Forecasting Global Stock Market Crisis

Our group is divided into three sub-teams according to different division of work: data acquisition group (3035675373 Zou Linzuo, 3035675074 Liang Yunshu), data processing group (3035675713 Wu Junhao, 3035675725 Zeng Ni), modeling group (3035675153 Li Jiatao, 3035675969 Wang Jingting, 3035674733 Wang Wanliu). Everyone in the group participated in writing the report.

Section 1: Executive Summary

The main goal of our project is to develop a solid forecasting mechanism for predicting a possible global stock market crisis, which performs as a global systemic early warning tool. Specifically, pricing patterns in liquid markets are interpreted to identify local or/and regional market shocks, which may evoke a worldwide economic crisis. We collect daily financial data of 39 countries from stock, bond and currency markets where the effect of extreme events can be better depicted. Through a set of machine learning algorithms for selecting the most relevant variables out of proposed ones, we figure out some important features for predicting the stock markets crisis such as the continuously compounded rate of MSCI stock index in England and that of 10-year bonds in China. We also find some significant evidence of interdependence and cross-contagion effects among stock, bond and currency markets based on covariance analysis on selected features. In the meantime, we employ oversampling methods for adjusting the imbalanced nature of the stock crisis. Logistic regression, Decision Tree, Random Forest, XGBoost, and two different neural network models are used to predict the probability of Global crisis or classify whether there will be a global crisis the next day or the next 20 days. Among all the models, XGBoost shows the highest accuracy both predicting crisis on the next day and the next 20 days due to its meticulous methods of handling overfitting. By verifying that the time frame of the stock market crash and the time frame of the major event are roughly fixed, we can create a reliable warning system. Therefore, this valuable tool can be utilized in banking, business, and other areas. Central banks over the world may use this system to early adjust their monetary policy in case of any possible financial crisis, and ensure financial stability.

Section 2: Introduction

At the start of the year, most investors expected the 11-year bull market to continue in 2020, only to be shockingly disabused of that notion by the spread of COVID-19. As a result, the Dow fell from record highs to bear-market territory in a matter of weeks. Not only the investors, but also the central bank need a way to price in risk and predict the status of the stock market in terms of so many possible significant events. Our intention is to create a warning tool for stock markets as soon as a major event is identified. The classical approaches for predicting stock crisis are using Bayesian model averaging for identifying financial figures of crises (Babecký et al. 2014) and linear models such as linear regression (De Haan 2017). However, the normal methods cannot capture dynamics figures and predict the status of the stock market in an out-of-sample setup.

Particularly, The reliability of predicting performance of early warning systems decreases substantially when attempting to figure out a rough date frame of the event.

Our project offers a multi-component modelling framework for global crisis events forecasting, driven by the merits of a series of techniques including Classification Trees, Support Vector Machines, Random Forests, Neural Networks, Extreme Gradient Boosting, and Deep Neural Networks. (SP. chatzis 2018)

The remarkable improvement is

- We combine multiple machine learning techniques by leveraging ensemble methods in order to minimize model risk and increase accuracy. (SP. chatzis 2018)
- We use daily financial market data, which tend to be more responsive than macro-data, since they are reported in higher frequencies. (SP. chatzis 2018)
- Compared with traditional methods of predicting stock crash, we use advanced machine learning techniques such as Deep Learning and Extreme Gradient Boosting which overcome problems such as the vanishing gradients problem, overfitting tendencies and also offer higher flexibility in learning nonlinear dynamics in large datasets.

Section 3: Data

We collect data of stock market indices, 10-year treasury bonds, exchange rates between main currency pairs and common macroeconomic indices of countries in Asia, America, Europe and Australia from several databases such as Investing, Wind, Yahoo Finance and Federal Reserve Economic Data. Web crawlers and databases are used in this process. We choose MSCI indices of different countries instead of local indices to set a relatively stable basis for judgement. However, MSCI indices in some countries such as China were absent until 2005 and thus are replaced by local indices. Troubles also occur in 10-year treasury bonds since prices of some species are not updated every day, but cubic spline Interpolation helps to fix it. Besides, by drawing the graph of the inverse covariance matrix, it is quite clear that some emerging markets have less connection with other markets, and we remove some of them seems to be acceptable.

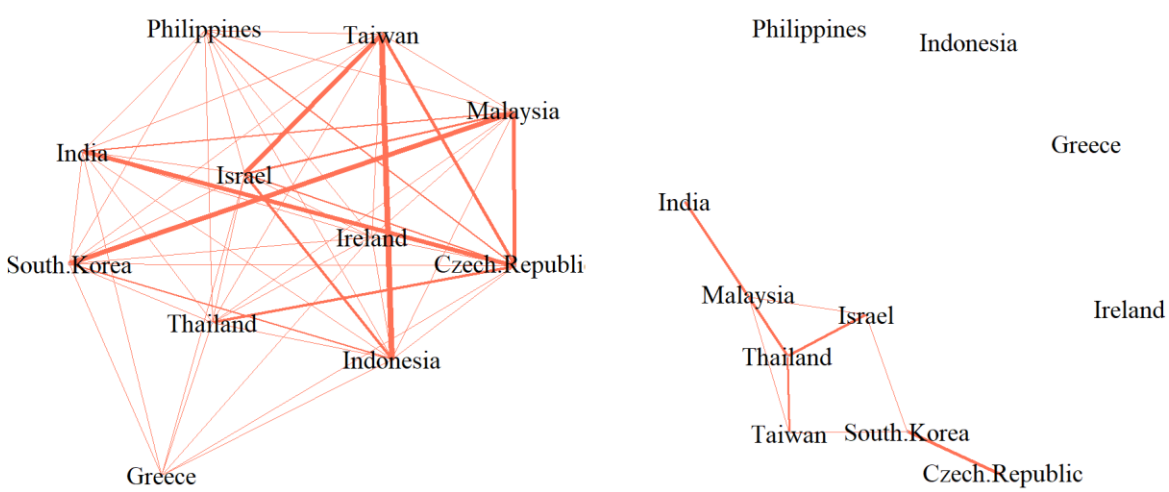

Fig 1. the inverse covariance matrix.

Section 4: Analysis

A. Variable exploitation

All the collected data must be aligned according to the date in the same length of time and the same quantity. Therefore, we use the observations covering the period 04/17/2002-03/18/2020.

a. Changes. Based on the stock index, bond, currency and etc. data of 39 different countries, we calculated the daily continuously compounded return rate as the independent variables.

$$\ln return = [\ln(y_t) - \ln(y_{t-1})]$$

b. Volatility. Next we calculated the volatility of return rates which were calculated on the previous part, simply by squaring the log return.

$$\text{The daily volatility of return} = (\ln return)^2$$

c. Identifying "crisis event", "Single Crisis event" was identified if the return of the variables under the categories of stock indices, bond yields, currency exchange rates and additional variables was below the 0.6 percentile of the associated empirical distribution of returns for selected countries at each working day. We employ a percentile-based mechanistic method of crisis detection to avoid the ad-hoc selection of crisis episodes and connect them to high volatility levels (SP. chatzis 2018).

The initial empirical distribution of returns were calculated based on the returns of the first 500 observations (covering the period 17/04/2002–03/12/2004). Then, for each subsequent record (e.g. day in the initial empirical distribution), we recalculated the following empirical distribution of the subsequent 500 observations in order to incorporate the new observation.(e.g. For 18/04/2002, the empirical distribution is 18/04/2002–04/12/2004. For 19/04/2002, the empirical distribution is 19/04/2002–05/12/2004.) Then, an event was identified if the return was below the 0.6 percentile of these empirical distributions. The reasons for choosing 0.6 percentile as a distinction point are A. Reducing the sensitivity of the warning system where only the effect of extreme events would be recorded B. Minimizing model risk and increasing accuracy.

For independent variables including bond yields, currency exchange rates and additional other variables, we only need to calculate the number of crises per working day in terms of each variable, and then we calculate the aggregate number of events on the global scale.

For the dependent variable involving stock index, we not only identify extreme movement events at country level, but also summarize the aggregate number of events for each day on a regional scale(America, Europe and Asia) and subsequently on the global scale. Then we derived a set of binary predictor variables, identifying whether the number of events per day exceeded a threshold for separate regions in order to derive the status of stock markets on both regional scale and global scale. (e.g. If the number of events exceeded the given threshold, we define it as stock crisis)

The threshold is determined by the following process:

Country	Threshold
America	At least 3 events per day out of a total of 7 countries.
Asia	At least 6 events per day out of a total of 13 countries.
Europe	At least 8 events per day out of a total of 19 countries.
Global	At least 2 regions are undergoing stock crises on a daily basis.

Tab 1. summary of the regional and global threshold.

d. Creating dependent variables. 'The status of stock markets in a 1-day predictive horizon' and 'The status of stock markets in the 20-days predictive horizon' were the two main binary dependent variables. These dependent variables were created as indicators of stock market crisis on the basis of the following hierarchical process, which was described in Fig 2.

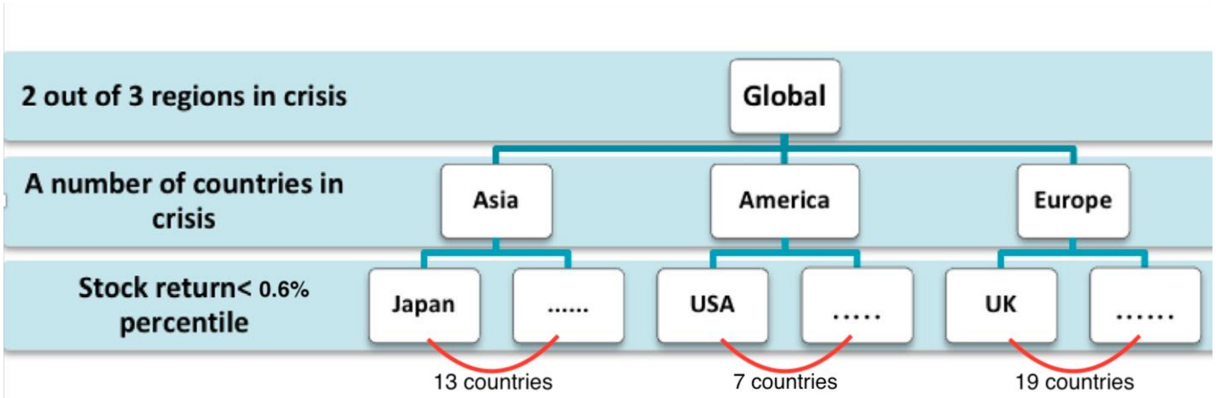

Fig 2. Outline of the dependent variable construction process.

Based on the dependent variable construction process, we summarized the time frame of the stock crisis during the period from 18/04/2002 to 18/03/2020. Figure 2 demonstrates the global stock crisis on a daily basis. Generally the time frame when stock crises happened is highly in line with the time period when signature events happened. (i.e the financial crisis in 2008 led to significant stock markets crash around the world)

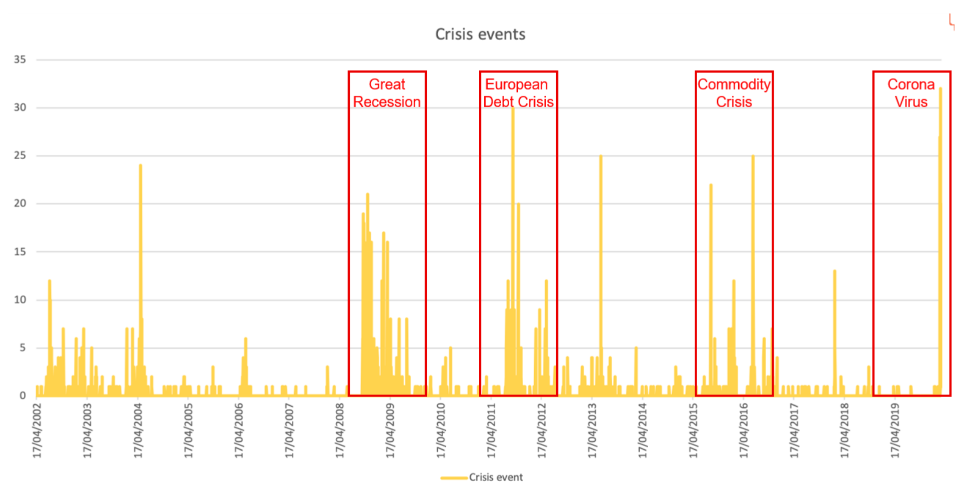

Fig. 3 Number of extreme stock markets fall over the world(exceedance less than 0.6% percentile of the empirical distribution).

e. Lag variables. Besides those above country-specific variables, we also explored whether the variables in a specific country or region have any statistical correlation with observed crises in the past. To this end, we created additional exploratory variables that constitute lagged values of the extracted predictive indicators. These variables can be considered systemic shock indicators, inclusion of which may potentially facilitate capturing contagion effects among different regions, as well as of crisis clustering effects. Therefore, in an effort to capture subtler dynamics and dependencies, we consecutively computed the following transforms. First, we calculated the lagged variables on a daily basis for variables including return rate, volatility and the identified crisis, in 2 days horizon, 3 days horizon... 5 days horizon, 20 days horizon, 40 days horizon, 60 days horizon (named as crr_lag2, crr_lag3… crr_lag60). Those figures were also chosen as independent variables to indicate the trend of return in a relatively long term period. Second, We computed the average number of significant events for the stocks market during the last 5 working days and the last 20 days, based on previous values of the binary predictor variable which identifies whether the number of events exceeded a threshold at some day. Then, we also computed the average number of events during the last 5 working days and the last 20 days, based on the total number of events on a daily basis for all the variables under the categories of stock indices, bond yields, currency exchange rates and additional variables.

Based on the above outcomes, we created a set of 2284 predictors and two classes of target variables to forecast covering an adjusted time period 07/10/2002-03/18/2020:

- The first measures whether there is a significant global crisis event in the next working day,
- The second measures whether there is a significant global crisis event during the next 20 working days .

Stock Variables	**Currency and Bond Markets**	**Other Variables**
Sample countries stock market returns and volatilities Binary variables of extreme movement events in country level Discrete variables counting number of extreme events on a day globaly and by region(e.g. Asia) Binary variables representing regional or past global events on different horizons (5,20, 40 and 60 days).	10 year sovereign bound yield changes. Changes in basic exchange rates Volatilities of sovereign bond yield and basic currencies Binary variables of extreme movement events in country level for sovereign bonds and currencies Discrete variables counting number of extreme movements (>94% percentile) on a day globaly	Libor rate VIX index Gold Price TED spread Oil price Effective Federal Funds Rate High yield bond returns

Tab. 2 Summary of the raw measurements for variable exploritation

B. Variable selection

After the variable exploitation, we constructed a whole dataset which comprises an excessive number of independent variables. One remarkable character of our data set is that the number of independent variables is disproportional to the size of the dataset, i.e. the observations in the sample; specifically, we are dealing with around 2285 variables over around 4676 days (observations). Fitting a machine learning model to such a high dimensional data (relative to the size of the dataset) would be highly possible to undergo dimensionality problems. Specifically, the fitted classifier may seem to yield a perfect performance in the training dataset, but it turns out to generalize a very poor performance outcome in the test data. Thus, it is necessary to implement a robust feature selection process in order to avoid dimensionality problems. Besides, the reduction on dimensionality of data would increase the computational efficiency for the following machine learning algorithms.

The feature selection process mainly comprises three phases:

In the first step, we employ three different methodologies that independently assign importance to the available features: Boruta (Kursa & Rudnicki 2010), LASSO (Tibshirani 1996), and a normal variable importance method. In the second step, we calculated the importance score for every feature based on previous methods. We assign relative high scores for the feature selected by Boruta due to the extensive analysis of the features in the dataset (0.6 score for every feature selected by Boruta, 0.2 score for features selected by Lasso and normal methods). In the third step, we rank the importance score and choose the features which are higher than 0.4. Generally, We eventually selected 64 explanatory variables.

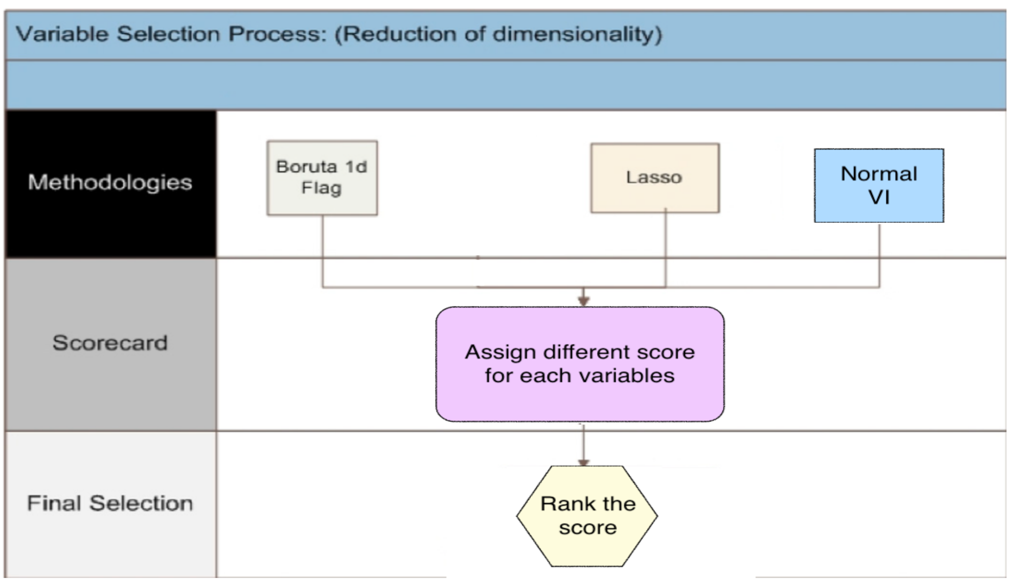

Fig. 4 Variable scoring and selection process.

C. Fixing dataset imbalances

Since the distinction point (0.6%) is relatively small compared to volatility of the stock index, we can only figure out a very small fraction of the binary dependent variables whose values equal to one (indicating a crash event). This is especially true when it comes to the dependent variable which measures whether there is a significant global crisis event in the next working day(only around 1% of the available sample corresponds to crash events). It can discard useful information about the data itself which could be necessary for building rule-based classifiers such as Random Forests. To fix this problem, we employed the ROSE (Random over-Sampling Examples) package in the R language. ROSE allows for generating balanced artificial datasets, by leveraging sampling methods.(SP. chatzis 2018)

D. Model description

Here we didn't choose the rolling basis instead of cross validation, because the crisis happens so unusual that the accuracy would be very low, since there's only a few Global Crisis data in our data set. We balance the data which causes the disturbance of its time series property. This is also the reason why we didn't choose the RNN model.

a. Logistic regression

First, we build logistic regression with l2 penalty to select features. KFold(5-folds) is used to train models and make average prediction to reduce overfitting. This simple model has no hyperparameter needs to be tuned.

b. Decision Tree

Second, decision tree(CART) is built, with gini as the splitting criterion. For this model, maximum depth and minimum impurity decrease are tuned to reduce overfitting.

c. Random Forest

Third, random forest is trained, with gini as the splitting criterion. We tune maximum depth and number of estimators to control overfitting and the behavior of the model.

d. XGBoost

Then, we build eXtreme Gradient Boosting with log loss metric. Hyperparameters like learning rate, maximum depth of each tree, gamma, subsample are tuned by hand to improve the behavior of the model.

e. Neural Network

Here we build fully connected Neural Network models of three hidden layers. K-fold is utilized to choose optimal parameters: numbers of cells of each hidden layer, activate function, and learning rate. To avoid overfitting problems, we add weight decay and dropout parts and only use the 68 features which were selected before. Two output cells are built for classifying the Global Crisis, and a softmax function is used so that the outputs are in range 0-1, representing the probability of having a crisis or not.

f. Deep Neural network

To conduct Deep Neural networks, we built a 5-layer perceptron. It is composed of 4 hidden layers. Hyperparameters are selected using 5-fold cross-validation with the purpose of minimizing CrossEntropy loss. Relu activation function is adopted and we also use regularization methods, such as weight decay and dropout to reduce the model' s overfitting tendencies. Also, it is worth mentioning that the trained model for 1 day prediction is more complicated than that for 20 days in terms of dropout times and BatchNorm times. Main reason is that the number of 1 value in the training data for one-day prediction is much smaller than that for 20-day prediction data.

E. Model result

We employ three measurements to evaluate the performance of six different models we built: ROC curve, Precision, and Recall.

a. ROC

Here we plot the ROC curve (receiver operating characteristic curve) and also show the AUC(area under curve), which can objectively reflect the comprehensive prediction ability of positive and negative samples, and eliminate the influence of sample skew. Class imbalance often occurs in the actual data set, that is, there are many more negative samples than positive samples: there are so few Global Crisis -- less than 1%. When the distribution of positive and negative samples in the test set changes, the ROC curve can remain unchanged. How to determine the result according to ROC: which classifier is closer to the upper left corner. At the same time, according to roc, we can determine where the probability boundary of positive samples is suitable.

i. Global Crisis in the next day

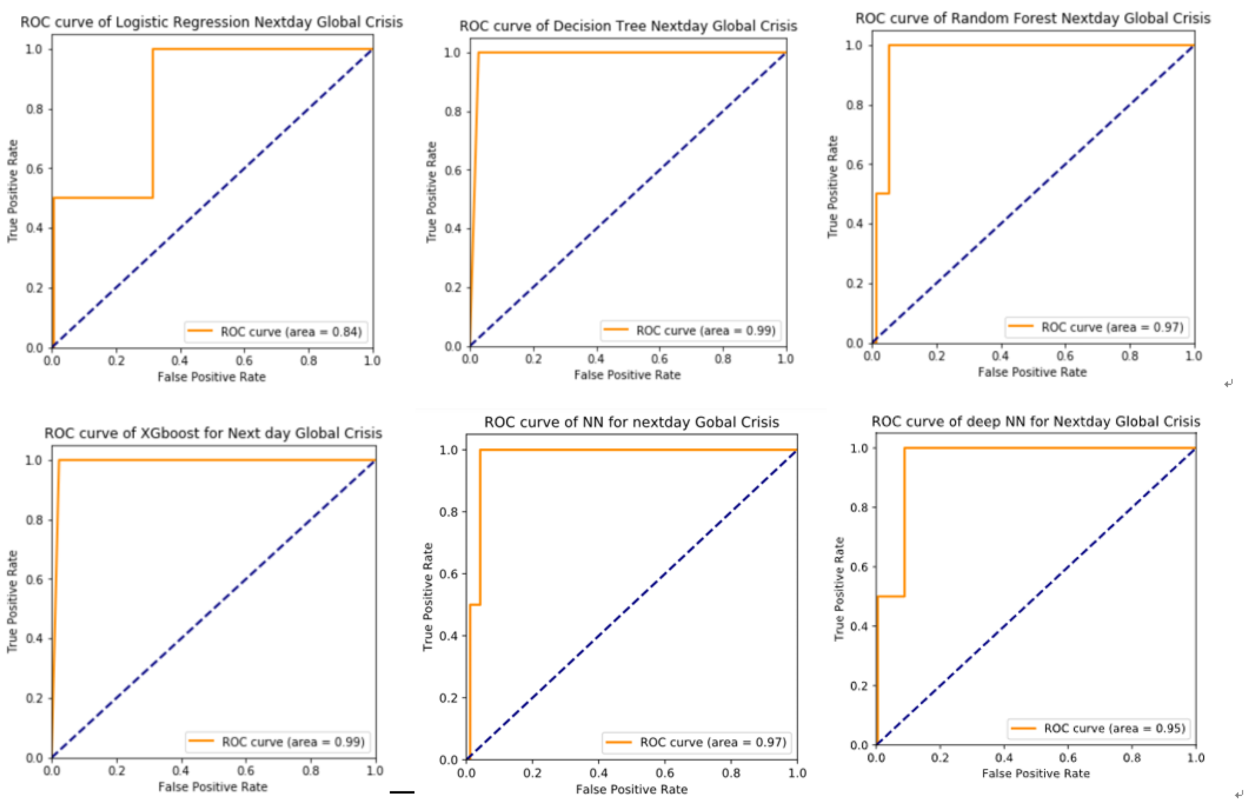

Fig. 5 ROC curves for 'global crisis in the next day'.

From the ROC curve we plot above we can assert that Decision Tree and XGboost curves are closer to the upper left corner, representing better performance than others when predicting Global Crisis happened the next day. While, the Logistic Regression and DNN have the worst performance.

ii. Global Crisis in the next 20 days

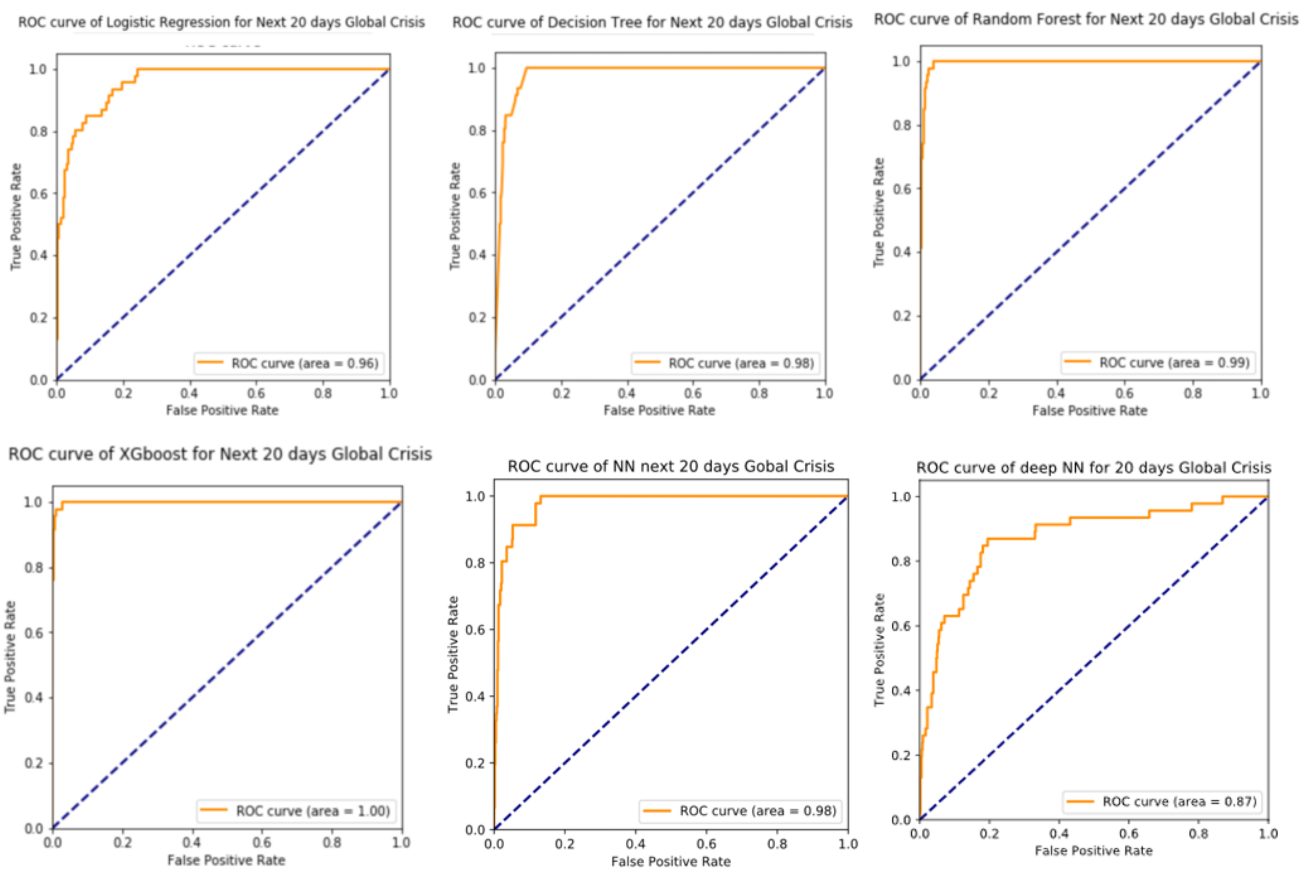

Fig. 6 ROC curves for ‘global crisis in the next 20 days’.

From the ROC curve we plot above we can assert that Random Forest and XGboost curves are closer to the upper left corner, representing better performance than others when predicting Global Crisis happened in the next 20 days. While, the Logistic Regression and DNN have the worst performance.

b. Precision and Recall

For the sake of evaluating the result of different models, we here use two values: precision--at the result of predicted to be a crisis, how many percent of them are true, and recall-- when there is a crisis whether the model can actually predict it.

	Nextday Global Crisis		Next 20 days Global Crisis	
Models	precision	recall	precision	recall
Logistic Regression	3.23%	50.00%	41.30%	82.61%
Decision Tree	9.52%	100.00%	42.20%	100.00%
Random Forest	9.09%	50.00%	62.16%	100.00%
XGboost	9.09%	50.00%	86.54%	97.83%
Neural Network	0.59%	100.00%	13.73%	100.00%
Neural Network with adjusted threshold	3.13%	100.00%	24.19%	97.83%
Deep Neural Network	0.78%	100.00%	13.61%	100.00%
Deep Neural Network with adjusted threshold	1.29%	100.00%	17.69%	100.00%

Tab. 3 Summary of precision and recall

In general, predicting Global Crisis in the next 20 days has both higher precision and recall, which might be caused by the skewed data on the test set when predicting a crisis in a day: only 2 days have a stock crisis the next day. When predicting next day crisis Decision Tree has a better performance in general, while the XGboost is better when predicting the next 20 days Global Crisis which is basically corresponding to the ROC findings.

The two neural networks models tend to predict a crisis too often, but seldomly lose a crisis. Adjusting the threshold and raising the standard of a crisis could lead to a better performance.

Section 5: Conclusion and future work

A. Conclusion

Based on the predicted results, we found that it is not easy to predict the global crisis that happens the next day because the precision is less than 10% in most of our models. Though decision trees get a 100% recall, it is not convincing because there are only two positive labels that can not provide an efficient test set. And too few positive training samples also account for poor behaviour of all the models. Predicting global crisis in the next 20 days is relatively easy. On the one hand, there is relatively more training data for predicting 'global.crisis.20''. On the other hand, 20 days is longer than one day, so that it is much easier to predict crises that occur in 20 days than on a specific day.

Concluded from the models, we notice that XGBoost behaves the best. With the limitation of 'global.crisis.1' prediction, we only comment on the results of 'global.crisis.20'. More specifically, logistic regression makes a normal prediction which is not too bad. Decision tree successfully predicted all the crisis but with relatively low precision(less than 50%). Random forest helps improve precision from 42% to 62% without sacrificing any recall. Neural networks behave relatively poorly because this kind of machine learning algorithm asks for a larger amount of data. Our final model, XGBoost, dramatically improved precision to 86.5% while only sacrificing 2.2% recall. With this model, we can efficiently predict most of the global crisis with confident precision.

B. Limitation and Future Work

Too few positive samples result in poor prediction on the problem of Global.Crisis.1. But it is not an issue that can be solved in the future since the global crisis would not happen frequently.

Although the combined machine learning algorithms for selecting important features does reduce the dimension of the data, we are supposed to employ covariance analysis and shrinkage to avoid strict multicollinearity. We can improve it through linearly combining the independent variables, such as adding them together or perform an analysis designed for highly correlated variables, such as principal components analysis, binned aggregation and elastic net.

a. The empirical distribution for indicating the stock crisis tends to ignore some persistent but not significant crisis. For example, the subprime mortgage crisis in 2008 leads to a decline of stock markets which lasted more than a year. In such cases, the empirical distribution is not capable of sorting those events out, it is only able to sort out events when there is great volatility in the stock markets.

b. To create a solid warning system for the stock markets, it is necessary to take macro economical figures into account. (i.e. the trend of the unemployment rate during the crisis event, GDP for 39 countries and etc.) Only combine both the macro economical and micro economical figures can the accuracy of the warning system be improved.

c. The missing value in the bonds and additional variable leads to some complicated issues related to generating the variables and predicting results. It is not suitable to use imputation methods to fill the missing value since the historical financial data must be as precise as possible. In the future work for this project, we aim to collect a more complete data set in order to avoid the missing value issues.

Reference

1. Sotirios P. Chatzis, Vassilis Siakoulis, Anastasios Petropoulos (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Systems With Applications, 112 (2018) 353–371.
2. https://fred.stlouisfed.org/
3. https://www.investing.com/
4. https://finance.yahoo.com/