Title: Fraud Detection Model using Categorical Features

Overview:

In this data science project, we aimed to develop a fraud detection model using categorical features to identify fraudulent and non-fraudulent transactions. The project involved analyzing the distribution of fraud and non-fraud cases across categorical columns, scaling the data using Standard Scaler, applying dimensionality reduction techniques, and ultimately building a classification model.

Data Exploration:

During the initial data exploration phase, we observed that fraudulent cases appeared to be uniformly distributed among the various classes in the categorical columns. There were no clear or glaring predictors that could easily distinguish fraudulent from non-fraudulent transactions. Surprisingly, a similar uniform distribution was observed for non-fraudulent cases.

Data Preprocessing:

To prepare the data for modeling, we performed the following preprocessing steps:

Standard Scaling: We used the Standard Scaler to standardize the numerical features, ensuring that they all have a mean of 0 and a standard deviation of 1. This step was essential for models sensitive to feature scales.

Dimensionality Reduction: We applied dimensionality reduction techniques to reduce the complexity of the dataset while retaining as much information as possible. This step helped to mitigate the curse of dimensionality and improve model performance.

Model Building:

After preprocessing the data, we proceeded to build a classification model. The primary goal was to accurately identify both legitimate and fraudulent transactions. The model achieved an accuracy of approximately 82%, indicating that it can effectively classify transactions into these two categories.

Conclusion:

In conclusion, this data science project focused on developing a fraud detection model based on categorical features. Despite the absence of clear predictors in the categorical columns, our model, which included data scaling and dimensionality reduction, demonstrated a promising accuracy rate of 82% in identifying both legitimate and fraudulent transactions. Further refinements and enhancements can be explored to improve model performance and adapt to evolving fraud patterns.