Hello Everyone,

My topic for classification module is company bankruptcy prediction

The goal of this project is use financial ratios to predict bankruptcy risk or potential financial distress of public companies

The target users of this project are bank creditors or investors.

It is important for them to know a company’s financial distress or bankruptcy risk before make the decision of granting a loan or investing of a company

Here is a quick overview of the data and methodology

The data used for training and testing the classification model is from Taiwan Journal Economics.

It contains financial ratios of 800 - 900 listed companies from 1999 to 2009 with nearly 7000 rows and 96 features.

The data was trained and compared through different machine learning classification models.

After building the model, I interested in applying the model on the SP500 companies and predicting what companies has higher risk of bankruptcy in the most recent three years. I collected the SP500 from Merging online database with financial ratios from 2018 to 2020

Next, we can see that the data is highly imbalanced

True class indicating: bankrupt

False class indicating No bankrupt

From the creditor or bank point of view, they don't want to suffer much risk from giving money to a company that has high bankruptcy risk.

That means the users of the model has low tolerance to high False Negative, and relatively higher high tolerance on False Positive.

Therefore, I choose Recall and roc\_auc as evaluation metrics for selecting and training the models (ROC: Receiver operating characteristic)