CREDIT CARD FRAUD DETECTION

INTRODUCTION:

Use of credit cards can occasionally lead to considerable financial losses and is not always advantageous for users. Frauds with credit cards is on the rise. progressively growing each day as internet usage shifts toward digital. The use of debit cards is increasing along with the use of credit cards for purchases. purchases made online. There are many different forms of fraudulent transactions that can happen with anyone, anywhere, using a range of conditions and methods. To uncover unauthorized purchases of goods by customers, credit card firms must be able to identify credit card fraud transactions. Data Science and machine learning are now assisting in the identification of these Transactions that are fraudulent Transactions involving fraud are common and helping the financial institutes to make an informed decision whether to approve credit card for a user or not.

PROBLEM STATEMENT:

The project aims to develop a machine learning-based system that analyses transaction data in real-time, effectively detecting credit card fraud while minimizing false positives. This solution will help financial institutions protect against fraudulent transactions, reducing financial losses and ensuring customer trust.

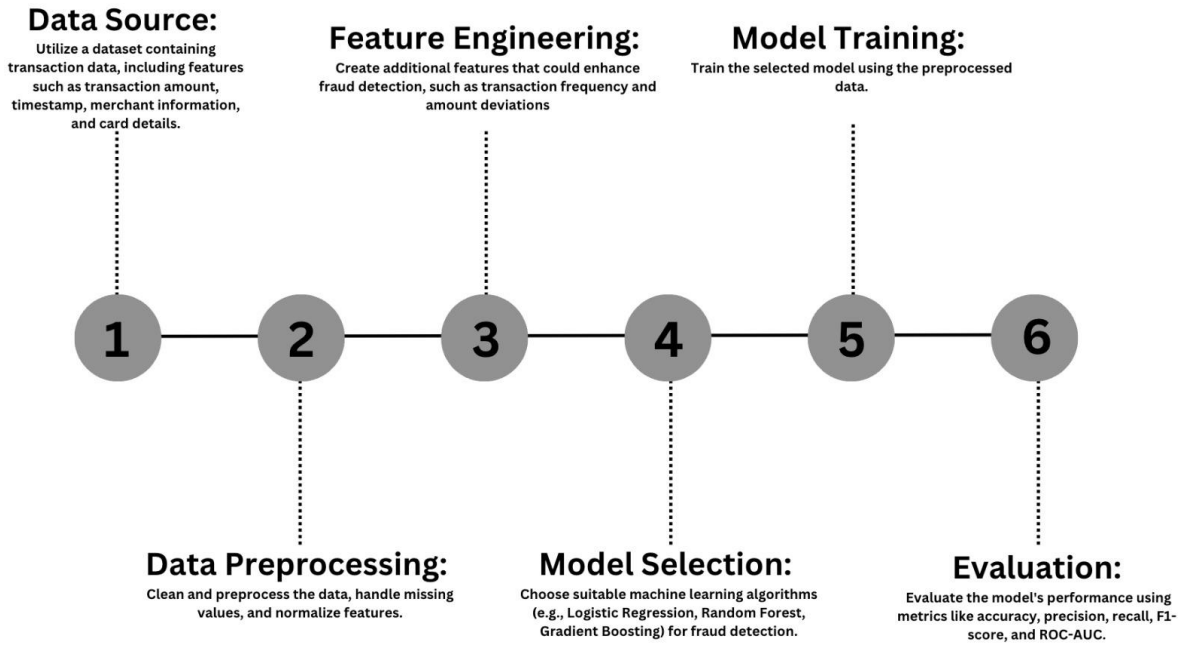
OBJECTIVE:

The problem is to develop a machine learning-based system for real-time credit card fraud detection. The goal is to create a solution that can accurately identify fraudulent transactions while minimizing false positives. This project involves data preprocessing, feature engineering, model selection, training, and evaluation to create a robust fraud detection system.

ABSTRACT:

The proposed system aims at analysing the number of fraud transactions that are present in the dataset. In proposed System, we use random forest, Decision tree and Logistic Regression to classify the credit card dataset. The dataset is classified into trained and test dataset where the data can be trained individually, these algorithms are very easy to implement as well as very efficient in producing better results and can able to process large amount of data. Even for large dataset these algorithms are extremely fast and can able to give accuracy of about over 90%.

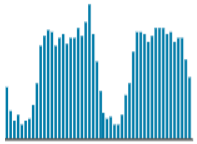
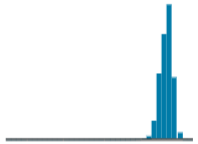
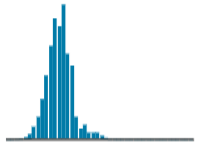
DESIGN THINKING:



DATASET INFORMATION:

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

BigQuery Table bigquery-public-data.fraud_detection.comments

# Time	# V1	# V2	# V3	# V4	# V5	# Amount	# Class
Number of seconds elapsed between this transaction and the first transaction in the dataset	may be result of a PCA Dimensionality reduction to protect user identities and sensitive features(v1-v28)					Transaction amount	1 for fraudulent transactions, 0 otherwise
 0 173k	 -56.4 2.45	 -72.7 22.1	 -48.3 9.38	 -5.68 16.9	 -114 34.8	 0 25.7k	 0 1
0	-1.3598071336738	-0.0727811733098497	2.53634673796914	1.37815522427443	-0.338320769942518	149.62	0
0	1.19185711131486	0.26615071205963	0.16648011335321	0.448154078460911	0.0600176492822243	2.69	0
1	-1.35835406159823	-1.34016307473609	1.77320934263119	0.379779593034328	-0.503198133318193	378.66	0
1	-0.966271711572087	-0.185226008082898	1.79299333957872	-0.863291275036453	-0.0103088796030823	123.5	0
2	-1.15823309349523	0.877736754848451	1.548717846511	0.403033933955121	-0.407193377311653	69.99	0
2	-0.425965884412454	0.960523044882985	1.14110934232219	-0.168252079760302	0.42098688077219	3.67	0
4	1.22965763450793	0.141003507049326	0.0453707735899449	1.20261273673594	0.191880988597645	4.99	0
7	-0.644269442348146	1.41796354547385	1.0743803763556	-0.492199018495015	0.948934094764157	40.8	0
7	-0.89428608220282	0.286157196276544	-0.113192212729871	-0.271526130088604	2.6695986595986	93.2	0

ALGORITHMS USED:

1) RANDOM FOREST
2) DECISION TREE
3) LOGISTIC REGRESSION

RANDOM FOREST:

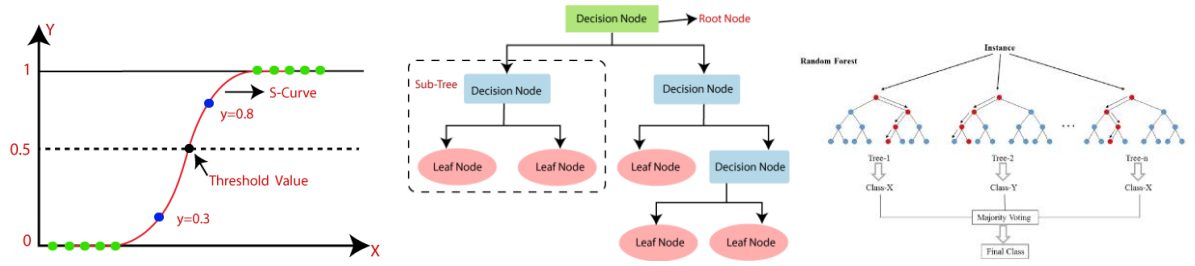
Random Forest is applied to credit card fraud detection by training on historical transaction data, considering features like amount and frequency. After preprocessing and model training, it's evaluated on a testing set using metrics such as precision and recall. Continuous monitoring and periodic retraining ensure the model's effectiveness in real-time fraud detection.

DECISION TREE

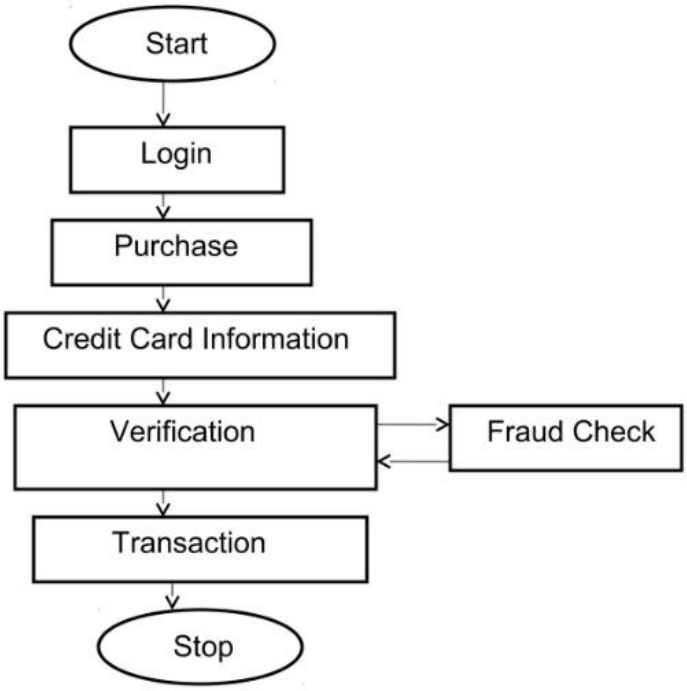
Collect and preprocess credit card data, addressing missing values and scaling features. Opt for a decision tree algorithm, such as Random Forest, balancing simplicity with interpretability. Train the model, assess its performance using metrics like precision and recall. Deploy the model, continuously monitor and update for effective credit card fraud detection in a production environment.

LOGISTIC REGRESSION:

Logistic Regression is employed for credit card fraud detection by training the model on imbalanced data after preprocessing, including feature scaling. Evaluation metrics such as precision, recall, and AUC-ROC score assess the model's performance on the testing set. The adjustment of decision thresholds allows for customization, and feature importance analysis provides insights into the significant predictors. Upon satisfactory results, the model is deployed for real-time fraud detection in a production environment.



FLOW DIAGRAM:



LITERATURE SURVEY:

By, Ghosh S., and Reilly D.L.,

-Due to a rapid advancement in the electronic commerce technology, the use of credit cards has dramatically increased.

By, Syeda M., Zhang Y. Q., and Pan Y.,

-As credit card becomes the most popular mode of payment for both online as well as regular purchase, it provides cashless shopping. It will be the most convenient way to do online shopping, paying card has also been increasing.

By, Stolfo S.J., Fan D. W., Lee W., Prodromidi A., and Chan P. K.,

-In the existing credit card fraud detection system, fraudulent transaction will be detected after transaction is done. It is difficult to find out fraudulent and regarding loses will be barred by issuing authorities.