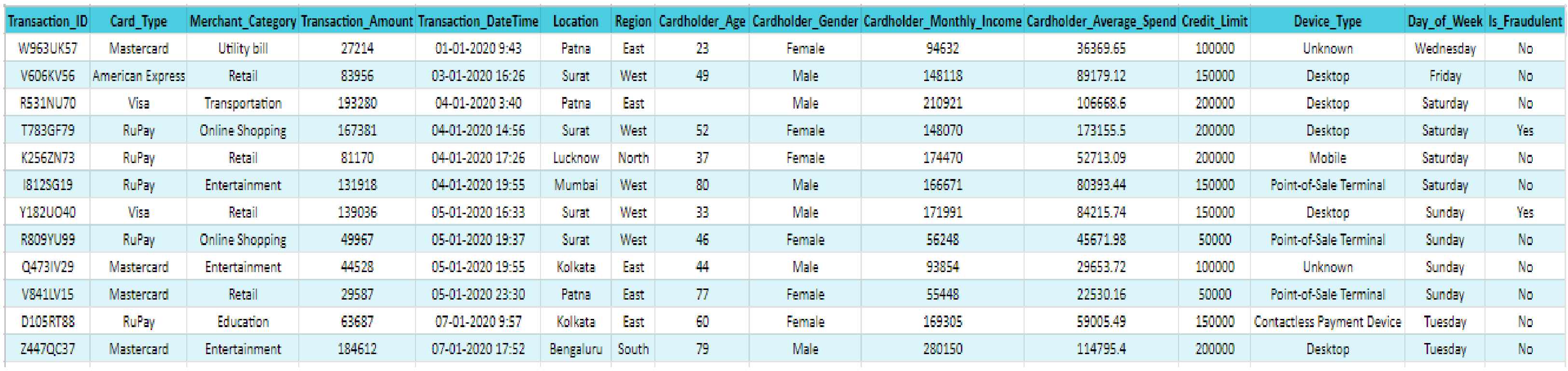
Project Title: *Credit Card Fraud Detection*

Problem Statement:

With the increasing prevalence of online transactions, ensuring the security of credit card transactions is very important. The objective of this project is to develop a robust machine learning model capable of accurately detecting fraudulent credit card transactions in real-time. Utilizing a dataset containing transaction details such as transaction amount, merchant category, cardholder information, transaction location, the aim is to build a predictive model that can effectively differentiate between legitimate and fraudulent transactions. By employing advanced machine learning algorithms and feature engineering techniques, the goal is to create a system that enhances fraud detection capabilities, thereby minimizing financial losses for both cardholders and financial institutions while maintaining a low false positive rate. Ultimately, this project seeks to contribute to the development of proactive measures for securing credit card transactions and safeguarding the financial interests of stakeholders in the digital economy.

Dataset Overview:



Transaction_ID	Card_Type	Merchant_Category	Transaction_Amount	Transaction_DateTime	Location	Region	Cardholder_Age	Cardholder_Gender	Cardholder_Monthly_Income	Cardholder_Average_Spend	Credit_Limit	Device_Type	Day_of_Week	Is_Fraudulent
W963UK57	Mastercard	Utility bill	27214	01-01-2020 9:43	Patna	East	23	Female	94632	36369.65	100000	Unknown	Wednesday	No
V606KV56	American Express	Retail	83956	03-01-2020 16:26	Surat	West	49	Male	148118	89179.12	150000	Desktop	Friday	No
R531NU70	Visa	Transportation	193280	04-01-2020 3:40	Patna	East		Male	210921	106668.6	200000	Desktop	Saturday	No
T783GF79	RuPay	Online Shopping	167381	04-01-2020 14:56	Surat	West	52	Female	148070	173155.5	200000	Desktop	Saturday	Yes
K256ZN73	RuPay	Retail	81170	04-01-2020 17:26	Lucknow	North	37	Female	174470	52713.09	200000	Mobile	Saturday	No
I812SG19	RuPay	Entertainment	131918	04-01-2020 19:55	Mumbai	West	80	Male	166671	80393.44	150000	Point-of-Sale Terminal	Saturday	No
Y182UO40	Visa	Retail	139036	05-01-2020 16:33	Surat	West	33	Male	171991	84215.74	150000	Desktop	Sunday	Yes
R809YU99	RuPay	Online Shopping	49967	05-01-2020 19:37	Surat	West	46	Female	56248	45671.98	50000	Point-of-Sale Terminal	Sunday	No
Q473IV29	Mastercard	Entertainment	44528	05-01-2020 19:55	Kolkata	East	44	Male	93854	29653.72	100000	Unknown	Sunday	No
V841LV15	Mastercard	Retail	29587	05-01-2020 23:30	Patna	East	77	Female	55448	22530.16	50000	Point-of-Sale Terminal	Sunday	No
D105RT88	RuPay	Education	63687	07-01-2020 9:57	Kolkata	East	60	Female	169305	59005.49	150000	Contactless Payment Device	Tuesday	No
Z447QC37	Mastercard	Entertainment	184612	07-01-2020 17:52	Bengaluru	South	79	Male	280150	114795.4	200000	Desktop	Tuesday	No

Data Attributes:

Field Name	Description
Transaction_ID	Unique identifier for each credit card transaction.
Card_Type	Type of credit card used in the transaction (Visa, Mastercard, RuPay, American Express).
Merchant_Category	Category of the merchant where the transaction took place (Retail, Online Shopping, Dining, Entertainment, Healthcare, Education, Transportation, Utility bill).
Transaction_Amount	The amount of money involved in the transaction.

Transaction_DateTime	Date and Time of the transaction.
Location	City where the transaction occurred (Mumbai, Delhi, Bengaluru, Kolkata, Chennai, Hyderabad, Kochi, Pune, Ahmedabad, Surat, Jaipur, Lucknow, Patna).
Region	Region where the transaction occurred (North, South, West, East).
Cardholder_Age	Age of the Cardholder in years.
Cardholder_Gender	Gender of the cardholder (Male or Female).
Cardholder_Monthly_Income	Monthly income of the cardholder.
Cardholder_Average_Spend	The average amount spent by the cardholder.
Credit_Limit	Maximum amount of credit extended to the cardholder by the issuing bank (50000, 100000, 150000, 200000).
Device_Type	Type of device used for the transaction (Mobile, Desktop, Point-of-Sale Terminal, Contactless Payment Device, Unknown).
Day_of_Week	Day of the week when the transaction took place.
Is_Fraudulent	Indicator of whether the transaction is fraudulent or not (Yes/No).

Dataset Download:

https://raw.githubusercontent.com/ArchanaInsights/Datasets/refs/heads/main/credit_card_transactions.csv

Project Steps and Objectives:

1) Exploratory Data Analysis (EDA):

a) Analyze the distribution of categorical features such as Card_Type, Merchant_Category, Location, etc.

b) Explore numerical features like Transaction_Amount, Cardholder_Age, Cardholder_Monthly_Income, and Cardholder_Average_Spend. Use descriptive statistics to understand their central tendency and spread.

c) Conduct bivariate and multivariate analysis to identify potential relationships between the features as well as with the target variable (Is_Fraudulent).

d) Visualize the distribution of transaction amounts for fraudulent vs. non-fraudulent transactions using histograms or box plots.

e) Investigate whether certain features are more susceptible to fraud.

2) Data Preprocessing - Data Cleaning:

a) Handle missing values if any, using appropriate techniques such as KNNImputer; mean or median imputation for numerical features, and mode imputation for categorical features.

b) Check for outliers in numerical features using statistical methods like Z-score or IQR (Interquartile Range) and remove them if necessary to ensure data quality.

c) Assess skewness in numerical features by calculating the skewness score. If any features are highly skewed, consider applying transformations such as square root or log transformation to improve their distribution before scaling, if needed.

3) Feature Engineering:

a) Identify the categorical features in the dataset.

b) Encode categorical features to numerical using techniques like one-hot encoding or label encoding techniques to prepare the data for machine learning algorithms.

4) Feature Selection:

a) Select relevant features that have the most impact on predicting fraudulent transactions.

b) Identify and remove redundant or irrelevant features that do not contribute significantly to the prediction task.

5) Split data into training and testing:

a) Divide the dataset into training and testing sets to evaluate the model's performance.

b) Ensure that both sets maintain the same distribution of fraudulent and non-fraudulent transactions to avoid data leakage.

6) Feature Scaling:

a) Scale numerical features to ensure that they have the same magnitude, preventing some features from dominating others during model training.

b) Common scaling techniques include Min-Max scaling or Standardization (Z-score normalization).

7) Build the Machine Learning Model:

a) Import the necessary modules and libraries for building and evaluating machine learning models.

b) Define a list or dictionary of classifiers to be evaluated; including Logistic Regression, Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, and SVM. Then, compute the accuracy score and F1-score for each classifier.

c) Select a machine learning algorithm for binary classification with the highest accuracy or F1-score from the above step.

d) Train the selected model using the training dataset and evaluate its performance using appropriate metrics like confusion matrix, accuracy, precision, recall, and F1-score.

e) Validate the model's performance on the testing dataset and interpret the results to assess its effectiveness in detecting fraudulent transactions.