**Documentation**

**README.md**

|  |
| --- |
| #Project 1-Fraud Detection  ## Overview   * The main objective of this project (e.g., fraud detection based on historical data). * A brief explanation of the approaches used (Random Forest, Logistic Regression, and AdaBoost).   ## Dataset   * Data source: https://drive.google.com/drive/folders/1tOUKAgKpE897iqKLo2b8yGmL2W-BdNj8 * Description of key features ('Unnamed: 0', 'trans\_date\_trans\_time', 'cc\_num', 'merchant', 'category', 'amt', 'first', 'last', 'gender', 'street', 'city', 'state', 'zip', 'lat', 'long', 'city\_pop', 'job', 'dob', 'trans\_num', 'unix\_time','merch\_lat', 'merch\_long', 'is\_fraud').   ## Tools & Technologies Used   * **Environment:** Google Colab / Jupyter Notebook / Local Python Environment. * **Modeling Tools:** * Random Forest using from sklearn.ensemble import RandomForestClassifier * Logistic Regression using from sklearn.linear\_model import LogisticRegression * AdaBoost using from sklearn.ensemble import AdaBoostClassifier(estimator=DecisionTreeClassifier(max\_depth=1), n\_estimators=50, random\_state=42)   ## Data Preprocessing   * Feature normalization / standardization. * Encode variable * Splitting data into train and test sets.   ## Model Development   * Logistic Regression. * Random Forest: * AdaBoost   ## Model Evaluation & Comparison   * Evaluation metrics: Model, Resampling, Accuracy, Precision, Recall, F1-score, Confusion Matrix * Performance comparison of Importance Score. |