A Fair and Interpretable Model for Credit Assessment

Mohammad Toutiaee

Credit Risk Assessment

Borrower

Credit Risk

Loan Repayment

- Affect borrower's financial history
- Loss of investor's capital and interest
- Damaged reputation as a responsible lender

Modeling Process

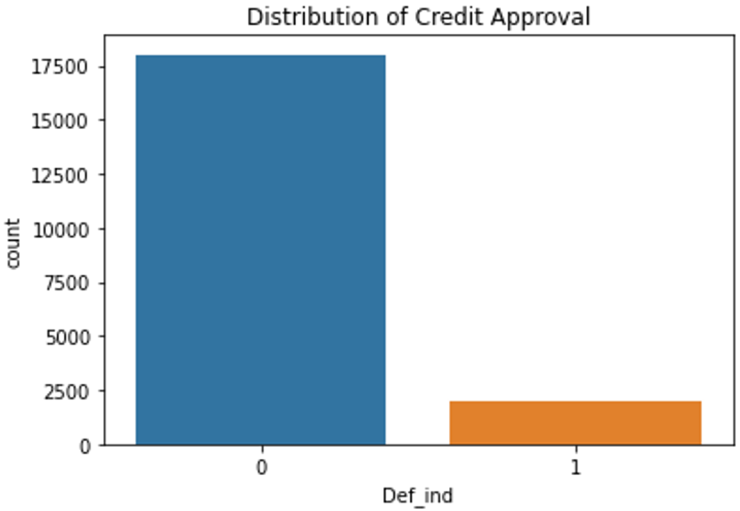
- Collecting Dataset
- Target Definition
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Modeling Process

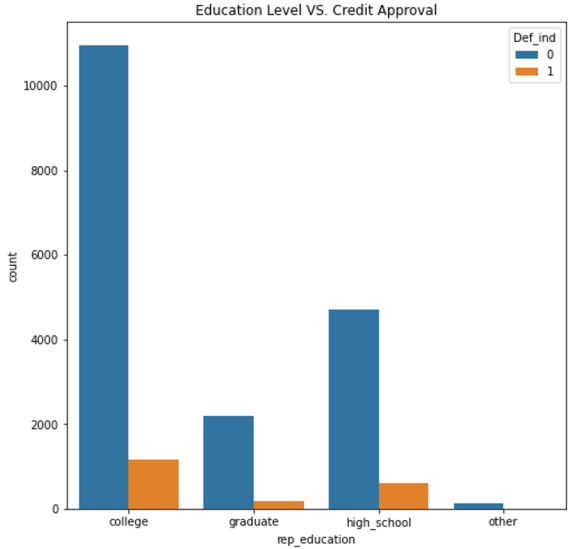
- **Collecting Dataset**
- Target Definition
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Data

Training: 20K, Test: 5K, 20 Predictors, 1 Binary Response



Distribution of Credit Approval

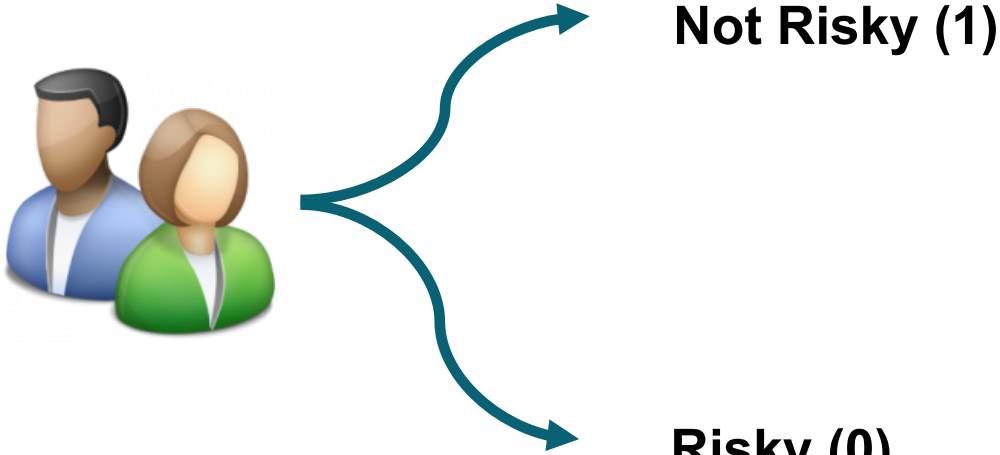


Number of Accounts per Education Level

Modeling Process

- Collecting Dataset
- **Target Definition**
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Target Definition



Modeling Process

- Collecting Dataset
- Target Definition
- **Data Pre-processing:**
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Data Pre-processing (Categorical)

Missing Values

- Impute by Sample of Data
- Impute by Mode
- Add a Feature for Nulls

Rare Values (Outliers)

- Replace with Sample of Data
- Replace with Mode
- Add a Feature for Outliers

Convert to Numbers

- Replace by Frequency
- Assign Order for Ordinal
- One-hot Encoding
- Target Encoding

Data Pre-processing (Categorical)

Missing Values

- Impute by Sample of Data
- Impute by Mode
- Add a Feature for Nulls

“Level of Education"

Mode: ”College”

Rare Values

- Replace with Sample of Data
- Replace with Mode
- Add a Feature for Outliers

Convert to Numbers

- Replace by Frequency
- Assign Order for Ordinal
- One-hot Encoding
- Target Encoding

“Level of Education”

Other =	0
High School =	1
College =	2
Graduate =	3

Data Pre-processing (Numerical)

Missing Values

- Impute by Sample of Data
- Impute by Mean or Median
- Impute by Extreme Values in the Distribution

Outliers (Only for Linear Models)

- Replace with Sample of Data
- Replace with Mean or Median
- Replace with Extreme Values in the Distribution

Data Pre-processing (Numerical)

Missing Values

- Impute by Sample of Data
- Impute by Mean or Median
- Impute by Extreme Values in the Distribution

Outliers (Only for Linear Models)

- Replace with Sample of Data
- Replace with Mean or Median
- Replace with Extreme Values in the Distribution

Data Pre-processing (Transformation)

Normalization (Standardization) [-1,1],[0,1], etc.

- Linear Models
- Neural Networks

Linearization (Linear Models)

- Transform by log, sqrt, etc.
- Rebinning (Replace by the output of Shallow Tree)

Data Pre-processing (Transformation)

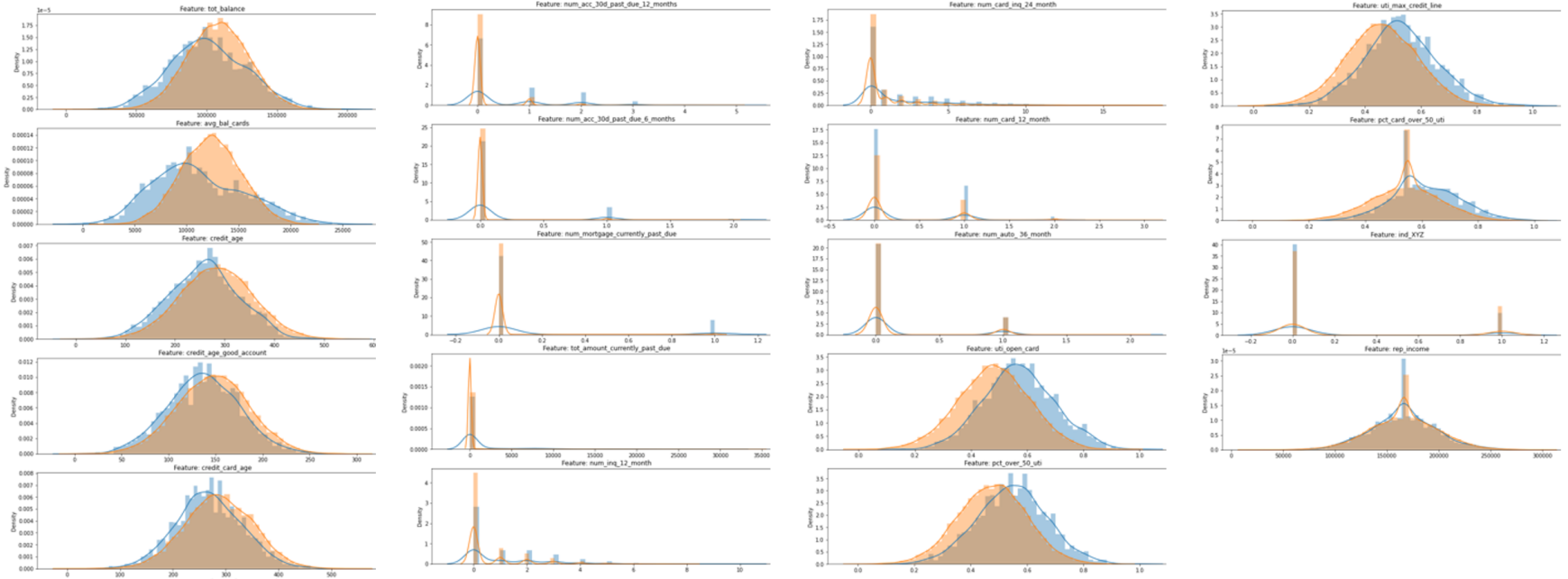
Normalization (Standardization) [-1,1],[0,1], etc.

- Linear Models [0,1]
- Neural Networks

Linearization (Linear Models)

- Transform by log, sqrt, etc.
- Rebinning (Replace by the output of Shallow Tree)

Feature Selection

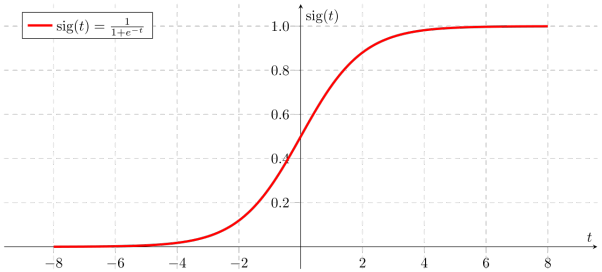


Modeling Process

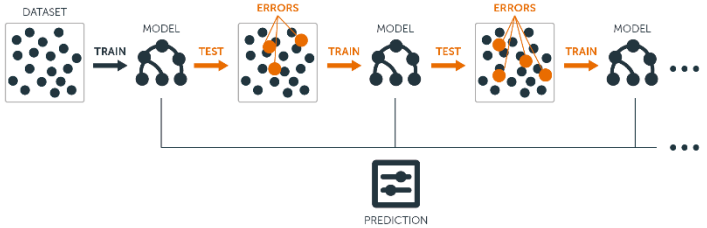
- Collecting Dataset
- Target Definition
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- **Model Building and Optimization**
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Model Building

Lasso Logistic Regression

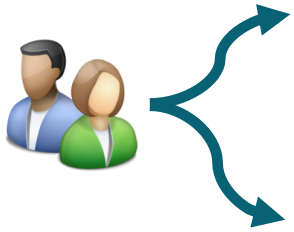


Xtreme Gradient Boosted Trees (XGBoost)



The Models Learn to Assign:

- **High Probability on 1**
- **Low Probability on 0**

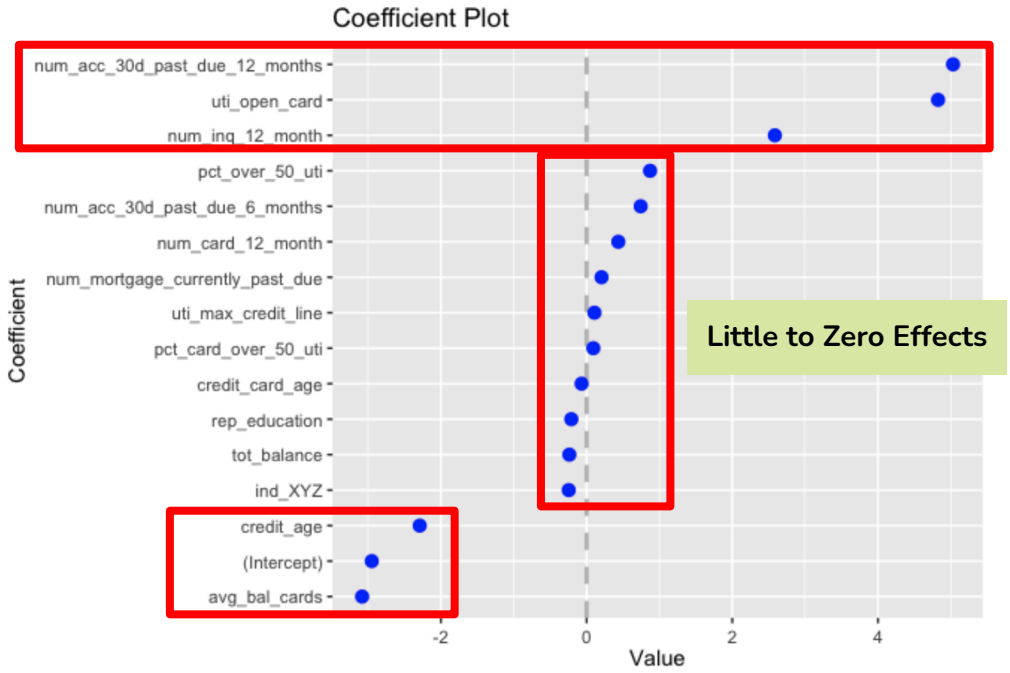


Not Risky (1)

Risky (0)

Results

Lasso Logistic Regression Coefficients



Little to Zero Effects

XGBoost achieved a significant better prediction due to the fact that XGBoost benefits from a complex training.

XGBoost

Actual	Predicted: Yes	Predicted: No
Yes	267	236
No	193	4307

Logistic Regression

Actual	Predicted: Yes	Predicted: No
Yes	96	404
No	60	4440

Model	AUC	Recall
L1 Logistic Regression	0.58	0.18
XGBoost	0.87	0.52

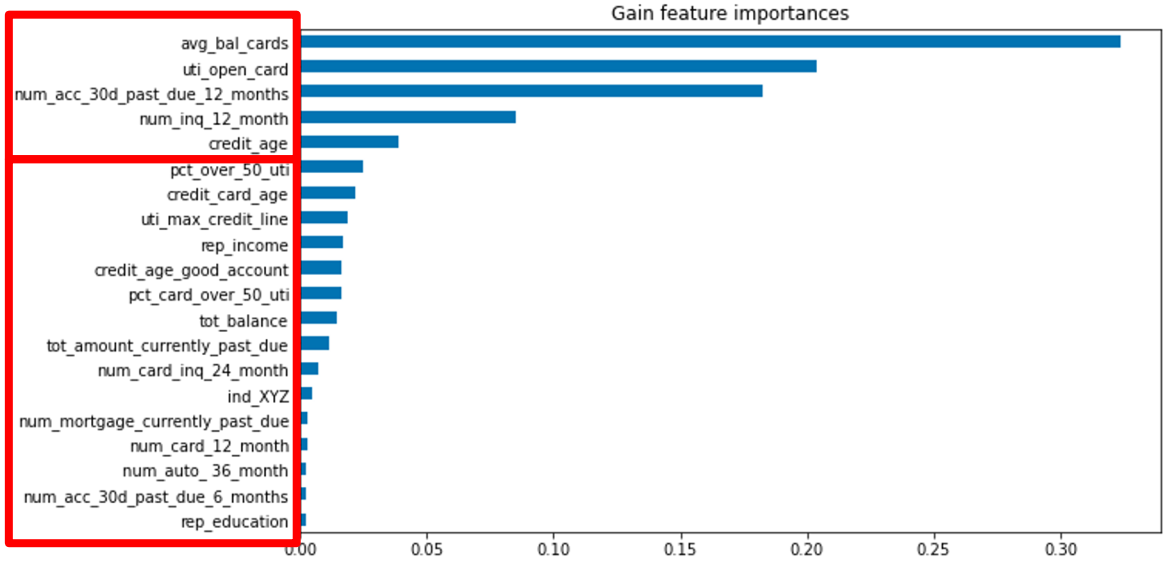
YES: "Def_ind" = 1

NO: "Def_ind" = 0

Modeling Process

- Collecting Dataset
- Target Definition
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- **Model Interpretation**
  - What-If Scenarios
  - Fairness Investigation
- Model Improvement
- Conclusion

Machine Learning Global Interpretation



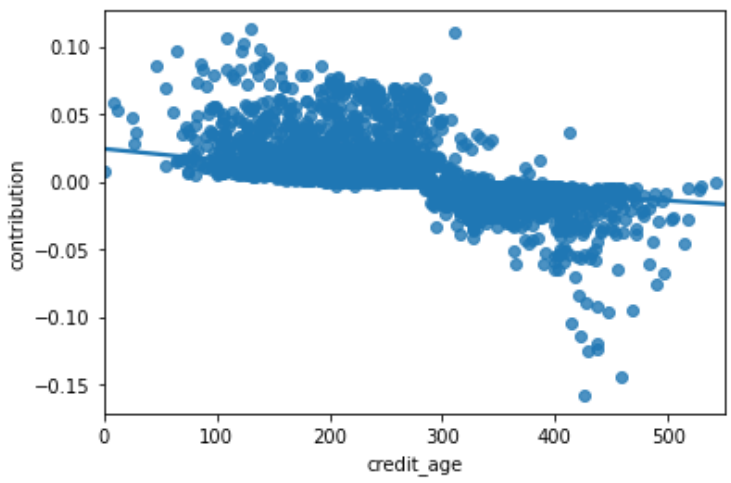
Feature Importance for XGBoost Model

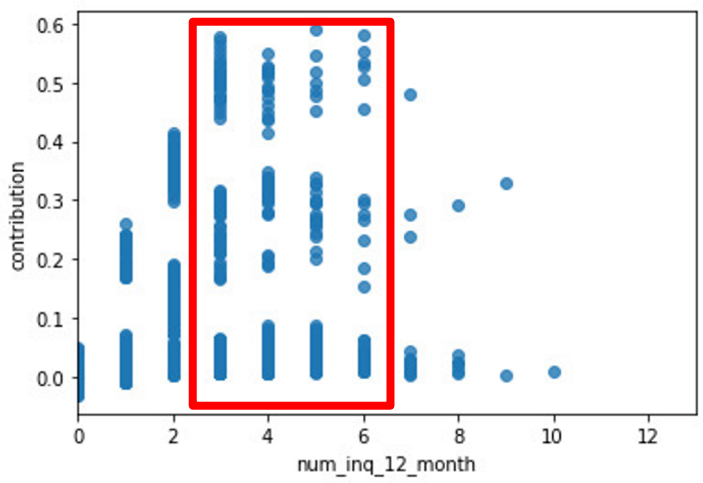
Gain-based Feature Importance - **XGBoost**

- **average balance for all credit cards**, most important feature for the XGB model
- The higher the number of average balance for all credit cards, the higher the probability of the applicant being approved.
- This same relationship could also be observed with
  - **“uti_open_card”**
  - **“num_acc_30d_past_due_12months”**
  - **“num_inq_12_month”**
- **education level** is the least important factor in credit application approval

Directional Feature Contributions (DFCs)

- Contribution of the credit age is almost constant across all the age groups
- If the number of inquiries in last 12 months were between 3 and 6, they would contribute the most in the prediction of the model.
- The clustered impacts among the values suggest that other factors could possibly add influence to the impact, which can be found in a further investigate





Machine Learning Local Interpretation

Shapeley Explanations

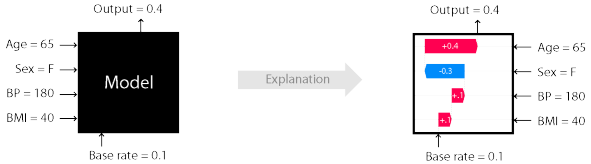
A unified approach to interpreting model predictions

S Lundberg, SI Lee - arXiv preprint arXiv:1705.07874, 2017 - arxiv.org

Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy in many applications. However, the highest accuracy for large modern datasets is often achieved by complex models that even experts struggle to interpret, such as ensemble or deep learning models, creating a tension between accuracy and interpretability. In response, various methods have recently been proposed to help users interpret the predictions of complex models, but it is often unclear how these methods are related and …

☆ Cited by 2430 Related articles All 10 versions

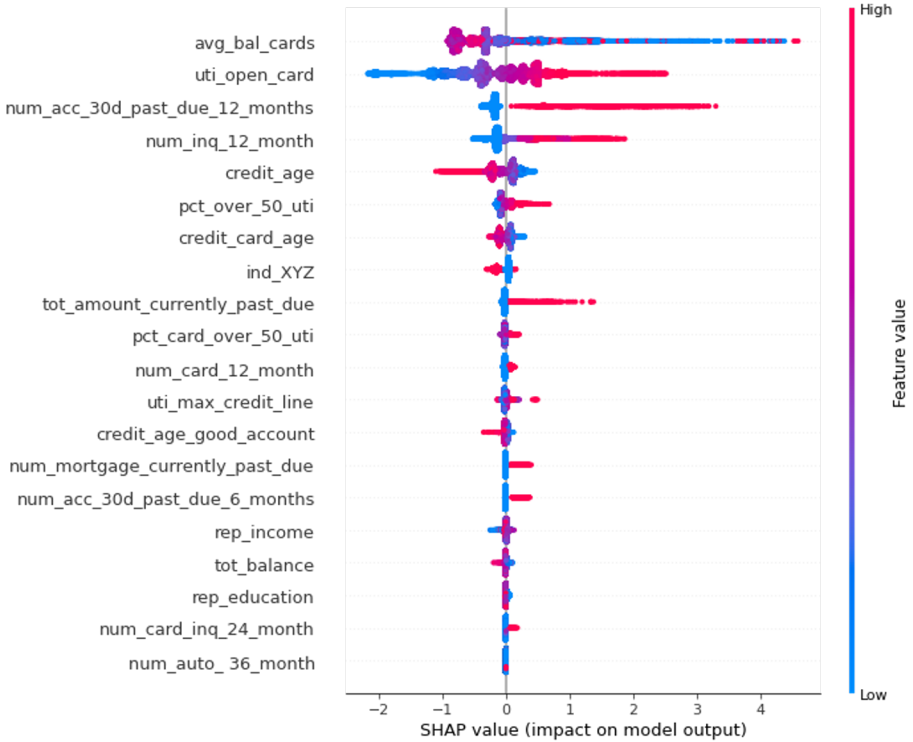
SHAP



$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left[f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right]$$

- Enables game theory to connect to local explanations, where each feature value of the instance is a "player"
  - *S* is a subset of the features used in the model, *i* is the vector of feature values of the instance to be explained and *F* is the set of all features.

SHAP; Machine Learning Interpretation

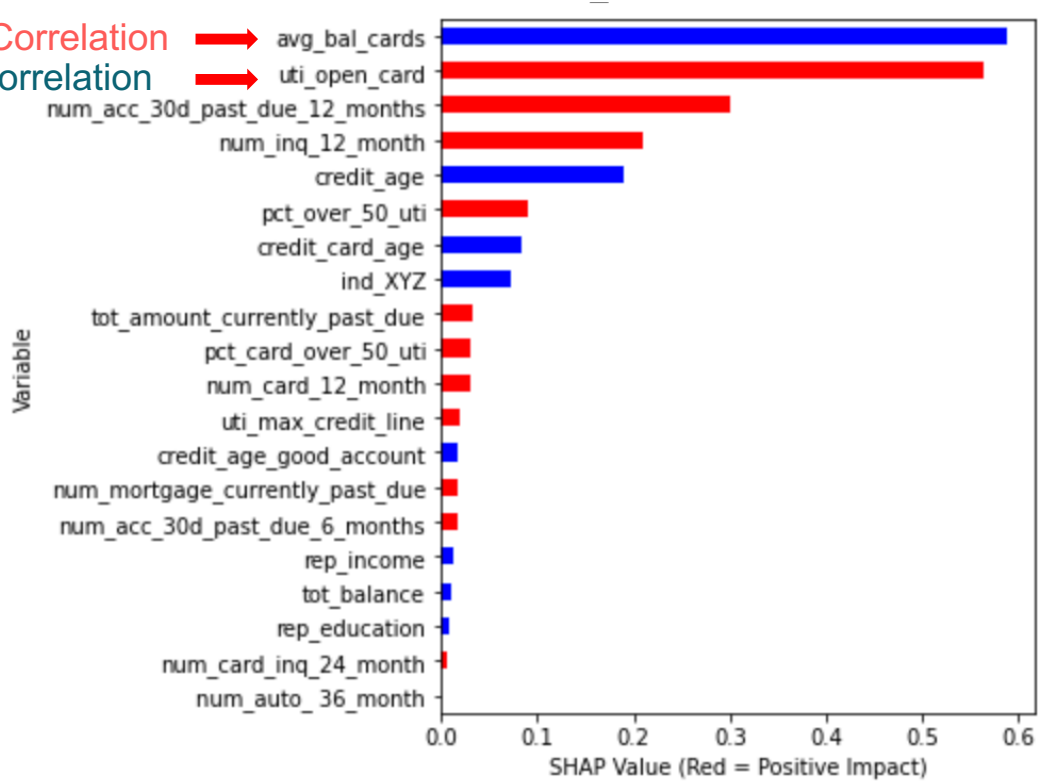


SHAP; Global Interpretation:

- Feature Importance
- Impact
- Original Value
- Correlation

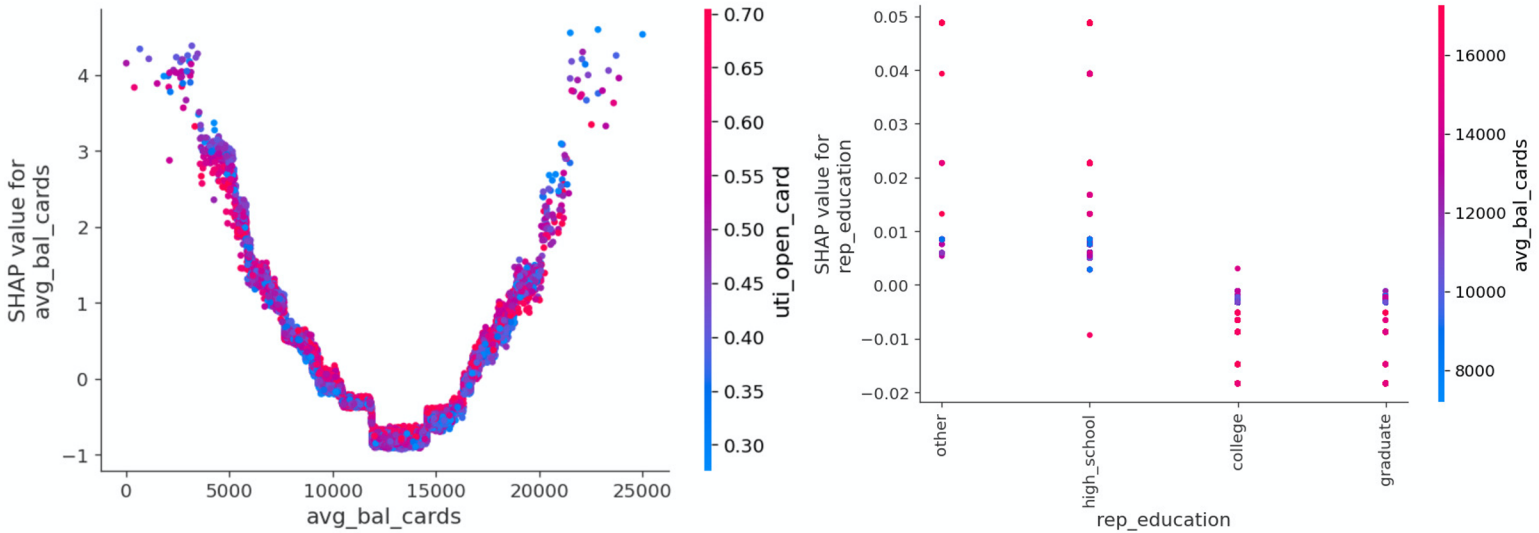
SHAP; Machine Learning Interpretation

Negative Correlation →
Positive Correlation →



Machine Learning Local Interpretation

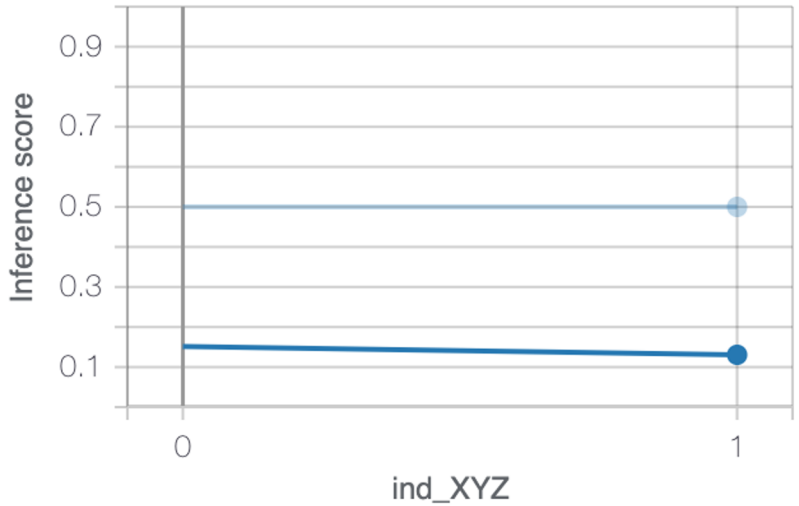
SHAP Dependence Plot



- Accounts for the interaction effects present in the features.
- A larger amount of balance would not necessarily lead to credit denial.
- If the ratio of balance divided by credit limit is low enough, a higher chance for credit approval.
- Higher level of education does not contribute to the credit approval, as we can see a flipped trend where the lower education level would receive more benefit.

What-If Tool

Do customers who already have an account with the financial institution XYZ receive any favorable treatment in your model? NO!



PDP for the credit approval prediction modeland bank XYZ.

Rejection Application Scenario

LIME

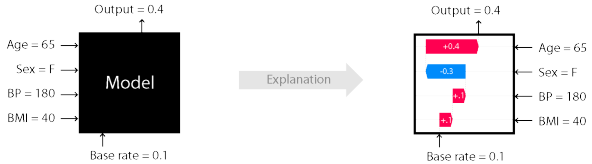
" Why **should** i **trust you**?" Explaining the predictions of any classifier

MT Ribeiro, S Singh, C Guestrin - Proceedings of the 22nd ACM …, 2016 - dl.acm.org

Despite widespread adoption, machine learning models remain mostly black boxes. Understanding the reasons behind predictions is, however, quite important in assessing **trust**, which is fundamental if one plans to take action based on a prediction, or when …

☆ Cited by 4635 Related articles All 26 versions

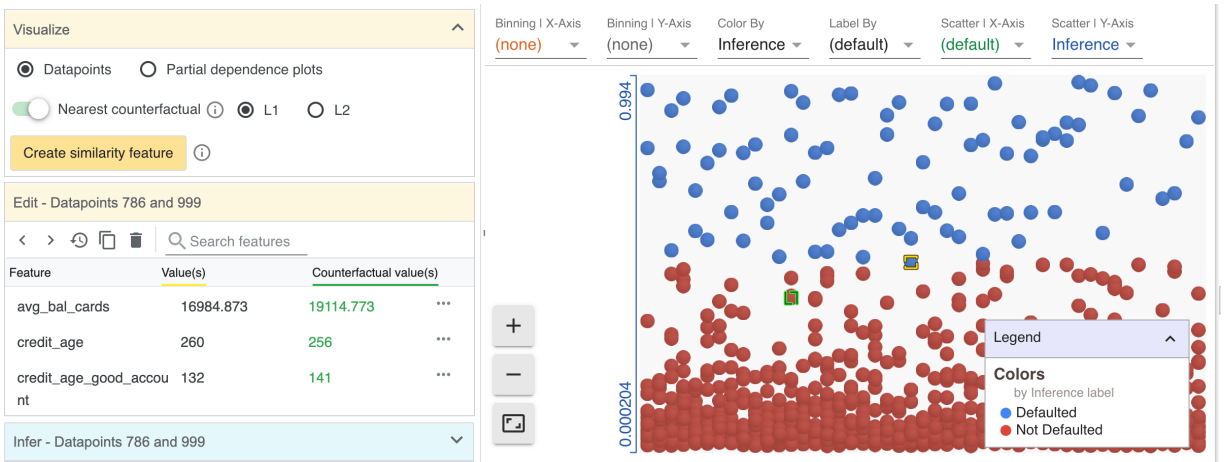
SHAP



Counterfactual Analysis For Rejected Application Scenario

What-If Tool

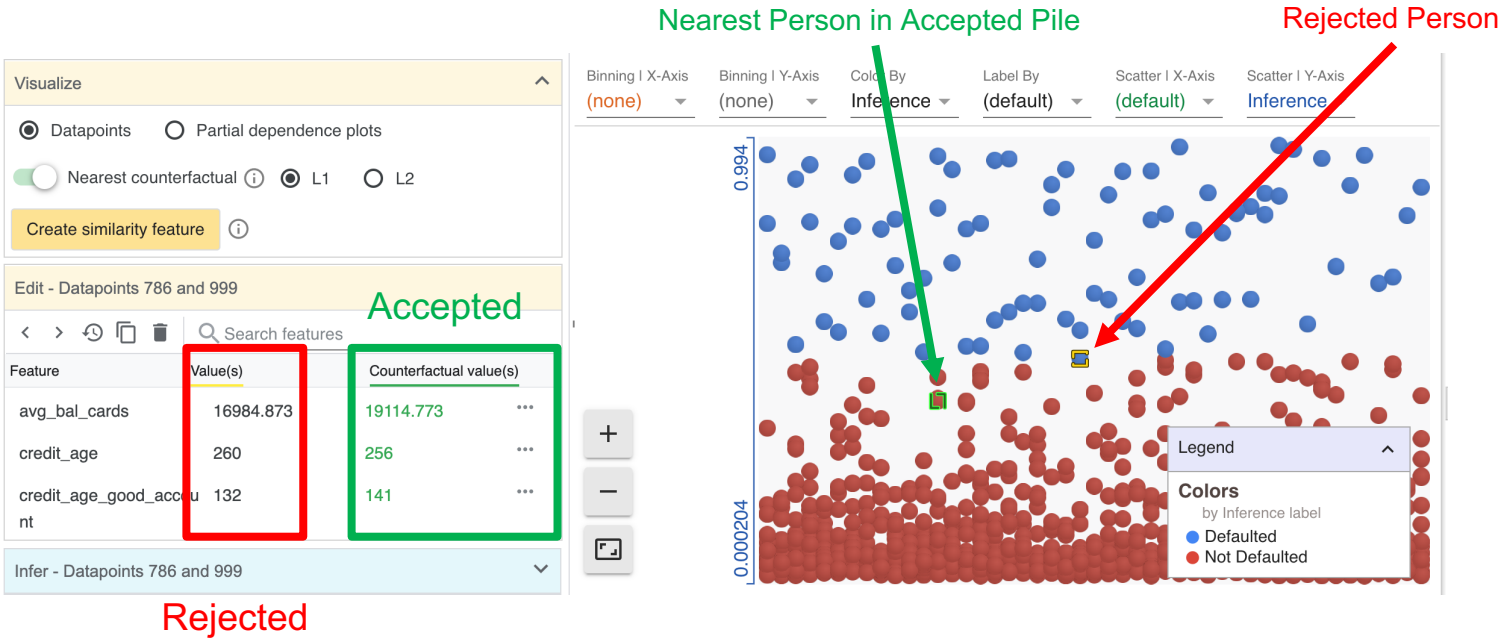
Rejection Application Scenario



Counterfactual Analysis For Rejected Application Scenario

What-If Tool

Rejection Application Scenario



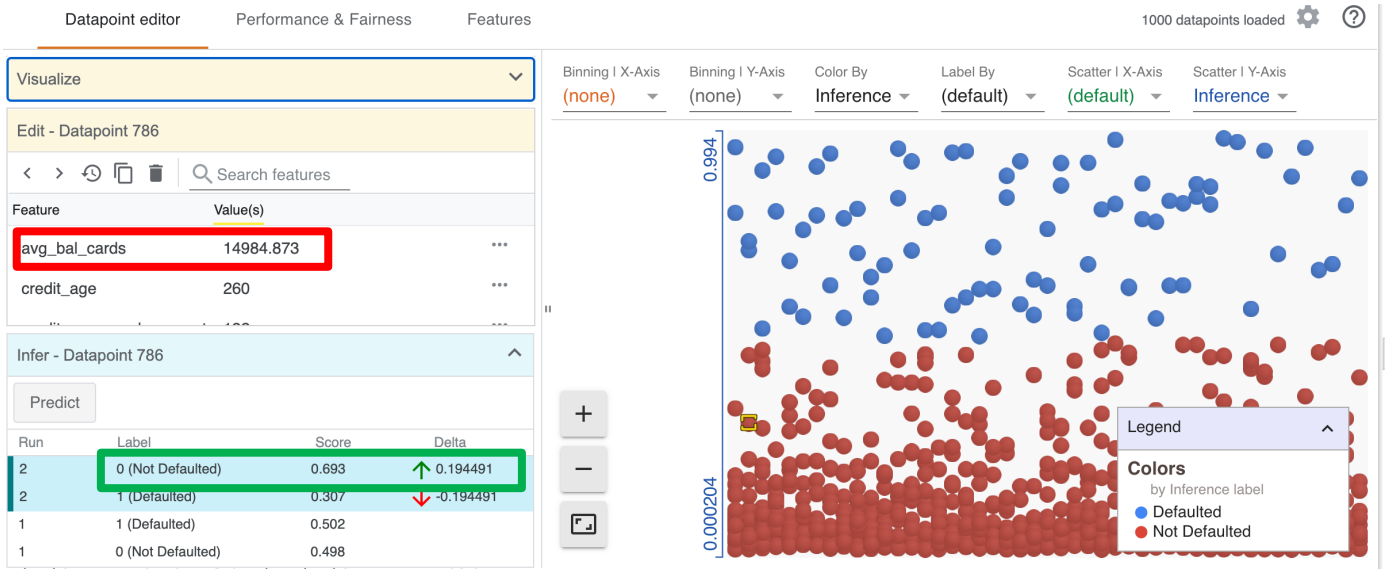
Counterfactual Analysis For Rejected Application Scenario

What-If Tool

Rejection Application Scenario

-2000

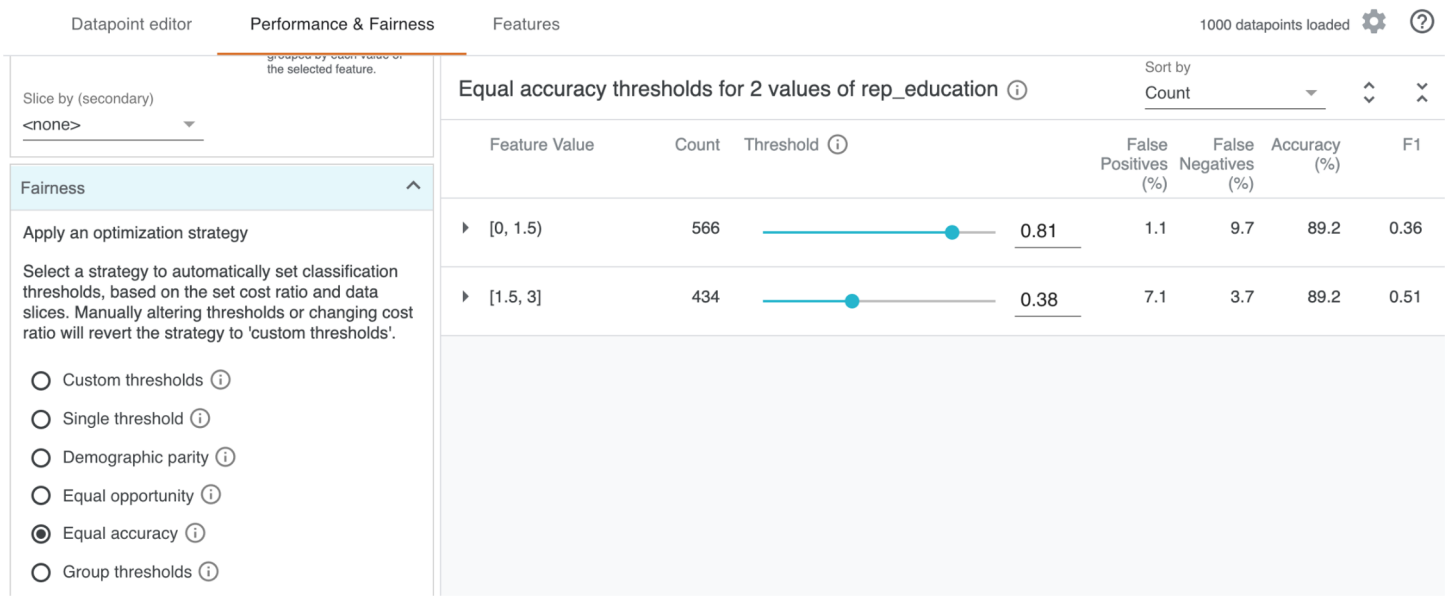
Accepted



Counterfactual Analysis For Rejected Application Scenario

What-If Tool

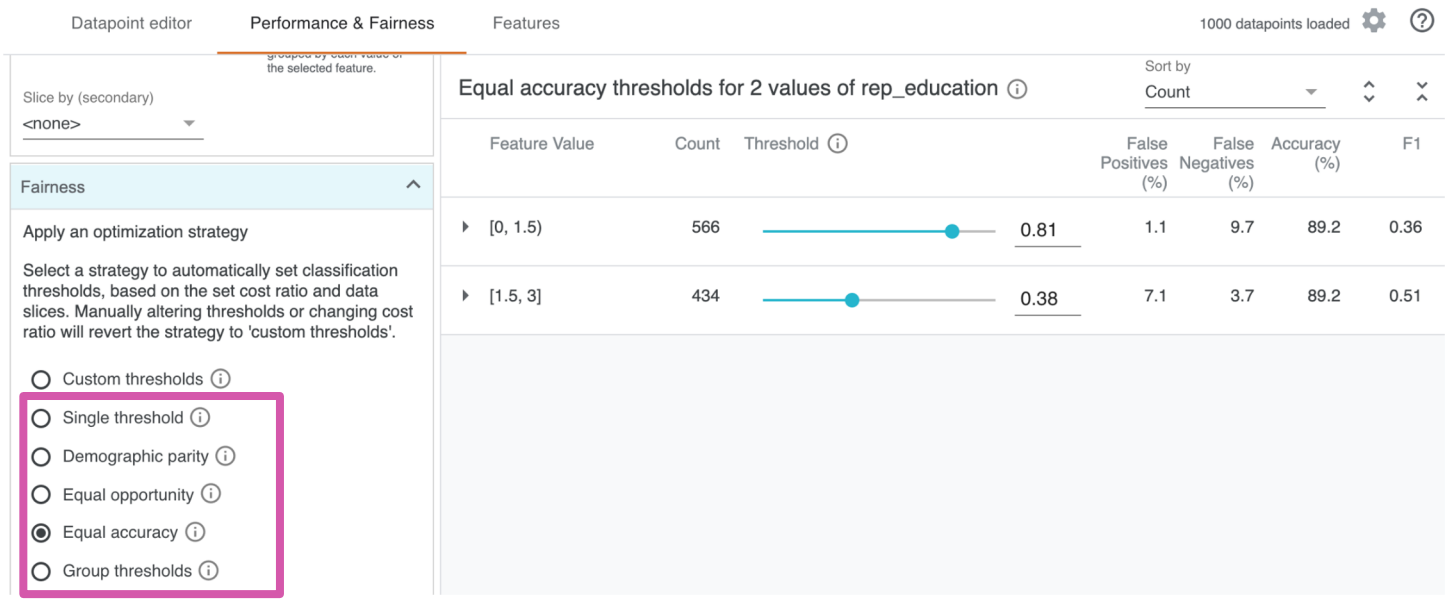
Does the Model Discriminate Against Low Educated Applicants?



Fairness Investigation Through What-If Tool

What-If Tool

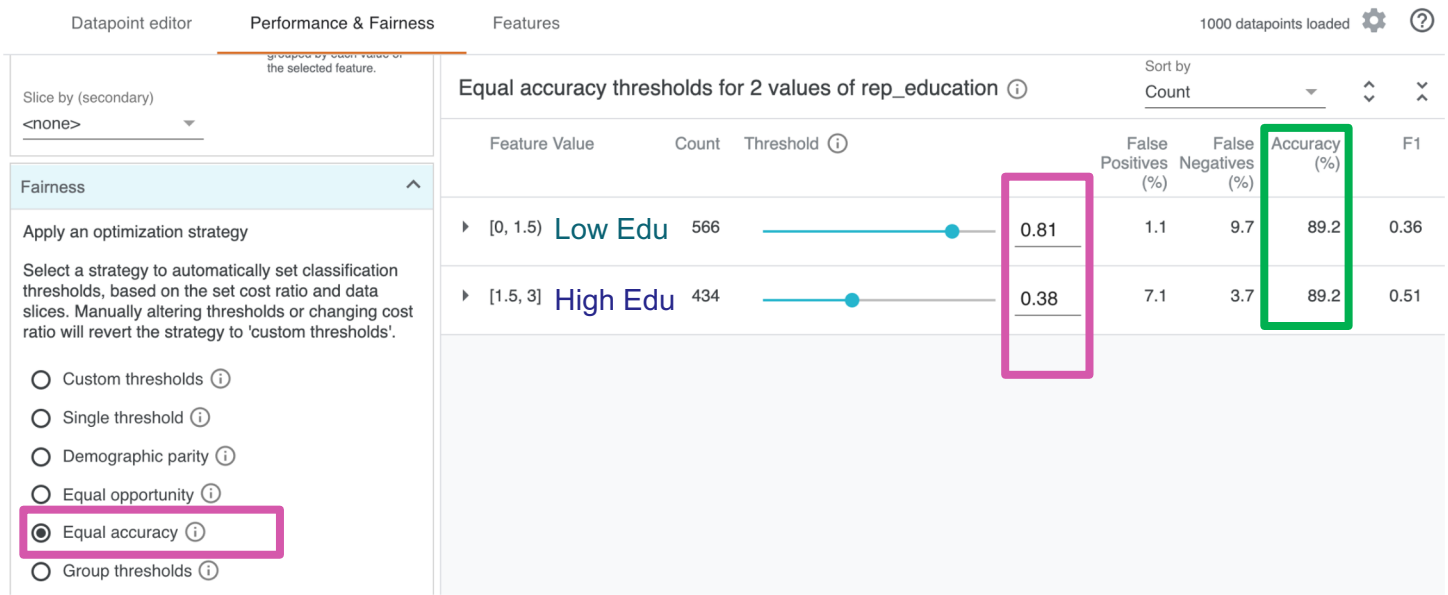
Does the Model Discriminate Against Low Educated Applicants?



Fairness Investigation Through What-If Tool

What-If Tool

Does the Model Discriminate Against Low Educated Applicants?



Fairness Investigation Through What-If Tool

Modeling Process

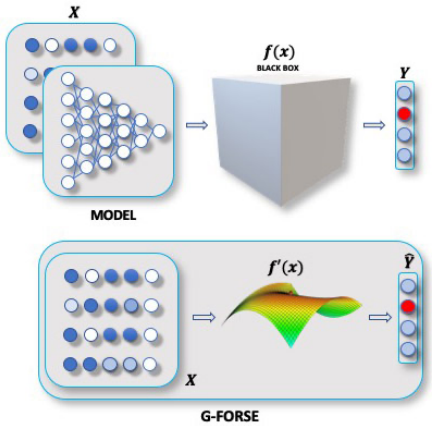
- Collecting Dataset
- Target Definition
- Data Pre-processing:
  - Handling Missing Values
  - Data Transformation
  - Feature Selection
- Model Building and Optimization
- Model Interpretation
  - What-If Scenarios
  - Fairness Investigation
- **Model Improvement**
- Conclusion

Model Improvement

Adding More Data in Training

Clustering Data by K-Means

Metamodeling: Model Stacking + 2-D ML Interpretation
(G-Forse, M Toutiaee et al, IEEE Big Data 2020)



Conclusion

- Machine Learning methods can help us assess the credit risk easier.
- Nonlinear trainer (like XGBoost) can outperform other linear models.
- Machine Learning Interpretation (MLI) toolsets can help us identify Fairness, Accountability and Transparency (FAT) in the model.

Toolsets

- Python
- R
- TensorFlow
- Google Colab
- Scikit-Learn
- What-If Tool
- SHAP

Thank you for your attention!

Questions ...