Springboard
Data Science track
Capstone project I

Predicting the likelihood of company bankruptcy using machine learning

By Qiwei Lu
Aug 2020

Table of Contents

1. Introduction *2*

2. Methodology *3*

2.1 Data acquisition **3**

2.2 Data exploration **3**

2.2.1 Financial ratios explanation 4

2.2.2 Outliers 6

2.2.3 Features collinearity 7

2.2.4 Feature selection 8

2.2.5 Data imbalance 9

3. Model selection and training *10*

4. Modeling results and analysis *12*

5. Rethinking on the feature selection *19*

6. Tuning classification threshold *21*

7. Application of the test result *22*

8. Assumptions and limitations *24*

9. Future work *24*

10. Conclusion *25*

11. References *25*

1. Introduction

This is a time when “too big to fail” has become a myth. “Nothing is impossible” is probably a better summary for the market. From the financial giants trembled in 2008 to the retailing empires like Sears, Forever21 and Toys R Us foundered in recent years. People may wonder, who will vanish and who will survive?

While most people will only be surprised by a company’s vulnerability when the news hits the frontpage news, there are shareholders who have to monitor and assess the company’s insolvency on a regular basis:

a. The management of companies exposed to bankruptcy risk: the c-suite executives have to monitor the balance of its overall financial and operational performance to prevent any adverse sign that may lead to bankruptcy. Although, under current law frame, company’s claiming bankruptcy sometimes means opportunities like restructure while continuing to operate, suspend or reconfigure debt payment, and get back on its feet, the switch of management style, massively laying off employees and closing locations can be harmful to the brand image and interests of its investors. It is best if the management can be alerted of some red flags before they hit the restart button.
b. Financial service providers such as independent external auditors and consultants: external auditors have an obligation in assessing a company’s going concern issue before the client acceptance and before issuing a report. The financial consultant may need to provide professional advice towards the overall health of a company.
c. The investors: in the event that a publicly-listed company declares bankruptcy, the company's shareholders may be entitled to a portion of the liquidated assets, depending on which shares they hold and how much liquid assets are leftover. However, the stock itself will become worthless, leaving shareholders unable to sell their defunct shares. That being said, the investors want to be the first one to jump off the sinking boat and sell their stocks before the stock price takes a nose down.
d. The regulators: regulators such as the SEC need to issue guidance for the public accounting firm to do the external audit every year to consider the going concern in the protection for the investors. They will also have to monitor the fraud and any regulation violation in the market. Since going concern is the most basic assumption for any financial reports, public companies will be committing fraud if a going concern issue is noticed internally but is not properly disclosed.

Given the fact that the judgement of going concern can influence various kinds of groups, research on bankruptcy prediction can be dated back as early as 1932, when FitzPatrick published a study of 20 pairs of firms, one failed and one surviving, matched by date, size and industry, in The Certified Public Accountant. He did not perform statistical analysis as is now common, but he thoughtfully interpreted the ratios and trends in the ratios. Today, with the power of machine learning technologies, a more thorough and statistical analysis becomes possible and necessary.

While the chain reaction in a financial crisis or short recession caused by COVID-19 pandemic may be sudden and difficult to foresee, people with business charisma always know that the incident reported on the newspaper is always the last straw rather than the main cause. The underlying risks are there hidden in the financial reports of the company long before the bankruptcy, waiting for you to explore. The problem I want to solve in this research is how to predict the bankruptcy of a company given its financial health indicated by financial ratios using machine learning, and to interpret the indications drawn from the results from the models for the business world.

2. Methodology

2.1 Data acquisition

The dataset I used for this research includes financial ratios and bankruptcy information of some Polish companies. The data was downloaded from a Kaggle competition. The initiator of the competition has collected, cleaned and randomly selected posted dataset from UCI Machine Learning Repository. The original data is from Emerging Markets Information Service (EMIS), which is a database containing information on emerging markets around the world. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013.

2.2 Data exploration

Within the dataset, 64 financial ratios are given for each company and the class column is the target, with 1 to be bankrupt and 0 to be not bankrupt yet. 10,000 companies are included in the dataset.

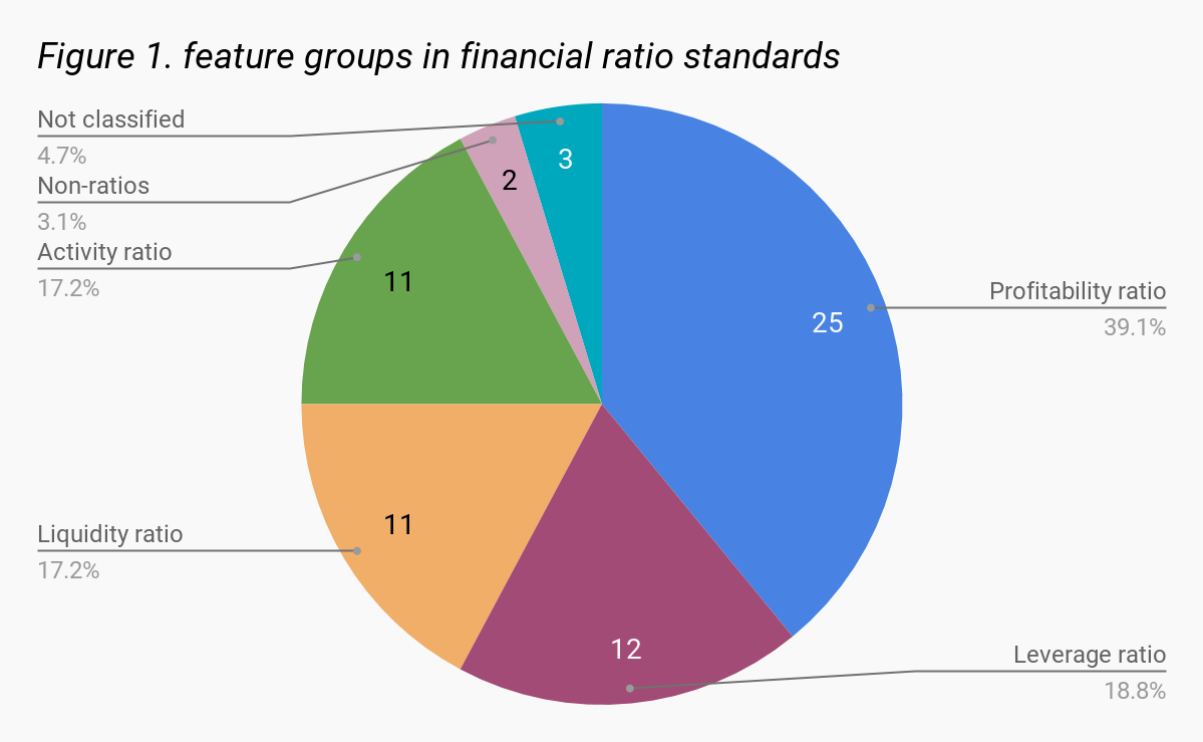
Since the data was cleaned beforehand, there is no missing data. While browsing the data, I noticed that two of the 64 attributes(predictors) are not actually financial ratios. One is attribute 29: logarithm of total assets and another is attribute 55: logarithm of working capital. These two attributes will be paid more attention to because the absolute value of an asset or working capital is not comparable among companies of varying sizes and industries.

2.2.1 Financial ratios explanation

While python is the language for data stories, accounting is the language for business stories. Before fitting the features to the models, it is always beneficial to have some domain knowledge. In accounting context, financial ratios can be generally divided into five categories:

A. liquidity ratios, which demonstrate a company's ability to pay its debts and other liabilities. If it does not have enough short-term assets to cover short-term obligations, or it does not generate enough cash flow to cover costs, it may face financial problems.
B. Activity ratios, which demonstrate a company's efficiency in operations. In other words, you can see how well the company uses its resources, such as assets available, to generate sales.
C. Leverage ratios, which demonstrate a company's ability to pay its long-term debt. These ratios examine a company's dependence on debt for its operations and the likelihood it can repay its obligations.
D. Performance/profitability ratios, which tell investors about a company's profit, which explains why they are frequently referred to as profitability ratios.
E. Valuation ratios, which rely on a company's current share price, provide a picture of whether or not the stock makes a compelling investment at current levels.

The valuation ratios, however, were not used in the dataset. I have categorized all the financial ratios under the rest 4 groups and found out while most attributes fall in the 4 groups, 2 attributes, like I mentioned before, are not ratios and another 3 ratios cannot be properly classified into any group based on my knowledge. The distribution is shown in figure 1.



All the ratio classification is based on personal best judgement

The most important financial ratios, which most accountants will be checking, are current ratio, debt/equity ratio, cash flow to debt ratio and operating cash flow to sales ratio. The first two are included in the features (attr4 and attr8 respectively). There are, however, no ratios involving any value from the cash flow statement, which includes crucial indicators for bankruptcy. Since the cash flow is a change value rather than a point value, without knowing values like the change of working capital, it is impossible to get the ratios of cash flow through feature combination. On the other hand, under the accounting standards, although some attributes can be grouped under the 4 groups, they are not good ratios to represent the company's financial situations. So after I grouped all the ratios, I realized the dataset is not a perfect whole package of commonly used financial ratios.

When looking at the formulas to calculate the ratios in the description, I found that the dataset used the same accounting values, like total asset from the balance sheet or gross income and net profit from income statement, on calculating multiple ratios. There is a big chance that ratios containing the same accounting ratios may have a strong correlation. And when we think about it, since there are multiple ratios explaining the same aspect of the company's financial situations, there could be a tendency that the data are correlated with each other to some degree. There will be a deeper look into this when we discuss feature collinearity later in section 2.2.4.

2.2.2 Outliers

To get a general idea of the outliers, I used z-score, which describes any data point by finding their relationship with the standard deviation and mean of the group of data points. The formula is $z = (x-\mu)/\sigma$, where x is the raw score, μ is the population mean, and σ is the population standard deviation. In most of the cases, a threshold of 3 or -3 is used. That means, if the Z-score value is greater than or less than 3 or -3 respectively, that data point will be identified as outliers.

After calculating z-score to all the data points, I found 2409 data points whose z scores are over 3, among which 22 companies declared bankruptcy. Company wise, there are 1038 companies that have one or more financial ratios that are over 3 and 519 that have two or more financial ratios that are over 3.

After calculating z-score to all the data points, I found outliers are not rare. All the 64 attributes contain outliers in their columns and 1038 companies out of 10,000 have one or more outliers. The top two attributes that have outliers are attribute 16 and attribute 26. The boxplots of these two attributes are shown below:

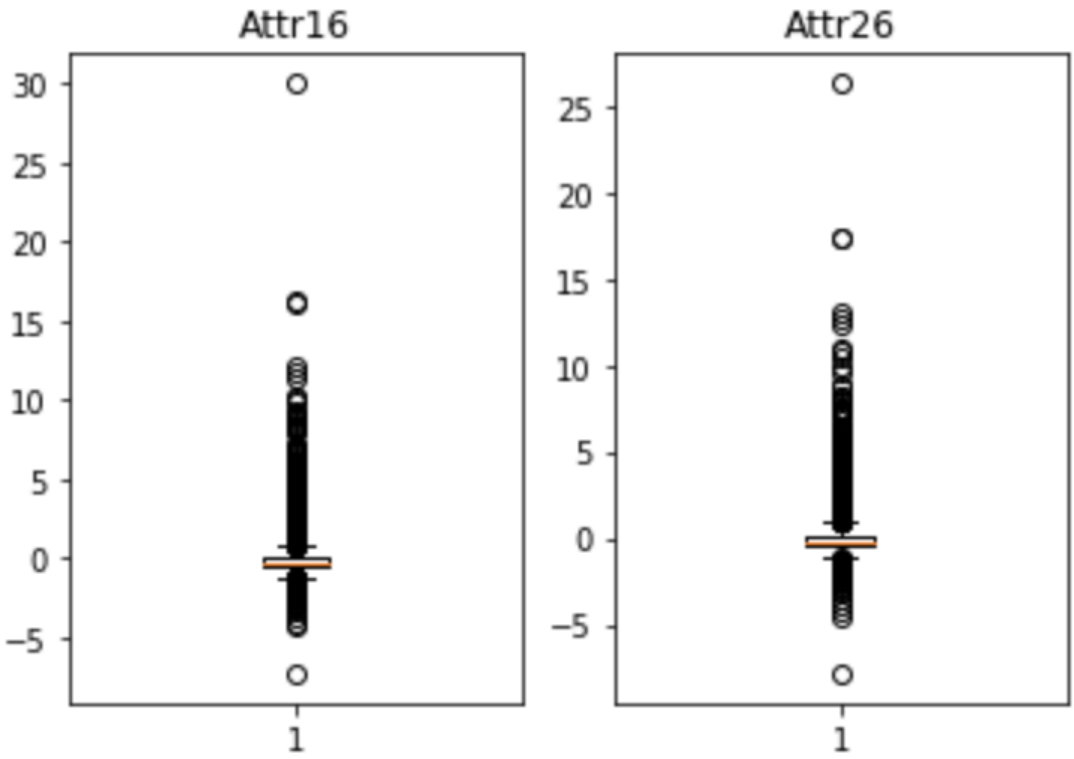


Figure 2, box plots for the two features that have the most outliers

We can see that for these two attributes, the outliers seem like a normal part of the whole distribution span. Only two or three extreme values are at the far top or end. Other value points where z-score is over 3 are displayed evenly in the span. The two attributes that are not ratios as we mentioned above: the “logarithm of total assets” and the “logarithm of working capital” are among the top 20 attributes that have the most outliers, confirming my intuition.

2.2.3 Features collinearity

As I have discussed above, since most of the ratios can be divided into 4 groups to represent the financial performance of a company. If a company has been running well, it should have a decent profit (profitability ratio), a decent amount of borrowings to finance its operation (leverage ratio), can pay back its short-term debt (liquidity ratio) and high efficiency in operation and production(activity ratio). Thus, the features in the dataset may be related to each other to a certain degree. Let’s take a deeper look using heatmap for the correlation among all the attributes here.

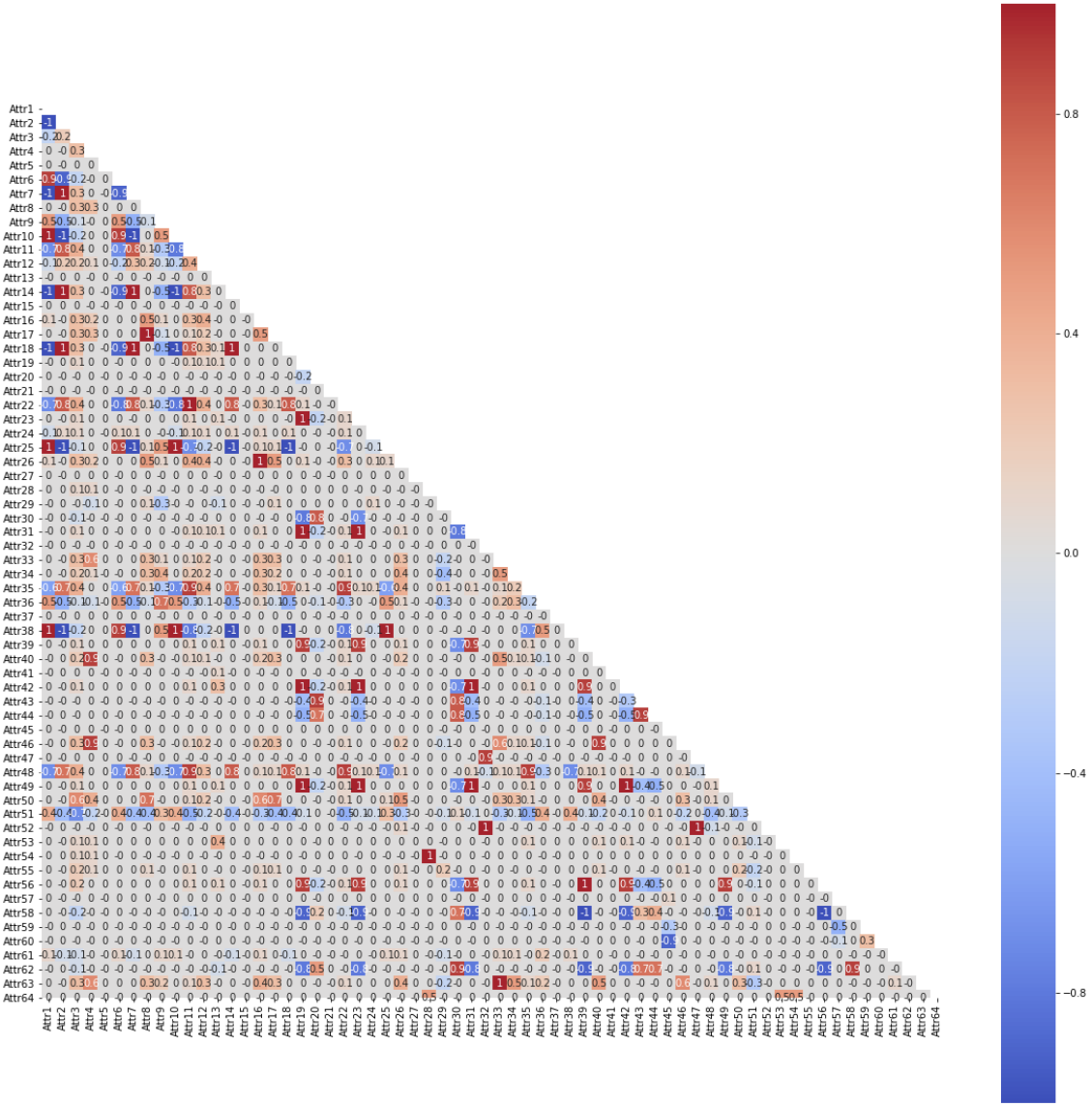


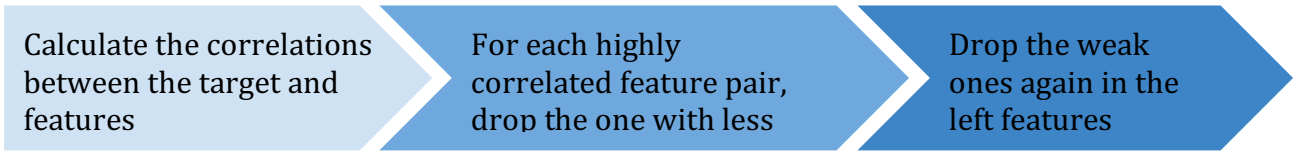
Figure 3, heatmap for 64 features

As we can see from the heatmap, there are a lot of attributes that have strong linear correlation with other attributes, which can negatively impact our model performance in the future. We should apply appropriate feature selection methods to better fit the models.

2.2.4 Feature selection

The feature selection methodology is as follows:

Figure 4, procedures for feature selection



Because the features are continuous values and the output target is categorical values, Anova F-test will be used to calculate the scores between features and the target.

Next step I go back to the heatmap and find the feature pair that has a correlation over 0.95, which is a very high threshold to filter all the feature pairs that have a strong level of correlation relationship. We can change this threshold later after we find the best model and try different sets of features by adjusting how strict the filter is.

I drop the feature that has less explaining power to the target and keep the one that has a higher Anova F-test score to the target. After this process, 17 attributes have been dropped. To continue with the process, I again calculated the Anova F-test for the 47 attributes left and found a proper threshold: 0.1 to filter out the weak features. The threshold may be changed later but for 0.1, I found another 12 weak features to drop.

After the whole process is completed, 35 features are left that are not highly correlated among each other but all have a strong relationship with the target. These 35 features will be the basis we split the train and test set.

2.2.5 Data imbalance

Just like many classification problem datasets, there is a big imbalance in the data. Among 10,000 companies, only 203 companies are labeled as positive. That is a 2% positive rate. This bias in the dataset can influence many machine learning algorithms, leading some to ignore the minority class entirely. To deal with the highly skewed dataset, I took several steps through the research:

A. Use stratified split for the train test split and cross validation to make sure the two labels show up proportionally in train, validation and test set.
B. Use random oversampling, random undersampling and Synthetic Minority Oversampling Technique (SMOTE) methods to resample the data and train on

the chosen models respectively and see which method will lead to an ideal performance when tested on test data.

C. Use multiple evaluation metrics to assess models. In the research, accuracy score, recall score, precision score and ROC AUC score are all calculated and are compared among models. Using only the accuracy score can be highly misleading since the model can blindly predict all the labels to be negative and still get a high accuracy score. Within the four metrics, the recall score and ROC AUC are the most important benchmark. Since bankruptcy is a huge red flag, as much as true positive should be captured.

D. Tune the threshold for classification. For those classification problems that have a severe class imbalance, the default threshold o.5 can result in poor performance. After we pick the best model, we can tune the threshold to improve the accuracy of the classification even more.

3. Model selection and training

I have chosen 7 different models in this research ranging from simple model to more complex ensemble models:

A. Extreme Gradient Boosting classifier
Extreme Gradient Boosting classifier(Xbgoost) implements the gradient boosting decision tree algorithm. This algorithm goes by lots of different names such as gradient boosting, multiple additive regression trees, stochastic gradient boosting or gradient boosting machines. Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made. XGBoost dominates structured or tabular datasets on classification and regression predictive modeling problems. The evidence is that it is the go-to algorithm for competition winners on the Kaggle competitive data science platform.

B. Support Vector Machine(SVM)classifier
The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

C. Logistic Regression classifier
Logistic Regression classifier is a linear model for classification. In this model, the possibilities describing the possible outcome of a single trial are modeled using logistic regression.

D. Decision Tree classifier
A Decision Tree is a simple representation for classifying examples. It is a Supervised Machine Learning where the data is continuously split according to a certain parameter.

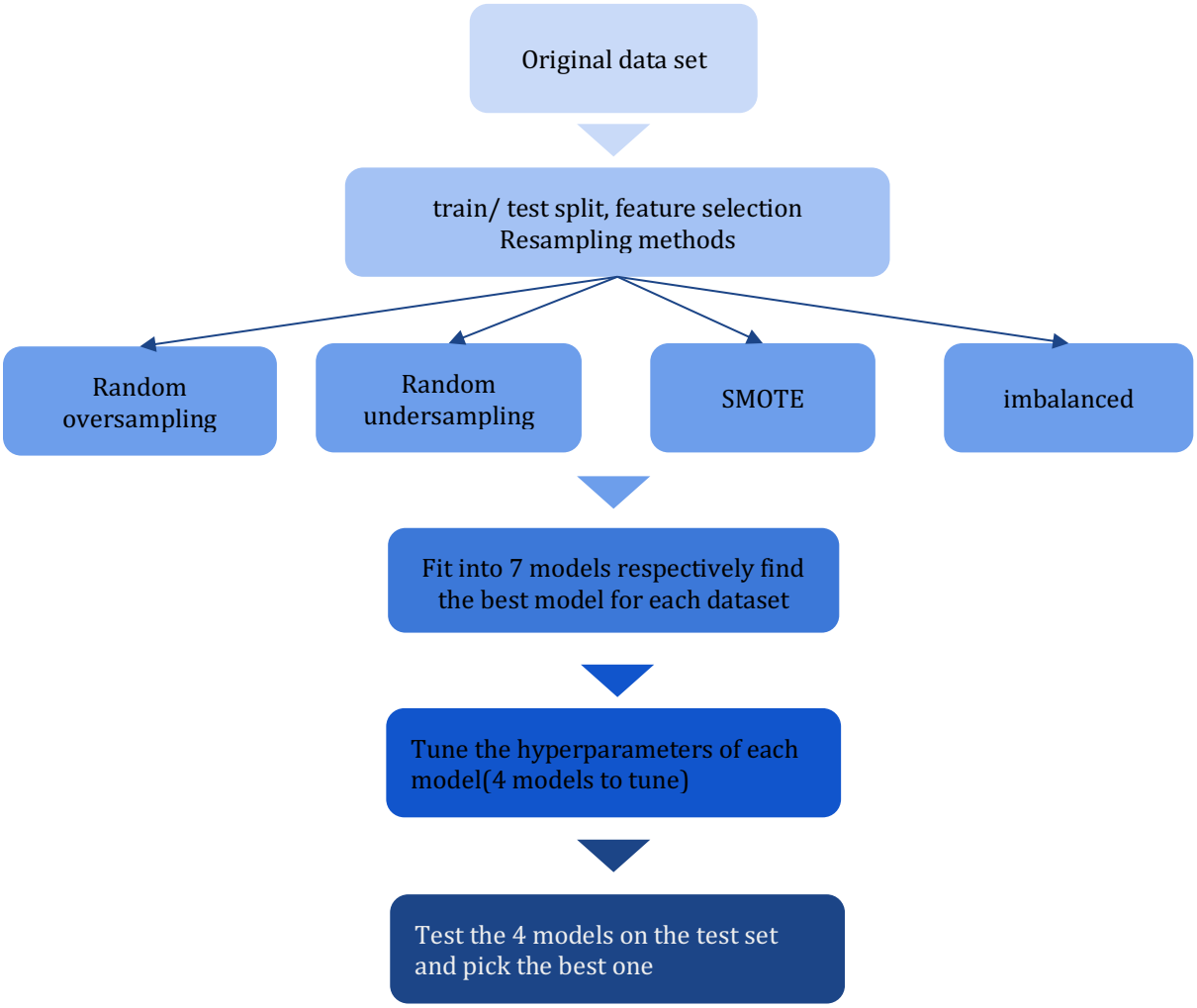
E. Adaboost classifier
Ada-boost or Adaptive Boosting is one of ensemble boosting classifiers. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get a high accuracy strong classifier. The basic concept behind Adaboost is to set the weights of classifiers and training the data sample in each iteration such that it ensures the accurate predictions of unusual observations.

F. Random Forest classifier
Random Forest classifier is an ensemble tree-based learning algorithm. The Random Forest Classifier is a set of decision trees from randomly selected subset of training set. It aggregates the votes from different decision trees to decide the final class of the test object.

G. Balanced Bagging classifier
Bagging is the application of Bootstrap procedure in a high-variance machine learning algorithm. This implementation of Bagging is similar to the scikit-learn implementation. It includes an additional step to balance the training set at fit time using a RandomUnderSampler.

We use stratified K fold cross validation to get the performance metrics on the 7 models trained by 3 training datasets that have been processed with 3 different resampling methods (random oversampling, random undersampling and SMOTE) and another untouched original imbalanced dataset. We fit these four datasets into 7 models separately and get the 4 scores to evaluate, within each dataset, which model works the best. We then have 4 best performed models and we will tune the hyperparameters to even boost their performance. After we get the four models equipped with their best parameters, we will proceed to the testing part, where we test the 4 models with the hold-out test data we set aside earlier and see which model finally wins the game. The process is shown in figure 5.

Figure 5. flow chart of overall procedures



4. Modeling results and analysis

After training the models on the data, let's take a look at the results and pick the best model for each dataset.

A. Random oversampling

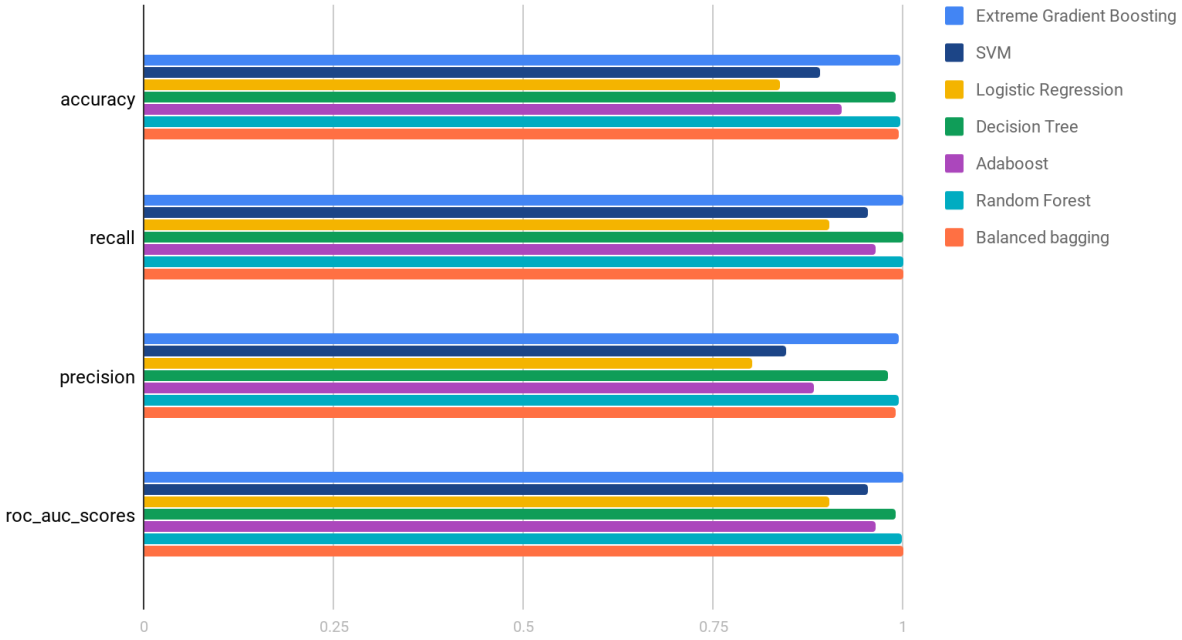
The evaluation metrics for 7 models are shown as below:

Table 1. oversampling scores

	Extreme Gradient Boosting	***SVM***	***Logistic Regression***	***Decision Tree***	***Adaboost***	***Random Forest***	***Balanced bagging***
accuracy	*0.997385*	*0.889831*	*0.838926*	*0.989793*	*0.918346*	*0.996938*	*0.994833*
recall	*1*	*0.953432*	*0.903038*	*1*	*0.964532*	*1*	*1*
precision	*0.994798*	*0.84586*	*0.800433*	*0.979998*	*0.882978*	*0.993914*	*0.989777*
Roc auc scores	*1*	*0.95487*	*0.902844*	*0.989793*	*0.963197*	*0.99949*	*0.999936*

It can be seen that the Extreme Gradient Boosting classifier is the best we can pick.

Figure 6. Oversampling metrics



B. Random undersampling

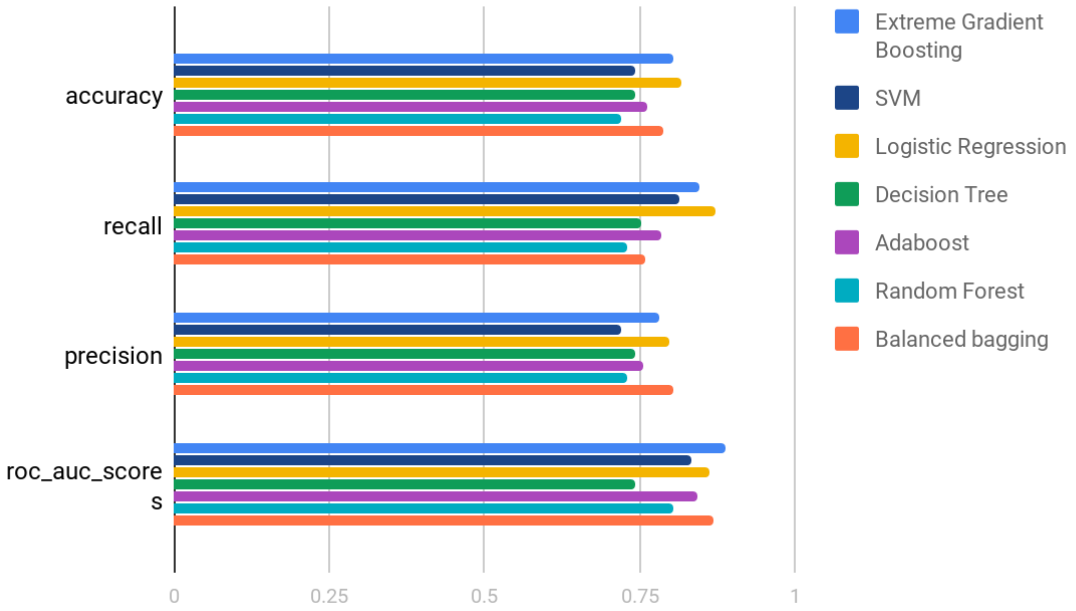
The metric scores are shown below:

Table 2. undersampling scores

	Extreme Gradient Boosting	***SVM***	***Logistic Regression***	***Decision Tree***	***Adaboost***	***Random Forest***	***Balanced bagging***
accuracy	0.8025	0.743606	0.817788	0.74375	0.7625	0.719087	0.786875
recall	0.845833	0.814394	0.870644	0.752652	0.784659	0.728977	0.75947
precision	0.782982	0.719542	0.796078	0.741112	0.754912	0.729585	0.804525

Roc auc scores	0.887743	0.834221	0.861245	0.743845	0.842353	0.804841	0.868901

Figure 7, undersampling metrics



We can see there is a competition between Extreme gradient boosting classifier and logistic regression classifier. Within the four metrics, recall score and roc auc score are the most important ones to consider. So this round, logistic regression performed a little bit better than Extreme Gradient Boosting classifier.

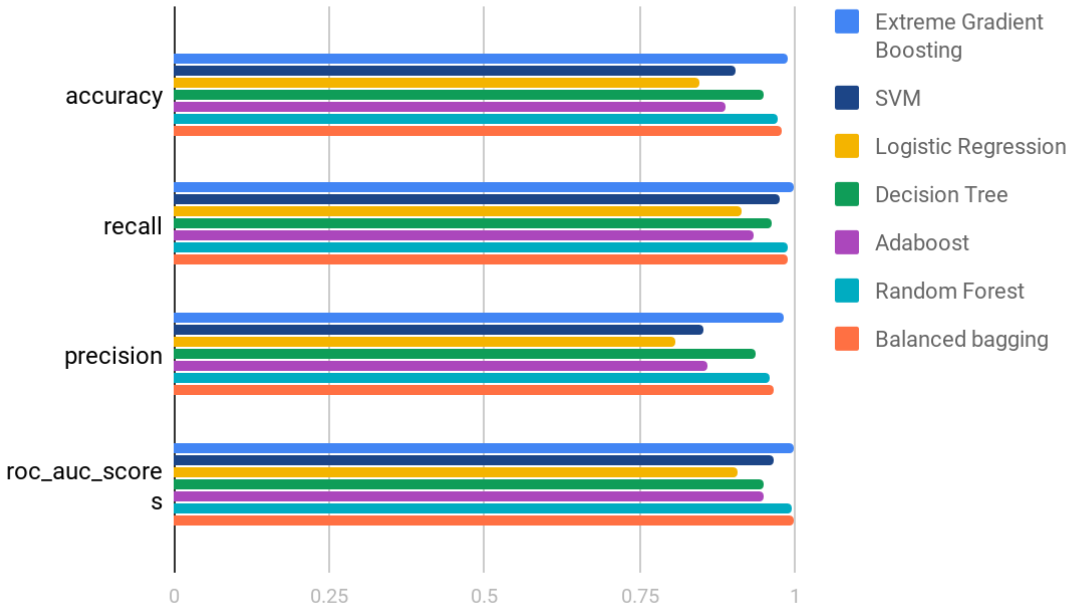
C. SMOTE

The metric score are shown below:

Table 3. SMOTE scores

	Extreme Gradient Boosting	SVM	Logistic Regression	Decision Tree	Adaboost	Random Forest	Balanced bagging
accuracy	0.989793	0.903738	0.847219	0.948584	0.889002	0.972952	0.977418
recall	0.99898	0.975505	0.914646	0.963766	0.932382	0.988517	0.989921
precision	0.980966	0.853118	0.805992	0.9354	0.858035	0.958694	0.965775
roc auc scores	0.999523	0.965399	0.905984	0.948584	0.948384	0.994488	0.996632

Figure 8. SMOTE metrics



Extreme gradient boosting classifier has undoubtedly won this time.

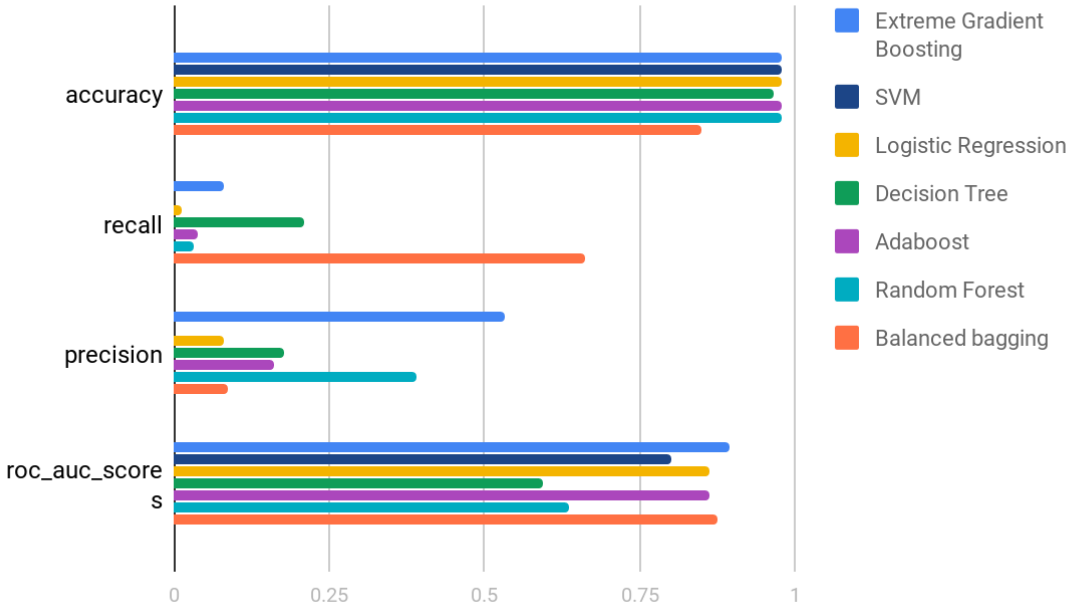
D. Imbalance

The metric scores are shown below:

Table 4. imbalance scores

	Extreme Gradient Boosting	SVM	Logistic Regression	Decision Tree	Adaboost	Random Forest	Balanced bagging
accuracy	0.98	0.97975	0.97825	0.9645	0.977875	0.978625	0.8505
recall	0.080114	0	0.0125	0.209091	0.036932	0.030492	0.660606
precision	0.533333	0	0.08	0.178147	0.16	0.39	0.085753
roc auc scores	0.894078	0.79991	0.862369	0.594594	0.863814	0.636536	0.873928

Figure 9, imbalance metrics



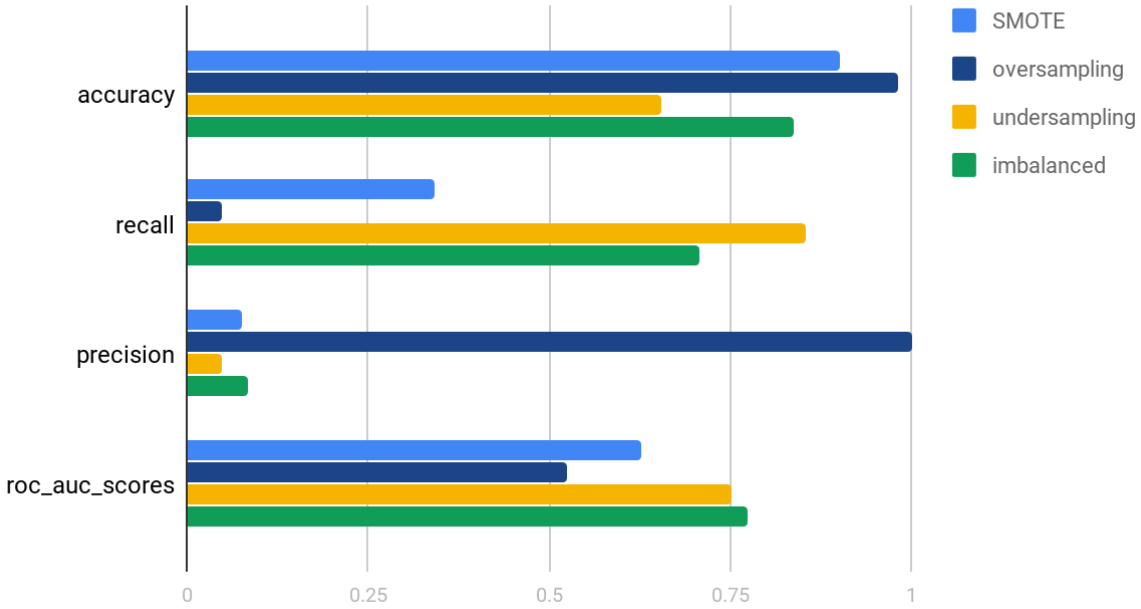
We can see there is a difficult decision to make here: extreme gradient boosting classifier performed very well except for the recall score, which is very low, compared to balanced bagging classifier. Because we have determined that recall score is more important than accuracy score, we will choose the balanced bagging classifier which has the best recall score and good roc auc score.

After we got the four best models: Extreme Gradient Boosting classifiers for both SMOTE and oversamping datasets; For undersampling, logistic regression performed a little bit better; for the imbalanced dataset, balanced bagging model is the best. Next step, I will fine-tune the parameters of these models within the different datasets using Grid search. After the best parameters are found, the hold-out test data will be fit into the four models and the results will be compared.

Table 5, test set scores

	SMOTE	oversampling	undersampling	imbalanced
accuracy	0.9	0.9805	0.6545	0.8355
recall	0.341463	0.04878	0.853659	0.707317
precision	0.074866	1	0.048611	0.083815
Roc auc scores	0.626577	0.52439	0.751995	0.77275

Figure 10, final test set metrics

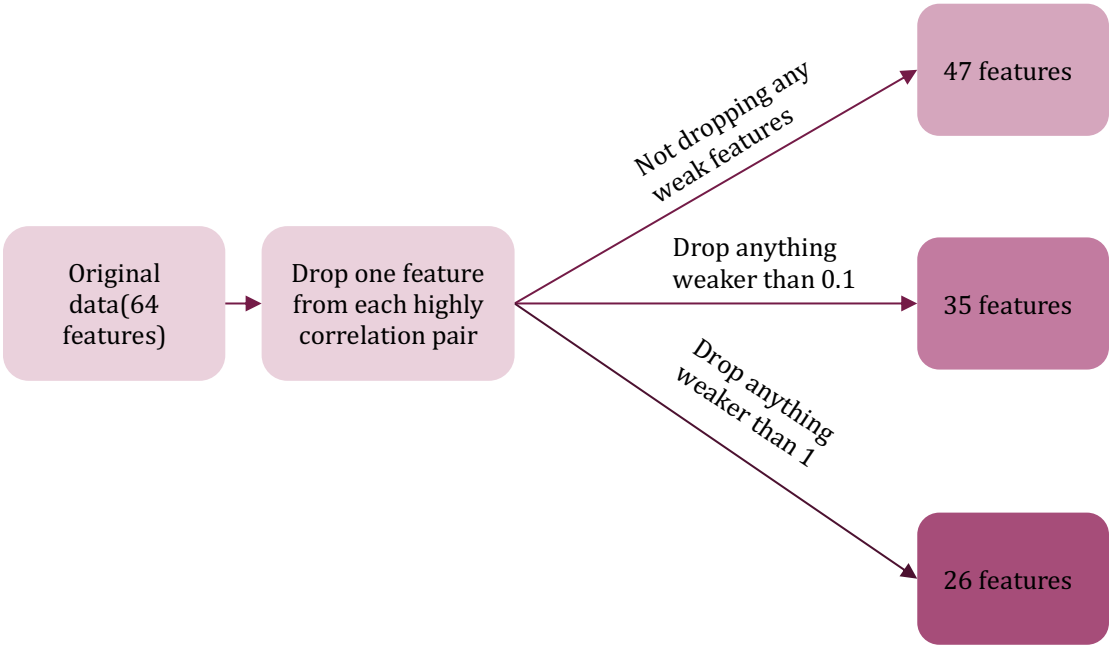


In deciding the final pick, there is a dilemma between the results of undersampling and imbalanced datasets. Although the undersampling's recall score is higher than the imbalanced's, imbalanced's recall score is relatively reasonable and the other scores are all higher than undersampling. To make a small compromise, we get an overall premium performance from the final pick: Balanced bagging classifier using the imbalanced data set.

5. Rethinking on the feature selection

When we rewind and go back to when we did the feature selection, we set the threshold of Anova f-test score to be 0.1, so any feature's score less than 0.1 were dropped. And we got 35 features. What if we move the threshold up and down? Now we have picked the best model, we can use the balanced bagging model to test on different features numbers and see which performs better. The procedures are shown as figure 11.

Figure11, changing feature selection threshold

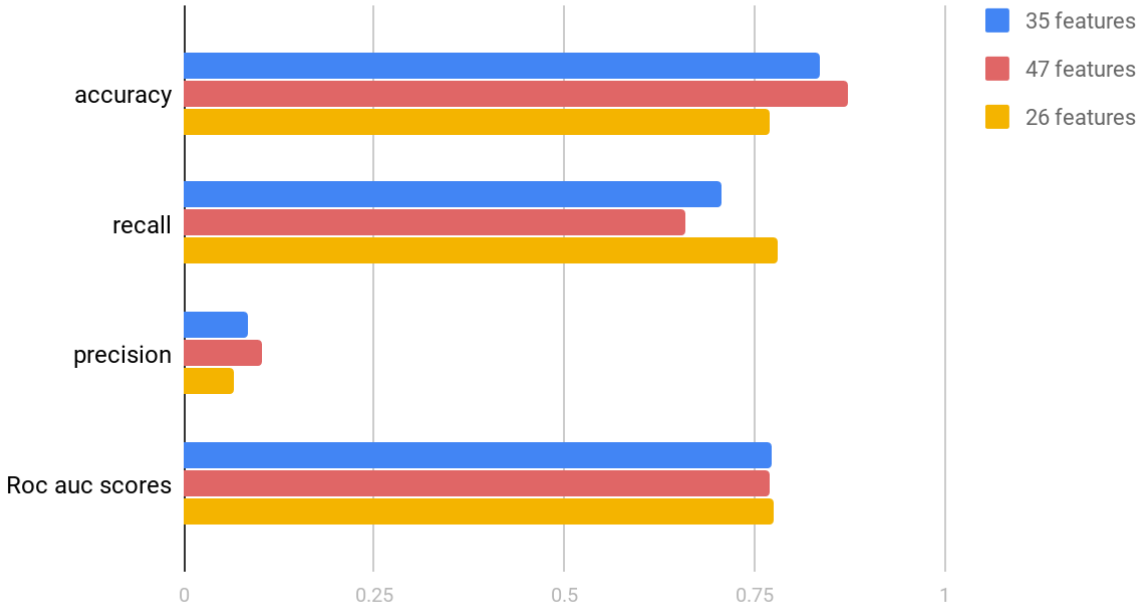


Through the process of moving the threshold for dropping the weak features, we got two new datasets with 47 features and 26 features respectively. We will fit them into the best model we picked through testing on the test set - the balanced bagging classifier, and to compare with the 35 features training set to see if there is any improvement.

Table 6, scores change caused by feature selection

	35 features	47 features	26 features
accuracy	0.8355	0.8735	0.769
recall	0.707317	0.6585365854	0.7804878049
precision	0.083815	0.1015037594	0.06597938144
Roc auc scores	0.77275	0.7682677822	0.774623688

Figure 12, score comparison between different feature sets



As we can see from the score above, changing the feature selection threshold did improve the performance from the original 35 features. 47 features have better accuracy score and precision score and 26 features have better recall score and roc auc score. Since recall and roc auc scores are the most important metrics, balanced bagging classifier model with 26 features is our final pick.

6. Tuning classification threshold

As we have mentioned in 2.25 data imbalance, the default threshold for interpreting probabilities to class labels is 0.5. But using the default threshold in a severe class imbalanced dataset can result in poor performance. Therefore, we have to tune the threshold just like we tune other parameters. We are using the ROC curve to find the optimal threshold. The metrics here to search for the optimal is the Geometric Mean or G-Mean, which is a metric for imbalanced classification that, if optimized, will seek a balance between the sensitivity and the specificity. The statistics are shown below:

Sensitivity = TruePositive / (TruePositive + FalseNegative)

Specificity = TrueNegative / (FalsePositive + TrueNegative)

Where:

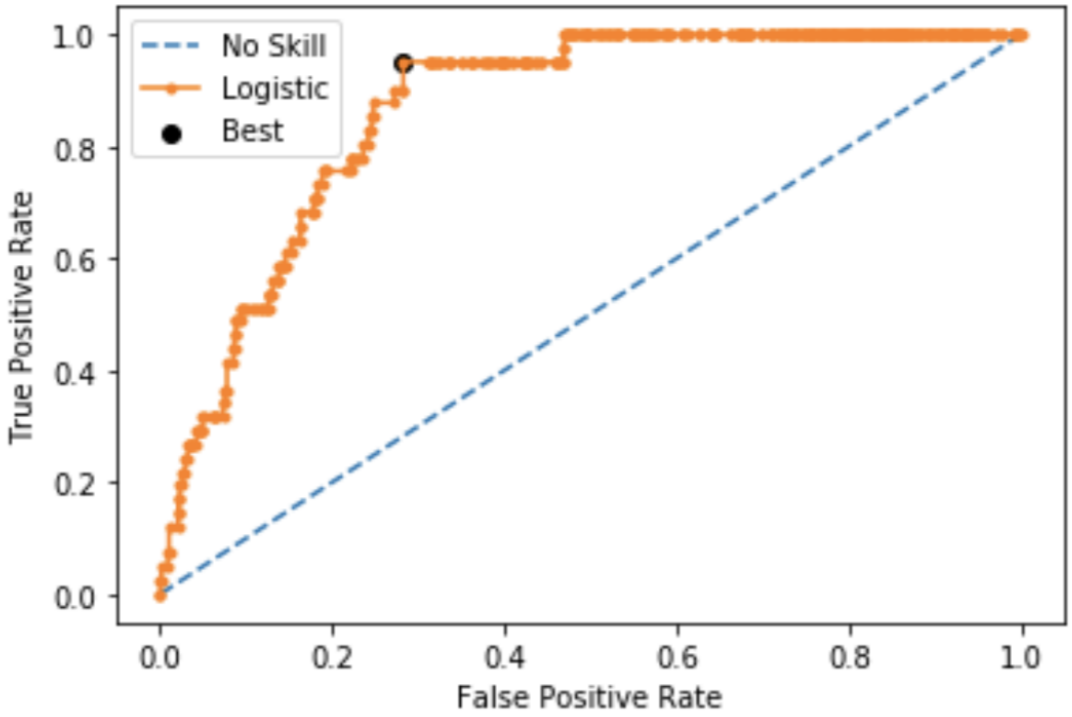
Sensitivity = True Positive Rate

Specificity = 1 – False Positive Rate

G-Mean = sqrt(Sensitivity * Specificity)

After searching through the value between 0.1 to 1 for the threshold, we found the best threshold=0.442130 and it gives the G-Mean the max value: 0.826. The best point is shown below in the figure.

Figure 13, tuning the classification threshold and finding the best point



7. Application of the test result

In the business world, we are sometimes more interested in possibilities of a company will be classified as bankruptcy rather than if it will be labeled in that way. Even if we have tuned the threshold to make it more reasonable, a company of 0.441 possibility will be labeled as not bankrupt and another company with 0.444 will be labeled as bankrupt. To provide a label of 0 or 1 is not helping the use of the model to assess what they should do with the results. To better understand this, let's take a look at how the auditors will process the information from this model.

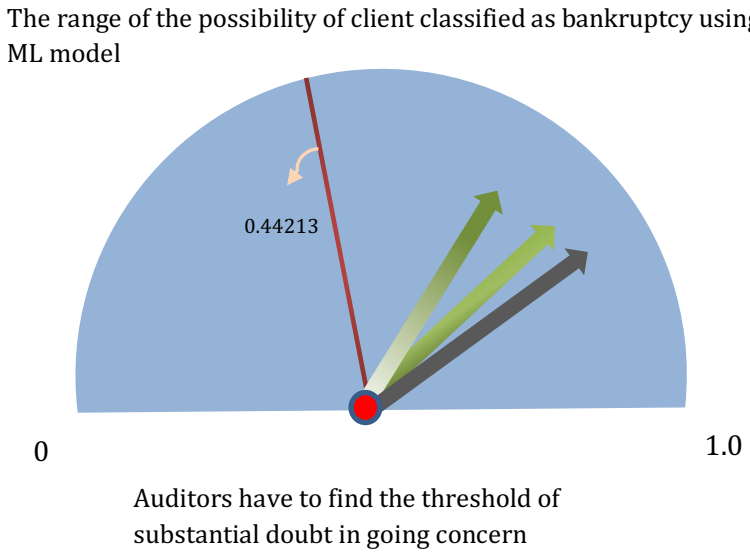
In FASB's standards, management is responsible for determining whether preparing the financial statements on a going concern basis is appropriate for the entity. Management needs to assess whether there is substantial doubt about the entity's ability to continue as a going concern for that 12-month period. Management then concludes whether preparation of the financial statements as a going concern is appropriate.

The auditor is required to consider the evaluation that has been performed by management and then to come to his or her own conclusion on whether the use of the going concern basis is appropriate for preparation of those financial statements. Another requirement is for the auditor to consider the adequacy and the appropriateness of the disclosures around the conditions and events relative to going concern. The first step of the evaluation, of course, is to consider, from the auditor's perspective, whether there are any conditions or events that cause or raise substantial doubt about the ability to continue as a going concern. Based on the judgement and management's plans for mitigating substantial doubt, auditors can make correspondent action from issuing an adverse opinion to including an emphasis-of-matter paragraph in the report.

Given the process of how auditors are assessing a company's going concern issue, the predicted probabilities may be more useful than a black-or-white crisp prediction of labels. The potential usage of this machine learning model can be shown as below. The management can use this model to evaluate the company's current situation and come up with a reasonable plan to mitigate the substantial doubt of going concern. Auditors, on the other hand, have to make their decision on all the information on going concern assumption. They can use this model to make a bankruptcy possibility prediction and have to figure out a threshold for possibility to be defined as substantial doubt. The tuned threshold of classification: 0.44213 can be a good reference.

One important thing to note, although both parties are able to use the machine learning model provided by this research, auditors always have to be independent and cannot rely on the model results provided by the management.

Figure 14, how the possibility prediction can be used by auditors



8. Assumptions and limitations

Based on the "No Free Lunch" theorem, states that there is no one model that works better than another without assumptions. There are some assumptions based on how the whole datasets were generated from the originally posted datasets on UCI Machine learning repository. The original dataset contains five separate files for companies filing for bankruptcy in 1 to 5 years from the financial ratios observed. The dataset I am using for this research, however, did not indicate how the data was randomly selected throughout the five files and combining the different years' data together can mislead the model.

Another assumption is that we assume the companies included are from the similar industries or have similar financial structures. The dataset does not provide industry information. However, different industries sectors can have different benchmarks for healthy financial ratios. Manufacturing companies, for example, normally have higher debt ratios because they are in capital-intensive industries. To make sure we can use all the data from the dataset, we have to assume that the financial ratios provided are comparable.

The biggest limitation from the data is that, as I have mentioned, there is no information provided in the financial ratios about cash flows, which are very important indicators of the going concern issue. There are some ratios provided that are not efficient indicators overall either.

9. Future work

The future research can be continued on the below aspects:

A. Adding more efficient financial indicators like cash flow numbers
B. Adding more information such as the management change, stock price fluctuation and industry information to make the research more thorough
C. tracking down companies' financial ratios for 3-5 consecutive years rather than just one set of financial data

There are many ways to improve the quality of the data so that future search can yield a better result.

10. Conclusion

In this research, we have used domain knowledge to understand the given dataset, used machine learning tools to select features, deal with data imbalance, train and test 7 models and found the best one with good performance: balanced bagging classifier model. We then adjusted the feature selection threshold and tuned classification threshold and boosted performance even more.

We have also explored the possible application of the model to the real business world where the positive possibilities will be a better reference than the crisp predicted labels for management and auditors to judge for threshold of the substantial doubt for going concern.

We have looked at the assumptions and limitations of this research and we are aware that by eliminating the limitations and adding more information in the future work, we will obtain a more solid and thorough result for bankruptcy prediction with our machine learning tools.

11. References

[1] Sai Surya Teja Maddikonda, Sree Keerthi Matta(2018): Bankruptcy prediction: mining the Polish bankruptcy data. Data mining techniques final project report.

[2] Vikram Devatha(2019): Predicting bankruptcy using Machine Learning.https://towardsdatascience.com/predicting-bankruptcy-f4611afe8d2c

[3] Joanna Wyrobek:Predicting Bankruptcy at Polish Companies: A Comparison of Selected Machine Learning and Deep Learning Algorithms.e-ISSN 2545-3238

[4] Wenhao Zhang(2017):Machine Learning Approaches to Predicting Company Bankruptcy.Journal of Financial Risk Management Vol.06 No.04(2017), Article ID:81016,11 pages

[5]Imbalanced Classification on Bankruptcy Prediction.https://rstudio-pubs-static.s3.amazonaws.com/336831_55c98ceeb234439b80d844da949ff1f4.html#data-set

[6] Jacky C. K. Chow(2017):ANALYSIS OF FINANCIAL CREDIT RISK USING MACHINE LEARNING

[7] Pawełek, Barbara (2019) : EXTREME GRADIENT BOOSTING METHOD IN THE PREDICTION OF COMPANY BANKRUPTCY, Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, NY, Vol. 20, Iss. 2, pp. 155-171, http://dx.doi.org/10.21307/stattrans-2019-020

[8] Bob Dohrer, Ken Tysiac(2020): Going concern tips for auditors during the pandemic. The journal of accountancy. April 3, 2020