Deep Science Publishing
https://doi.org/10.70593/978-93-49910-19-5_5

Chapter 5: Artificial intelligence in credit risk modeling, underwriting, and customer behavior prediction

5.1. Introduction to Artificial Intelligence in Finance

The financial services industry has recently undergone a significant transformation due to advancements in digital technology. The easing of traditional barriers of entry associated with regulation, high capital, and information asymmetries has enabled non-financial technology companies, as well as new regulations, data sources, and data technology tools, to challenge the traditional oligopolistic structure of the industry. Traditionally dependent on human decision-making, financial services are now under pressure to maximize efficiency, lower costs, reduce risk, and improve customer experience, as in any other industry. In this process, artificial intelligence technologies have become increasingly important. Financial institutions now use AI to boost sales in some areas, directly or indirectly enhance customer experience, automate back office operations, and improve analyses for risk management, especially credit risk. Non-financial companies use machine learning for fraud detection and use risks, employ partner firms to provide additional financial services, and then use AI to enhance marketing (Galindo & Tamayo, 2000; Khandani et al., 2010; Lessmann et al., 2015).

As AI capabilities continue to develop, there now seems to be no limit to the potential applications of AI technologies. Now AI is rapidly based everywhere in financial institutions: front office, middle office, and back office, as well as used by operations and human resource functions of financial institutions. Overall, a growing number of financial industry practitioners believe that AI will be one of the most disruptive technologies in finance in the 2020s, with the potential to change the entire business processes and organizational structure of traditional financial institutions and challenge the current business models (Moro et al., 2014; Sirignano & Cont, 2019).

5.1.1. Background and Significance of AI in Financial Services

For centuries, the financial industry has relied on cutting-edge technology to execute business tasks. Investment banks were the early users of computerized systems that allowed them to execute a large volume of trading with high speed and lower costs. They also pioneered the use of quantitative models to value new products. These quantitative models relied on sophisticated mathematics that were developed to better understand financial markets. With advances in information processing, the nature of financial business underwent two major changes. The first change was the overwhelming growth of the amount of data generated from financial transactions. The second change was the lowering of costs due to technological advancement. With the lower costs of storing digital information, more financial data are being stored in electronic format all over the world than in paper format. This phenomenon provides financial institutions with unique business and financial competitive advantages. It also generates new challenges, or risks, for financial players.

Fig 5 . 1 : Credit Risk Models with Machine Learning

With the onset of financial crises during the late 1990s, and realizing the inadequacies of statistical models, risk management reviews by both private sectors and regulators have awakened the need for more specialized tools and techniques for prudential supervision of risks faced by the financial industry. Consequently, many senior executives and risk officers exert serious efforts to build sophisticated systems to predict possible future scenarios. Such efforts not only lead to improved risk management but also deliver a significant improvement in the bottom lines of those financial institutions that implemented models in managing their performance. Those attempts have, nevertheless, suffered from a certain level of push-back, due mostly to the reluctance of senior risk managers to use models in day-to-day management tasks. The perception by risk officers is prompted by two primary concerns. First, it is believed that the black-box nature of many techniques offers little or no explanation about the potential outputs or inferences from the model.

5.2. Overview of Credit Risk Modeling

Credit risk measures the likelihood of loss due to a borrower's inability to repay a loan or meet contractual obligations. It is the risk that a lender will not get paid back, which may result in the loss of interest or principal or, in the worst-case scenario, absorbed losses exhausting capital. To quantify credit risk, lenders identify the risk involved with lending money to a borrower and use this to calculate credit risk spreads or the premiums charged on defaultable bonds relative to default-free bonds. Spreads are determined by market forces, which reflect the judgment of bondholders about the risk of default. Spreads compensate investors for the default risk premium, expected losses, as well as liquidity and exchange risk premiums. The default risk premium compensates bondholders for the probability of default multiplied by the expected loss given default. The expected loss given default is, in turn, determined by market forces, based on the borrower's characteristics, such as credit history, outstanding debt, and debt repayment ability. The demand for credit risk pricing and credit risk prediction, as reflected in credit migration matrices, rating agencies, and the incorporation of credit risk in the capital reserves of banks, has led to an enormous body of literature. Most of this literature has focused exclusively on one of the two aspects of the problem: either credit spreads or credit ratings. However, despite the apparent differences between the two pricing and modeling approaches, there is a stark similarity between them in terms of the information they provide and the techniques typically used for their construction. A major issue in credit risk modeling concerns the fact that credit data are infeasible for any analyst to manage because of their limited size. Credit modeling, however, is similar to survival analysis, a much broader statistical model that has been applied successfully to other areas.

5.2.1. Key Concepts and Frameworks in Credit Risk Modeling

Credit risk modeling borrows many of its key concepts from more established disciplines in science and economics. These cognitive inputs have been essential for structuring the development and subsequent refinement of two primary tasks in credit risk modeling, namely credit risk driving and default prediction. The bulk of contemporary academic research into credit risk modeling has focused on developing better solutions to the task of default prediction.

Credit Risk Drivers

Default prediction itself is an analytical and incremental discovery step that proceeds by searching for predictive signals correlating with an unobserved abstract variable of interest, the probability of default at a point in time. More formally, the probability of default at a point in time can be expressed as: PDt = P{eventual default | information at time t} Whereas, the event of eventual default is defined as eventual default = {default within a management-determined follow-up horizon or time-to-default | default is from the set of events defined as Credit Event: dispossession transfer or debt obligation cure due to insolvency (distress)}.

When used in a purely statistical sense, the term default does not encapsulate the unique nature of financial default being studied here. This is because events classified as defaults in finance exist at the interface between default prediction and the fields of longitudinal event data analysis and time-to-event analysis. In addition, the model mechanics for credit risk, non-default event assignment training, and validation signatures are generally different, when compared with the other two fields mentioned above. Hence the term non-default is used in the stricter sense, where the concept of non-event is absent.

5.3. The Role of AI in Credit Risk Assessment

The expansive growth of accessible data, coupled with significantly enhanced computing capabilities, has propelled the refinement of numerous risk measurement frameworks. This shift has had profound implications for credit risk methodologies triggering the transition from credit scoring to credit risk modeling. Traditional credit risk evaluation methodologies employed by banks, pension funds, leasing or factoring companies, and firms providing both financial services and products have limitations. Such assessments typically utilize statistical modeling techniques including Moving Average, Regression Analysis, Analytical Hierarchy Process, Linear and Non-Linear Programming, Logit Analysis, Decision Trees, and Neural Networks to analyze various deterministic and stochastic risk impacts, whichever is suitable. Presently, the realm of credit and operational risk management is experiencing an evolution with the emergence of Artificial Intelligence solutions such as Big Data Analysis, Automated Machine Learning, Text Mining, Natural Language Processing, and Cloud Computing. These developments have triggered a reassessment of previously held principles and opinions.

Advanced Machine Learning techniques such as Neural Networks with Deep Learning, Random Subspace, Hypercube, Gradient Boosting, and Support Vector Machines offer appealing, effective, and accurate predictive modeling solutions for significant problems across various industry sectors. These concerns include predicting default events, bankruptcy or credit rating changes, mapping and by-passing on-balance-sheet and off-balance-sheet underlying modeling techniques, and finally determining the risk parameters of Loss Given Default, Probability of Default and Exposure at Default, along with corresponding Early Warning Systems. The current essay builds on the ongoing debate by examining the potential of AI for Credit Risk and Risk Parameter Assessment.

5.3.1. Integrating AI Techniques into Credit Risk Assessment Models

While traditional credit scoring relies on simplistic binary distinctions between "good" and "bad" clients, the process of classification must account for degrees of riskiness. A lender might decide to help a person classified as "bad," especially when that person is near the cutoff point for being labeled a "good" person, and believes a loan will solve a crisis. Contrarily, lenders would want to deny loans to those people who lie far away from the cutoff point. This smoothing effect can be applied through the use of artificial intelligence techniques. AI techniques help to enhance data analysis and model performance relative to traditional statistical techniques. Therefore, AI techniques have become an important tool for dealing with large amounts of data, which is a prerequisite for the proper identification of risk segments.

Several benefits can arise from the integration of AI techniques into credit risk assessment models. Enhanced accuracy levels, lower default risk as a result of better risk differentiation, improved segmentations that help price different client characteristics more adequately, better explanations of performance and pricing decisions, and greater resource savings. The main AI techniques used in practice for credit risk assessment are decision trees and their extensions, as well as artificial neural networks, genetic algorithms, support vector machines, and fuzzy logic. Hybrid systems that integrate AI techniques into statistical models have also garnered considerable interest, as modeling objectives can be solved using these alternatives. These hybrid systems are characterized as systems that, through the use of AI techniques, will assist the analyst with the completion of a statistical task.

5.4. Machine Learning Techniques for Credit Scoring

A wide range of machine learning techniques have been applied to credit scoring problems. These techniques can be categorized into supervised learning, unsupervised learning, and hybrid approaches that build on both supervised and unsupervised ML algorithms. Supervised learning approaches rely on labeled training data with known credit outcomes. The most recognized supervised ML approaches for credit scoring include decision trees and associated techniques, support vector machines, neural networks, logistic regression, and ensemble techniques. Various unsupervised learning techniques are commonly applied to credit scoring problems including cluster analysis, self-organizing maps, and association rule mining.

Fig 5 . 2 : AI-based credit scoring

The major distinguishing feature of hybrid approaches is that the final model is based on a supervised ML method that involves customer segmentation or clustering as a preparatory step. By segmenting or clustering customers into homogenous subsets, the differences in customer behaviors and credit risk in the various segments can be modeled separately using different supervised ML models. In this way, the supervised ML techniques can achieve improved accuracy. The benefits of hybrid approaches are heightened when the segments identified via the unsupervised approach can be validated as being stable over time. Hybrid approaches consist of the independent cluster algorithm, clustering K-nearest neighbor, and ensemble clustering K-nearest neighbor.

Hybrid approaches can also build on supervised methods such as logistic regression, Bayesian belief networks, Gaussian mixture models, and decision trees.

5.4.1. Supervised Learning Approaches

Supervised machine learning techniques search for an explicit function that maps the observed characteristics of a borrower to a classification that delineates different default or non-default categories and or types. The function defined with the training sample is then used to classify the remaining validation sample, which assists in selecting the best prediction functions or classifiers based on the evaluation criteria. The selected classifiers are subsequently operationalized to validate and implement credit scoring models. These techniques can be broadly classified into three categories: regression techniques, decision rules-based classifiers, and probability density function-based classifiers.

The regression techniques, which predominantly model the probability of default, are the traditional logistic regression and the more recent applications of probit regression and inverse logit transformation. Decision rules-based classifiers partition the input space using decision rules derived from the observed data to classify borrowers. Notable members of this family include decision trees, functional trees, ensemble classifiers, etc. Probability density function-based classifiers estimate the unconditional probabilities of the observed classes and classify borrowers based on maximum likelihood, or minimum distance criteria. This family includes kernel density estimators, neural networks, support vector machines, etc. Decision rule-based classifiers and probability density function-based classifiers are more flexible and powerful and often model the default risk patterns in a better way than the older regression-type models. However, parametric regression-type models are still preferred as the first choice due to their interpretability, simplicity, and the existence of long-running empirical compilations supporting their better-predicting accuracy.

5.4.2. Unsupervised Learning Techniques

Numerous researchers have investigated the potential of utilizing unsupervised learners for credit scoring. Early endeavors anticipated the worth of hidden information to be discovered by these techniques as characteristics supplementary to standard credit scoring inputs. In particular, self-organizing feature maps used by the default customers were adopted to enhance the model performance of logistic regression. In an extensive dataset, the incremental benefit was measured using statistics calculations. Using clustering, hidden structures in the borrows default segments were discovered and it was postulated that the clustering results be folded back into a traditional credit scoring model. A similar cluster-based strategy was additionally followed. Indirectly, rules were generated by breaking down the clustering results into sub-segment rules. Interested in the description of the segments rather than predicting goals. Thereafter, the relevance of hidden segment structures to credit scoring was pointed out.

More recent contributions have incorporated clustering strategies into the phases of fraud detection, risk rating, and risk event prediction. Noticeably, a combination of clustering and tree induction methods was designed to deal with issues of their own. Although several of the above works bear the problems of the credit scoring process, none of them has examined the use of unsupervised learning in general credit scoring. The popularity of the area may rely upon the results of the previous works, especially the seemingly questioned any added predictive power if such models use improper methods or small datasets. Computer memory and time necessities may retain the strategy from wider acknowledgment on the end of the lender, especially if the credit decisions were based on more than money aggregate agency systems.

5.5. Data Sources for Credit Risk Modeling

Data availability is one of the critical factors for successfully applying AI in credit risk modeling. The use of various publicly available data has recently gained prominence in the AI and credit risk community. To illustrate the data issues, we present the specific sections of data types primarily used in credit risk modeling. In this section, we will mainly use an SME credit-rating modeling context to explore the data sources. The focus on SME rating models is justified by the fact that SMEs do not have publicly available credit-rating agency ratings, but other sources of external ratings are available.

Data utilized in credit risk modeling can be grouped into two broad categories: traditional and alternative data sources. Traditional data sources consist of annual financial statements, interim financial statements, internal management assessments of SME company risk, and the sample-based weighted transition matrices, while alternative data sources range from quantitative sources, such as web traffic counts and credit card transactions, to qualitative sources, such as social media data. The qualitative alternative data sources need to be classified further into data without sentiment information (plain data) and data with sentiment information (sentiment data) regarding companies' risk. The qualitative alternative data sources also differ in terms of degree of availability and need for data scraping. In credit risk modeling, qualitative data have seen less use compared to quantitative data. In addition, quantitative data, especially unrestricted web traffic sources, are currently among the fastest-growing alternative data sources in credit risk modeling. Moreover, many of these quantitative alternative data sources can be updated in real-time, compared to traditional data sources that are published yearly and have significant delays.

5.5.1. Traditional Data Sources

In general, a data source can be defined as a resource from which the data is obtained. For the models analyzed in the present paper, the data required are either called 'traditional data sources' or 'alternative data sources'. Traditional data sources refer to bank transactional data from institutions that use the models. For consumer behavior modeling, usually, companies have transactional data from previous customers who decided to acquire the product. The two main examples of traditional data sources are the following: (1) Credit history data: Credit history data can be defined as the one that monitors and assesses individual and corporate financial commitments, such as payments imposed by loans. This data is usually used together with other traditional data sources in any statistical modeling for credit risk management. Credit history data are generally provided in credit bureaus depending on the origin country's laws and regulations. (2) Behavioral data from bank transactional data: Real bank processing transactional data is the primary source of data that some companies must have, especially for consumer behavior modeling (targeting, customer acquisition risk, etc.). Traditional data sources rely on the fact that all operations went through accounts considered in the data. The more operations available, the better quality of the risk estimates, and the higher significance of the estimates.

For banks, transactional data is an essential and primary input for predictive modeling, monitoring client risk, or estimating expected loss. Transactions detail customer purchases, whether by payroll, bank transfer, check, or purchases or withdrawals in ATMs. It comprises detailed data about the account, including the amount, currency, date of the transaction, account balance, and whether the transaction is a debit or credit.

5.5.2. Alternative Data Sources

There is a large literature on the use of alternative data sources. It is possible to group them into four groups. The first category contains non-bank transaction data. Several firms collect information on transactions people do in different establishments and create panels available for specific markets. Some companies analyze the search of individuals or a company. Monthly indicators, such as the number of searches related to hiring employees, used car prices, and propane/gas prices, have proven helpful as signals of bad debt. The third category contains data from corporate and customer behavior. From a corporate point of view, business and supply chain information such as the movement of goods, company tax payments, and ports' activity table, are examples of alternative data that banks might leverage to enhance their scoring models. From a customer features perspective, works scrape the social media and gathering mobile phone and broadband service usage of consumers. Finally, the fourth group embeds hybrid data sources. Around 40 percent of the consumers had at least one frozen credit report, mainly in the three main credit bureaus. Having frozen their reports means that these customers would be excluded from traditional data models. However, a method was developed that helps to use the frozen data. This method generates a new ID that parses names, addresses, e-mails, and phone numbers to get a new unique ID that centers on that type of reference. Using dumpster data from credit cards, specifically branded credit cards, is also an example of hybrid data being used in a market.

5.6. AI Algorithms in Credit Risk Modeling

The credit risk modeling task, in which an applicant is classified into one of two groups – risks or non-risks, the so-called binary classification task, has become one of the major application areas of AI. Various algorithms have been developed in credit risk modeling. The research began with a single Decision Tree induction algorithm. A decade later, several new AI algorithms and enhancements of existing algorithms were developed. In particular, two types of neural networks were applied, i.e., the classical FeedForward Neural Networks with BackPropagation training algorithms and winner-takes-all networks. Other new AI algorithms included two multi-layer designs for Decision Trees, AdaBoost, and Tiling and others were Multiple Hidden Layer Feedforward Neural Networks with Back-Propagation Algorithms for Learning, Adaptive Boosting Algorithms for Learning Decision Trees, Randomization Techniques for Learning Multiple Hidden Layer FeedForward Neural Networks and others, Support Vector Machines was also applied.

The application of AI Algorithms in credit risk modeling is explored in many other studies. Specialized ensembles of Decision Trees are shown to outperform classic statistical methods. Specifically, Bagging and Boosting Ensembles of Deep Decision Trees outperform other popular classic statistical methods for the binary classification task, such as Logistic and Linear Discriminant Analysis. In the area of Neural Networks, Hybrid Neural Networks are shown to outperform other popular classic statistical methods. Neural Networks with Weight Decay are shown to perform better than other popular classic statistical methods for the binary classification task for large datasets, while Multi-Layer Feed-Forward Neural Networks with Back-Propagation Algorithm for Learning perform better than other popular classic statistical methods for the binary classification task for large training samples. Support Vector Machines are shown to outperform other classic statistical methods for credit risk modeling datasets characterized by high input dimensions.

5.6.1. Decision Trees

There is a vast number of algorithms that can be used to create predictive models for credit risk and customer behavior. The models range from very simple procedures to advanced software tools and approaches. Traditionally, regression has been the only technique for analyzing credit risk. Banking institutions successfully adopted logit and probit models as advanced analysis tools for predicting the likelihood of events happening, such as attrition, credit default, and so forth. New techniques are believed to be lacking flexibility, especially when data is a high-dimensional mix type. Cleansing often removed critical information and implications for how information is related in some instances, this resulted in dimensionality increased, i.e. data realistic with the large customer base. More recently, techniques like CPM, latent class models, MNP, ICC, and MDA, can also be used to analyze these issues.

However, these models majorly assume that the dependence/relationship between the response variable and independent covariates are known a priori. In the meantime, they cannot model interaction effects. They cannot deal with the case when observations are plenty but a few samples in some areas form nonlinear effects. Recently, AI has been gaining ground in different industries specifically in finance. Although some skepticism still exists, AI is impacting customer behavior modeling and credit risk modeling positively. Decision trees have a simple structure and are easy to interpret. Once estimating the decision tree, we can compute the estimates of the conditional mean, determine the locations of breakpoints, and determine the importance of the independent variable. Growing decision trees include CART, C4.5, and CHAID. BDT has been incorporated into decision trees to overcome the limitations of classical decision trees, especially for classification settings.

5.6.2. Neural Networks

Neural networks are a class of AI algorithms based on a common biologically-themed architecture that has shown great promise in various modeling tasks. An artificial neural network is a weighted graph connecting a set of nodes across multiple layers. Nodes in a layer are referred to as neurons, and each artificial neuron operates as a simple logic function. In a neural network model the weighted sum of a neuron's input is conditioned by a nonlinear activation function, called the transfer function. The transfer function outputs the final weighted response signal from that neuron to the subsequent layer. Jointly optimizing all the weights in a neural network model to minimize the output prediction losses on a training set is generally a non-trivial empirical task. Neural networks are capable of approximating almost any smooth nonlinear function to an arbitrary degree of accuracy using a sufficiently large set of hidden neurons. Other important properties of neural networks include: their great parallelizability; a relatively low computational cost to evaluate after training; and the fact that they are approximation algorithms—they do not offer an explicit description of the approximated function.

Artificial neural networks were first proposed about 70 years ago, in the context of simulating neurobiological processes in the human brain. However, in practice, the first models were criticized for being quite limited in terms of the complexity of tasks they could solve. Interest in neural networks revived in the late 1990s and over the following decades a series of breakthroughs were leveraged to develop deep learning architectures which made these algorithms much more versatile and efficient.

5.6.3. Support Vector Machines

Support Vector Machine (SVM) is a machine learning technique often utilized in classification tasks. Its main aim is to determine the optimum hyperplane that can distinguish two classes while maximizing the margin between them. Regardless of its underlying theory, it performs well in practice and is both easy to use and has solid theoretical foundations, features that recommend it as a first-choice method. The hardest part is the selection of the kernel function and the function parameters. It consists of a method for solving supervised learning problems about assigning the correct label to a given observation that belongs to one of two classes, based on a training set of examples containing observations with known labels.

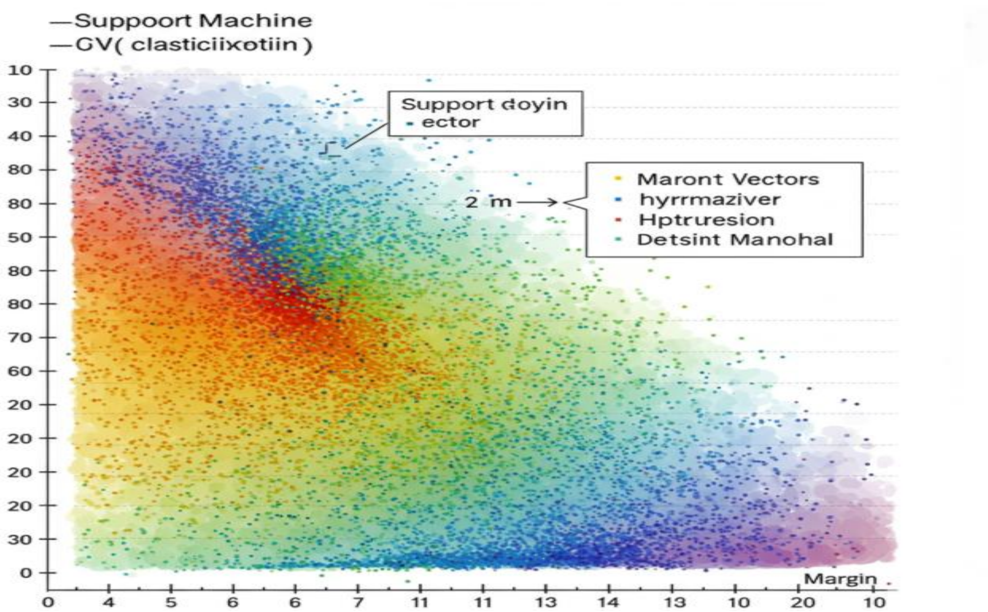

Fig 5 . 3 : SVM Classification Distribution

SVM makes an important assumption by identifying that data can be separated by a hyperplane. In a two-dimensional space, it is represented by a straight line; in three dimensions, it is a flat surface and, in the generalized case it is a hyperplane. The line (or hyperplane in the general case) is chosen to maximize the distance from it to the nearest data point, this distance is referred to as the margin. If points from the two classes cannot be separated, then SVM will find the hyperplane that provides the largest possible margin, but this margin may not be entirely free of points. It will allow a few of the points to be within the margin and even added to the wrong side of the hyperplane, thereby introducing some errors. Those points that lie exactly on the margin, or, in case of error, on the wrong side of the hyperplane are called support vectors and these are of particular importance because they affect the SVM model. Any data that is far from the margin does not affect the position of the hyperplane and, therefore, does not affect the model.

5.7. Conclusion

The modern banking systems we know today were created in the last three centuries and are responsible for a huge part of the worldwide economy. Especially when it comes to risk, banks are the first line in the prevention of money laundering and terrorist funding. However, each year those systems become more complex, as customers are not only opening bank accounts but are also working with credit cards, investments, and even insurance. It is vital to understand customer behavior and the next steps they will take, which can be seen by efforts in the digital banking experience provided to them. Therefore, banks must start leveraging more technology to increase their efficiency and improve their models, considering that the data used in risk and behavioral modeling is becoming dirtier and dirtier. We believe that Artificial Intelligence can help with that.

Nonetheless, that does not mean technology should be blindly trusted, and auditors must keep providing checks and balances to avoid wrong judgments and hence trillions of dollars in losses. We demonstrate that Artificial Intelligence can surely help with credit risk modeling and customer behavior predictions, with its advantages over traditional methods in certain tasks. Data quality, using rewards as a target variable especially, and fluctuating modelizations for customer behavior prediction are limitations that may be circumvented whenever the AI techniques are used correctly. And as this market increases, new algorithms will tend to emerge, making models better and better. These are exciting times and the future of Artificial Intelligence in Banking looks very promising.

5.7.1. Final Thoughts on the Future of AI in Credit Risk Management

Credit risk management in its traditional sense is forever changed as the adoption of artificial intelligence in credit risk modeling and customer behavior prediction is not only inevitable; it is already happening by the early adopters who are investing heavily in it. This journey to model-building nirvana where machine learning automation tools, such as self-taught machine learning, automated data preparation, automated machine learning, and model validation tools with the best-practiced governance is not only reassuringly easy but fast and customer-centric, is facilitated by an explosion of advances in hardware, affordable storage, software, and storage for large and diverse data with increasing dimensionality.

When it comes to decisions for businesses or individuals on the threshold of bankruptcy in this high-risk environment of volatility and uncertainty opportunities, such as the push for electric and autonomous vehicles, clean energy sources, or the pandemic-inspired push for self-reliance and local business, the accuracy, interpretability, and speed of use of AI algorithms to classify customer creditworthiness or their propensity to repay must be validated across cohort sizes over the quickening business cycles and then be deployed as part of a closed-loop decision system that captures customer behavioral changes for adaptability overtime to soak up the new-age data with the increasing future uncertainty that includes the macroeconomic indicators guiding credit risk behavior stability over the business cycle such as the company or sectoral GDP that need to be monitored. Thus, it is reassuring to conclude that the marriage of machine learning with traditional credit risk models will benefit banks in their quest to increase lending revenues while controlling loan defaults, in this era of thin margins, where faster and more accurate decisions make all the difference.

References

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (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

Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, 62, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer Credit-Risk Models via Machine-Learning Algorithms. Journal of Banking & Finance, 34(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Galindo, J., & Tamayo, P. (2000). Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications. Computational Economics, 15(1-2), 107–143. https://doi.org/10.1023/A:1008699112516

Sirignano, J., & Cont, R. (2019). Universal Features of Price Formation in Financial Markets: Perspectives from Deep Learning. Quantitative Finance, 19(9), 1449–1459. https://doi.org/10.1080/14697688.2019.1571683