

Credit Risk Assessment

Team Credit Guard

Soumya Ganguly

Tanuj Gupta

Tushar Pandey

Aravind P. Ravi

Ankur Roy

[Link to Github Repository](#)

Project Goals

- Supervised ML models => Identify credit risk of applicants based on historical financial records
- Three most important features contributing to delinquency => Personalized suggestions for improving credibility

Data Processing

Data Source: [Kaggle - Give Me Some Credit](#)

age	NumberOfTime30-59DaysPastDueNotWorse	DebtRatio	MonthlyIncome	NumberOfOpenCreditLinesAndLoans	NumberOfTimes90DaysLate	NumberRealEstateLoansOrLines	NumberOfTime60-89DaysPastDueNotWorse
45	2	0.802982	9120.0	13	0	6	0
40	0	0.121876	2600.0	4	0	0	0
38	1	0.085113	3042.0	2	1	0	0
30	0	0.036050	3300.0	5	0	0	0
49	1	0.024926	63588.0	7	0	1	0

Initial Data Cleaning:

- Remove NA values
- Remove rows where age is 0

Data visualization:

- Set thresholds for features to remove outliers
- Plot correlation matrix

Correlation = 0.98

age	NumberOfTime30-59DaysPastDueNotWorse	DebtRatio	MonthlyIncome	NumberOfOpenCreditLinesAndLoans	NumberOfTimes90DaysLate	NumberRealEstateLoansOrLines	NumberOfTime60-89DaysPastDueNotWorse
45	2	0.802982	9120.0	13	0	6	0
40	0	0.121876	2600.0	4	0	0	0
38	1	0.085113	3042.0	2	1	0	0
30	0	0.036050	3300.0	5	0	0	0
49	1	0.024926	63588.0	7	0	1	0

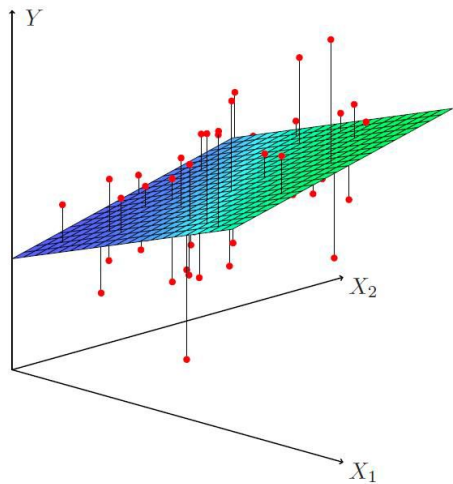
Correlation = 0.98

Correlation = 0.99

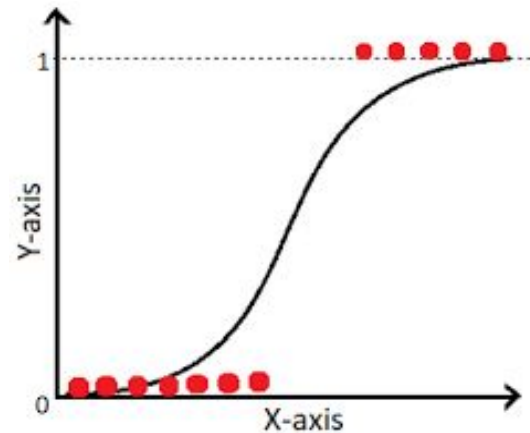
We create a weighted average of the three columns

age	DebtRatio	MonthlyIncome	NumberOfOpenCreditLinesAndLoans	NumberRealEstateLoansOrLines	NumberOfDependents	NumberOfTimeLate
45	0.802982	9120.0	13	6	2.0	0.333333
40	0.121876	2600.0	4	0	1.0	0.000000
38	0.085113	3042.0	2	0	0.0	0.666667
30	0.036050	3300.0	5	0	0.0	0.000000
49	0.024926	63588.0	7	1	0.0	0.166667

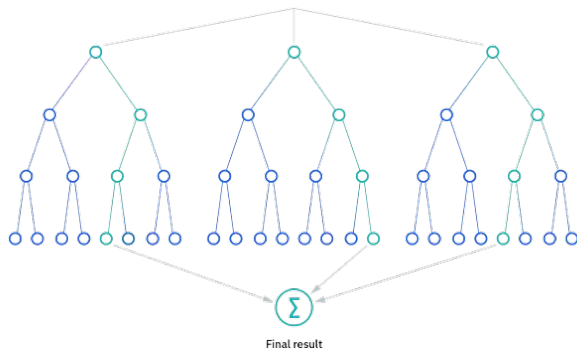
Models we used



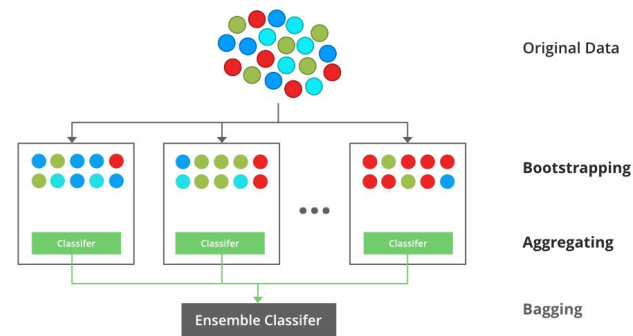
Linear Regression



Logistic Regression



Random Forest Classifier



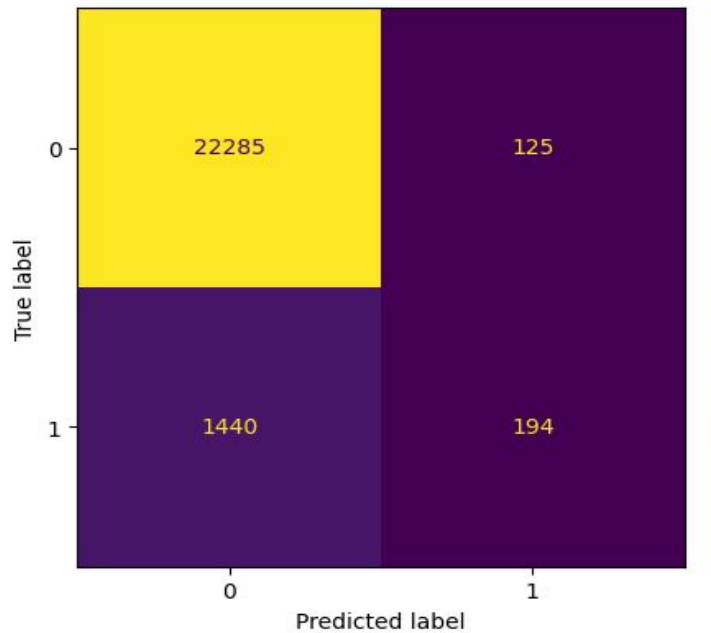
XGBoost

Comparing Models

Comparing the Models

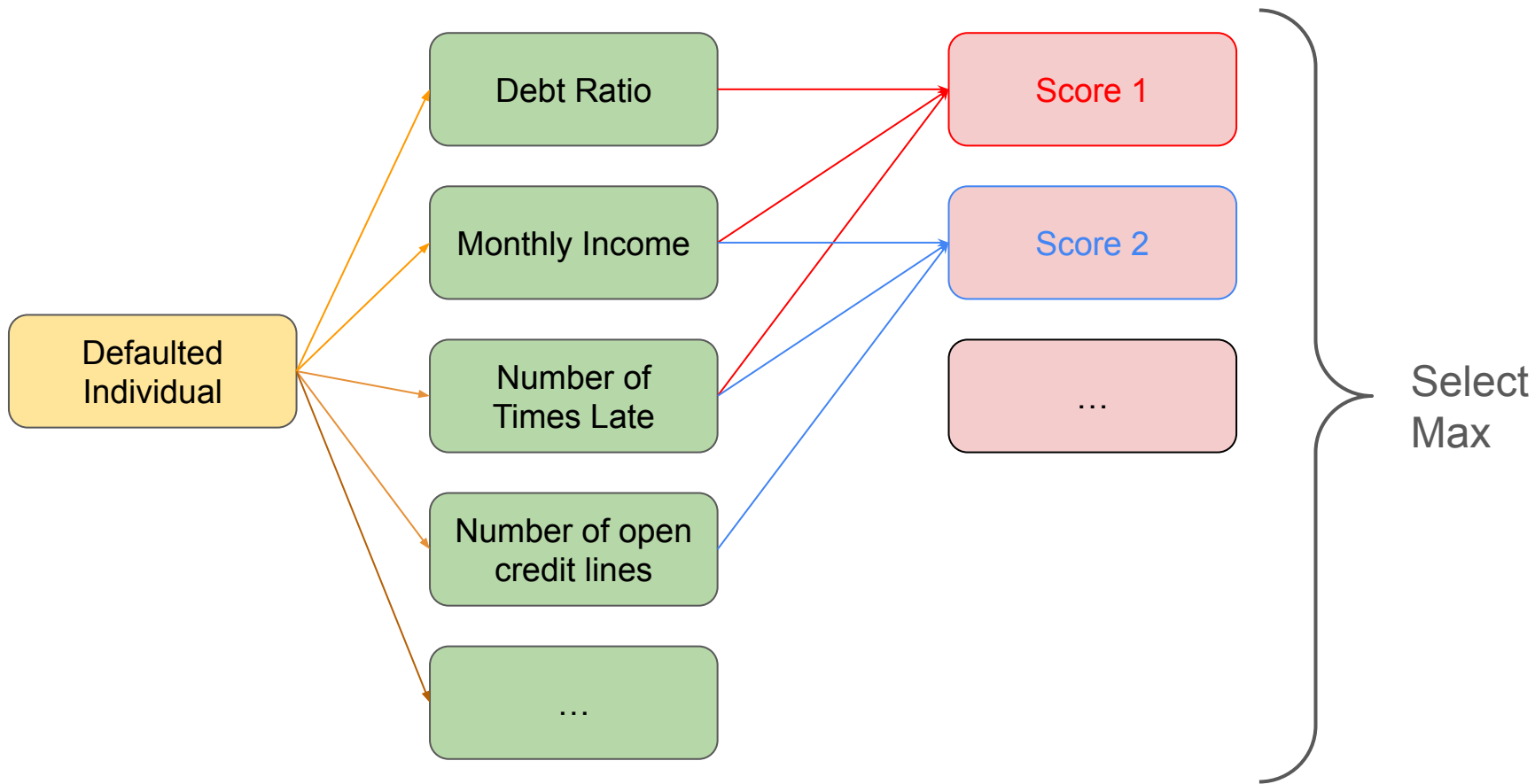
Linear Regression	0.66
Logistic Regression	0.76
Random Forest Classifier	0.81
XGBoost	0.84

AUC Score



Confusion Matrix

Personalized Recommendations



Sample Output

Features of Interest:

```
2 = RevolvingUtilizationOfUnsecuredLines;  
4 = DebtRatio;  
5 = MonthlyIncome;  
6 = NumberOfOpenCreditLinesAndLoans;  
7 = NumberRealEstateLoansOrLines;  
9 = NumberOfTimeLate
```

865

Delinquent Index: People from test data prone to delinquency based on XGBoost

```
[0, 54, 66, 76, 91, 110, 111, 117, 123, 125, 136, 177, 184, 206, 224, 259, 278, 279, 340, 380, 418, 515, 538,
```

Enter the index corresponding to the likely delinquent person from the delinquent_index array: **1065**

Person 1065 is likely to default as predicted delinquency score of 0.517001748085022 is greater than 0.4

Person 1065 can reduce chances of serious delinquency by taking care of the features: **[5 6 9]**

Thank You!

- <https://www.linkedin.com/in/soumya-ganguly-mathmae/>
- <https://www.linkedin.com/in/tanujmath/>
- <https://www.linkedin.com/in/tpmath/>
- <https://www.linkedin.com/in/apazravi/>
- <https://www.linkedin.com/in/ankur-roy1/>