Part of Dun & Bradstreet's
Worldwide Network

D&B'S

AI LAYERED CREDIT SCORING SYSTEM

INTEGRATING ARTIFICIAL INTELLIGENCE WITH CREDIT SCORING – STRUCTURALLY ADVANCING BUSINESS RELEVANT PREDICTIVE PERFORMANCE THROUGH INNOVATION IN CREDIT SCORING.

Barry de Goeij & Joris Peeters
Dun & Bradstreet B.V., June 2023

Preface

As part of the service and products offered to its customers, D&B [1] has been developing statistically based credit scoring systems for more than 30 years. Every day, these credit scores drive thousands of creditworthiness B2B decisions, across the world.

D&B has always strived to provide its customers with the most predictive scoring systems, given the data at hand, enabling its customer to do as much business as possible, for the least level of future business bankruptcy or delinquency of payment.

In this light, Machine Learning and Artificial Intelligence are nothing new to D&B. In effect, the company – and in extensio its customers – have been heavy users of these technologies for many years.

Given however the increased attention to, and possibilities provided by (big) data, machine learning and artificial intelligence, combined with the request of its customers for ever more predictive scoring systems, in 2019 D&B[2] achieved a breakthrough in its R&D trajectory on exploring how artificial intelligence may enhance the classical credit scoring systems.

Today, we can proudly state that the results of this R&D investment have not only proven to significantly promote the predictive power of D&B's scoring systems, but that in our view they represent an entirely new avenue for future (credit) scoring developments. So far, our learning curve has been quite steep, and we believe we have created a very solid basis. Much remains to be explored and uncovered, turning this solid base into an avenue for more and exciting future developments.

During the development of this new methodology for (credit) scoring, a substantial time and effort was invested in ensuring that this new scoring methodology provides a system which is controllable, explainable, and scalable & maintainable. Logic, subject matter expertise and economic soundness are fully integrated. In this, we remained true to the basic, fundamental principles of credit scoring.

This document provides an overview, explanation and insight into the new scoring methodology as developed by D&B. Out of intellectual property protection reasons, no intricate details can be shared. However, we believe that it does give an adequate level of insight for the user to grasp the underlying concepts, reasoning, and outcomes.

[1] Dun&Bradstreet Inc., headquartered in the USA
[2] Dun&Bradstreet BV, headquartered in the Netherlands.

We hope you enjoy reading this document as much as we enjoyed – and keep on enjoying – our AI Layered scorecard developments.

Joris Peeters,

Barry de Goeij,

Rotterdam, June 2023

Table of Contents

1. Background

D&B has been developing and marketing credit scorecards for over 30 years. The objective of the D&B scoring systems, is to predict a future bankruptcy or delinquency of payment event, based on the data and information in D&B's database.

Over the years D&B has invested continuously in advancing the predictive power, relevance and logic of its scoring systems.

Some fundamental R&D breakthroughs achieved by D&B since the early 1990's are as follows[3]:

- Adoption of logistics regression as the standard econometric estimation technique
- Segmentation of the population in relevant, logical risk groups
- Advanced stratified sampling to avoid bias caused by dormant or non-trading entities in the database [4]
- Modular-for-macro to allow the integration of a forward-looking component.
- Integration of an adaptive logic to ensure the self-adjustment of the model to the macro-economic environment and evolutions.

Today, D&B's scoring systems are driving automated risk assessments in 41 countries around the world. They cover all companies, from sole traders to large multinational companies, with or without the presence of financial account data.

On a daily basis, thousands of companies are re-scored, as the data in the D&B databases is continuously updated and enriched. D&B's scoring coverage is the largest in the market, with the D&B Scores driving thousands of decisions, every day.

2. Some insights in credit scoring

Credit Scoring was originally developed by Fair, Isaac in the US, to facilitate consumer lending decisions.

[3] Dr. R. Barazesh, J. Peeters, D&B Global Scoring Group, Bisnode Scoring Group, S. Hasbi & J. Peeters
[4] This applies specifically to business failure models

By their very nature, scorecards are transparent and logically explainable. This is mainly due to the principles of credit scoring:

(1) Credit scoring relies on statistical and regression analysis to produce an estimate of a future event (now a.k.a. machine learning or AI)
(2) The explanatory variables, typically are 'binned' or grouped into logically, risk level bases, ranges (now a.k.a feature engineering)
(3) The coefficient estimates are converted into logical 'points'
(4) The sum of the points produces a total "score". The value of the score is in itself a representation of the likelihood with which a future event is expected to take place.

It must be noted that such transparency is an existential element for credit scoring. At any time, a score given to an individual or company must be explainable to a user of the score, as well as to the scored person or company.

Below is a graphical representation of such a scorecard. The points of each variable will be summed to arrive at the total score.

Age	< 2 y. *0*	< 5 y. *12*	< 10 y. *15*	>= 10 y. *23*
SIC	< 1000 *10*	< 5000 *0*	< 7000 *15*	< 9999 *30*
Profit/Loss	< 0 *0*	<= 300.000 *10*	> 300.000 *20*	> 500.000 *50*
%change in Profit/Loss	no change *5*	< 0% *0*	<= 5% *10*	> 5% *20*
Net Worth	< 0 *0*	< 100.000 *5*	< 200.000 *15*	> 200.000 *24*
Paydex	< 40 *0*	< 60 *1*	< 70 *15*	>= 70 *31*

Figure 1: Scorecard

The higher the overall score (i.e., the sum of points), the lower is the likelihood of the future event happening. Likewise, the lower the overall score, the higher the likelihood of the future event happening.

3. The development of the scorecard with an integrated neural network (AI) layer[5]

Already in the last century, the credit scoring industry explored the use of artificial neural networks ('ANN') for the development of credit scorecards. Generally, however, it was found that the increase in predictive power was not sufficient to off-set the inherent 'black-box' characteristic of these ANN's.

Consequently, as transparency is of utmost importance for credit scoring, the industry lost interest in using ANN's for credit scoring.

However, since the middle of the last decade, as social media platforms and "new economy"-companies have started using AI systems on their platforms, algorithms have become and more mainstream. As such, they appear – for better or for worse – more and more "in public" and receive widespread attention.

Typically, they are seen as a key towards any future development, and have as such also attracted attention from regulators and legislators.

As a consequence, the credit scoring industry also took a renewed interest in Artificial Intelligence as a means to develop credit scorecards. Some recent developments on providing transparency on AI, a.k.a. or Explainable AI, also supported this renewed interest.

As the Benelux market is quite advanced, D&B's customers in this market also showed a great interest in the potential of the use of artificial intelligence for credit scoring. Of particular interest was the objective of increasing substantially the predictive power hidden within the data, using deep learning (as neural networks are now often known) or other recently developed ML or AI techniques.

In parallel to its customers, D&B had already started to explore the use of artificial intelligence, with 2 main objectives:

(1) To substantially increase the predictive power of the scoring systems
(2) To ensure the explainability and transparency of the system

[5] B. de Goeij & J. Peeters, 2019

This goal was achieved in May 2019, with the development of the first credit scoring system incorporating a layer of artificial intelligence.

Other than some other recent approaches witnessed in the credit scoring industry, D&B's approach was to not have ANN's directly produce credit scores. Instead, the artificial intelligence has been encapsulated as a "layer", in-between two other layers:

(1) Layer 1: a logic layer, which provides steerage and guidance for the artificial intelligence. This layer is human-controlled, and driven by logical (subject matter expertise), economical and bad rate driven considerations. In addition, it holds the feature derivation and preparation of the AI layer.

(2) Layer 3: this is a traditional scorecard layer, following the output of the artificial intelligence layer. As we cannot expect the AI to be 'perfect', this layer serves to enhance the results from the AI layer and produces the final score. The scorecards in this layer are logistics regression based.

The results achieved with the "AI Layered" approach were very encouraging. In short, this approach enables to structurally advance the relationship between the number of "bad" versus the number of "good": for the lower scoring ranges, this means more "bads" captured, and less "goods". For the higher score ranges, more "goods" and less "bads" captured.

The below provides an insight into how this AI layered system enhances the good-bad ratio (a.k.a. odds-ratio) on a sub-portfolio of one of D&B's customers: the new system can capture the same 60% of defaulted cases from a new client acquisition process, for only 5% of the goods. This, compared to 24% of the goods using a classical scorecard development approach.

New Method

To capture			
5%	of the bads,	0.00%	of the goods are captured
10%	of the bads,	0.00%	of the goods are captured
15%	of the bads,	0.00%	of the goods are captured
20%	of the bads,	0.04%	of the goods are captured
25%	of the bads,	0.11%	of the goods are captured
30%	of the bads,	0.26%	of the goods are captured
40%	of the bads,	0.57%	of the goods are captured
50%	of the bads,	2.91%	of the goods are captured
60%	of the bads,	5.66%	of the goods are captured
70%	of the bads,	10.65%	of the goods are captured
80%	of the bads,	20.61%	of the goods are captured
90%	of the bads,	40.09%	of the goods are captured

Classical Method

To capture			
5%	of the bads,	0.15%	of the goods are captured
10%	of the bads,	0.23%	of the goods are captured
15%	of the bads,	0.26%	of the goods are captured
20%	of the bads,	0.79%	of the goods are captured
25%	of the bads,	1.70%	of the goods are captured
30%	of the bads,	3.25%	of the goods are captured
40%	of the bads,	6.53%	of the goods are captured
50%	of the bads,	14.08%	of the goods are captured
60%	of the bads,	23.78%	of the goods are captured
70%	of the bads,	36.13%	of the goods are captured
80%	of the bads,	47.15%	of the goods are captured
90%	of the bads,	62.89%	of the goods are captured

Figure 2: Performance Improvement

These results show how the addition of the AI layer can help D&B's customers advance the amount of business they can do, without increasing bad debt.

As for the architecture of this new system, the below highlights graphically the high-level difference between the current systems, and the new AI Layered system:

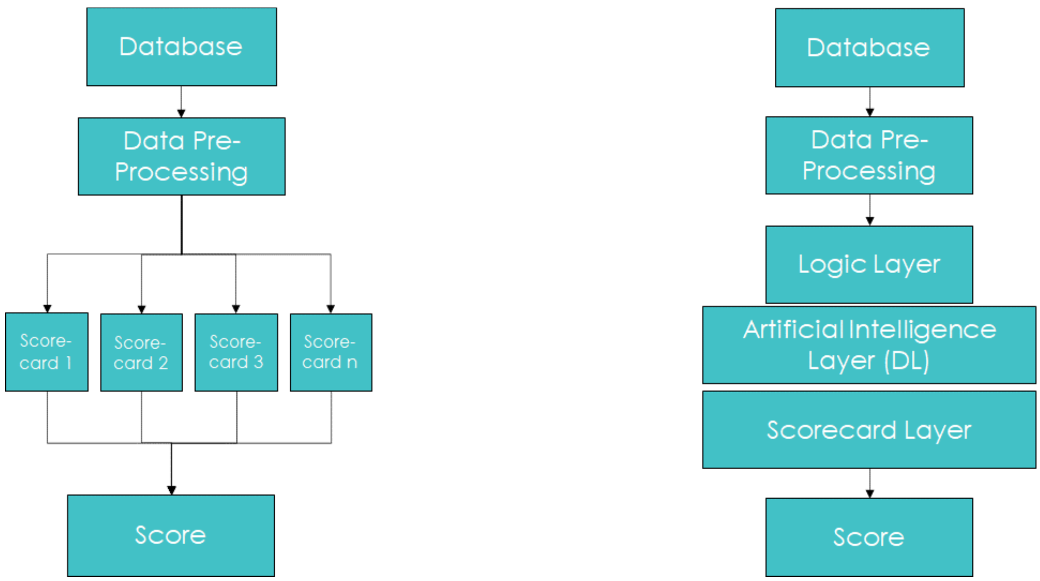


Figure 3: Current Scoring System Architecture (left hand side) versus AI Layered Scoring System Architecture (right hand side)

In addition to the development of a new logic and architecture, D&B invested heavily in the explainability, the logic, and the controllability of the system.

In the following sections, each of the Layers will be discussed in some level of detail, and appropriate explanations provided.

In a final section, an overview will be provided of the explainability and controllability algorithms developed by D&B.

It is important to note that the entire development is driven by:

- Human oversight and subject matter expertise
- The economic relationship between the data variables at hand, and the event to be predicted

- The use of data variables which satisfy the D&B DUNSRight® process[6]

3.1 Layer 1 – The Logic Layer

As described above, the first layer of the AI Layered scoring system, is the Logic Layer. This layer serves two main purposes:

(1) To segment the population in relevant sub-segments, and
(2) To provide guidance and steerage to the ANN's in the subsequent Layer

3.1.1 Segmenting the population in relevant sub-segments

Many business and real-life problems involve populations which are heterogeneous in nature. If and when this is detected, or otherwise known, it is always beneficial to employ either a stratified sampling methodology, or to segment the population along one or more dimensions which can separate the heterogenous population into more homogenous groups.

Given the extreme wide range of companies present in the D&B's databases, the business population on which any modeling takes place is by its very nature nearly always heterogeneous. To this end, D&B has a tradition of segmenting the inherently heterogeneous business universe at hand, into relevant, homogenous sub-segments. Credit Scorecards are then developed for each sub-segment.

This approach both enhances the predictive power of the models and ensures a proper treatment and incorporation of relevant underlying variables, in function of the segment at hand.

Some have argued that AI systems, such as ANN's, may not require a user to segment the development data set into relevant sub-samples. Instead, the going principle is that the neural network should by itself be sensitive to the underlying segments.

Unfortunately, D&B's research has shown this latter not to be the case. In developing an ANN on the entire data set of cases, it became clear that the network was not capable of sufficiently grasping the heterogeneity present in the data set(s). This was true even in

[6] See section 8 for a description of the DUNSRight® process, and what it means for the data used.

case the network was enlarged by either adding numbers of hidden layers, or adding extra neurons to the layers, of a combination of both.

Instead, the best results are achieved when developing ANN's on the sub-segments present in the underlying data set.

This raises the obvious question of which segmentation is the most relevant to be chosen. In some ways, this may be driven from subject matter expertise, but it may also be driven by statistical considerations.

D&B's segmentation approach, developed in the 1990's and further enhanced during the past decades, was to combine subject matter expertise with a statistical foundation. This approach has proven its relevance over time and is as such a standard feature of any D&B scoring system.

For the development of the AI Layered system, it was therefore again opted to combine, and fine-tune, two levels of segmentation:

- a subject matter expertise driven logical segmentation: this is based upon company typology, data coverage and risk level. Such a segmentation approach has proven its worth in a large range of scoring systems which were developed by D&B for its customers over the past decade[7]. It will not be elaborated further on in this document.
- within each logical segment, a second risk-group based segmentation is performed. This latter is also driven by business logic but is primarily based upon both statistical and data considerations.

Both segmentation approaches are combined in sequence, to generate a set of segments, upon which the Artificial Intelligence Layer is developed.

Before proceeding to the description of the Artificial Intelligence Layer, the secondary level of segmentation is discussed.

3.1.2 Statistical & Data Driven Segmentation: Risk-Based Segmentation

As indicated above, the secondary layer of segmentation is based upon statistical and data considerations. As the objective is to develop a system to predict the risk of

[7] J. Peeters,2010

bankruptcy or delinquency of payment, this secondary segmentation essentially is a risk-based segmentation process.

In the initial stages of the development of the AI Layered system, a manual process was used to generate relevant risk-based segments. This approach has now been optimized, whereby both logic as well as the underlying statistics (i.e., risk levels) are the basis of this segmentation.

It is also for this risk-based segmentation, that we deploy a deep learning AI process. Stated simply, the role of the deep learning is to assign to each record in the dataset, the risk segment to which this record most likely should belong.

The training performed is supervised, meaning that each for each record, the ex-post segment to which the records should belong, is known. The deep learning can then be trained to try and classify each record in the appropriate risk-based segment.

The most important element here is not so much the use of the neural network. Instead, the generation of the relevant risk segments is key and critical for a successful development of the AI Layered system.

In short, the risk-based segments should be:

- relevant,
- logical,
- separable, and
- provide a sufficiently different risk level,

both from a statistical as well as a subject matter expert perspective.

If the above dimensions are not respected, D&B's research has shown that no AI technique will be able to provide a satisfactory, ex-ante prediction to which segment a record will belong.

Furthermore, from the research conducted by D&B's, it was clear that using raw data to feed the network, led to results which are often chaotic and incomprehensible. Instead, a good degree of feature engineering is required to avoid these pitfalls.

Additionally, D&B's research also showed that feeding even the engineered data, led again to chaotic and incomprehensible results, when all available data fields are presented to the network.

In order to overcome these challenges, D&B developed a process encompassing three steps:

- determining the risk-based segments using primarily a logical approach
- a good degree of feature engineering on the raw data
- Chi square based metrics and ANOVA to determine by which data dimensions the segments can best be defined

The process is mathematical by nature, and will generate a range of reports, which help to understand whether a given segmentation is appropriate or not.

Ultimately, the outcome (or report) produced is essentially a list of "eligible" variables, which help to identify ex-post and ex-ante, the risk segment to which a record [should] belong.

An example of the Chi square metric and the eligibility table can be found in Appendix A.

It is using these "feature-engineered", eligible data fields, that superior classification results were achieved, both ex-post and ex-ante, in-sample and combined "out-of-sample-out-of-time".

During the development of the AI layered system, it became clear that an overly dynamic selection of the segments may lead to hardly any eligible variables. The network may provide a very nice fit, but then clearly it overfit, of finds patterns which are not generalizable.

On the other hand, a too cautious selection of the risk-based segments may lead to lost opportunities for improving the predictive performance of the overall system.

Through the R&D performed by D&B, a way was found to combine both dimensions of a dynamic selection vs. a cautious selection of risk-based segments. The experience and expertise gained has resulted in an algorithm which will allow for the segmentation process to run automatically, and which will provide us with the most optimal segmentation into relevant risk groups, for the data set at hand. Or at least, provide us with a series of segmentation options, for the data scientist and the subject matter experts to choose from.

Regardless of the degree of automation chosen to deploy on the segmentation into relevant risk classes, human oversight over this process will always be performed by D&B. This achieved through controlling:

- how the data is engineered prior to being fed to the neural network

- whether a segmentation abides to the rules set out above
- the performance of the network, in sample, out-of-sample and out-of-time
- how the network functions and can be controlled

For D&B, the last point has been a pivotal element in its R&D efforts. In contrast to the general current discussions on artificial intelligence, whereby explainability is high on the agenda, the research by D&B has shown the concept of controllability was, and continues to be, even more important.

Rather than "explaining" a neural network after the estimation process, the aim should be to fully control it. And this control should be performed prior to the network estimation process, rather than afterwards. Controlling the network also ensures explainability and understanding of how it works. In addition, controlling also provides the means to make the network do what the data scientist wants it to do, rather than the user being subject to the network's functioning.

3.2 Layer 2 – The Artificial Neural Network

As indicated above the neural network's prime role is to estimate, for each record, to which risk risk-based segment a record would belong.

The deep learning algorithms were developed using the Python Keras - Tensorflow package, in which a traditional, feed-forward backpropagation network was developed. The network is a multiclassifier type, as the aim was to develop an algorithm which classifies the records into their respective risk-based segment.

The selection of the variables to be used was supported by the insight generated by the above-mentioned controllability and explainability algorithms.

As mentioned, the reporting algorithms allow to effectively control the neural network. For example, the research of D&B showed that the removal of variables which the reporting indicates not to add to the predictive power, lead to more stable, robust, and explainable models.

The choice of the number of hidden layers, and the number of hidden neurons, was initially performed manually. This methodology was driven by the basic premise that full control over the modeling process was required, and to avoid overfitting.

The learning curve for this human oversight was very steep, as this approach allowed D&B to test and understand how variables and network structure affected the results produced by the system. Going through this learning curve has provided D&B with the necessary expertise to develop a fully automated process whereby, through an iterative process, an algorithm will be able to suggest an optimal model structure (or possibly a range of structures, from which either the system, or a developer, may then make a choice).

Such a hyperparameter-algorithm will rely strongly on the outcome of the explainability algorithms, but also on the stability demands which we have on the model produced by Keras[8].

As stability is very important to the users of the D&B scores, the "optimal" model structure was not simply the model which would generate the most accurate predictions on the development sample. Rather, it was the model which would provide a combination of 2 key metrics: predictive accuracy as well as stability over the out-of-time, out-of-sample, samples.

This 2-metric demand meant that some deep learning architectures which produced a very good accuracy, were rejected because of lack of stability. Likewise, stability can be achieved by a less powerful or accurate model. But then the lack of predictive power would make the model not usable for credit scoring purposes.

Feeding the model with the appropriate feature engineered data fields, and then testing for accuracy and stability, lead to a model architecture which conformed to the three key metrics:

(1) Model accuracy
(2) Model Stability, over the out-of-time and out-of-sample data sets (i.e. avoiding overfitting)
(3) Model logic and explainability

It was only after these three metrics were achieved, that the models were released for the next Layer in the AI Layered structure.

It must be noted that this approach was performed for each of the segments as identified in the Logic Layer. Each logical segment therefore benefits from its own deep learning algorithm, and each network takes explicitly the characteristics of the segment, in terms of relevant predictive data fields and the 'behavior' of the segment, into consideration.

3.3 Layer 3 – The Scorecard Layer

[8] D&B does not yet use such an hyperparameter-algorithm, but the building blocks and expertise are available already.

Following the Artificial Intelligence Layer, the third and final Layer consists of traditional credit scorecards.

As such, this Layer consists of the two main scorecard development components:

(1) The Binning of the variables, otherwise known as the 'univariate analysis'
(2) The Logistics Regression, otherwise known as the 'multivariate analysis'.

The outcome of this layer is a set of "classical" scorecards, which produce a "classical" credit score.

For the Logistics regression, Elastic Net was used (automatic scaling), which from D&B's experience in using this technique for more than 10 years, has proven to be very efficient in estimating the appropriate coefficients of the relevant variables.

Classical metrics such as correlation and significance tests of variables were included, as per D&B's modeling standards.

Each scorecard was developed on the set of records which were allocated by the neural network to a given risk-based segment. In extension, this set of records therefore (1) belongs to a logically based main company-type segment, and (2) within this logical segment, to a risk-based segment.

The coefficients as estimated by the regression, were converted to "points" using D&B's proprietary conversion algorithm[9].

The advantage of this approach is that the final score is transparent and fully explainable through the points-scoring system. In addition, as these scorecards are built on subsets of the records after allocation to a risk-based segment by the artificial neural network, it allows an additional human oversight, as the modeler developing the scorecards, will work on the set of records as allocated by the AI Layer. Any anomalies, or any illogical relationships between the data variable and the event to be predicted, or any finding which may not be explained from an economical or subject matter expertise perspective, observed during either the binning stage or the logistics regression stage, will become apparent during this scorecard Layer development phase.

As with the deep learning layer, the scorecards were tested for accuracy (as measured by GINI, KS, KL,… metrics), as well as stability over time. Also, for this layer, variables which proved to be non-stable over time, were removed from the scorecard algorithm.

[9] D&B Global Scoring Group, 1992 – 2003

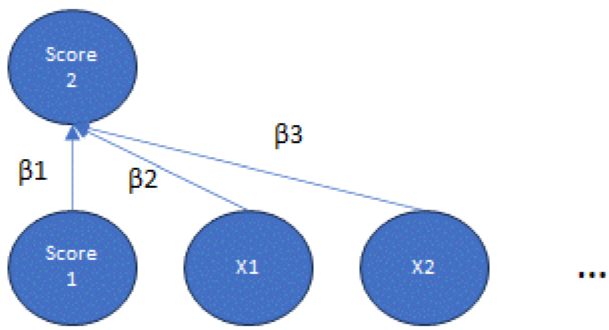
4. The Modular Scoring Approach

Over the more than 30 years of developing AI and Machine Learning models, D&B has developed an approach which has become known as 'modular scoring approach'.

Under this approach, one or more predictive variables may be omitted from use, either in the logistics regression, and/or in the artificial intelligence layer.

These variables are then later added back onto the calculated score, to generate the final score as computed by the scoring system.

Graphically, this can be depicted as follows:



Most often, this modular approach is chosen for the sector of activity variable (SIC Code). The rationale behind this choice is that this makes the model suited for easily integrating macro-economic forecasts ("forward looking") into the scoring system.

As per the R&D previously conducted[10], and as per the current 1-100 Dutch scoring system[11], the easiest way to integrate such forward looking is by entering this via:

(1) The industry sector or SIC code
(2) In a modular manner

[10] Bisnode Scoring Group, 2007-2008
[11] S. Hasbi & J. Peeters, 2011-2012

For the Benelux, a suite of macro-economic business bankruptcy models was developed when the Covid pandemic hit the region[12]. These models provide an assessment of the future probability of failure, given the current and past macro-economic developments.

Integrating the forward-looking component via the modular SIC code approach into the scoring system, provides:

(1) a direct calibration of the future, to-be-expected PD levels[13],
(2) the number of bankruptcies to be expected,
(3) an adjustment over the D&B Ratings[14]

Next to the SIC code, which for larger developments will nearly always be added in a modular manner, also other data fields can be considered.

In a standard-integrated-custom ("ST_I_CU") solution, these may also include customer specific variables, or variables to which a special handling needs to be foreseen.

The outcome of the Modular Layer is a new score, which combines the initially calculated score, with the data fields as added via the modular approach.

5. Controllability and Explainability

This section describes the outcome of the R&D efforts performed by D&B, in light of ensuring a full controllability and explainability of the artificial neural network.

To this end, D&B has used its extensive knowledge and expertise on credit scorecard development, to generate a range of controllability and explainability algorithms which provide a very detailed control and insight into the neural networks used as part of the AI Layered scoring methodology.

5.1 Controllability

As detailed above in section 4.2, the R&D efforts brought D&B to the point that a very clear way of controlling the functioning of the neural network has been achieved. D&B

[12] B. de Goeij & J. Peeters, 2020
[13] Bisnode Scoring Group, 2007-2008; Peeters, 2009
[14] J. Peeters, August 2020

now understands how it can, starting from the data in its comprehensive databases, make the network run according to how and what D&B wants it to do.

For D&B, this latter comes down to a system which enables a powerful risk classification, which is fully transparent and logically explainable, and stable through time at a high level of predictive performance.

Controlling the network does, however, not mean that D&B pre-determines its functioning. D&B still lets the deep learning do what it does best: to run through the data, and provide the user with a relevant, reliable, and predictive AI model. Rather, through the control on the network estimation process, D&B is able to envision the output of the process, at the point when the data is fed to the deep learning training process.

To date, D&B's approach hereby has been to avoid as much as possible that the network would produce inacceptable results. An example of such unacceptable behavior would be if the network were to place entire groups of records in the dataset into one or just a few high risk segments, when there is no logical explanation for it. Mostly, given the appropriate level of feature engineering, and selection of variables, this can be avoided prior to training the network.

Even so, controlling the network does not exonerate one from explaining how and what the network is producing as output. For this reason, D&B has invested a considerable amount of R&D, into developing its own metrics of explainability.

This explainability will be described in the next section.

5.2 Explainability, neural network control metrics, and stability

In order to gain an understanding and insight into an artificial neural network, practitioners and academic researchers often use explainability metrics such as Lime and Shapley. In addition, the data set is mostly split in a train and a test dataset (for generalization and stability testing), and neural networks are often visually inspected on the coefficient level.

D&B also started out by using these typical stability and explainability metrics. However, very soon it became clear that these metrics did not always provide the degree of insight and explainability required by D&B.

Against the background of D&B's traditionally fully transparent credit scorecards, and which tend to be very stable in their performance through time, D&B needed to have a full understanding of:

- the importance of each variable in the network, and in relation to the output produced
- whether the economic logic of the variable, in terms of its predictive power and pattern, was maintained
- how variables interact with each other in the network, and how variables influence each other

Developing appropriate insight metrics into the above three dimensions, represents a key investment of the R&D efforts performed by D&B during the development of the AI Layered credit scoring methodology. During this development process, D&B leveraged its expertise and know-how in developing credit scorecards. As a result, a range of analytical algorithms and metrics, reflected in easy-to-read and easy-to-comprehend reports, were developed. These algorithms effectively provide D&B with the understanding and the insight it was aiming to achieve.

5.2.1 D&B's Explainability Reports

Traditional credit scoring relies on two main underlying components to ensure model logic, control and explainability:

- a requirement of monotonicity, or a clear and quasi linear relationship between the variables and the target to be predicted
- a binning of the variables in 'buckets', and the transformation of the estimated coefficients from the underlying regression, into "points"

For the purpose of rendering its deep learning models explainable, D&B's research focused primarily on the monotonicity requirement. As for the 'binning' concept, this was leveraged to gain an in-depth understanding of the functioning of the deep learning algorithm.

One option would have been to convert the coefficients of the deep learning algorithm into points. This was however not pursued as it would not have helped in the advancement towards the objective of gaining insight into the model functioning in an easy and efficient manner[15].

[15] As is generally known, even in a 'classical' scorecard the 'points' of the variables do not provide a suitable insight into the model performance and behaviour.

D&B therefore orientated its research away from the classical metrics such as LIME, Shapley and network visualization, towards gaining an understanding of the relationship between the input parameters of the network and the outcome of the neural network.

By leveraging the knowledge on credit scoring techniques, a set of mathematical-analytical algorithms were developed, which were afterwards summarized into easy-to-read reports. This latter is important to also allow non-technical colleagues to be able to read and understand the functioning of the deep learning layer.

The reports were as follows:

- **Report I: Variable Monotonicity**

  As mentioned above, traditional credit scoring functions on the premise of a monotonic relationship between the level of the variable at hand, and the target to be predicted. This is not performed on the detailed level of the variable values, but rather on a 'binned' version.

  For judging the performance of the neural network, this concept of binning was leveraged to gauge the monotonic relationship between the level of the independent variable used in the model, and the outcome/prediction.

  Therefore, for each variable in the system, the algorithm assesses the monotonicity of the relationship between each 'bin' (or range of values) of a variable and the event rate, and this at the level of the prediction as generated by the neural network.

  The nature of relationship is then assessed through a traditional OLS regression on the observed relationship (after the prediction) . From this, the monotonicity can easily be observed and judged as follows:

  - The higher the R^2 value, the higher the monotonicity in the variable
  - The higher the value of the coefficient of the regression, the higher is the slope inclination of the monotonicity

  Both of these metrics are calculated for each variable and are represented graphically as well as in a summarized tabular format. The monotonicity algorithm has been designed to run both in a positive sloped and negative sloped relationship. It can handle concave and convex (a.k.a. "U-shaped") relationships as well, as for some of the variables the predictive value is U-shaped.

In addition to judging and observing the monotonicity of each variable, the ranking of the variables by their R^2 and coefficient value, provides the necessary insight into the importance of each variable in the model.

Variables entered into the model which, after running this report, fail to adhere to the requirement of monotonicity and/or have a too low coefficient value, are obvious candidates for removal from the model.

D&B's research has shown that removing variables from the model on these grounds, leads to improvements in the model performance, reduces the risk of overfitting, and therefore improves the overall model stability through time.

The report also allows subject matter experts to review, and comfort themselves with the fact that the relationship as used by the neural network to produce the prediction, are logically and economically sound and plausible.

As the network is designed to assign records to a relevant risk segment, the monotonicity assessment is performed, and reported, by the risk segment allocation as predicted by the neural network.

An example of the output of the Monotonicity Report can be found in Appendix B.

- **Report II: The Event Rate**

The "event rate" used as part of the monotonicity testing, is extracted from the calculation of the Monotonicity, and further processed and presented in an Event Rate report.

This report is rather straightforward in nature, as it just calculates the average event rate per prediction as produced by the neural network. As per the monotonicity report, this average event rate is calculated per predicted risk segment.

Next to the simple arithmetic average, also a weighted average event rate is calculated.

In line with the risk-based segmentation, the event rates allow to gauge and judge that the average event rate per variable, both weighted and unweighted, differs between the predicted risk segments.

The general expectation is that as a company is allocated – predicted- to belong to a higher risk segment, then the average event rate should be higher for those records falling into a higher risk segment. Likewise, for companies allocated to a lower risk segment, the average event rates should be lower.

To the extent that the neural network can effectively classify records into their respective risk-based segments, the event rate report can help to understand, on a variable level:

- whether an adverse relationship exists for that variable, between the average event rate of the variable and the risk group to which it is predicted to belong
- the relative size of the average event rates per predicted outcome segment

Variables for which an adverse relationship is observed, are candidates for removal from the model. As for the monotonicity, D&B's research has shown that removing such variables advances the model performance, logic and stability.

In addition, the relative size further helps to understand the relative contribution a variable will have in allocating a company to its relevant risk group.

An example of the output of the monotonicity report can be found in Appendix C.

- **Report III: Variable Interaction Report**

A key requirement placed upon any scoring system, is to understand the interaction between the variables used by the model.

The above reports, Monotonicity and Event Rate, provide an excellent and unique insight into how each variable influences the outcome of the model. However, they do not provide an insight into how variables interact with each other in the neural network.

To gain an understanding and insight into this latter, D&B leveraged its credit scoring expertise and developed an algorithm which provides a detailed insight into how variables interact inside the deep learning system.

Essentially, this algorithm assesses for each variable, the influence of each other variable on its functioning in the overall outcome (i.e., the prediction). At its raw form, for larger datasets where quite a larger number of independent variables are present, this produces a lengthy number of tables.

Therefore, an additional algorithm was developed, which summarizes the detailed output tables, into an easy-to-use index. This allows to provide a concise and direct, easy-to-use insight in the interaction between the variables, in a single table format.

When summed, the total of the individual indexes on how a variable interacts with the other variables, additionally provides a rank ordering of which variable is the most influential across all other variables.

As with the insight generated from the other reports described above, the research conducted by D&B showed that variables which have little interaction with the other independent factors, and which show individually that they add little to the predictive power, are prime candidates to be removed from the model.

Again, research showed that doing so improves the overall model performance, stability, and logic.

An example of the output of a variable interaction report can be found in Appendix D.

- **Report IV: Distribution Span Report**

As mentioned above, a key component of D&B's AI Layered scoring system is the risk-based segmentation. As indicated above, it is the prime objective of the AI Layer to classify records into one of the risk-based segments.

The above-mentioned explainability algorithms and corresponding reports provide a solid and in-depth understanding of how the deep leaning model uses the different data fields to classify companies into one of the risk segments. As such, they report on a horizontal, unified view of the importance and role of a variable in the AI model, and how these variables interact to generate the prediction.

However, next to the 'horizontal' view, it is also critical to understand how variables are vertically distributed within each risk-based segment. This in order to understand how levels of variables may be concentrated into one, or more, segments.

To this end, the Distribution Span algorithm was developed. This metric assesses the distribution of a variable for each of the risk segments. It does so by comparing the range (or 'bin') of the variable with the highest frequency of occurrence, with the range (or 'bin') of the variable with the lowest frequency of occurrence.

The absolute difference between these two frequency counts, provides what can be seen as a "span".

If a for a variable in a segment, a large span is detected, then this indicates that there is a significant degree of concentration of a certain range of the variable, for a given risk-based segment.

Combined with the average and weighted average Event Rate for the variable, given the range and the segment, an immediate understanding of the functioning of the neural network is gained.

For example, if for a low risk-based segment there is a significant distribution span (i.e. concentration) in a variable with an average event rate which is low to very low, then it follows immediately that the network has assigned companies to this low risk segment, using the part of the variable where the event rate is very low to minimal.

Obviously, this also works in the opposite direction (i.e., for high event rate ranges and segments).

Likewise, the lower the distribution span, the less concentration on a particular range (or ranges) of the variable will be present in the given risk segment.
Furthermore, the research conducted indicated that the distribution span is also related to the type of variable at hand. For example, a variable which is typically indicative of an increased level of risk, may show a larger distribution span in the higher risk segments, whereas its distribution span for the lower risk bands may be far less.

When summed, the total of the measured spans across the risk segments provides an insight ranking of the variables used in the deep learning model: the higher is the total sum, the bigger is the underlying distribution concentration and span. The lower is the total sum, the lower is the underlying distribution concentration and span. Best practice meanwhile has it that whilst the total sum provides a good starting point, it is important to look at the detailed distribution spans, combined with the average event rate.

An example of the output of a distribution span report can be found in Appendix E.

5.2.2 Model Stability and Maintenance

As indicated above, model stability (i.e. avoiding overfitting) is a very important characteristic of any credit scorecard. The reason for this is simply that users of the output the credit scoring system, will expect the model to behave in line with the performance achieved during the development of the model.

Model which behave poorly after development, or which produce widely varying performances through time, will not satisfy the needs of the users.

At the same time, models may need maintenance over time. To date, this often is a cumbersome trajectory.

On both topics, D&B's development of the AI Layered scoring system, has provided guidance and insight, and some advances towards the objective is providing stable models, driven by artificial intelligence, which are also easier to maintain than classical systems.

5.2.1.1 *Model Stability – Avoiding overfitting*

Credit Scorecards, as members of the family of econometric predictive models, generally show a lower predictive value on validation datasets than what was measured on the development sample. D&B's experience has shown that such a loss may typically range around 10 points on the GINI measurement. And this even if the traditional 70-30 rule was used during the development of the model.

The loss itself may be a disappointment for the data scientists, but D&B's experience has shown that after the "first loss", the model can continue with a stable performance for a prolonged period[16].

Many modelers will use the classical 70% development (or 'train') – 30% validation (or 'test') rule to split their data set. The general underlying concept is that a model developed on the train dataset, and which is stable in performance on the test dataset, is assumed to be not-overfitted and stable.

D&B's experience, however, has shown that this is not typically the case. Specifically not when performed on the large data sets on which D&B develops models. And specifically not when measured through time.

In this sense, D&B has leveraged its expertise in model development, and has followed the following rigorous path to ensure that the AI System would not overfit on the development sample, and that consequently would provide a stable performance through time:

- As indicated above, a good level of feature engineering was performed on the variables used in the development of the system. Expert matter expertise was as essential in this process, as any statistical measures

[16] This is particularly so in case the underlying company population remains stable in its characteristics, there is no loss of predictive data due to legal data constraints being added (e.g., such as GDPR), and an adaptive model structure is used.

- The system was built on a development sample, and immediately tested for stability through time.
- Variables causing a larger degree of instability where identified, isolated, and removed from the model

In line with the experience gained in building customized scoring systems over the past decades, D&B has, in this way, successfully developed a highly advanced credit scoring systems, which holds stable through times[17] .

5.2.1.2 Model Maintenance

As with any machine learning models or artificial intelligence models, regular maintenance and updates to the models are also to be performed on credit scoring systems.

Typically, such updates may be quite time consuming and are also resource and investment intensive.

With the advances in the processing power of IT, the increased availability of data, and the coming of age of machine learning, principles such as self-learning models have gained stronger interest over time.

Whilst there still is sometimes reluctance to set up such a self-learning AI system, it would help to overcome models which are no longer up to date, but for time and resource reasons are not being updated.

This need for easy, flexible, and effective updating was also on D&B's radar during the development of the AI Layers credit scoring system.

Whilst the choice was made not to pursue the development of a self-learning system for the development of the new AI Layered scoring system for generic scoring systems, D&B's research did point out a number of highly interesting development routes:

[17] Specifically in the case of Corona, where government lockdowns and support for businesses, have the potential to vastly impact any machine learning or AI development.

- To update the system, it may suffice to update only the deep learning algorithms. This would reduce in an important manner other adjustments and the corresponding investment in time and resources. Moreover, the time-to-market if an adjustment will also be reduced in a significant manner.

  This route surely will be pursued by D&B in case maintenance and updating is required.

- Given the controllability, explainability and stability algorithms and approaches developed by D&B for the AI Layered scoring system, D&B feels confident that several of the steps may be further automated. Crucial to this is that D&B has identified the critical parts of the system where human oversight remains essential, whilst for some other automation may help alleviate the task of updating.

  For now, this route will not be pursued by D&B for the generic scorecards but may be pursued for other applications.

6. Score Overrides[18]

Since the development of generic scorecards by D&B, D&B has employed the principles of adding score override to the system. These score overrides can be added to the system once the statistical modeling has taken place.

Score overrides are based upon events or underlying data on the companies, which either cannot, or either was not properly taken into consideration by the statistical scoring system.

Often, such overrides are based upon subject matter expertise, logic, and an in-depth understanding and experience of the functioning of a scoring system.

Previously, such overrides were performed by a rule-based approach, on the D&B Rating (which itself was derived from the D&B 1-100 Score).

[18] Manual overrides performed by business analysts on individual cases, are left out of scope in this section

Since 2008, however, R&D indicated that for many overrides, a bad rate-based approach could be followed[19].

This bad rate-based approach relies on statistical analysis, and re-set the score at the most granular level, to the score with the appropriate bad rate for the group of companies under consideration.

It must be noted that such bad rate based overrides, will only be applied to a small number of well-identifiable records.

[19] Bisnode Scoring Group, 2008

7. The Final Score & Rating

After the score override layer, the outcome generated is a Final Score.

Depending on the scale of even odds and points doubling the odds, the final score will be produced by the system.

This final score will run on a granular scale, with each score representing the probability of the future event taking place.

The granular score will most often be transformed into a "Rating" scale, which is easier to use, interpret and "read" than the detailed score. This transformation is performed by placing cut-offs on the score range.

If required, Rating overrides may also be set. Often, these will be the result of a manual intervention in the system. The frequency of occurrence will be very small in any event.

Once completed, the final Score and Rating can be provided to the user, for decision making, portfolio management, or analysis purposes.

8. Interaction with regulatory frameworks

As indicated above, the data which is held in D&B's database is governed by the proprietary D&B DUNSRight® process.

The DUNSRight® process stipulates that data must come from sources which are relevant, structured, reliable, complete and timely. It represents the core principle of D&B's global data operations and applies all data collected and stored by D&B[20].

More information on DUNSRight® can be found on:

- https://www.altares.nl/en/our-data/data-quality/
- https://www.dnb.com/about-us/data-cloud/the-dunsright-quality-process.html

As a major player in the data industry, the D&B database complies with all applicable laws and regulations.

The AI Layered Scoring system complies more specifically with:

- The EU Directive on Data Protection ('GDPR') and the country data protection laws:
    - Whilst D&B may hold some personal data in its database (e.g., certain details on the owner(s) or the management of a business), D&B does not use such data fields in its models.
- The new EU AI Act[21]:
    - D&B has welcomed this Act and has provided subject matter feedback and input into the discussions in the European Parliament.
    - Whilst not directly impacted (as D&B does not provide consumer credit scoring), D&B's standards in developing

[20] Some of the data used by D&B may come from unstructured data sources. All of them will however comply with the principles of the DUNSRight® process.
[21] EUR-Lex - 52021PC0206 - EN - EUR-Lex (europa.eu), pdf (europa.eu), TA (europa.eu)

machine learning models and AI based systems, ensure that D&B complies to the rules set forth for high risk AI systems.

- Machine Learning for IRB models, European Banking Authority[22]: D&B complies with the rules set forth by the EBA, notably on overfitting, explainability, human oversight, and relevant expertise.

22 https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Reports/2023/1061483/Follow-up%20report%20on%20machine%20learning%20for%20IRB%20models.pdf

9. Software used for the development of the AI Layered system:

For the development of the AI Layered Scoring System, D&B has used a variety of industry standard software:

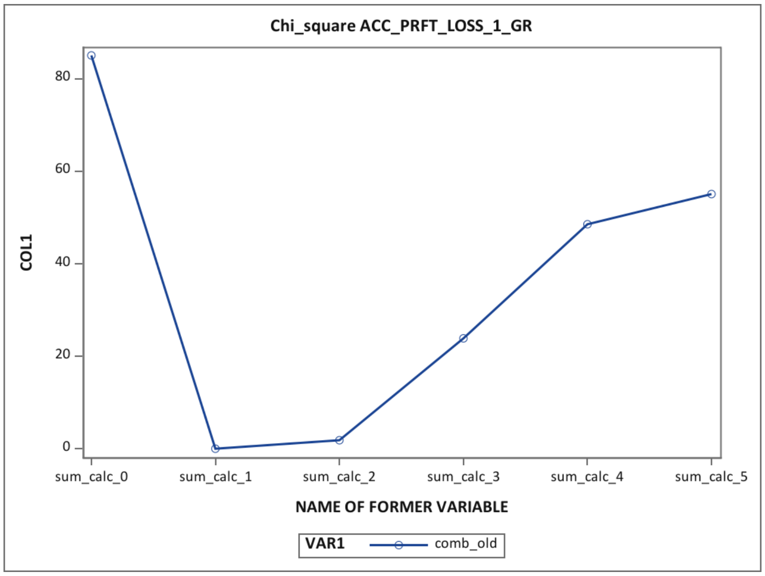
SAS Base/Analytics Pro	Data preparation & processing, feature Engineering
Python Numpy, Statsmodel	Data preparation & processing, feature Engineering, elastic net logistics regression
Keras Tensorflow	Artificial neural networks, deep learning
Quantforce	Optimized binding
SQL Server/Azure Data Studio	Data extraction
Excel	Result presentations, post-processing

10. About Dun & Bradstreet

Altares Dun & Bradstreet is market leader Benelux in collecting, processing and delivery of business enterprise data. As a business data specialist and part of the Dun & Bradstreet (NYSE:DNB) global network, offering their clients access to data from more than 500 million companies in 220 countries. Altares Dun & Bradstreet's Data Cloud solutions deliver insights that enable clients to mitigate risk, increase revenue, reduce costs, and thus improve business performance.

Appendix A - Controllability Report

The below infographic provides an example of the Chi square report, for a particular risk-based segmentation (in this example, 6 segments), for the variable 'Retained Earnings' [23]:



The below table provides a view of the list of available data fields, and their eligibility:

Base Variable	sum_chi_sq	sum_count	sum_chi_sq_bads	sum_count_bads	eligible	eligible_bads
age_bus_GR	1.721	1.000	1.545	1.000	0.000	1.000
age_summon_GR	2.809	2.000	0.557	1.000	1.000	1.000
AVG_CURR_OVR_BAL_GR	13.026	2.000	1.987	1.000	1.000	1.000
AVG_CURR_OVR_BAL_L24M_GR	13.149	2.000	2.190	1.000	1.000	1.000
CURR_PAYD_SCR_GR	9.928	2.000	4.715	1.000	1.000	1.000
EXPR_L1_L30_L12M_GR	11.680	1.000	10.079	1.000	0.000	1.000
EXPR_L1_L30_L24M_GR	11.608	1.000	10.079	1.000	0.000	1.000

Note: only 7 variables are shown. The variables are not ordered in any particular way.

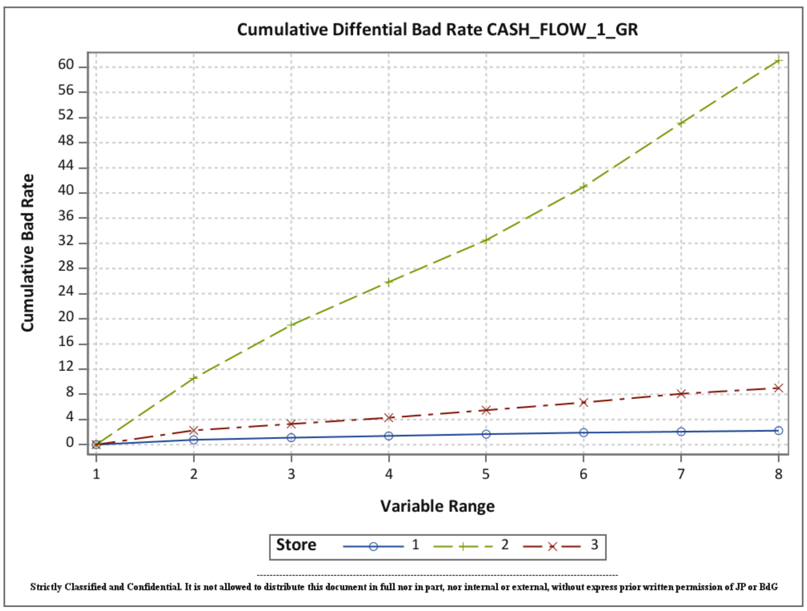
[23] Source: generic scorecard development, D&B Belgium, 2023

Appendix B - Variable Monotonicity Report

The below infographic provides an example of the Variable Monotonicity Report, for the variable 'Cash_Flow'[24] .

A.1: Monotonicity metrics

Risk Group	Coefficient of slope	Rsquare
Label of model	rownumber_I	R-squared
m_CASH_FLOW_1_GR1	0.2945238	0.9346125
m_CASH_FLOW_1_GR2	8.3683333	0.9954671
m_CASH_FLOW_1_GR3	1.2325	0.9896917

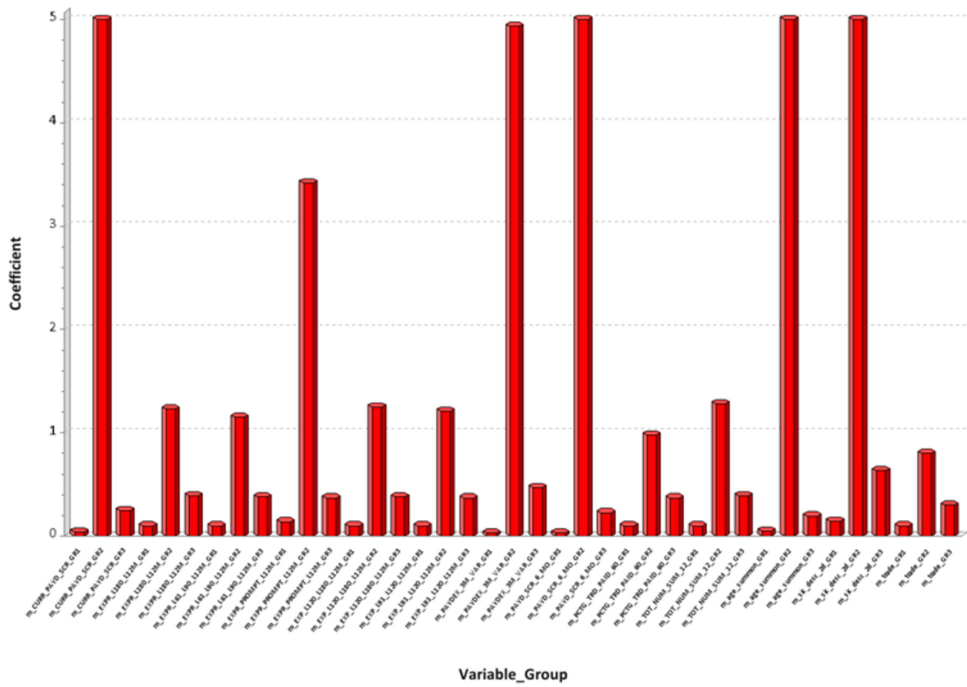


[24] Source: generic scorecard development, D&B Belgium, 2023

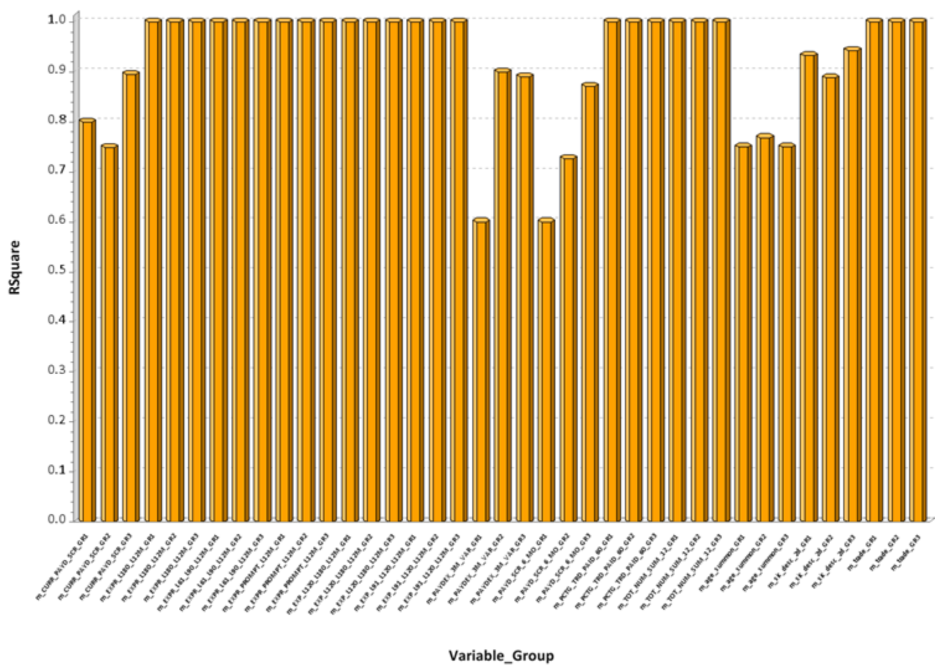
A2: Anova Table

Comparisons significant at the 0.01 level are indicated by ***.				
prediction Comparison	**Difference Between Means**	**Simultaneous 99% Confidence Limits**		
0 - 2	0.056600	0.043377	0.069824	***
0 - 1	1.294631	1.278609	1.310653	***
2 - 0	-0.056600	-0.069824	-0.043377	***
2 - 1	1.238031	1.219058	1.257003	***
1 - 0	-1.294631	-1.310653	-1.278609	***
1 - 2	-1.238031	-1.257003	-1.219058	***

A3: Coefficient Overview Graph for all variables in the neural network



A4: Rsquare Overview Graph for all variables in the neural network



Appendix C – Event rate report

The below infographic provides an example of an Event rate report[25].

The higher is the value in the cell for a given variable, the lower is the event rate for that variable in the cell.

The lower is the value (or the more negative), the higher is the event rate for that variable in the cell.

In this example, for the prediction of business failures in for Belgium, there are three risk-based segments. The numbers in the cells provide an insight into the average event rate per variable, per segment.

For example, for the variable Evolution of Working Capital over a 3-year historical period, the event rate for the least risky segment is 0.46. In contrast, for the segment with the highest risk level, the value is -0.42. This shows a clear difference in the variable allocation to the relative segments.

For the segment in between the least risky and the riskiest segment, the value is 0.44. This indicates that for this variable, the network is less able to separate this segment from the other two.

VAR	_0	_1	_2
Evol_WrkCap_y1y3_1_GR	0.4659887	-0.422126	0.4432407
ICR_jp_1_GR	0.5039976	-0.571124	0.5799194
NET_OP_PROFLOSS_1_GR	0.5774828	-0.617041	0.628553
NET_PROFIT_LOSS_1_GR	0.6025741	-0.607131	0.6550832
NET_WORTH_1_GR	0.8281382	-0.64048	0.4912046
PREPAY_ACRRUED_1_GR	0.1948874	-0.155143	0.107227
PROF_LOS_BF_TAX_1_GR	0.6039087	-0.616503	0.6399426
QK_RATO_1_GR	0.6124346	-0.575681	0.56725
RES_AFT_TAX_1_GR	0.5945039	-0.607335	0.6574195
RTN_ASET_PRETAX_1_GR	0.5812708	-0.620899	0.6413705
SOLV_RATO_1_GR	0.5998227	-0.580422	0.5792969
Surv_rat_y1_y2_1_GR	0.0355215	-0.001217	0.0412965
Surv_ratio_2ND_1_GR	0.0521363	-0.023838	0.0343506
Survival_ratio_1_GR	0.0820876	-0.015188	0.1088024
TANGIBLE_ASET_1_GR	0.4230679	-0.428703	0.2167012
TANG_NET_WRTH_1_GR	0.817543	-0.631958	0.4771626
[illegible]	[illegible]	[illegible]	[illegible]

[25] Source: generic scorecard development, D&B Belgium, 2023

Note: only 16 variables are shown. The variables are not ordered in any particular way.

Appendix D – Variable Interaction Report

The below infographic provides an example of a variable interaction report[26].

The values in the cell indicate how, in the network, the 'EVALVAR', is influencing the 'Base variable'. The higher the cell value, the higher the influence.

The term 'influence' relates to the 'loss of events' for the Base Variable, for various levels of this variable, due to the values if the Evalvar. For ease of representation and interpretation, the 'loss values' are indexed to a single metric.

The Sum-line, is a simple sum for each Evalvar. The higher is the value, the more the variable will 'influence', in the network, the other variables. Likewise, the lower the value, the less the variable will 'influence', in the network, the other variables.

In this example, the D&B Paydex® has the highest total sum. This indicates that this variable, when present, will have a strong influence on the model performance. By contrast, the Paydex Trend seems (far) less powerful.

The variables related to Social Security Summons seem less important. However, from a subject matter expertise, these variables known to be very predictive. The low Sum is due to the sparse occurrence of this variable on the Belgian business universe[27].

	Variable Section						
	EVALVAR						
Base Variable	CURR_PAYD_SCR_GR	Paydex_trend_GR	TOT_NUM_SUM_36_GR	TOT_NUM_SUM_60_GR	age_bus_GR	age_summon_GR	sic_desc_2d_GR
CURR PAYD SCR GR	.	-0.133	0.279	0.386	1.017	0.400	1.061
EXPR PROMPT L12M GR	1.321	0.154	0.201	0.295	0.748	0.301	1.243
EXP L120 L180 L12M GR	2.305	0.348	0.177	0.233	0.177	0.245	0.597
EXP L120 L180 L24M GR	1.791	0.180	0.185	0.251	0.707	0.261	0.821
EXP L91 L120 L12M GR	1.456	0.018	0.168	0.210	0.486	0.215	0.709
NO OF EXPR L12M GR	1.923	0.146	0.159	0.223	0.833	0.227	1.124
PAYD SCR 12MA GR	1.317	0.452	0.190	0.269	0.901	0.271	0.974
PAYD SCR 6 MO GR	0.839	0.138	0.192	0.276	0.975	0.282	0.978
PCTG TRD PAID 60 GR	1.507	-0.015	0.144	0.193	0.825	0.197	1.029
PCT TRDPAID60 L24M GR	1.355	0.020	0.137	0.192	0.781	0.196	0.998
Paydex trend GR	2.147	.	0.214	0.299	0.682	0.293	0.735
TOT NUM SUM 36 GR	1.641	0.176	.	0.033	0.853	0.241	1.241
TOT NUM SUM 60 GR	1.671	0.113	0.944	.	0.839	0.198	0.926
age bus GR	0.398	0.064	0.236	0.385	.	0.386	1.225
age summon GR	1.717	0.215	0.176	0.007	0.580	.	0.631
sic desc 2d GR	0.669	0.091	0.250	0.391	1.030	0.392	.
SUM:	22	2	4	4	11	4	14

Note: only a few variables are shown. The variables are not ordered in any particular way.

[26] Source: generic scorecard development, D&B Belgium, 2023

[27] The importance of these variables will be less visible in this report. In the other reports, they will however be shown to be very predictive. Hence the need to review the Reports in combination.

Appendix E - Distribution span report

The below infographic provides an example of a distribution span report[28].

The higher the values in the cell, the higher is the degree of concentration, across the different ranges (values) of a variable, for each risk-based segment.

The 'Sum' column is the simple sum of the individual cell values for a given variable.

In this example, it can be seen that a variable related to the social security summons, has the highest Sum of the distribution spans. As the table below is ordered by the 'Sum' (in descending order), this suggests that this variable has the strongest influence in the functioning of the neural network/deep learning (note: as would be expected by the subject matter experts).

VAR	_0	_1	_2	sum0
age_summon_GR	69.1761	50.130268	41.132289	160.43866
TOT_NUM_SUM_12_GR	66.627289	48.430328	18.196961	133.25458
TOTAL_LIABS_1_GR	48.371207	35.968566	45.909201	130.24897
TOT_ASSETS_1_GR	48.387916	32.734265	45.688063	126.81024
TOT_NUM_SUM_36_GR	62.171778	32.65389	29.517889	124.34356
NET_WORTH_1_GR	49.040491	34.530942	40.515708	124.08714
solv_rat_jp_1_GR	47.150419	33.332925	41.985201	122.46855
DEG_IDTD_1_GR	48.789944	32.780371	40.009574	121.57989
TANG_NET_WRTH_1_GR	46.182403	34.062464	40.283037	120.5279
TOT_NUM_SUM_60_GR	58.402707	24.18169	34.221017	116.80541
cash_flow_jp_1_GR	47.654936	34.257322	34.369236	116.28149
EBITDA_jp_1_GR	47.791573	34.303733	32.816473	114.91178
CASH_FLOW_1_GR	46.040017	33.903279	33.320952	113.26425
solv_rat_jp_2nd_1_GR	44.505489	29.733305	37.093105	111.3319

Note: only 13 variables are shown. The variables are ordered by the column 'Sum0', descending.

--------------- End of Document ---------------

[28] Source: generic scorecard development, D&B Belgium, 2023