Model Card - Credit Prediction

Model Details:

- Developed by Shashank Vankadari, a graduate student at Lehigh University Data Science program.
- This project designs various models and chooses the best binary model that can predict if a person will be able to repay the loan or not based on the historical data on which the model is trained.
- This is a revised version which is developed based on fairness assessments and value audits done on the previously developed model. It now ensures the data is cleaned properly as per various privacy and ethics principles, makes sure the model's fairness is assessed across different groups and thus gives suitable recommendations. Also, the model selection is done considering multiple metrics rather than considering one single metric.
- The Random Forest model is found out to be the best for predicting loan repayment.
- Developer can be contacted at shv222@lehigh.edu

Intended Use:

- Intended to be used by financial institutions to determine person's worthiness of receiving a loan.
- Intended for any person with proper identity (Native to United States or international resident with proper immigration status).
- Not intended to make judgments on any individual.

Metrics:

- The performance metrics such as accuracy, P-value, sensitivity, specificity, balanced accuracy was considered **relevant factors** in choosing the best performing model and Random Forest model is found to be the best and most effective. All metrics were evaluated from the confusion matrix obtained after testing the model. You can find the results of model evaluation under Quantitative analysis section.
- A P-value less than 0.05, sensitivity and specificity close to 1 and balanced accuracy close to 100% is considered.
- For evaluating the fairness of the Random Forest model, we considered Precision, Recall, False positives and made sure the fairness is improved on Race and Gender compared to earlier version.

Ethical considerations:

- The sensitive attributes like race, sex, gender is not considered while model building following privacy and ethics principles. This is done in accordance with Equal Credit Opportunity act which considers sex, race, or country as protected attributes.
- The dataset used has class bias on sensitive attributes and since they were not considered for model training, the model can be less biased.

Training data:

- The credit dataset contains occupation, personal and financial information necessary to predict loan repayment and thus making it more effective in building the models.
- The individual's native country (attribute) is removed from the data set before the data split, as we have improper distribution of records for some countries. Due to this, the test data may have some records which were not present in train data and vice versa. This creates a problem while evaluating the model.
- The historical credit dataset is split into 70% of training data and 30% of testing data as

this distribution ensures sufficient data points for model training and testing.

- The data split is done after necessary pre-processing (like converting attributes to suitable data types, ordering of categorical data, removal of certain unwanted attributes etc.)

Testing Data:

- 30% of the data in the data split was used to evaluate the performance of the models.

Quantitative analysis:

```
Female   Male
  9782  20380

Amer-Indian-Eskimo Asian-Pac-Islander              Black              Other              White
               286                895               2817                231              25933

                  Cambodia                     Canada                      China                   Columbia
                        18                        107                         68                         56
                      Cuba         Dominican-Republic                    Ecuador                El-Salvador
                        92                         67                         27                        100
                   England                     France                    Germany                     Greece
                        86                         27                        128                         29
                 Guatemala                      Haiti         Holand-Netherlands                   Honduras
                        63                         42                          1                         12
                      Hong                    Hungary                      India                       Iran
                        19                         13                        100                         42
                   Ireland                      Italy                    Jamaica                      Japan
                        24                         68                         80                         59
                      Laos                     Mexico                  Nicaragua Outlying-US(Guam-USVI-etc)
                        17                        610                         33                         14
                      Peru                Philippines                     Poland                   Portugal
                        30                        188                         56                         34
               Puerto-Rico                   Scotland                      South                     Taiwan
                       109                         11                         71                         42
                  Thailand              Trinadad&Tobago              United-States                    Vietnam
                        17                         18                      27504                         64
                Yugoslavia
                        16
```

Fig-1: Class bias on sensitive attributes – Sex, Race and Country

```
Confusion Matrix and Statistics

   predict_modell
       0    1
  0 5406  221
  1 1049  865

               Accuracy : 0.8316
                 95% CI : (0.8229, 0.84)
    No Information Rate : 0.856
    P-Value [Acc > NIR] : 1

                  Kappa : 0.4814

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.8375
            Specificity : 0.7965
         Pos Pred Value : 0.9607
         Neg Pred Value : 0.4519
             Prevalence : 0.8560
         Detection Rate : 0.7169
   Detection Prevalence : 0.7462
      Balanced Accuracy : 0.8170

       'Positive' Class : 0
```

```
Confusion Matrix and Statistics

   y_predn
       0    1
  0 5364  263
  1 1313  601

               Accuracy : 0.791
                 95% CI : (0.7817, 0.8001)
    No Information Rate : 0.8854
    P-Value [Acc > NIR] : 1

                  Kappa : 0.3263

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.8034
            Specificity : 0.6956
         Pos Pred Value : 0.9533
         Neg Pred Value : 0.3140
             Prevalence : 0.8854
         Detection Rate : 0.7113
   Detection Prevalence : 0.7462
      Balanced Accuracy : 0.7495

       'Positive' Class : 0
```

```
Confusion Matrix and Statistics

   y_predictr
       0    1
  0 5282  345
  1  725 1189

               Accuracy : 0.8581
                 95% CI : (0.85, 0.8659)
    No Information Rate : 0.7966
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.5991

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8793
            Specificity : 0.7751
         Pos Pred Value : 0.9387
         Neg Pred Value : 0.6212
             Prevalence : 0.7966
         Detection Rate : 0.7004
   Detection Prevalence : 0.7462
      Balanced Accuracy : 0.8272

       'Positive' Class : 0
```

Fig-2: L-R Model **Fig-3: Naïve Bayes model** **Fig-4: Random Forest**

Description: df [2 × 5]

Gender <fctr>	Nums <int>	Precision <dbl>	Recall <dbl>	False_Positives <dbl>
Male	5079	0.7751303	0.8428801	0.06321959
Female	2462	0.7748691	0.9392338	0.01777594

2 rows

Fig-5: Fairness assessment based on gender

Description: df [5 × 5]

Race <fctr>	Nums <int>	Precision <dbl>	Recall <dbl>	False_Positives <dbl>
White	6478	0.7740768	0.8701197	0.05060006
Black	709	0.7951807	0.9408946	0.02456647
Asian-Pac-Islander	226	0.7833333	0.9156627	0.06103286
Other	50	0.6666667	0.9090909	0.04166667
Amer-Indian-Eskimo	78	0.7500000	0.8918919	0.01298701

5 rows

Fig-6: Fairness assessment based on race

Caveats and Recommendations:

- Since there is bias in the data, it cannot be possible to achieve complete fairness. Bias can be reduced at data collection stage by adjusting the sampling process. Once bias is removed, fairness can be improved as per legal/ethical principles of the place.
- In the fairness metrics, the false positives vary across various groups, so there is risk of improperly predicting the outcome based on race / sex. This can raise concerns about fairness of the model.
- We can collect data points in such a way that class bias is low or use advanced techniques such as up sampling / down sampling to reduce bias and improve fairness.