Project Overview: Credit Card Fraud Detection  
Dataset Details  
Dataset of 120,091 transactions, each described by 30 numerical features and the 'class' label - 1 for fraud  
Skewed dataset since there are 249 fraud cases out of a total of 120,901 transactions.  
Machine Learning Approach Data Preprocessing:  
• Missing values: Some records had missing values in V25, V26, V27, V28, and Amt. They were treated prior to training of model  
• Scaling features : standard scaler for normalization.  
• Imbalance Handling - SMOTE - Synthetic Minority Over-sampling Technique was used for balancing the dataset.  
  
• A random forest classifier with 100 trees was used.  
• The model was trained on a subset of 10,000 samples for faster training.  
• It is done on an 80-20 train test ratio.  
  
Accuracy: 99.5%  
Precision & Recall: The model performed well in identifying fraud cases.  
Confusion Matrix:  
Low number of false positives and false negatives.  
Fraudulent transactions mostly fell under a correct classification  
  
SMOTE greatly enhances the model's performance by balancing the fraud to normal transaction ratio. The Random Forest Classifier showed high accuracy but can be further optimized by hyperparameter tuning. Feature importance analysis could help identify key transaction patterns indicative of fraud.