Machine Learning Credit Analytics for Trade Finance

TRADETEQ WHITE PAPER
APRIL 2018

WHY MACHINE LEARNING?

Credit analytics is an important part of trade finance. Unlike many other lending activities, trade finance is mostly needed for smaller companies, with relatively small loan amounts ($20k to $200k), short terms (2-3 months), and borrowers lacking formal credit ratings. Tradeteq's mission is to expand access to trade finance for SMEs by making trade finance exposures investable. This requires better transparency of risks.

The traditional approach to trade finance would require a lot of antiquated and labour-intensive bank processes to assess SME credit. With banks optimising their books and cutting branch networks in many countries, many information flows that the traditional approach is relying on are now broken.

The existing bank credit underwriting and credit scoring approaches are very rigid and require the availability of specific accounting information from each applicant.

Tradeteq's approach to credit analytics:

- Leverages a broad set of available and emerging data sources.
- Uses many different company features as inputs, but does not place a hard requirement on the availability of most of them.
- Applies a rigorous evidence-based credit scoring process.

This approach allows for both better credit decisions and the improved access to credit for many SMEs.

A TRANSITION PREDICTION PROBLEM

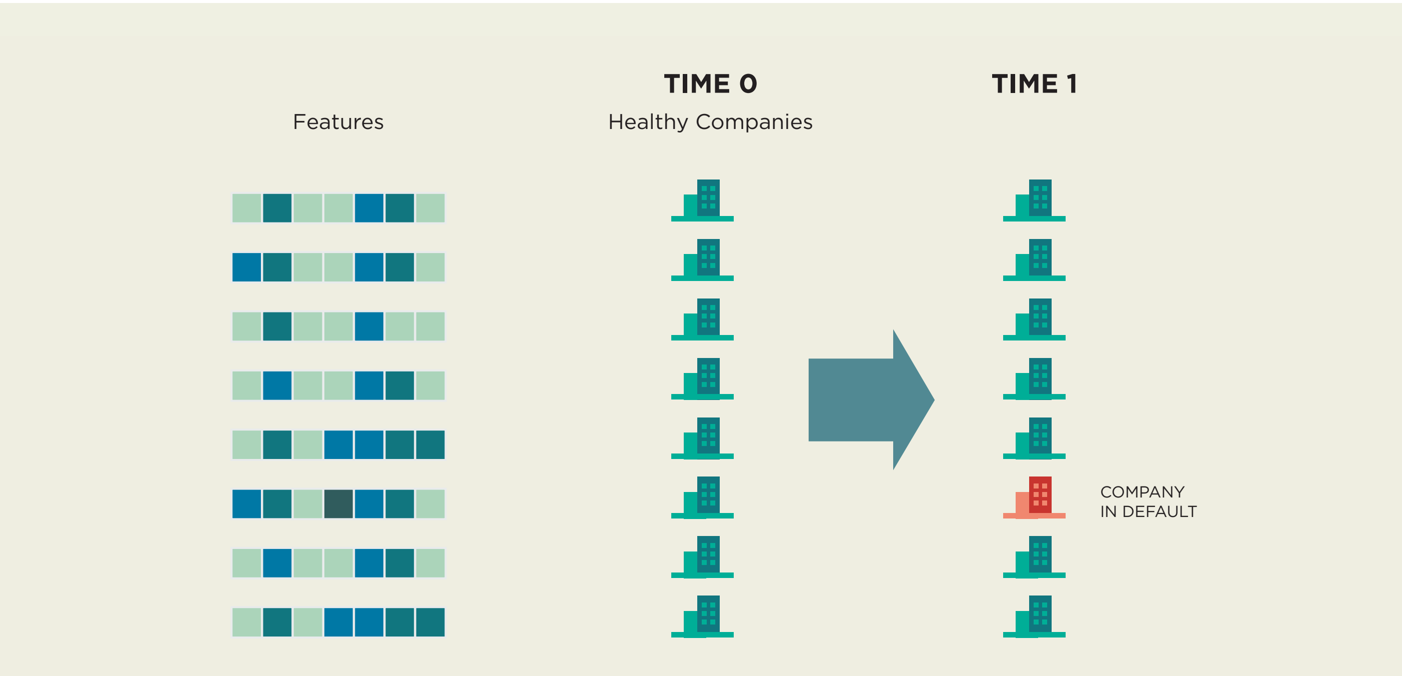
Fundamentally, credit scoring is a *transition prediction problem*. We are observing a universe of companies with data on each company, its operations, and the environment in which the company operates and make predictions for each company based on our assessment of credit event likelihood in a given time period for each company. A credit event can be administration, receivership, bankruptcy proceedings, creditors' meetings, or late payment of a single trade receivable.

> *"Fundamentally, credit scoring is a transition prediction problem."*

For a given class of credit events, and for each period, this is a binary classification problem. As no prediction will be perfect, it is more informative to show predictions as scores that express our assessment of credit event probability.

The traditional, and still widely applied, credit scoring model is the Altman Z-score and its modifications and variations. This approach uses linear discriminant analysis based on several accounting indicators [Altman et al, 2014][1]. It has been in use since 1968 and improved over the years by many researchers expanding the set of accounting features, recalibrating on alternative datasets, and applying non-linear predictors, including neural networks (see [Louzada et al, 2016][2] for an overview.)

Figure 1. Predicting credit transition based on observed features (for illustration purposes only).

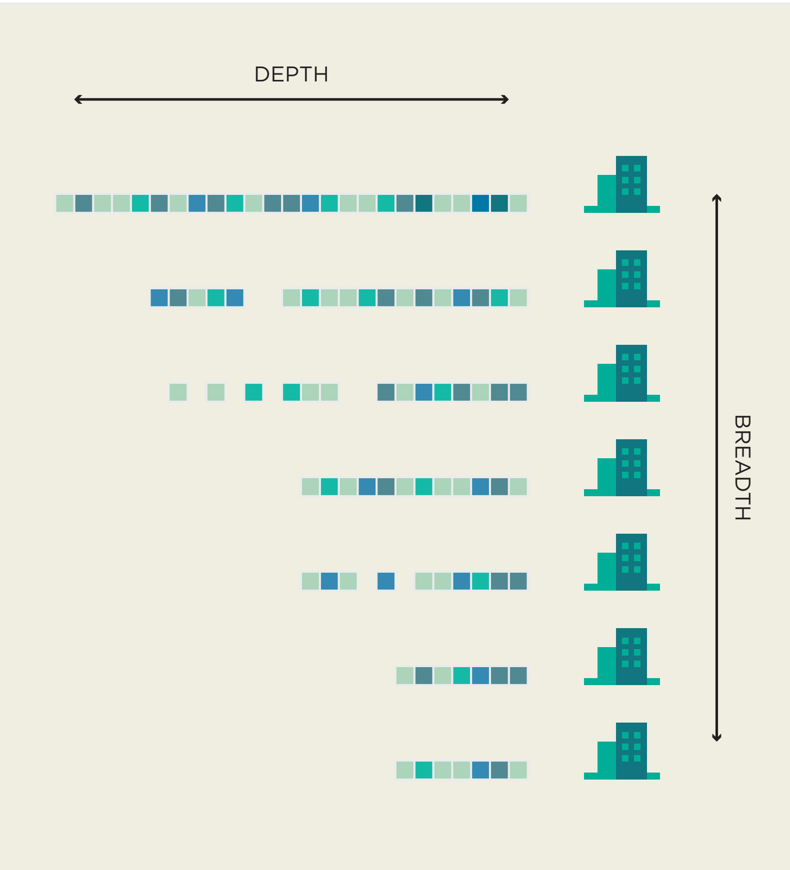


1. [Altman et al 2014] Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K. and Suvas, A., 2014. Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model.

2. [Louzada et al 2016] Louzada, F., Ara, A. and Fernandes, G.B., 2016. Classification methods applied to credit scoring: Systematic review and overall comparison. Surveys in Operations Research and Management Science.

DATA: BREADTH VS DEPTH

Figure 2. Data Breadth vs. Width (for illustration purposes only)



In company information datasets we can distinguish between data breadth – how many companies this data is available for - and data depth – how much information is available for each company.

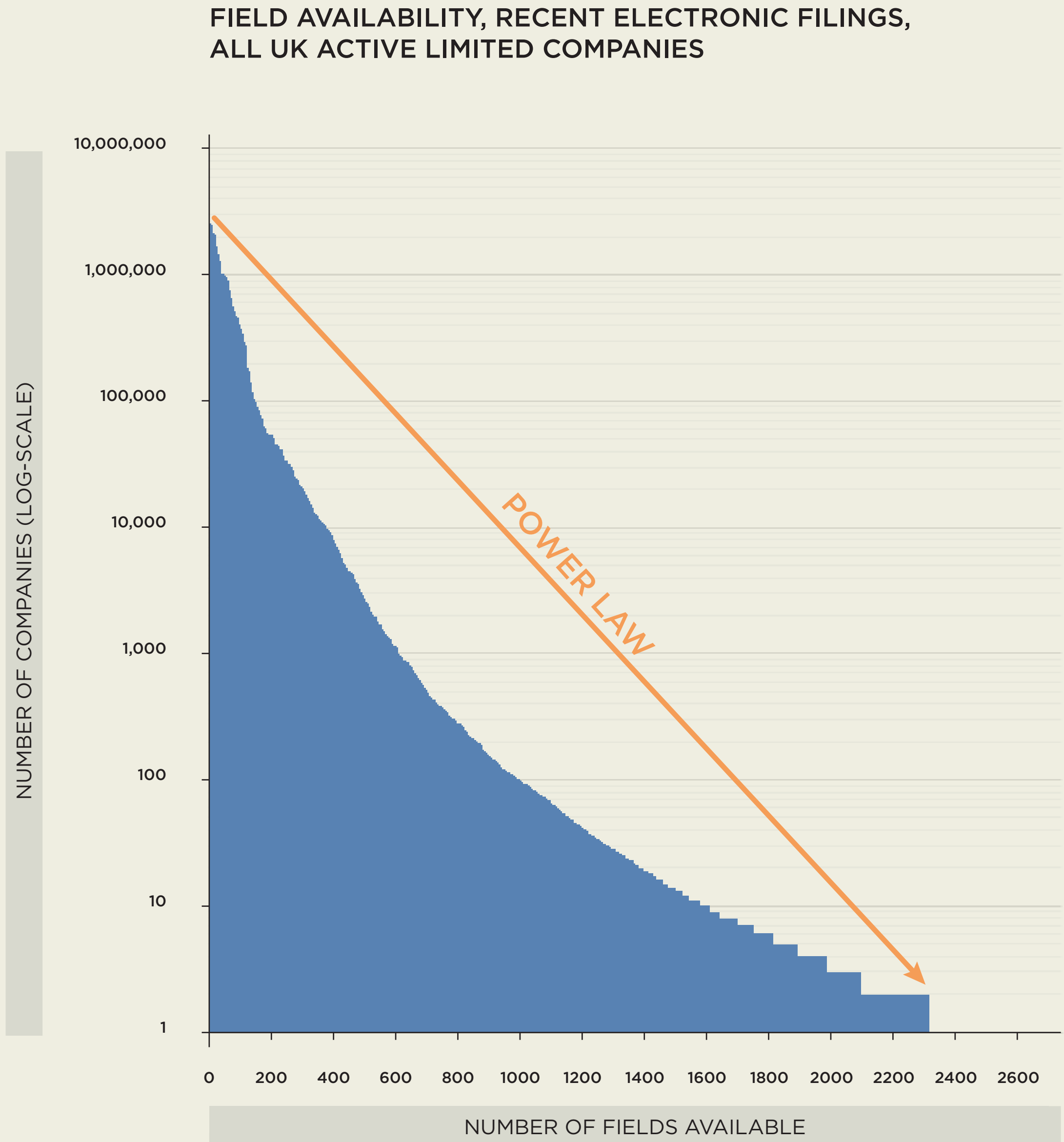
Traditional credit scoring requires a small number of accounting entries and ignores most information available in typical company accounts, as well as any non-accounting information that is available for a company. At the same time, these models place a hard requirement on each of the entries used – the score cannot be calculated for any company that misses even one of these entries. That is, the model imposes a fixed depth requirement on the dataset, thus posing a severe restriction on its breadth.

This combination of small number of entries used and hard requirements does not combine well with the patterns of data availability. Each additional data field required reduces the number of companies covered and does not allow the incorporation of many data sources that do not cover the full company universe. The faster the number of companies shrinks when one increases the number of required entries, the more important it is to relax hard entry requirements. It turns out that data availability often follows extremely fat-tailed patterns, with coverage breadth (number of companies) declining faster than at a power law with required data depth (number of features per company. See figure 3.) This is a much steeper decline than the one observed by Mandelbrot[3] for many social phenomena.

This means that good predictive credit models should be able to accommodate varying data availability across companies. If a certain entry is missing for a company, the absence of this entry may give useful information about a company and a good model would incorporate this lack of information into the score rather than discarding the company. This approach allows us to use quite deep datasets, each of which may have just moderate breadth.

3. Mandelbrot, Benoit B. The fractal geometry of nature. Vol. 173. New York: WH freeman, 1983.

Figure 3.
Distribution of UK limited companies by the number of fields available for bulk access from the Companies House



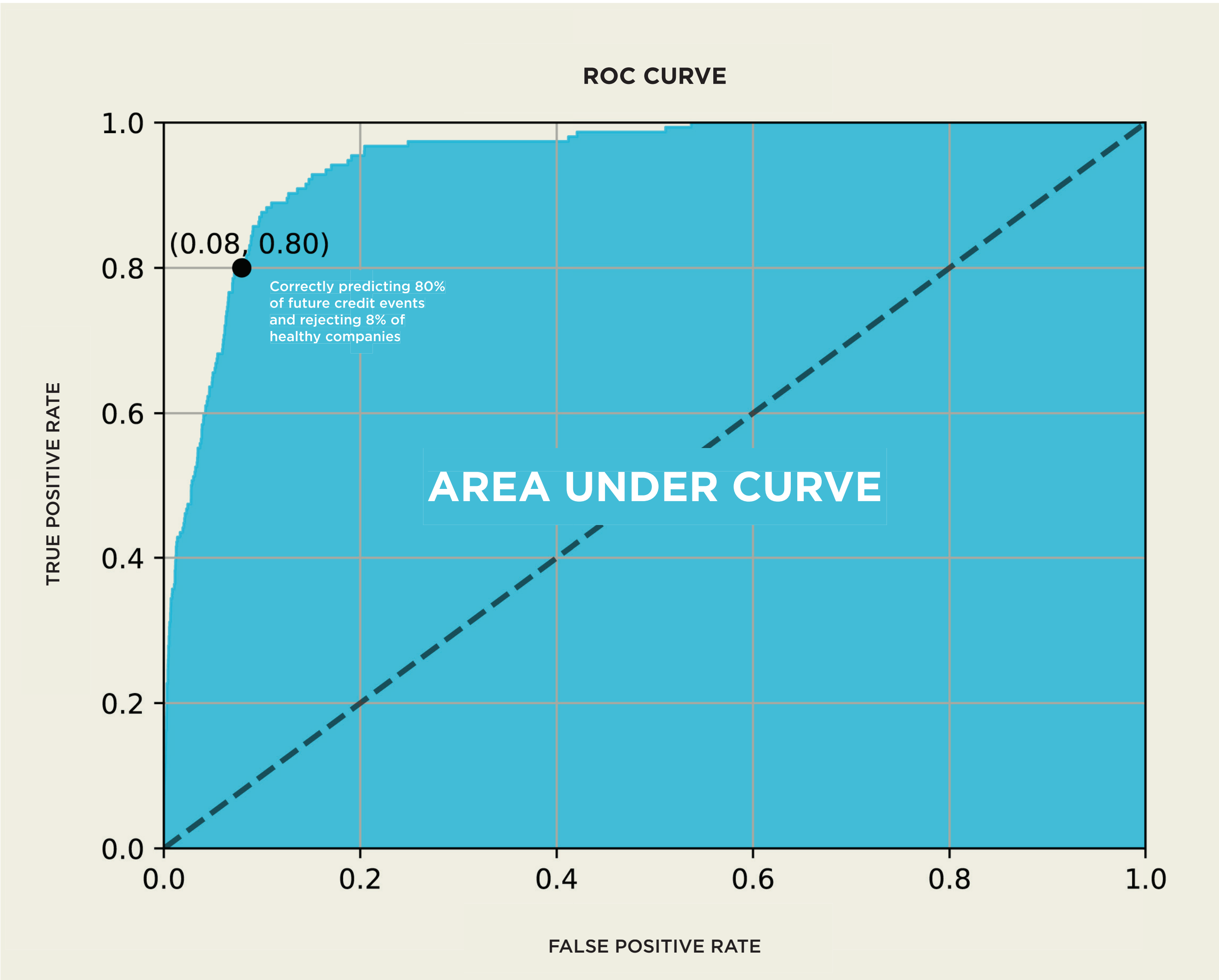
“Tradeteq’s model is able to incorporate a variety of higher frequency granular features.”

Another problem with approaches based on accounting measures is the lack of timely data: most companies file accounts once a year or even less frequently, so the predictions may be based on data which is up to 2-3 years out of date. To improve prediction quality, one needs access to higher frequency and more granular data. Such datasets are available to corporates themselves and to certain counterparties, e.g. banks, large customers, electronic invoicing companies, marketplaces, etc. The challenge is to gain access to this private data, accumulate it, and to utilise it efficiently for better credit event prediction. Tradeteq’s model is able to incorporate a variety of higher frequency granular features.

MEASURING MODEL PERFORMANCE

Typically, a credit model assigns each company a measure of riskiness – a credit score. This score represents model assessment of company risk. For each company, it can be used to make pricing or acceptance/rejection decisions. These decisions can be relatively complex and depend on many other factors and inputs, like the opportunity cost and market competitiveness, so it makes sense to separately measure model performance and credit decision efficiency.

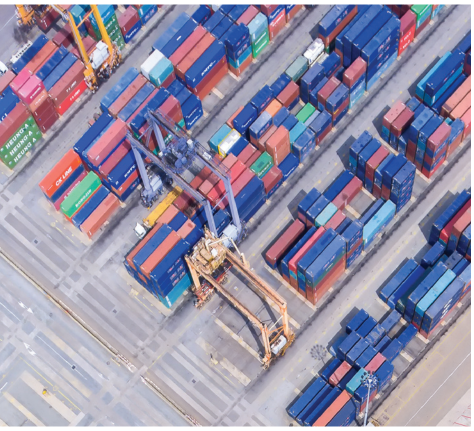
Figure 4. Tradeteq Neural Network UK limited company credit model, v.2, test sample ROC curve. Area Under Curve (AUC) 0.92.



The standard way to measure classification model performance involves calibrating the model on a training dataset, and then letting the model assign scores to companies from another disjoint data set, called the test set. We can then set a certain credit score threshold and classify all test set companies with a credit score above this threshold as “risky” and all companies with a credit score below the threshold as “safe”. Model performance can then be described by two numbers: the share of companies that had a credit event and were classified as “risky”, called the true positive rate (TPR), and the share of companies that did not have a credit event and were classified as “risky”, called the false positive rate (FPR). Ideally, we want a high true positive rate, and a low false positive rate. If we move the score threshold high, we will increase both the TPR and the FPR, and if we move the threshold low, we will decrease them both. Charting TPR versus FPR for different values of the score threshold gives us our model’s Receiver Operating Characteristic (ROC) Curve, see figure 4 for an illustration. Depending on the costs of missing a risky company and the benefits of lending to an additional safe company, one may be interested in different points on the ROC curve. An often used integral characteristic of the curve is the Area Under Curve (AUC). This metric does not depend on the choice of a particular threshold and can be used to describe predictive model performance across a range of risk/return preferences. AUC of 0.5 corresponds to a “coin flipper” model that simply randomly assigns companies to the “risky” and “safe” classes based on a coin flip. A perfect “time machine” model that never makes mistakes has an AUC of 1.0.

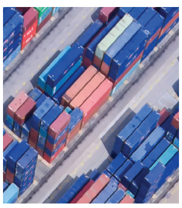
CURRENT APPROACHES AND BENCHMARKS

The traditional method of credit scoring is the famous Altman Z-score and its variations, still widely used today. For private non-listed companies, this score is calculated as a sum of several financial ratios, such as the ratios of retained earnings and working capital to total assets, with pre-determined weights.



Over the years since its introduction, there were many attempts to improve this score, by adding new financial ratios or replacing the linear discriminant approach by other models, including neural networks. Some of these models restrict themselves to certain types of companies (e.g. listed companies only or industrials only) but they usually don't take into account the wealth of non-financial information available for each company, such as its industry, registered address, history of renaming or mortgage charges. A frequent reason for non-inclusion of this data is that it is not as such obvious why a history of name changes may be relevant for company credit risk, for example.

The latest published large-scale test of the Z-score and its variations was performed by Altman with co-authors in 2014 [Altman et al, 2014]. This test covers private limited companies in 35 countries. For European countries, the test was based on an accounting database with near-complete coverage, but the authors excluded smaller companies from their study "because financial ratios in very small firms are generally too unstable for a failure prediction model". For the UK, their total sample was around 340k firms, or around 13% of the full limited company universe. On this dataset, the authors report AUC for Z-score and several tested variations with additional accounting data between 0.70 and 0.74.



TRADETEQ CREDIT MODELS

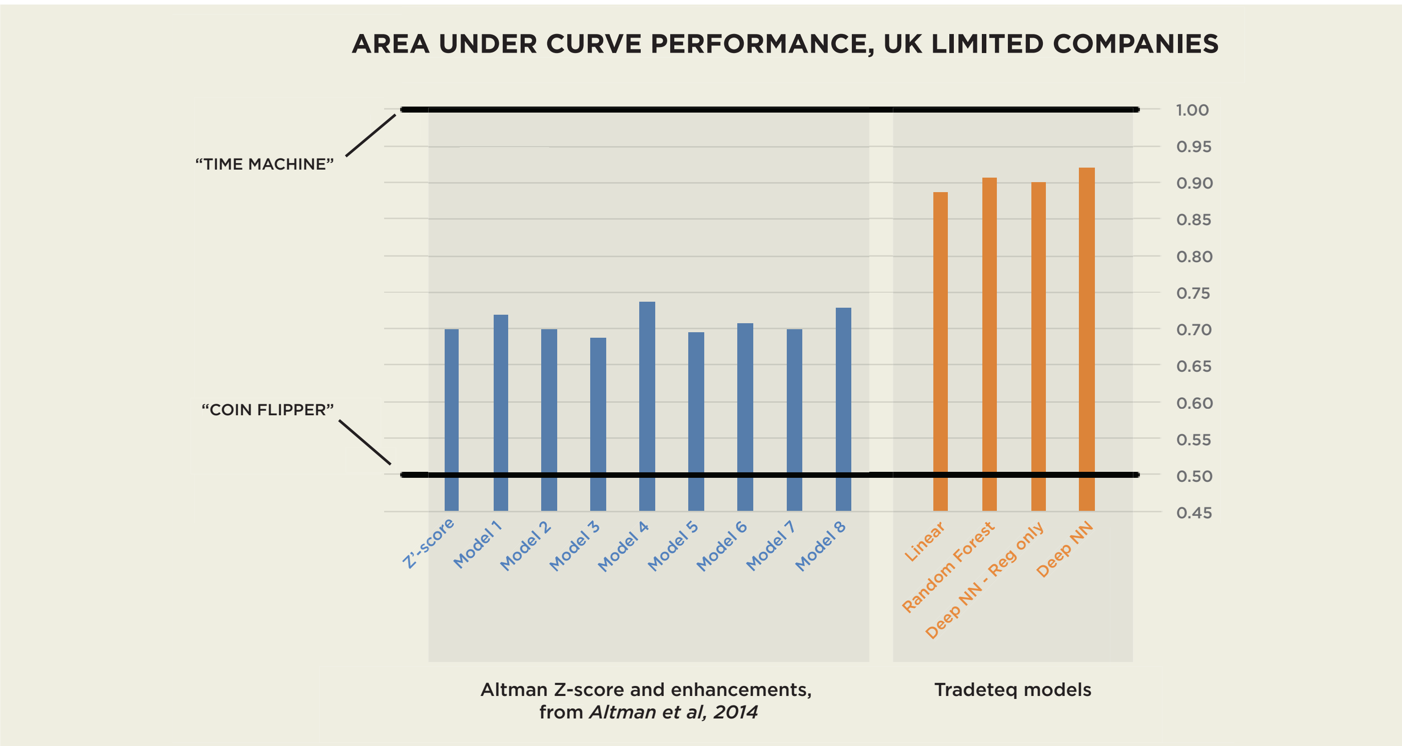
Tradeteq's credit models combine registration and accounting information about companies and enrich this data with geographical and socio-economic information. We recognize data depth and breadth by carefully handling any missing entries. This means that if a company is missing a certain accounting entry, we are not excluding it from the training and test sets, but instead note the absence of data and learn from this pattern of absences.

We are enriching our data by mapping the company's registration address to socio-economic area classification and census data. This information proves to be useful for a number of industries.

At the moment, we have a UK limited company model in production and live. This model combines inputs from Companies House, London Gazette, Office of National Statistics, and provides credit scores for all the 3.4m UK active limited companies. We have tested a number of classification model approaches, including linear models and random forests. Presently, our best performing model is a deep neural network with 4 hidden layers.

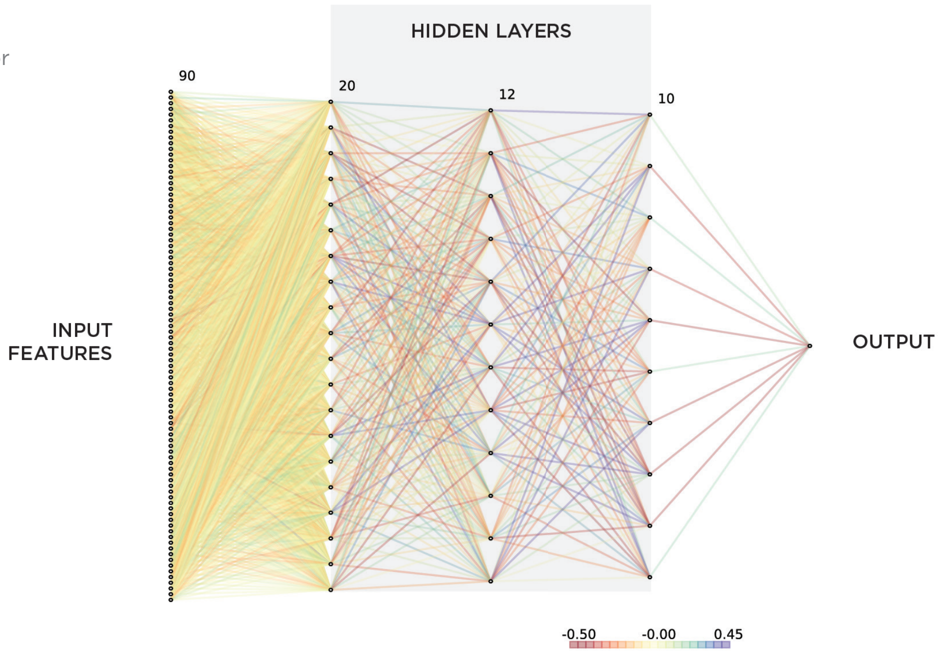
Tradeteq's models have much wider coverage than the models tested in Altman et al, 2014 and never reject a company just because its assets are too low. A combination of machine learning techniques with deep and broad data coverage allows the outperformance of traditional Z-score and similar models even on pure registration data, without using any accounting inputs. Neural networks prove to be efficient for this kind of data.

Figure 5. AUC of models from Altman et al, 2014 vs. Tradeteq models, UK limited companies. Tradeteq models are Linear, Random Forest (120 trees), Deep Neural Network restricted to company registration data only, and an unrestricted Deep Neural Network with 4 hidden layers.



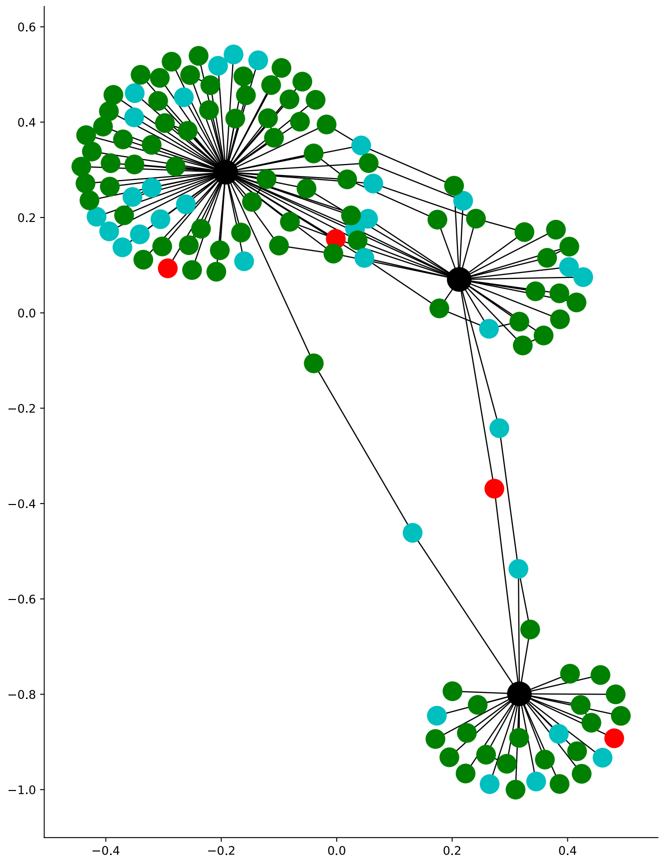
"A combination of machine learning techniques with deep and broad data coverage allows the outperformance of traditional Z-score and similar models even on pure registration data."

Figure 6.
A classifier Deep Neural Network with 4 hidden layers. Vertice color corresponds to network weight



NETWORK DATA MODELLING

Figure 7.
Supply chain as a graph (for illustration purposes only.)



To make credit event prediction more timely and precise, one needs deeper and higher frequency datasets than just registration and accounting data.

It is well known that the networks that an individual belongs to can be a good predictor of her creditworthiness [Wei at al, 2015][4]. This is used in a number of consumer credit scoring systems. It is also well known that payment and supply chains are important transmission channels for credit shocks [Kolay et al, 2016][5].

By the nature of Trade Finance, private data will be often available in batches by trade network, with underlying credit exposures linked by common clients, suppliers, or bank relationships. This allows us to utilise this network data by applying both supervised and unsupervised learning.

Network data will allow timely inference that would not be possible to make in a company-by-company analysis [Sen at al, 2008][6]. For example, for trade finance data, it allows us to spot features not visible in simple cross-sectional analysis, such as irregularities in repayment patterns relative to other customers of a given supplier, or changes in trade payments relative to other similar companies from the same industry/region.

Tradeteq is looking for partnerships and collaborations to work on transaction-level trade finance datasets. We will be combining our expertise in deep data analysis and credit scoring with our partners' broad data to produce state-of-the-art credit analysis.

Contact info@tradeteq.com if you want to know more about our credit analytics offering.

4. [Wei at al, 2015] Wei, Y., Yildirim, P., Van den Bulte, C. and Dellarocas, C., 2015. Credit scoring with social network data. Marketing Science, 35 (2), pp. 234-258.

5. [Kolay et al, 2016] Kolay, M., Lemmon, M. and Tashjian, E., 2016. Spreading the misery? Sources of bankruptcy spillover in the supply chain. Journal of Financial and Quantitative Analysis, 51 (6), pp. 1955-1990

6. [Sen at al, 2008] Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B. and Eliassi-Rad, T., 2008. Collective classification in network data. AI magazine, 29 (3), p. 93.

tradeteq

Tradeteq

119 Marylebone Road
London NW1 5PU
United Kingdom
info@tradeteq.com
+44 20 3965 6900

Our Partners

CREATIVE DESTRUCTION LAB

GTR VENTURES