Structured Data Extraction and Regulatory Narrative Generation: Machine Learning Models for Automated Financial Reporting and Disclosure

Dr. Xiaojing Wang, Professor of Electrical and Computer Engineering, University of Illinois Urbana-Champaign (UIUC)

1. Introduction

Reliability, automation, and accuracy in financial reporting have become more pertinent than ever with rising regulatory requirements, rapid expansion in business activity, and increasing complexities, as these require analysis of larger volumes of financial data than ever before. Artificial intelligence (AI) could be a savior in this world to drive efficient, automated, and compliant financial reporting. AI can surpass human capabilities in analyzing enormous data sets, identifying patterns and trends, enabling real-time insights on business performance, identifying fraudulent activities, and ensuring compliance. AI has the ability to present the results in real-time to enable efficient decision-making. Timely and effective decision-making gives a company an edge over others. AI models are trained to learn from the data provided to evolve their model, ultimately increasing their precision, improving outcomes over time, and helping in better decision-making as new data is captured and analyzed.

However, integrating AI is also fraught with its own perils and bottlenecks, including data protection risks. This essay introduces you to the world of intelligent financial reporting and subsequently discusses the cues to AI in finance, including risks identified by regulators, the functional areas across organizations where AI is being implemented, and survey results on AI impact and challenges. The use of AI to support professionals handling investments and transactions is an increasingly popular option to give firms increased confidence in outcomes and a competitive advantage. Further, the essay provides examples of the use of AI in finance. Finally, the essay provides a critical analysis, considering questions such as—Can AI be trusted? Will jobs be lost? Why consider the risks in developing AI models? What is the need to train governance analysts in AI testing?

2. Background and Significance of Automated Financial Reporting

Traditional approaches to financial reporting are time-consuming, strongly depend on manual processes, and are often prone to error. This is increasingly seen as a problem in fast-changing financial markets. The increase in the pace of businesses, their geographical spread, and developments in business laws and regulations lead to numerous financial data reporting requirements. As a result, corporations, from either external requirements to internal, require twice the amount of financial reporting than they did 15 years ago. In response to this, many technical advancements have been made in fully automated and semi-automated financial reporting in the past two decades. The main factor contributing to this progress is continuous improvements in AI and machine learning capabilities alongside software and hardware infrastructure.

Several initiatives have driven the transition from manual to automated financial reporting. In the 1970s, stock markets began to move towards high-frequency trading, and more recently many regulators now require real-time financial disclosures. The first paper describing an automated financial disclosure system was published in 1986. In 1997, the SEC began publicly releasing financial filings through EDGAR, allowing processing of financial filings electronically. Companies also publish real-time financial disclosures on the internet, which are increasingly in demand. The use of IFRS worldwide and their specific parliamentary financial reporting requirements have driven the automation of financial reporting in many developing countries. Financial data is an essential element of the success of any organization because with this data they can understand if they are losing or making profits. Data that are delayed, unavailable, incomplete, or inaccurately produced could lead to critical wrong decisions and a drop in the company and its share prices, as was evident in the subprime mortgage crisis. Automated financial reporting contributes to the success of an organization by improving decision-making capabilities. Providing investors with accurate information in real time will reduce the amount of investment due diligence by these and result in more efficient financial markets. Automated reporting reduces bureaucracy and therefore corrupt practices that can also negatively affect a firm. It provides an equitable mechanism for an organization to manage their costs and resource allocation. Automated reporting systems that generate reports by retrieving data from across the company provide transparent information to an organization's shareholders, auditors, and other stakeholders regarding how the data that are subject to the financial statements were obtained, classified, and recorded. They separate management judgment from these processes. Automated reporting systems reduce the risks of private interest in financial reporting data. Whether errors in financial reporting systems occur due to fraud or lack of funds, or due to judgment or interpretation of a business's policies, financial reports that are not transparent must necessarily contain these errors. Automating parts of financial reporting systems reduces this risk. That is, automated reporting systems can remove human error for the part of the process that does not rely on management judgment.

3. Machine Learning Techniques for Financial Data Analysis

Nowadays, with the large amount and complex structure of financial data, it is necessary to understand how to explore the data to dig potential laws between items, subsidiary companies and ultimate controllers, and short-term liabilities and losses. In the process of financial statement analysis, machine learning is a general technology for extracting useful information from financial big data. This section introduces the commonly used machine learning techniques in financial reporting.

Supervised learning, also known as recognition-based learning, can predict financial outcomes with post-confirmatory evidence; that is, the relationship between samples and categories has been trained. A typical research method for supervised learning-based financial reporting is ratio analysis.

Unsupervised learning, also known as reasoning-based learning, has advantages in identifying features from large data without labeled targets. Through unsupervised learning, one can qualify hidden patterns and anomalies without agreement and positive approval of designated evidence.

Given the complex structure and large amount of financial data, deep learning is an effective method to automatically extract features from raw data. Characterized by complex neural units, deep learning can reduce the tedious selection of algorithms through specialties. Deep learning technology is becoming more and more popular, and it enriches the design of financial report processing.

As the above shows, there is no silver bullet of machine learning technology suitable for all financial report processing. Rather, we should bear in mind the pragmatism that different machine learning technologies are suitable for different financial report working processes. In the following, we summarize the relative strengths and limitations of these machine learning techniques.

3.1. Supervised Learning in Financial Reporting

Predictive modeling fundamentally focuses on the area of supervised learning. Supervised learning can be defined as building predictive models from transaction data with a focus on identifying outputs from specific data inputs. Hence, the migration from paying vendors for business intelligence reports to the deployment of AI and machine learning to uncover insights from real data describing organizational activity. The supervised learning process involves an algorithm turning the labeled dataset into a trend line. Labeled datasets are a set of training pairs in which there is a conceptual or theoretical answer underscoring the actual answers. The crowd or organizational objectives would include the recognition of some information from the input data, estimating how much, and the identification of anything wrong from the data. The business is left to offer a conceptual answer to this using human cognition to label examples.

A total of 95% of predictive models built in a banking or trading application are built using algorithms to acquire intelligence from artificial data to inform a decision, like in the fraud detection process. Many algorithms used in predictive modeling focus on decision-making to acquire intelligence about potential outcomes, like decision trees. Another technique frequently included is algorithms like regression analysis with output predictions placed on a continuous scale, like income and age. The use of supervised models is increasingly being included in auto-financial reporting to assist in automated risk assessment, fraud detection, or credit scoring. Predictive models can help leverage the likelihood of future activity, like car insurance, having better data analysis create predictive pricing. This is seen in the severity and likelihood models which are constructed using regression analysis. Meanwhile, decision tree models attempt to predict the probability of what activity should be undertaken when out of the normal. In evaluations of predictive models in fintech, automatic predictions account for exceeding a 10% reduction in the frequency and net bad on-book experience. Rather than extract these trends, predictive models are taking into consideration the relationships and interactions among several variables or sets of data to render them more accurate. Ten develop predictive models.

3.2. Unsupervised Learning for Pattern Recognition

You may use unsupervised learning to recognize patterns in a dataset that can then be used as a basis for automated cash flow statement forecasting. This way, cash flows with similar descriptions will be forecast similarly, which cannot be achieved with a regression model. In fraud detection, unsupervised learning is also frequently employed to find patterns distinguishing fraudulent transactions among the vast majority of non-fraudulent ones. Moreover, unsupervised learning is relevant for credit scoring models as it can be used by issuers to classify borrowers with similar characteristics into groups that share the same creditworthiness. Firms can also use unsupervised learning to identify process inefficiencies based on patterns in their transaction data so that RPA can be implemented to cut down on the human effort necessary for regulatory reporting or management accounting and increase the overall analytical capability and level of detail of a firm's financial reporting. Unsupervised learning is frequently used for anomaly detection because it does not require labeled data. It can be applied in a variety of areas, such as finding unusual transactions that might be the result of fraud. Nevertheless, determining the effectiveness of an unsupervised learning model is often rather complicated because of the lack of a "gold standard," a pre-labeled dataset to compare with. This technology is used in transaction scoring. To dramatically cut the number of alerts and "false positives" that require additional attention from compliance officers, broader analytical profiling of customer and product segments will be done. In risk analysis, it has been shown that applications of unsupervised learning to the detected anomalies showed key customer segments and product interactions. Income and spending behavior analysis of retail customers may also involve using unsupervised learning.

3.3. Deep Learning Applications in Financial Data Analysis

Deep learning, or adaptive machine learning, marks the modern-day era of truly data-driven artificial intelligence. Its pioneer is back-propagation neural networks, which explore linear and non-linear relationships within the input data using layers of artificial neurons – and therefore the model's name: neural networks. By stacking up neurons, the deep learning model leverages big data over time to learn various complex patterns within multi-modal networks as a master learner. This growing massive body of knowledge results from cumulative dynamic interactive learnings of all neural network learners and operations in every layer as a time-evolving system. This makes the model continuously learn and adapt to new data over time, capable of performing an almost unbounded scope of data manipulation and mapping, prediction, and decision making. Each island of subsequent chains of individual computation is either called a dense layer or a skip connection. This is exactly why deep learning models can handle unstructured data, such as images, videos, natural language, audio, etc., without – or with minimal – pre-processing required.

Deep learning technology has begun to find extensive application in the finance sector. It has been employed for predicting the volatility of exchange rates in algorithmic trading. In quantitative finance, it has been employed to predict future prices for commodities, stocks, and derivatives in either internal proprietary trading or forms of SaaS and B2B systems. For strategic financial planning and credit risk assessment, deep learning boosts the decision-making capabilities of lenders by identifying potential defaulters using a credit scoring system in mortgage finance, corporate banking, credit lending, and leasing. Alternatively, or complementarily, by quickly assessing loan applicants and advancing scores. Financial institutions could lower insurance premiums by auditing user emotions on social media regarding sentiments about a product. Emotions of tweets on financial text data are used by institutional traders and market researchers for sentiment analysis of stocks. Given the practical applications above, this implies that deep learning can also be used to standardize the performance of financial processes within the presentation of financial data at automated intelligent reporting systems. However, deep learning, as a neural network and a form of soft computing, faces different challenges related to financial artificial intelligence, which may be exacerbated or precipitated by the COVID-19 pandemic. Such challenges include the ever-changing market observations in computer vision, prediction, and language of deep learning in finance, no clear financial data quality in the literature, blaming deep neural networks with the help of explainable AI, the highest computational cost of building, training, and testing deep neural networks, and the lean idea of asset data acquisition in different fields. In the next sections, we review the deep learning employed to address how adaptive and soft intelligence could be used to automatically report AI-generated intelligent and explainable insights from financial data by automating the three processes of financial data analysis. Finally, in order to substantiate the expected research pattern, a case study will present a modern deep neural network architecture and explain how it can be re-employed in different learning tasks for financial data analysis.

4. Challenges and Limitations in AI-based Financial Reporting

AI-based financial reporting faces several limitations and challenges. From a technical perspective, data quality is often a key problem as poor or inconsistent underlying data make it harder for any AI to provide reliable reports. Moreover, regulations within the financial sector are complex. Achieving compliance with an algorithm presents two challenges: first, the algorithm must be adaptable to changes in the law, and second, the law may be subjective. For example, laws that stipulate that banking AI must be 'fair' are drawing ethical considerations directly into regulation. Building ethical responsibility into AI algorithms is challenging, too. Many financial stakeholders worry about introducing new automated decision-making processes, fearful of the unintentional harm they may cause to their customers or processes. Arguably, this is a rational fear, witnessed in notable cases in AI where AIs have not acted in line with even loosely stated ethical rules.

AI requires an algorithm, data, and expertise to run effectively. Many smaller financial organizations simply lack the software expertise required to implement AI. Although AI technologies are constantly improving, they are neither infallible nor omnipotent. Automated AI systems are only as good as the humans who create them, the data they are fed, and the models and rules they follow. Moreover, AIs need to be updated continuously in a process that requires new data, testing, and iterations. AI continues to pose operational risks as it is integrated into financial services systems and business operations. AI and machine learning also pose new challenges to 'fairness', which has implications for consumer protection and societal well-being generally.

5. Future Trends and Opportunities in Automated Financial Reporting

This section examines the rapid changes currently taking place in the automated financial reporting landscape, detailing future opportunities for both practitioners and accounting academia. AI is a rapidly evolving field, leveraging machine learning to develop pattern recognition and anomaly detection tools across a wide range of industries. Notable sub-technologies include natural language processing to automate narrative and commentary, and robotic process automation to identify repetitive financial consolidation and adjustment tasks. In response to these innovations, vendors are integrating previously separate financial, managerial, and compliance systems into end-to-end solutions, with certain existing technologies now capable of producing and delivering fully integrated financial statements, balance sheets, and cash flows in a single process.

Internal and external collaboration is also increasingly digitized, to the extent where banks and investors can access real-time financials over secure online platforms. Forward-thinking organizations are recognizing the potential of real-time stream analytics and associative workflows, particularly in industries like logistics where managing big data leads to decreased idle time and improved efficiency. Furthermore, data buzzwords like "Artificial Intelligence," "Big Data," and "Machine Learning" are appearing in organizational strategic management system conversations with increasing frequency. Financial professionals can respond by upskilling from automatable finances and traditional accountancy services to become invaluable strategic advisors, partners, and change agents. Accounting academia must similarly evolve to teach the emerging skills most relevant to today's firms that seek both ethical personal behavior and improved business models.

Notably, accountancy-related regulatory bodies are on the brink of some significant shifts, mainly toward cloud-based governmental services, with many implementing or experimenting with applications through both domestic and international frameworks. This suggests new regimes of compliance, so the safe bet would be to reflect upon how best to prepare for these forthcoming regulations. Future trends in compliance specifically related to automated financial reporting or data might provide some equally exciting opportunities for academia and practitioners. Finally, recent developments are increasingly demonstrating professional value in the informed ethical use of sensitive personal and talk-based relational, emotional, and social data. Furthermore, many significant trends are connecting the increasing sophistication of automated financial reporting to the wider potential for a data-driven future society. Data analytics and machine learning will likely be crucial accounting topics down the track as the profession continues to evolve, remaining relevant in the economic structure.

Furthermore, professionals who bring their accounting and finance insights to discussions of privacy, society, and governance will likely develop a competitive advantage over those who do not.

6. Conclusion

The purpose of this Essay is to provide a detailed analysis of the industry's current standing with respect to automated financial reporting and financial machine learning models. By reviewing various machine learning models and how they are purported to function, key understandings have been presented to the Financial Profession and the greater industry. A number of points can be taken away, the primary one being the fact that AI and ML models are seen as a useful tool in accounting procedures, capable of speeding up the financial reporting process and increasing the precision of the final financial reports.

Despite the advancements in the technology, a careful integration of the system is needed as the very specifics of machine learning models can call for both formal and moral consideration. A good knowledge of each model is beneficial, as well as the type of transaction-sets most-suited with each AI system. Challenges and limitations must be taken in account in the integration of AI in the financial reporting. This essay also demonstrates the emerging nature of accountancy and financial reporting in light of the AI evolution.

While a greater need in AI specialists is identified, financial professionals possess the data required in automated financial reporting models. Accountants are encouraged to keep up with future developments in the area to stay relevant in their role. Unlike older techniques, reinforcement learning and Portfolio optimization are AI systems that require ongoing attention and investment to get best results. Like a diver with a watchful eye to the time and air supply, the financial professional working in this area needs to be attuned to the continuous behaviour of these programs; collaboration between accountants and data specialists is key. With ever-advancing technological developments, the operating environment for finance professionals is changing. As the biggest users of financial datasets, those working in financial reporting are challenged to perceive, accept and use AI data models before they are overlooked.