Universität Zürich[UZH]

University of Zurich

Department of Finance

Innovation in Risk Management: Three AI-Enabled Solutions for Swiss Banking

May 8, 2025

Editors: Prof. Dr. Walter Farkas[1,2], Dr. Daniel Fasnacht[3]
Scientific Coordinator: Patrick Lucescu[1]
Contributors: Ekaterina Frolenkova[4], Tanja Srindran[2], Florian Pauschitz[2], Dorina Ababii[5]

[1]Department of Finance, University of Zurich
[2]Department of Mathematics, ETH Zürich
[3]Swiss FinTech Innovation Lab, University of Zurich
[4]Geneva School of Economics and Management, University of Geneva
[5]Portfolio Manager, HCP Asset Management

Abstract

Driven by advances in technology, banks are encountering increasingly complex and dynamic risk landscapes. In response, a group of Swiss banks defined key challenges in three areas, i.e. operational risk classification, internal fraud detection, and regulatory compliance automation. This report presents innovative solutions that directly support the financial sector's need for enhanced efficiency, precision, and adaptability. It highlights the power of academic-industry collaboration in delivering practical, forward-looking innovations for the banking sector.

Chapter 1 introduces an AI-powered framework for operational risk classification and forecasting. By integrating data cleaning, Large Language Models (LLMs), and Transformer-based time-series analysis, the approach significantly improves classification accuracy and enables timely, comprehensive risk reporting. For banks, this represents a scalable tool to reduce losses and strengthen operational oversight.

Chapter 2 tackles internal fraud through the application of an Isolation Forest model to identify anomalous transaction behavior – particularly among relationship managers. Requiring no labeled data and offering a customizable alert threshold, the solution enables compliance teams to prioritize investigations and better allocate resources, enhancing fraud prevention capabilities.

Chapter 3 presents a multilingual compliance tool based on DistilBERT, designed to automate regulatory document comparison and change detection. Tested on English and German texts, the model demonstrates strong potential to reduce compliance effort and translation costs – addressing a growing challenge for banks operating across jurisdictions.

Collectively, these solutions illustrate how targeted AI applications can modernize risk management, strengthen fraud detection, and streamline compliance processes – while preserving essential human oversight. By improving the agility, reliability, and cost-efficiency of risk practices, Swiss financial institutions can further solidify their position as global leaders in innovation and regulatory excellence.

Contents

Introduction

RiskON represents a groundbreaking initiative that merges academic excellence with industry practice. Conceived under the leadership of the University of Zurich's (UZH) Department of Finance—with Professors Walter Farkas and Dr. Daniel Fasnacht spearheading the project—and the UZH Innovation Hub in collaboration with the N9 House of Innovation, RiskON has emerged as a platform where theory meets practice in the context of risk management. With strategic support from distinguished financial institutions including Julius Bär, LGT Switzerland, and Pictet Group, RiskON serves as an incubator for innovative solutions to risk management challenges in the financial sector.

The aim of this report is to present the challenges of RiskON 2024 and discuss the innovative solutions developed by the students.

RiskON is structured as a hackathon-style event, typically spread over two full days. The schedule includes a range of activities: from opening sessions, pitch presentations, structured hackathons addressing pre-defined challenges to social activities such as networking events and the award ceremony.

The Innosuisse AI Innovation Booster accelerates the transfer of AI-driven practices from theory to real-world application by fostering open innovation. It provides a structured environment where multidisciplinary teams collaboratively develop, refine, and validate AI-based innovation ideas. The event's format ensures an intensive period of ideation and collaboration, with tasks designed to push participants to generate tangible concepts that can evolve into minimum viable products (MVPs) or proofs of concept (PoCs) within six months following the event. The systematic organization—the inclusion of registration, team formation, idea development phases, and final pitch presentations—underscores RiskON's commitment to both academic thoroughness and practical feasibility.

The core of the RiskON initiative is its challenge-based competition model. In the 2024 academic cycle, three prominent financial institutions—Pictet Group, Liechtenstein Global Trust (LGT), and Julius Bär—collaboratively developed strategic research challenges designed to explore critical intersections between emerging technologies and risk management practices. These challenges were carefully constructed to enable student teams to select and engage with complex problem domains, specifically focusing on: *Use of New Technologies for Risk Events Classification and Anal-*

ysis (Pictet Group), *Fraud Risk Detection – How Artificial Intelligence Can Help?* (LGT), *Enhancing Efficiency for CROs Using Artificial Intelligence* (Julius Bär). By providing students with industry-defined research challenges, the initiative facilitates a structured approach to applied academic research, encouraging innovative problem-solving and interdisciplinary collaboration.

An integral component of the RiskON hackathon is the role of the expert jury. The jury is composed of leading figures from the Swiss finance industry, including former regulators and partners from prominent financial institutions. Prominent members include experts from PwC Switzerland, the former Vice Chairman of FINMA, and representatives from the Swiss Risk Association, granting the evaluation process significant credibility and professionalism. The jury not only evaluates the feasibility and innovativeness of the proposed solutions but also provides immediate, actionable feedback, thus serving both as evaluators and mentors. The rigorous assessment process, based on strict criteria related to strategic relevance, technical feasibility, and potential for real-world application, ensures that the winning proposals reflect both academic excellence and market readiness.

The primary goal of this paper is to provide an academic perspective on the findings derived from the three winning teams of RiskON 2024. Each team addressed a unique challenge using a combination of innovative methodologies and forward-thinking strategies.

The remaining of this paper is structured in three chapters, one for each of the three challenges. Each chapter presents the specific industry-defined challenge, critically analyzes the winning team's proposed solution, and provides a comprehensive academic discussion of the solution's potential implications, limitations, and broader applicability in the contemporary risk management landscape.

Chapter 1

Use of new technologies for risk events classification and analysis

Financial institutions are regularly exposed to operational risks, ranging from human errors to systemic technical failures, which can have severe financial and reputational consequences (Alonso et al., 2024). A notable example is a 2012 trading scandal, where a misclassified portfolio led to a $6 billion loss (Silver-Greenberg and Craig, 2012). This case exemplifies the shortcomings of conventional risk event classification methods which often exhibit delays in detecting and addressing critical risks, thereby precipitating significant adverse outcomes.

Mandated by regulatory bodies, financial institutions have to report and investigate risk events, which is usually done through a structured three-step process. First, an employee logs the event by summarizing it in a brief narrative and recording associated metadata, including the time of occurrence and any losses incurred. This initial step takes approximately 1 to 5 minutes, according to data provided by the Pictet Group. Next, a human reviewer classifies the incident based on its causes, such as trading, operations, or investment-related issues. This classification process, which evaluates the incident across multiple dimensions (e.g., identifying the impacted business unit and the type of risk incurred, such as financial or regulatory), generally requires about 15 minutes. Finally, members of the risk division conduct a thorough analysis of significant risk events and compile the main insights for management, a task that typically involves several weeks of labor.

This traditional approach exhibits two primary shortcomings: it is both labor-intensive and inefficient. The delay between the occurrence of a risk event and the generation of actionable insights can extend over several weeks—a period during which ongoing risk events may continue to cause unmitigated harm. Moreover, the substantial daily volume of risk events renders their comprehensive classification and analysis prohibitively resource-intensive for risk management teams. As a result, at the Pictet Group, only 5–10% of risk events are fully classified and communicated

to management, despite this subset accounting for approximately 85–90% of total losses.

The 2024 RiskON Challenge 1, proposed by the Pictet Group, calls for innovative technological solutions to streamline the classification and analysis of risk events. By harnessing emerging technologies, the goal is to overcome the current delays and inefficiencies inherent in traditional processes, thereby enabling financial institutions to swiftly generate actionable insights. Such innovations hold the promise of significantly enhancing risk management practices, ensuring that critical risks are promptly identified and mitigated.

1.1 Literature

The integration of Artificial Intelligence (AI) in the banking sector has emerged as a significant area of academic inquiry over the past decade (Leo, Sharma, and Maddulety, 2019). Machine learning techniques offer novel approaches to processing complex datasets, potentially augmenting traditional risk assessment methodologies. While AI technologies may provide additional analytical insights, their effectiveness remains contingent upon careful implementation and ongoing human oversight. Financial institutions are thus exploring strategic approaches to incorporate AI-driven methods within existing risk management frameworks.

The classification of operational risk events was discussed by Pakhchanyan, Fieberg, and Metko (2022). They evaluated the efficacy of machine learning algorithms in classifying textual descriptions of operational loss events into Basel II event types. Their study employed support vector machines and multinomial naïve Bayes algorithms, which generate satisfactory accuracy with minimal costs. Valli (2024) explore diverse applications of predictive analytics in risk mitigation. The authors demonstrate how machine learning models can be combined with common statistical methods to predict risks across diverse sectors such as finance, healthcare, manufacturing, and energy. A common use of machine learning in finance is the analysis of historical data and the prediction of future operational and market risks in order to inform effective mitigation strategies. The authors, however, identify limiting factors that hinder the adoption of these technologies. The main factors are data privacy regulations and the high complexity of model interpretation, which might hinder the correct auditing of financial institutions using these tools.

AI is not only applied within the risk divisions of banks. For instance, Jain (2023) review the use of machine learning tools across the entire banking sector. They found that AI is used in the form of robo-advisors, who directly interact with customers by providing them with financial recommendations. The use of chatbots to provide 24/7 support services is also common. Beyond that, AI techniques have been identified to detect fraud in an automated manner. Moreover, AI can also give

banks an edge in creditworthiness analysis. Crucially, AI techniques can minimize operational costs and partially replace humans in error-prone work.

Outside of the financial sector, the use of Large Language Models (LLMs) in classification tasks is discussed by Wang et al. (2024). The authors propose a system leveraging pre-trained LLMs for efficient and adaptive text classification by fine-tuning the models for specific text categorization tasks. Fine-tuning LLMs involves adapting a pre-trained model by further training it on task-specific or domain-specific data, thereby aligning its outputs more closely with the desired application and improving performance. Wang et al. conclude that fine-tuned LLMs can capture more complex semantic structures than simpler models. Moreover, this approach offers banks the advantages of developing tailored solutions across various operational areas at a lower cost than training models from scratch.

In the context of time-series forecasting, numerous machine learning approaches have been presented to enhance predictive capabilities. More recently, Li et al. (2019) propose Transformer models[1] as a viable alternative. To circumvent issues such as locality awareness and memory bottleneck, Li et al. use a LogSparse attention rather than the traditional uniform attention mechanism in the construction of the Transformer model. Using existing datasets, the authors conduct a number of experiments demonstrating LogSparse Transformer outperforms traditional machine learning techniques in terms of accuracy, training efficiency, and scalability.

1.2 Proposed Approach

The analysis of risk events represents a necessary but time consuming task that leaves banks potentially exposed to major financial losses. It is thus imperative to further improve the capabilities of risk management teams.

The winning team identify three major areas of potential improvement. First, the response time to risk events should be reduced. Second, the risk of human-errors stemming from the labor-intensive process should also be minimized. Third, the goal should be to ignore no risk events. The students propose an AI framework able to classify risk events and predict future risks. By combining AI models with human supervision, their framework could lead to a reduced response time to risk events, an increase in classification accuracy, and a smaller proportion of risk events being ignored.

[1]For an introduction to Transformer models we refer the reader to Turner (2023).

1.3 Methodology

To ensure the accurate classification and prediction of risk events, the students' framework is composed of three integral components. The first step focuses on cleaning historical data to ensure that it is accurate, consistent, and complete. This process involves filtering out noise, correcting identified errors, and flagging discrepancies for subsequent human review to ensure a reliable dataset for analysis. In the second phase, risk events are systematically classified according to predefined categories. Advanced machine learning algorithms are utilized to assign events to these categories, enabling a structured and efficient approach to risk segmentation based on the underlying causes and associated factors. Lastly, the third module involves predicting and reporting potential risks. By using state of the art Transformer models, the students aim to use the inherent time-series structure of classified risk events to find patterns in the data and report the findings in an interpretable format.

1.3.1 Data Cleaning

Financial institutions like Pictet Group have access to decades of risk-events data which is crucial for training AI models. However, as events are classified by humans, the dataset might contain a significant number of errors. Additionally, bias could be present within the dataset, with events reported at the end of the day more susceptible to human errors.

Training AI models on erroneous or biased datasets can lead to suboptimal models that compromise predictive accuracy and decision-making efficiency. To ensure the reliability of their model, the students use an AI-free algorithm designed to flag potentially false entries within the dataset. After an observation is deemed questionable, the algorithm generates the proposed changes which are then accepted or refused by a human supervisor.

The algorithm uses a bag-of-words approach to flag potential errors. Using a predefined dictionary that maps keywords to risk event classes (e.g. cause of event, type of risk, sub-type of risk), the algorithm extracts relevant words from the event description and metadata. If more than one keyword is identified, then the most probable risk-event class is selected for each of the classification dimensions. If there are discrepancies between the algorithm's findings and the historical data, the entry is flagged, and a human supervisor takes the final decision.

1.3.2 Risk event classification

With a clean dataset of risk events available, the students employ a LLM tailored to classify events based on their brief description and accompanying metadata. Initially, the model tokenizes the textual description and metadata, generating con-

textual embeddings that capture both semantic nuances and structured information necessary for precise risk categorization. The embeddings are then passed through a deep neural network, which learns the relationships between the embeddings and the risk-event classes. Implicitly, it learns the relationship between the words in the descriptions, as well as the metadata, and the risk-event classes. Similarly to the data cleaning step, the risk-event classes are predefined. On the other side, both the number of tokens and the token-embeddings are directly learned by the neural network.

1.3.3 Prediction and reporting

The third and final component of the students' framework shifts emphasis from the earlier stages of cleaning and classifying individual events. Instead, it concentrates on identifying patterns within the data, forecasting potential future risk events, and presenting these insights in a format that humans can readily understand and interpret. To achieve this, the students employ the LogSparse Transformer model developed by Li et al. (2019), and instead of looking at risk events on an individual basis they construct time series of classified risk events.

While Transformers were initially developed to cope with natural language processing tasks, adapting them to time-series data is feasible and merely requires modifying the input layer. This allows Transformers to learn time-series data with long-term dependencies, which can lead to the discovery of patterns that span over many years. However, the patterns in financial risk events may evolve significantly with time due to changing financial regulations or customer habits. Therefore, it is beneficial to modify the Transformer model to prioritize short-term dependencies, while not completely ignoring long-term dependencies. This is achieved by using a LogSparse self-attention mechanism. Additionally, the historical data might contain interesting patterns that have only existed during a certain period. To learn such dependencies, the students add a restart attention mechanism to the Transformer model.

The Transformer model returns relevant patterns and predictions, which can be adapted to the specific needs of each institution by directly modifying the output layer of the model. This approach is particularly advantageous as the overall architecture remains unaltered, enabling institutions to customize solely the final layer to accommodate domain-specific requirements. While training the model will require a considerable amount of time, which is dependent on the size of the historical dataset, the subsequent model evaluation process can be executed with remarkable computational efficiency. Compared to the traditional risk analysis framework, where the analysis of risk events can take several weeks, the Transformer model is be able to generate a condensed summary of the most important insights in a fraction of the

time.

1.3.4 Results

To evaluate the performance on the classification task, the students compared the accuracy of the AI-free and the AI-based classification algorithms on a small test case which contains real risk events provided by Pictet Group. For the training of the AI model, an artificial training set was generated. The students observed that the AI-integrated algorithm demonstrated superior performance compared to the AI-free algorithm. In the preliminary test set, the respective accuracy rates were 90% and 70%. It is important to note that these figures may not precisely represent the true predictive capabilities of the algorithms. Nevertheless, under the condition of sufficient high-quality historical data for model training, the students hypothesize that the AI-based algorithm will consistently outperform its non-AI counterpart.

1.4 Proposed Implementation Strategy

The students identify two challenges that are key to the successful integration of the AI framework into financial institutions' daily operations.

Firstly, the maximal accuracy of AI models is closely linked to the quality and quantity of the training data. Hence, resources must be allocated in order to adequately clean historical datasets biases. For example, this might require the institutions to extend the historical training data with artificial event logs, especially for low-occurrence risk categories. It is thus crucial to conduct an initial assessment on the quantity and quality of the data. For institutions where data quantity is less than desirable, a clear data acquisition mandate is also advisable.

Secondly, the students note that the accuracy of AI models might never reach 100%, with more frequent errors in certain risk categories (e.g. where training data is not sufficient). The crucial step that companies will need to tackle before using this tool is to educate their employees to these shortcomings and train them to minimize overconfidence in the model's accuracy.

1.5 Discussion

The integration of AI into risk management processes will represent a transformative shift for financial institutions. By automating the classification and analysis of risk events, AI-driven models will enhance the efficiency of risk management teams, leading banks to obtain a better understanding of their risk levels.

The solution proposed by the students takes a holistic approach and addresses the entire risk management pipeline. The three sequential parts of their framework

are the clean-up of historical data, the LLM-supported risk-event classification, and a risk prediction method using a LogSparse Transformer. Students advocate that such a solution can substantially reduce costs and increase predictive capabilities of risk divisions in financial institutions. However, they also warn that their solution is not meant to replace the human factor entirely but rather to act as a complementary tool, with human supervisors having the final decision. The methodology demonstrates significant potential for practical implementation, offering financial institutions a sophisticated, data-driven approach to risk management that balances technological innovation with human expertise. However, the proposed solution is not without drawbacks, which will be discussed in the following paragraphs.

To ensure accurate historical data, the students rely on a bag-of-words approach to map risk events to risk classes. While the accuracy of such a method might not be as high as an AI solution, its simplicity ensures high interpretability for human analysts, allowing them to transparently understand the data cleaning process and rationale behind each risk event mapping. However, the success and accuracy of the algorithm critically depend on the quality and comprehensiveness of the keyword-to-class mapping. As such, financial institutions must allocate substantial time and resources to build and continuously refine this lexical framework, recognizing it as a critical investment in their risk management infrastructure.

The classification of risk events is based on a LLM that uses the risk event description and associated metadata to categorize events into predefined risk event classes. As in the case of the data cleaning process, financial institutions must ensure a comprehensive and well-defined set of risk classes. Additionally, such a model requires a sizable training data to ensure accuracy. The performance of the model is also dependent on any biases present within the training data. For example, the model might not correctly classify black swan events, as they do not appear in the training data. Hence, human supervision is still advised. Despite these caveats, the use of AI for classification is advantageous. It allows to classify risk events at a much faster speed with minimal labor-costs. In addition to this, it can lead to an increased classification accuracy compared to traditional human classification.

Finally, the prediction and reporting of risk events takes advantage of the time-series nature of the dataset and uses a LogSparse Transformer model to generate insights and predictions. These can then be used by risk management teams to lead their investigations and formulate data-based recommendations. Due to its near instantaneous evaluation, the Transformer model can lead to a reduction in labor costs and classification time while also analyzing all risk events, which is in contrast with currently employed frameworks where only 5 to 10% of risk events are fully analyzed. The biggest disadvantage of such a model is its lack of interpretability which might not be desirable by both financial institutions and regulatory bodies. To facilitate the effective integration of AI models in risk management frameworks,

financial institutions should establish rigorous model validation protocols that comprehensively assess the algorithmic capacity to discern nuanced and multifaceted risk event characteristics. Furthermore, the models should be deployed in a phased manner, initially running parallel to existing classification methods to validate its performance and gradually increasing its autonomy in risk event analysis.

References

Alonso, Asunción et al. (2024). "Basel core principles for effective banking supervision: an update after a decade of experience". In: *Financial Stability Review* 46.

Jain, Neetu (2023). "Artificial intelligence technology in banking sector: A systematic literature review". In: *Journal of management & entrepreneurship*.

Leo, Martin, Suneel Sharma, and Koilakuntla Maddulety (2019). "Machine learning in banking risk management: A literature review". In: *Risks* 7.1, p. 29.

Li, Shiyang et al. (2019). "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting". In: *Advances in neural information processing systems* 32.

Pakhchanyan, Suren, Christian Fieberg, and Daniel Metko (2022). "Machine learning for categorization of operational risk events using textual description". In: *Journal of Operational Risk*.

Silver-Greenberg, Jessica and Susanne Craig (2012). "JPMorgan trading loss may reach $9 billion". In: *New York Times*.

Turner, Richard E (2023). "An introduction to transformers". In: *arXiv preprint arXiv:2304.10557*.

Valli, Latha Narayanan (2024). "Predictive Analytics Applications for Risk Mitigation across Industries; A review". In: *BULLET: Jurnal Multidisiplin Ilmu* 3.4, pp. 542–553.

Wang, Zhiqiang et al. (2024). "Adaptable and Reliable Text Classification using Large Language Models". In: *arXiv preprint arXiv:2405.10523*.

Chapter 2

Fraud Risk Detection – How artificial intelligence can help?

According to the Association of Certified Fraud Examiners (ACFE) 2024 report, organizations lose approximately 5% of their annual revenue to fraud. Of all types of fraud, financial statement fraud is the least common but most costly (ACFE, 2024). Employee fraud alone has grown significantly, with a reported increase of 73% over eight years (Banking Frontiers, 2024). In Switzerland, there are 78 cases of white-collar crime, including COVID-19 related fraud which was resolved in 2022 with total reported losses of CHF 581 million (KPMG, 2023).

Fraud's impact extends beyond direct financial losses, encompassing significant reputational and regulatory risks. Effective transaction screening requires adaptive detection systems that can evolve alongside emerging threat landscapes. While emerging technologies offer potential enhancements to risk mitigation strategies, successful implementation demands continuous methodological refinement and robust analytical frameworks.

Against this backdrop, the 2024 RiskON Hackathon Challenge 2, proposed by Liechtenstein Global Trust (LGT), aims to harness the potential of AI to mitigate these risks. The challenge focuses on the design of a pragmatic AI-driven fraud detection tool, both for external and internal fraud. With the rapid growth of big data and AI, the challenge is to exploit these technologies for fraud detection in the private banking sector (RiskOn, 2024). Students were asked to conceptualize a solution, identifying the target users, their needs and current challenges. Deliverables included a one-page solution that briefly outlines the value proposition, key technologies and feasibility; and a presentation showing the solution, its benefits, technical implementation and future extensions in a private banking context.

2.1 Literature

The evolution of fraud detection has undergone significant transformation over recent decades, moving from early statistical models to sophisticated AI-driven approaches. Early foundational work by Bolton and Hand (2002) established the statistical frameworks, which were later extended with the advent of machine learning methods. An in-depth overview of this transition—from rule-based systems to modern, AI-enhanced solutions—is provided by West and Bhattacharya (2016). Contemporary research further demonstrates how advanced machine learning techniques are reshaping fraud detection in the financial sector, as discussed by Boukherouaa et al. (2021) and explored in relation to regulatory implications by International Monetary Fund, Monetary and Capital Markets Department (2024).

Initial studies primarily focused on two main statistical approaches: logistic regression and neural networks. Sohl and Venkatachalam (1995) were among the pioneers in applying neural networks to assess financial statement fraud, and further evidence of their efficacy was presented by Fanning and Cogger (1998), who compared these methods with traditional statistical approaches in detecting management fraud. Complementing these efforts, Rezaee (2002) conducted an analysis specific to financial statement fraud while Bolton and Hand (2002) provided a broader examination of statistical learning techniques in fraud detection.

Decision trees also emerged as a popular tool for fraud detection due to their capacity to model complex decision-making processes and offer interpretable insights. Bai, Yen, and Yang (2008) demonstrated their application to financial statement fraud in the Chinese market, while comparisons between decision trees and alternative methods such as support vector machines were undertaken by Whitrow et al. (2009). In addition, random forests have been applied in various contexts: Xuan et al. (2018) focused on credit card fraud detection and C. Liu et al. (2015) applied them to predicting financial statement fraud. More recently, John and Naaz (2019) and Ounacer et al. (2018) have employed isolation forests for anomaly detection in credit card fraud, highlighting the continued expansion of model diversity in the field.

Despite the historical prevalence of supervised learning in fraud detection, contemporary research is exploring the potential of unsupervised and hybrid algorithmic strategies. These models have been proven to detect fraud more accurately when compared to human experts or existing algorithms (Hilal, Gadsden, and Yawney, 2022). Paula et al. (2016) use AE to detect money laundering in Brazilian organizations involved in foreign trade that have irregular and suspicious patterns in their behavior. Kuna (2023) outlines the utilization of AI-enhanced fraud detection systems employing a hybrid model, which integrates supervised and unsupervised methods (SVM and AE) into a deep learning architecture. S. Visbeek, E. Acar, and

F. d. Hengst (2023) implement a hybrid model combining recurrent neural networks (RNN) with genetic programming (GP) for fraud detection. The RNN model initially generates a set of expressions (or fraud candidates), while GP is employed as an inner optimization loop to facilitate selection, recombination and mutation on these candidates. The model is used on a data set of simulated bank transactions and demonstrates good precision and recall scores. DeLise (2023) uses a semi-supervised model on a small labeled dataset alongside a large unlabeled dataset to compare it to a purely unsupervised algorithm. Their findings suggest a hybrid model handles sparse and imbalanced datasets more effectively than traditional unsupervised approaches, while also reducing false negatives and improving precision. Vashistha and Tiwari, 2024 explore the use of Hyper Ensemble ML (HEML) with synthetic bank account datasets. By combining multiple unsupervised and semi-supervised models, HEML reduces computational costs and human interventions and is able to detect new fraud patterns that were not included in the training data.

While many studies concentrate on external fraud—encompassing areas such as money laundering, credit card fraud, and insurance fraud—internal fraud remains less explored. Internal fraud, involving illicit activities by employees or insiders, is challenging to detect due to the inherent trust placed in these individuals. For example, Holton (2009) combined text mining with Bayesian belief networks to identify potential insider threats, and Yuan et al. (2018) applied Long Short-Term Memory Deep Neural Networks to detect insider risks. Phua et al. (2010) offer a practical perspective on fraud detection that includes discussions of internal fraud and hybrid approaches.

2.2 Proposed Approach

Relationship managers (RMs) are bank representatives directly involved with the bank's clients. Their knowledge of the bank's internal controls places them in a key position where they can discover and take advantage of weaknesses in the bank's control systems and processes. Mismanagement of clients from the RMs side for fraudulent reasons can cause reputational and financial damages to the bank. It is thus crucial to oversee initial transactions screening techniques and detect suspicious transactions.

The winning team proposes an innovative approach targeted at internal fraud detection, specifically at Compliance officers (COs) who want to detect fraud connected to RMs. Their solution incorporates an AI anomaly detection model to flag suspicious transactions. This approach allows the user to independently choose the proportion of flags the model will output; the choice of the positive rate depending on the bank's resources. By doing so, the model stays flexible and does not require extensive maintenance as it is expected to automatically adapt to the evolution of

fraud.

2.3 Methodology

The students conducted an in-depth evaluation of current fraud risk detection methodologies, concentrating on the integration of advanced AI systems within financial institutions. Their findings highlighted an opportunity to implement an AI-based tool that could seamlessly integrate with existing infrastructures to automate the identification and analysis of anomalies across financial transactions. This integration not only promises to enhance operational efficiencies but also aims to improve the overall accuracy of fraud identification, aligning with the strategic goals of mitigating risk.

Furthermore, the evaluation underscored the necessity for an adaptable AI approach capable of evolving with increasingly intricate financial environments. By aligning the AI tool's capabilities with current systems, the research laid a critical foundation for developing a more proactive and intelligent fraud risk detection framework that minimizes risks associated with manual monitoring and data discrepancies.

2.3.1 Data

The students consider a large dataset consisting of 23 characteristics divided in three distinct groups: information about the customers, RMs, as well as specific transaction made by RMs which will be used as a proxy for RMs' behavior. The client's variables are biographic, their risk profile, wealth, knowledge and experience, years of relationship with a bank, associated RM, and frequency of communication with RM. RM's variables are RM's biographic, rating, number of clients, years of employment, fitting transactions rate (%), variable to fixed salary rate (%). Finally transaction data is represented by amount, date, liquidity, volume, client referred, riskiness, fitting transaction (Yes/No), Fitting asset class (Yes/No), Fitting geography (Yes/No). In the current setting variables starting with "Fitting" indicate whether the transaction is in line with the past transactions of the client in terms of magnitude, asset class, geography. Note that the dataset puts high importance on the RM's characteristics as they are the ones susceptible to fraudulent behavior. Based on this characteristics, the students simulated a synthetic dataset of 7,114 observations.

2.3.2 Model Selection

Unsupervised anomaly detection models are widely used to detect fraudulent activity in financial domains. They focus on identifying patterns that significantly deviate from the norm without requiring old labeled data. The algorithm are able to work

on the data they needs to classify directly without needing to see any training data prior to that. A key advantage is their ability to detect emerging or unknown fraud patterns that may not have been previously observed (Butvinik, 2020).

The underlying model chosen by the students is the Isolation Forest which is an algorithm for data anomaly detection using binary trees developed by F. T. Liu, Ting, and Zhou (2008). Such algorithms have already been used in external fraud detection, as shown by John and Naaz (2019) and Ounacer et al. (2018). Isolation Forest is explicitly designed for anomaly detection by isolating observations through recursive splitting. Unlike Random Forest, it aims to separate anomalies early in the splitting, which naturally isolates rare and unusual data points. This algorithm is computationally efficient and works well on large datasets. It assigns anomaly scores based on the number of splits required to isolate a point, with fewer splits indicating a higher likelihood of an anomaly. This method is particularly effective in high-dimensional spaces and in detecting outliers.

Consider a dataset $X = \{x_1, \ldots, x_n\}$, where each data-point x_i is d-dimensional, with characteristics given by $q_1, \ldots q_d$. Each Isolated Forest is made up of individually constructed Isolated Trees (iTree). Each iTree is defined as a data structure where for each node T in the Tree, T is either a leaf (external node) with no child, or an internal node with exactly two child nodes (T_l and T_r) and a splitting case. Each internal node has an associated splitting case characterized by a characteristic $q \in \{q_1, \ldots q_d\}$ and a splitting value p that determines how the dataset is split. To build an iTree, the algorithm recursively splits the dataset X by randomly selecting a characteristic q and a splitting value p until either the node has only one data point or all the data at the node have the same value.

Once the iTree is fully grown, each data point in X is isolated at one of the external nodes. The fundamental premise of the algorithm rests on the observation that anomalous points require fewer partitions to be isolated from the general distribution. For any point $x_i \in X$, the path length $h(x_i)$ is defined as the number of edges traversed from the root node to the external node, i.e. the number of dataset splits required to isolated x_i. Based on the path length, we can create an anomaly score and average it out over all the iTrees in the Isolated Forest. Data points with the lowest isolation score can then be identified as outliers.

Finally, a detection percentage is predefined as a parameter to determine how many points to label as anomalies (F. T. Liu, Ting, and Zhou (2008)). The students focused on 1% anomaly detection, but they note this parameter can be adjusted—either increased or decreased—depending on the application's specific requirements and tolerance for false positives versus false negatives. In practice, the ideal percentage may vary based on factors such as the underlying data distribution, the cost of misclassification in the particular domain, and the expected contamination rate of the dataset. Therefore, fine-tuning the detection percentage is essential

to achieve optimal performance in different scenarios.

2.3.3 Results

The students analyzed fraud detection at 1% level on a dataset of 7,114 observations. Out of these, 72 transactions were flagged as suspicious, which corresponds to 1.01%—aligning with the specified parameter. Figures 2.1 and 2.2 display the frequency distributions of transaction amounts for non-anomalous and anomalous data, respectively. When comparing the two frequency distributions, it is evident that a higher number of transactions under 1.0 billion are flagged. This trend aligns with the non-anomalous data, where transactions below 1.0 billion are substantially more common. Moreover, for transactions exceeding 1 billion, the detection of suspicious activity does not appear to be influenced by the transaction amount, implying that other features likely played a more significant role in flagging these cases.

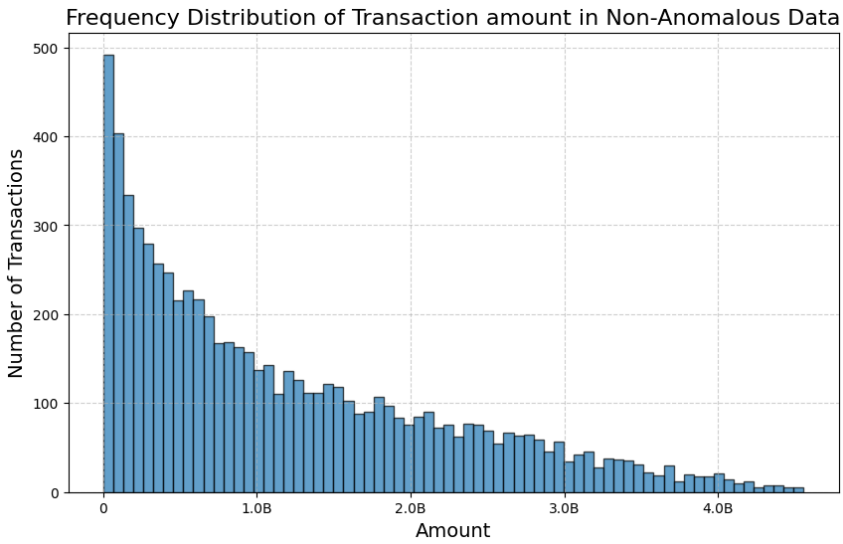

Figure 2.1: Amount of Non-anomalous transactions and their frequency

2.4 Proposed Implementation Strategy

The students propose a simple implementation procedure, where the Isolation Forest model is integrated within transaction monitoring systems to identify potentially fraudulent activity by RMs. The algorithm would analyze data on clients, RMs and transactions and flag transactions considered to be fraud. These flagged transactions can be reviewed by compliance officers for further investigation. This allows them to efficiently focus on high-risk transactions. This approach not only improves fraud detection, but also provides a time-saving and constantly evolving RM-driven fraud detection tool.

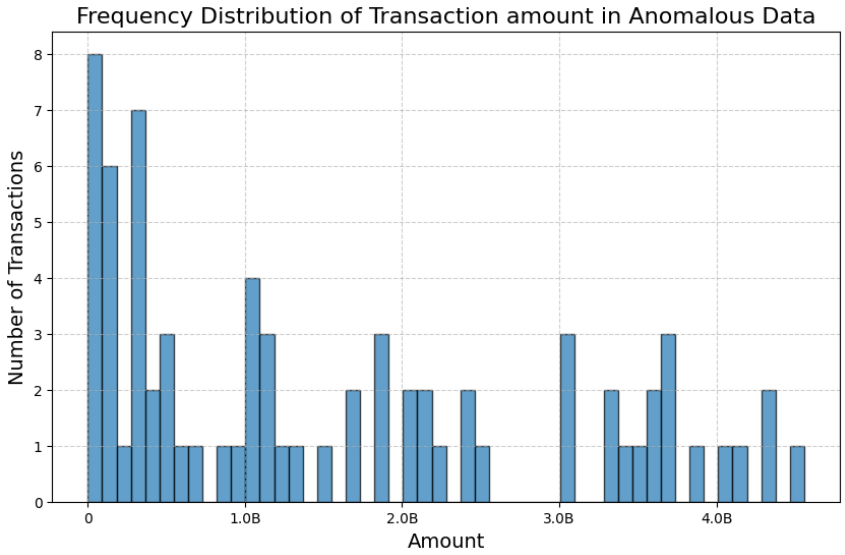

Figure 2.2: Amount of anomalous transactions and their frequency

2.5 Discussion

The key challenge for the banking sector is that the definition of fraud can vary considerably. It is therefore necessary to gain an understanding of what should be considered as an internal fraud, namely that it should not be characterized as an 'extreme' client's behavior. As a result, an extreme client's behavior has a tendency to be misidentified as a false positive in the model. The priority is to be able to distinguish between genuine and false positives as excessive false positives can lead to customer dissatisfaction, service disruption and the need for manual revision of processes which is both costly and time-consuming. Shifting the focus to RMs raises the challenge of determining appropriate input factors. The algorithm cannot rely solely on 'fitting' variables, as poor design of these variables can also lead to false positives.

The solution proposed by the students focuses on an unsupervised learning algorithm which overcomes the necessity of labeled that. This approach should be appealing to the industry as it does not rely on a dataset of labeled data. Rather, the algorithm can capture the dynamic nature of frauds, not relying on past frauds to detect new ones as they may evolve through time (Samantha Visbeek, Erman Acar, and Floris den Hengst (2024)).

The students argue that their approach is not intended as a definitive solution for internal fraud detection but as a preliminary approach. The use of anomalies detection eliminates the necessity to define a fixed profile for a fraudulent transaction, given that such profiles change dynamically. In their setting, any transaction considered unusual is flagged. The model's principal advantage is the reduction in time required for COs, who should therefore take the final decision. The time saved can be invested in reallocating resources to less mechanical tasks, such as more specific

investigations involving layering, churning or other techniques.

While their solution presents certain advantages, there are some key issues that will make industry adoption challenging. Although the proposed algorithm is relatively robust to feature scaling, it still faces challenges in effectively handling categorical variables, which may prevalent in banking transactions. Moreover, the method's assumption that anomalies are more susceptible to isolation may not hold true for sophisticated internal fraud schemes, where perpetrators often deliberately manipulate transactions to appear normal. Additionally, the requirement to specify a threshold level for anomaly classification introduces subjectivity into the detection process, potentially leading to either excessive false positives or missed fraud cases, depending on the chosen threshold value.

Unlike other AI models, Isolation Forest models do not have temporal awareness. Rather than consider the sequential nature of the transactions, the model processes them individually. A potential enhancement would be to incorporate time series analysis into the current model. Combining the Isolation Forest algorithm with approaches that consider transaction sequences, such as the method proposed by LaRock et al. (2020), could improve the detection of fraudulent behavior patterns while reducing false positives from unique but legitimate transactions.

Finally, the use of synthetic data in this study, while practical, raises questions about the model's performance with real-world data. Testing the algorithm against historical fraud cases would provide valuable insights, though accessing such data presents challenges due to confidentiality concerns and the possibility of undiscovered fraud cases in historical records.

References

ACFE (2024). *Occupational fraud 2024: A report to the nations*. Association of Certified Fraud Examiners.

Bai, Belinna, Jerome Yen, and Xiaoguang Yang (2008). "False financial statements: characteristics of China's listed companies and CART detecting approach". In: *International journal of information technology & decision making* 7.02, pp. 339–359.

Banking Frontiers (May 2024). *Employee fraud rises 73% in 8 years*. `https://bankingfrontiers.com/employee-fraud-rises-73-in-8-years/`.

Bolton, Richard J and David J Hand (2002). "Statistical fraud detection: A review". In: *Statistical science* 17.3, pp. 235–255.

Boukherouaa, El Bachir et al. (2021). *Powering the digital economy: Opportunities and risks of artificial intelligence in finance*. International Monetary Fund.

Butvinik, Danny (2020). *Online Machine Learning: Incremental Online Learning Part 2*. Insights Article, NICE Actimize. https://info.nice.com/Data-Science-Resources.html.

DeLise, Timothy (July 2023). "Deep Semi-Supervised Anomaly Detection for Finding Fraud in the Futures Market". In: CISMF Research Paper Series. https://marketsurveillance.esg.uqam.ca/wp-content/uploads/sites/161/Deep-Semi-Supervised-Anomaly-Detection-for-Finding-Fraud-in-the-Futures-Market.pdf.

Fanning, Kurt M and Kenneth O Cogger (1998). "Neural network detection of management fraud using published financial data". In: *Intelligent Systems in Accounting, Finance & Management* 7.1, pp. 21–41.

Hilal, W., S. Gadsden, and J. Yawney (May 2022). "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances". In: *Expert Systems With Applications* 193. https://www.sciencedirect.com/science/article/pii/S0957417421017164, p. 116429. DOI: 10.1016/j.eswa.2021.116429.

Holton, Carolyn (2009). "Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem". In: *Decision support systems* 46.4, pp. 853–864.

International Monetary Fund, Monetary and Capital Markets Department (Oct. 2024). *Global Financial Stability Report: Uncertainty, Artificial Intelligence and Financial Stability*. International Monetary Fund. URL: https://www.elibrary.imf.org/display/book/9798400277573/CH003.xml#CH003ref59.

John, Hyder and Sameena Naaz (2019). "Credit card fraud detection using local outlier factor and isolation forest". In: *Int. J. Comput. Sci. Eng* 7.4, pp. 1060–1064.

KPMG (July 2023). *KPMG Fraud Barometer*. https://kpmg.com/ch/en/media/press-releases/2023/07/fraud-barometer.html.

Kuna, Siva Sarana (2023). "AI-Enhanced Fraud Detection Systems in Digital Banking: Developing Hybrid Machine Learning Models for Real-Time Anomaly Detection and Customer Behavior Analysis". In: *Journal of Artificial Intelligence Research and Applications* 3.2. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=3cCORa8AAAAJ&citation_for_view=3cCORa8AAAAJ:MXK_kJrjxJIC, pp. 1086–1130.

LaRock, Timothy et al. (Jan. 2020). "HYPA: Efficient Detection of Path Anomalies in Time Series Data on Networks". In: *Proceedings of the 2020 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, pp. 460–468. ISBN: 9781611976236. DOI: 10.1137/1.9781611976236.52. URL: http://dx.doi.org/10.1137/1.9781611976236.52.

Liu, Chengwei et al. (2015). "Financial fraud detection model: Based on random forest". In: *International journal of economics and finance* 7.7.

Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou (2008). "Isolation forest". In: *2008 eighth ieee international conference on data mining*. IEEE, pp. 413–422.

Ounacer, Soumaya et al. (2018). "Using Isolation Forest in anomaly detection: the case of credit card transactions". In: *Periodicals of Engineering and Natural Sciences* 6.2, pp. 394–400.

Paula, Ebberth L. et al. (Dec. 2016). "Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering". In: https://www.researchgate.net/publication/313454543_Deep_Learning_Anomaly_Detection_as_Support_Fraud_Investigation_in_Brazilian_Exports_and_Anti-Money_Laundering.

Phua, Clifton et al. (2010). "A comprehensive survey of data mining-based fraud detection research". In: *arXiv preprint arXiv:1009.6119*.

Rezaee, Zabihollah (2002). *Financial statement fraud: prevention and detection*. John Wiley & Sons.

RiskOn (2024). *RiskOn Hackation*. https://www.riskon.ch/.

Sohl, Jeffrey E and AR Venkatachalam (1995). "A neural network approach to forecasting model selection". In: *Information & Management* 29.6, pp. 297–303.

Vashistha, A. and A. Kumar Tiwari (2024). "Building Resilience in Banking Against Fraud with Hyper Ensemble Machine Learning". In: *SN Computer Science* 5.3. https://link.springer.com/article/10.1007/s42979-024-02854-w, pp. 28–54. DOI: 10.1007/s42979-024-02854-w.

Visbeek, S., E. Acar, and F. d. Hengst (Dec. 2023). "Explainable Fraud Detection with Deep Symbolic Classification". In: *3rd International Workshop on Explainable AI in Finance*. https://arxiv.org/abs/2312.00586. 4th ACM International Conference on AI in Finance (ICAIF), p. 12. DOI: 10.48550/arXiv.2312.00586.

Visbeek, Samantha, Erman Acar, and Floris den Hengst (2024). "Explainable Fraud Detection with Deep Symbolic Classification". In: *World Conference on Explainable Artificial Intelligence*. Springer, pp. 350–373.

West, Jarrod and Maumita Bhattacharya (2016). "Intelligent financial fraud detection: a comprehensive review". In: *Computers & security* 57, pp. 47–66.

Whitrow, Christopher et al. (2009). "Transaction aggregation as a strategy for credit card fraud detection". In: *Data mining and knowledge discovery* 18, pp. 30–55.

Xuan, Shiyang et al. (2018). "Random forest for credit card fraud detection". In: *2018 IEEE 15th international conference on networking, sensing and control (ICNSC)*. IEEE, pp. 1–6.

Yuan, Fangfang et al. (2018). "Insider threat detection with deep neural network". In: *Computational Science–ICCS 2018: 18th International Conference, Wuxi, China, June 11–13, 2018, Proceedings, Part I 18*. Springer, pp. 43–54.

Chapter 3

How can AI be leveraged to enable CRO employees to work more efficiently?

In today's rapidly evolving global financial ecosystem, regulatory compliance has emerged as a critical and increasingly intricate challenge for financial institutions. As organizations navigate an increasingly complex landscape of international regulations, they must develop sophisticated strategies to manage and adapt their compliance documentation effectively. The stakes are particularly high for multinational financial entities, which must simultaneously address varying regulatory requirements across different jurisdictions, making the task of maintaining comprehensive and up-to-date compliance frameworks more demanding than ever before.

In Switzerland, the Swiss Financial Market Supervisory Authority (FINMA) enforces rigorous regulations that align with international standards, while the European Union's Markets in Financial Instruments Directive (MiFID II) and the General Data Protection Regulation (GDPR) introduce additional regulatory layers in countries such as France and Germany. Similarly, Asian markets such as Singapore and Hong Kong have their distinct regulatory challenges, each with robust financial regulatory frameworks.

This diverse range of jurisdiction-specific requirements, coupled with overarching mandates such as Basel III, necessitates a sophisticated approach to document management and revision. Financial institutions must not only ensure compliance with these extensive frameworks but also manage the frequent modifications to these regulations and address potential discrepancies across different regulatory regimes. Consequently, the deployment of an AI-driven solution is not merely advantageous but essential.

The 2024 RiskOn Hackaton Challenge 3, proposed by Julius Bär, aimed to develop an innovative AI-powered proof-of-concept tool designed to optimize regula-

tory document revision processes within the financial sector. The challenge sought to critically evaluate large language model (LLM) multilingual compliance verification capabilities, with a specific focus on demonstrating the potential of artificial intelligence to significantly reduce time and resource expenditures associated with complex regulatory documentation management. By exploring advanced analytical techniques, the challenge intended to provide a comprehensive assessment of AI's transformative potential in streamlining compliance procedures, while simultaneously addressing the intricate challenges faced by Chief Risk Officer (CRO) departments in managing regulatory documentation across diverse linguistic and jurisdictional contexts.

3.1 Literature

Regulatory compliance in the banking sector has become increasingly complex, particularly after the 2008 global financial crisis. Banks must adhere to a wide array of rules imposed by national and international regulatory bodies to mitigate risks, ensure financial stability, and protect consumers' interests.

The modern era of banking regulation is often traced to the Basel Accords. Basel I introduced standards for minimum capital requirements, which aimed to level the playing field across international markets (Supervision (2011)). Subsequent frameworks, including Basel II and Basel III, sharpened the focus on capital adequacy, supervisory review, and market discipline (Banking Supervision (2017), Vives (2016)). These measures responded to concerns about heightened systemic risk and underscored the need for robust internal control and risk management systems. Goodhart (2011) notes that these evolving international guidelines reflect ongoing attempts to harmonize rules while acknowledging the unique risk profiles of diverse banking institutions.

Despite such harmonization efforts, significant variations persist across jurisdictions. National policymakers frequently adapt global standards to align with domestic priorities, creating divergence in issues such as capital calculation, leverage ratios, and liquidity requirements (Schoenmaker (2013)). Howarth and Quaglia (2016) underscores that even within regions aiming for closer financial integration, variations in implementation can result in uneven compliance landscapes. his divergence poses particular challenges for multinational banks, which must navigate a mosaic of different supervisory expectations.

Compliance in banking is inherently tied to risk management. Scholars emphasize embedding compliance within an enterprise risk management framework to guard against reputational harm, fines, and disruption of operations (Adeniran et al. (2024)). An integrated, proactive compliance structure allows banks to move beyond a reactive stance, ensuring that risk identification, monitoring, and mitigation

become continuous processes. For institutions operating globally, organizational approaches to compliance may vary between centralizing functions at headquarters or granting greater autonomy to local subsidiaries (Ferner et al. (2004)). The choice depends on balancing centralized oversight with the need to respond effectively to local regulations.

Compliance is not solely a cost center. Studies indicate that robust compliance frameworks can foster long-term stability, bolster market confidence, and minimize the potential for disruptive financial events (Bird and Park (2016)). Nonetheless, banks face challenges in maintaining compliance with rapidly shifting rules, adopting new reporting requirements, and allocating sufficient resources to compliance staff and technologies.

With the explosion of financial technology (fintech) and the increasing sophistication of regulatory requirements, banks have turned to RegTech solutions to enhance compliance processes (Arner et al. (2017)). These technologies rely on automation, artificial intelligence (AI), and data analytics to streamline know-your-customer (KYC) procedures, anti-money laundering (AML) monitoring, and stress testing. Subsequent studies showcased the prowess of natural language processing (NLP) techniques in parsing regulatory texts to extract pertinent clauses tailored for financial institutions (Olawale et al. (2024)). Despite these advances, challenges persisted, particularly in processing multilingual documents and offering language-specific recommendations.

Natural Language Processing (NLP) has experienced significant strides in recent years, largely propelled by transformer-based architectures (Vaswani (2017)). The introduction of BERT (Bidirectional Encoder Representations from Transformers) has been a breakthrough, enabling deeper contextual interpretation of text (Lee and Toutanova (2018)). Parallel efforts have led to other notable models, such as GPT-2, GPT-3, and GPT-4, which further expand generative capabilities, refining tasks like text completion, summarization, and translation (Radford et al. (2019), Brown et al. (2020), Achiam et al. (2023)). Moreover, mT5 (multilingual Text-To-Text Transfer Transformer) highlights the power of a unified text-to-text framework designed explicitly for multiple languages, thereby enhancing the transferability of learned representations across diverse linguistic contexts (Xue (2020)). As a result, transformer-based models now not only achieve higher accuracy and efficiency in text analysis but also extend their impact across multilingual environments. Such developments have further democratized NLP research, as open-source frameworks have become increasingly reusable and adaptable, resulting in an expanded range of real-world applications.

Given the global nature of financial operations, the need for solutions adept at navigating multiple languages is paramount. Recent studies highlight the efficacy of multilingual transformer models can be used in analyzing financial sentiments across

different languages (Araci (2019)). Advanced models, such as GPT-4, have also demonstrated potential for streamlining tasks such as automated contract review and policy interpretation, thereby reducing time and cost for international financial entities (Kumar and Roussinov (2024)). However, while their efficacy in standard cross-lingual and sentiment-analysis tasks is increasingly evident, there is still a pronounced gap in systematically applying these models to the dynamic requirements of real-time regulatory document revision and compliance verification across multiple languages. Bridging this gap will require targeted studies that incorporate region-specific regulatory frameworks, domain-specific lexicons, and multi-document analysis strategies to fully exploit the advantages of multilingual NLP in the global financial industry.

Within the framework of RiskON 2024, the students extend the existing body of research by integrating a deployable DistilBERT model to refine regulatory compliance within the banking sector. They present a pragmatic, dual-purpose methodology that not only identifies requisite amendments but also autonomously generates language-specific recommendations aligned with the relevant regulatory framework. Furthermore, they propose a robust implementation strategy that effectively merges automation with critical human oversight, addressing notable gaps in both academic literature and real-world practice. As a result, this work has the potential to usher in a transformative approach to enhancing regulatory compliance across the financial industry.

3.2 Proposed Approach

Regulatory compliance in the financial sector is an increasingly complex and resource-intensive process, requiring organizations to invest substantial time and financial resources to navigate intricate legal frameworks, conduct detailed risk assessments, and implement comprehensive monitoring systems. The continuous evolution of regulations, such as anti-money laundering (AML) and know-your-customer (KYC) requirements, can demand significant annual expenditures, with large financial institutions often spending hundreds of millions of dollars and dedicating entire departments to ensuring strict adherence to regulatory standards.

To tackle the ever increasing cost of compliance, the winning team has focused on assessing the efficacy of large language models in identifying changes within regulatory texts. Specifically, they employ the DistilBERT model to automate the revision of regulatory documents, identifying changes and providing targeted recommendations across multiple languages. Their results indicate LLMs can significantly accelerate compliance processes and reduce translation costs, and clearly show the potential of AI to improve both efficiency and accuracy within the banking sector.

3.3 Methodology

The students conducted an in-depth evaluation of Julius Bär's regulatory compliance procedures, focusing on the bank's current levels of automation, data storage practices, and data governance frameworks. This assessment revealed several critical areas where the existing systems could benefit significantly from enhanced automation and AI integration. Specifically, it was noted that the bank's approach to managing regulatory changes and document revisions, especially across multiple languages and jurisdictions, posed substantial challenges in terms of efficiency and accuracy.

Their findings highlighted a clear opportunity for an AI-based tool to integrate seamlessly with Julius Bär's existing infrastructure, addressing regulatory compliance procedure challenges by automating the detection and analysis of changes in regulatory documents. The AI tool's ability to process multilingual data and provide targeted, actionable insights would directly support the bank's needs for more dynamic and robust compliance processes. This integration promises not only to enhance operational efficiencies but also to improve the overall accuracy of compliance practices, aligning with the strategic goals outlined by Julius Bär.

This evaluation underscored the need for an advanced AI solution capable of adapting to intricate regulatory environments and enhancing compliance operations, thereby reducing risks linked to manual processes and discrepancies in interpretation. By demonstrating the interoperability of their AI solution with the bank's current systems, the students established a critical foundation for the implementation of a more agile and efficient compliance management framework.

3.3.1 Model Selection

Given the constraints of time and computational resources, students opted for a smaller model that could deliver quick results while maintaining acceptable performance levels. DistilBERT emerged as an ideal candidate, being a streamlined variant of the well-known BERT architecture, specifically designed to provide a more efficient yet effective solution for natural language processing tasks. This compact model consists of 6 transformer layers, halving the number of layers compared to the original BERT's 12 layers. This reduction results in a model with 40% fewer parameters, significantly decreasing the computational load and improving the efficiency of model training and inference processes (Sanh (2019)).

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model introduced by Devlin (2018). It leverages the Transformer architecture to capture deep bidirectional contextual information from text, enabling a more comprehensive understanding of language nuances. The base version of BERT consists of approximately 110 million parameters and is trained on a large corpus

that includes English Wikipedia and the Toronto Book Corpus. In contrast, DistilBERT, first introduced by Sanh (2019), is a distilled version of BERT that retains approximately 97% of BERT's performance while utilizing 40% fewer parameters. DistilBERT simplifies the architecture by removing token-type embeddings and the pooler layer, which are less critical for performance in many tasks, and reducing the number of layers by a factor of 2. These modifications make DistilBERT an ideal choice for deploying state-of-the-art NLP models in environments where computational resources are limited or when faster processing times are needed without significantly sacrificing performance. This makes it a perfect candidate for Julius Bär's challenge.

3.3.2 Testing Pipeline

To tackle the challenge, the students focused on assessing DistilBERT's efficacy in identifying modifications within regulatory texts. They consider two contracts, one in English and one in German, to test the model's multilingual capability in a controlled proof-of-concept environment. Additional input data comprised of antecedent regulation (baseline document) and an updated regulation (revised version). These documents were provided in either plain text or PDF format, simulating a real-world scenario where legal teams compare historical and updated regulations.

To streamline the performance assessment, a straightforward testing pipeline was implemented. Given the complex nature of the bank's PDF documents, the students utilized Google's advanced DeepOCR technology to efficiently capture text, tables, and images. Following data extraction, the documents were then tokenized using DistilBERT's tokenizer to prepare them for NLP analysis. DistilBERT was subsequently utilized to perform semantic comparisons between the old and new versions of the law effectively identifying and highlighting notable changes, especially those with potential compliance implications. This method enabled them to efficiently pinpoint critical areas needing attention in the contracts.

To evaluate the model's multilingual performance capabilities, the identical analytical framework was systematically applied to both English and German contractual documents. This assessment was conducted without extensive fine-tuning on a diverse multilingual corpus, particularly one that extends beyond the financial domain, thereby testing the model's inherent out-of-the-box functionality. By employing a uniform methodological approach across these two languages, the study aimed to determine the model's ability to generalize its analytical processes in different linguistic contexts without the necessity for domain-specific adaptations. This strategy not only assesses the robustness and versatility of the model in handling multiple languages but also highlights its potential limitations when operating in diverse and less structured linguistic environments.

3.3.3 Results

The model generated outputs that highlighted text segments where changes were identified, providing preliminary evaluations of these changes—such as additions, deletions, or modifications. To ensure the validity of these results, a manual review was conducted in conjunction with ChatGPT-4, aimed at assessing the relevance and accuracy of the identified changes.

The findings from this validation process confirm that DistilBERT possesses a high degree of efficacy in processing multilingual datasets while also delivering actionable recommendations. The ability of DistilBERT to operate effectively within a multilingual framework substantially enhances its utility in international regulatory contexts, offering invaluable support for organizations operating within diverse and complex regulatory environments. Consequently, the model's capabilities not only improve operational efficiency but also reinforce the reliability and scalability of regulatory compliance mechanisms on a global scale.

3.4 Proposed Implementation Strategy

To effectively integrate AI-powered regulatory compliance tools with the banks' current infrastructure, the students advocate for a phased implementation strategy. This approach is designed to mitigate risks, build stakeholder trust, and ensure a seamless adaptation of the technology within existing systems.

The initial phase involves a pilot program where the AI tool is deployed to manage a select group of non-critical documents. This stage serves two primary purposes: first, it demonstrates the tool's capabilities in a controlled environment, and second, it begins to foster trust among the stakeholders by showcasing tangible benefits and operational enhancements. The success of this phase is critical as it sets the groundwork for broader acceptance and integration of the AI tool.

Following a successful pilot, the next step is a parallel running phase, where the AI system functions alongside the existing manual processes. This dual-operation allows for a comprehensive comparison between the AI-driven and traditional methods, providing insights into the efficiency, accuracy, and reliability of the AI solution. During this phase, the system's performance can be fine-tuned based on real-time feedback, ensuring that the model adapts well to the practical challenges of Julius Bär's operational environment.

As confidence in the AI system solidifies, the students recommend gradually extending its application to encompass a broader range of document types and regulatory areas. This phased scaling helps in managing the complexities associated with wider deployment, such as increased variability in data and heightened regulatory scrutiny. It also allows the organization to manage change more effectively,

minimizing disruption to existing workflows.

The final goal is full integration, where the AI tool is seamlessly embedded into Julius Bär's standard workflow management systems and document repositories. This phase marks the complete transition from traditional processes to a more sophisticated, AI-enhanced compliance framework. Full integration would streamline operations, reduce redundancy, and potentially lead to significant cost savings in regulatory compliance processes.

Critical to the success of each phase is a comprehensive training program tailored for compliance officers, legal teams, and other relevant staff. Training will focus on the operational aspects of the new system, as well as on the nuances of AI-driven decision-making. Equally important is clear and continuous communication about the capabilities and limitations of the AI system to set realistic expectations and ensure transparency.

To balance efficiency gains with risk management, establishing clear protocols for human oversight is imperative. These protocols should define the circumstances under which human intervention is necessary and outline the steps for such interventions. Ensuring that the AI tool enhances rather than replaces human expertise is crucial, maintaining rigorous standards of compliance and ethical responsibility.

As compliance regulations are always changing, the AI-powered tool needs a comprehensive strategy for continuous improvement. Regular retraining of the model with the latest regulatory updates and data is crucial. This ongoing adaptation allows the model to stay current with new compliance requirements and shifting market conditions. Retraining ensures that the tool's recommendations remain relevant and that its analytical capabilities evolve in line with the latest regulatory frameworks. Additionally, implementing robust feedback loops from users is essential to refine and enhance the model's performance. Feedback collected from end-users, who apply the model in real-world scenarios, provides invaluable insights. These insights help in fine-tuning the model's algorithms and improving the accuracy and relevance of its recommendations.

Periodic audits should be conducted to ensure the tool's compliance with advancing regulatory standards. These audits are critical to verify that the tool remains in strict alignment with current regulations and to identify any areas where adjustments may be required. This process not only supports regulatory compliance but also reinforces the tool's credibility and reliability.

By adopting these strategies, the model will continue to serve as a dynamic and responsive tool in the complex landscape of financial regulations. The continuous improvement framework ensures that the model does not just respond to current conditions but is also poised to adapt to future changes, maintaining its operational efficacy and compliance integrity in a perpetually evolving market.

3.5 Discussion

The students' approach was subject to several limitations due to the inherent time constraints of the challenge. First, DistilBERT was not fine-tuned on domain-specific or extensive regulatory texts, nor on specialized company vocabulary, which could potentially enhance its performance for specific compliance tasks. Second, the analysis was confined to only two documents, which may not fully capture the complexity of real-world applications. Lastly, the evaluation predominantly focused on qualitative insights instead of quantitative metrics, such as precision or recall, which could provide a more rigorous assessment of the model's performance.

Implementing AI-driven solutions within a banking context involves navigating several challenges, particularly concerning legacy systems, regulatory compliance, and data privacy. Many banks still rely on older infrastructures that may not seamlessly integrate with advanced AI technologies. It is necessary to assess the maturity of the bank's data culture and ensure existing systems can interface effectively with modern AI solutions. This may require upgrades to existing hardware or middleware to bridge older and newer technologies.

Another crucial consideration is data privacy and security. Banks must implement robust measures to safeguard sensitive financial information against unauthorized access or breaches. This can include encrypting data both in transit and at rest and employing secure protocols for data handling and analysis to maintain confidentiality and integrity.

Regulatory compliance introduces further complexity. Banks must navigate a dense regulatory landscape to implement AI tools for compliance monitoring, including securing necessary approvals from financial regulators. Each tool must be rigorously tested to confirm it meets risk management and data protection requirements mandated by governing bodies.

Additionally, it is important to prepare the bank's personnel for the transition to AI-enhanced systems. Employees need training not only in using the new tools but also in recognizing and addressing possible errors or anomalies that AI might introduce. Ongoing education will help empower employees to collaborate effectively with AI solutions.

Looking ahead, while the initial integration of AI can handle general contract analysis, adapting these solutions to address unipersonal contracts remains an area for future development. Incorporating real-time regulatory updates will enable the tools to provide immediate adjustments, which is particularly critical in volatile financial environments where requirements can shift rapidly. Such adaptations typically demand sophisticated contextual understanding of individual circumstances and regulatory directives.

Successfully addressing these challenges can facilitate a smoother transition to

AI-driven compliance strategies within banking. However, the evolution of AI applications will be influenced by changing regulations and the degree to which banks can accommodate new technologies. Ongoing research and development will be essential to extend the capabilities of AI systems in the financial sector and beyond.

References

Achiam, Josh et al. (2023). "Gpt-4 technical report". In: *arXiv preprint arXiv:2303.08774*.

Adeniran, Ibrahim Adedeji et al. (2024). "Strategic risk management in financial institutions: Ensuring robust regulatory compliance". In: *Finance & Accounting Research Journal* 6.8, pp. 1582–1596.

Araci, D (2019). "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". In: *arXiv preprint arXiv:1908.10063*.

Arner, Douglas W et al. (2017). "FinTech and RegTech: Enabling innovation while preserving financial stability". In: *Georgetown Journal of International Affairs*, pp. 47–58.

Banking Supervision, Basel Committee on (2017). "Basel III: Finalising post-crisis reforms". In: *Bank for International Settlements*.

Bird, Robert C and Stephen Kim Park (2016). "Turning corporate compliance into competitive advantage". In: *U. Pa. J. Bus. L.* 19, p. 285.

Brown, Tom et al. (2020). "Language models are few-shot learners". In: *Advances in neural information processing systems* 33, pp. 1877–1901.

Devlin, Jacob (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Ferner, Anthony et al. (2004). "Dynamics of central control and subsidiary autonomy in the management of human resources: Case-study evidence from US MNCs in the UK". In: *Organization Studies* 25.3, pp. 363–391.

Goodhart, Charles (2011). *The Basel Committee on Banking Supervision: A history of the early years 1974–1997*. Cambridge University Press.

Howarth, David and Lucia Quaglia (2016). "Banking on stability: the political economy of new capital requirements in the European Union". In: *Redefining European Economic Governance*. Routledge, pp. 139–152.

Kumar, Bimal and Dmitri Roussinov (2024). "NLP-based Regulatory Compliance–Using GPT 4.0 to Decode Regulatory Documents". In: *arXiv preprint arXiv:2412.20602*.

Lee, JDMCK and K Toutanova (2018). "Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* 3.8.

Olawale, Olufunke et al. (2024). "Risk management and HR practices in supply chains: Preparing for the Future". In: *Magna Scientia Advanced Research and Reviews* 10.02, pp. 238–255.

Radford, Alec et al. (2019). "Language models are unsupervised multitask learners". In: *OpenAI blog* 1.8, p. 9.

Sanh, V (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv preprint arXiv:1910.01108*.

Schoenmaker, Dirk (2013). *Governance of international banking: The financial trilemma*. Oxford University Press.

Supervision, Banking (2011). "Basel committee on banking supervision". In: *Principles for Sound Liquidity Risk Management and Supervision (September 2008)*.

Vaswani, A (2017). "Attention is all you need". In: *Advances in Neural Information Processing Systems*.

Vives, Xavier (2016). *Competition and stability in banking: The role of regulation and competition policy*. Princeton University Press.

Xue, L (2020). "mt5: A massively multilingual pre-trained text-to-text transformer". In: *arXiv preprint arXiv:2010.11934*.