

Fraud Detection in Financial Transactions

Microsoft Machine Learning Engineer



TEAM MEMBERS

- Abdelrahman Ashraf Abdelkarim
- Mohamed Abdelhamid Mohamed
- Mahmoud Hesham Shaban
- Yassen Mohsen Yassen Ali

Project Supervisor: Eng Mahmoud Safian

Group Code : FYM1_AIS5_M1e

Traning Company : GK



ABOUT TEAM

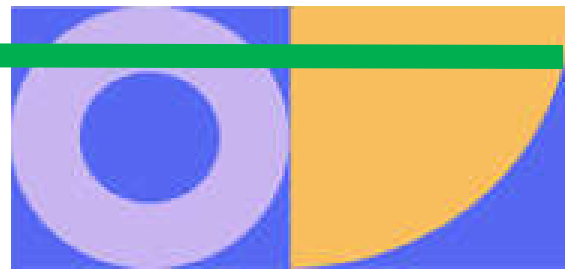


All team members contribute to all tasks.

- 1- Data Searching and Collection
- 2- Data Preparation, Cleaning, and Preprocessing
- 3- Statistical Analysis and Data Visualization
- 4- Model Architecture Design
- 5 -Applying NLP Techniques to Text Datasets
- 6 -Model Deployment and Monitoring
- 7- Testing Models in Real-World Applications
- 8 -Report and Presentation

Agenda Outline

- 1- Introduction
- 2- Problem Statement
- 3- Challenges in Fraud Detection
- 4- Project Aims and Objectives
- 5- Methodology
- 6- Week 1 (work Flow) : Data Collection and Preprocessing
- 7- Week 2 (work Flow) : Statistical Analysis and Machine Learning
- 8- Week 3 (work Flow) : Integrated NLP Techniques
- 9- Week 4 (work Flow) : MLOps and Final Presentation
- 10- Results
- 11- Evaluation Matrices
- 12- Future Works
- 13- Conclusion





Introduction & Background

01

Transaction Monitoring

Implement AI algorithms to continuously monitor banking transactions in real-time, enabling immediate detection of suspicious activities and reducing false positives in fraud alerts.

02

Customer Behavior Analysis

Utilize machine learning models to analyze customers' spending patterns, identifying anomalies that may indicate fraudulent behavior while enhancing customer trust and satisfaction.



Problem Statement

Data Imbalance

Fraudulent transactions are rare, causing model prediction bias.

High Bias

Models initially achieve high accuracy but fail to detect fraud.

Null Values

Missing data affects model performance and integrity.

Novel Fraud Detection

Traditional systems struggle to identify new and unforeseen fraudulent activities.

Robustness Requirement

Effective detection requires advanced AI techniques for diverse datasets.



Challenges in Fraud Detection and Proposed Solutions

	Challenge	Solution	Tools/Techniques
Imbalanced Data	Original dataset was imbalanced	Employed SMOTE and random sampling	Python, Scikit-learn
Handling Null Values	Presence of null values	Removed rows with excessive NAs or used mean imputation	Python, Pandas
Outliers Management	Outliers distorting performance	Applied techniques for outlier handling	Python, Scikit-learn
Initial Model Performance	High initial accuracy misled results	Balanced dataset improved Recall and F1-Score	Python, Scikit-learn, MLflow



Project Aims and Objectives Overview



Data Processing

Clean and preprocess the dataset for model training performance.



Statistical Analysis

Analyze feature distributions relevant to detecting fraudulent transactions.



Model Development

Construct machine learning models focused on fraud detection tasks.



NLP Techniques

Employ NLP to extract insights from transaction descriptions and notes.



MLOps Implementation

Utilize MLflow for comprehensive tracking and management of models.

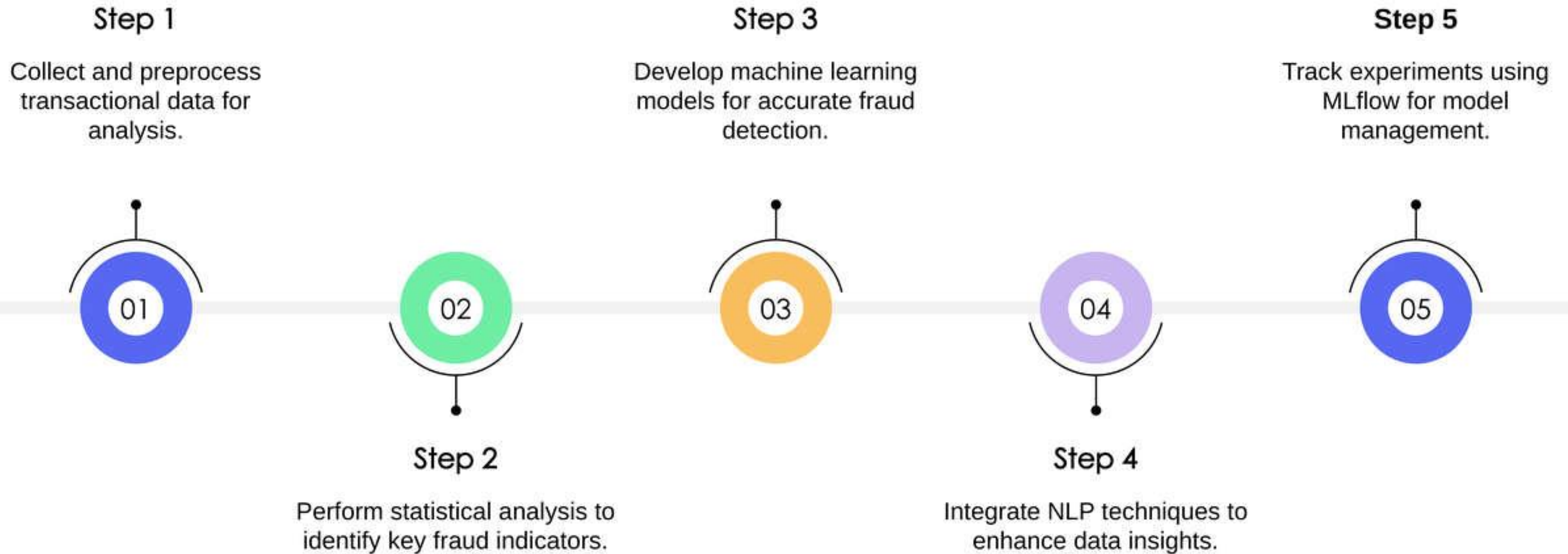


Performance Metrics

Evaluate models using Recall and F1-score for performance accuracy.



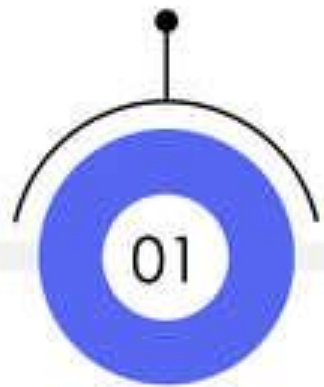
Methodology: Step-by-Step Project Plan



Week 1: Data Collection and Preprocessing

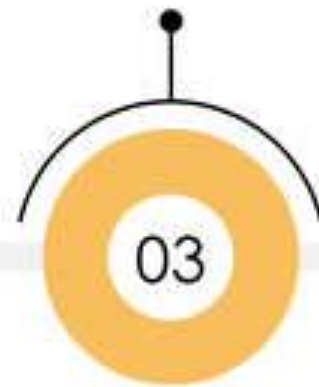
Data Acquisition

Gather labeled transaction data from relevant sources.



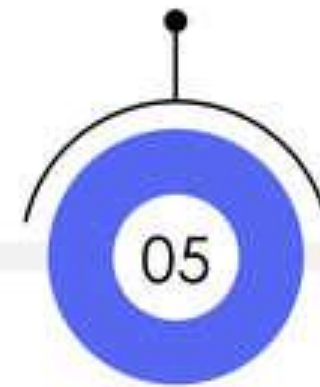
Normalization

Standardize data formats for consistency across features.



Missing Values

Identify and address missing values in the dataset.



Final Review

Conduct a thorough review of the preprocessed data.



Data Cleaning

Remove duplicates and irrelevant records from dataset.



Data Exploration

Visualize data to understand distributions and trends.



Feature engineering

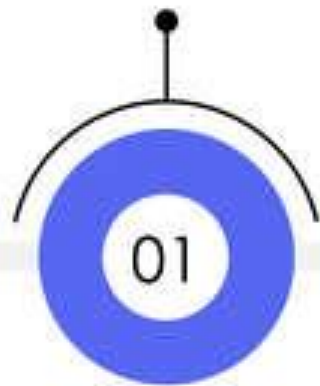
Create new features beneficial for fraud detection.



Week 2: Statistical Analysis and Model Development

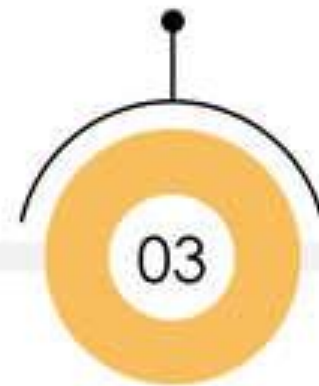
Check Distribution

Examine feature distributions in the dataset.



Feature Selection

Identify key features influencing fraud detection.



Data Visualization

Utilize plots to reveal feature relationships.



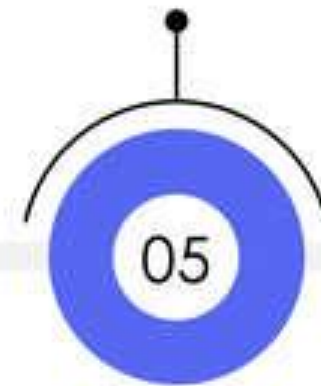
Model Selection

Choose suitable algorithms for fraud detection.



Model Training

Fit models using training dataset for learning.



Model Testing

Evaluate model performance on test datasets.



Evaluate Model

Calculate metrics like accuracy, recall, F1-score.



Results Analysis

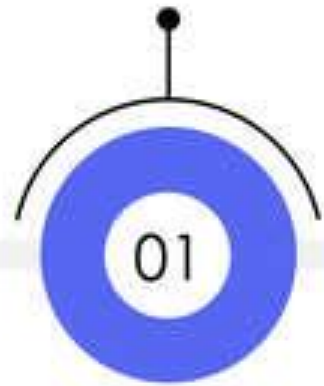
Analyze results and prepare for adjustments.



Week 3: Integrating NLP Techniques

Data Preparation

Clean and preprocess transaction descriptions to ensure quality input for NLP analysis.



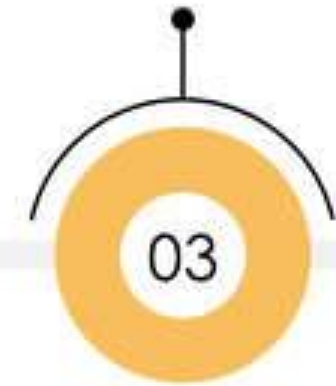
Feature Extraction

Utilize NLP methods to extract relevant features from text, enhancing the model's understanding of patterns.

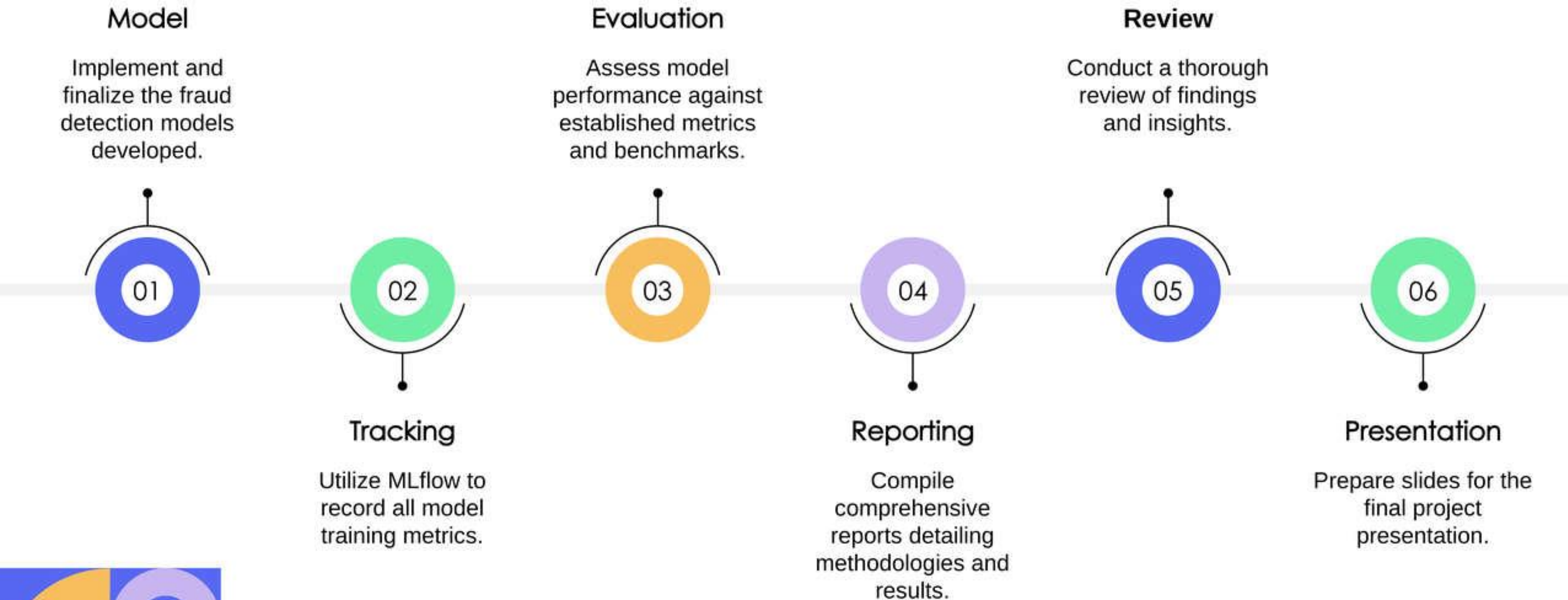


Model Integration

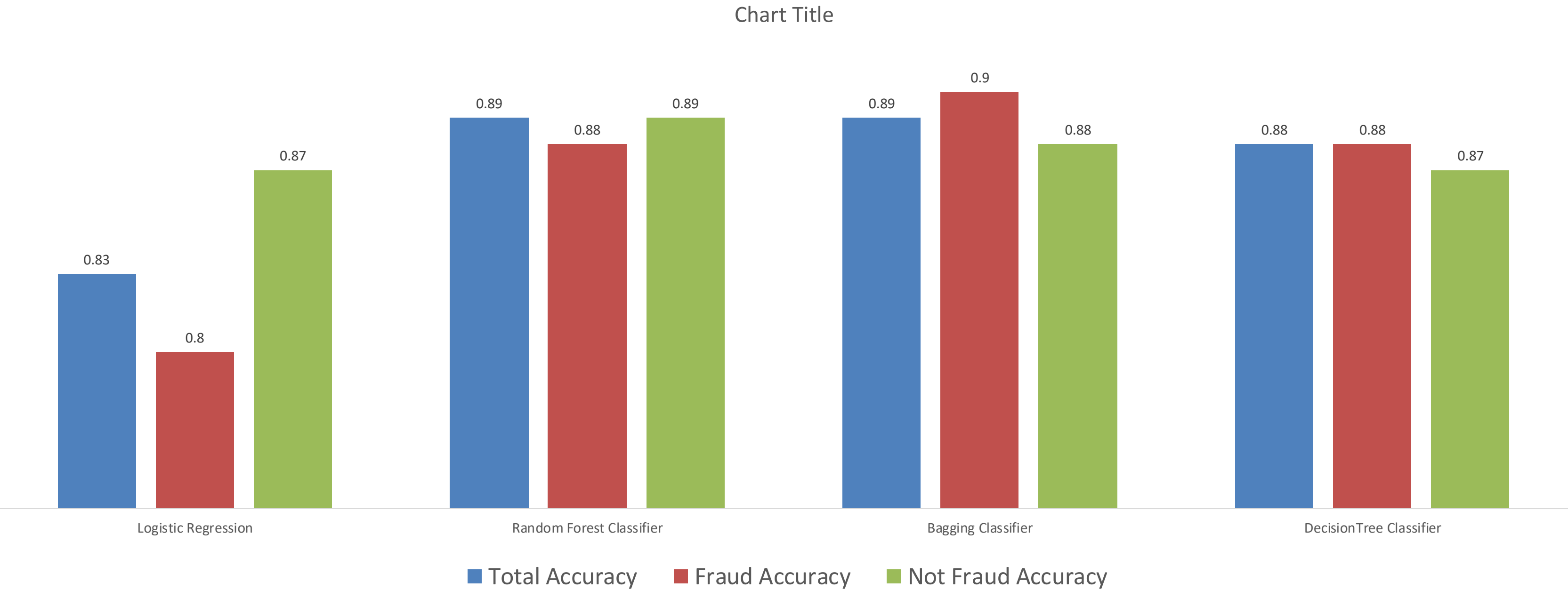
Incorporate extracted features into the machine learning models for improved fraud detection accuracy.



Week 4: MLOps and Final Presentation Preparation



Fraud Detection ML Models



Fraud Detection NLP Models



MLFlow Results

<input type="checkbox"/>	Run Name	Created	Dataset	Duration	Source	Models
<input type="checkbox"/>	Fraud_detection_NLP(9)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	25.9s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(8)	21 hours ago	dataset (ee8fc596) , dataset (1c0...	19.9s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(7)	21 hours ago	dataset (ee8fc596) , dataset (1c0...	30.2s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(6)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	23.7s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(5)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	22.7s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(4)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	21.7s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(3)	21 hours ago	dataset (ee8fc596)	22.6s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(2)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	28.5s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_NLP(1)	21 hours ago	dataset (1c0b7ef3) , dataset (ee8f...	25.2s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_ML(5)	21 hours ago	dataset (6ff932be) , dataset (4ec8...	16.6s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_ML(4)	21 hours ago	dataset (4ec87259) , dataset (6ff...	17.3s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_ML(3)	21 hours ago	dataset (4ec87259) , dataset (6ff...	9.4s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_ML(2)	21 hours ago	dataset (6ff932be) , dataset (4ec8...	7.0s	c:\Users\...	sklearn
<input type="checkbox"/>	Fraud_detection_ML(1)	21 hours ago	dataset (6ff932be) , dataset (4ec8...	15.1s	c:\Users\...	sklearn



Conclusion

This project successfully developed a fraud detection system that effectively handles imbalanced data and achieves reliable results. The initial model displayed high accuracy, but balancing techniques such as SMOTE significantly improved recall and F1-score, reducing bias. MLflow proved to be a valuable tool in managing experiments, tracking models, and maintaining transparency. Future improvements could include using deep learning techniques like RNNs to enhance the system's ability to analyze sequential data and integrating real-time data streams to improve responsiveness.

Future Work

Deep Learning

Implement LSTM networks to enhance sequential data analysis capabilities.

Real-Time Monitoring

Integrate a system for continuous fraud detection during transactions.

Feature Engineering

Explore additional features from transaction metadata to boost model performance.

Anomaly Detection

Utilize unsupervised learning techniques to identify novel fraudulent patterns.

Cross-Validation

Employ k-fold cross-validation to assess model performance rigorously.

User Feedback

Incorporate user feedback loops to refine fraud detection algorithms effectively.

THANK YOU

Any Questions?