Capstone project 1- Project proposal to predict credit card approval

Section 1: Questions to Answer

1.What questions do you want to answer?

- Can we predict credit card approval based on customer information effectively?

- How can different customer attributes influence credit card approval or rejection?

- Which machine learning model is the most suitable for predicting credit card approval?

- How does feature engineering impact the model's performance?

- What is the potential impact of improving credit card approval prediction for a bank?

2. Why is your proposal important in today's world? How predicting a good client is worthy for a bank? How predicting a good client is worthy for a bank? How is it going to impact the banking sector? 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 proposal is important in today's world for several reasons:

- **Risk Mitigation:** Predicting credit card approval accurately helps banks minimise the risk of approving cards for customers who may default on payments. This reduces financial losses for the bank.

- **Customer Experience**: Accurate predictions ensure that customers who are likely to be responsible cardholders are approved, leading to a better customer experience.

- **Cost Reduction**: Effective prediction can reduce the cost of managing and recovering debts from defaulters.

- **Competitive Advantage:** Banks that can efficiently evaluate credit card applications are more competitive in attracting responsible customers.

- **Data-Driven Decisions:** The proposal emphasizes the importance of data-driven decisions, which is increasingly crucial in the banking sector

The impact on the banking sector includes:

- Improved Profitability: Banks can increase profitability by approving more credit cards for low-risk customers.

- Enhanced Customer Relations: A streamlined approval process based on data analysis can lead to improved customer relations

- Risk Management: Better credit card approval prediction can enhance risk management strategies.

- Compliance: Banks can ensure compliance with regulatory requirements by making objective decisions based on data.

The knowledge gap proposal addresses is the application of the advanced data analysis and machine learning techniques to the problem of credit card approval. It can serve as a template for other banks in India or around the world looking to optimize their credit approval processes using data-driven methods.

Section 2: Initial Hypothesis (or hypotheses)

Initial Hypothesis:

- **Hypothesis 1:** Customer attributes such as age, income, and employment history will significantly influence credit card approval. Younger individuals with stable jobs and higher incomes are more likely to get approved.

- **Hypothesis 2:** Machine learning models, such as Random Foret and XGBoost will help in predicting credit card approval due to their ability to capture complex relationships in the data when compare to other models.

- **Hypothesis 3:** Feature engineering techniques such as logarithmic transformation of income and square root transformation of family members, improved model performance by making data distributions more suitable for modelling

Section 3: Data analysis approach

- **Approach:** Started with exploratory data analysis (EDA) to understand the distribution of variables, identify outliers, and found the correlations between features and the target variable. Later used visualisations and found the relationships using graphs and plots.

- **Feature Engineering**: Applied feature engineering techniques such as logarithmic transformation and square root transformation to relevant variables based on data distribution and skewness analysis.

- **Statistical Tests:** Conducted statistical tests to validate initial hypotheses, including t-tests and ANOVA to compare means of different groups

- **Visualization:** Used visualization techniques like box plots, histograms and heatmaps to uncover patterns and relationships in the data.

- **Model Evaluation:** To justify, used evaluation techniques with machine learning models using accuracy, precision, recall, F1-score as metrics.

In the given project, both the initial given datasets were loaded and inspected. Data preprocessing steps were performed, including handling missing values, calculating age and employed years and addressing outliers in key features. Exploratory data analysis (EDA) was conducted to visualise the data distributions, identify outliers and understand feature relationships and performed feature engineering techniques. Categorical columns were one-hot encoded to prepare the data for machine learning models.

Section 4: Machine learning approach

- **Method**: Used various machine learning models including Logistic Regression, Random Forest, XGBoost and Decision Tree to predict credit card approval based on customer information.

- **Justification:** Random Forest and XGBoost are expected to perform well due to their ability to capture non-linear relationships. Logistic Regression will serve as a baseline. Decision Tree may overfit but will provide insights into feature importance

- **Hyperparameter Tuning:** Need to perform hyperparameter tuning using GridSearchCV to optimize the Random Forest model.

- **Model Comparison**: Compared all four models based on accuracy, precision, recall and F1-Score. The best performing model will be selected as the final model.

- **Final Model**: The final model will be evaluated on the test set, and a detailed classification report will be provided.

In this approach, Random Forest model after hyperparameter tuning demonstrated the highest accuracy and was selected as the final mode.

Conclusion

- The project successfully demonstrated the feasibility of predicting credit card approval based on customer information using machine learning techniques.

- The Random Forest model with appropriate hyperparameter achieved an accuracy of approximately 92.58% on the test data, making it a reliable choice for credit card approval prediction.

- Feature engineering played a crucial role in improving model performance by transforming data distributions and reducing the impact of outliers. This project highlights the importance of data analysis and machine learning in streamlining the credit approval process for banks resulting in reduced risks, enhanced customer experience and improved profitability.