Githublink:https:https:https://github.com/sathishkumarsk2005

TITLE:GUARDINGTransactionswithAI-PoweredCreditCardFraudDetectionand Prevention

1:ProblemStatement:

Creditcardfraudisasignificantissueforfinancialinstitutions,merchants,and consumersglobally.Withtheincreasingvolumeofonlineandofflinecreditcard transactions,thepotentialforfraudulentactivityhasalsorisen.Traditionalfraud detection systems, relying on rule-based algorithms, often fall short in identifying new andsophisticatedfraudulentschemes.Thischallengeisexacerbatedbythevast numberoftransactionsthatmustbeprocessedquickly,theevolvingnatureoffraud tactics,andtheneedforreal-timedetectionwithoutnegativelyimpactinglegitimate user experiences.

To address this, there is a need for advanced, AI-powered credit card fraud detection andpreventionsystemsthatcanadapttoemergingfraudtacticswhileminimizingfalse positives and optimizing the transaction experience for legitimate users.

2. ProjectObjectives

- Buildamachinelearningmodelthatcanreliablydetectfraudulenttransactions.
- Utilizesupervisedandunsupervisedlearningtechniquestodevelopaclassification model capable of differentiating between legitimate and fraudulent transactions.
- Trainthemodelusinglabeleddatasetswithbothfraudulentandnon-fraudulent transactions.
- Implementanomalydetectiontechniquestoidentifyemergingfraudpatternsthat have not yet been encountered in historical data.

3:FlowchartoftheProjectWorkflow

|Start |

|DataCollection&Integration|

|-Collecttransactiondata|

|-Integrateexternaldatasources(e.g.,deviceinfo,geolocation)|

|DataPreprocessing |

|-Datacleaning&normalization|

|-Featureextraction&selection|

|ModelSelection |

|-ChooseappropriateMLmodels(e.g.,supervised,unsupervised,RL)|

|-Selectalgorithms(e.g.,decisiontrees,SVM,deeplearning)|

|ModelTraining&Evaluation|

|-Splitdataintotraining&testingsets|

|-Trainmodelonhistoricallabeleddata|

|-Evaluatemodelperformance(precision,recall,F1-score)|

|-Hyperparametertuning |

|-Cross-validationforrobustness|

|-Adjustforfalsepositives/negatives|

|Real-TimeFraudDetection|

|-Deploymodelforreal-timescoringoftransactions|

|-Assignfraudriskscoretoeachtransaction|

|ActiononFraudulentTransactions|

|-Flagsuspicioustransactions|

|-Sendalertstocustomersorinstitutions|

|-Initiateverificationprocessifnecessary|

|ContinuousLearning&FeedbackLoop|

|-Monitormodelperformance(e.g.,falsepositives,detectionaccuracy)|

|-Updatemodelwithnewfraudpatternsanddata|

|-Retrainmodelperiodicallyforcontinuousimprovement|

|Compliance&Security |

|-Ensureprivacy(GDPR,PCI-DSS)|

|-Dataencryptionandsecurestorage|

|End |

4. DataDescription

- DatasetName:StudentPerformanceDataSet
- Source:UCIMachineLearningRepository
- TypeofData:Structuredtabulardata
- RecordsandFeatures:395studentrecordsand33features(numeric+categorical)
- TargetVariable:G3(finalgrade,numeric)
- StaticorDynamic:Staticdataset
- AttributesCovered:Demographics(age,address,parents'education),academics (G1, G2, study time), and behavior (alcohol consumption, absences)
- DatasetLink:https:https://github.com/sathishkumarsk2005/Project--phase2.git

5. DataPreprocessing

1. DataCollection

Thefirststepinpreprocessingistogathertherawtransactiondata.Thistypically includes:

- TransactionFeatures:
  - o TransactionID
  - o Cardholderdetails(userID,cardnumber,etc.)
  - o Merchantdetails(merchantID,merchantcategory,location,etc.)
  - o Transactionamount

- o Transactiontime(timestamp)
- o Transactiontype(online,offline,etc.)
- o Devicedetails(deviceID,IPaddress)
- o Geolocation(latitude,longitude)

2. DataCleaning

Data cleaning involves handling missing values, removing duplicates, and dealing with any inconsistencies or errors in the raw data.

Actions:

- MissingValues:
  - o Handlemissingdatapointsusingtechniqueslikeimputation(mean,median,or mode) or dropping rows/columns with excessive missing values.

6. ExploratoryDataAnalysis(EDA)

- UnivariateAnalysis:
  - Mean,Median,Mode
  - StandardDeviation&Variance
  - Min&Max
  - Histograms,BoxPlots,DensityPlots

- Bivariate&MultivariateAnalysis:
  - Correlationmatrix
  - ScatterplotsofG1vsG3andG2vsG3
  - Groupedbarcharts

- KeyInsights:
  - G1andG2arethestrongestindicatorsofG3

- Morestudytimecorrelateswithhigher G3
- Studentswithmorefailuresorabsencestendtoscorelower

7. Feature Engineering

Transaction-BasedFeatures

- TransactionAmountDifferences:
  - o Amountvs.AverageTransaction
  - o Formula:TransactionAmount-AverageTransactionAmount

8. ModelBuilding

- AlgorithmsUsed:
  - LinearRegression
  - RandomForestRegressor

- ModelSelectionRationale:
  - LinearRegression:interpretableandfast
  - RandomForest:robusttooverfitting,handlesmixeddatatypes well

- Train-TestSplit:80%training,20%testing

- EvaluationMetrics:
  - MAE,RMSE,R^2Score

9. VisualizationofResults&ModelInsights

- FeatureImportance:BarplotsfromRandomForest
- ModelComparison:MAE,RMSE,andR^2forbothmodels
- ResidualPlots:Predictionerrorsvs.actualgrades

- UserTesting:IntegratedmodelintoGradiointerface

10. ToolsandTechnologiesUsed

- ProgrammingLanguage:Python3
- NotebookEnvironment:GoogleColab
- KeyLibraries:pandas,numpy,matplotlib,seaborn,plotly,scikit-learn,Gradio

11. TeamMembersandContributions

Datacleaning:(B.THIRUPATHI)

- Mean,median,ormodeimputationfornumericalfeatures
- Modeimputationforcategoricalfeatures

EDA:(M.SENTHILKUMAR)

- Classimbalanceawareness
- Biasdetection

Featureengineering:(S.SATHISHKUMAR)

- Averagetransactionamountperuser
- Transactionfrequency

Modeldevelopment:

- Algorithmselection,handlingclassimbalance
- Performancemetricsanalysis