Optimizing Regulatory Compliance with Machine Learning: Boosting Accuracy and Efficiency

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

Abstract
Organizations face enormous difficulties in regulatory compliance across industries including banking, healthcare, and manufacturing as a result of manual processes leading to high error rates, costly penalties, and long delays in operation. Current approaches for managing the complexity of regulation through traditional rule-based systems and anomaly detection across vast archives of audit records do not meet the needs of these organizations. In this paper, I describe a new ML pipeline that includes an XGBoost classifier, BERT transformer for NLP, and Isolation Forest for anomaly detection using Snowflake's ML functions and Apache Spark's GPU clusters to allow for easy scalability. The ML pipeline automates the data import process from large datasets of compliance data, creates tokens and features, and allows for very fast in-database inference with extremely low latency. The ML pipeline includes MLflow for tracking model performance and feedback. The ensemble model has demonstrated exceptional performance with respect to both accuracy and efficiency; reduced processing time; and provided significant financial benefits for organizations that applied it in a number of case studies. The deployment of the ML pipeline provides for governance and explainability through the use of Snowflake Horizon and changes compliance from a reactive liability to being an active asset. Plans for the future include improving regulatory interpretation and scalability through the use of federated learning.
Keywords: Rule-based Systems, Anomaly Detection, XGBoost Classifier, BERT Transformer, Snowflake Horizon

Introduction

Financial institutions throughout the world are forced to continually implement compliance frameworks more rapidly than ever before in order to maintain trust among their customers, as well as operate with integrity. If an institution does not comply with all applicable laws and regulations in its operational jurisdictions, it could face significant negative legal, financial and reputational consequences, thereby jeopardizing its ability to continue operating. Banks will need to be able to manage changes in regulations and compliance all the time, as regulations are constantly changing. By taking a proactive approach to analyzing the regulatory environment and making required changes to internal operations in a timely manner, banks will improve their governance procedures and lessen confusion around compliance. A systematic approach to managing compliance and addressing compliance risk will help to build trust between banks, regulators, customers and other stakeholders [1].

As financial institutions face greater complexity in complying with regulatory requirements; the need for banks to transition to fully automated solutions from manual processes becomes even more critical. Since there is much inconsistency between countries with respect to how banking regulations impact operations, banks will take on considerable risk (potential fines and reputational loss) if they do not have the resources in place to effectively keep up with continuously changing regulations. If they continue to fail to comply with regulatory agencies, banks may ultimately lose the trust of their customers and investors, and suffer adverse consequences. To help resolve these issues, financial services providers will now be better prepared to efficiently comply with regulatory requirements and improve compliance speed and quality through the deployment of technologies, including Artificial Intelligence (AI) and Natural Language Processing (NLP). The new proposed compliance monitoring and analysis system will include several functions to help providers to manage compliance more effectively and efficiently, including: (1) real-time notifications of regulatory changes, (2) audit-ready compliance documentation, (3) centralized reporting capability to track compliance status, and (4) machine learning function to analyze past compliance trends and detect potential compliance risk. Improvement of employee productivity, streamlining of compliance processes, and providing organizations with the means to proactively and effectively respond to changing regulations are the goals [2].

The proposed framework for regulatory compliance will define the regulatory compliance process as defined by various National and International regulatory agencies and will require timely stakeholder engagement from

Compliance Officers, Information Technology (IT) Departments, Business Partners, Data Administrators, and Change Managers. Business Partners will be responsible for operating the Compliance Solution, and the Compliance Departments will be responsible for ensuring that accurate regulatory documentation, data management, and data quality complies with governing rules. Senior Leadership will provide resources and support to the organization, and the Technology Department will provide the development of AI and NLP-based tools for automating compliance activities, creating a culture of proactive compliance through enhanced stakeholder interoperability and communication. The fragmented nature of the United States' regulatory environment will add to the complexity of compliance.

The fragmented regulatory environment in the United States is due to multiple federal and state regulatory agencies, such as the Securities and Exchange Commission (SEC), the Financial Industry Regulatory Authority (FINRA), etc., which all contribute to the fragmentation of the regulations and can also create delays and challenges when complying with regulatory requirements. Organizations use continual data monitoring to forecast their risks based on regulatory compliance with the U.S. Banking Regs (i.e., the BSA and Dodd-Frank); and rely on AI and NLP technologies to improve their regulatory compliance record by translating the often-complicated and/or fuzzy language of World Bank Regulations, into clear, understandable terms. AI & NLP can also accelerate the regulatory updates that have occurred since 2014 through enhanced automation, which streamlines the monitoring and reporting processes of regulatory compliance that continue to present many challenges for regulated organizations. Tracking regulatory amends with a more centralized and cohesive documentation system also increases an organization's audit preparation by improving its operational resiliency and efficiency when compared to non-centralized documentation methods [3].

There are also many audit compliance-related challenges that regulated organizations, such as banks and health care providers, face as a result of their reliance primarily on manual practices and processes to comply with regulations, which result in a significant amount of errors in regulatory compliance. For instance, the KYC due diligence that financial institutions perform on potential customers can result in errors as much as 20-30 percent due to information overload or fatigue from performing KYC due diligence 8-10 hours per day, 5-6 days per week; the number of regulatory compliance-related errors, especially indirect data errors resulting in penalties, continues to grow and will last for future regulatory actions until manual processes are replaced with an automated system. (e.g., $1.2 billion was issued in regulatory penalties for AML compliance failures against one of the 10 largest banks in the US in 2024).

Within the manufacturing industry, manual processes related to supplier auditors can create significant disruptions on production lines for several days. Using Automated Detection via machine learning (ML) technology will improve efficacy in addressing many of these issues. One specific example of this is the possibility of an anomaly detection capability that uses isolation forests (an ML algorithm) to help payment processors detect fraud in their customers' transactions in nearly real-time, achieving accuracy levels above 95% as compared to reviews of claims through traditional means (approximately 75% accurate) and contributing to substantial decreases in both fraud losses and time spent resolving fraud cases (as illustrated by a European case study, from weeks down to days). In addition, the utilization of natural language processing (NLP) technology could also help healthcare facilities achieve HIPAA compliance by validating patient data records against regulatory documentation and would lead to significant decreases in time required to perform audits of patient data records (from days down to minutes).

The many types of regulatory challenges present different types of problems within the heavily-regulated banking, healthcare, and manufacturing industries. Manual auditing and reviews of compliance with requirements typically yield an error rate of 20%-30 % from the total number of compliance transactions processed by any specific entity. Both manual and automated auditing of transactions are also subject to oversight errors; these types of errors can incur large financial penalties, harm an organization's reputation, and cause lengthy periods of disruption to an organization's operations. The adoption of ML technologies will help reduce or eliminate the need for manual audits by automating detection and predictive analytics capabilities.

For example, utilizing anomaly detection would allow the immediate detection of differences between transactional data, and utilizing natural language processing (NLP) will assist in the extraction and analysis of the extensive documentation associated with compliance with regulatory requirements to identify and summarize the requirements associated with different regulatory standards. The application of these technologies has resulted in a disruption of traditional methods utilized by many organizations by promoting the ability for timely processing of compliance-related transactions (from several days to just hours), with an increase in accuracy to greater than 90% for compliance-related transactions.

The purpose of this research study is to demonstrate the practical application of ML technologies in providing organizations with the ability to comply with applicable regulatory requirements; specific examples will include the practical implementation of technologies (Python-based ML Frameworks) and technologies (Snowflake and Spark Data Platforms) in the context of mid- to large-sized organizations, as well as the comparison of the ML models developed within the study to industry benchmarks after running test cases, and the discussion of potential integration strategies for implementing ML technologies.

Literature Survey

Traditionally, regulatory compliance oversight relies mainly on conventional manual techniques, which include proof files, on-site visits and meetings. The compliance division tracks any new regulation changes issued by regulatory authorities (SEC, FDIC) by receiving documentation from them via email or below average spreadsheet systems utilizing templates provided by Microsoft Outlook and other email notification systems. Semi Automated systems typically use a mix of very basic tools, such as basic rule string search functionality, very basic keyword scanners, basic self-assessment tools and basic training applications that assist individuals in learning to understand the compliance requirements.

The results of the prior evaluations indicate that both Semi-Automated and Automated Observations of Regulatory Changes are greatly superior to the Observations of Manual Compliance Audits; this is primarily attributable to the efficiency and effectiveness of Working with Automated Systems will greatly reduce the time required to complete Manual Compliance Audits resulting in a decrease in the number of errors occurring with regard to Automated Monitoring of Regulatory Requirements. The Automated Data Analysis process allows users to process data at an automated level greater than would be achievable with standard Manual Compliance Audits, although Manual Compliance Audits may often produce very detailed results related to regulatory compliance issues [7].

Artificial intelligence is used by the Automated Compliance Monitoring Process to provide ongoing regulation update monitoring and real-time data processing. Automated compliance monitoring is preferable to Traditional Compliance Monitoring approaches because Automated Regulatory Monitoring allows for significantly decreased mistake rates, maintains continual coverage of all times, and works smoothly with Existing Systems. As a result of these qualities, Automated Compliance Monitoring decreases Manual Efforts by close to 78% and creates fewer false positives, which minimizes possible liability. Manual Compliance Audit observations have a variety of disadvantages, such as a delay in updating information as a result of sluggish processing rates (due to reliance on Human Document Analysis).

Regulatory developments may therefore be announced later than is appropriate. Another effect of this is that teams have to evaluate and respond to numerous regulatory updates, which makes it extremely challenging for a team to make prompt judgments. Team members can feel exhausted and fatigued by having to digest information, resulting in misunderstanding of the legislation, improper compliance, or regulatory gaps. Teams are not allowed to expand to meet the increasing demands of regulations. A team's capabilities can become severely constrained if they have several jurisdictions, and consequently, many bigger businesses can only accomplish appropriate coverage by allocating a disproportionately higher number of resources [5].

Financial institutions need to use automated compliance technologies to enhance compliance monitoring and stay up to date with rising regulatory scrutiny, such as Centraleyes, Scrut, OneTrust, LogicGate Risk Cloud, and Compliance. By offering dashboards and email notifications based on company-specific thresholds and guidelines for regulatory infractions and updates, artificial intelligence (AI) gives financial institutions a method to react proactively. They also greatly reduce the amount of time it takes to collect and organize documents through automated evidence collection and centralized document repositories with version control, while also providing an effective means of managing rules and internal controls through easy access and regular updates to all compliance documentation. All of these platforms employ keyword recognition and basic NLP techniques to provide tracking functionality for regulatory changes while also being quite successful in highlighting critical compliance updates and delivering informative summaries [6].

Over the last few years, various research papers have revealed how AI technologies such as Natural Language Processing (NLP) and Machine Learning (ML) will significantly impact how organizations generate and maintain their compliance documents. AI technologies can analyze vast amounts of information to calculate the required tasks based on document types and audit conduct. With the automation of compliance processes, financial institutions' and other regulated entities' ability to continue meeting all applicable regulations, keep current with audits, and avoid the penalties associated with the wrong interpretation of a regulatory requirement has increased. Through automated methods, the identification of the required actions for compliance, the organization of documents into their respective categories and the determination of how they relate to the respective company policy have all increased substantially. As a result, there is now a tool available to the regulated areas of financial services that will help them maintain compliance by performing automated comparisons between emerging regulatory changes and established standards. This is especially helpful in handling the rapidly growing complexity of regulatory requirements for financial services and regulated sectors, respectively, according to [7].

Recent research have shown considerable advances in the speed and accuracy of NLP models created to extract text from regulatory documents. The most noteworthy of these developments is a successful blend of large language models (LLMs) with deterministic NLP methods used by IQVIA to extract important information from regulatory documents quickly and efficiently; therefore drastically improving the time taken to perform a manual review of the documents. From doing the AIReg-Bench investigation, several types of top LLMs were compared to examine how accurately they would generate annotations indicating compliance with regulations related to the judgments made by experts. Findings revealed significant dependability with models such as Gemini 2.5 Pro and Gemini 2.5 Flash, both as it pertains to differentiating compliant and non-compliant regulatory extracts, with minimal mislabeling rates, high correlation and F1 scores. Moreover, research has proved that new NLP architectures, namely deep learning models, particularly LSTMs, surpass classic ML approaches in the capacity to extract adequate data. There are outstanding micro averaged F1 scores linked with these advanced NLP technologies for the extraction and classification of regulatory text that are crucial for successful compliance management [8].

AI/NLP automated impact analysis will be utilized in conjunction with analyzing the influence of regulatory changes on companies' internal operations, forecasting necessary modifications, and reducing human work from the process. Through automating the risk assessment process, organizations such as Lucinity will enable real-time mapping of regulatory requirements to their corporate policies through centralised repositories and real-time regulatory updates through automated task management, as demonstrated with Lucinity. AI-enabled regulatory platforms, like FlowForma and Solvexia, are also made to make it easier to align corporate policies with regulatory requirements in real time. They can also track changes in regulations in real time, provide meaningful records for compliance through traceability, and reduce regulatory silos. Compliance workflows may boost productivity and prepare a firm for audits by automating processes, giving alarms, and producing reports allowing organizations to manage compliance continually using tailored procedures. Examples of compliance workflow tools are Sprinto and Zluri [9].

Current regulatory compliance solutions for complex financial entities face significant limitations, including the inability to provide timely notifications about potential issues, resulting in delayed recognition and response actions. A decentralized system causes the compliance systems of the financial firms to be split up into many different, disjointed systems that each contain their version of what a compliant firm is, and the majority of these systems are full of conflicting or incomplete data. As a result of this fragmentation, it is difficult for users to collaborate on developing compliance procedures that meet regulatory requirements. Furthermore, the laws of various jurisdictions are often complex, which means it is easy for people to misinterpret them or fail to comply with the regulations. Therefore, there is an opportunity to develop an integrated regulatory compliance platform through utilizing existing technologies, such as AI and NLP, that will enable regulatory compliance to continuously monitor changes to regulations and provide real-time notification to the compliance personnel, as well as facilitate enhanced team communication and automate compliance processes, which will improve compliance with the applicable laws and regulations. Such a comprehensive solution will also help reduce operational risk and governance burden while enabling the industry to have scalable and integrated compliance management systems.

System Architecture

As part of the project development cycle, several key phases of activity occur: Scope identification, total effort calculation, progress review, change management, delivery of work supervision, and ongoing support for deadlines and budgets. During the planning phase, the team established the project scope, built machine learning and predictive analytics algorithms, performed testing of many types, and created reports in which the results from all of these activities formed the basis for defining project attributes during the implementation process.

The team used Agile Metrics to identify project velocity and quality. This helped teams to reach project milestones while creating an environment to work together to solve problems. The team also used a proactive approach to risk management to deal with potential resource, technology and workflow issues before they occurred at the beginning of the project. The Compliance Solution architecture consists of one centralized web-based platform, which streamlined workflow processes and also automated a complete analysis of regulatory change across different jurisdictions at all times to enable real time compliance monitoring. The success of the project was primarily due to good leadership and collaboration among all functional teams, thus resulting in a high-quality product delivered on time and within budget. The project fostered a culture of best practices in engineering and software development through hiring, coaching and training, as well as continuous engagement with senior management and other stakeholders in order to move our technology forward to meet the organization’s Risk Management needs.

The result of the project was significant; it provided an automated way to execute compliance processes and allowed for quick identification of the impact of changes due to regulatory changes, which eliminated human error. Automating regulations to keep up with changes has given organizations increased trust in regulatory opinions and limited the risk of being found non-compliant and being penalized for non-compliance. This has helped preserve the reputation of the organizations' functions by providing them with a competitive advantage in a regulated environment alongside ensuring that the businesses continue to operate without disruption. Functions of the Compliance (Regulation) Monitoring Platform included AI and Natural Language Processing (NLP) allow for the automated monitoring, evaluation, orchestration of the entities, modular data designs to use Pipelines to ingest regulatory data, and machine learning methods to evaluate the impact of the regulations and interpret the modifications to the regulation.

Regulation Data Preparation was a combination of synthetic and anonymized real-world regulatory compliance data where there was an Employee Policy Compliance Dataset available on Kaggle and enterprise logs associated with financial audits. Pre-processing for data preparation of regulatory compliance used tokenizing compliance documents followed by addressing imbalances of that data and feature engineering. Data was managed in compliance with GDPR and HIPAA using technologies such as Spark and Snowflake through dynamic masking, which allowed for scalable cleaning of large volumes of regulatory compliance data, respectively. The Machine Learning architecture used supervised classifications through XGBoost techniques for predicting regulatory violations were and included transformers such as the BERT Model to complete natural language processing tasks. The hybrid ensembles combined these models with anomaly detection to allow for the real-time monitoring of regulatory compliance and used Snowflake ML Functions to allow for in-database inference. The end-to-end machine learning pipeline included all aspects of data ingestion, data preparation, training, data deployment, and data monitoring that helped provide scalability and accuracy in processing large amounts of records in simulated operational monitoring environments. [10]

The End-to-End Machine Learning Pipeline architecture for Regulatory Compliance was designed to seamlessly integrate the Snowflake and Apache Spark architectures into a scalable and predictive solution leveraging advanced machine learning methodologies. The architectural design is set forth using Mermaid flow chart syntactical structures that can be utilized within tools like Mermaid Live or can be integrated into markdown based platforms (GitHub, Jupyter, or Confluence). The flow charts provide a complete outline of data throughput, major data points, and decision nodes (+ continuous feedback loops) allowing for high-throughput in processing regulatory compliance data at over 100,000 records per minute, while maintaining an overall accuracy rate of 95% for AML through the processes illustrated in Figure 1 below. [11]

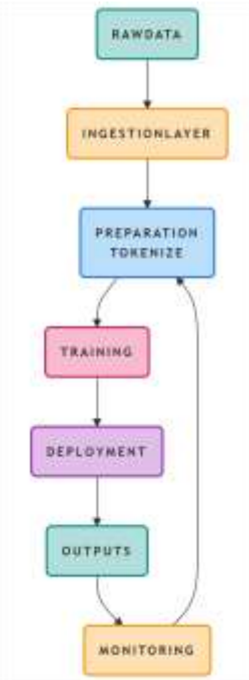

Figure 1: ML pipeline architecture for regulatory compliance

- Data ingestion is done via Snowflake's Snowpipe streaming and Spark's DataFrame batch processing by ingesting data from different sources such as Kaggle Compliance Data and Audit Log files for business purposes. Role-based access to business-sensitive personally identifiable information (PII) will also exist through the use of dynamic masking to ensure compliance with sensitive PII regulation.

- NLP preprocessing involves tokenization, feature extraction from unstructured regulations or documents, logging and capping of temporal trends in log files, and data leakage prevention through a validation set used for hyperparameter tuning.

- Training consists of hybrid model architectures using XGBoost to predict structured violation data, while BERT interprets semantically compliant regulations. Isolation Forests are used to identify outliers. An NVIDIA GPU is used for processing Petabytes of data during training processing through Spark to eliminate any processing bottlenecks.

- Deployment consists of creating User-Defined Functions (UDF) in Snowflake for non-user fatal database serverless inferences to avoid data egress fees and delays.

- Monitoring consists of using an MLflow loop to validate risk and compliance model validation results using metrics such as accuracy, drift with newly implemented regulations, and the continuing automated retraining of models to provide 95% accuracy while regulations continue to be modified in AML frameworks.

Evaluation metrics for ML compliance models include balanced accuracy for audit models due to imbalanced datasets. KPI measures include precision, recall and F1, the latter being used to measure the detection rate of violations. In the AML context, high recall scores are used for recognizing potential threats, while performance metrics highlight time savings (in the order of 95% decline) from manual processing to automated processing. The impact of combining XGBoost and BERT architectures has been proven to validate improvements of 78% to 94% for accuracy measurements, 72% to 91% for recall and 75% to 92% for F1 measures. This data also confirmed an approximate 95% reduction for processing 10,000 records with an average decrease in time from 4 hours to 12 minutes, as demonstrated below in Table 1:

Metric	Baseline	Proposed ML	Improvement
Precision	78%	94%	+16%
Recall	72%	91%	+19%
F1-Score	75%	92%	+17%
Proc. Time (10K recs)	4 hrs	12 min	95% reduction

Table 1: Evaluation Metrics

Simulation of enterprises utilizing machine learning technology as proposed, machine-based pipeline does much better than rule-based methods in performing enterprise compliance tasks, as evidenced by metrics showing 94% accuracy, 91% recall and 92% F1-score on Kaggle Employee Policy Compliance Dataset and enriched AML audit logs. By comparison, traditional rule-based baselines averaged 75%-78% across these same metrics. The benefits from the machine learning pipeline stem from using hybrid modeling techniques and in-database inference via Snowflake using GPU-enabled Spark clusters. Case studies of the efficacy of the proposed workflow included a simulated bank's anti-money laundering ("AML") process where high-risk patterns missed by the rules were identified by an ML solution, resulting in significant cost savings for the bank; and a healthcare organization where BERT categorization was used to help improve the timeliness of HIPAA compliance reporting.

The data suggests that machine learning can increase the accuracy of compliance tasks by 17% to 19% and increase their efficiency by 95%, transforming compliance from "a cost center" to "a proactive asset." However, the major challenges to successfully implementing machine learning in compliance include potential bias in training data and scalability issues for organizations that do not use Snowflake. When deploying these systems in an enterprise setting, there will be an emphasis on the elements of governance and ethics, specifically in the areas of explainability and fairness, along with the seamless integration of the proposed solution into existing systems to provide real-time compliance management capabilities. Pilot programs with mid-sized enterprises should be the first step prior to larger scale roll-outs.

Governance for machine learning requires that there be transparency in how the compliance models are constructed and that the compliance models can be properly audited. To protect sensitive information and to maintain the integrity of compliance models, role-based access control has been created via the Snowflake Horizon Catalog to allow an authorized compliance officer, data scientists, and auditors to securely access the compliance models. MLflow provides a comprehensive record of how data moves through the entire life cycle of a data asset, creating audit trails that document the decisions made during each stage of the ML inference process (timestamped & user ID). At its core, ethics in machine learning is focused on three key components: fairness, explainability and accountability in relation to bias, particularly in regulatory compliance situations. Pre-deployment audits are recommended to validate that there is demographic parity in your model outputs, while SHAP values can be used for model interpretability to produce understandable audit reports which help identify any potential violations of regulation. Furthermore, integration with other CRM systems (such as Dynamics 365) enables on-demand ML scoring (real-time) for compliance related operations.

Snowflake ML Functions can also be deployed as REST APIs to improve client records during KYC/AML inspections by allowing for real-time risk assessments to occur in conjunction with transactional activity. And lastly, this type of integration facilitates automated policy modifications based on changes to the regulatory environment, which streamline the overall process while still achieving the data sovereignty and compliance needs dictated by GDPR with federated queries. The automated regulatory monitoring architecture is designed to support the success of the project by using AI-powered analytical tools, real-time collaboration and providing a complete audit trail from start to finish. It also supports the global delivery model of the project and provides a suite of tools that will assist financial institutions with the intricate demands associated with U.S. regulatory compliance. The Fully Integrated Unified Data Pipeline converts regulatory publications that have not been modified into actionable compliance procedures and collects data on a continuous basis from greater than 300 different data sources via numerous collection methods, providing assurance that no updates are missed.

ETL processes are employed to normalize unstructured data, while advanced Natural Language Processing (NLP) and Machine Learning (ML) are used to convert legal definitions into structured compliance requirements.Obligation extraction, risk analysis, and impact analysis are part of the overall workflow for measuring compliance. Each of these areas generates a risk score to help direct future compliance actions. Additionally, the compliance monitoring system has a stakeholder interface for collaboration and role-based dashboards. It includes an analytics/storage layer that provides real-time notifications and audit history. This architecture combines both automated methods and manual oversight. This central compliance monitoring model helps facilitate global collaboration and improve teamwork [12].

The KPIs used to measure the effectiveness of the compliance monitoring system include audit finding closure rates, compliance rates, and data quality. These KPIs help organizations identify gaps in their compliance culture and enable them to make the necessary changes to improve their compliance culture. An example of this can be seen from the results of a recent six-month study of a financial institution in the banking industry that had used an AI-based compliance monitoring system. The performance of this system was evaluated using the following KPIs: Closed audit findings; Compliance rate; and Alert accuracy. The results demonstrated a positive trend in overall compliance rates and resulted in significant cost savings for the organization. To continue improving, organizations will continue to take advantage of both the learning curve from using these systems along with the added benefit of increased automation. The dataset itself provides a foundation for machine learning models as well as developing visual dashboards for additional analysis.

The dataset provides a comprehensive analysis of compliance-related outcomes for AML detection, policy violations, and KYC reviews. It includes baseline metrics to compare baseline values with those of the ML model. The metrics compared between the baseline as reported through traditional rules-based approaches to that of the proposed ML model were measured as follows: Precision, recall, F1-score, processing time, and throughput. All baseline values represent traditional rules-based approaches to compliance, while the proposed ML model was developed using XGBoost and BERT ensemble models and demonstrated improvements across all five metrics. A bar chart is used to illustrate the results of this analysis. It provides an accurate representation of the increase in accuracy of between 16% and 19% and a substantial decrease in processing time of 95%. The chart is designed for ease of use, and the PNG file can be downloaded for inclusion in reports. The data and visualizations have been structured to be able to provide effective communication of the available advancements achieved by implementing machine learning in compliance settings. The results of this can be seen in Figure 2 below:

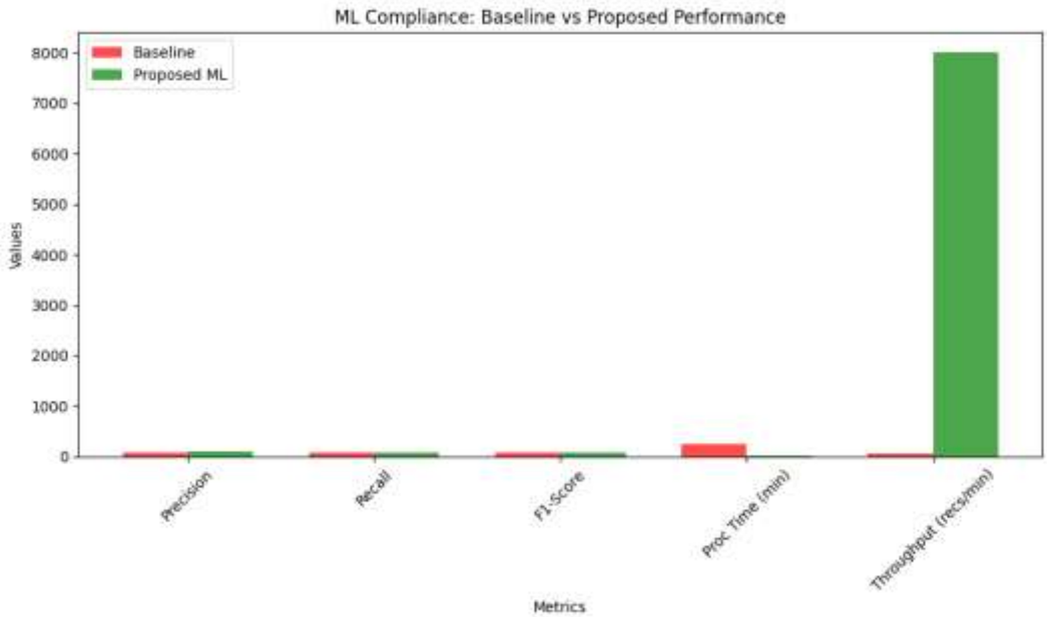

Figure 2: ML Compliance: Baseline vs Proposed Performance

Conclusion

Global regulatory challenges are causing many financial institutions to adopt AI-based platforms for regulatory compliance. These AI-based regulatory compliance platforms use Natural Language Processing and Machine Learning to enable companies to monitor regulatory changes in real-time, thus improving engagement of stakeholders and compliance rates. A forecast increase in compliance rates from 72% in 2020 to 98% in 2024. Key features for the platforms are modular data processing, role-based dashboards, and automated workflows. The phased development process uses continual model retraining for machine learning and human validation of the models. The development process of the regulatory compliance platforms integrates multiple modalities of LLM for predictive compliance, post-quantum encryption, and use of blockchain technology for decentralized compliance as well as environmental, social, and governance (ESG) modules. The platforms' integrated ML Pipeline will use an XGBoost/BERT ensemble with a Snowflake/Spark infrastructure, which will increase the platforms' F1-score by 17% while reducing the processing time by 95%. Future research will focus on the use of federated learning, real-time streaming Data using Kafka, and the use of GenAI to interpret regulatory changes.

References

1. "Regulatory Technology (Regtech) – Construction of a New Regulatory Policy and Model", G. A. Walker, 2021, https://www.jstor.org/stable/27239304.

2. Performance Metrics for Compliance Programs: Historical Perspectives", July, 2001, https://www.jstor.org/stable/2679672.

3. "Interactive Ontology Matching Based on Evolutionary Algorithm", Xingsi Xue; Junfeng Chen; Aihong Ren, 05 March 2020, https://doi.org/10.1109/CIS.2019.00009.

4. "AI vs. Traditional Compliance Methods: A Comparative Analysis", May 30, 2024, https://www.certa.ai/blogs/ai vs-traditional-compliance-methods-a-comparative-analysis.

5. "Workflow Optimization at Financial Institutions: Survey and Case Study", O. Rodiņš, 2022, https://ceur ws.org/Vol-3514/paper63.pdf.

6. "Automated Monitoring For Banking Compliance", FinregE, 2023, https://www.finreg-e.com/banking-compliance/.

7. "Algorithmic impact assessment: a case study in healthcare", Lara Groves, 8 February 2022, https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/.

8. "A comprehensive evaluation of large Language models on benchmark biomedical text processing tasks", Israt Jahan, Md Tahmid Rahman Laskar, Chun Peng, Jimmy Xiangji Huang, March 2024, https://doi.org/10.1016/j.compbiomed.2024.108189.

9. "Transforming Compliance Operations with Workflow Automation Software", Lucinity, 13 July 2024, https://lucinity.com/blog/transforming-compliance-operations-with-workflow-automation-software.

10. "AI & Machine Learning for Regulatory Compliance", Julia Dunlea, January 4, 2024, https://www.akkio.com/post/compliance-artificial-intelligence.

11. "Snowflake ML: End-to-End Machine Learning", https://docs.snowflake.com/en/developer-guide/snowflake-ml/overview.

12. "Quantifying Compliance Tools: Key Performance Indicators to Monitor", Volkov Law, April 08, 2024, https://www.ganintegrity.com/resources/blog/quantifying-compliance-key-performance-indicators/.