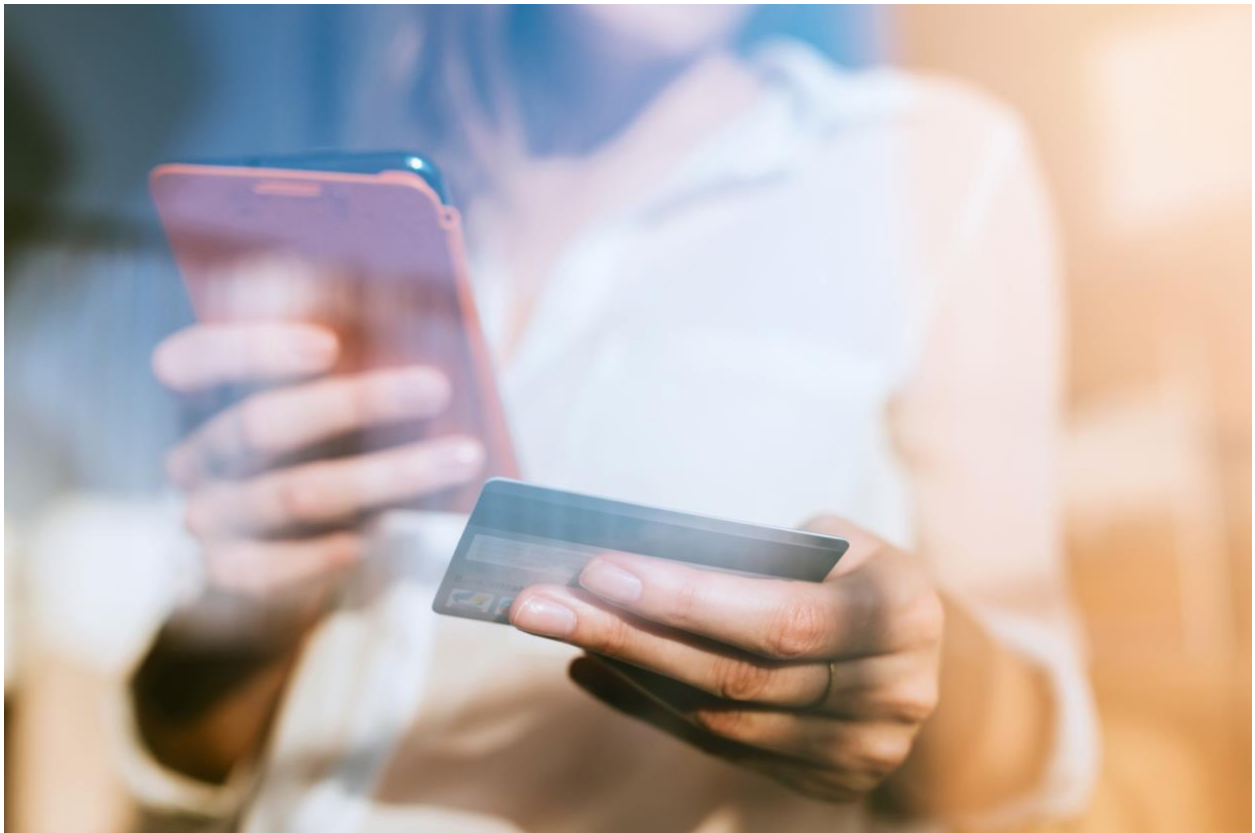




*Predictive classification model for identifying fraudulent credit card transactions*



*Project Proposal by: Himani Kaushik*

- **Question/need:**
  - What is the question behind your analysis or model and what practical impact will your work have?
    - The purpose of the model is to predict fraudulent credit card transactions using various classification methods. A credit card fraud involves obtaining the credit card information without the proper authorization of the account holder to engage in an illegal financial transaction. This results in financial loss as well as the loss of creditability and trust and affects both the customers and the credit card company. This model can be implemented to detect credit card fraud and prevent the said losses.
  - Who is your client and how will that client benefit from exploring this question or building this model/system?
    - A credit card company, CreditCards, has decided to proactively monitor and implement fraud prevention mechanisms. It plans to build machine learning models to predict the fraudulent transactions to not only prevent financial losses, but also reduce chargebacks and fees, save time on lengthy manual reviews and decrease denials of legitimate transactions.
- **Data Description:**
  - What dataset(s) do you plan to use, and how will you obtain the data?
    - The data will be obtained from the Kaggle website - <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>.
    - The dataset contains credit card transactions made by European cardholders in September 2013. It consists of 284,807 transactions, out of which 492 are frauds. The dataset is highly unbalanced and the positive class (frauds) account for 0.172% of all transactions.
  - What is an individual sample/unit of analysis in this project?
    - The individual sample includes time, amount, class and other numerical input variables V1, V2,...V28, which have been transformed due to confidentiality issues.
  - If modeling, what will you predict as your target?
    - The feature 'Class' is the response variable and indicates whether the transaction is fraudulent or not.
- **Tools:**
  - How do you intend to meet the tools requirement of the project?
    - Following Python libraries would be used:
      - Pandas and Numpy: For EDA and cleaning data.
      - Scikit-learn: For implementing various classification models and performing cross validation, regularization, hyperparameter tuning and feature engineering.
      - Matplotlib and Seaborn: For visualizing the data.

- **MVP Goal:**
  - What would a minimum viable product (MVP) look like for this project?
    - MVP for the project would be the implementation and comparison of various classification models like k-nearest neighbors, logistic regression, random forest, etc. to predict the likelihood of a credit card transaction being fraudulent.