics powered by artificial intelligence to anticipate potential fraud trends, enabling proactive measures to be

implemented before new threats fully materialize.
Blockchain Technology:

Exploration of blockchain technology for secure and transparent transaction verification, minimizing the risk of fraudulent activities and enhancing the traceability of financial transactions.

Collaborative Threat Intelligence:

Increased collaboration and information sharing among financial institutions, businesses, and security agencies to create a comprehensive network for identifying and responding to emerging fraud patterns collectively.

Biometric Authentication:

Wider adoption of biometric authentication methods, such as facial recognition and fingerprint scanning, for secure user identification and reducing reliance on traditional credentials.

Explainable AI in Fraud Detection:

Integration of explainable AI techniques to enhance the transparency of decision-making processes in fraud detection systems, ensuring a clear understanding of why certain transactions are flagged as potentially fraudulent.

Cross-Industry Collaboration:

Collaboration between different industries, including finance, technology, and cybersecurity, to create standardized frameworks and share best practices for combating fraud on a broader scale.

Regulatory Developments:

Evolving regulatory frameworks that address the growing challenges of online fraud, ensuring a balance between user privacy, security, and the seamless flow of digital transactions.

Enhanced User Education:

Greater emphasis on user education and awareness programs to empower individuals with the knowledge to recognize and report potential fraud, creating a more informed and vigilant online community..

12. APPENDIX

SOURCE CODE:

https://colab.research.google.com/drive/1GXDZPmAjZxSRSOONagyDfS444ULGMWMr

GITHUB LINK:

https://github.com/smartinternz02/SI-GuidedProject-613741-1700710990.git