**High level design (HLD)**

****

**PW SKILLS: FULL STACK DATA SCIENCE PRO**

***Submitted by: POONAM SAINI***

**Credit Card Default Prediction**

**INTRODUCTION**

In today's increasingly digital financial landscape, managing credit risk is more important than ever. A credit card default occurs when a cardholder doesn't make the agreed payments on their credit card debt. This can occur due to various reasons, such as unemployment, medical emergencies, or financial difficulties. Defaulting on credit card payments can have serious consequences, including harm to credit scores, accumulating debt from interest and penalties, and the possibility of legal action by creditors.

Predicting credit card defaults is vital as it helps financial institutions identify at-risk customers early. By analyzing data like demographics, payment history, and spending behavior, predictive models can estimate the likelihood of defaults. This allows banks to take steps like offering financial guidance, adjusting credit limits, or restructuring payment plans to lower the risk of default and minimize financial losses.

**PROBLEM STATEMENT**

Financial threats are displaying a trend about the credit risk of commercial banks as the incredible improvement in the financial industry has arisen. In this way, one of the biggest threats faced by commercial banks is the risk prediction of credit clients. The goal is to predict the probability of credit default based on the credit card owner's characteristics and payment history.

**Data Insights**

This dataset includes details on default payments, demographic characteristics, credit information, payment history, and bill statements for credit card clients in Taiwan, covering the period from April 2005 to September 2005.

**Data Variables Overview**

This dataset consists of 25 variables, described as follows:

**ID**: Unique identifier for each client.

**LIMIT\_BAL**: Total credit amount provided in NT dollars (includes both individual and family/supplementary credit).

**SEX:** Gender of the client (1 = male, 2 = female).

**EDUCATION**: Educational background (1 = graduate school, 2 = university, 3 = high school, 4 = others, 5 = unknown, 6 = unknown).

**MARRIAGE:** Marital status (1 = married, 2 = single, 3 = others).

**AGE:** Client's age in years.

**PAY\_1:** Repayment status for September 2005 (-1 = pay duly, 1 = payment delayed by one month, 2 = payment delayed by two months, and so on up to 9 = payment delayed by nine months or more).

**PAY\_2:** Repayment status for August 2005 (same scale as above).

**PAY\_3:** Repayment status for July 2005 (same scale as above).

**PAY\_4:** Repayment status for June 2005 (same scale as above).

**PAY\_5:** Repayment status for May 2005 (same scale as above).

**PAY\_6:** Repayment status for April 2005 (same scale as above).

**BILL\_AMT1:** Amount on the bill statement for September 2005 (in NT dollars).

**BILL\_AMT2:** Amount on the bill statement for August 2005 (in NT dollars).

**BILL\_AMT3:** Amount on the bill statement for July 2005 (in NT dollars).

**BILL\_AMT4:** Amount on the bill statement for June 2005 (in NT dollars).

**BILL\_AMT5:** Amount on the bill statement for May 2005 (in NT dollars).

**BILL\_AMT6:** Amount on the bill statement for April 2005 (in NT dollars).

**PAY\_AMT1:** Amount of payment made in September 2005 (in NT dollars).

**PAY\_AMT2:** Amount of payment made in August 2005 (in NT dollars).

**PAY\_AMT3:** Amount of payment made in July 2005 (in NT dollars).

**PAY\_AMT4**: Amount of payment made in June 2005 (in NT dollars).

**PAY\_AMT5:** Amount of payment made in May 2005 (in NT dollars).

**PAY\_AMT6:** Amount of payment made in April 2005 (in NT dollars).

**default.payment.next.month:** Indicates whether the client defaulted on payment in the next month (1 = yes, 0 = no).

**Tools and Technologies Used**

For building the Credit Card Default Prediction model, we utilized several tools and technologies:

* Python Programming Language: The core language used for developing the model.
* NumPy: Employed for efficient numerical computations and handling arrays.
* Pandas: Utilized for data manipulation and preprocessing.
* Scikit-learn: Applied for model building, training, and evaluation.
* Matplotlib: Used for creating static, animated, and interactive visualizations.
* Seaborn: Leveraged for statistical data visualization and enhanced graphical representations.
* Streamlit: Implemented for deploying the model as an interactive web application.

**Development Details:**

**Workflow and Deployment:**

**Conclusion:**

This project is built using Flask, making it easily accessible to all users, regardless of their technical background. By implementing the design process outlined above, banks and loan lenders can accurately predict whether customers are likely to default on credit card payments. This enables the relevant departments to take proactive measures, such as adjusting credit limits or offering financial counseling, based on the model's predictions.

The user interface is intentionally user-friendly, ensuring that even those with minimal technical expertise can navigate the system with ease. Users only need to input the required information to obtain accurate results, making the tool both practical and efficient for real-world applications. Additionally, the model's flexibility allows it to be adapted and scaled to meet the needs of various financial institutions, ultimately helping to reduce the risk of defaults and improve overall financial stability