Default Prediction with Knockout Model

CEU Capstone project Summary

Áron Hartvig

Supervisor: Zoltán Papp

Contents

I. Description of the project

Company: Start-up that provides liquidity for SMEs. The company's vision is to enable SMEs to increase their liquidity at the ease of clicking a button, at low cost, full transparency, completely online.

Key problem: The (credit and fraud) risk model developed so far needs further improvements. The objective of the capstone project is to further develop knockout (KO) criteria model by producing research and developing a tool for KO criteria testing.

II. Knockout model

The company is exposed to default risk. The liquidity provider may suffer losses if the debtor goes bankrupt. Furthermore, the start-up has to filter fraud cases as well. Therefore, it monitors closely its client's and their partners.

The company is filtering the risky companies in two ways. The first line of defence is the knockout model, and the second is a probability of default model. The knockout (KO) classification model predicts default based on several criteria. Originally the KO model applied thresholds based on expert opinion for 17 financial metrics. If the corresponding value of the company exceeds (or below) the threshold level the model predicts default for that company. Although individual thresholds for the metrics may not be the most accurate way of classification, interpretability is a key aspect of the KO model. The analyst must understand and record why the model predicts default for a client.

III. Area Under the Curve

To measure the performance of a classification method the most commonly used tool is the receiver operating characteristic (ROC) curve. The curve plots the true positive and false positive rates for all classification thresholds. Still, to describe the model with one value we can calculate the area under the curve (AUC) that is the entire two-dimensional area underneath the entire ROC curve. Without building any model, just using the aprior probabilities of the categories the AUC value is 0.5, consequently any model that has lower AUC than 0.5 has worse predictive power than applying simple probabilities.

III.1 Threshold-based model

The first step to enhance the model was to estimate new thresholds with data-driven method. I built a Python model that calculates cross-validated thresholds for the metrics and maximizes the AUC value of the prediction. Then, as the company applies clustering on the companies (6 clusters), I used these clusters to separate the data and recalculate the thresholds for each cluster. Finally, I combined the metrics, so the model predicts default for a company if its financial metrics exceeds all the thresholds. To compare the original model to the new one I calculated the AUC values of the models. In general the performance of the KO model enhanced a lot, but the predictive power of the new model largely differs between the clusters.

III.2 Random Forest model

As a next step I used Random Forest, a more sophisticated model, to predict default. The previously mentioned KO model applying thresholds is basically a simple decision tree. Random Forest is an ensemble learning method for classification that combines bagging and random feature selection.

After finetuning the Random Forest model achieved similar AUC value to the previous model's performance on average. Still, the lack of interpretability is a serious disadvantage of the Random Forest model. Analysts frequently check the importance of the features in the Random Forest model to see which ones have the greatest effect on the dependent variable in general. On Figure 1 I present the feature importance.

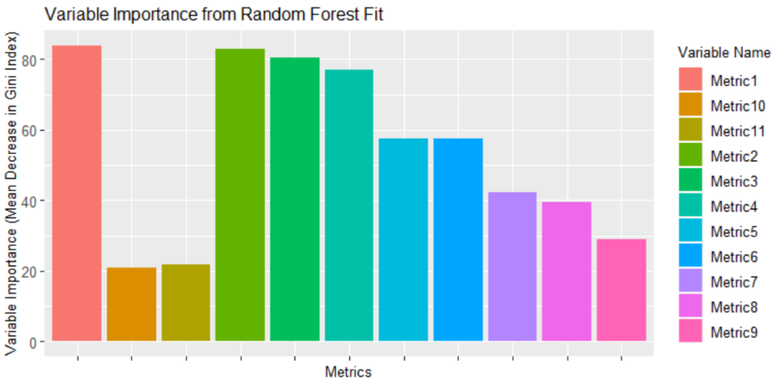

Figure 1 Random Forest feature importance

From the figure we can conclude that the most important features are Metric1 and Metric2. Still, feature importance only represents the overall impact of a given feature, it does not provide any information why a specific company was predicted default.

III.2.1 Shapley

Machine Learning models have the disadvantage that they are hard to interpret. Consequently, a main aspect of the KO model is violated. To overcome this drawback of the model, I use the Shapley method to make the results more interpretable. Shapley value is the average marginal contribution of a feature value across all possible coalitions. From the Shapley values we can also calculate the SHAP (SHapley Additive exPlanations) values that explain individual predictions. On Figure 2 I plotted the SHAP values for all the observations in the test dataset. The greatest advantage of SHAP that it can be calculated for a specific company and the analyst can see which features dominated the prediction.

To conclude, by applying Shapley method to the Random Forest model the Random Forest model is also able to provide accurate and interpretable result. Although the AUC values are not very high in neither of the models, still further improvements can be done, like combining the threshold-based and the Random Forest model.

Both models were optimized to maximize the AUC values. AUC, however the most often used measure for classification performance, has a serious drawback that questions whether it is the best choice for the KO model.

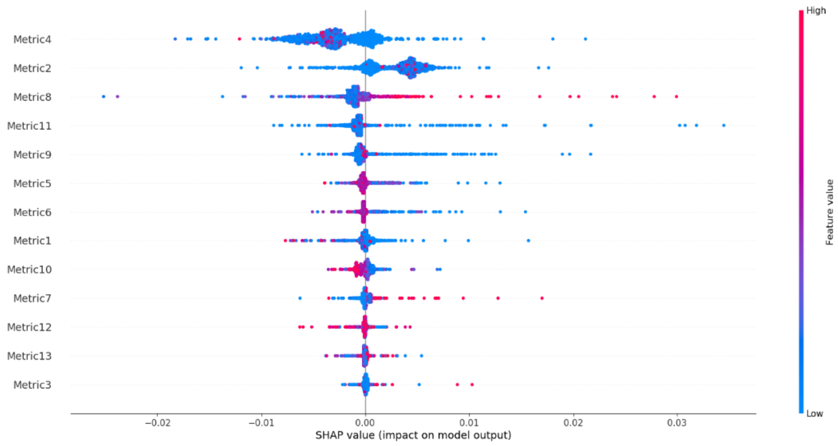

Figure 2 SHAP values of Random Forest model

AUC is classification-threshold invariant that sometimes makes AUC appealing, but in cases where there are wide disparities in the cost of false negatives versus false positives, it may be critical to minimize one type of classification error. AUC is not a useful metric for this type of optimization. The cost of false negatives and false positives are not equal, since false positives leads to revenue loss, while false negatives can be filtered by the probability of default model. An appropriate alternative of AUC for this case is the H-measure.

IV. H-measure

In the literature H-measure becomes more and more accepted as an alternative of AUC. It is applicable to cases where the cost of false negatives and false positive are not equal. It naturally requires a severity ratio, which represents how much more severe misclassifying a non-default instance is than misclassifying a default instance.

I calculated the H-measure and AUC values for Metric1 with several default thresholds based on the corresponding percentile values and the difference between the two measures is clear, H-measure suggests a stricter threshold. As the cost of false positive predictions are set higher than the false negatives the model only predicts default in cases where there is high probability of default.

V. Lessons learnt

The cooperation with the client was a great experience. We had weekly meetings which were extremely useful as I got almost instant feedback on my work. At the beginning some major tasks were outlined to be done, but the client was flexible and as the project evolved, they were open to my suggestions. In my opinion this was essential, since research had come to a dead end many times and without the will to experiment the value of the project would have been much lower.

Furthermore, the capstone project enabled me to experience the unique environment of start-ups. I had the chance to see how the company works, the tools it has, and the impact a good project can make. I felt my work valuable and the client also helped me in that. To conclude, I learned a lot about process management, risk management and data science as well.