Credit Line Increase Model Card

Basic Information

- **Person or organization developing model**: Pawan Kumar, "pawanmarine81@gwu.edu"
- **Model date**: October, 2022
- **Model version**: 1.1
- **License**: MIT
- **Model implementation code**: DNSC 6301 Project.ipynb

Intended Use

- **Primary intended uses**: This model is an *example* probability of default classifier, with an *example* use case for determining eligibility for a credit line increase.
- **Primary intended users**: : Mostly by the Graduate student, especially the data and Business analysts' students
- **Out-of-scope use cases**: Any use beyond an educational example is out-of-scope.

Training Data -Data dictionary:

Name	Modeling Role	Measurement Level	Description
ID	ID	Int	unique row indentifier
LIMIT_BAL	input	float	amount of previously awarded credit

Name	Modeling Role	Measurement Level	Description
SEX	demographic information	Int	1 = male; 2 = female
RACE	demographic information	Int	1 = hispanic; 2 = black; 3 = white; 4 = asian
EDUCATION	demographic information	Int	1 = graduate school; 2 = university; 3 = high school; 4 = others
MARRIAGE	demographic information	Int	1 = married; 2 = single; 3 = others
AGE	demographic information	Int	age in years
PAY_0, PAY_2 - PAY_6	inputs	Int	history of past payment; PAY_0 = the repayment status in September, 2005; PAY_2 = the repayment status in August, 2005; ...; PAY_6 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment

Name	Modeling Role	Measurement Level	Description
			delay for two months; ...; 8 = payment delay for eight months; 9 = payment delay for nine months and above
BILL_AMT1 - BILL_AMT6	inputs	float	amount of bill statement; BILL_AMNT1 = amount of bill statement in September, 2005; BILL_AMT2 = amount of bill statement in August, 2005; ...; BILL_AMT6 = amount of bill statement in April, 2005
PAY_AMT1 - PAY_AMT6	inputs	float	amount of previous payment; PAY_AMT1 = amount paid in September, 2005; PAY_AMT2 = amount paid in August, 2005; ...; PAY_AMT6 = amount paid in April, 2005
DELINQ_NEXT	target	Int	whether a customer's next payment is delinquent (late), 1 =

Name	Modeling Role	Measurement Level	Description
			late; 0 = on-time

- **Source of training data**: GWU Blackboard, email `jphall@gwu.edu` for more information
- **How training data was divided into training, validation and test data**: 50% training, 25% validation, 25% test
- **Number of rows in training and validation data**:
  - Training rows: 15,000
  - Validation rows: 7,500

Test Data

- **Source of test data**: GWU Blackboard, email `jphall@gwu.edu` for more information
- **Number of rows in test data**: 7500
- **State any differences in columns between training and test data**: None

Model details

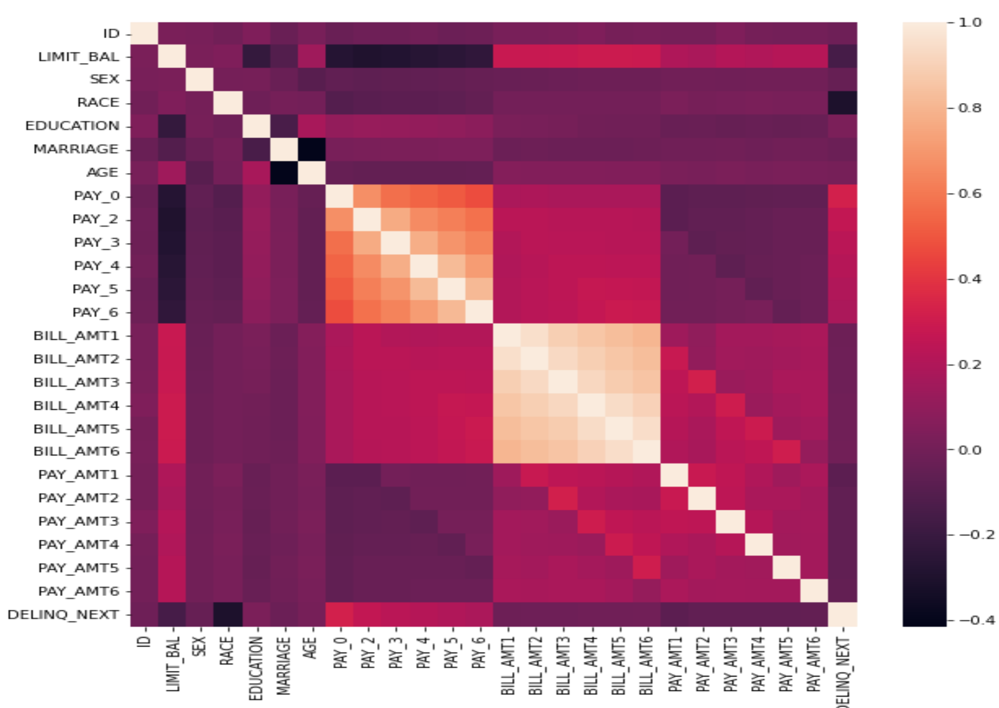
- **Columns used as inputs in the final model**: 'LIMIT_BAL', 'PAY_0', 'PAY_2', 'PAY_3', 'PAY_4', 'PAY_5', 'PAY_6', 'BILL_AMT1', 'BILL_AMT2', 'BILL_AMT3', 'BILL_AMT4', 'BILL_AMT5', 'BILL_AMT6', 'PAY_AMT1', 'PAY_AMT2', 'PAY_AMT3', 'PAY_AMT4', 'PAY_AMT5', 'PAY_AMT6'
- **Column(s) used as target(s) in the final model**: 'DELINQ_NEXT'
- **Type of model**: Decision Tree
- **Software used to implement the model**: Python, scikit-learn
- **Version of the modeling software**: 3.10.7, 1.1.2
- **Hyperparameters or other settings of your model**:

```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',
                       max_depth=6, max_features=None,
max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort='deprecated',
                       random_state=12345, splitter='best')
```

Quantitative Analysis and Model Work Flow

1. Load and analyze date

The create line increase was loaded in google colab. Basic data analysis was performed to identify the shape of data, get column names, find missing values, and generate descriptive statistics. The correlation matrix was calculated to find pairwise correlation of the columns in the data. All columns are visually represented as histograms. A correlation heatmap figure was generated to represent the correlation matrix.



2. Train a decision tree model

The data is partitioned into training, validation, and test sets (50%, 25%, 25% respectively) to accurately evaluate the model. Testing data which is used to evaluate the trained model to test how the model will perform on real data. 12 different models are trained using decision trees and calculated the ROC AUC for each model.

AUC ROC

An ROC curve (Receiver Operating Characteristic curve) is a graph that shows performance of a classification model that is captured at all classification thresholds.

True Positive Rate (TPR): It is called recall and it is measured as:

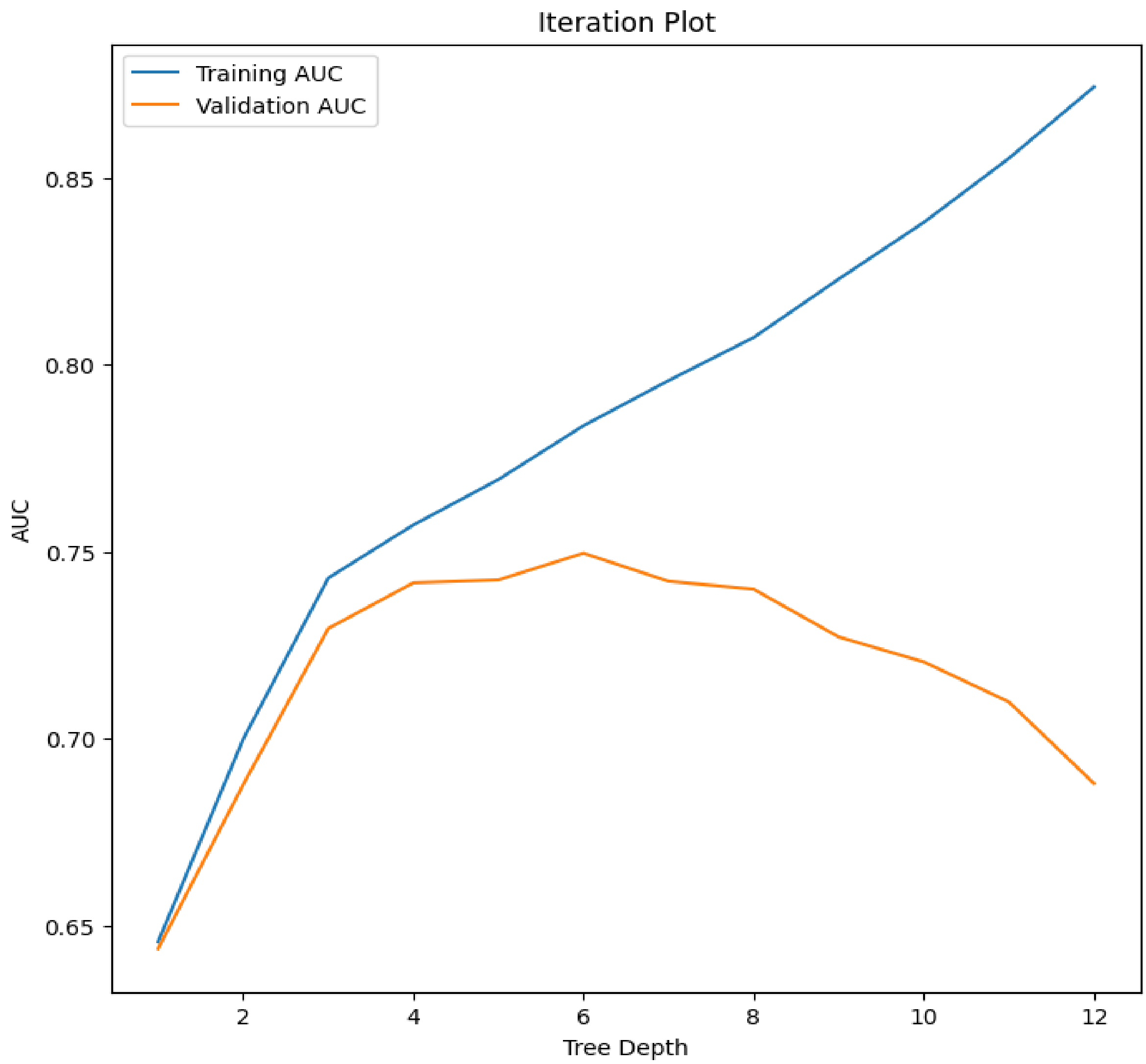
TPR = TP/(TP+FN)

False Positive Rate (FPR):

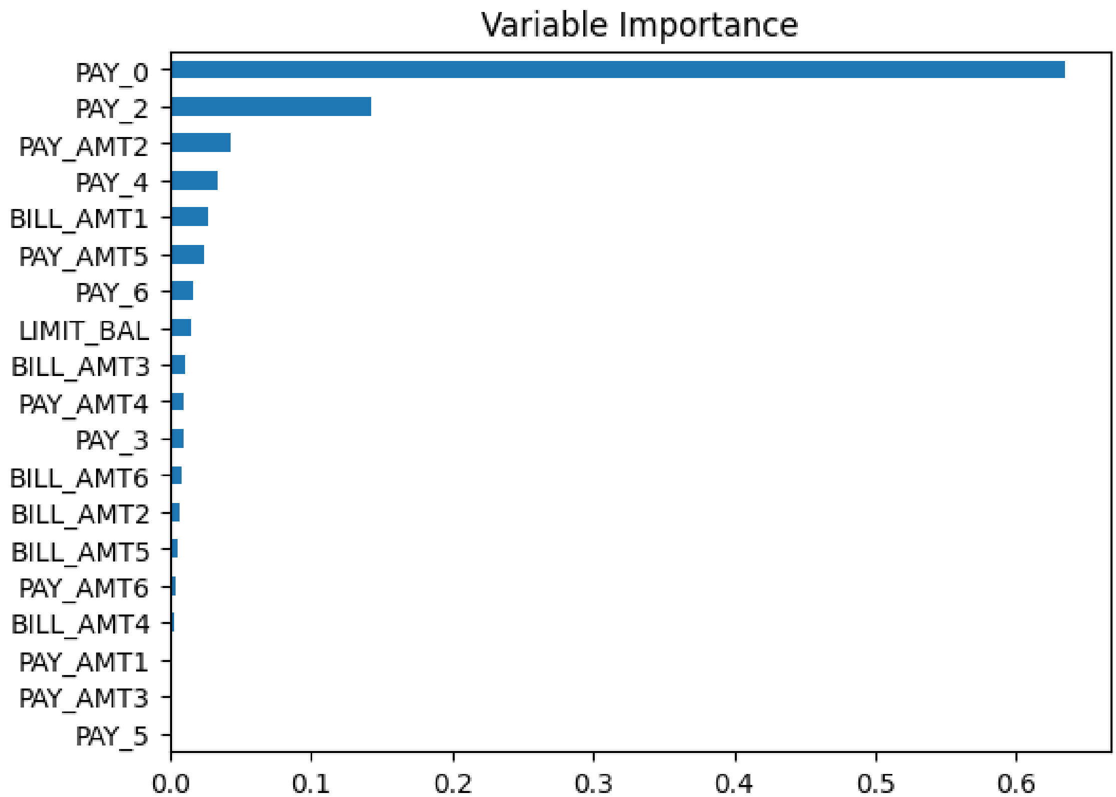
FPR = FP/(FP+TN)

ROC curve plots FPR on the x-axis and TPR on the y-axis at different classification thresholds. If threshold level is reduced, more items will be considered as positive. Thus, both False Positives and True Positives will increase.

The AUC curve is Area Under the ROC curve. The higher the AUC, the better the model can predict the target variable. The below graph shows that the maximum validation AUC is at depth 6.



The below plot is the plot of variable importance. It provides the variable importance that provides a list of the most significant variables in descending order. Pay_0 contribute most to the model and also have high predictive power in classifying the target variable.



Lastly, the test AUC is calculated:

DATA TYPE	**AUC**
Training Data	0.7837
Validation Data	0.7496
Test Data	0.7438

3. Test the model for discrimination

It is very important to consider ethical measures such as fairness and security.

ADVERSE IMPACT RATIO (AIR):

Adverse impact is the negative effect an unfair and biased selection procedure has on a protected class. It occurs when a protected group is discrimination against during a selection process, like a hiring and promotion decision.

The fourth-fifths rule state that if the selection class for a certain group is less than 80% of that of the group with the highest selection rate, there is adverse impact on that group.

RACIAL BIAS:

The protected groups for racial bias testing are Hispanic, Black, and Asian. The reference group in this dataset is White. From the confusion matrices and AIR calculation, the Hispanic-to-White AIR is below the benchmark of 0.8. On the other side, the Black-to-white impact ratio is greater than 0.80. In this case, additional supporting evidence since the ratio is between 0.8 and 0.9(according to Office of Federal Contract Compliance Programs). The white-to-Asian AIR is 1.0 and thus favorable.

AIR COMPUTATION:

White proportion accepted: 0.568

Hispanic proportion accepted: 0.434

Hispanic to white AIR: 0.76

Overall	Actual: 1	Actual: 0
Predicted: 1	623	1200
Predicted: 0	211	1729

WHITE: Reference Group

Overall	Actual: 1	Actual: 0
Predicted: 1	176	813

HISPANIC – Protected Group

Overall	Actual: 1	Actual: 0
Predicted: 1	447	387

Predicted: 0	72	1228

Predicted: 0	139	501

AIR COMPUTATION:

White proportion accepted: 0.568

Black proportion accepted: 0.465

Black-to-white AIR: 0.82

Overall	Actual: 1	Actual: 0
Predicted: 1	625	1161
Predicted: 0	229	1765

WHITE: Reference Group

BLACK – Protected Group

Overall	Actual: 1	Actual: 0
Predicted: 1	176	813
Predicted: 0	72	1228

Overall	Actual: 1	Actual: 0
Predicted: 1	449	348
Predicted: 0	157	537

AIR COMPUTATION:

White proportion accepted: 0.568

Hispanic proportion accepted: 0.568

Hispanic to white AIR: 1.00

Overall	Actual: 1	Actual: 0
Predicted: 1	999	1597
Predicted: 0	61	2445

WHITE: Reference Group

ASIAN – Protected Group

Overall	Actual: 1	Actual: 0
Predicted: 1	176	813
Predicted: 0	72	1228

Overall	Actual: 1	Actual: 0
Predicted: 1	186	784
Predicted: 0	59	1217

GENDER BIAS:

The protected group for gender bias testing is female, and the reference group is male. The AIR value is favorable for women because it exceeds the best scenario by 0.06. This indicates that a higher number of females are awarded a loan as compared to males.

AIR COMPUTATION:

Male proportion accepted: 0.503

Female proportion accepted: 0.533

Female-to-white AIR: 1.06

Overall	Actual: 1	Actual: 0
Predicted: 1	1258	2332
Predicted: 0	427	3483

MALE: Reference Group

Overall	Actual: 1	Actual: 0
Predicted: 1	546	905
Predicted: 0	179	1292

FEMALE – Protected Group

Overall	Actual: 1	Actual: 0
Predicted: 1	712	1427
Predicted: 0	248	2191

4. REMEDIATE DISCOVERED DISCRIMINATION

It is a common scenario that Black and Hispanic groups face difficulty in getting approval for home loans as compared to White and Asian people. In 2015, 27.4% of the Black applicants and 19.2% of Hispanic applicants were denied mortgages, compared with about 11% of White and Asian applicants which can be also observed in our initial model. The biased behavior of ML models has adverse effects on society. With our initial probability cutoff of 0.15, the Hispanic-to-White AIR fell below the minimum acceptable value of 0.80 and the Black-to-White AIR was just over 0.80 by 0.02.

Notice that the cutoff may influence the result. Changing the cutoff of 0.18 rather than 0.15, we tried to remediate biases by recalculating AIR and confusion matrices. Next, we again run the model search by training decision trees with validation-based early stopping. Instead of picking the best model defined by AUC, we worked on 12 different models and observed the trade-off between performance and fairness indicators. The model balanced between two factors was chosen. The following table shows that the AIR value of

Hispanic-to-White and Black-to-White was impacted positively after applying 0.18 of the cutoff rate.

White proportion accepted	0.735
Hispanic proportion accepted	0.613
Hispanic to White AIR	0.83

White proportion accepted	0.735
Black proportion accepted	0.626
Black to White AIR	0.85

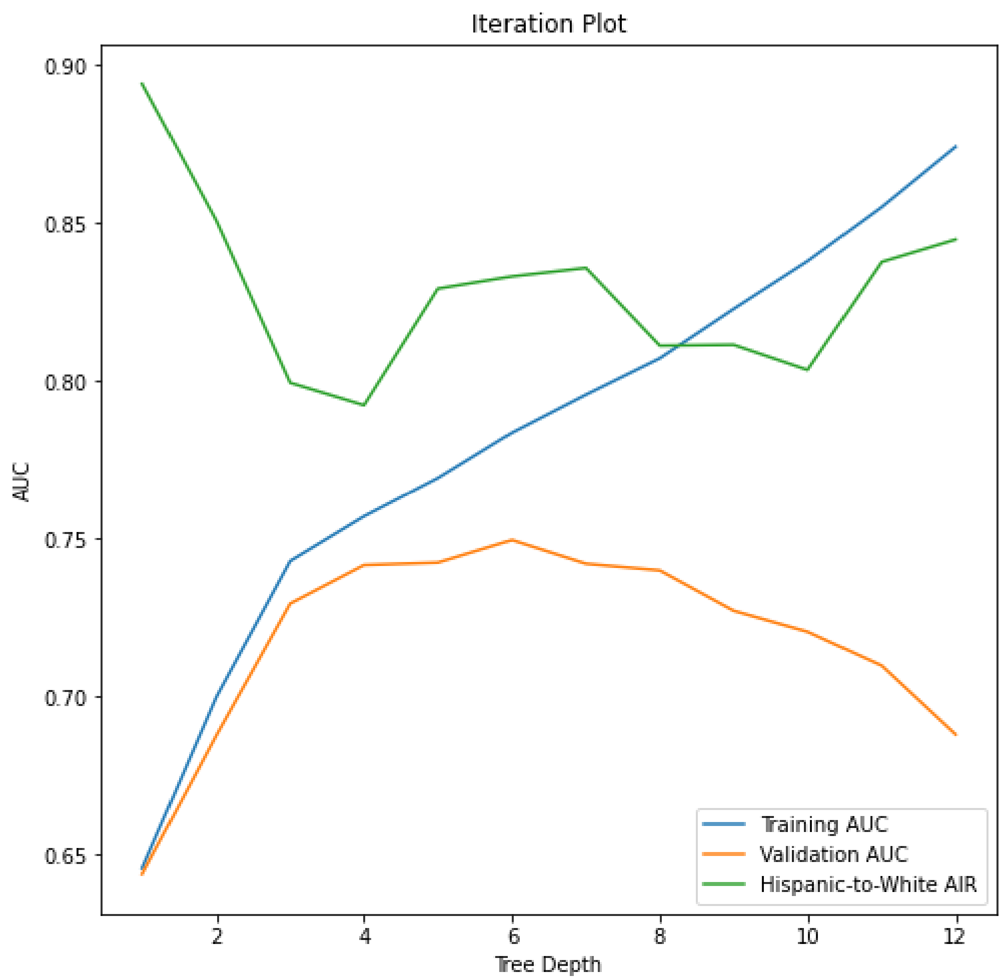
White proportion accepted	0.735
Asian proportion accepted	0.739
Asian to White AIR	1.00

The below table indicates the final value of the metrics for all data: training, validation, and test data.

Data Type	AUC
Training Data	0.7837
Validation Data	0.7496
Test Data	0.7438

Race	AIR (0.15 cutoff)	AIR (0.18 cutoff)
White to Hispanic	0.76	0.83
White to Black	0.82	0.85
White to Asian	1.00	1.00

Gender	AIR (0.15 cutoff)	AIR (0.18 cutoff)
Male to Female	1.06	1.02



Ethical considerations:

1. Negative impacts of using the model:

Math/Software Problems: There is always a high risk of attaining inconsistent result if results are based on recent payment trend while the long-term trends of are overlooked. One of the software problems in this model also the variable is carrying significant importance too.

Real world risks: Because of inconsistent result, it is likely that the customers' credit limit may remain plateau, if not dipped, i.e., increase in credit limits of customer's' are almost unlikely. Since, the model is little biased, applying it to real world could have severe negative impact.

2. Potential uncertainties relating to the impacts of using your model:

Math/Software Problems: A classic example of software problem is that it is quite possible because of one or few low metrics among several high scoring metrics may cause an unjustified result, eventually causing low credit limit for customers.

Real world risks: If the output of model will be implemented by the users, then the customers will not be granted additional credit what they deserve. In bigger prospect it will not only hamper the individual but also will adversely impact the economy, a negative aftermath of bias model.

3. Other unexpected results:

There is a very high probability that real-world data will have missing values. We have seen that there is no missing value in our data. It is evident from the variable importance chart that PAY_0 variable has the most critical role in the model. Thus, the model is dependent on the most recent payment instead of the consistent payment history. Still, the above-unexpected results do not disqualify the model, however, it is important to be aware of them.

SOURCES:

1. Desilver, D., & Bialik, K. (2017, January 10). Blacks and Hispanics face extra challenges in getting home loans. Pew Research Center. https://www.pewresearch.org/fact-tank/2017/01/10/blacks-and-hispanics-face-extra-challenges-in-getting-home-loans/

2. Hall, P. (n.d.). Increase Fairness in Your Machine Learning Project with Disparate Impact Analysis using Python and H2O.

Jupyter. https://nbviewer.org/github/jphall663/interpretable_machine_learning_with_python/blob/master/dia.ipynb

3. U.S. Department of Labor. (n.d.). Practical Significance in EEO Analysis Frequently Asked Questions. https://www.dol.gov/agencies/ofccp/faqs/practical-significance