GithubLink: **https://github.com/panneerselvam1230/project-Guarding-transactions-with-AI.git**

ProjectTitle:Guardingtransactionswith AI -powered credit fraud detection and prevention

PHASE-2

- **ProblemStatement**

With the rapid growth of digital payments and online transactions, credit card fraud has become a significant challenge for financial institutions and customers alike. Traditional frauddetectionmethodsoftenrelyonrule-basedsystems,whicharelimitedintheirability to adapt to evolving fraud patterns and can lead to high false-positive rates, frustrating legitimatecustomers.ThereisapressingneedforanAI-poweredsolutionthatcananalyze large volumes of transactional data in real time, accurately identify suspicious activities, and prevent fraudulent transactions without disrupting the user experience. This project aimstodevelopandimplementanAI-drivencreditfrauddetectionandpreventionsystem that enhances security, reduces financial losses, and improves customer trust.

- **ProjectObjectives**

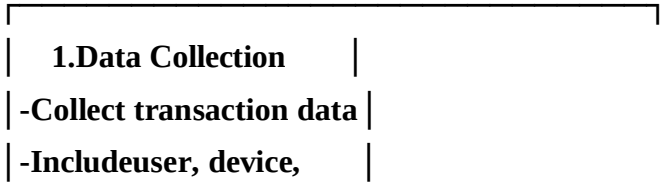
Toanalyzeexistingpatternsofcreditcardtransactionsandidentifykeyindicatorsof fraudulent behavior using historical data.

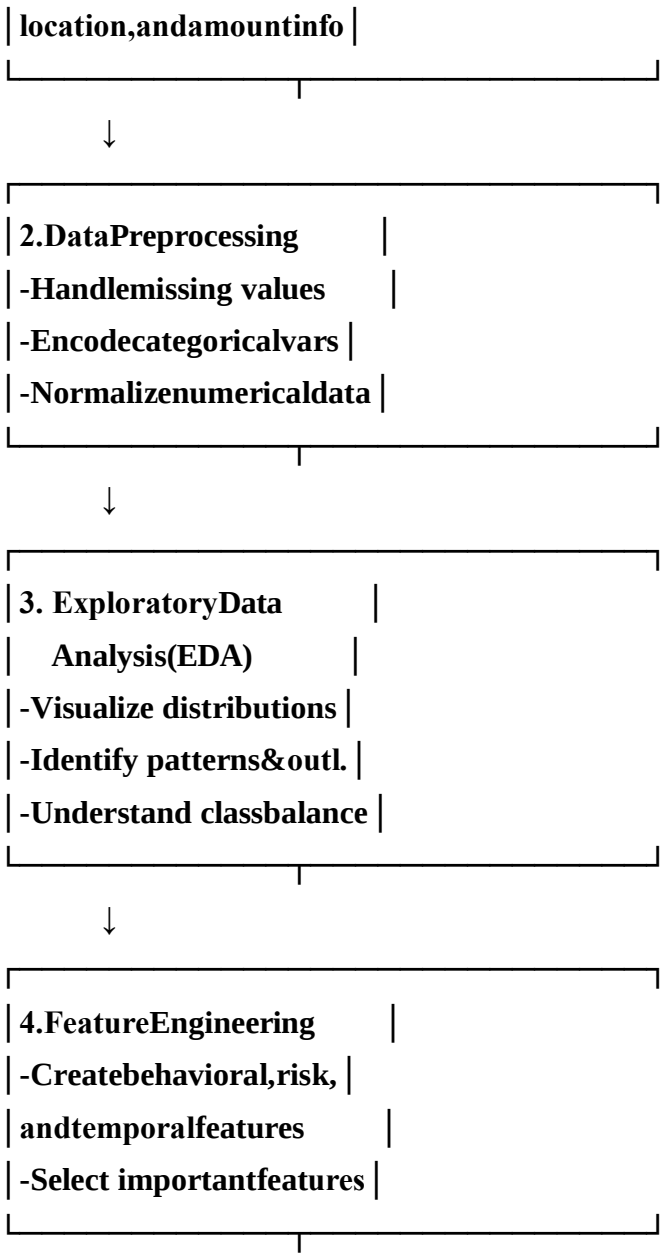
Todesignanddevelopamachinelearningmodelcapableofdetectingandpredicting fraudulent transactions in real time with high accuracy.

Tominimizefalsepositivesandfalsenegativesinfrauddetection,ensuringthatlegitimate transactions are not wrongly flagged while maximizing the detection of fraudulent activities.

TointegratetheAI-basedfrauddetectionsystemintothetransactionprocessingpipeline without significantly impacting transaction speed or user experience.

- **FlowchartoftheProject Workflow**





- **DataDescription**

Transaction Data

Transaction ID → Unique identifier for each transaction

Timestamp→Dateandtimewhenthetransactionwasmade Amount

→ Total value of the transaction

MerchantID→Identifierofthemerchantorstore

Transaction Type → Purchase, withdrawal, transfer, etc.

PaymentMethod→Creditcard,debitcard,onlinewallet,etc.

Currency → Currency used in the transaction

Datesetlink:https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

- **DataPreprocessing**

1. Datacleaning

Removeduplicates→Dropduplicatetransactionsthatmayresult fromsystemerrors.

Handle missing values →

Imputemissingvaluesusingstatisticalmethods(mean,median)ordomainrules. Remove

records with excessive missingness if needed.

Correctdatatypes→Ensurenumeric,categorical,anddatefieldshavecorrect formats.

2. DataTransformation

Converttimestamps→Extractusefultimefeatures(hourofday,dayofweek, weekend/weekday).

Normalizeorscalenumericfeatures→Applymin-maxscalingorstandardizationonfeatures like amount to handle wide value ranges.

- **ExploratoryDataAnalysis(EDA)**
-
- **1.UnderstandDatasetOverview**
- Checkdatadimensions→Number ofrows(transactions) andcolumns(features).
- Examinefeaturetypes→Identifycategorical, numerical,anddatetimevariables.
- Checkmissingvalues→Percentageofmissingdataper feature**.**
- **2.SummaryStatistics**
- Numericalfeatures→Calculatemean,median,min,max,standarddeviationforamount, transaction time, etc.
- Categoricalfeatures→Countfrequenciesandproportionsfortransactiontype,merchant ID, payment method.
- **FeatureEngineering**
- **1.BasicFeatures(from rawdata)**
- Transactionamount→Rawamountofthetransaction.
- Transactiontype→Purchase,transfer, withdrawal,etc.
- Timestampfeatures→Extract:
- Hourofday
- Dayofweek

- Weekendvs.weekday
- Holidayflag
- Merchantcategory→Typeofmerchant(e.g.,retail,travel, electronics).
- Payment method→Creditcard,debit card,digitalwallet.
- **ModelBuilding**

1. DefinetheModelingObjective

Goal:Predictwhetheratransactionisfraudulent(fraud=1)orlegitimate(fraud=0). Type of

problem: Binary classification.

2. SelectMachineLearning Algorithms

Chooseoneoracombinationofthefollowingalgorithms:

✅**BaselineModels(forbenchmarking)**

LogisticRegression

Decision Trees

✅**AdvancedModels(forbetterperformance)**

RandomForest

Gradient Boosting (XGBoost, LightGBM, CatBoost)

NeuralNetworks(especiallyforlarge,complexdatasets)

- **VisualizationofResults&ModelInsights**
- **What itshows:**
- **N1. ConfusionMatrix**
- Numberoftruepositives(fraud correctlypredicted)
- Numberoffalsepositives(legitimateflaggedasfraud)
- Numberoftruenegatives(legitimatecorrectlypredicted)
- Number offalsenegatives(fraud missed)
- Howtovisualize:
- Usea heatmapto show countsor percentages**.**
- ✅**2. ROCCurve(ReceiverOperating Characteristic)**

- **What itshows:**

- Trade-offbetweentruepositiverate(recall)andfalsepositiverateatvariousthresholds.

- **Toolsand Technologies Used**
-
- ✅**ProgrammingLanguages**
- Python→Mainlanguagefor dataprocessing,modeling,anddeployment
- R→(Optional)Forstatisticalanalysisandvisualization
- ✅**DataProcessingand Analysis**
- pandas→Datacleaning,manipulation,and analysis
- NumPy→Numericaloperationsandarrayprocessing
- SQL→Extractingtransactionaldatafromdatabases
-
- **TeamMembersandContributions**

1. M.NANDHAKUMAR—DataScientist

Performed data cleaning, preprocessing, and feature engineering

Builtandtunedmachinelearningmodels(RandomForest,XGBoost)

Conductedexploratorydataanalysis(EDA)andextractedactionableinsights Evaluated

modelperformance using precision, recall, F1, and AUC metrics

✅**2.S.PRAKASHRAJ—MachineLearningEngineer**

Implemented ML pipelines for training and validation

Integrated model with backend systems using Flask API

Optimized model for real-time prediction and scalability

DevelopedDockercontainersfordeploymentincloudenvironments

✅**3.K.PANNEERSELVAM—DataAnalyst/VisualizationExpert**

CreateddetailedvisualizationsforEDAandmodelinsights(confusionmatrix,ROCcurves, feature importance)

Developed dashboards using Plotly and Tableau for monitoring fraud patterns

Providedreportsonkeyfindingsandrecommendationsforbusinessstakeholders

✅**4. M.PERUMAL — Project Manager / Domain Expert**

Coordinated team activities, timeline, and task management

Actedasliaisonwithfinancialdomainexpertsandstakeholders

Defined problemstatement, project scope, and success criteria

Reviewedfinaldeliverablesand ensuredcompliancewithregulatorystandards