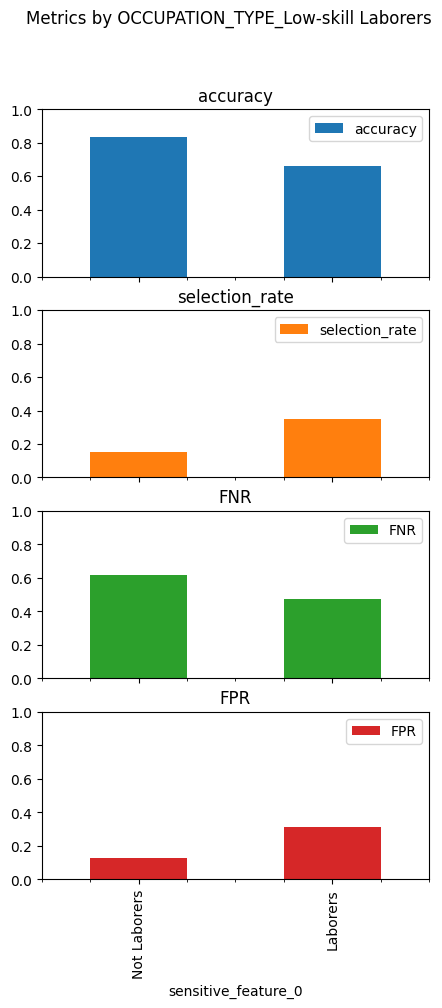
Outcomes

a. Since the primary objective of the ADS is to predict loan repayment ability, the accuracy metric is a critical evaluation metric for the system. Accurate models are highly valuable to stakeholders like banks and financial agencies. The accuracy of an ADS lies in its capacity to pinpoint individuals who are at a higher risk of defaulting on their loans. To analyze the accuracy of the ADS across different subpopulations, we can compare the accuracy rates of the three models on various subpopulations. These subgroups are drawn based on the level of education, financial stability, property ownership, family status, and occupation type. For the baseline logistic regression model, we observe an accuracy ranging from 0.72 to 0.84, indicating variability in accuracy across different subpopulations; for the random forest baseline model, we observe accuracy ranging from 0.82 to 0.85, indicating a fairer and more accurate performance; at last, we observe a similar performance in the random forest model with poly-features included. This indicates that the poly-features chosen by the author do not improve prediction accuracy, though it can have an effect on fairness.

b. To evaluate the fairness of the model, we can utilize various fairness metrics, such as demographic parity difference and demographic parity ratio. Demographic parity difference measures the discrepancy in the proportion of positive outcomes between different subgroups, while demographic parity ratio represents the ratio of positive outcomes between subgroups.

After examining the three models based on different sub-populations, it appears that loan requests for lower amounts and low-income individuals are more likely to be rejected. The most unfair attribute across all three models is by far occupation type. the ADS exhibits discrimination against physical laborers, with low-skilled labor having the highest Demographic Parity Ratio with all models. It has a demographic parity ratio of 0.263, while the random forest model and the poly-feature random forest model has slightly higher demographic parity ratio of about 0.43. Even on our fairest model, the selection rate of low-skill laborers doubles the selection rate of non-low-skill laborers, meaning the model is twice as likely to reject a low-skill laborer. Plot 1 and 2 below shows fairness metrics based on occupation type of low-skill laborers, compared to an example of a non-physical labor heavy occupation, high-skill tech staff.

Figure 1 Figure 2

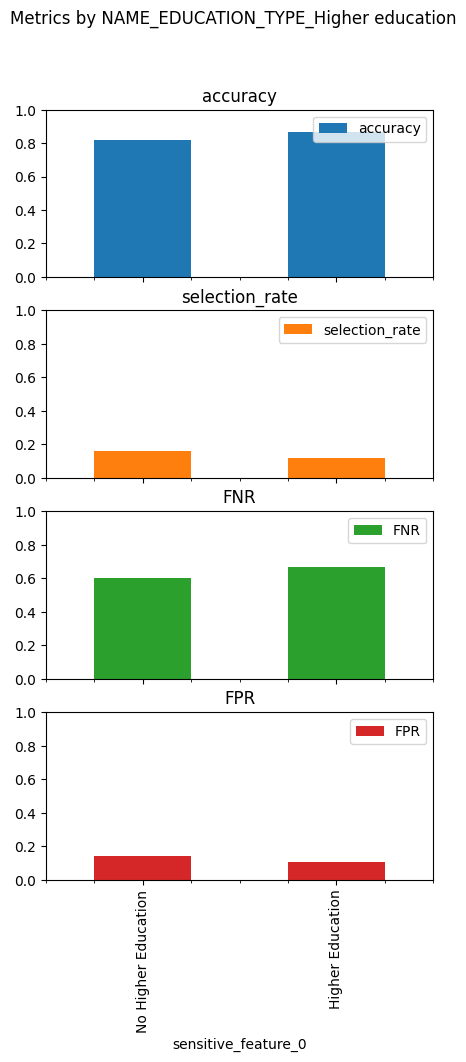
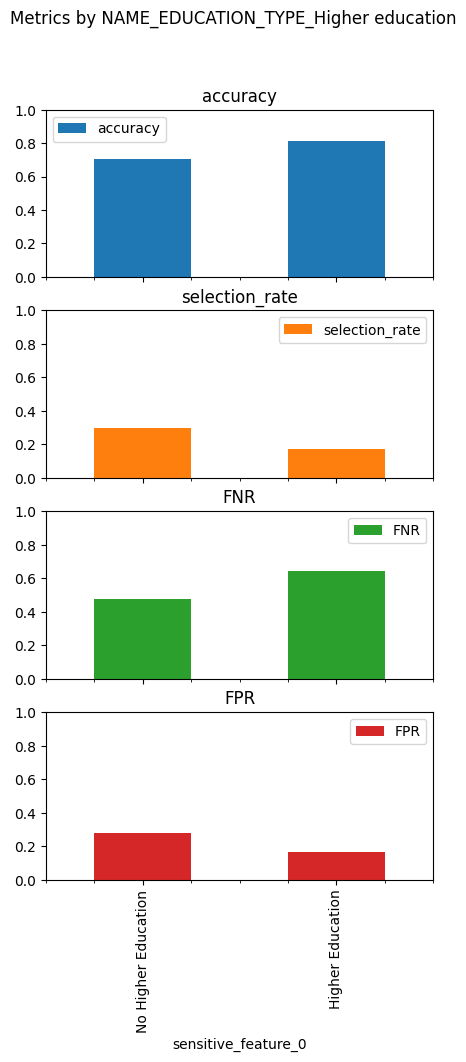


The model appears to demonstrate a similar likelihood of rejecting both high-skill tech staff and their non-high-skilled counterparts. However, there's a noticeable bias against low-skill laborers, who are rejected more frequently compared to their non-low-skilled counterparts. This discrepancy is further emphasized by the imbalance in the false negative and false positive rates within the low-skill laborer subgroup, with the former experiencing a notably higher rate of false rejections.

Additional inconsistencies exist, though they are not as severe as those related to occupation type. For example, AMT\_CREDIT, representing the requested loan amount, shows a tendency of the model to reject smaller loans, leading to a demographic parity ratio of 0.423.

When education level is treated as a sensitive attribute, a marked improvement in fairness metrics can be observed with the poly-feature random forest model. The fairness metrics associated with higher education and secondary education as sensitive attributes are clearly demonstrated in the subsequent plots 3 and 4.

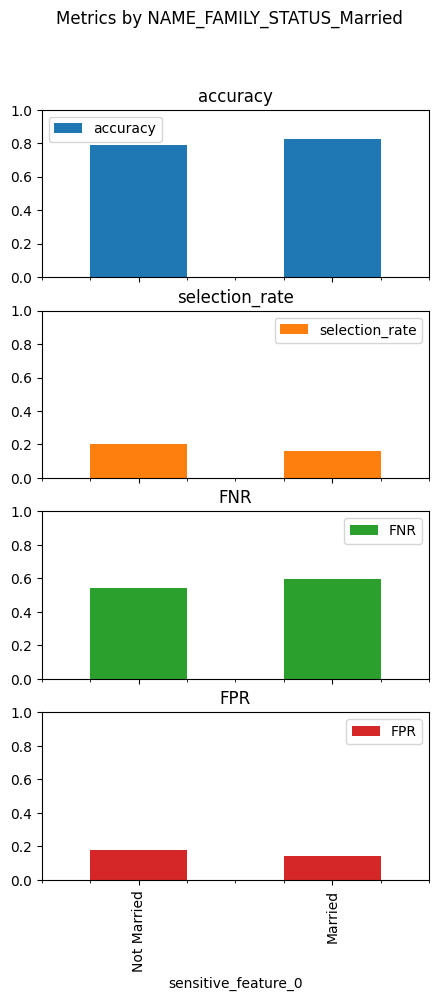
Figure 3: Log regression model metrics Figure 4: Poly-feature model metrics



We observe a notable enhancement in the fairness of the selection rate, False Negative Rate (FNR), and False Positive Rate (FPR) for subgroups with higher education. A similar trend is also seen for those with secondary education, indicating that the model's predictions show a preference for individuals with higher levels of education.

Plots 5 and 6 present metrics when family status is applied as a sensitive attribute. Plot 5, which utilizes the feature CNT\_CHILDREN, indicates that those with children are more likely to be denied a loan compared to those without children. However, Plot 6 reveals that marital status has minimal impact, with a demographic parity difference of only 0.044. These results were generated using poly-feature models and show minimal improvements in fairness metrics compared to both the baseline logistic regression model and the baseline random forest model.

Figure 5: Figure 6:



c. Property ownership has a higher demographic parity ratio, indicating that the basic logistic model is more biased towards property ownership, which is evident in both car and real estate ownership. Therefore, an alternative model is necessary to address this parity difference. One limitation of the basic logistic model is its assumption of linear relationships between input features and output variables. In reality, many relationships between the features and output may be non-linear and complex, leading to reduced model performance and stability. The random forest model can overcome this limitation by capturing non-linear and complex relationships using multiple decision trees with randomized feature subsets. This approach can improve performance and stability, while the use of SHAP or LIME techniques can aid in identifying potential weaknesses or biases in the model and explaining the model's predictions.

The plots of top shap values for each model are shown below.

Figure 7: Shap values for logistic regression model

A picture containing text, screenshot, diagram, font

Description automatically generated

Figure 8: Shap values for baseline random forest model

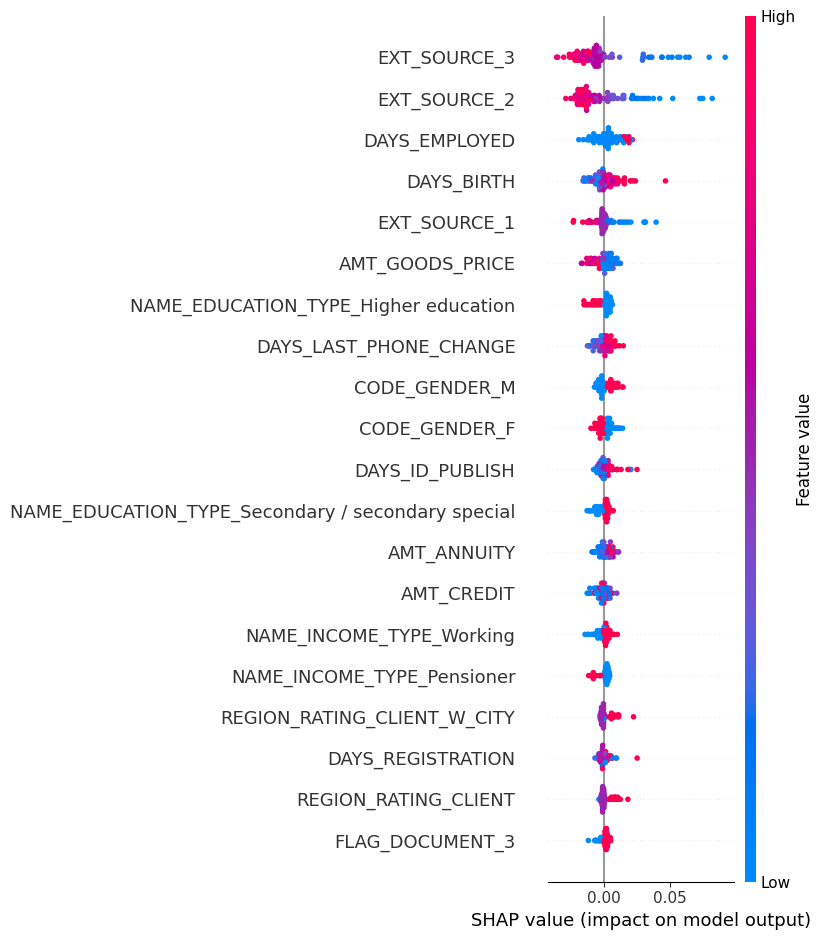
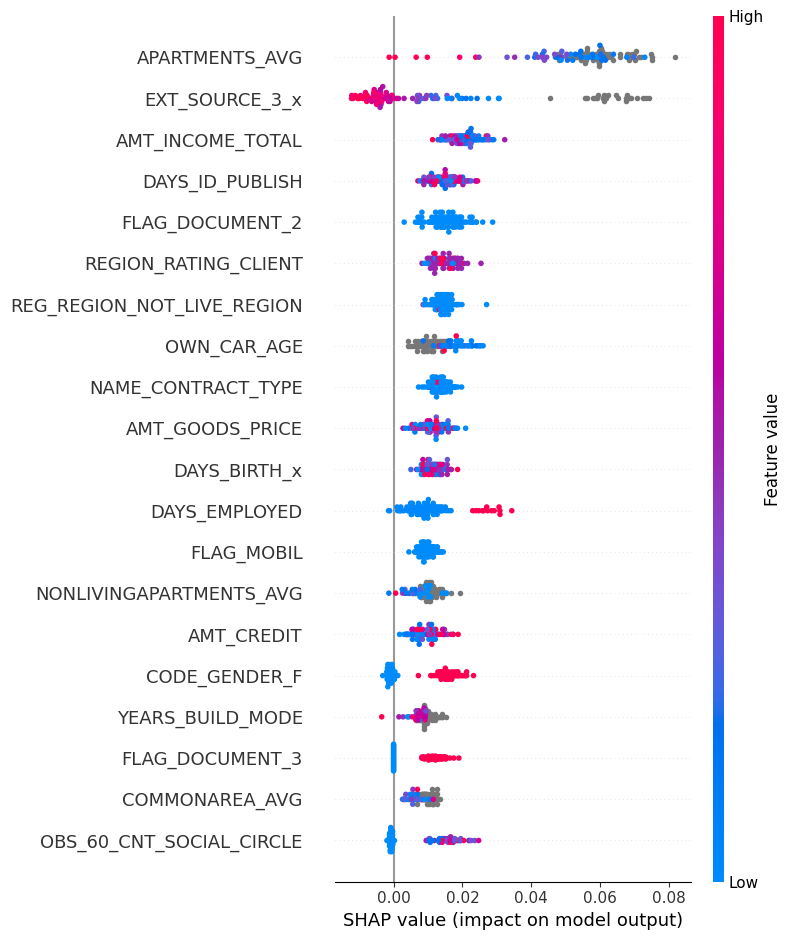


Figure 9: Shap values for poly-feature random forest model



We find that the most important features for each model is very different. The logistic regression model does not capture complex relationships in features such as EXT\_SOURCE as well as the random forest models. Gender is the top feature of the logistic regression model, which is a sensitive predictor that should be avoided. On the other hand, the poly-feature “APARTMENT\_AVG” has the largest impact on the poly-feature random forest model. It's interesting to note that neither of the binary features, FLAG\_OWN\_CAR or FLAG\_OWN\_PROPERTY, appear in the list of top SHAP values for the latest model. Instead, features like APARTMENT\_AVG and OWN\_CAR\_AGE are present, which might be due to these features already encapsulating the information represented by the aforementioned binary features. Overall, we find that the poly-feature random forest model is more robust because the SHAP values is more spread-out instead of being consistent across diverse instances. This shows the model is not overly reliant on any feature. We also find the two random forest models to be more stable than the logistic regression model. As a complex dataset with more than 200 features, the random forest models is better at capturing complex relationships.

Summary

1. The information utilized for the ADS was suitable for determining credit default risk. An extensive amount of data, including details on borrowers' past credit histories, demographic data, and other financial data, has been made available by the competition's organizers. Additionally, the data has been preprocessed and cleaned to remove outliers and missing values, which ought significantly increase the model's accuracy.
2. Since predicting loan repayment ability is the ADS's main goal, the accuracy metric is a crucial system evaluation statistic. It gauges how well the ADS can pinpoint those who are more likely to fall behind on their loan payments. We can compare the accuracy rates of the model based on various criteria in order to examine the accuracy of the ADS across diverse subpopulations. We can use a variety of fairness indicators, such as the demographic parity difference and demographic parity ratio, to assess how fair the model is. Demographic parity ratio represents the ratio of positive outcomes between subgroups, whereas demographic parity difference evaluates the variation in the proportion of positive outcomes between various groupings. Accurately identifying people who are more likely to fail on their loans depends on the ADS's accuracy. However, robustness and fairness are not always implied by accuracy alone. Even if a logistic model is highly accurate, it may still be biased towards particular subpopulations, producing unfair results. A random forest model, on the other hand, is more reliable and accurate and does better when dealing with complex interactions between input data and the output variable. We can determine significant features and make sure the model is equitable across various subpopulations by applying SHAP or LIME approaches. The polyfeature technique can also be used to achieve fairness, robustness, and stability across several subpopulations. While a more reliable and fair ADS benefits the loaner by minimizing bias and obtaining higher stability, it may not necessarily be accurate for the bank. More accurate ADS may not always be robust and fair, but it benefits the bank by lowering the amount of defaulted loans. To guarantee the ADS serves both the loaner and the bank, it is crucial to strike the correct balance between accuracy, robustness, and fairness.
3. I wouldn't feel confident using this ADS in the public or private sectors based on the examination of the fairness indicators. Indicating probable discrimination or unfairness, the demographic parity difference and demographic parity ratio demonstrate that there are notable disparities in the acceptability rates between certain groups for specific attributes. Based on variables including gender, education, occupation, and property ownership, the ADS may unintentionally discriminate against specific subpopulations. However, we can reduce bias and guarantee that the model is fair across different subpopulations by using fairness metrics like demographic parity difference and demographic parity ratio. We can prevent the ADS from treating subpopulations that may be more susceptible to prejudice unjustly by using an intersectional approach to fairness. Unfair ADS deployment can have detrimental effects on people and society, such as maintaining existing imbalances and transgressing moral and legal norms. To avoid these undesirable effects, it is crucial to confirm that ADS is impartial and fair before deployment. Additionally, utilizing an unfair ADS might harm the standing of the company or institution using it, resulting in a decline in credibility and trust. If the impacted people or groups file a lawsuit against the organization, it may also result in legal and financial obligations.
4. To prevent potential biases and discrimination, sensitive features like gender can be excluded from the ADS. Loan approval rates have been demonstrated to be significantly influenced by gender in particular, and discrimination against specific groups may result. The ADS can be made more impartial and fair by eliminating gender as a characteristic. However, it is crucial to keep in mind that eliminating sensitive elements could also result in a decline in the model's precision and effectiveness because these features might contain crucial data for forecasting loan repayment. Therefore, while choosing which features to include in the ADS, it is crucial to carefully analyze the trade-off between fairness and accuracy. ADS biases and prejudice can be addressed while maintaining accuracy using additional techniques like data augmentation or synthetic data generation.