Prediction of Asset Credit Rating Model of Listed Companies

**1. Goal**

Create well-performing models to accurately predict and classify the asset credit ratings of listed companies

**2. Algorithm**

The model that predicts and classifies whether the rating will decline, considers the two-category label, that is, the rating is reduced from high to "1", and if the previous condition is not met, it is "0". Establish a series of algorithm models, such as decision trees, random forests, gradient boosting trees, and logistic regression, draw ROC (trainee test curve) graphs, calculate the AUC area of ​​the model, and compare the AUC area to determine the best performance Model and put it into use

**3. Data analysis**

*①Factor acquisition*

Obtain the financial statement information of the original 7 source listed companies, and find the relevant modeling factors according to the Beneish M-Score model and the F-Score model. Each factor is as follows:

• DSR: Accounts Receivable Amount Index

• GMI: Gross Margin Index

• AQI: Intangible Assets Percentage Index

• SGI: Sales Growth Rate Index

• SGAI: Sales Administration Proportion Index

• LGVI: Leverage Ratio Index

• TATA: Cash Flow Accrual Index

• DEPI: Depreciation rate index

• ROC: Liquid Assets Proportion Index

• ROL: Ratio of assets and liabilities

• ROS: Percentage of flexible assets

• ROA: asset share

• Revenue: the company’s gross revenue

• TOT\_SIZE\_ASSETS: company capital scale

• TOT\_Profit\_label: company gross profit index

*②Label establishment*

Obtain the source rating table of listed companies and search for listed companies whose ratings were AAA or above in the previous year and dropped below AAA in the next year. If the conditions are met, give the label ‘1’, otherwise the label is ‘0’

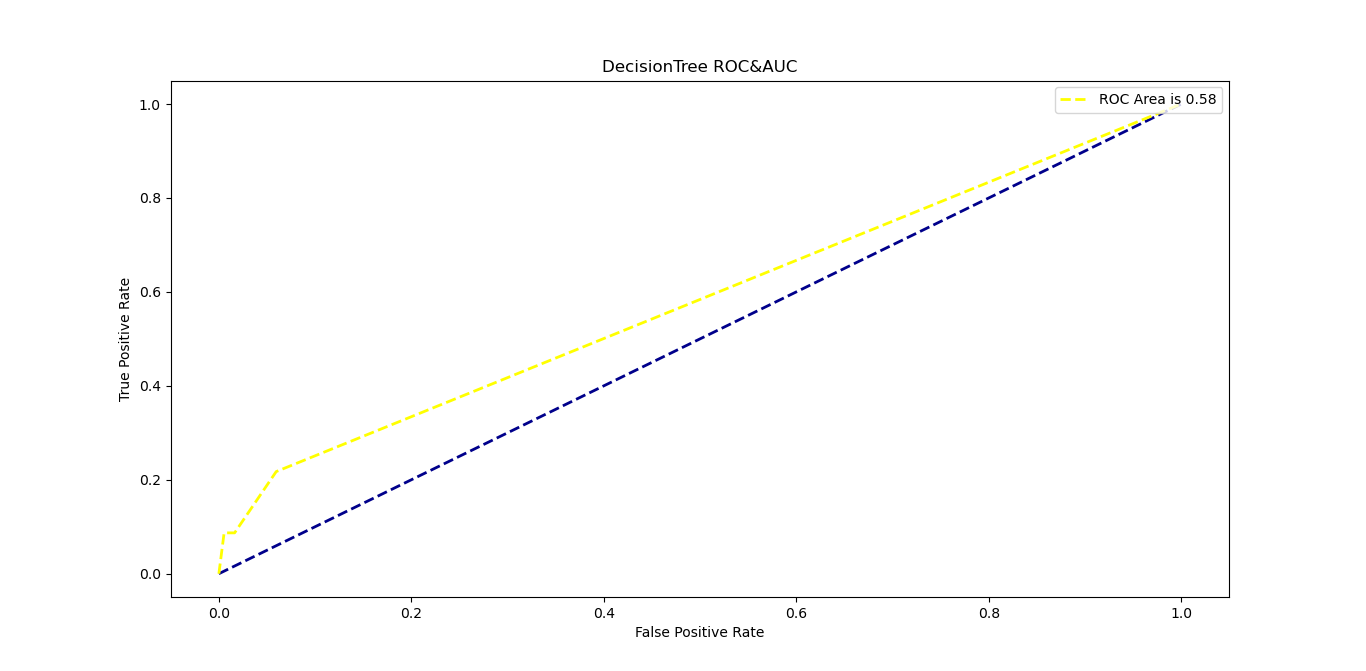
*③Data cleaning & preprocessing*

After obtaining all the factor and label tables through the two steps of ①②, filter out the factor items with a null value rate of more than 80% and remove them, and use the median filling method to remove the null values ​​for the remaining null rate factor items. Finally get a table with no missing values

**4. Model creation and analysis**

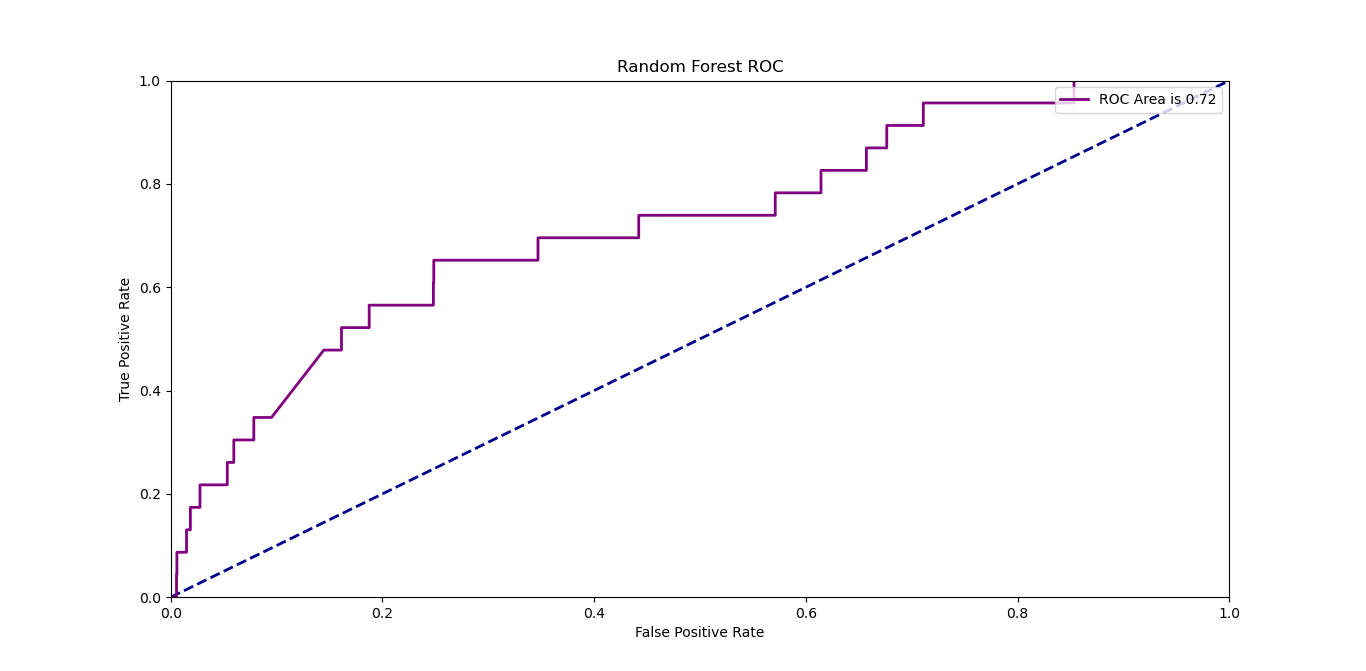
Refer to the two classification algorithm, establish a series of classification models, draw ROC curve and calculate AUC area. The usage model and corresponding ROC curve are as follows:

*① Decision tree model (DecisionTree)*



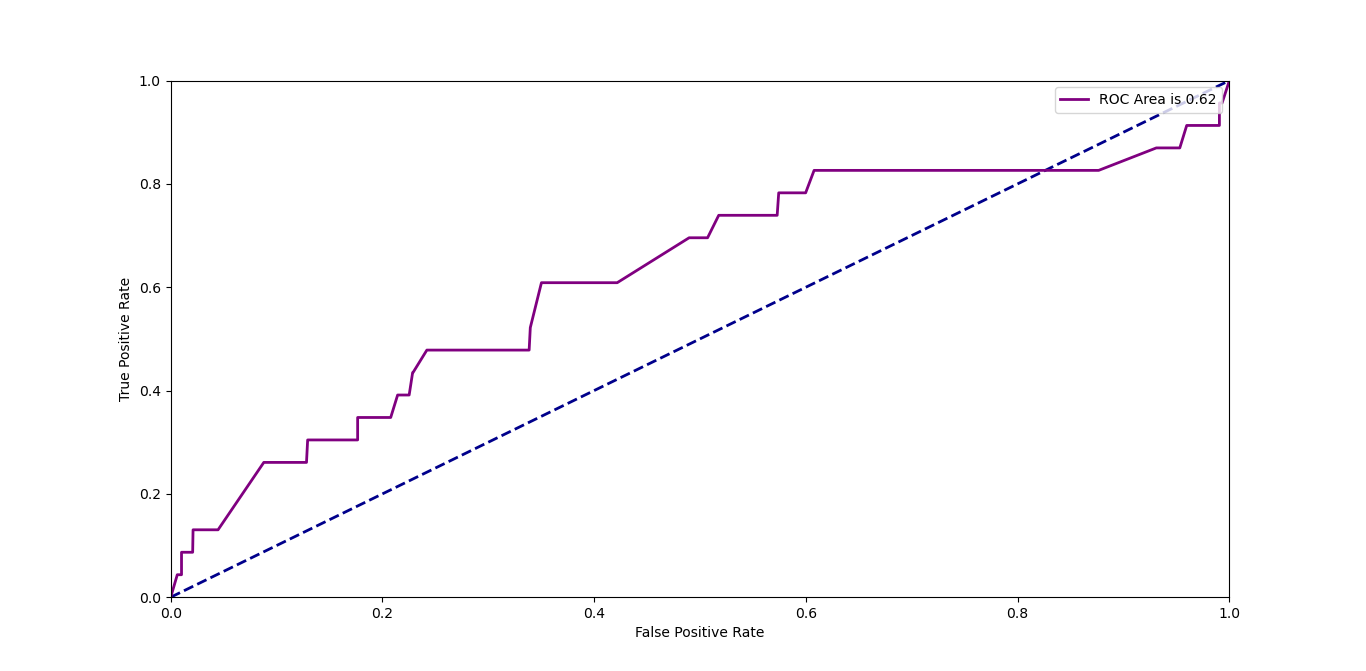
AUC=0.58

*②Random Forest Model (RandomForest)*



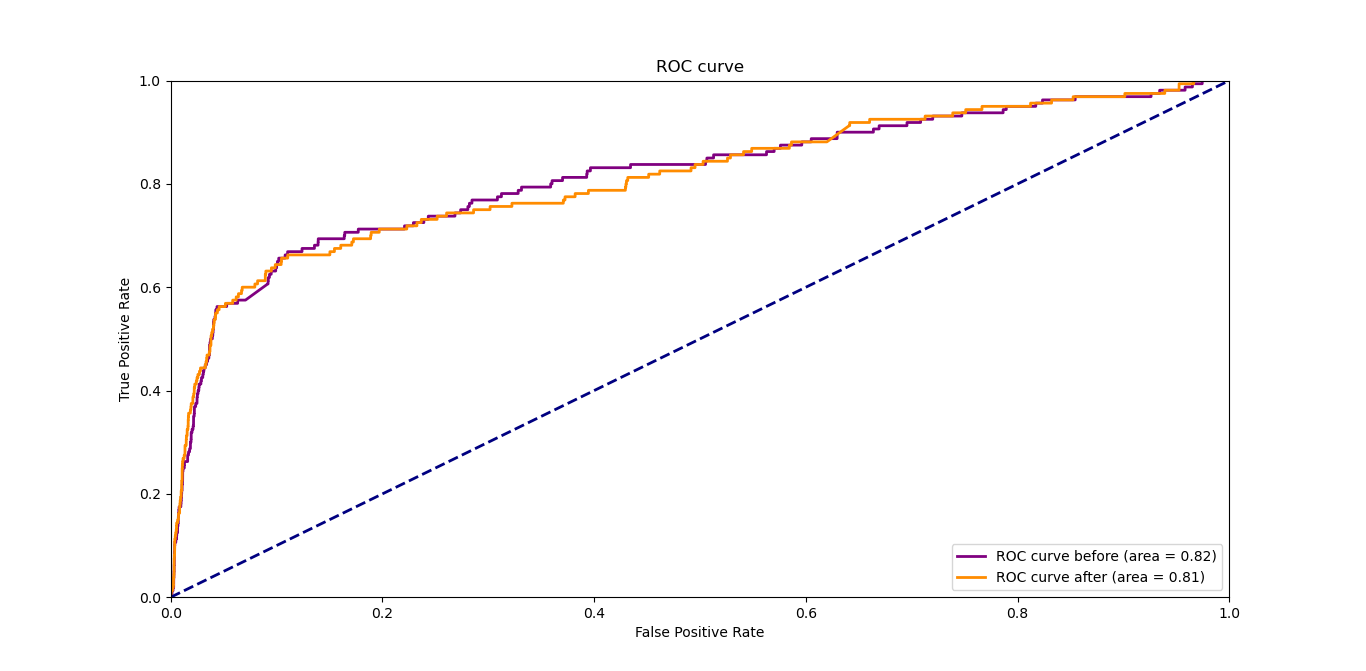
AUC=0.72

*③Gradient Boosting Decision Tree Model (GBDT)*



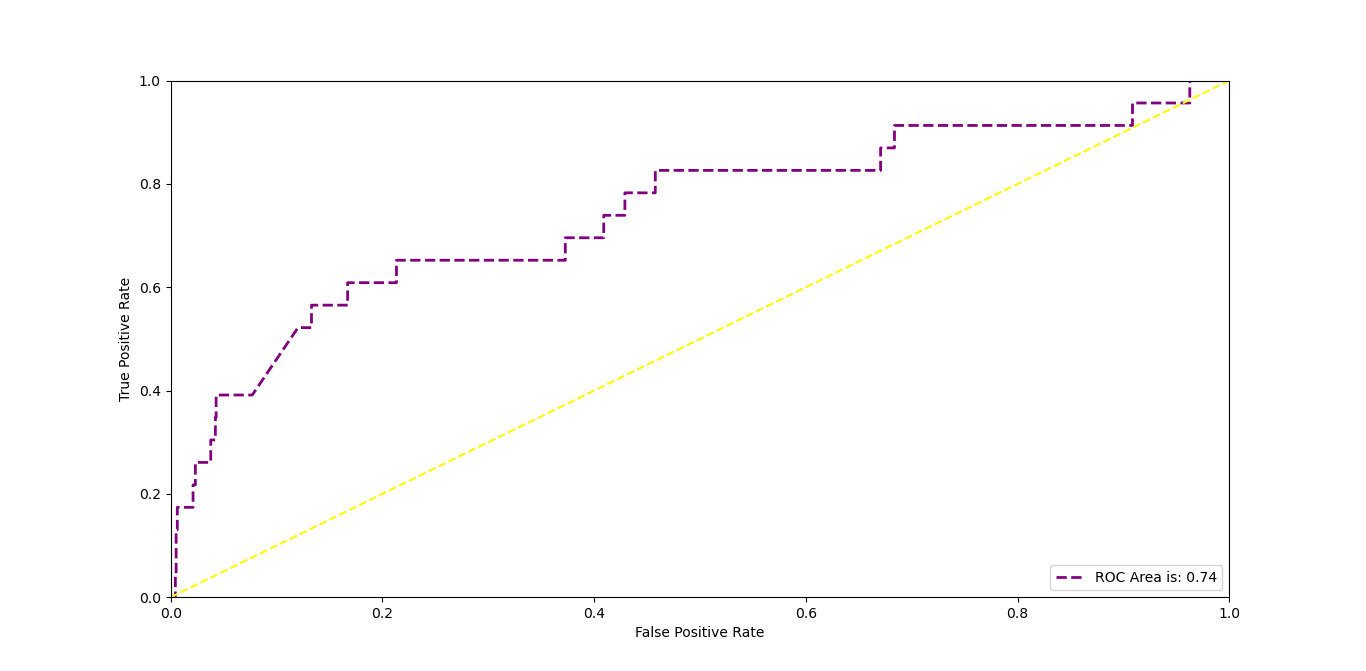
AUC=0.62

*④ Logistic regression model (LogisticsRegression)*



AUC=0.81

*⑤ XGBoost tree model*



AUC=0.74

**5 Conclusion**

According to the model creation in the previous chapter and the comparison of the AUC area, the logistic regression model has the best form, with an AUC area of ​​0.81