**Credit Card Fraud Detection**

**Description:**

Credit card fraud detection involves monitoring transactions to spot signs of fraud transactions. The goal is to catch and prevent fraudulent transactions and minimizing false alarms to avoid inconveniencing real customers.

**Business Value:**

Implementing a credit card fraud detection model brings following significant business values to organizations:

* **Fraud Prevention:** Stops unauthorized transactions, saving money from chargebacks.
* **Cost Reduction:** Automates detection, lowering operational expenses.
* **Customer Trust:** Enhances security, boosting customer loyalty and card usage.
* **Competitive Edge:** By safeguarding transactions, the business shows it's trustworthy, earning a good name in the market.

**Research:**

Credit Card Fraud Detection is a popular dataset containing credit card transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset.

We believe that CRISP-ML Q (Cross Industry Standard Process for Machine Learning with Quality Assurance) will ease obtaining efficient results as it follows step-by-step process. Starting by understanding the business and data, preprocessing the data then model building it and finally evaluate the model to make sure it’s performing well.

**`CRISP-ML(Q)`** process model describes six phases:

1. Business and Data Understanding
2. Data Preparation
3. Model Building
4. Model Evaluation
5. Deployment
6. Monitoring and Maintenance

**Business and Data Understanding**

**Objective(s):** Maximize the convenience of Credit Card Fraud Detection.

**Constraint(s):** Minimize the False Negative’s and False Positive’s to avoid unnecessary inconvenience to legitimate customers.

**Success Criteria:**

**Business Success Criteria:** Reduce the Credit Card Fraud transactions and improve the user experience for legitimate customers.

**Machine Learning Success Criteria:** Achieve model accuracy at least 90%. Balanced performance to minimize both false positives (legitimate transactions flagged as fraud) and false negatives (fraudulent transactions not flagged).

**Economic Success Criteria:** Financial organization who provide credit card services should achieve increase in business revenue when compared to previous business.

**Data Collection/Description:**

**Data:** The Credit Card Fraud Detection Dataset is provided by Kaggle and publicly available to use.

**Data Dictionary:**

- Dataset contains 31 columns/features

- Dataset contains 284807 records

**Description:**

* **Id** - Unique identifier for each transaction
* **Time** - Number of seconds elapsed between this transaction and the first transaction in the dataset
* **V1 - V28** - PCA dimensionality reduction data to protect user identities and sensitive features
* **Amount** - The transaction amount
* **Class** - Binary label indicating whether the transaction is fraudulent (1) or not (0)

You can download the dataset from here [Click here](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud)

The dataset contains transactions made by credit cards in September 2013 by European cardholders.  The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'.

**Exploratory Data Analysis:**

As part or EDA, we used D-Tale, which is an Auto EDA library which gives the complete insights on data. We can able to see Descriptive Statistics and Heat map which shows the correlation among the variables.

**Data Preprocessing/Data Wrangling/Data Cleaning**

The provided dataset is already PCA transformed, there isn’t much preprocessing needed. There are no missing/Null values in the dataset. Duplicates are identified and rectified. We don’t have any values near zero or zero variance.

**Graphical Representation:**

Histograms is used to identify the patterns and distribution of the data.

**Model Building/Modelling**

**Logistic Regression:** Logistic regression is a practical and reliable choice for credit card fraud detection, offering a good balance between simplicity, performance, and ease of use. Logistic regression is inherently designed for binary classification tasks, making it a natural fit for fraud detection where the outcome is typically fraudulent or not fraudulent.

**Training:** Data training is achieved by splitting the data into two parts using class feature where target value is 1 or 0. If it is 1, it is a fraud transaction and if it is 0, it is valid transaction.

Checking total count of valid and fraud transaction can give an overview of whether the data is biased or unbiased. In this case, data is biased, so we considered to take a sample data from the original data which consist of 473 records of fraud transaction and 473 records of valid transactions to make it balanced.

We used train\_test\_split method to split the data into training and test data. Training data is fitted to Logistic regression algorithm, which is imported from Sci-kit learn.

**Metrics:** By deriving the Classification report, we can able to see the accuracy scores, precision and Recall. We can also use Confusion matrix to know the accuracy of the model.

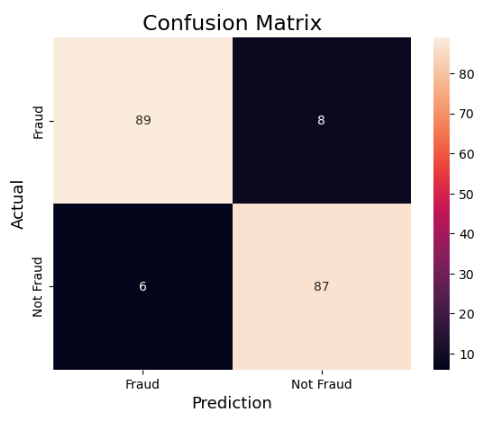
Here, we evaluated the model using various metrics, and here are the results:

**Accuracy score:** 92.63%

**Precision:** 91.57%

**F1 Score:** 92.57%

**Confusion Matrix:**

****

**Conclusion:**

In conclusion, this work has demonstrated the potential of logistic regression as a valuable tool for credit card fraud detection. The developed model achieved [accuracy, precision, F1 score] in identifying fraudulent transactions. By analyzing various features of credit card transactions, the model was able to distinguish between legitimate and fraudulent activities with a high degree of accuracy.