SvedbergOpen
DISSEMINATIONOFKNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Credit Risk Assessment in Modern Banking: A Data-Driven Perspective

Vijayalakshmi Duraisamy

Senior Consultant in Data Analytics, Independent Researcher India. Email: vijayalakshmi02@gmail.com

Abstract
Credit risk assessment is a fundamental component of lending operations within financial institutions. The growing volume of customer data generated through digital banking channels has transformed the way lenders evaluate borrower risk. Data analytics enables banks to make informed lending decisions by identifying patterns in customer behaviour, predicting default probabilities, and optimising portfolio performance. The integration of statistical techniques, machine learning algorithms, and predictive modelling has significantly improved the accuracy and efficiency of credit risk management. This paper examines the role of data analysis in credit risk assessment, focusing on customer-level evaluation and portfolio-level risk management. It reviews the main data sources, analytical methodologies, performance metrics, regulatory frameworks, challenges, and emerging trends that shape modern credit risk practice, and it demonstrates these through an applied example drawn from a centralised credit analysis engine and an illustrative portfolio framework. The analysis concludes that institutions combining data-driven insight with strong governance are better positioned for long-term resilience.

Keywords: Credit risk analysis; Data analytics; Credit scoring; Machine learning; Probability of default; Portfolio risk management; Stress testing.

I. Introduction

Lending is one of the main ways banks earn money, but it also carries an unavoidable risk. Not every borrower repays a loan as agreed, and when defaults occur, banks can face serious financial losses that weaken their capital strength and overall profitability. For this reason, the ability to assess credit risk accurately has become a central part of running a stable and sustainable financial institution.1

In earlier times, lending decisions were based mostly on experience, personal judgement, and basic financial checks. While this worked to some extent, it often lacked consistency and was difficult to apply at scale. Today the situation is very different. With the help of data analytics, banks can process large volumes of customer information quickly and effectively. By using statistical tools, predictive models, and artificial intelligence, they can estimate the likelihood of default, identify risky borrowers early, and manage loan portfolios more efficiently. In many respects, data has become the backbone of modern credit decision-making.2,3

Understanding Credit Risk

At its core, credit risk is the possibility that a borrower will not repay a loan as agreed. It is one of the most important risks banks face because it directly affects their financial health.

This risk can arise in several ways. A borrower may miss payments or stop repaying altogether, which is known as default risk. In other cases, a bank may lend too heavily to one industry, region, or group of customers, which creates concentration risk and increases vulnerability to economic shocks. There is also counterparty risk, where one party to a financial agreement fails to meet its obligation. Settlement risk occurs when one side completes a transaction but the other does not. On a broader level, country risk arises from political instability, economic downturns, or policy changes that affect a borrower's ability to repay.

Understanding these different types of risk helps banks build safer and more balanced lending strategies.

Role of Data Analysis in Credit Risk Assessment

Data analysis plays a major role throughout the entire lending process, from the moment a customer applies for a loan to the time the loan is fully repaid.

When a customer applies for credit, banks use data to assess how risky that borrower may be. Factors such as income, employment stability, existing debt, and repayment history are considered. During the approval stage,

these insights help determine whether to approve the loan, what interest rate to charge, and how much credit to offer.

Once the loan is granted, the process does not stop there. Banks continue to monitor borrowers over time to detect early signs of financial stress, allowing them to act before small issues turn into major losses.

Overall, this data-driven approach makes lending decisions more accurate, faster, and more consistent. It also helps reduce defaults and allows banks to design better products for different types of customer.

Data Sources for Credit Risk Analysis

Sound credit decisions depend heavily on good data. To understand a borrower properly, banks combine information from several sources.4

Internal data includes customer demographics, account activity, repayment records, transaction history, and past interactions with the bank. This shows how a customer behaves financially over time.

External data provides a wider view. Credit bureau reports show credit scores, outstanding loans, and past defaults. Economic indicators such as inflation, unemployment, and interest rates help banks understand the broader financial environment.

In recent years, banks have also begun using alternative data such as utility bill payments, online shopping behaviour, mobile usage patterns, and other digital footprints. This is especially useful for individuals who lack a long credit history but still demonstrate financial responsibility.

Data Preparation and Quality Management

Before any model can be built, the data must be clean and reliable. If the data is incomplete or inconsistent, the results will not be trustworthy.5

This stage involves correcting errors, removing duplicate records, and ensuring that all data is in a consistent format. Missing values are handled using estimation techniques, while unusual values are examined carefully because they can distort results.

Another important step is feature engineering, which means creating more meaningful variables from raw data. For example, ratios such as debt-to-income or credit utilisation often reveal more about a borrower's financial health than raw figures alone.

When data is properly prepared and well managed, the final analysis becomes far more reliable and meaningful.

Exploratory Data Analysis

Exploratory data analysis is usually the first step in understanding the data before any predictive models are built.

At this stage, analysts seek to uncover patterns and trends in borrower behaviour. They may observe differences in income levels, repayment habits, or signs of financial stress within certain customer groups. These early insights help shape how risk is understood before more advanced techniques are applied.

Exploratory analysis is also important for identifying problems in the data, such as missing values, inconsistencies, or outliers. Addressing these early ensures that the final results are more accurate and dependable.

In simple terms, exploratory analysis turns raw and inconsistent data into something meaningful that can support better lending decisions.

Statistical Models in Credit Risk Analysis

Statistical approaches continue to form the backbone of credit risk evaluation because they are structured, interpretable, and dependable. Although modern machine learning techniques have gained popularity, traditional models remain widely used in banking and fully acceptable to regulators because of their clarity.6

A widely used method is logistic regression, which estimates the probability that a borrower will default. It evaluates key borrower attributes such as income level, employment stability, and repayment history. Its main advantage lies in its simplicity and interpretability, as it clearly shows how each factor influences the risk outcome.

Other established techniques include discriminant analysis, which separates borrowers into risk categories, and survival analysis, which focuses on predicting the timing of potential default events. Together, these models support both risk prediction and effective portfolio decision-making.

Machine Learning in Credit Risk Assessment

The rapid expansion of digital financial services has produced an explosion of data, opening the way for machine learning applications in credit risk analysis. These methods are particularly useful for uncovering hidden and complex relationships within large datasets.7,8

Unlike traditional models, machine learning systems do not depend on strict statistical assumptions. Instead, they learn directly from historical data, allowing them to detect non-linear patterns and subtle indicators of financial distress. Techniques such as random forests, gradient boosting, and XGBoost are commonly used because of their strong predictive performance.

Beyond accuracy, these models can adapt to changing customer behaviour and economic conditions. However, their complexity often reduces interpretability. As a result, financial institutions must ensure that such models remain explainable, transparent, and compliant with regulatory standards.

Credit Risk Metrics and Performance Measurement

Measuring risk accurately is fundamental to effective credit risk management. Financial institutions rely on specific metrics to quantify exposure, estimate potential losses, and support strategic decisions.9

The core building blocks are the probability of default (PD), loss given default (LGD), and exposure at default (EAD). Combined, these form the expected loss framework, which provides an estimate of potential credit losses.

At a broader level, portfolio indicators such as delinquency trends, recovery performance, and non-performing loan ratios offer valuable insight into overall credit health. Monitoring these indicators regularly enables institutions to detect deterioration early and take corrective action before losses escalate.

Portfolio-Level Credit Risk Analysis

Assessing individual borrowers is important, but the overall health of a lending institution depends on how risk is distributed across the entire portfolio. Portfolio-level analysis focuses on aggregate exposure and highlights concentration risk across sectors, geographies, or borrower groups.

Dividing the portfolio into segments allows institutions to identify areas where risk is building. For example, excessive lending exposure to a single industry can create significant vulnerability during an economic downturn.

Tools such as vintage analysis and risk migration studies help track how credit quality changes over time. These insights are essential for improving capital allocation and maintaining a well-diversified lending portfolio.

Stress Testing and Scenario Analysis

Stress testing assesses how a loan portfolio would behave under extreme yet plausible economic shocks, such as recessions, interest rate spikes, or financial crises.

Rather than relying only on historical performance, these tests simulate future adverse scenarios to reveal hidden weaknesses in the portfolio. Scenario analysis complements this by examining how different risk conditions could affect financial stability.

These techniques play a critical role in strategic planning, helping institutions determine appropriate capital buffers and prepare mitigation strategies. Over time, stress testing has evolved from a compliance requirement into a core risk management practice.10

Regulatory Framework and Compliance

Credit risk management operates within a strict regulatory environment designed to safeguard the stability of the financial system. Compliance is not only a legal obligation, it also strengthens trust and institutional credibility.

Global frameworks such as Basel II and Basel III set guidelines for capital adequacy, risk measurement, and governance. Basel III, in particular, introduced more stringent requirements for liquidity, leverage, and capital buffers to enhance financial resilience.11

In parallel, IFRS 9 has transformed credit loss accounting by requiring institutions to adopt expected credit loss models that incorporate forward-looking information. This encourages banks to consider future economic conditions rather than relying solely on past data.12

Challenges in Credit Risk Data Analysis

Despite technological progress, credit risk modelling still faces several practical limitations. One major challenge is data quality, as incomplete or inconsistent records can significantly reduce model reliability.13

Another issue is data fragmentation, particularly in large organisations where information is spread across multiple systems. In addition, shifts in economic conditions can weaken the performance of previously reliable models.

Machine learning introduces further complexity through its limited interpretability, which can make decision logic difficult to explain. Alongside this, concerns related to cybersecurity and data privacy regulation continue to grow. Addressing these challenges requires strong governance, continuous validation, and robust monitoring frameworks.

Emerging Trends in Credit Risk Analysis

Credit risk management is undergoing rapid transformation, driven by technological innovation and increasing data availability. Artificial intelligence and machine learning are enabling faster, more precise, and real-time risk evaluation.

The growth of big data allows institutions to incorporate both structured and unstructured information from diverse sources. At the same time, alternative credit scoring methods are expanding financial inclusion by assessing individuals who lack a traditional credit history.

Explainable AI is becoming increasingly important as organisations seek transparency in complex models. Combined with cloud computing and real-time analytics, these advances are reshaping credit risk management into a more dynamic and intelligent function.14,15

Applied Example: A Centralised Credit Analysis Engine

In a project that implemented a centralised credit analysis engine, a test scenario was used in which a bank receives a large corporate loan request from a mid-sized technology services company planning expansion across several countries. Instead of relying on manual review, the application is processed through a centralised credit engine that integrates machine learning models, policy rules, regulatory requirements, and structured decision workflows.

As part of the process, the system builds a data lineage trail, which tracks the origin, movement, and transformation of every data element used in the credit decision. The bank can therefore trace financial inputs back to their original sources, such as audited financial statements, tax filings, credit bureau reports, transaction systems, and internal exposure databases. The system also records how the data was cleaned, transformed, and used in risk calculations, ensuring transparency and auditability.

The credit engine then consolidates this lineage-tracked data together with external inputs such as industry risk trends, macroeconomic indicators, and cross-border exposure information. During this step, the system identifies that the company is linked to other corporate entities already partially exposed to the bank, which increases overall concentration risk across the portfolio.

Machine learning models next evaluate the borrower's risk profile by calculating key indicators such as the probability of default, loss given default, and exposure at default. Although the company appears financially stable, the combined analysis shows elevated portfolio risk due to sector concentration, large exposure size, and international operations.

The system then applies internal credit policies and regulatory constraints. Even though the borrower is acceptable on a standalone basis, the credit engine flags that group exposure limits are approaching regulatory thresholds. Because every input is traceable through data lineage, risk officers can see clearly which data sources influenced each part of the decision, improving transparency and explainability.

Instead of a simple approve or reject outcome, the credit engine generates a structured recommendation that includes reducing the loan amount, adjusting pricing, tightening repayment terms, and requiring additional collateral. The decision is then reviewed by a credit committee, supported by a full lineage-backed audit trail and an explainability report.

After approval, the system continues to monitor the borrower in real time. Any deviation in financial behaviour is detected early, and because all historical decisions are linked through data lineage, the bank can quickly trace how earlier risk assumptions were made.

Case Study: Credit Risk Analysis for a Retail Lending Portfolio

Consider a bank managing a retail loan portfolio of approximately 100,000 customers. The primary objective is to reduce default rates while maintaining profitability.

The bank gathers detailed customer information, including income, credit scores, repayment patterns, and transaction data. After preparing the dataset, analysts apply both statistical and machine learning models to estimate default probabilities.

The analysis typically shows that borrowers with low credit scores, high debt burdens, and unstable employment histories are more likely to default. Based on these insights, the bank refines its lending criteria, strengthens monitoring systems, and adjusts pricing strategies. Over time, these improvements lead to reduced defaults and stronger portfolio performance.

Table 1 presents an illustrative expected loss framework applied across a set of hypothetical portfolio segments.

Portfolio segment	**PD**	**LGD**	**EAD ($bn)**	**Expected loss ($bn)**

Corporate lending	1.2%	45%	500	2.70
Retail mortgages	2.0%	25%	750	3.75
Credit cards	3.5%	85%	220	6.54
Commercial real estate	1.8%	40%	300	2.16
Investment loans	0.8%	50%	150	0.60

Note: Figures are illustrative and provided for demonstration only. Expected loss = PD × LGD × EAD.

Recommendations

To strengthen credit risk management, financial institutions should invest in robust, well-integrated data ecosystems that ensure accuracy and consistency.

While advanced analytics should be adopted, it is equally important to maintain transparency, fairness, and regulatory compliance. Continuous monitoring and regular stress testing should be embedded within risk frameworks to enable early detection of potential issues.

The responsible use of alternative data should also be encouraged to improve credit access, but it must be governed ethically. Successful implementation depends on close collaboration among risk teams, data scientists, and compliance units.

Conclusion

Modern credit risk management is increasingly shaped by data-driven technologies and advanced analytics. Statistical models, machine learning techniques, and modern data tools have significantly improved the ability to assess borrower risk and predict defaults.

Although challenges such as data limitations, model interpretability, and regulatory constraints remain, continued innovation is strengthening the field. Institutions that combine data-driven insight with strong governance will be better positioned for long-term resilience and sustainable growth. Future work could extend this analysis with empirical validation on real portfolio data and a comparison of model performance across market conditions.

Use of artificial intelligence

[To be completed by the author before submission. If artificial intelligence tools were used in preparing this manuscript, name each tool and describe how it was used, and include the statement required by the journal: 'I have reviewed and edited the content produced by the artificial intelligence tools and am fully responsible for the information contained in the article.' If no such tools were used, state: 'No artificial intelligence tools were used in the preparation of this manuscript.']

References

1. Basel Committee on Banking Supervision, (2017), 'Basel III: A global regulatory framework for more resilient banks and banking systems', Bank for International Settlements, Basel, available at https://www.bis.org/bcbs/basel3.htm (last accessed 3rd July, 2026).
2. Baesens, B., Van Vlasselaer, V., Verbeke, W., (2015), 'Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques', John Wiley & Sons, Hoboken, NJ.
3. Lessmann, S., Baesens, B., Seow, H. V., Thomas, L. C., (2015), 'Benchmarking state-of-the-art classification algorithms for credit scoring', European Journal of Operational Research, 247, 1, 124–136.
4. World Bank, (2018), 'The Global Findex Database: Financial Inclusion and Digital Payments', World Bank, Washington, DC, available at https://www.worldbank.org/en/publication/globalfindex (last accessed 3rd July, 2026).
5. Hand, D. J., Henley, W. E., (1997), 'Statistical classification methods in consumer credit scoring: a review', Journal of the Royal Statistical Society: Series A, 160, 3, 523–541.
6. Thomas, L. C., Edelman, D. B., Crook, J. N., (2002), 'Credit Scoring and Its Applications', Society for Industrial and Applied Mathematics, Philadelphia, PA.
7. 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.
8. Lessmann, S., Baesens, B., Seow, H. V., Thomas, L. C., ref 3. above.

9. Basel Committee on Banking Supervision, ref 1. above.
10. Ibid.
11. Basel Committee on Banking Supervision, (2006), 'International Convergence of Capital Measurement and Capital Standards (Basel II Framework)', Bank for International Settlements, Basel, available at https://www.bis.org/publ/bcbs128.htm (last accessed 3rd July, 2026).
12. International Accounting Standards Board, (2014), 'IFRS 9 Financial Instruments', IFRS Foundation, London, available at https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/ (last accessed 3rd July, 2026).
13. Financial Stability Board, (2017), 'Artificial intelligence and machine learning in financial services', Financial Stability Board, Basel, available at https://www.fsb.org/wp-content/uploads/P011117.pdf (last accessed 3rd July, 2026).
14. Organisation for Economic Co-operation and Development, (2020), 'Artificial intelligence, machine learning and big data in finance', OECD Publishing, Paris, available at https://www.oecd.org/finance/ (last accessed 3rd July, 2026).
15. Rudin, C., (2019), 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', Nature Machine Intelligence, 1, 206–215.