**Underwriter Model Project**

**Overview**

The provided dataset contains over 500,000 transaction records from a loan mortgage account.

**Dataset Details**

* Confidentiality: The dataset is strictly confidential.
* Size: Over 500,000 records.
* Content: Contains various loan mortgage transactions.

**Data Preparation**

* Data Cleaning: The dataset includes some missing values. Candidates are expected to handle data cleansing and use only non-redundant columns.
* Feature Selection/Engineering: Candidates should perform feature engineering to enhance the predictive power of the model.

**Objective**

Build a prediction model to identify loan "Default" status.

Loan Status Categories:

\*\*Charged Off (Bad loans)

\*\*Fully Paid\*\* (Considered as good loans)

\*\*Current\*\* (Could turn into good or bad loans in the following month)

\*\*Default\*\* (Bad loans)

\*\*Late\*\* (Subjective interpretation)

- Only "Fully Paid" loans should be considered good.

- "Current" status comprises about 70% of the dataset and must be used creatively rather than discarded.

**Deliverables**

Formats Accepted: Jupyter Notebooks, native Python code, or a combination of both.

**Additional Resources**

Data Dictionary: An attached data dictionary provides explanations for the columns in the CSV records.

**Tips for Solution**

1. We value your personal time. This is not a timed test, but we expect the project to be completed within a day.
2. Framing the Problem: Focus on how you translate the business problem into a data science problem, and how you convert the results back into solutions that address the underlying business needs.
3. Utilizing 'Current' Loan Data: Ensure you use and extract insights from data points in the "Current" loan category. Models that drop all "Current" loan status data will be disqualified.
4. Exploratory Data Analysis (EDA): Full EDA is not required. Conduct EDA just enough to assist with feature selection and/or feature engineering. Instead of extensive plots, provide one or two key plots that summarize your findings.
5. Hyperparameter Tuning: You can fully skip spending time on hyperparameter tuning.
6. Modeling Approach: Rather than focusing on delivering a highly accurate model, concentrate on completing the data science project workflow. You may entirely skip building any deep learning models.

By following these guidelines, we aim to assess your ability to approach, structure, and solve a data science problem within a practical timeframe. Good luck!