Results

Prediction

Combining the models and the special consideration to the imbalanced data, I fit five sets of models to the data: logistic regression, using prevalence as the decision threshold (logit-P); logistic regression, using re-weighted data (logit-RW); RF, using CV-tuned decision threshold (RF-CV); RF, using re-weighted data (RF-RW); RNN + BiLSTM, with a focal loss function (RNN-BiLSTM-focal). Their performances are shown in Table~\ref{tab:res}.

The RF model with a CV-tuned decision threshold performs the best, having a balanced accuracy of 93\%. The sensitivity and specificity show that the model correctly identifies 100\% of the bubble cases and 86\% of the non-bubble cases. Most other models perform decently well, with balanced accuracy rates around 70\%. The RNN model is a disappointment. It performs no better than a naïve model. I will attempt to explain the predictive power of the models in the discussion section.

Figure~\ref{fig:cv\_threshold} and Figure~\ref{fig:rfcv\_full} provide more information about the RF-CV model. Figure~\ref{fig:cv\_threshold} shows the process of determining the decision threshold. Cross-validation chooses the decision threshold of 0.12 since it yields the lowest balanced error in the CV data. Figure~\ref{fig:rfcv\_full } shows the prediction results (red area) of the RF-CV model when imposed on the complete data set. As we can see, the RF-CV model correctly classifies almost all cases.

When reading in the most recent data when this research is written (April 2021), the RF-CV model gives the prediction of “bubble.” It suggests that we are likely in a bubble, and financial crashes are likely in the next six months. This result is not financial advice, and I do not recommend the reader to make risky financial decisions based on this result. If the reader still wants to take this result as a reference, I highly recommend the reader to read the discussion section for the limitation of this research before making any decisions.

Inference

There are some further insights that we can draw from the research besides the prediction results. We can identify important factors for prediction. I use two of the best performing models for inference: the RF-CV model and the logit-RW model.

Figure~\ref{fig:varImp} shows the variable importance in the RF-CV model. We see that consumer confidence, long-term investment returns, e.g., 5-year S\&P returns, and fundamental indicators, e.g., market capitalization-to-GDP ratio, are the most important. On the other hand, the short-term S\&P 500 return and short-term macroeconomics data are among the least important features.

Table~\ref{tab:logit} shows the regression table for the logit-RW model. It mostly agrees with the results of the RF-CV variable importance plot. The fundamental indicators and long-term investment returns are significant at the 99.9\% confidence level. The 3- and 6- month market returns are not significant. However, the logit-RW model finds inflation and one month market return significant for prediction.

The inference favours the investment philosophy held by value investors, at least in the context of avoiding market crashes. If the stock prices are too high for corporate earnings and national production, it would be wise to heed the risks of market crashes. Also, this inference explains why the machine learning models from previous studies do not perform well. As mentioned in the background section, most of the previous machine learning studies for predicting financial crashes only include short-term trading data as features. Those features have the lowest predictive power among all the input variables I select.