S&P Global
Market Intelligence

Authors:

Luka Vidovic & Lei Yue
Analytical Development Group
S&P Global Market Intelligence

Machine Learning and Credit Risk Modelling

Machine Learning (ML) algorithms leverage large datasets to determine patterns and construct meaningful recommendations. Likewise, credit risk modelling is a field with access to a large amount of diverse data where ML can be deployed to add analytical value. In the following analysis, we explore how various ML techniques can be used for assessing probability of default (PD) and compare their performance in a real-world setting.

Machine Learning in Finance

A recent publication by the Bank of England (BoE) and the Financial Conduct Authority (FCA) reports the results of a survey on the use of ML in United Kingdom (UK) financial services.[1] Results show that two-thirds of respondents use ML in some form. The use cases have passed the development stage and are starting to enter into the deployment stage. The banking and insurance sectors are advanced with respect to deployment, and ML is most often used in anti-money laundering and fraud-detection applications. The survey also notes that ML may amplify existing model risk, while validation frameworks still need to evolve to cope with the complexity of ML applications.

As ML is becoming more represented and influential in finance, it is important to recognize its benefits and drawbacks to prudently evaluate its performance. ML models have the potential to uncover subtle relationships, capture various nonlinearities, and process unstructured data. For example, applications such as fraud-detection analysis or textual data analytics benefit from not needing to predefine structure, that is, the theory behind finding patterns and extracting meaningful outputs. ML can do this without the need for humans to derive theoretical models with accompanied assumptions, and the data is empirically driving the ML model.

[1] Bank of England, Financial Conduct Authority: "Machine learning in UK financial services", October 2019.

However, ML may still contain assumptions, such as the dataset does not contain. This can pose a significant challenge when analyzing noisy historical financial data and may lead to poor model performance. Imposing constraints on the model to control for model biases or counterintuitive behavior can also be an onerous task for some ML techniques. In addition, decomposing ML models can be complicated, thus creating issues when there is a need to explain the model's functionality in detail.[2] [3] [4] [5]

Background

We analyze the performance of selected ML algorithms for the prediction of PD. To make this analysis relevant and material, we use a real-world example of constructing a default prediction model for private companies. To that end, we collected a global sample of private companies across various industries.[6] Private companies are a particularly relevant example for our analysis for a number of reasons. The universe of private companies is large and highly heterogeneous, as it includes large international corporations, as well as local small- and medium-sized enterprises. The composition of a global sample captures companies from various macroeconomic environments, thus introducing additional macroeconomic risk components. Additionally, private companies tend to publish limited and infrequent financial disclosures, which reduces the scope of available information.

The characteristics of private companies create a need for a default prediction model to be well designed in order to capture the heterogeneity of private companies and achieve good performance under the data availability constraints. We leverage the S&P Capital IQ platform to collect annual financials for private companies globally from 2002 to 2016. Our final sample includes a total of 52,500 observations, of which 8,200 companies have defaulted.

Feature Engineering: We 'pre-treat' the financial data by calculating relevant financial ratios to express various risk dimensions, such as profitability, leverage, and efficiency. We also include a Country Risk Score (CRS) and Industry Risk Score (IRS) as additional variables to help the model capture systemic risk components of various countries and industries. We also standardize the ratios to make them comparable and limit the impact of outliers, thus enabling the algorithms to achieve better performance.

Variable Selection: To account for the limited availability of private company financial data, we only use ratios that have sufficiently good coverage across the S&P Capital IQ platform, while also ensuring the representation of relevant risk dimensions. Such parsimonious construction simplifies the use of the model in deployment, as it requires fewer inputs and less data handling, and increases the model coverage. This is especially important for private companies, where

[2] Bazarbash, M.: "Fintech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk", IMF Working Paper, 2019.
[3] Bracke, P., Datta A., Jung C. and Sen S.: "Machine learning explainability in finance: an application to default risk analysis", Staff Working Paper No. 816, Bank of England, 2019.
[4] Rasekhschaffe, C. K. and Jones, C. R.: "Machine Learning for Stock Selection", Financial Analysts Journal, 2019.
[5] Addo, M. P., Guegan, D., Hassani, B.: "Credit Risk Analysis Using Machine and Deep Learning Models", Risks, 2018.
[6] Financial sector is excluded from the analysis.

financial data is generally more infrequent and less comprehensive. Table 1 contains the final list of selected variables used to train the PD model with various ML algorithms.

Table 1: List of variables used to train PD models for private companies

Risk Type	Variable	Risk Dimension
Financial Risk	Total Equity / Total Assets	*Capital Structure*
Financial Risk	Current Liabilities/Net Worth	*Short-term Leverage*
Financial Risk	Return on Net Capital	*Profitability*
Financial Risk	Cash & Short-term Investments / Total Assets	*Liquidity*
Financial Risk	Net Income/Total Liabilities	*Debt Service Capacity*
Business Risk	Total Revenues	*Size*
Business Risk	Net Income/Total Revenues	*Efficiency*
Business Risk	Property, Plant and Equipment (PPE)/Total Assets	*Operating Flexibility*
Business Risk	CRS	*Country Risk Score*
Business Risk	CPI Growth	*Consumer Price Index (CPI) Growth*
Business Risk	IRS	*Industry Risk Score*

Source: S&P Global Market Intelligence. As of January 21 2020. For illustrative purposes only.

In-Sample and Out-of-Sample Analysis: We split the dataset of private companies into two samples to help assess the performance of the model in a real-world deployment. The in-sample portion (90%) represents our training dataset and is used to develop the model, while the out-of-sample portion (10%) is used to evaluate the model. We also make sure that the two datasets are similar with respect to the default rate and other descriptive properties (such as industry sectors and revenue size).

Different ML Algorithms

There are several ML algorithms available, and selecting the optimal algorithm is not straightforward. Algorithm selection depends on various factors, such as data type and features, transparency and interpretability, and model performance characteristics. We selected the following classification and regression algorithms for further analysis:

- **Altman Z-score:** The Z-score is an established model that leverages a linear combination of financial ratios to estimate the likelihood of financial distress. The model is based on the discriminant analysis technique to optimize model parameters.
- **Logistic regression:** A logistic regression is a statistical model that uses a logit function to model a binary dependent variable. It is a classical and widely used technique to

model the PD. The optimization function usually tends to include a regularization term (e.g., lasso, elastic net, or ridge) to limit the overfitting.

- **Support Vector Machine (SVM):** A SVM is similar to logistic regression and constructs a hyperplane multidimensional surface to separate two classes in the dataset. Inputs are transformed using a kernel function, allowing SVM to model nonlinear classification problems. However, by using a nonlinear kernel, the SVM becomes a black box because each prediction is not easily attributable to an individual variable.
- **Naïve Bayes:** Naïve Bayes is a classification technique that utilizes Bayes' theorem with an assumption of independence among predictors. Although this assumption is often violated in practice, naïve Bayes still tends to perform well. The technique is relatively robust and easy to implement, however, strong violations of the independence assumptions and nonlinear classification problems can lead to poor performance.
- **Decision Tree:** A decision tree model produces a flow chart structure where model prediction is obtained through a sequence of nodes and branches. While decision trees are a highly flexible tool, their usability may be hindered by poor out-of-sample performance as a result of overfitting. Various techniques exist to reduce overfitting by controlling the size of decision trees, such as pruning. We opted to contain the tree size by setting a limit of 50 observation per node.

Results

We tested the performance of the described ML algorithms using our global sample of private companies and accompanied variables, listed in Table 1. We implemented the analysis using Statistics and ML Toolbox™ functions in MATLAB®, and applied default algorithm settings to train the PD models and calculate their performance statistics.[7]

We evaluated the ML models using the receiver operating characteristics (ROC) curve and corresponding area under the curve (AUC). Table 2 shows the in-sample and out-of-sample AUC performance statistics. In-sample, the decision tree model exhibits superior performance with a near-perfect classification of defaulted and non-defaulted companies. Logistic regression and SVM are similar techniques and exhibit equally excellent performance, while the other two approaches demonstrate good or fair performance.[8]

Out-of-sample AUC, however, demonstrates a more realistic measure of the model's performance in a real-world situation. While the decision tree method still shows the best performance, it is only marginally better than logistic regression. It is worth noting that the performance of the decision tree deteriorates considerably out-of-sample compared to in-sample, indicating lower reliability of this method in a real-world application. In comparison, the other approaches exhibit more consistent performance.

[7] MATLAB and Statistics and Machine Learning Toolbox 2019b, The MathWorks, Inc., Natick, Massachusetts, U.S.
[8] Typically, AUC values between 70% and 80% are considered fair, values between 80% and 90% are considered a sign of good discriminatory power, and values above 90% are considered excellent.

Table 2: AUC using various ML models

	Z-score	Logistic Regression	Support Vector Machine	Naïve Bayes	Decision Tree
In-sample	74.9%	93.1%	92.9%	89.0%	99.8%
Out-of-sample	79.4%	93.6%	93.1%	89.8%	94.8%

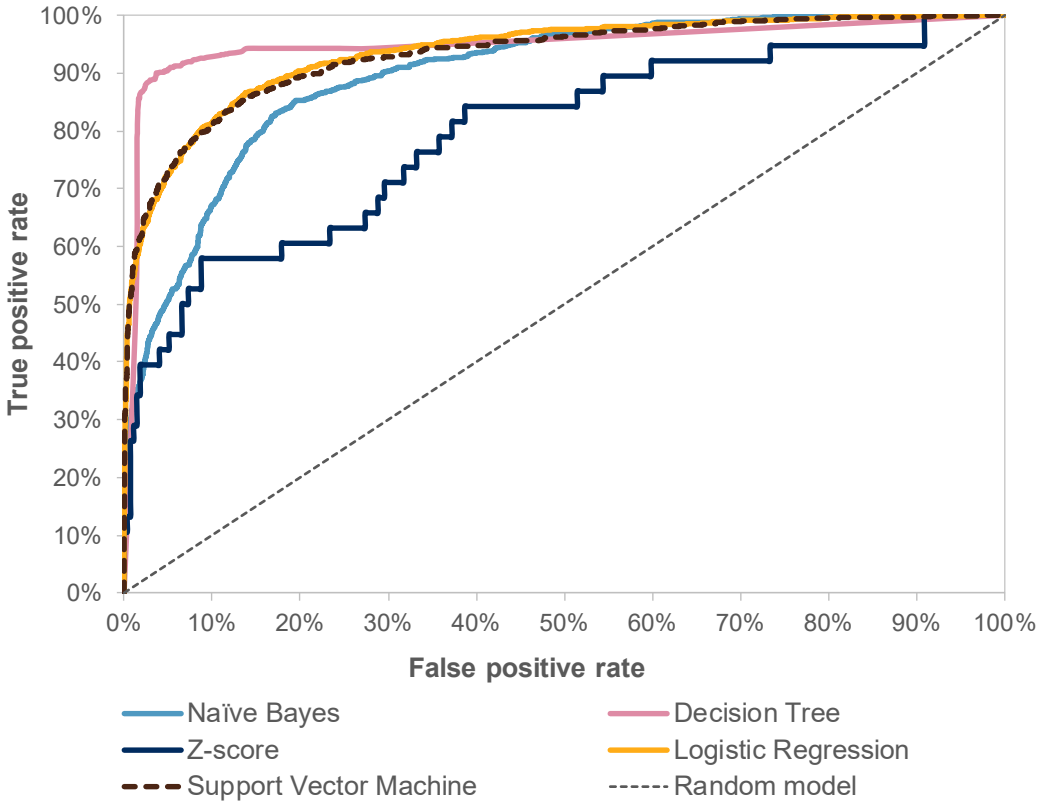
Source: S&P Global Market Intelligence. As of January 21 2020. For illustrative purposes only.

In Figure 1, we depict the out-of-sample ROC curves for the analyzed ML models. While two models may have the same AUC, the shape of corresponding ROC curves may be very different. For example, the decision tree and logistic regression have very similar out-of-sample AUCs, but their corresponding ROC curves are very distinct and cross at the low false positive rate and the high true positive rate. This reflects the Type I error and Type II error characteristics of the two models.[9] The decision tree outputs are rather binary, i.e., producing PD estimates of either 0% or 100, resulting in a more abrupt shape. The logistic regression, however, produces much more granular and continuous estimates of PD, resulting in a much smoother shape of the ROC curve.

Selection of the optimal model also depends on the use case. For example, Type I Error is more relevant when the goal is to minimize the incorrect classification of borrowers as creditworthy. Type II error, on the other hand, is more relevant when the goal is to minimize denying a loan to a creditworthy customer. If users focus on identifying defaults among the worst companies, they might prefer the decision tree model. However, those interested in good overall performance and differentiation among low-, medium-, and high-risk companies might favor the logistic regression model.[10]

[9] Type I error (false positive rate) is the probability of assigning a low PD to an obligor that will default. Type II error (false negative rate) is the probability of assigning a high PD to an obligor that will not default.
[10] Stein, M. R.: "Benchmarking default prediction models: pitfalls and remedies in model validation", Journal of Risk Model Validation, 2007

Figure 1: Out-of-sample ROC curve for various ML models



Source: S&P Global Market Intelligence. As of January 21 2020. For illustrative purposes only.

In addition to model performance, transparency and interpretability also play a vital role in the model evaluation. Namely, understanding drivers and the sensitivity of model predictions to changes in the input is an important aspect of model usability. In that aspect, logistic regression is preferred to SVM as it is more straightforward to analyze and interpret. The logistic regression also enables users to incorporate various constrains easily, thus making this technique highly controllable and adaptable.

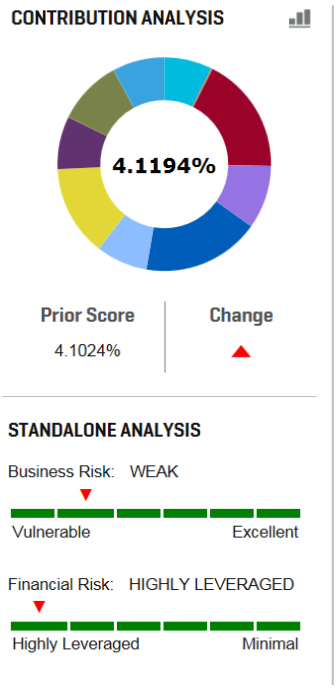
S&P Global Market Intelligence's Approach

At S&P Global Market Intelligence, we developed PD Model Fundamentals (PDFN) - Private Corporates, a statistical model that produces PD values for all private companies globally. The model is based on the maximum expected utility (MEU) theory and employs a logistic regression algorithm with ridge (Tikhonov) regularization.[11] [12] The methodology includes a number of data handling techniques to support robust treatment of financial ratios and management of extreme values. The process of variable selection leverages a k-fold Greedy Forward Approach to support a good out-of-sample and out-of-time performance. The transparent, 'glass-box' model structure of PDFN - Private Corporates enables users to understand the model behavior and easily analyze sensitivity and contributions of model inputs.

[11] Friedman. C and Sandow S.: "Learning Probabilistic Models: An Expected Utility Maximization Approach." Journal of Machine Learning Research, 4, 2003.
[12] S&P Market Intelligence: "PD Model Fundamentals - Private Corporates", White Paper, 2018.

Figure 2 shows an example of PDFN - Private Corporates outputs for Neiman Marcus Group, Inc. ('Neiman Marcus'), an omni-channel luxury fashion retailer primarily located in the U.S. Based on the latest available financial data, the company's PD of 4.1% implies a credit score of 'b'.[13] The in-depth analysis of the model drivers reveals that the retailer is highly risky from a financial and business point of view. The contribution analysis shows that low profitability and high debt are the main drivers of the PD estimate. The sensitivity metrics indicate that Neiman Marcus's credit score is highly sensitive to any adverse changes in industry and country risk factors.

Figure 2: PDFN - Private Corporates outputs for Neiman Marcus Group, Inc.



	RATIO VALUE	MEDIAN	CONTRIBUTION		SENSITIVITY	
BUSINESS RISK			RELATIVE	ABSOLUTE	VALUE	RANK
Corporate Industry Risk Score	moderately high risk	moderately high risk	0.0000%	7.2738%	0.2110	
Country Risk Score	aaa	aaa	0.0000%	0.0000%	0.0799	
CPI Growth	1.64%	2.09%	1.4205%	0.2055%	0.0009	
Net Income / Total Revenue	(3.49%)	0.40%	93.2172%	17.7961%	0.0148	
PPE / Total Assets	20.84%	42.87%	17.9844%	9.6123%	0.0119	
Total Revenue	4,688.2*	613.0*	(61.0268%)	17.7961%	0.0203	
FINANCIAL RISK						
Cash / Total Assets	0.52%	3.69%	6.5532%	7.8489%	0.0017	
Current Liabilities / Net Worth	199.28%	79.37%	27.8856%	13.7255%	0.0150	
Net Income / Total Liabilities	(2.34%)	1.97%	7.2414%	8.0078%	0.0011	
Return on Net Capital	(102.37%)	1.97%	54.6397%	9.8850%	0.0000	
Total Equity / Total Assets	5.12%	31.81%	12.1260%	7.8484%	0.0022	

Note: Industry median calculated based on a sample of department stores in the U.S.
Source: S&P Global Market Intelligence, as of January 21 2020. For illustrative purposes only.

Summary

A prudent approach includes reviewing and assessing various techniques for the problem at hand. While all presented models could be further refined and optimized to achieve better performance, the knowledge of the end application should also be factored into the decision-making process. In a real-world environment, this includes taking into account data availability limitations, model transparency requirements, the granularity of model outputs, and ease-of-use.

[13] S&P Global Ratings does not contribute to or participate in the creation of credit scores generated by S&P Global Market Intelligence. Lowercase nomenclature is used to differentiate S&P Global Market Intelligence PD scores from the credit ratings used by S&P Global Ratings.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global Ratings does not contribute to or participate in the creation of credit scores generated by S&P Global Market Intelligence. Lowercase nomenclature is used to differentiate S&P Global Market Intelligence PD credit model scores from the credit ratings issued by S&P Global Ratings.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its Web sites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.