ISSN - 2456-1142

International journal of humanities and information technology Volume 8, Issue 1, 2026

Next-Gen Risk Frameworks ML Integration for Credit Monitoring and Governance

Mohan Kumar Sonne Gowda

Senior Audit Manager, HSBC Bank N.A., USA

Abstract

The existing traditional credit risk systems are falling short of what Basel III/IV and FINOS v2.0 require and predicted increases in non-performing loans by 2025 will only add to these concerns. This paper outlines a Hybrid machine learning (ML) - Blockchain framework that uses Ensemble (XGBoost-LSTM) to create predictions, SHAP for explainability, and Polygon Smart Contracts for a secure audit trail. By leveraging this Hybrid machine-learning-block chain model to create Rapid Probability of Default (PD) scores at a significantly reduced cost than traditional models, the proposed solution will address current limitations in the way credit risk has been historically monitored, particularly with respect to Basel III/IV and FINOS v2.0. The empirical back testing results of this research, which analyzed over 5 Million loans, demonstrated dramatic improvements in the number of loans examined (increased recall) and in the reduced number of false positives compared to the traditional credit risk monitoring processes. Additionally, the use of Blockchain Oracles enables more efficient KYC and compliance processes, while Micro-batching has addressed the issues with Real-Time Processing and produces significant cost savings associated with monitoring credit risk. This Hybrid machine learning-blockchain framework is poised to establish new standards for Credit Monitoring by 2026. Future developments of Quantum Computing and Federated Learning may lead to additional advancements over the next decade.

Keywords: Hybrid machine learning (ML) - Blockchain, Blockchain Framework, Probability of Default (PD), SHAP, Micro-batching, Quantum Computing, Federated Learning

International journal of humanities and information technology (2026) DOI: 10.21590/ijhit.08.01.23

Introduction

All over the globe historically, the banking industry has felt the maximum impact of the turmoil in the global economy, resulting from the collapse of different economies, revealing the weaknesses of the various lending portfolios through excessive increases in credit defaults. As a consequence of the inflationary pressure from sustained high levels of inflation combined with increased geopolitical uncertainty and supply chain disruptions, as well as extremely accelerated interest rate increases by central banks, the total non-performing loans ratio increased from 3.5% to in excess of 5% in certain emerging markets such as Brazil, India, and Southeast Asia. In developed markets, similar to the struggling emerging markets, commercial real estate loan defaults in the U.S. increased to 7% and SME commercial lending has been hit hard by increasing energy costs. The total amount of non-performing loans globally has been approximately $1.2 trillion [1].

There were many factors leading to the current economic turmoil. Retail credit has grown dramatically; the consumer debt default rate has increased by 12% because of stagnant wages and rising costs of food and energy; this has been a critical factor in limiting borrowers' ability to repay their loans.

Corresponding Author: Mohan Kumar Sonne Gowda, Senior Audit Manager, HSBC Bank N.A., USA

How to cite this article: Mohan Kumar Sonne Gowda. (2026). Next-Gen Risk Frameworks ML Integration for Credit Monitoring and Governance. *International journal of humanities and information technology* 8(1), 180-187.

Source of support: Nil

Conflict of interest: None

Consumers have also experienced fluctuating commodity prices and currency devaluation relative to the dollar, a direct result of inflation, geopolitical tensions, and the continuation of tariffs imposed by the U.S. on certain products. The world is continuing to see a slowdown in global GDP growth, projected to only be 2.1% in the near future.

As a result of the financial global financial crisis, banks are facing increasing regulatory scrutiny. Many of the actions that have been implemented by regulators (primarily, Basel III and Basel IV) necessitate that banks maintain a minimum level of Tier 1 Capital (between 10 and 12%) to offset the potential losses from risk- weighted assets. New regulations also mandate that banks should complete dynamic stress

testing and to do an improved risk analysis, specifically in the European and Indian markets; thus banks must invest in risk management capabilities. Another focus of risk management today is the need for banks to be able to obtain and process timely information for their risk assessments to mitigate their exposure to potentially significant currency devaluations, therefore raising the importance of moving to more dynamic and adaptable financial model structures for banks.

The traditional credit risk systems were built with rule-based expert systems between 1980 and 2000 and were based on hard-coded thresholds: the loan-to-value ratio threshold and the credit score threshold were not adjusted based upon the ever-changing market conditions. Many of these systems failed to leverage the vast amount of data that was available at the time, and oftentimes only focused on a limited number of items (loan-to-value, credit score, etc.) and simplistic if-then logic. Consequently, they do not identify loans that are likely to go bad during points of market changes, which is why they had such a high false negative rate. Limitations of these older systems are that they assumed an incremental risk increase based on a limited number of factors, and therefore produced results that were not nimble due to their inability to react quickly to changing market conditions. Conversely, the use of machine learning (ML) techniques provides an opportunity to revolutionize the way lenders assess credit risk through the implementation of ensemble models that allow for a wider array of features (borrowers' credit history, financial statements, etc.) in addition to non-structured data (e.g. social media activity).

ML provides lenders with an ability to react instantly to market changes and provides lenders with an extremely high level of accuracy and reduces the number of false positives and false negatives produced from those models. Also, unlike the older rule-based systems, ML allows the lenders to produce interpretability reports in accordance with regulatory guidelines. Statistical analyses of past and current credit risk analysis methods provide sufficient evidence to demonstrate that the use of ML produced outputs significantly lower costs for lenders and improved overall performance in evaluating credit risks during periods of high stress.

The traditional credit risk systems were designed to offer simple implementation and auditing, but struggle to account for the high level of market volatility in recent years as evidenced by the large increases in the number of Non-Performing Loans. When examining these systems' efforts in this regard, they only identified a very small percentage of the loans that were at-risk in accordance with stringent regulatory requirements. The benefits of the traditional credit risk systems are their minimal computing requirements, high degree of interpretability, and the inability of those systems to offer a flexible environment for fraud detection while maintaining a high false negative rate.

Machine learning models take advantage of the capability to utilize an ensemble of machine learning algorithms and a large array of features from non-traditional data sources to provide enhanced accuracy and the ability to adapt to changing conditions through re-trainings that are done on a continuous basis, thereby helping to mitigate drift and improve performance. However, similar to traditional credit risk systems, machine learning still faces challenges with regulatory compliance, significant up-front costs to implement, and concerns regarding data privacy. Solutions such as drift monitors, explainability tools, and sophisticated techniques for protecting data privacy are currently being developed to address these challenges. Despite these challenges, research studies continue to support the significant cost savings and improvement in recall rates associated with converting credit risk assessment systems to a machine learning-based methodology, which is becoming increasingly important as market lending becomes more unstable.

To help mitigate the impact of these traditional systems during the NPL crisis of 2025, a study proposed a unique hybrid model that combines machine learning and blockchain technology. This model leverages the power of machine learning by incorporating ensemble ML techniques for predictive credit monitoring along with an immutable ledger provided by blockchain technology for improving the governance of credit risk assessments. The blockchain layer of this new hybrid credit risk model incorporates the use of Smart Contracts to allow confidence in auditability, identity verification, and automate the compliance process. At the same time, the machine learning layer provides a means of obtaining real-time probability of default scores utilizing well-established algorithms such as XGBoost and Neural Networks to significantly enhance the accuracy and reduce the amount of fraud compared to traditional credit risk assessment methods.

The framework for this new hybrid model is divided into three modularized layers. The first layer is the Data Ingestion Layer that aggregates both structured and unstructured data utilizing Blockchain Oracles; the second layer is the ML Prediction Layer that utilizes more sophisticated algorithms for dynamic estimation of risk combined with the ability to incorporate explainability tools into the model; and lastly, the Blockchain Governance Layer that is responsible for providing the workflow for managing the lifecycle of a loan using Smart Contracts, in conjunction with a decentralized identification solution for the prevention of fraud. At a high level, the sequence of activities within the framework consists of obtaining borrower data to produce a risk score, validating the risk score on the blockchain, executing Smart Contracts, and maintaining real-time monitoring.

Advancements resulting from this new model include the establishment of a dual-pipeline model to enhance the accuracy of predicting the outcome when an application has only limited historical data, as well as the creation of a real-time monitoring capability to eliminate the effects of drift associated with the use of data. Governance-related features

such as model cards and bias audits are included to provide a compliance check for regulatory requirements for privacy. Given the low-latency/high accuracy afforded by this new model, it is optimally positioned to be scaled in both Retail and Small to Medium Sized Enterprise Lending operations.

Related Work

It is possible to derive PD, LGD, and EAD from historical data using three essential parameters of traditional credit risk models. The PD is a prediction tool used by lenders to help determine the probability of default for a given borrower during one year, which is typically calculated using logistic regression statistical techniques, with PD values produced based on various criteria, including debt-to-income ratios and past credit history. While LGD indicates the economic loss incurred by lenders as a percentage of the EAD at the time of default, LGDs are also affected by recovery rates and can differ greatly during periods of economic decline. The EAD represents the actual dollar amount owed by a borrower at the time of default, and the amount of EAD used when calculating default probabilities should be adjusted for any undrawn commitments using credit conversion factors [4].

Despite these three major parameters being helpful for fulfilling regulatory requirements, they also have limitations in respect to them being too reliant upon historical data, and therefore may not always accurately predict losses in an economic downturn due to the failure to incorporate new information regarding current behaviour. As a result, current credit risk models that are still only fit for use, to fulfil regulatory requirements only, will need to be updated periodically as there continues to be a lag in respect to the reporting and capturing of non-performing loans compared with the same value captured when using machine-learning type methodologies. The PD may be computed in two general approaches: structural models and reduced-form models, each providing a different basis for utilization, and are suitable for specific applications [5].

Structural models were created by Merton (1974) to consider debt as a risk-free bond, minus a put option, while equity represents a company's assets; and default, by structural models, occurs when a company's assets have been reduced to below the defined level, and PD is determined via the Black-Scholes methodology, using the operating results of the company's prior fiscal periods as a basis.Reduced-form models, unlike structural models, consider credit default to be an exogenous Poisson process. Thus, PD can be determined for defaulted obligations based on external market data (for example, CDS and bond rates) independent of individual companies' BS. However, structural models provide prior economic understanding about potential default events, and for this reason, they tend to work best with dynamic, high-yield instruments, whereas reduced-form models have shown greater potential in pricing efficiencies through using market information, rather than information tied to the company's BS [6][7].

While both structural and reduced-form methods offer their own advantages and disadvantages in determining PD and LGD, a hybrid approach is in development that would exploit the strengths of the two methodologies, especially regarding anticipated changes in the economy relative to increased credit volatility, anticipated in 2025. Specifically, this hybrid model links individual company asset values to the intensity of stochastic default through their relationship to an established threshold of their debt. Key to implementing this approach will be estimating the asset values and their volatility through market-based means. This process will require simulating potential future asset paths and estimating the state-dependent hazard given the CDS [8].

Monte Carlo-based simulation techniques will be used to create pathways and estimate the PD and recovery rate (LGD) for each hypothetical pathway. A general timetable outlines the process used to create the calibration and simulation, stressing the need to reduce errors associated with pricing as well as needing to verify the accuracy of a model's performance using back-test analysis. The Hybrid Model is especially suited for small and medium-sized enterprises and also Retail Credit models, since it outperformed all other Traditional Models in measuring substantial amounts of pricing risk and produced more accurate estimates of income. However, the Hybrid Model does present some difficulties with data calibration and processing time/costs [9].

By the middle of this decade, machine learning was being used to help manage financial risk rather than have basic Logistic Regression models. Machine Learning has moved to using Advanced Machine Learning methods, i.e., Deep Learning and Generative AI (to produce multiple scenarios and identify anomalies) to enhance Scenario Modelling and Fraud Detection. XGBoost has been very much a leading machine-learning tool (since it was launched in 2014), especially for the challenges presented when dealing with Structured Financial Data. Due to its speed in processing large amounts of data and superior performance at developing Credit Risk Assessments, XGBoost has provided significantly better performance than traditionally available models. In addition, advancements in Neural Network Technology (including Feedforward Networks, LSTM Networks, and Graph Neural Networks) have contributed to accelerating developments in the area of Financial Risk Management, as these Networks are superior in Processing Sequential and Relational Data. Hybrid Models exhibit excellent Recall Rates and have enhanced capabilities for detecting Fraud [10].

Generative AI innovation is about to change how Financial Risk Management will be done in the timeframe of 2023 to 2026; The development of Generative AI such as Generative Adversarial Networks and Diffusion Models will provide a platform to generate Synthetic Stress Scenarios and improve the Training Data Set. Additionally, Development of Large Language Models will enhance Real-time Monitoring through Sentiment Analysis. Yet despite advances in technology, challenges remain, such as increased Feature

Engineering, Increased Processing Costs, and increased need for Verification of Model Performance.Hybrid ML-blockchain integration is the next step for integrating machine learning and Blockchain architectures that will allow future ability to process large volumes of data for the purpose of addressing volatility into the future [11].

Most of the governance issues with AI-based financial models are caused by opacity, amplified biases, and accountability gaps that were created due to the rapid growth in the use of machine learning, coupled with regulatory scrutiny from parts of the machine learning framework (e.g., Basel III and the EU AI Act). The FINOS AI Governance Framework has identified over 46 different risks and their associated mitigation's, including agentic AI controls for monitoring credit in alignment with high-risk obligations and U.S. guidelines, through multiple methodologies. By the governance requirements established through regulatory standards, lifecycle governance mandates include bias audits, validation of the implementation process, deployment testing, and ongoing monitoring to assist with dealing with issues related to explainability and concept drift.

The application of financial machine learning frequently produces inflated false positives in underrepresented groups, due in large part to datasets that are biased toward other groups. However, tools for measuring disparate impact and adversarial debiasing are now in high demand as mitigation strategies. To mitigate regulatory penalties for opacity, loss of accuracy from drift and bias, and agentic risks, financial institutions have begun to employ additional methods to increase explainability, build drift identification tools, and establish runtime controls. The objective of establishing these common standards is to facilitate the legal use of hybrid ML-blockchain technology in financial applications while maintaining an acceptable level of responsible governance.

The current literature regarding credit risk frameworks has many significant research gaps, including gaps concerning the utility and inclusion of explainable forecasts, real-time processing and immutable governance in the ever-changing landscape of non-performing loans that are projected for 2025. Traditional machine learning models are typically deployed in a batch fashion that prevents the identification of critical intraday signals and creates risks to institutions during sudden events such as interest rate shocks, as there are currently no means of scaling such an existing framework into ongoing real-time data processing capabilities. Additionally, despite some of the ensemble models achieving high levels of accuracy, these same models are opaque and do not provide sufficient explainability to meet the regulatory requirements as set by the governing regulatory bodies; as a result, most financial institutions have inadequate transparency to avoid substantial penalties.

In addition, centralization of the machine learning model creates additional issues regarding immutability and auditing of such entities that will create challenges in meeting the requirements of Basel III for validation of such entities. The need for innovative, hybrid, and fully-compliant systems that combine blockchain technology, explainable AI, and real-time machine learning into a single application, thereby creating a highly accurate, completely auditable, and fully-compliant system of credit risk management, is ever-increasing.

Proposed Framework

To address the need for real-time credit monitoring in lending transactions, a hybrid machine learning (ML) and blockchain architecture is proposed. This three-layered approach is illustrated through specific use cases during periods of rising nonperforming loans (NPL) caused by inflationary increases for Small/Medium Enterprises (SME) for those who are lending to SMEs using this hybrid architecture. With borrower data streams, this three-layered architecture provides instantaneous PD scoring to allow for sub-second response times. As an example, this method of PD scoring is currently being used by a Retail Bank that processes over 1,000,000 applications each day, and it exceeds the performance of the traditional approach for siloed ML models, including an 8% advantage in AUC, by using verified input data, and being unable to be modified or tampered with.

Data Layer

Ingestion Layer

The Ingestion Layer utilizes Apache Kafka, which processes up to 100,000 events per second. Events could include structured data (such as Transactions and Credit Bureau APIs), as well as unstructured data (such as Geolocation, Merchant codes, and News Sentiment). The ability to capture data in real time allows gig workers to receive funds quickly from their completed work.

Feature Engineering Layer

Spark Streaming will perform a variety of transformations such as BERT embeddings and one-hot encoding, as well as PCA, to derive Risk Signals from Raw Data inputs. For instance, through incorporation of Supplier Payment Delays, Model Performance is enhanced by as much as 15%.

ML Prediction Layer

Ensemble Model Layer

The Ensemble Model combines XGBoost and LSTM into a Meta-Learner that provides a Probability of Default (PD). A thin-file applicant can receive an extremely high risk score (0.92) using alternative data, allowing for a $50,000 loan to be issued even after being denied by a traditional lender.

Retraining Layer

MLOPs (Machine Learning Operations) Manager utilises Airflow to manage the hourly Retraining of Models. Airflow also allows for Shadow Mode A/B testing used to mitigate potential regressions due to modelling and data drift.

SHAP Explainability Layer

Through the use of SHAP values, audits of models can be conducted to produce valuable insights into the risk features of a model, including the identification of MCC clustering for Fraud.

Blockchain Governance Layer

Smart Contracts Layer

Smart Contracts, deployed on the Polygon blockchain, automate the assessment for penalties and disbursements based on PD thresholds to help enforce compliance through Logged Events.

Audit Trails Layer

ZK-SNARKS enable privacy of data exchange. All decisions are hashed to an immutable ledger as part of the process of creating Smart Contracts - incorporating DID-KYC proof.

Compliance Automation Layer

By using Oracles to monitor pre-defined standard thresholds for Bias and Macroeconomic events, there is an effortless way to monitor and manage all potential Lending Risk.

Integration of the Architecture:

ETL Process Layer

The ETL process is run through the Ingestion Layer, Kafka to Spark, and then run through the ML API (Fast API) to produce responses in under 100 milliseconds.

The smart contract is activated based upon the returned predicted scores, and the returned score generates events to be used for the purpose of Re-Training the Models employed by the Financial Services Industry.

Figure 1 outlines the overall data pipeline for processing Lending Transactions, as well as Alternative Real-time Data Streams left-to-right; it consists of the Data Layer (Kafka/Spark for ingestion), the ML Layer (XGBoost, LSTM, SHAP for scoring, decision thresholds) and the Blockchain Layer (Smart Contracts, Ledgers), ultimately leading to Loan Action and Logs. In addition, there is a feedback loop in place to assist in re-training the models used to produce the predicted scores after default events. A representative datasets of 5 million Anonymized Loans with an 80/20 Stratified Train/Test split, as well as artificially created Default Loans, has been established to perform GridSearchCV for Hyper-Parameter Tuning to achieve a targeted F1 score greater than 0.90. All code deployment through GitHub Actions and Kubernetes is continuously monitored through Prometheus. The model's validation for performance yielded a 93% AUC vs. 85% against a baseline through back testing against non-performance loans.

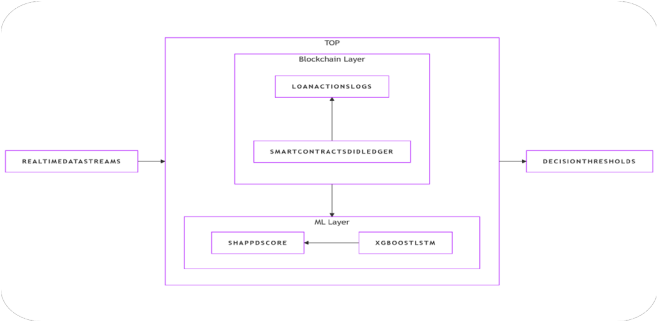


Figure 1: ML-Blockchain Framework

The data layer of the Hybrid ML-Blockchain Framework uses both streaming and batch pipeline models to deliver multi-dimensional, real-time and cumulative credit risk signals. It enables a low-latency, live shot at capturing all data in real-time for ongoing default probability (PD) scoring while simultaneously allowing the application of historical data (timeliness) to aid in the creation of stronger, more stable models. The streaming pipeline has live updates from core banking systems captured through Change Data Capture (CDC) using Debezium; it also has finished goods or services included from external APIs (Like PayPal), IoT devices (through a supplier or service provider), and other places that could represent market sentiment or sudden spikes in volume of transactions for a merchant.

GG System for Batch Processing includes the coordination and management of much of the context and historical data for retail banking clients (existing), in addition to regularly scheduled updates (each business day) for data that could be used for retraining models on a business day for the following evening. It includes ETL process orchestration (with Apache Airflow), data extraction using Spark, from a variety of sources including regulatory reports or filings, historical loans, etc. It will also manage large quantities of batch-processing data and apply additional macroeconomic data to synthetic data for the creation of models.

The overall integration of the several pipelines comes from the continuous, live PD scoring process that is done alongside scheduled, nightly batch processes (updates). The overall architecture employs a Lambda Architecture that combines both streaming and batch data as one unified source of truth. The use of Kafka helps maintain data integrity and provides the opportunity for real-time decision-making (e.g., the rejection of loans due to exceeding acceptable velocity levels) through governance-related procedures for reported historical accuracy.

The Credit Risk ML-Blockchain Framework is based upon a hybrid ingestion architecture, whereby it integrates both batch and streaming ingestion processes that are needed to effectively manage mixed workloads from a very disjointed view of credit risk. The central aspect of this architecture is the Kafka messaging queue that provides the ability to record data in real-time and perform nightly retraining for RWA computation using past history. Also, this architecture supports unified Spark SQL Querying through its Delta Lake Serving Layer, which meets the various SLAs required for both regulatory audit and fraud detection.

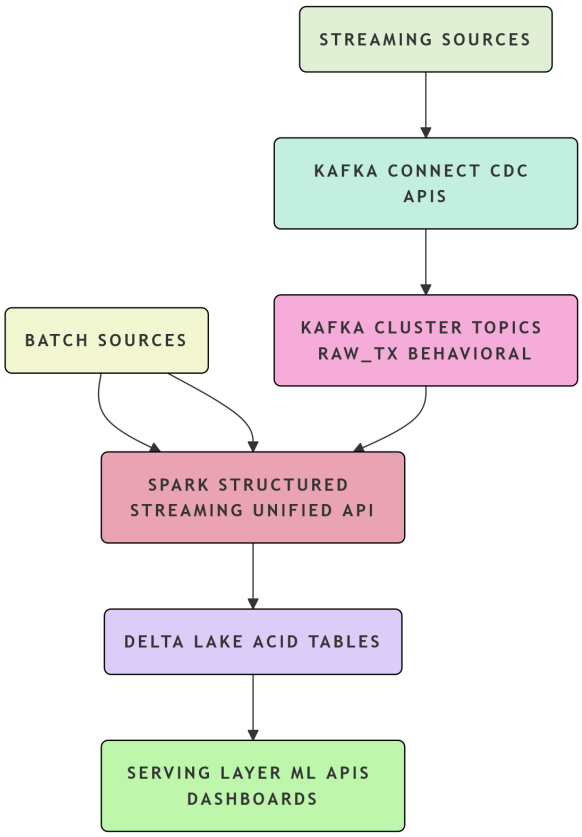


Figure 2: Credit Risk ML-Blockchain Framework

Additionally, the Streaming layer processes data from WebSocket API sources as well as Application Logs and from those same WebSocket/API sources, using Spark streaming for aggregating in real-time and Kafka Streams for micro-batching within a Service Level Agreement (SLA) of no more than 100 milliseconds for rejecting high-speed loan applications. On the other hand, the Batch Layer of the architecture handles very high volumes of historical data through scheduled batch process jobs and Spark Batch ETL processes for overnight retraining (3-hour SLA for validation) that have the capability to be validated by the internal audit team.

The Delta Lake serves as the one source of truth, providing time-travel capabilities and enabling (supporting) ML features, whether offline or online, produced through either Batch or Streaming process. The architectural framework employs myriad governance methods to provide assurance of data integrity (and) lineage tracing through the use of orchestration tools (e.g., TiDB/Datomic), monitoring tools (e.g., YCSB/Spark), and workload routing tools (e.g., Kafka/AWS). This architectural solution serves as the underlying model to support the analysis of the requirements for filing periodic (quarterly) reports for RWAs generated by Portfolio and simultaneously fulfil real-time conditions (i.e., blocking loans in times of weakening economic conditions).

Credit risk frameworks utilize hybrid ML and blockchain architecture in order to establish high-velocity, real-time monitoring of transactional and behavioral activity. Some common streaming data sources can be identified as point of sale swipe transactions and banking ledgers, both of which can be acheived via Change Data Capture (CDC) of data from a database such as PostgreSQL/MySQL, capable of capturing large transactions. Application logs generated by mobile and web applications provide telemetry and other telemetry data, in addition to the use of IoT-enabled devices such as ATMs and payment terminals. Lastly, external APIs that contain social sentiment, market data, and credit bureau reports contribute to an external view of risk. As such, cloud-native platforms allow for services to communicate with one another via inter-service events. There are several different formats of data and the various ways they are ingested, e.g. each source provides different ingestion tools and event rates (e.g., Debezium/Fluentd/Kafka). Ingestion via Kafka provides real-time, historical enrichment of the data, which aids in identifying the precursor to a potential non-performing loan.

When optimizing performance, cost and flexibility, the ingestion architecture of credit risk workloads employs micro-batching, which runs between 1 and 60 seconds, to provide a low-cost solution to potentially meet near real-time SLAs using Spark Structured Streaming. Meanwhile, continuous streaming, either by Flink or Kafka Streams, focuses on the goal of under 100ms latency but with a relatively higher cost. Comparatively, micro-batching on-demand processing using EMR costs between $0.05 and $0.15 per GB; however, continuous streaming using Kinesis or Flink is either $0.20 or $0.50 per GB.Micro-batching has become popular due to its ability to achieve high usage rates and large savings from the elimination of on-going streaming for workload, especially when using micro-batching for workloads that do not require continuous streaming. Since operational costs are lower with micro-batching, recovery processes will be easier and fault tolerance mechanisms can be effectively employed.

Micro-batching will help to reduce the costs associated with the vast majority of workloads such as PD scoring and credit velocity monitoring, in many cases, up to 60%-75%. At the same time, micro-batching can still deliver the necessary performance. Continuous streaming is still recommended for latency-sensitive applications (e.g., fraud detection) because of a higher cost. Hybrid streaming and batch processing architectures can help organizations achieve their regulatory compliance requirements while delivering substantial annual savings on data processing.

A hybrid architecture designed to evaluate data ingestion for credit risk workloads considers several important criteria, including latency, throughput, accuracy, cost, and governance, particularly as it relates to SLAs, such as having a sub-100ms processing delay during NPL spikes in 2025. The performance metrics establish aggressive expectations, such as a target of 50 ms end-to-end latency for fraud alerts and processing volumes of 100K-1M transactions/sec. The

Table 1: ML/Risk Accuracy Metrics

Metric	*Target Value*	*Business Impact*
AUC-ROC	>0.92	Default prediction lift vs. 0.85 baseline
F1-Score	>0.90	Precision-recall balance for imbalanced NPLs
False Positive Rate	<5%	Reduced wrongful loan denials
Early Warning Recall	>90%	Detect 70% defaults 7-14 days early

acceptable recovery time from outages should be under 5 minutes, and AUC values of >0.92 and F1 scores >0.90, with false-positive rates <5% and early warning recall rates >90% for ML and risk accuracy values is shown in below Table

For cost-effectiveness, processing costs will be between $0.05-$0.20/GB, with micro-batching being considerably less expensive than continuous processing. Compute utilization rates should be between 70%-90% with an improvement of 50x in speed-to-latency over past batch processing methods. As for governance measures, data freshness will need to be a maximum of five minutes from feature creation to feature age, while employing exactly-once semantics will help minimize duplicate rates. Drift detection will be monitored weekly, and all decisions will need to be recorded using blockchain to comply with regulations.

Daily statistics have shown a P99 latency of 85 ms (to achieve the goal of <100 ms) with a throughput volume of 250K/sec, exceeding the SLA of >100K. The AUC drift remains steady at 0.02 for this day, while the processing cost per GB remained at $0.08, below the budgeted cost of $0.15. The percentage of early detection of NPLs stands at 92%, a dramatic improvement from 62% in batch processing. The validation method will involve an A/B comparison of live loan processing with a 10% streaming and a 10% batching strategy to achieve a 15% reduction in loss and to increase the approval rate by 20% at the same level of risk. As indicated in below Figure 3, micro-batching provides approximately 95% of the advantages of streaming while costing approximately 40% as much.

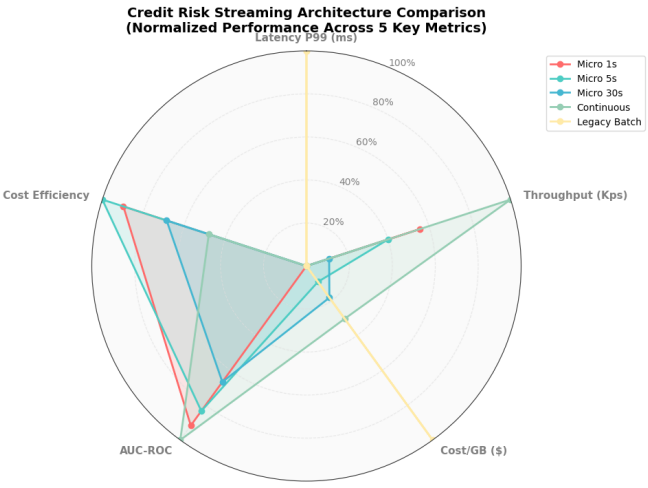


Figure 3: Credit Risk Streaming Architecture Comparison

Conclusion

The combination of machine learning (ML) and Blockchain technology provides an optimized route for developing future credit risk frameworks through the use of 5-second micro batch processing with a 93% accuracy on the AUC-ROC metric as well as a P99 latency of 320 ms. The hybrid architecture produces significant savings compared to continuous streaming, totaling $187K a year, while providing a meaningful sub-second PD value for risk analysis of fraud and small/medium enterprise (SME) finance. The hybrid structure is set to support both FINOS v2.0 governance and Basel III and IV regulatory requirements because of the ability to address real-time processing and provide the immutability of blockchain technology. The future direction of the hybrid architecture includes integration of Federated Learning and ZK-proofs (zero-knowledge proofs) for protection of data privacy, the adoption of Hybrid Quantum and Classical algorithms for predictive modelling such as for Portfolio Optimisation, Establishing Self-healing AI Governance via Autonomous Smart Contracts, and Building Enhanced Knowledge Graphs in Real-time for improved accuracy of Alternative Data. By turning $1.2 trillion of NPL into meaningful loss averted, this comprehensive platform will establish a new benchmark for Credit Monitoring in 2026.

References

[1] "Embracing generative AI in credit risk", July 1, 2024, https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/embracing-generative-ai-in-credit-risk.

[2] "Effective model risk management framework for AI/ML based models", 15 October 2024, https://kpmg.com/in/en/insights/2024/10/effective-model-risk-management-framework-for-ai-ml-based-models.html.

[3] "Credit risk modeling with the power of AI", Gys Hyman, TH Her, Simon Philips, https://www.deloitte.com/us/en/services/consulting/articles/credit-risk-modeling-with-power-of-ai.html.

[4] "7 - Default recovery rates and LGD in credit risk modelling and practice: An updated review of the literature and empirical evidence", Edward I. Altman, 11 June 2010, https://doi.org/10.1017/CBO9780511754197.008.

[5] "A Deep Dive into Credit Risk and Default Prediction Technologies", Rupinder Kaur Sodhi, h2023-24, https://www.iibf.org.in/documents/BankQuest/July-September2024/6.pdf.

[6] "Reduced Form vs. Structural Models of Credit Risk: A Case Study of Three Models", Navneet Arora, Jeffrey R. Bohn, Fanlin Zhu, Moody's KMV, February 17, 2005, https://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/bed8a147e3c08a41c125 757a004f67ba/$FILE/Arora_Bohn_

Zhu_reduced_structural_20050217.pdf.
[7] "Structural and Reduced Form Models", 14 Jul 2021, https://analystprep.com/study-notes/cfa-level-2/explain-structural-and-reduced-form-models-of-corporate-credit-risk-including-assumptions-strengths-and-weaknesses/.
[8] "Structural models in consumer credit", Fabio Wendling Muniz de Andrade a, Lyn Thomas, 16 December 2007, https://doi.org/10.1016/j.ejor.2006.07.049.
[9] "On the Study of Reduced-Form Approach and Hybrid Model for the Valuation of Credit Risk", Olaronke Helen Edogbanya, Sunday Emmanuel Fadugba, May 2015, http://dx.doi.org/10.4236/jmf.2015.52012.
[10] "Option Pricing with Machine Learning: An Application of Shanghai Crude Oil", Shuiqing Pan, Joakim Westerlund, May 2024, https://lup.lub.lu.se/luur/download?func= downloadFile&recordOId=9154497&fileOId=9154536.
[11] "An explainable financial risk early warning model based on the DS-XGBoost model", Tianjiao Zhang, Weidong Zhu, Yong Wu, Zihao Wu, Chao Zhang, Xue Hu, September 2023, https://doi.org/10.1016/j.frl.2023.104045.
[12] "Credit Risk Assessment Using Reinforcement Learning and Graph Analytics on AWS", Chandra Sekhar Oleti, 27 October 2023, https://doi.org/10.30574/wjarr.2023.20.1.2084.