Loan Tap Case Study

Problem Statement

LoanTap deploys formal credit to salaried individuals and businesses 4 main financial instruments:

1. Personal Loan
2. EMI Free Loan
3. Personal Overdraft
4. Advance Salary Loan But the main focus is to interpret the underwriting process behind the Personal Loan only

Given a set of attributes for an Individual, determine if a credit line should be extended to them. If so, what should the repayment terms be in business recommendations?

- Additional views
  - We need to track the users previous credit line history and repayment status.
  - Analyzing the previous loans tenure and the total liability.
  - As we are focusing more on salaried individuals, we need to take the salary of the person into consideration.

Solution - Colab Link

Actionable Insights & Recommendations

- 80% of customers have fully paid their loans, while 20% have defaulted.
- The trained model predicts a 94% f1-score for the negative class (Fully Paid) and a 62% f1-score for the positive class (Charged Off).
- Cross-validation accuracy and testing accuracy being nearly the same indicate the model's decent performance and reliability on unseen data.
- To enhance the model's performance, options like collecting more data, employing a more complex model, or fine-tuning hyperparameters can be considered.
- The ROC AUC curve, with an area of 0.73, suggests the model correctly classifies around 73% of instances, signifying a good but improvable performance.
- The precision-recall curve visualizes the trade-off between precision and recall based on varying thresholds, enabling identification of the optimal point for specific application needs.
- Post balancing the dataset, significant improvements were not observed in precision and recall scores for both classes.
- The Logistic Regression Classifier achieves an accuracy of 0.891 on the test set, reflecting a decent performance, not merely by chance.

Questionnaire

1. **What percentage of customers have fully paid their Loan Amount?**
   Around 80.38% of customers have fully paid their Loan Amount.

2. **Comment about the correlation between Loan Amount and Installment features.**
   There is a very high correlation between loan amount and installment. pearson coefficient is ~ 0.95. This indicates high multicollinearity between these two features. Hence the installment column is dropped while creating the model.

3. T**he majority of people have home ownership as _____.**
   Mortaged and then followed by rent. With 50.07% and 40.35% in order.

4. **People with grades 'A' are more likely to fully pay their loan. (T/F)**
   True. Out of all people with grade 'A', ~ 92-93% got their loan approved.

5. **Name the top 2 afforded job titles.**
   Teacher & Manager

6. **Thinking from a bank's perspective, which metric should our primary focus be on..**
   a. ROC-AUC : Not a good metric to consider as we have highly imbalanced data.
   b. Precision: Consider when only want to reduce NPA
   c. Recall: Consider when we do not want to miss good opportunity for providing loan to customer
   d. F1-score: Good metric for us, because we want to consider both Precision and Recall.
      Answer - F1-score

7. **How does the gap in precision and recall affect the bank?**
   - The disparity between precision and recall impacts risk assessment for lending decisions.
   - Higher precision ensures more accurate identification of potential defaulters, aiding in cautious decisions.
   - Lower recall could result in missing actual defaulters, leading to missed opportunities.
   - Balancing precision and recall affects credit denials and default rates.

- Striking a suitable balance is crucial for risk management aligned with business goals.

8. **Which were the features that heavily affected the outcome?**
   Zip_code ,grade ,dti , open_acc ,revol_util ,loan_amnt ,purpose_small_business ,home_ownership ,int_rate
9. **Will the results be affected by geographical location? (Yes/No)**
   Yes. Few states have high changes of getting loan compare to other states.