Deep Science Publishing
https://doi.org/10.70593/978-93-49910-16-4_2

DeepScience
Open Access Books

Chapter 2: The evolution of credit data and the role of machine learning in modern credit scoring

2.1. Introduction

Establishing a stable and healthy financial system is vital for the government, banks and other financial institutions. Credit risk is one of the issues that banks face as it can arise from bad commercial loans without proper management. Instead of imposing high interest rates on all applicants to compensate for potential bad loans, banks can benefit from retrieving informative data about prospective applicants and classify the creditworthiness based on the data. High speed computers and growing availability of large databases, with respect to efficiency, bring forth Machine Learning (ML) methods to credit scoring and evaluation. Credit scoring can be automated by these methods, but deemed too complex for understanding rationality of output. Therefore, simplifying the model and scoring are questioned where the trade-offs can be challenging. The classification of credit scoring is regarded on 3 aspects; (1) the algorithms and techniques used, (2) how to assess and compare the performance of the classification models, and (3) practical levels on which this evaluation can be done. Majority of ML methods such as Neural Networks (NN), Extreme Learning Machines (ELM), Support Vector Machines (SVM), Gradient Boosting Trees (GBT) are regarded as black-box methods and difficult to interpret. Decision Trees (DT) are evaluated with integrated techniques such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanation (SHAP), K-nearest neighbor (KNN), Naïve Bayes models or linear logit models where nonlinear transformations can be stored out. Non-linear methods and transforms are regarded as hazier based on combos of large numbers of predetermined linear and non-linear factors, and GVFs are still used [2]. Evaluation of performance of credit scoring classification methods is primarily taken with binary classification measures such as True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR), in addition to Domain Specific Typos and

related metrics compared up on European Union (EU) policy driven thresholds. Cost-benefit ratio and its effects and propagation on ranking of classifiers are also included with regards to decreasing data and increased bank profitability(Gupta & Soni, 2023; Agarwal & Verma, 2023; Chen & Zhao, 2024).

Fig 2.1: Credit Scoring

2.1.1. Background and Significance

With the increasing affordability of many products, there is a rise in the consumption of products on credit. Where lending institutions can sell goods on credit, there is a possibility of default. The lending industry needs to examine a new customer's profile to determine whether they are trustworthy customers or possible defaulters. This profiling requires substantial foundational data on each customer, and most methods require almost a complete profile of each customer with extensive data gathering covering all aspects of customers' lives. In the absence of such comprehensive data, companies need to build a profile based just on the most basic information consisting of only eight variables. It is to examine the ability of the dataset to withstand the challenge faced by the lending industry in assessing the risk of credit defaults to win and retain customers.

Credit cards are a very popular mode of payment. Banks issue credit cards based on the individual's credit history. Credit risk is defined as the risk of financial loss to design if the borrower fails to repay a loan or otherwise comply with the contractual obligations. Credit score refers to the numerical expression of a person's creditworthiness. There is a rise in credit card delinquency rates in the United States of America, which imposes a financial risk for lending institutions. Credit risk has shown exponential growth in the 1990s against dramatic economic and technological change. The number of defaults has

risen significantly and has cost commercial banks millions of dollars. This has become critical for banks and lending financial institutions to use robust mechanisms to forecast probabilities of credit defaults before lending it to customers. In many cases, it has to be done where the customers have limited or no credit history. In these cases, only pure financial factors will be available for formulation. The dividend and cash flow are two balance sheet items that play a decisive role in predicting companies' bankruptcies for many industries. But application of the pure high cardinality model depends on the firm type and the firm age. Due to the high imbalance ratio, financial ratios alone may not yield good results for whitelisting. Since no fine-tuning is done to the classifier, there may be somewhere room for further improvements. Considerable data preprocessing and re-sampling are also carried out to improve the performance of classifiers. Credit scoring agencies are now in a better position to enable the banks by providing extensive credit analysis of the customers. The most common data used by scoring agencies for analysis is demographic, financial, behavioral and statistical. Data is at present being continuously collected by companies which can relatively easily harnessed to a high credit risk worthiness determination. It is now possible to build a robust credit risk model with fair prediction accuracy using machine learning techniques without any credit history data.

2.2. Historical Overview of Credit Data

Credit scoring is one of the oldest applications of analytics where lenders perform statistical analysis to assess the creditworthiness of potential borrowers. Fair Isaac was founded in 1956 as one of the first companies offering retail credit scoring services in the US. Its well-known FICO score has been used by financial institutions, insurers, utilities companies, and employers [5]. The first corporate credit scoring models date back to the late sixties with Edward Altman developing the z-score model for bankruptcy prediction. In Europe, the first corporate credit scoring models were developed in the eighties by various banks. Originally, these models were built using limited data and were based on simple classification techniques such as linear programming, discriminant analysis, and logistic regression. The importance of these credit scoring models increased due to regulatory compliance guidelines such as the Basel Accords and IFRS. As a consequence, financial institutions rely on these models to both comply with regulations and allow them to grant loans and operate profitably.

Due to their early establishment, the majority of corporate credit scoring models are now outdated. Research on developing high-performing credit scoring models has stalled, leaving a huge estimate of losses due to "bad" clients. Simultaneously, a huge group of potential borrowers remain untapped (the financing gap). As expectations for transparency and explainability have risen in recent years, there is a renewed interest in

credit scoring with the intention of building models that comply with these requirements. However, as cost-effective and efficient as possible, there is strong opposition to developing proprietary scoring models (Sharma & Bansal, 2022; Singh & Mehta, 2023).

The best investment is to leverage innovative Big Data sources. These sources, such as mobile phone data, are being intensely explored in the consumer space. Not only does this present the opportunity to profile potential borrowers using a wider representation of behavior, but it also presents an ethical challenge. It is thus crucial for stakeholders involved in data sharing, model development or assessment, as well as those who possess advanced scoring models or who otherwise profit from it, to think about how to leverage and develop ethical, explainable and transparent Big Data consumer scores.

2.2.1. Early Credit Practices

Credit scoring is used to assess the creditworthiness of current or prospective borrowers. Its intention is to characterize how likely a borrower is to repay an issued loan within the defined loan term, reducing the risk of unexpected defaults. It started in 1890 when the first consumer credit reporting agency was founded in the U.S., and Everitte, the firm's owner, began the use of individual credit reports. The process can be summarized in three stages: credit report assessment, credit risk classification, and modeling. The assumptions utilized in the process are that past behavior is an indication of future behavior, that risk aversion and risk-sharing are characteristics of all agents, and that risk can be assessed. Initially, credit decisions were based entirely on reports manually collected by agents and traded in a secured environment across individuals such as merchants, bankers, and property owners. Later, information became centralized in credit bureaus that reported to third-party companies, and the data to assess the latter became more structured than originally.

The first initiative to assess credit repayment probability more formally was taken in 1956 when the Fair Isaac Corporation (FICO) was founded. At first, credit risk classification merely consisted of generalized linear models using broad socio-economic income and demographic information as explanatory variables. In the 1980s, computerized solutions took off and allowed for a broader set of variables to consider, significantly refining the scoring process itself. In the 1990s, most lenders began to set up in-house analytics departments, implementing tailor-made accounts' lead scoring solutions. As a result, the analytical paradigm relating to pre-qualification techniques became similar across industries. In these systems, the first collection strategy is implemented much like a lead score, utilizing readily-available customer contact information. In subsequent treatments, the models, rules, and channels become more intricate.

2.2.2. Development of Credit Bureaus

The setup of credit data providers is usually formed by a decentralized market structure where multiple private firms, known as credit bureaus, may operate in the same country. Private credit bureaus collect and distribute consumers' credit data across many credit grantors of different industries, such as telecommunication companies, banks, and retailers. In developing countries, however, the entry of private credit bureaus is usually restricted by the government due to concerns of monopoly and corruption. In these countries, a centralized public credit registry is usually constructed instead, which is operated and maintained by national banks of the government. Credit registries only collect limited credit information of credit grantors in some regulated industries, like banks only. In comparison, credit bureaus gather a more complete set of credit data including a wider population of credit grantors (including those in unregulated industries such as utility companies). This results in a richer set of baseline information to form credit scores. The usability of credit data in credit scoring is greatly influenced by the availability and the development of credit bureaus. A weighted-average indicator was computed to summarize the development of credit bureaus using the number of credit bureaus, the amount of credit information available per capita, the percentage of adults covered by a credit bureau, and the public availability of credit and default information. To properly value the extent to which credit data are useful for credit scoring, the development of credit bureaus must be taken into account. This is done by setting a modified base credit score when credit bureaus are absent. The differences in the development of credit data were investigated throughout the world. It is also hypothesized that the improvement in predictive performance will be more apparent for a longer sample period.

2.3. Current Landscape of Credit Scoring

The current state of credit risk modeling is characterized as stagnant, with limited improvements in accuracy or new development frameworks over the years. On the other hand, there is growing pressure from regulators to improve the blinding practices of current modeling approaches. Recent advancements in machine learning, specifically in the area of tree-based ensemble methods, present an opportunity for significant improvements in modeling quality. However, the nature of current modeling frameworks inhibits the fruitful implementation of these methods. On one hand, phase wise modeling, where variable selection is carried out before model development, creates a barrier between data understanding and the modeling process that dissuades data scientists from fully utilizing their knowledge on modern, advanced machine learning techniques. This leads to an unfortunate state of affairs where much of the modeling

process is either completely automated or, at the other end of the spectrum, is barely touched by modern idea thinkers and researchers.

In addition, there is an additional hurdle to overcome due to the complex regulatory environment and data governance restrictions that inhibit research like this from being widely pursued. Credit risk modeling plays a pivotal role in both technical and regulatory environments. A major requirement for credit scoring models is to provide a maximally accurate risk prediction. Regulators demand these models to be transparent and auditable. Therefore, simple predictive models such as logistic regression or decision trees are still widely used. Significant potential is missed, leading to higher reserves or more credit defaults. This paper presents a framework for making black box machine learning models transparent, auditable and explainable. During the last two decades, a lot of new modeling techniques and paradigms have emerged that allow the development of more powerful and advanced models to address and solve various business problems, usually leading to a lower prediction error.

2.3.1. Traditional Credit Scoring Models

The analysis phase of credit data used in commercial credit scoring after a company has filed for bankruptcy is critical to ensure the future performance of proposed scoring decisions. This can be provided through the inclusion of alternative credit data features, capturing real-time credit risk information, and the use of complex ML approaches. Banking financial services prepare B2B credit scores needed for their loan conditions and monitoring. The banking companies with these budgets are large commercial lenders in Turkey, and two datasets are available for modeling. Data preprocessing takes place for delivering a bank-ready score model. While a robust financial stability score model can be produced through simple ML methods with accuracy above 80% and interpretability, more exclusive models can be alternatively suggested that implement state-of-the-art ML methods.

While keeping for a NL discrimination score, therefore, a more favorable modeling consideration is offered to banks, starting with general preparation and simple modeling then allowing access to larger features and more exclusive modeling with elaborate features. Though, just the corporate credit worthiness phase is covered, future works can be conducted concerning monitoring. Some banking financial service companies may also own lots of data with no internal scoring effort. They would intrinsically benefit from this suggestion too, including a gap for future works from a custom bank-specific perspective. The alternative credit features are provided with relevant checks, delineating the suitable aspects of the problem description for a novel offering of an under-researched banking or non-banking need.

Fig 2.2: Credit Scoring Models

2.3.2. Limitations of Conventional Approaches

The credit scoring model has been crucial to the commercial bank's risk management and acceptance of high-risk customers during its long history. With the development of technological advancements and big data, banks have started utilizing machine learning (ML) methods in constructing credit scoring models considering the large amount of social media data and other added transnational data. However, banks have mainly focused on the predictive power of the ML model outputs without thorough examination of the characteristics of the data set used, potentially leading to the safety and sustainability of the banking system being adversely impacted [2]. There are great concerns about using machine learning models due to its lack of transparency and auditability, especially in high-stakes decision making. ML methods have been recently focused on in the academic field.

Each ML technique has its advantages and use cases in fields such as finance and credit risk. For example, while LSTM networks outperform others in relation to repeat transactions, logistic regression is a widely used classical model, providing direct assessment against regulators' regulations. In regard to credit scoring, though historical pre-post samples have been widely used, there are still some mistakes in applying ML models, forecasting the market rationally, and tightening credit conditions. When ML techniques are adopted, it is crucial to know enough about the sample itself, i.e. the structure, number of data points, possible overfitting, etc. Given its advantages in handling high dimensional and non-linear data, random forests (RF) have been adopted in various scoring applications. In financial markets, transaction recoverable information costs significant amounts, while analyst coverage leads to favorable information pricing. However, it is partially unclear whether better performance can be achieved with the consideration of controlled data bias.

2.4. Introduction to Machine Learning

Machine learning (ML) methods have become an important tool to determine the characteristics of variables with respect to a phenomena because of their wide variety of approaches taking into account minimal assumptions and associations among each other. However, it has been shown that prediction accuracy of these methods is sensitive to the selection of the features, methods, size of the data and transformation properties. Proper implementation of ML requires investigation of the most suitable and current ML classification algorithms with respect to the aforementioned aspects in order to attain robust predictions.

ML is a set of mathematical techniques which improve a model through experience by learning from past outcomes. Over the past decade, development of high speed computers providing parallel architecture made the implementation very attractive. ML methods have been used in many interesting fields such as marketing, finance, electronics, communication, biology and medicine. In the finance sector, forecasting stock markets, credit scoring, detecting fraud, risk analysis, portfolio allocation analyses and improving financial strength of companies are few applications of ML.

2.4.1. Definition and Key Concepts

Machine Learning (ML) methods are one of the most effective approaches of recent years to assess (and estimate) the developed characteristics and specifications of variables with respect to their associations. However, it is hard to determine a nonlinear relationship in a simple mathematical formula. Moreover, it is obvious that prediction accuracy of the ML methods utilized is sensitive to the selection of the features, the methods, the size of the data, and the transformation properties. Following the developments in computer technologies and the diffusion of high-speed computers, basically, the abundance of Big Data provides new avenues for the implementation of the ML. The financial sector is one of the widely affected sectors benefiting from the employment of the ML methods. There are various markets in the financial sector, among them, the banking sector is doubtlessly one of the key markets. Since financial risk is uncertain and complex, it is difficult to predict. On the other hand, banks are prone to the danger of high default risk, mainly caused by credit risk. Credit risk is the possibility of a monetary loss from a debtor that fails to repay a loan or otherwise meet a contractual obligation. It is problematic for banks assessing a bank's creditworthiness and creating a lending policy. Following the subprime crisis of 2008, many banks all over the world nearly collapsed. Hence, comprehensive regulatory frameworks like Basel II, III, IV, and V were developed to regulate risks in the banking sector.

Monitoring current and future risk levels of a bank's loan portfolio is a key factor for its competitiveness and profitability. Basically, banks assess the credit risk with respect to their own expert judgment by checking the history of the client as well as the overall economic conditions. In addition to expert judgment, banks assess (and estimate) credit risk employing credit scoring, an established method for assessing the financial strength of a person. Credit scoring has a great importance for banks since it can be employed for different usage areas such as bulk credit applications and automating credit limits. The stricter regulatory frameworks, the dynamic competitive environment in the banking sector, increasing competition amongst banks to acquire creditworthy customers, and improving data science and data analytics motivate banks to modernize (and also develop) their credit scoring process not only for competitive but also for sustainability purposes. In the credit risk assessment, the risk management system of a bank should be robust and modernized in parallel to the change in data science. There are four key aspects of successfully predicting the risk of data with respect to the credit assessment application of the ML methods. These aspects are: the ML method(s) chosen, feature selection, data transformation, and validation structure.

2.4.2. Types of Machine Learning Algorithms

Machine learning models can be classified into three categories based on whether labels are provided or generated [6]. In supervised learning, predictions must be made about the labels of instances based on examples with known labels. In unsupervised learning, the instances have attributes, but no labels. The goal is often clustering, i.e., grouping instances based on similarity, to highlight patterns in the data for visualization. In semi-supervised learning, data comes with both labeled and unlabeled instances. As a consequence, clustering algorithms may help in uncovering related structure concerning the labels in the partially labeled data.

There exist different definitions and variations of supervised learning and unsupervised learning. For instance, self-supervised learning is that form of supervised learning where the labels are generated automatically. In contrast, in generative modeling, one tries to generate new data that match the training data as closely as possible. Generation in this context can either be unconditional, i.e., aiming at producing instances without any constraints, or conditional, meaning that some conditions on the data must be satisfied. Finally, some variants of generative modeling concern classification tasks and focus on approximating conditional distributions instead.

The cost of making a wrong prediction is crucial in supervised learning. An area of research called cost-sensitive learning tries to alleviate the shortcomings of learning algorithms with respect to applications with imbalanced classes by adjusting the considered error costs in different ways. Computer-aided prediction support is the task

of assisting human experts, focusing on methods that very reliably recommend predictions that are close to the predictions of a human expert. Ensemble learning is the idea of constructing a hypothesis representation based on multiple simple component hypotheses.

2.5. Application of Machine Learning in Credit Scoring

The heterogeneity of various regulatory frameworks within the European Union (EU) limits the comparability of machine learning models, their predictions and business implications across European banks. Banks accustomed to developing, deploying and maintaining their own credit risk models will be forced to adapt their processes to the new regulations, especially with respect to model validation and documentation. That said, banks that have dealt with machine learning models so far will have a head start over their peers and will most probably face the forthcoming regulations with far less effort needed for a well-intended implementation. As such, this report discusses key machine learning methods relevant for credit risk professionals and outlines recent developments regarding machine learning and the EBA guidelines.

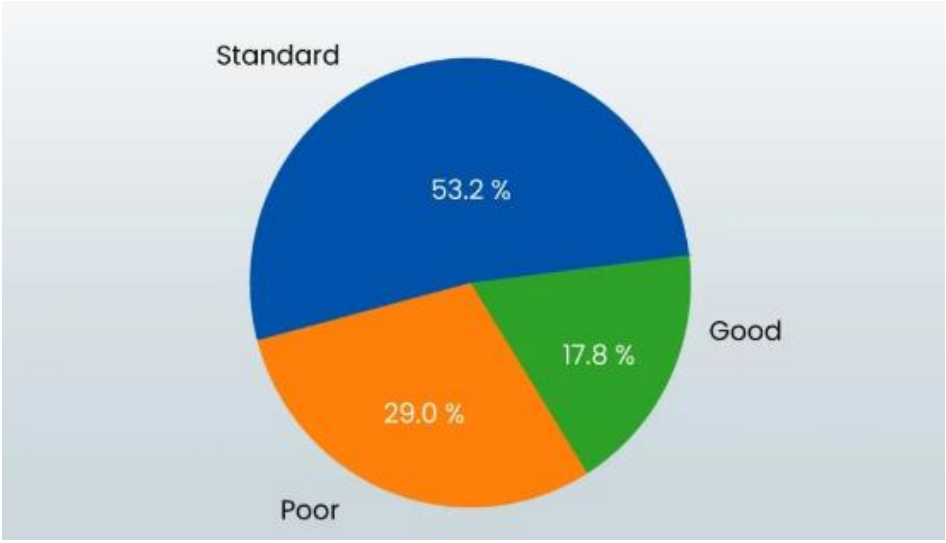

Fig : Develop a Credit Scoring Model with Machine Learning

For a long time, credit scoring systems have been dominated by the application of simple scoring models because they can easily satisfy the requirements of model auditability, whose results can be traced backwards into an audit trail which consists of a series of deterministic functions (or mathematical combinations). Therefore, data, information, and odds ratios are often expressed in simple terms of linear functions of the scorecard; consequently, the weights and cut-offs to form a score can easily be explained. With the

rapid development and automated application of machine learning, primarily tree-based models, data-driven models now appear to behave more like “black boxes” in which there are many complicated calculations in terms of various functions. It is difficult to explain how the input information is transformed into the score, even though all model outputs are interpretable estimation results. In addition, numerous studies have shown that non-credit scoring or actuarial applications of machine learning can deliver a more accurate prediction than scoring based on tree-based models. Recent advancements in model interpretable techniques have allowed the water-shed cognitive gap between prediction accuracy and model transparency to be bridged. Indeed, the game of interpretability with respect to model explanation rather than prediction has been sidestepped. The attempt at machine learning interpretability should basically satisfy the requirements for model input auditability and output interpretability.

2.5.1. Predictive Modeling Techniques

Credit scoring aims at separating good and bad risks in the context of the granting of credit, giving the letter a higher score. The task is typically solved in two stages. In the first one, the risk is modelled using a set of applicant's data, and the model is used to score a new applicant. In a second stage, the score is classified cut-off as an approval or rejection of the credit application. A number of different statistical modelling techniques can be used to tackle the modelling stage in credit scoring. These include approaches based on regression, relative risk, classification trees, and neural networks. All of these techniques have been successfully applied to credit scoring. Nevertheless variants of regression seem to be the models that are more frequently applied, due to good predictive performance, greater availability and ease of interpretation. Yet the availability of very large datasets for credit scoring has been pushing lenders towards trying other techniques, such as classification trees, neural networks, and support vector machines. Due to the good performance of a number of new learning algorithms for building these models, this trend is likely to continue on. However with these new techniques, lenders are confronted with the problem that they are often not so easy to interpret and the risk is not so easily decomposed into explanatory features.

The present paper investigates the option of using boosting to enhance the accuracy of classification trees in the context of credit scoring. Boosting and classification trees are briefly described, and then the performance of the resulting boosted trees is compared with that of neural networks. Results show that boosted classification trees are very competitive in generalization accuracy with neural networks, and find support in the literature. However neural networks generally yield better performance on some datasets. This suggests that classification trees have advantages over neural networks

with respect to the interpretability and insight into the process they provide to the modeller.

2.5.2. Data Preprocessing and Feature Engineering

Machine learning (ML) technologies have become widely researched and utilized due to impressive performance regarding predictive accuracy. Nevertheless, many fields apply these technologies. ML is not inherently interpretable but allows for explanation discovery by applying domain knowledge and theory afterwards. Feature selection is the process of selecting a subset of relevant features (attributes or variables) for building a model. Feature transformation is the process of transforming features into a different domain. ML methods can be grouped into four categories. The content-based approach selects the best features based on metrics and clustering (binary, numeric, text) characteristics. The wrapper approach forms groups of features based on a selection algorithm, then evaluates feature concepts by utilizing them in a learning algorithm. The embedding approach integrates feature selection as part of the learning algorithm. The hybrid approach combines two or more of the other approaches. Despite an increasing number of regulations and guidelines advancing ethical AI and trustworthy machine learning, implementations of robotic process automation (RPA) for scoring and credit scoring systems based on ML models are adopted relatively slowly in traditional banks. This is especially the case for Germany, where processes are still carried out manually, which can be time-consuming and laborious. The best combination of three aspects may vary depending on multiple factors. However, regardless of what these factors are, the number of features impacts the success of a feature selection method. Banking environments discourage the testing of a wide variety of methods for each problem candidate as a machine-hungry approach. Due to its usage in identifying an optimal subset, which can be significantly smaller than the original set, subset-selection methods have gained popularity in predictive modeling. As the focus has shifted from developing features to developing feature selection methods, the number of filter-based feature selection metrics and methods has exploded in the last few years. Highly imbalanced data is a problem associated with many domains in the ML literature. Designing methods that mathematically characterize the imbalance or choosing suitable evaluation measures (particularly with high score) and combining them with methods that identify informative variables are critical steps in ML.

2.6. Case Studies in Machine Learning Credit Scoring

In the first case study, the problem of default prediction on unsecured lending in the credit card context is tackled. A dataset set in the credit industry and containing

characteristics and payment history of defaulted and non-defaulted borrowers is analyzed. From the available data set of 20 variables, the approach involves the selection of features and tuning of hyper-parameters for machine learning classifiers to predict defaults. Several machine learning classifiers such as logistic regression (LR), support vector machines (SVM), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), and light gradient boosting machines (LGBM) are tested. High-class accuracy and an area under the curve (AUC) scores are achieved for all models (LR - 0.8396, SVM - 0.8478, DT - 0.765, RF - 0.8385, XGBoost - 0.8596, and LGBM - 0.8835); all models are able to detect hard defaults quite effectively.

The second case study addresses the issue of reliability and interpretability of machine learning (ML) technologies in credit scoring. The stress on transparency, audibility and applicability of models, which gained momentum with the PROGRESS regulation, is analyzed and current implementations in the credit scoring domain are reviewed. Since most professional credit scoring models are statistical models, the large explainable models provide auditability of the development process and applicability. However, with the advent of ML, their mathematical basis became increasingly complex. Some implementations are proposed for decision trees or neural networks, but for most learning strategies, credit scoring algorithms remain black boxes. In addition to evaluating internal supranational standards or guidance by regulators, a catalogue of methods for ensuring credibility of ML models, a template of processes to encourage sharing of best practices, and approaches for introducing rules of thumb to comply with standards are presented.

2.6.1. Successful Implementations

The most impactful machine learning applications in credit scoring have been either in alternative (non-traditional) credit scoring or in augmenting existing scorecard solutions. There has been a noteworthy trend toward providing credit scores using alternative datasets in many developing markets. Mobile phone data has allowed credit providers to assess risk in markets that were hard to reach before. Companies emerged that score millions of borrowers with this novel approach. Social network analytics has been used for the same purposes and advantageously combined with mobile phone data. These innovative use cases have allowed previously unscareable borrowers to be served with credit, enabling financial inclusion.

At the same time, latent techniques to enhance existing credit scoring systems have come to the forefront of many consumer lenders' transformations. Incorporating techniques such as gradient boosting, extra trees, and deep learning into the first stage of scorecard development has become mainstream since a few studies showed that these machine learning techniques yield better performance than traditional approaches. This, along

with implementations of model interpretation and benchmarking techniques, is the most widely recognized instance of machine learning's second wave of adoption within the credit industry. Tools in the same domain have also emerged in the European market, albeit at a different stage of maturity and sophistication as compared with the North American vendors. However, the corporate credit scoring realm has seen less adoption by machine learning applications, opened up only in recent years.

2.6.2. Challenges and Failures

Various challenges remain for continuous political involvement in credit data privacy, including concerns about advances in data science, widespread datafication across various domains, and the debate over automation and algorithmic governance. Given these dynamics, any public policy would need to be both advanced and flexible, enhancing privacy rights while accounting for the dynamic and context-dependent nature of publicly sensitive credit scoring algorithms. Beyond suggestions concerning discretion over data access and use, policy options for expanded credit data access by alternative providers, clarity regarding data ownership, accountability, ethics, and democratic process are highlighted. Given advances in online datafication, consequential prediction, and dissemination live and across platforms, data protection would need to apply comprehensively across credit data sources. Examples of actionable consumer-facing remedies include the right to audit credit scores through disclosure for broader transparency and contestation rights concerning disputable scoring, filtering and infusion, and fine-tuning of public scores industry-wide. Finally, given how algorithmic credit scoring and scoring institutions are core to an increasingly affluent economy, indeed of the market, the patenting of scoring algorithms is an avenue worth developing towards public accountability, consumer auditability, novelty-sharing, and standards development. Such applications would also benefit long-established proprietary algorithms in industries with associated risks and potential for harm. The recent years have seen the remarkable maturity of machine learning tools to improve credit risk scorecards, especially regarding data capacity and interpretability. Nevertheless, most mature financial institutions still hold credit risk models that rely on a handful of logistic regression variables, significantly under-stacking the potential of predictive power improvement. Also, machine learning instruments are often perceived as a challenge to compliance rather than on-par or preferable instances on explainability.

2.7. Conclusion

The earliest known record of commercial banking dates back to the 15th century B.C. during the reign of the Babylonians. The modern banking system started in the 17th

century, when the Bank of England introduced the various lending facilities that banks use today. Every transaction involves lending money, be it a salary, a loan or bill payment. If a person does not pay money borrowed from another party, the transaction is said to have defaulted. In banking parlance, this transaction is known as a 'delinquency', while the person defaulting is known as a 'delinquent'. Financial institutions (FI) face significant risks when lending money to customers. With the increasing digitization of financial services, financial institutions must predict defaults accurately. However, collecting data for predicting defaults has become an increasing challenge. First, banks lack enough historical records for newly enrolled customers. Second, banks can randomly check a small percentage of customers' records; frequently checking records for all customers will frighten them away and ruin relationships with loyal customers. In recent decades, the growing power of computing technology and data availability has resulted in an explosion of recorded data. In banking, data are essential assets that may provide insights regarding financial behavior. Banks can analyze customers' borrowing and payment history collected over the years to assess their credit history. Prior to this evolution, scoring was a relatively simple process, limited to only a few input variables. The scoring mechanism was typically a linear combination of these variables, which were relevant to the institution and its experience with customers similar to the one under analysis. Today, institutions can handle far more parameters that define their portfolio in far more complex combinations. New end-user technology that produces a stream of pooled data allows institutions to analyze patterns and connections that were previously unseen or uncalculated. There are numerous options among modern computing options, from simple rolling regression models to complex adaptive learning algorithms. The latter are usually called "black box" models because they produce estimates that are often impossible to decompose into their parts. Together with data availability, computing technology has produced a meaningful improvement in credit risk assessment. However, the advanced machine learning approach reshapes the nature of the task, and frequent regression coefficients in simple linear models are not enough to capture the nature of the new data and produce a clear perceived risk. The changing nature of the quality assessment task makes meaningful monitoring obligatory. Prior to the evolution of data availability and computing possibilities, monitoring was performed with a limited set of measures related to Gini coefficients. The new variables today encompass a far greater array of possibilities and notions. However, simultaneous monitoring of the entirety of new variables is impractical. Therefore, clear dimensions need to be defined that describe the quality and behavior of the model. On the level of modeling, aside from the frequently emphasized need for additional information, a clear strategy regarding concurrent techniques is also necessary.

2.7.1. Future Trends

Regulatory challenges regarding the non-accredited use of bank-provided credit scoring models have recently gained traction in Europe and the United States. At the same time, bank-sponsored FinTechs use sophisticated machine learning approaches to provide more value-oriented credit assessments. Such regulatory policies have not followed the rapid FinTech advancement yet. Manual inspections of machine learning risk mitigation strategies are not scalable and incentivize bad behavior. Therefore, regulatory challenges will arise in the future. To this end, regulatory authorities must introduce tighter regulations regarding machine learning credit scoring models and risk mitigation techniques.

Machine learning firms will suffer direct consequences resulting from such regulatory interventions. In particular, these firms are likely to either reduce model complexity adequately or abandon their usage altogether. Mechanistic credit scoring models will make a comeback at the expense of firms presently using algorithms with greater complexity. Because of their robustness, such models will display good generalization and require comparatively simple risk mitigation strategies. Hence, FinTech firms will be at a disadvantage compared to banks. If such a trend arises, regulatory challenges concerning non-designated models will emerge, too. Demystifying model complexity will provide firms with a competitive advantage.

The abrupt introduction of regulation following the rapid advancement of FinTech services poses challenges for supervision. Regulator-initiated reactions to design system risk mitigation policies would go a long way to alleviate concerns regarding accountability and explainability while improving credibility among supervisory authorities. However, asking for stricter regulation without careful considerations risks losing highly competent FinTech firms and/or causing capital flight. Future-oriented intervention strategies must strike a delicate balance between regulatory scope options and execution timing. Audience-specific response policies are required. Broad stakeholder participation and risk incentives involving governance moderation are necessary to make investment- and innovation-friendly regulation pre-empt.

References

Singh, A., & Mehta, S. (2023). RegTech: The Role of AI and Big Data in Regulatory Compliance. Springer.Chen, L., & Zhao, Y. (2024). Machine Learning in Credit Scoring: Techniques and Applications. Springer.

Gupta, V., & Soni, A. (2023). AI-Driven Fraud Detection in Financial Services. Wiley.

Sharma, R., & Bansal, S. (2022). Cloud-Based Financial Services: Architecture and Security. Elsevier.

Agarwal, P., & Verma, R. (2023). Digital Transformation in Finance: The Role of AI and Big Data. Springer.