**Credit Card Transactions Fraud Detection Dataset Project**

**Introduction:**

Credit card fraud has become a common problem in our increasingly digital age, costing people and businesses a lot of money. This issue is being addressed by the Credit Card Transactions Fraud Detection Dataset project, which intends to produce a reliable dataset to make it easier to develop and test machine learning models for detecting credit card fraud. Because it helps to increase the security of financial transactions and protect the interests of both consumers and businesses.

**Objectives:**

The primary objectives of the Credit Card Transactions Fraud Detection Dataset project are as Collecting the dataset of credit card transactions, Cleaning and Preprocess the dataset to make suitable for analysis, Extract relevant features from the data (Feature Engineering), Finalizing the final data table with all proper labels and data, Doing analysis using the dataset, Finding the correlations between the dataset columns, Create the model to find the final outcome (fraud transaction or not), Split the dataset for train, test data and run the model, Check the outcomes of the model. And finally predict the future possibilities of fraud transactions.

**Scope:**

The Credit Card Transactions Fraud Detection Dataset project will have a specified scope and restrictions. It will include credit card transactions made over a specified period, including authorized and fraudulent ones from various merchant types and geographical areas. By eliminating personally identifying information and information connected to the investigation of fraudulent transactions, exclusion criteria will guarantee the privacy protection of cardholders. The dataset will, however, be static and not updated in real-time, so it will only be a historical snapshot. It might not take into consideration how fraud strategies change during the data collecting period and won't provide details on how external economic or market factors affect fraud rates.

**Data Sources:**

Data for the Credit Card Transactions Fraud Detection Dataset project will be sourced from Kaggle. It has around 1.3 million records in Train dataset and 556k in Test dataset. The dataset has columns like Trans date, CC\_Num, Merchant, Category, Amt, Gender, Name, City, State, Zip, Job etc.…

**Methodologies:**

The project will employ a range of data science techniques to create and validate the dataset. This includes data cleaning, which involves the removal of duplicates, handling missing values, and ensuring data consistency. Feature engineering will be employed to extract meaningful attributes such as transaction amount, merchant category, and time of day, enhancing the dataset's quality. To prevent class imbalance issues, the dataset will be carefully balanced between genuine and fraudulent transactions. Additionally, anonymization techniques will safeguard cardholders' privacy, while comprehensive documentation will be created to accompany the dataset. Lastly, data distribution will occur via a secure platform or repository with controlled access, facilitating responsible and secure use by researchers and practitioners.

**Timeline:**

The Credit Card Transactions Fraud Detection Dataset project will follow a carefully planned timeline for execution. We will complete the whole project end of the November.