Task for ML Intern at Dunlin.ai

Objective:

Use the provided data in the CSV file to clean it thoroughly and then predict the values of the `Split` column. Candidates can use any approach they deem fit but must document the entire process, including the techniques used for data cleaning, forming an algorithm, generalizing the data, and other important steps.

Instructions:

1. **Data Cleaning:**
   - Load the dataset and inspect its structure.
   - Handle missing values appropriately.
   - Ensure consistency in data types and formats.
   - Remove or correct any erroneous data entries.
   - Normalize the data if necessary.
2. **Exploratory Data Analysis (EDA):**
   - Conduct a thorough analysis of the dataset to understand its distribution and patterns.
   - Visualize the data to identify trends and outliers.
3. **Feature Engineering:**
   - Create new features if necessary to improve the predictive power of your model.
   - Select relevant features for the model.
4. **Model Development:**
   - Split the data into training and testing sets.
   - Choose an appropriate machine learning model to predict the `Split` column.
   - Train the model using the training data.
   - Evaluate the model using the testing data.
5. **Documentation:**
   - Write a detailed document explaining the entire process:
     - How the data was cleaned.
     - The techniques used to handle missing values and outliers.
     - The feature engineering process.
     - The choice of model and the rationale behind it.
     - The training and evaluation process.
     - Any challenges faced and how they were overcome.

- ○ Include visualizations and code snippets where necessary.
6. **Submission:**
   - ○ Share the Colab notebook link and the document link via email to shubham@saasden.club.
   - ○ Ensure proper permissions are given to access the documents.

Data Overview:

The dataset consists of the following columns:

- Date: Date of the transaction.
- Transaction Type: Type of transaction (e.g., Deposit).
- Name: Name associated with the transaction.
- Memo/Description: Description or memo for the transaction.
- Split: The column to predict.
- Amount: Amount involved in the transaction.
- Balance: Balance after the transaction.

Feel free to use any tools or libraries you find appropriate for this task. Good luck, and we look forward to your submission!

If you have any questions, please reach out to shubham@saasden.club.

InternData

Also if you need any insights on the data you can watch some of the bookkeeping videos listed down below :-
- The BOOKKEEPING BASICS for BEGINNERS
- Learn Financial Accounting form from Basics to Advance - YouTube