Project Proposal: Predicting Credit Card Approval

Section 1: Question and Answers

1. Why is your proposal important in today’s world? How predicting a good client is worthy for a bank?

* Predicting creditworthiness is crucial for banks to mitigate risks associated with credit card approvals. In today's world, with increasing financial transactions and data availability, leveraging predictive analytics can enhance the efficiency of credit assessments, reducing the likelihood of defaults and improving overall portfolio performance.

1. How is it going to impact the banking sector?

* Implementing an effective credit card approval prediction model can lead to better decision-making, minimizing financial losses due to defaults. It enhances customer satisfaction by ensuring that credit cards are offered to individuals who are more likely to manage their credit responsibly. This can contribute to a healthier credit portfolio, increased customer trust, and long-term profitability.

1. If any, what is the gap in the knowledge or how your proposed method can be helpful if required in the future for any bank in India?

* The current credit approval processes might have limitations in accurately assessing an individual's creditworthiness. By leveraging advanced data analysis and machine learning, the proposed method aims to bridge this gap. If successful, it can serve as a template for banks in India to enhance their credit evaluation systems, making them more robust and adaptive to changing economic conditions.

Section 2: Initial Hypothesis

For Data Analysis (DA) Track:

Initial Hypothesis: Certain demographic features such as annual income, education level, and employment status may exhibit patterns that correlate with credit card approval or rejection.

For Machine Learning (ML) Track:

Initial Hypothesis: Machine learning models, trained on historical credit card application data, can effectively predict credit card approval. Features like income, education, and employment history are likely to play significant roles.

Section 3: Data Analysis Approach

1. What approach are you going to take in order to prove or disprove your hypothesis?

* Conducting exploratory data analysis (EDA) to identify patterns and correlations.
* Analyze summary statistics of key features to understand their distributions.
* Visualizing relationships between features and the target variable.
* And exploring the impact of different demographic factors on credit approval using statistical tests.

1. What feature engineering techniques will be relevant to your project?

* One-hot encoding for categorical variables.
* Log transformation for skewed numerical variables.
* Handling missing values through imputation.
* Creating new features based on existing ones, like age from birthday\_count.

1. Please justify your data analysis approach.

* EDA helps understand the data's structure and relationships, guiding feature engineering decisions.
* Visualizations provide insights into potential patterns and help in making informed decisions during the modeling phase.
* Identifying important patterns in the data using the EDA approach to justify findings.

1. Explore income distribution across different demographics.

* By analyzing the relationship between education level and credit approval.
* Investigating the impact of employment history on credit decisions.

Section 4: Machine Learning Approach

1. What method will you use for machine learning based predictions for credit card approval?

* Random Forest Classifier due to its ability to handle complex relationships and feature importance analysis.

1. Please justify the most appropriate model.

* Random Forest is an ensemble method that performs well on diverse datasets. It handles non-linearity and provides insights into feature importance, aiding interpretability.

1. Please perform necessary steps required to improve the accuracy of your model.

* Hyperparameter tuning using grid search.
* Feature scaling and engineering.
* Handling class imbalance through oversampling.

1. Please compare all models (at least 4 models).

* Logistic Regression, Gradient Boosting, and K-Nearest Neighbors will be considered for comparison.
* And for evaluation metrics i include accuracy, ROC AUC, F1 score, and precision-recall curves.