CREDIT SCORING REVIEW

Best Practices in Credit Scoring

Credit comes from a Latin “cred” that means “to believe” or “to suppose.” Thousands of lenders are already establishing their scorecards in-house, making credit scoring truly international. In addition, several more banks utilize third-party solutions to generate and then use credit risk scorecards in-house. However, the pattern has shifted away from outsourcing scorecard development towards internal designs.

A Basel II Accord has been the single most important factor encouraging banks to take credit scorecard creation in-house during the last decade (Abrahams & Zhang, 2009). Banks that chose to (or have been forced to) adhere with Basel II’s Foundation or Advanced Internal Ratings Based methods were required to produce PD (Probability of Default) estimations internally (and also estimates for LGD (Loss Given Default) and EAD (Exposure at Default).

Probability of Default means that a likelihood in which a borrower would be not able to complete scheduled repayments over such a particular time period, generally one year.

Expected losses allow decision makers to add and alter depending upon the specific needs of business internal rules and regulatory guidelines. As a result, it is utilised to develop analytics and decrease misconceptions or issues. The amount of money lost by a bank and other financial institution if a borrower fails on such a loan, expressed as a total exposure percentage just at moment of default, is known as the loss given default (LGD). Exposure at Default (EAD) defines that the expected amount of loss that a bank might face if a debtor fails on a loan.

These all three EAD, PD and LGD are used to calculate the expected losses. Just for argument, let us assume that loss per account for such a cell was $500. Whereas if default probability was 6% as well as loss given default was 97 percent of limit, therefore maximum credit limit with this cell could be computed as follows (Siddiqi, 2017):

Expected loss = EAD * PD * LGD

500 = EAD * 0.05 * 0.97

EAD = $10,309

The Basel EAD and LGD estimations were related to the severity of losses sustained during an economic downturn, whereas the accounting EAD and LGD models provide some neutral estimate projections of economic conditions. There are also other variations, like the idea that Basel LGD calculation includes collection expenses that is not added in the budgeting ECL models ("Regulatory treatment for accounting provisions", 2016).

Larger banks were forced to demonstrate their credit scoring competency as they increased their production and the use of credit scoring (Siddiqi, 2017).

Many organizations that have not been obligated to adhere to Basel II, like automotive loan companies and retail credit cards, make that choice anyway. Many would see it as a method to demonstrate their marketability and expertise and a full endorsement of their internal systems’ durability. Those who have seen Basel II adherence as a collection of best practices contributing towards an opportunity to improve their internal systems, instead of a mandated regulatory activity, benefited the most.

Credit scoring provided a simple and effective approach to use data to reduce losses and increase profits (R. Abrahams et al., 2000).

The SEMMA (Sample, Explore, Modify, Model, Assess) methodology is a good example of best practices used in credit scoring. SEMMA has been used as a structured, operational toolset, and is claimed to be so by SAS as part of the SAS Enterprise Miner programme. SEMMA refers to a sequential method for building machine learning algorithm. It is a world's largest producers in commercial statistical as well as business intelligence software. The sequential steps, of Sample, Explore, Modify, Model, and Assess, direct the growth of such a machine learning model ("SEMMA Model - GeeksforGeeks", 2021).

Therefore, various feature selection algorithms are used for the data pre-processing on credit scoring best practices. And the commonly used algorithms were relief, gain ratio, and chi-square (Piramuthu,2006).

Creditworthiness scoring has been among the oldest methods both in data analytics and risk management, and it is essential towards evaluating credit applications (Thomas et al., 2017). The systematic use can be traced back to the 1950s when it was first used to control risk and diversify a loan portfolio. Data mining was also utilized for credit scoring and supporting tools; it was among the first consumer behavioral data analytics applications. Credit scoring, however, was used much older, in 1820, according to the literature. Credit reporting started to develop in the 1820s due to the widespread adoption of lending and the necessity to adapt regulatory norms towards the new market environment. New bankruptcy rules were made loans a dangerous transaction, prompting efforts to standardize credit evaluation methods. The trade agency, founded by such a merchant Lewis Tappan in early 1841, sought information on debtors' assets from customers throughout the country - this was in the aftermath of the first economic crisis triggered by such a financial collapse (Morawski, 2003).

The credit score systems described above were completely subjective and relied on human judgment; while rating creditworthiness, analysts have only been influenced by their professional and experienced knowledge (Thomas et al., 2017). According to reports from the time, while awarding loans, racial criteria should be considered (Cohen, 2012). Such behavior was inappropriate by today's regulatory norms, and transparent quantitative metrics and characteristics must back up credit decisions. Furthermore, it must be noted that the credit industry was substantially smaller during this period, and businesses accounted for the majority of loans. At the time, loans were generally provided based on the principle of limiting losses: mainly safe transactions were co-financed by loans, and these transactions had a secure system (Thomas et al., 2017). Credit scoring originally describes the process of banks accepting loan applications (Oesterreichische Nationalbank, 2004).

On the other hand, credit scoring does not have to be associated solely with the banking approval procedure. It can be used in several different processes in which the client signs a contract, very often having committed to regular financial obligations (that is, telephone subscription, television), and it has to be pre-assessed throughout preparing the best contract terms such that the organization providing services does not risk far too much. The score analysis is a very well example of data analytics in a simple business approach in the form of Big Data.

Credit rating would be a pragmatism and empiricism-inspired activity that involves estimating a borrower's creditworthiness (in the context of modern macroeconomic conditions) (Thomas et al., 2017). The primary purpose of credit scoring would be to estimate credit risk projections in a way that is both successful (about the quality of returned predictions) yet efficient (throughout relation to the resources committed). Objectivity basis of statistical modeling methodologies and empirical data regarding consumer behavior has been the primary source of credit scoring improvement. The philosophical tendencies outlined above encourage the use

of accessible data about the consumer and his environment to boost the effectiveness of scoring models. Yet, certain regulatory standards expressly restrict the use of sensitive information (such as gender or ethnicity) like a criterion for assessing to choose whether or not to provide a loan.

Classical Scoring Approach

Credit scores were developed in the 1950s by such a group of mathematicians founded by Bill Fair then joined by Earl Isaac - as Fair, Isaac as well as Company, now recognized as FICO, is established to develop a uniform credit scoring model (Herron, 2013). The system's earliest versions were based on a manual, paper-based approach (now referred to as a scorecard). Because of the restricted development of computer science only, the scoring procedure can not be entirely automated; scorecard scores have mostly been calculated by groups of professionals (Thomas et al., 2017). The core principles of this system are aimed at increasing the objectivity of the scoring process while simultaneously reducing costs (resources and time). Lenders never longer wanted to hire a large number of analysts that undertake expert-based scoring — which used to take days or weeks in human-based scoring could then be completed in seconds or minutes (assuming data availability). As researchers know it now, the credit scoring system relates to a statistical method to scoring that usually entails the production of credit scorecards (Anderson, 2007; Matuszyk, 2004; Thomas et al., 2017). Generally, Classical approaches are based on statistics, probability, decision analysis, etc.

Credit scoring models were statistical predictive models used to anticipate future events, such as a customer's default. Specifically, the basic logistic regression is a function that inputs several variables into a credit score used by most banks. The logistic regression model's success stems from its ease of use and interpretation - a factor that is gaining increasing attention from both regulatory and business perspectives. Logistics regression has been frequently employed in banking, from IFRS 9 PD computation to capital needs assessment (PD models used in Basel III framework for Risk-Weighted Assets calculation) (Kamiński et al., 2020).

IFRS 9 outlines how a business may categorize and analyse financial assets and liabilities, as well as certain contracts to acquire and trade non-financial goods (Standards et al., 2021). Basel I would be a set of international banking laws developed by Basel Committee on Bank Supervision (BCBS) which sets the basic capital requirements for financial institutions in order to reduce credit risk. Internationally operating banks were required to have a minimum amount of capital (8 percent) calculated on the basis of risk-weighted assets. Basel II extended the standards of minimum capital requirements set by Basel I, a first international regulatory pact, and introduced a framework of regulatory scrutiny, and also transparency requirements for assessing banks' capital adequacy. Basel III would be a 2009 global regulatory agreement that enacted a series of measures aimed at mitigating danger as in global banking sector through mandating banks to maintain adequate leverage ratios to retain higher level of the reserve assets upon deposit ("Investopedia", 2021).

Following the discovery of the logistic regression model and the advancement of computers as the greatest analysis tools, practitioners began to employ scorecards based mostly on the logistic regression approach. The analyst may opt for a classic logistic model and perhaps a risk scorecard depending on their choices or bank policy.

Drawbacks of the Classical Approach

Apart from obvious benefits (i.e., ease of installation, suitability, and interpretability), classical credit scoring techniques have several drawbacks. Below are the few essentials (Johnston & DiNardo, 1997; Kennedy, 2003; Molnar, 2019):

- Classical models require a time-consuming procedure of variable servicing, correction, and selection to the model at the modeling stage; this typically requires the involvement of an entire group and even department throughout banks that handle those systems; furthermore, such efforts were also typically multiplied each of the numerous scoring models (i.e., products, clients) which are used in the bank (Kamiński et al., 2020).
- The models are linear and have a rigorous functional form. The basic linearity of the model assumes that characteristics and the target variable have a linear (typically affine) relationship. Such a property is widely regarded as a benefit, namely, model interpretability, well-known properties, simple model maintenance and implementation, additivity, and it is also widely regarded as a disadvantage, i.e., incapability to model complicated functional relations, which requires massive input transformations that represent non-linearities (Kamiński et al., 2020).
- Interactions among particular features that predict an intended outcome after taking into account individual feature impacts (for example, distinct consequences of Income and age variables which are not seized by the sum of partial total score); throughout principle, classical models permit for modeling interactions among features via simply introducing new features that were compositions. However, this leads to an increase in data handling and management challenges (as in behavioral ABT, the total actual number of variables – real-valued and dummies – might be in the thousands) (Kamiński et al., 2020).
- Some variables, particularly those changed from categorical to dummies, must be included or excluded in groupings within the credit scoring framework and needed model interpretability. Furthermore, the dummy coding, which employs the biggest cardinality, namely its most observations' coverage, is recommended as the reference category (Kamiński et al., 2020).

Weight of evidence

The Weight of Evidence (WoE) transformation was a very well concept (Siddiqi, 2012).

$$W_O E_k = \ln\left(\frac{G_k/G}{B_k/B}\right) == \ln\left(\frac{B}{G}\right) - \ln\left(\frac{B_k}{G_k}\right) \quad (1)$$

$$W_O E_k = Logit - Logit_k \quad (2)$$

As eqn. (1) & (2), was extremely comparable to logit value. In which k denotes every variable category, B, G represents the total number of bads and goods clients as in population, while B_k and G_k Denote the number of bads and goods clients. As a result, the difference between the overall population logit and the category logit has been the Weight of Evidence for such a category. The WoE method could also be used as a logit method.

Logistic regression is a technique for predicting the outcome of (LG). The logit model, suggested by Berkson (Berkson, 1944), considers a set of explanatory variables of $X = \{X_1 \dots \dots, X_P\}$, and a response variable having two categories. The logistic regression technique, $Y = \{y_1, y_2\}$ entails estimating a linear combination of X and also the logit transformation of Y. As a result, if y_1 is the category of interest for study, then the model can be formulated

as $\log\left(\frac{\pi}{\pi-1}\right) = X\beta$, wherein $\pi = P(Y = y_1)$ and β is also the vector holding the model's coefficients. The model could also be expressed as in eqn. (3),

$$\pi_i = \frac{\exp(X_i\beta)}{1-\exp(X_i\beta)} \quad (3)$$

wherein π_i would be the probability of ith individual belonging to a category y_1, dependent on X_i. The logistic regression model was a very well technique that is frequently compared to other methods (Li & Hand, 2002; Hand, 2005; Lee & Chen, 2005; Abdou et al., 2008; Yap et al., 2011; Pavlidis et al., 2012; Louzada et al., 2011) and utilized in technique mixtures [105]. Regularized logistic regression or local logistic regression are two further feasible and specific approaches to logistic regression.

WoE or Logit transformation

Every variable gets turned into an interval variable with WoE and Logit values in categories based upon those created categories. This concept was particularly useful for avoiding several issues in nominal coding variables, like missing imputation, the sensitivity towards outliers, and dummy variables. When opposed to dummy coding, the WoE transformation allows for more degrees of freedom to be saved — as in the case of behavioral scoring and application scoring relating to credit bureau ABT. Dummy coding can substantially lessen the number of observations that can be used in estimation. Furthermore, the WoE facilitates traceability or modification (namely, negative betas problems). The WoE enables a smoother evaluation of the volume of distribution in terms of low samples (dummies were generating more focus upon the tails).

Machine learning and its applications

On the other hand, machine learning is a set of techniques designed to solve computationally difficult pattern-recognition issues in massive datasets. Because of the high sample sizes and the intricacy of possible correlations between consumer transactions related attributes, such techniques, including radial basis functions, support vector machines, and tree-based classifiers, are excellent for The forecasting model would be created using any appropriate forecasting approaches after the pattern has been detected. Some of the applications of machine learning-based systems to consumer credit were examined in depth by (Shi et al., 2013) and (Bellotti & Crook, 2009).

Credit scoring models were developed to distinguish potential borrowers like creditworthy and default candidates based on financial situation or credit performance documented in such an application form or with a credit reference bureau. Because even a little increase in credit scoring accuracy can result in significant investments and savings (Huang et al., 2007). Statistical procedures like logistic regression (LR) or linear discriminant analysis (LDA) were basic and easy to understand, and they use a threshold upon that underlying likelihood of default that rejects customers who have a posterior probability of default below a certain threshold (Bellotti and Crook, 2009). Yet, their discrimination capacity remains debatable due to the non-linear relationship between default probability and credit patterns. Artificial neural networks (ANNs) as well as support vector machines (SVMs), for example, were particularly highly adapted for dealing with non-linear situations and frequently functioned in such a data-driven manner even without constraints on prior probability hypotheses (Pang et al., 2011; Wang et al., 2011).

Due to the solid theoretical foundation with attractive classification performance, SVMs, inspired by statistical learning theory (Vapnik, 1995) (Vapnik,1998), were used to various credit scoring difficulties. Using eight real-life credit scoring databases, (Lessmann et al., 2015)

investigated the effectiveness of different classifiers, while SVM surpassed many of the other strategies. Thomas et al. (2005) studied 17 consumer credit modeling methods and found that SVM had been the best. However, the classification result remained insufficient for practical usage. Huang et al. (2007) developed hybrid SVM-based credit scoring methods using three feature selection and parameter tuning procedures. Yet, removing characteristics may result in information losses and lower classification accuracy. Martens et al. (2007) derived rules from such a trained SVM that produced credit scoring models that were both efficient and understandable. However, compared to the black box system from which they are derived, the extracted rules lost a very small percentage of their accuracy.

SVMs effectively dealt with such a high number of defining characteristics, known as "features," but having too many features can lead to over-fitting modeling. (Pal et al., 2000; Guyon et al., 2002). Yet, because there are generally few features in the credit scoring issue to start with, and the input features were mutually independent, simply removing and eliminating any features is improper. Furthermore, when employing traditional SVMs to develop credit scoring models, all input features were considered equal and given the very same modeling contribution, regardless of their varying effects upon that output grant decision. A unique feature-weighting algorithm has emerged (Blum and Langley, 1997; Yeung and Wang, 2002; Wang et al., 2004).

Variable selection methods

Below are the most widely used variable selection methods. Therefore, the methods are to be discussed are given as:

Principal component analysis and autoencoders

The Principal Component Analysis approach is among the most widely used dimensional reduction techniques (PCA). It is dependent upon the eigenvectors of variance-covariance balanced features matrix generating a linear transformation among all variables. As a result, this strategy creates new variables, which are linear combinations of existing ones.

It should be noted that the kernel principal component analysis (kernel PCA), which is an extension of the traditional PCA, can be used in theory. This add-on has to do with the use of kernel functions (Hofmann et al., 2008). Whenever the basic PCA fails to reduce the dimensionality due to its non-linear correlations among features, the kernel PCA comes in useful. Kernel PCA was generally advised when basic PCA fails to function well. It depends upon the classical credit scoring approach.

Stepwise methods of variable selection

Backward stepwise selection (beginning with a whole feature set and reducing the least significant important features throughout subsequent steps) and forward stepwise selection (starting with no features, then adding common statistically significant in the following steps) are the two stepwise methods. These are also based upon the classical credit scoring approach. In theory, such strategies would result in an incremental improvement in the overall model's performance, and at the very least, maintain its current level. The algorithm must be turned off whenever the model's predictive power starts to deteriorate (test error rises) (Scallan, 2013; Scallan, 2011).

Decision trees and random forests

The non-linear relationship between the variables is considered (via the building of numerous tree models) in the random forest and gradient boosting models, including all variables. Gradient boosting would be especially good at detecting non-linear relationships among variables. Despite random forests, which perform best using deeper decision trees, gradient

boosting usually works with decision trees that are a little over two or three layers deep, which would be comparable to how ABT features are used. Hence, decision trees and random forests are based upon the modern credit scoring approach.

Comparison Between Classical And Modern Credit Scoring Approach

A machine learning framework demands that the initial data be divided into three samples: training, validation, and testing sets. The approach contrasts with classical econometrics. The original data set is separated into a training set and a testing set (thus, only one model was constructed from the start, and only one test set was conducted to assess its performance from out of the sample). Because the space of the hyperparameters in machine learning is larger than in classical models (namely, machine learning models had more average training parameters than the classical models), there must be extra data space wherein models can be evaluated out-of-sample the best one chosen.

In terms of the data missings (the MCAR as well as the MAR), it is recommended to a) the impute missings (there are several potential solutions discussed in the literature, such as population average, the median and dominant, the model fitting; testing several approaches as well as selecting one which gives the highest predictive power of the model was suggested) and b) form the dummy variable (0 was no missing for the particular characteristic, and also one is missing happens). This method also enables us to add potential MNAR information into models. On the other hand, machine learning methods, such as XGBoost, random forest, and decision trees have the benefit of handling missings consistently and efficiently. For instance, in XGBoost v1.2.0.1, missing values are learned by default throughout training (namely, branch directions with missing values were learned during training) (Obviously, one must be cautious if training and test sets have distinct missing distributions.)

The statistical approach for discriminant analysis, developed by Ronald A. Fisher, laid the scientific foundation for current credit scoring (Fisher 1936). Whenever the common features of group members were unobservable, discriminant analysis was a statistical approach that used distinguish among groups in the population using quantifiable attributes.

Traditional databases are used to hold structured data. A structured database, for example, could keep track of daily business operational transactions. Unstructured data does not normally have a predetermined sequence. Examples are freeform text, audio, video, social media information, photos, and unstructured data. Semi-structured data somehow doesn't follow the structure of structured data in any way. Instead, they have tags and marks on them.

Unstructured data's utility, objectivity, and quality for improving credit scoring techniques have yet to be proven (C Dudley & Knot, 2017). The legal utilization of unstructured data bears the promise of future investigation. Many regulatory agencies, including Australia, the UK, Singapore, and those Taiwan, and China, have created sandbox environments that encourage data innovation (FCA, 2021).

Machine learning approaches are less susceptible (namely, more resistant) to data-related difficulties than classical scoring systems (namely, collinearity and data occurrence gaps).

Stable variables, completely uncorrelated, having sufficient predictive power, and a low Variance inflation factor (VIF), were chosen as the classical approach towards the model building. The predictive ability for single variables and the lack of connections behind them are less important for non-classical machine learning approaches. Some of the model's criteria for variables can be avoided by employing regularisation methods and algorithms like decision

trees or random forests. In machine learning, systems with low individual prediction ability might perform well by combining several associated variables (Kamiński et al., 2020).

In classical procedures, one would have included or removed a certain variable from the model. In the case of methods discussed throughout this monograph, we enable (in certain situations) steady increases in parameter values while accumulating "material evidence" for predictive capability. A prevalence of near multicollinearity for conventional creditworthiness models would lead to unbiased results. However, it will be impossible to reliably convey to the client the exact rationale underlying individual credit choices (acceptance and rejection of the loan application). In this regard, as previously mentioned, ensuring reliable estimates of parameters (and lower variance of the model's parameters) is crucial and linked to the technical resilience of the scoring systems.

Obstacles Present in Classical And Modern Credit Scoring

Classical and alternative approaches for Validating scoring models

Due to the interpretability of credit decision requirements, non-classical analytical approaches that create scoring models are still quite rare. Alternative analytical methods, on the other hand, would be increasingly applied in risk assessment. It would be due to a combination of factors, including the models' great efficiency and the rapid development of techniques that enable us to comprehend and analyze the behavior of models previously viewed as black boxes. The European Banking Authority (EBA, 2021) report confirms this, stating that the growing usage of big data was intimately related to advanced analytics and machine learning methods. The validation of the resulting models is influenced to some extent by the use of non-classical analytical approaches.

The first distinction can be seen as early as the preparation and selection of variables in building the model. Even though the methods utilized for such a goal have not changed considerably, the approach to the variable selection problem changed. Investigating the association among variables and their discriminatory power was one of the parts of data validation during the model designing stage. The use of variables with high discriminatory power and not correlated one classical modeling approach required another; thus, these features were verified. Although strongly linked variables having low individual predictive power, non-classical approaches could be applied appropriately and successfully. The need for stability of variables used to generate the model remains the same regardless of the techniques used.

The procedures for validating the overall discriminatory power of the produced model are unaffected by various model creation approaches. All indicators or curves indicating discriminatory power described above are model agnostic, meaning they may be used in any classification model with such a binary dependent variable. It is similar to the Gini coefficient for variables. Shapley values associated with PFI calculations of variables are a new dynamic of validation that developed with the advent of non-classical modeling approaches. They demonstrate how individual variables contribute to the model's discriminating power.

A change was proposed for such calibration testing. It entails substituting the study underlying statistical significance for the analysis of the problem's materiality. This strategy also was discussed in the ECB publication (ECB, 2017). In the perspective of Risk-Weighted Assets %, the ECB calculates the overall applicability of quantitative effect as in portfolio. It is a novel approach to validation procedures. However, it is not the consequence of using alternative machine learning algorithms. The fundamental cause for modification is an issue with the actual realization of sample assumptions that underpin statistical testing, such as default independence within rating class and true DR stability.

The approaches for determining population or model stability are essentially the same — they involve comparing distributions over time. The use of Shapley values that evaluate the strength of the impact of variables mostly on the model's discriminatory power has been the most significant modification throughout this category of analyses connected with the use of the non-classical approaches.

Negative Betas Problem In Credit Scoring

Feature selection approaches are used when models were created like a classical logistic regression, including logit transformation with categorization among all variables (binning).

Linear regression (LR)

Although the response variable would be a two-class issue, linear regression analysis was applied for credit scoring systems. The technique establishes a linear relationship among borrowers' attributes $X = \{X_1 \dots \dots, X_p\}$ and also the target variable Y; such continues to follow:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \dots . . + \beta_p X_p + \epsilon \qquad (4)$$

In eqn. (4), where ϵ is the random error that is independent on X. The traditional approach for estimating $\beta = \beta_0, \dots \dots ., \beta_p W$here β would be the estimated vector, which is ordinary least squares. When Y becomes a binary variable, the conditional expectation E(Y|X) = x′ β can distinguish between good and poor borrowers. The result of the model cannot be taken as just a probability since -∞ < x′ β < ∞. Hand and Kelly (Hand, 2002) developed a linear regression-based super scorecard model. Scorecards are used in the credit scoring to provide a weight to every attribute in final customer score. Super scorecards are essentially multiplicative combinations of numerous additive models. In the simplest case, a two-component super scorecard will appear like this in eqn. (5):

$$S = S_1 S_2 = \sum_{i=1}^{p} a_{ir_i} \sum_{i=1}^{p} b_{ir_i} \qquad (5)$$

Where, a_{ir_i} & b_{ir_i} are the variable parameter's.

Karlis and Rahmouni (Karlis & Rahmouni 2006) provides Poisson mixture models to examine individual loans' credit-scoring behavior. Several researchers (Banasik et al., 2003; Efromovich, 2008) have worked on linear regression models and their generalizations for credit scoring.

Simpson's model states that the researcher may get opposite results depending on how well the data was classified or which variables were included in the model (Simpson, 1951). Simpson's paradox would be a statistical phenomenon in which a trend occurs in multiple independent sets of data then disappears even reverses when such groups are joined. It has been seen in several scientific disciplines for decades (Selvitella, 2017). The beta coefficients for the WoE transformation could be negative. The negative betas suggest inter-variable interactions that are undetectable as in single factor analysis used to calculate WoE. If negative betas were estimated, the variables positively affect that target like a single factor but have a negative impact when combined with other variables (named as Simpson's paradox).

In regression analysis, a beta coefficient represents the change as in outcome variable for such a unit change in an independent and predictor variable. The negative beta coefficient means that for unit change in independent variable, a dependent variable decrease. N-beta is the number of the negative betas in a logistic regression approach; this metric helps us capture

variable confounding; as in the case of the negative betas, we receive different answers for such a single variable design and the same variable for the model with multiple variables.

References

1) Abdou, H., El-Masry, A., & Pointon, J. (2007). *On The Applicability Of Credit Scoring Models In Egyptian Banks*. Banks and Bank Systems, 2(1), 4-20.
2) Abdou, H., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems With Applications*, *35*(3), 1275-1292. https://doi.org/10.1016/j.eswa.2007.08.030.
3) Abrahams, C., & Zhang, M. (2009). *Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors*. John Wiley & Sons.
4) Anderson, R. (2007). The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation. Oxford University Press.
5) Banasik, J., Crook, J., & Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal Of The Operational Research Society*, *54*(8), 822-832. https://doi.org/10.1057/palgrave.jors.2601578.
6) Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems With Applications*, *36*(2), 3302-3308. https://doi.org/10.1016/j.eswa.2008.01.005.
7) Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems With Applications*, *36*(2), 3302-3308. https://doi.org/10.1016/j.eswa.2008.01.005.
8) Berkson, J. (1944). Application to the Logistic Function to Bio-Assay. *Journal Of The American Statistical Association*, *39*(227), 357. https://doi.org/10.2307/2280041.
9) Blum, A., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97*(1-2), 245-271. https://doi.org/10.1016/s0004-3702(97)00063-5.
10) C Dudley, W., & Knot, K. (2017). *FinTech credit Market structure, business models and financial stability implications*. Fsb.org. Retrieved 5 October 2021, from https://www.fsb.org/wp-content/uploads/CGFS-FSB-Report-on-FinTech-Credit.pdf.
11) Cohen, M. (2012). Cotton, Capital, and Ethnic Networks: Jewish Economic Growth in the Postbellum Gulf South. American Jewish Archives Journal, 64, 112–36.
12) *EBA REPORT ON BIG DATA AND ADVANCED ANALYTICS*. Eba.europa.eu. (2021). Retrieved 5 October 2021, from https://www.eba.europa.eu/sites/default/documents/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf.
13) *ECB Guide on materiality assessment (EGMA) Materiality assessment for IMM and ACVA model extensions and changes*. Bankingsupervision.europa.eu. (2021). Retrieved 5 October 2021, from https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.egma_guide_201709.en.pdf.
14) Efromovich, S. (2008). Oracle inequality for conditional density estimation and an actuarial example. *Annals Of The Institute Of Statistical Mathematics*, *62*(2), 249-275. https://doi.org/10.1007/s10463-008-0185-1.
15) *FCA Innovation – fintech, regtech and innovative businesses*. FCA. (2021). Retrieved 5 October 2021, from https://www.fca.org.uk/firms/innovation.
16) FISHER, R. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals Of Eugenics*, *7*(2), 179-188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

17) Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). *Machine Learning*, *46*(1/3), 389-422. https://doi.org/10.1023/a:1012487302797.
18) Hand, D. (2002). Superscorecards. *IMA Journal Of Management Mathematics*, *13*(4), 273-281. https://doi.org/10.1093/imaman/13.4.273.
19) Hand, D. (2005). Good practice in retail credit scorecard assessment. *Journal Of The Operational Research Society*, *56*(9), 1109-1117. https://doi.org/10.1057/palgrave.jors.2601932.
20) Herron, J. (2013). *Yahoo is now a part of Verizon Media*. Finance.yahoo.com. Retrieved 5 October 2021, from https://finance.yahoo.com/news/fico-became-credit-score-100000037.html.
21) Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). Kernel methods in machine learning. The annals of statistics, 36(3), 1171–1220. http://dx.doi.org/10.1214/009053607000000677.
22) Huang, C., Chen, M., & Wang, C. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems With Applications*, *33*(4), 847-856. https://doi.org/10.1016/j.eswa.2006.07.007.
23) Investopedia. Investopedia. (2021). Retrieved 12 October 2021, from https://www.investopedia.com/.
24) Johnston, J., & DiNardo, J. (1997). Econometric methods. McGraw-Hill Companies, Inc.
25) Kamiński, B., Szapiro, T., & Kaszyński, D. (2020). *Credit scoring in context of interpretable machine learning*. SGH Publishing House.
26) Karlis, D., & Rahmouni, M. (2006). Analysis of defaulters' behaviour using the Poisson-mixture approach. *IMA Journal Of Management Mathematics*, *18*(3), 297-311. https://doi.org/10.1093/imaman/dpm025.
27) Kennedy, P. (2003). A guide to econometrics. MIT press.
28) LEE, T., & CHEN, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems With Applications*, *28*(4), 743-752. https://doi.org/10.1016/j.eswa.2004.12.031.
29) Lessmann, S., Baesens, B., Seow, H., & Thomas, L. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal Of Operational Research*, *247*(1), 124-136. https://doi.org/10.1016/j.ejor.2015.05.030.
30) Li, H., & Hand, D. (2002). Direct versus indirect credit scoring classifications. *Journal Of The Operational Research Society*, *53*(6), 647-654. https://doi.org/10.1057/palgrave.jors.2601346.
31) Louzada, F., Anacleto-Junior, O., Candolo, C., & Mazucheli, J. (2011). Poly-bagging predictors for classification modelling for credit scoring. *Expert Systems With Applications*, *38*(10), 12717-12720. https://doi.org/10.1016/j.eswa.2011.04.059.
32) Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal Of Operational Research*, *183*(3), 1466-1476. https://doi.org/10.1016/j.ejor.2006.04.051.
33) Matuszyk, A. (2004). Credit scoring: Metoda zarządzania ryzykiem kredytowym. CeDeWu.
34) Molnar, C. (2019). Interpretable Machine Learning: A guide for making black box models explainable.
35) Morawski, W. (2003). Kronika kryzysów gospodarczych. Wydawnictwo Trio.
36) Oesterreichische Nationalbank. (2004). Guidelines on credit risk management: Rating models and validation. Oesterreichische Nationalbank.

37) Pal, S.K., De, R.K., Basak, J., (2000). Unsupervised feature evaluation: a neuro-fuzzy approach. IEEE Transaction on Neural Network, 11(2), 366-376. http://dx.doi.org/10.1109/72.839007.
38) Pang, H., Dong, W., Xu, Z., Feng, H., Li, Q., & Chen, Y. (2011). Novel linear search for support vector machine parameter selection. *Journal Of Zhejiang University SCIENCE C*, *12*(11), 885-896. https://doi.org/10.1631/jzus.c1100006.
39) Pavlidis, N., Tasoulis, D., Adams, N., & Hand, D. (2012). Adaptive consumer credit classification. *Journal Of The Operational Research Society*, *63*(12), 1645-1654. https://doi.org/10.1057/jors.2012.15.
40) Piramuthu, S. (2006). On preprocessing data for financial credit risk evaluation. Expert Systems with Applications, 30(3), 489–497. doi:10.1016/j.eswa.2005.10.006.
41) R. Abrahams, C., R. Burnett, F., & Jung, J. (2000). *Using Data Mining Technology to Identify and Prioritize Emerging Opportunities for Mortgage Lending to Homeownership-Deficient Communities*. Analytics.ncsu.edu.
42) Regulatory treatment for accounting provisions. Iam.fmph.uniba.sk. (2016). Retrieved 12 October 2021, from http://www.iam.fmph.uniba.sk/institute/jurca/qrm/Chapter5.pdf.
43) Scallan, G. (2011). Selecting characteristics and attributes in logistic regression, In Credit scoring conference crc, edinburgh.
44) Scallan, G. (2013). Marginal Kolmogorov - Smirnov Analysis: Measuring Lack of Fit in Logistic Regression. Credit Scoring Conference CRC, Edinburgh. https://www.scoreplus.com/assets/files/Marginal-KS-analysis-Measuring-lack-of-fit-in-logistic-regression-Edinburgh-conference-Aug-2013.pdf.
45) Selvitella, A. (2017). The ubiquity of the Simpson's Paradox. *Journal Of Statistical Distributions And Applications*, *4*(1). https://doi.org/10.1186/s40488-017-0056-5.
46) SEMMA Model - GeeksforGeeks. GeeksforGeeks. (2021). Retrieved 12 October 2021, from https://www.geeksforgeeks.org/semma-model/. SEMMA Model - GeeksforGeeks. GeeksforGeeks. (2021). Retrieved 12 October 2021, from https://www.geeksforgeeks.org/semma-model/.
47) Shi, J., Zhang, S., & Qiu, L. (2013). Credit scoring by feature-weighted support vector machines. *Journal Of Zhejiang University SCIENCE C*, *14*(3), 197-204. https://doi.org/10.1631/jzus.c1200205.
48) Siddiqi, N. (2012). Credit risk scorecards: Developing and implementing intelligent credit scoring. John Wiley & Sons.
49) Siddiqi, N. (2017). *Intelligent credit scoring* (2nd ed.). John Wiley & Sons.
50) Simpson, E. (1951). The Interpretation of Interaction in Contingency Tables. *Journal Of The Royal Statistical Society: Series B (Methodological)*, *13*(2), 238-241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x.
51) Standards, I., Standards, L., & Instruments, I. (2021). IFRS - IFRS 9 Financial Instruments. Ifrs.org. Retrieved 12 October 2021, from https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/.
52) Thomas, L., Crook, J., & Edelman, D. (2017). Credit Scoring and Its Applications, Second Edition. https://doi.org/10.1137/1.9781611974560.
53) Thomas, L., Oliver, R., & Hand, D. (2005). A survey of the issues in consumer credit modelling research. *Journal Of The Operational Research Society*, *56*(9), 1006-1015. https://doi.org/10.1057/palgrave.jors.2602018.
54) Vapnik, V., (1995). The Nature of Statistical Learning Theory. Springer Verlag, New York.
55) Vapnik, V., (1998). Statistical Learning Theory. John Wiley & Sons, New York.

56) Wang, D. (2011). Binary tree of posterior probability support vector machines for hyperspectral image classification. *Journal Of Applied Remote Sensing*, *5*(1), 053503. https://doi.org/10.1117/1.3553800.
57) Wang, X., Wang, Y., & Wang, L. (2004). Improving fuzzy c-means clustering based on feature-weight learning. *Pattern Recognition Letters*, *25*(10), 1123-1132. https://doi.org/10.1016/j.patrec.2004.03.008.
58) Yap, B., Ong, S., & Husain, N. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems With Applications*, *38*(10), 13274-13283. https://doi.org/10.1016/j.eswa.2011.04.147.
59) Yeung, D., & Wang, X. (2002). Improving performance of similarity-based clustering by feature weight learning. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, *24*(4), 556-561. https://doi.org/10.1109/34.993562.