CS340 Assignment 3: Bias Mitigation

Overview

In this assignment, you need to build a prediction model to predict user loan status based on the provided data set. It should be noted that there is serious bias in the data set. If the model is directly trained using this, the prediction results may be significantly discriminatory. Therefore, you need to make your model's performance less biased.

About the dataset

Description

This is a loan prediction problem on Analytics Vidhya. There are two data sets. The training set gives some information about loan applicants and the results of their loan applications (allowed or rejected). The test set gives some information about loan applicants but does not have the results of their loan applications. It is necessary to train a classification model on this data and make predictions on the test set data.

We tweaked the dataset to delve deeper into possible bias issues in the model and focus on two sensitivity attributes: "gender" and "race". Among them, the attribute "gender" already exists in the original data set, while "race" is an additional attribute we introduced to analyze the bias issue more comprehensively. Through such adjustments, we expect to more accurately reveal possible bias issues in models when dealing with these sensitive attributes.

Data attributes

No.	Attributes	Description
0	Loan_ID	unique loan code
1	Gender	Male/ Female
2	Race	White/ Black
3	Married	Married or not (Y/N)
4	Dependents	Number of dependents
5	Education	Education level (Graduate/ UnderGraduate)
6	Self_Employed	Whether you are self-employed (Y/N)
7	ApplicantIncome	Applicant’s income
8	CoapplicantIncome	Co-applicant’s income
9	LoanAmount	loan amount
10	Loan_Amount_Term	Loan term (unit: months)
11	Credit_History	credit history
12	Property_Area	Urban/ Semi Urban/ Rural

Table 1 Description of dataset

Submission Guide

Please submit a zip file named <StudentID-Name-Assignment3.zip> to Blackboard, where the submitted zip includes:

- Executable code (including requirement file)

Please indicate the following in your code files：

- Prediction model file with bias exists (2 points)
- Bias test before mitigation, including test methods and test results (2 points)
- The model file with bias mitigated (2 points)
- Bias test after mitigation, including test methods and test results (2 points)

- **A written report**, which is a complete and coherent description on how you test and mitigate bias. (2 points)

Scoring criteria

- For the list mentioned in the submission guide, 2 points will be deducted for each missing item.
- The report has a total of 5 points, divided into three grades：

Level 1: 1-2points Level 2: 3points Level 3: 4-5points

Scoring criteria: completeness of the report, scientificity of evaluating and mitigating methods, and presentation of experimental results.

- The code part has a total of 5 points, divided into three grades:

Level 1: 1-2points Level 2: 3points Level 3: 4-5points

Scoring criteria: code completeness, executability, model performance, and code standardisation.

Total: 20points

How to get help

If you encounter any problems, please do not hesitate to reach out to the TA team by either sending posts on the Blackboard discussion board or making in-person appointments via QQ. We are here to help. Our ultimate goal is for all of you to acquire knowledge through proper training instead of overwhelming you :)

Reference

[1] Experiment Baseline: https://www.kaggle.com/code/sundarikonar/loan-prediction-gridsearchcv-feature-importance

Experiment Guide

Step1: Statistics and visualization of bias in data sets

It would help if you analyzed the dataset, understood its biases, and considered how these biases could impact your predictive model.

Example：

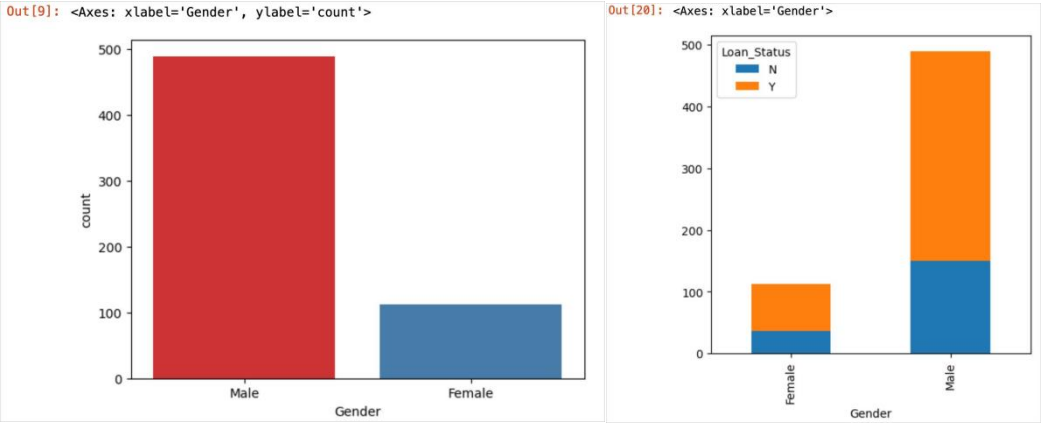


Figure 1 Gender feature in the data set(left), the relationship between gender and loans status(right)

Step2: Train a predictive model

Now, you need to train a loan prediction model. You can freely choose models such as logistic regression, decision tree, neural network, etc.
(Note: When performing feature engineering, the two attributes "gender" and "race" must be retained)

Step3: Evaluate model bias

Choose an evaluation metric from **Assignment 2** and perform a bias evaluation on your trained model. Introduce the bias situation through the evaluation results.

Step4: Mitigating model bias

You need to mitigate the bias phenomenon in **Step3**. You can use methods including but not limited to expanding data, using better models, using data enhancement, reinforcement learning, modify loss function, etc.
(Note: Thinking about and writing down the advantages and limitations of your chosen solution, which will help with subsequent projects)

Example:

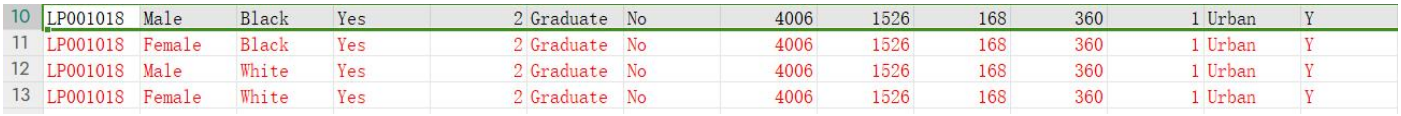
- expand dataset

You can expand data based on the original data set or find a more extensive, more comprehensive data set to supplement the model's training.
I.e., in the original data set, another piece of data with a different "gender" is generated for the same piece of data to balance the gender imbalance in the data set.

LP002086	Female	Black	Yes	0 Graduate	No	4333	2451	110	360	1 Urban	N
LP002505	Male	Black	Yes	0 Graduate	No	4333	2451	110	360	1 Urban	N
LP001603	Male	Black	Yes	0 Not Graduate	Yes	4344	736	87	360	1 Semiurban	N
LP001938	Male	Black	Yes	2 Graduate	No	4400	0	127	360	0 Semiurban	N
LP002236	Male	Black	Yes	2 Graduate	No	4566	0	100	360	1 Urban	N
LP001003	Male	Black	Yes	1 Graduate	No	4583	1508	128	360	1 Rural	N
LP001489	Female	Black	Yes	0 Graduate	No	4583	0	84	360	1 Rural	N
LP001577	Female	Black	Yes	0 Graduate	No	4583	0	112	360	1 Rural	N
LP002990	Female	Black	No	0 Graduate	Yes	4583	0	133	360	0 Semiurban	N
LP002367	Female	Black	No	1 Not Graduate	No	4606	0	81	360	1 Rural	N

161	LP002778	Male	White	Yes	2	Graduate	Yes	6633	0		360	0 Rural	N
162	LP001326	Male	White	No	0	Graduate		6782	0		360	Urban	N
163	LP002362	Male	White	Yes	1	Graduate	No	7250	1667	110		0 Urban	N
164	LP001370	Male	White	No	0	Not Graduate		7333	0	120	360	1 Rural	N
165	LP002301	Female	White	No	0	Graduate	Yes	7441	0	194	360	1 Rural	N
166	LP001043	Male	White	Yes	0	Not Graduate	No	7660	0	104	360	0 Urban	N
167	LP001562	Male	White	Yes	0	Graduate	No	7933	0	275	360	1 Urban	N
168	LP002315	Male	White	Yes	1	Graduate	No	8300	0	152	300	0 Semiurban	N
169	LP002473	Male	White	Yes	0	Graduate	No	8334	0	160	360	1 Semiurban	N
170	LP002067	Male	White	Yes	1	Graduate	Yes	8666	4983	376	360	0 Rural	N

Figure 2 Gender and race biases in the training set



10	LP001018	Male	Black	Yes	2	Graduate	No	4006	1526	168	360	1 Urban	Y
11	LP001018	Female	Black	Yes	2	Graduate	No	4006	1526	168	360	1 Urban	Y
12	LP001018	Male	White	Yes	2	Graduate	No	4006	1526	168	360	1 Urban	Y
13	LP001018	Female	White	Yes	2	Graduate	No	4006	1526	168	360	1 Urban	Y

Figure 3 Expansion of data sets in sensitive attribute dimensions

As shown in Figure 3, the 10th row is the data in the original data set. Data set expansion is used to expand different sensitive attribute values in the sensitive attribute dimension to weaken the impact of sensitive attributes on the inference results.

- select better algorithm

If the model you build using logistic regression shows a large bias, you could try algorithms such as random forests, deep learning, etc. Appropriate algorithms mitigate some of the bias problems.

Tips: If you have other methods, don't worry, feel free to write them out, you may get extra scores.

Step5: Bias evaluation of post-mitigation models

Reapply the evaluation metric chosen in step 3 to assess the bias of the model post-mitigate. Then, compare the evaluation outcomes before and after step 4.