NUML International Journal of Business & Management
Volume 21, Issue 1, June (2026)
Journal Home Page: nijbm.numl.edu.pk/index.php/BM
ISSN 2410-5392 (Print), ISSN 2521-473X (Online)

Integrating Disclosure Tone and Machine Learning for Financial Performance Prediction

Immad Ud Din Durrani[1], Hassan Raza[2*]

Article History:
Received: December 6, 2025
Revised: April 27, 2026
Accepted: June 16, 2026
Available Online: June 30, 2026

Keywords: Disclosure Tone, Sentiment Analysis, Machine Learning, Financial Forecasting, Pakistani Non-Financial Entities, Random Forest.

ABSTRACT

This study examines the effect of the tone of corporate disclosures (positive, negative, or uncertain) on the financial performance of non-financial firms in Pakistan using machine learning techniques. Most existing financial forecasting models rely solely on quantitative data and ignore qualitative data, such as corporate disclosures. To fill that gap, we examine key performance ratios alongside tone variables for major performance variables such as Return on Assets (ROA), Return on Equity (ROE), and Earnings per Share (EPS). The tone variables were computed using disclosures of annual reports, using a narrative-specific lexicon of finance, and the scores were used in combination with the financial ratios to train four learners: Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest (RF), and Long Short-Term Memory (LSTM). The panel comprises 164 firm-year observations from 40 non-financial firms on the PSX, over the period 2019–2022. Random Forest was the best model overall, achieving an accuracy of 0.82 for EPS, an ROA accuracy of 0.82, and an ROE accuracy of 0.85, while SVM and ANN were in the middle, and LSTM lagged (EPS: 0.64). This study proposes a new hybrid forecasting model that combines qualitative tone analysis with quantitative data to address the limitations of traditional forecasting models. It provides investors, analysts, and policymakers with useful insights in an environment of information asymmetry in emerging markets and helps advance the development of financial technology and corporate disclosure. Since positive tone had the greatest predictive significance for EPS, and litigious tone had the greatest predictive significance for ROE, the results suggest specific screening rules, such as screening reports that combine heavy use of litigious words with weak reported returns, to warrant a more thorough credit review.

[1] MS Scholar, Department of Management Sciences, SZABIST University, Islamabad.
[2] Associate Professor, Department of Management Sciences, SZABIST University, Islamabad. dr.hassan@szabist-isb.edu.pk

1. Introduction

Accurate prediction of financial performance is crucial for the strategic decision-making of investors, security analysts, and policymakers (Al-Okaily & Al-Okaily, 2024; Zheng et al., 2024). The accuracy of forecasts positively affects investment returns, resource allocation, regulatory compliance, and sound investment plans (Akinsulire et al., 2024), thereby influencing future performance (Liu et al., 2023; Olorunyomi et al., 2024). Most of these models are based on static assumptions that are not consistent with today's world. These contexts are highly dynamic and subject to rapid changes, creating significant challenges for real-time adaptability, particularly in emerging and rapidly growing economies, which are most prone to such contexts (Aljohani, 2023; Brunner et al., 2024).

Theoretically, the current literature in financial forecasting is largely grounded in the Efficient Market Hypothesis and classic econometric models that presuppose linearity and market rationality (Jin et al., 2020; Oyewole et al., 2024). These theoretical bases are inadequate for explaining how qualitative information, especially the tone embedded in corporate disclosures, can be used to predict financial performance. According to signaling theory, management language in annual reports provides information about the firm's future activities; however, this theoretical implication has not been systematically implemented in predictive modeling platforms, particularly in emerging market environments (Iqbal and Raza, 2021; Mousa et al., 2022). In practice, financial analysts and investors in markets such as Pakistan still rely on traditional ratio-based instruments that do not capture the non-linear relationship between quantitative and qualitative measures. The existence of this twofold divide, both theoretical and practical, indicates a need for a combined methodology to bridge the gap between narrative analysis and machine-learning-assisted financial prognostication.

Financial markets in countries such as Pakistan are still emerging due to regulatory changes, economic volatility, and the structures put in place regarding accounting standards (Kiani et al., 2023; Shah et al., 2024). Predictive analytics, as a science, strives towards this through the development of innovative tools. Advanced machine learning techniques have been implemented to enhance prediction accuracy by integrating quantitative financial variables with qualitative insights from text, especially corporate disclosures (Hajek & Munk, 2024; Mousa et al., 2022). Machine learning algorithms are becoming increasingly popular in the financial industry for identifying complex patterns and relationships that traditional models cannot capture (Sarker, 2021; Singh et al., 2022). The financial industry seeks to leverage the capabilities of machine learning algorithms because they can handle large volumes of structured and unstructured data (Nguyen et al., 2023; Patil et al., 2024). Although many studies have been reported on predicting bankruptcy and forecasting stock markets (Golbayani et al., 2020). These models typically focus on quantitative financial data (Teles et al., 2021).

In the current markets, there are several drawbacks to traditional forecasting tools. A ratio provides a backward-looking analysis of a firm's liquidity, profitability, and solvency characteristics; it does not necessarily measure current or future characteristics (Aljohani, 2023). The time-series methods, e.g., AutoRegressive Integrated Moving Average (ARIMA), are not well-suited to the non-stationarity and non-linearity of EMs because they assume stationarity and linear dependence (Jin et al., 2020). In theory, discounted cash flow models are a useful approach, but they rely on assumptions about cash flow timing and discount rates, and errors in the inputs grow significantly in situations of poor financial transparency (Bola et al., 2021). Furthermore, all these traditional approaches are strictly quantitative and miss qualitative information in the disclosures, including the tone of firms' MD&A segments, which have been shown to contain predictive value for subsequent firm performance (Loughran and McDonald, 2020; Mousa et al., 2022). The failure to handle unstructured text is a fundamental weakness, as much of the narrative content of annual reports contains premonitions of strategic changes, risk exposures, and management confidence that are not readily reflected in the numbers. This is even more severe in weaker economies such as Pakistan, where financial information can be limited, inconsistent, or subject to regulatory loopholes, and qualitative supplements are not only beneficial but necessary for making accurate predictions.

Machine learning methods offer high-quality financial predictions, yet remain relatively unexplored in emerging economies (Alvi et al., 2024; Patil et al., 2024). Many of these models are initially designed for developed markets and are heavily criticized for failing to capture the characteristics of emerging economies, thereby necessitating localization (He et al., 2023; Helden et al., 2021). Corporate behavior patterns, disclosure practices, and investor reactions in Pakistan differ markedly from those observed in developed countries; thus, it is essential to develop context-specific models (Anwar, 2021; Faisal, 2021). In addition, incorporating textual data, especially disclosure tone, into these models remains in its early stages in emerging markets such as India (Chakraborty & Bhattacharjee, 2020; Jain et al., 2023).

There has been great potential shown by deep learning and ensemble methods through machine learning techniques, including textual data in financial forecasting (Li & Pan, 2022; Patil et al., 2024). For example, textual data has been analyzed using deep learning models, namely extended short-term memory networks and convolutional neural networks, which help discover subtle linguistic patterns that may not be captured by traditional approaches. Integration of textual features with numerical financial data can enhance the accuracy of stock price predictions (Abimbola et al., 2024; Jain et al., 2021). Most research has focused on stock price movements rather than on financial performance measures such as profit and return stability (Almagtome & Abbas, 2020; Wirawan & Willim, 2023).

Traditional forecasting methods lack the ability to deal with the global financial markets in a highly competitive environment (Bustos & Pomares-Quimbaya, 2020; Thakkar & Chaudhari, 2021). The use of financial forecasting methods, including the integration of machine learning and sentiment analysis, is crucial for enhancing forecast precision and accurately reflecting the intricacies of financial markets in the modern era (Sharma et al., 2024; Bello, 2023). Financial ratios and historical data are the core of the traditional method for analyzing a company's future performance. The focus on the tone of reports, particularly in annual reports, can help anticipate financial results (Iqbal & Riaz, 2021; Mousa et al., 2022). However, integrating these non-financial variables with financial data into predictive models has been relatively underexplored, especially in developing countries such as Pakistan (Ahmed et al., 2021; Khan et al., 2023).

Large, complex datasets comprising quantitative financial data and qualitative textual data can be used to improve predictive accuracy with machine learning models such as random forests, deep learning, and support vector machines (Bansal et al., 2022; Nosratabadi et al., 2020). Previous studies have also demonstrated the potential of machine learning in financial prediction, but its specific application in predicting financial outcomes, such as profitability, earnings, and sustainability, using financial and textual information, such as disclosure tone, is limited and has been tried infrequently (Zhang & Zhang, 2024; Zhang et al., 2022). Although the existing literature focuses on the relationship between disclosure tone and the prediction of corporate risk, bankruptcy, and stock price movements, the relationship between disclosure tone and other financial variables remains unclear. (Iqbal & Riaz, 2021; Mousa et al., 2022).

This research examined the prediction of financial performance in Pakistan, where conventional tools, such as ratio analysis and regression models, fail to capture non-linear interactions and qualitative elements, such as the tone of disclosures. Emerging markets of this kind have few predictive models built around their own conditions, and Pakistan is no exception. We set out to fill that gap by feeding sentiment extracted from corporate disclosures into machine learning algorithms. Prior attempts to apply machine learning to financial forecasting have given limited attention to emerging markets and to the link between disclosure tone and finance. This research aims to improve predictive accuracy in erratic markets by incorporating both financial and non-financial variables.

Despite recent advances in using machine learning to predict company performance, more work is needed to bridge the gap between machine learning models and disclosure-based tone analysis (Ranta et al., 2023; Zhang et al., 2022). Previous research primarily relies on linear regression analyses for stock price prediction, which often fail to capture nonlinear relationships (Jin et al., 2020; Oyewole et al., 2024). Further, there is limited literature on applying machine-learning-based financial prediction in emerging economies, including Pakistan, compared to developed economies (Ashraf et al., 2020; Khalid et al., 2022).

There has been recent interest in research exploring the applicability of machine learning in emerging markets (Dana et al., 2022; Goodell et al., 2021). However, such studies are often limited to regional and industrial sectors. This research aims to overcome this limitation by using machine learning to predict financial narratives in a distinctive Pakistani context and by incorporating disclosure tone into the models. There is a positive relationship between optimistic tones in disclosures and future profitability (Alalwani & Mousa, 2020; Fisher et al., 2020). Financial decisions can reflect different tones, such as optimism and risk, and this study aims to explore how these tones relate to financial performance, especially in volatile markets (Alshorman & Shanahan, 2022; Hassan et al., 2022). The generalization of deep learning models has been discussed, but extensions to other financial measures remain relatively limited (Huang et al., 2020; Patil et al., 2024). Newly developed deep learning approaches are superior to conventional approaches, yet no comparative assessment has been conducted for other financial constraints (Mosavi et al., 2020; Thompson et al., 2020).

A research gap exists regarding whether the tone of disclosure affects long-term earnings and investors' sentiment (Pouryousof et al., 2022). It was studied by incorporating median absolute deviation in exposures and comparing it with the tone of disclosures (Elshandidy & Zeng, 2022). However, the long-term impact of tone on financial consequences has not been established (Lone & Bhat, 2022; Wu et al., 2021). The effect of disclosure tone on companies' financial performance and investors' attitudes will be examined over the duration of this research.

The impact of macroeconomic variables, including market fluctuations or new regulatory guidance, on machine learning model predictions is another largely unexplored area. Machine learning can predict FD without considering how external factors influence the results (Kim et al., 2020; Pirayesh & Sallan, 2024). By considering such factors, this research shall seek to improve the precision of machine learning models and give a more detailed picture of how disclosure tone affects financial anticipations.

2. Literature Review

The general limitations of traditional forecasting and the growing use of machine learning are addressed in the introduction, but this review takes a gap-oriented approach; the generalizations of the above theoretical lenses and their implications for disclosure tone are omitted, and instead, three theoretical lenses are presented that explain the why, rather than just the how, of disclosure tone carrying economic information. All these lenses make it worth examining the econometric and computational work that treats narrative tone as a predictor of a film's future financial performance, and these points provide context for the contribution of this present study in emerging markets like Pakistan.

Spence's (1973) signaling theory is the main theoretical basis for the use of disclosure tone in predictive models. Where information is asymmetric, managers have private information about the firm's prospects and choose linguistic cues in narrative disclosures to be informative to outside users. Consistency in the tone of the MD&A is systematically identified as having a significant relationship with future ROA, ROE, and financial risk in the study by Pouryousof, Nassirzadeh, and Askarany (2023), which establishes that the tone is an informative but potentially biased signal. A similar study conducted by Iqbal and Riaz (2021) also finds that managerial tone disclosure in Pakistani banks predicts future profitability, confirming that tone has more economic content than quantitative ratios. In a signaling sense, therefore, the use of positive, negative, and uncertainty words does not just signal the past, but also conditional expectations about future cash flows, risk exposures, and strategic commitments – machine learning models can be used to extract such expectations, given their ability to capture high-dimensional and non-linear relations between textual and numerical features.

The other lens is impression management theory, which suggests that managers can intentionally alter the tone of their narrative to influence stakeholder perceptions beyond what is supported by the underlying fundamentals (Merkl-Davies & Brennan, 2007). Beretta, Demartini, and Trucco (2019) show that the nature of the optimistic tone in an integrated report is correlated with ESG performance, both because it provides additional information and because of impression management motives. In some emerging-market contexts, however, managers are not just obfuscating but rather giving honest explanations, as found by Al Lawati, Hussainey, and Sagitova (2023), who find that a forward-looking tone in the chairman's statements is linked to a firm's actual performance. Diouf and Boiral (2021) report that tone management is a symbolic CSR disclosure strategy. Importance of the coexistence of signaling and impression management motives for predictive modeling: a well-trained machine learning model can learn the "equilibrium" tone–performance mapping that emerges in the system of truthful signals and strategic narratives, without needing to assume one-way causality, as a linear regression would.

A third strand is based on agency theory and behavioral finance. Linguistic cues may be used by managers under agency pressure to affect investor beliefs, and behavioral investors are less responsive to qualitative cues that are not fully priced in (Loughran & McDonald, 2020). This suggests that the independent contribution of tone-based variables should be added to the contribution of traditional ratios, especially in emerging markets where disclosure requirements are less consistent and information asymmetry is high (Tran, Tu, & Nguyen, 2021). To provide direct country-specific evidence supporting the present study's research design, Sufi, Hasan, and Hussainey (2024) examine the impact of narrative tone and governance indicators on firm performance using a sample of 125 non-financial firms in Pakistan for the period 2011-2020.

None of these are interchangeable lenses, and it is important to know when to use each one. Signaling theory posits that tone is informative: managers know more than do outsiders; managers communicate language that provides credible signals of future cash flows. The opposite is impression management, where tone may serve as strategic framing to influence perception; an upbeat report doesn't necessarily imply positive prospects. The two thus tug at the same words in opposite directions, and in any specific context, it is an empirical question who wins. Underlying both are agency theory and behavioral finance, respectively, to account for managers' incentives to distort disclosure and investors' time lags in adjusting prices to qualitative cues. In any real report, the credible signal and the strategic framing are combined, and that's the place for the current design: a linear model needs to have a single direction throughout the sample, while a non-linear learner can fit the net tone–performance relationship that arises after the two motives counterbalance.

The closest country-specific comparison is provided by Sufi, Hasan, and Hussainey (2024) as discussed below. They partner with a panel of 125 non-financial companies in Pakistan and find that machine-learning models outperform ratio-based measures in forecasting firm performance. There are three concrete aspects in which the present study differs from theirs. First, it includes multiple performance measures (ROA, ROE, and EPS). Second, it evaluates 4 different algorithmic learners (SVM, ANN, Random Forest, and LSTM), rather than a single family of learners. Third, it employs the full six–category Loughran–McDonald tone space, rather than governance-influenced sentiment, thereby allowing identification of the marginal effect of disclosure tone. It is important to note that the two studies are complementary: Sufi et al. have demonstrated that tone and governance both play a role in Pakistan, and this paper challenges the extent to which tone can be generalized across various outcomes and model classes.

However, there is no one-sided evidence, and it would be a one-sided review. In their survey of textual analysis, Loughran and McDonald (2016) warn that dictionary-based tone is a "noisy proxy," as many of the words in the dictionary are ambiguous in context, and the predictive power of measured sentiment is weak and unstable once standard controls are included. Does the tone in the text convey information or simply puffery? What do these tones predict about future firm value? Wu, Yao, and Guo (2021) ask these questions directly and conclude that the predictive value of textual tone for future firm value is weak and is subject to conditionality. The answer on the modeling side is that there is no evidence that the gains from more complex models over a well-specified linear baseline are large enough to conclude they are reliably better/smarter than simple models and are often sample-dependent. It is important to take these qualifications seriously, which is why this study reports accuracy, precision, recall, and F1 across four models rather than a single high rating.

Empirical work on tone–performance linkages has primarily examined developed markets and single-outcome settings, most often stock returns or bankruptcy, rather than modeling accounting-based performance indicators such as ROA, ROE, and EPS simultaneously. Acheampong and Elshandidy (2025) show that narrative disclosures improve analysts' forecast accuracy, while Mousa, Elamir, and Hussainey (2022) establish the utility of tone for machine-learning-based performance prediction. However, prior studies rarely bring together (i) multiple tone dimensions (positive, negative, uncertainty, litigious, superfluous, constraining), (ii) multiple accounting outcomes, and (iii) algorithmically diverse predictive models (SVM, ANN, Random Forest, LSTM) within a single emerging-market design.

The present study aims to fill this gap theoretically by drawing on the concepts of signaling and tone management, and methodologically by covering the entire tone variable space. Specifically in Pakistan, the disclosure practices of non-financial companies are less standardized than those of financial institutions, and disclosure quality varies across industries such as textiles, cement, pharmaceuticals, oil & gas, power, chemicals, automobiles, and fertilizers. Such heterogeneity makes tone information more valuable and more difficult to model with traditional econometric models; therefore, machine-learning models that can handle nonlinearity and interactions are well-suited for this context, as discussed by Sufi et al. (2024) and Iqbal and Riaz (2021). This theoretical and contextual background drives the following two hypotheses.
Building on prior research, the following hypothesis was developed:

H1: Machine learning models trained on financial ratios together with disclosure-tone variables will predict the financial performance indicators (ROA, ROE, and EPS) more accurately than *linear benchmarks, specifically OLS and logistic regression estimated on the same financial ratios without tone features.*

Predictive modeling with structured and unstructured data has enabled the analysis of vast amounts of information using machine learning (ML), transforming financial forecasting. Commonly used models comprise Random Forest (RF), an ensemble learning method that combines multiple decision trees to increase accuracy (Magloire, 2025). Further, Support Vector Machines (SVMs), which classify and regress by optimizing hyperplanes (Bhandary & Ghosh, 2025), as well as deep learning (DL) models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs). Fozap (2025) states that these are effective for time-series prediction and sentiment analysis. Stock price prediction, credit risk assessment, and financial performance forecasting are areas where these models have been extensively applied, outperforming traditional models in these capabilities.

The accuracy of ML models improves when financial variables and disclosure tone are used together. Magloire (2025) noted that sentiment scores derived from earnings reports were used in models that achieved lower prediction error. Similarly, accurate bankruptcy forecasting can be achieved by combining financial ratios with tone analysis of financial disclosures (Fozap, 2025). These results support the hypothesis that sentiment in text is more useful than financial data alone. Based on previous studies, the following hypotheses were made:

H3: The incorporation of positive tone, negative tone, and uncertainty tone in financial ratios with machine learning will improve the prediction accuracy of financial performance.

The qualitative information contained in the financial disclosures can be quite important to predicting firm performance. Three key variables are analyzed in the disclosure sentiment: Positive, Negative, and Uncertain Tone. The positive tone indicates optimism, potential growth, and confidence in management's plans, suggesting positive financial condition (Loughran & McDonald, 2020; Mousa et al., 2022). A negative tone signals risks, financial problems, and poor performance, prompting defensive responses from investors (Luo & Zhou, 2020). The tone is uncertain, with some propositions being ambiguous, thereby increasing volatility in market perception and financial modeling (Iqbal & Riaz, 2021). These sentiment categories will be measured by applying NLP techniques and sentiment analysis tools, such as Loughran and McDonald's Sentiment Dictionary (Loughran & McDonald, 2020).

Three additional categories, which are generally not mentioned in the earlier tone literature, are also used for the analysis. Some of the tone words related to litigation, claims, and liability are counted, and the number of such words increases when a firm is involved in litigation or under compliance risk (Loughran & McDonald, 2011). Tone helps to pinpoint words that convey restriction on management decisions, such as "must," "covenant," and "restrict," and has been associated with tighter financing constraints (Bodnaruk, Loughran & McDonald, 2015). When a high percentage of superfluous tone is used, it indicates redundant or filler language; with high percentages of superfluous tone, disclosure is less informative and is considered an indicator of obfuscation in the readability literature. The firm-year measure is used for each category and equals the number of matched words divided by the number of words in the firm-year, as was done for the positive, negative, and uncertainty categories. With all 6 included, the models can consider defensive and evasive language, as well as simple optimism and pessimism.

Lack of precise wording in financial documents raises risks associated with financial modeling and changes the anticipated path of a stock's price (Kang et al., 2020). Investors struggle to understand vague or noncommittal statements, leading to muddled expectations about an organization's financial future and further complicated by ambiguous disclosures. (Loughran and

McDonald, 2020). Heightened volatility in erratic investment behavior is how financial markets react to uncertainty (Nguyen et al., 2023). Empirical evidence indicates that applying uncertainty metrics to financial predictions using machine learning during times of turbulence enhances accuracy (Jain et al., 2021).

The tone of disclosure affects financial performance in various ways across developed and emerging markets (Iqbal & Riaz, 2021). Specifically, although in developed markets transparency and regulation characteristically cut off the uncertainties associated with tone, leaving fewer gaps about the accuracy of predictions (Mousa et al., 2022), in emerging markets, for instance in Pakistan, the weaknesses in corporate governance and asymmetrical information systems make financial forecasts extremely difficult, thus increasing reliance on tone analysis (Nguyen et al., 2023). The application of machine learning and sentiment analysis to text-based financial data improves analysis in both markets (Jain et al., 2021). It is posited that using text-based sentiment analysis to identify non-disclosed financial information improves accuracy, given inconsistent disclosure practices in some markets.

It is important that this study be conducted because it addresses critical gaps in the literature on predicting financial performance using machine learning techniques and disclosure tone analysis, especially in emerging markets such as Pakistan. This research could be relevant to three types of stakeholders: investors, analysts, and policymakers. Most prediction models of this kind rely exclusively on quantitative information, like financial ratios, and ignore qualitative information, such as statements made by the firm itself. The model presented here integrates both financial numbers and the tone of the narrative text to predict corporate results at a finer grain. Especially in emerging markets, reporting and transparency are not universal, and the information gap, lack of disclosure, and price volatility only exacerbate the issue. Many of the models currently on the market were developed for a more mature market and may not be suitable for this one. The models developed here are fitted to emerging-market conditions and can be transferred to other markets with comparable characteristics; the goal is a forecasting approach suited to local data rather than one "taken from the shelf" from developed markets.

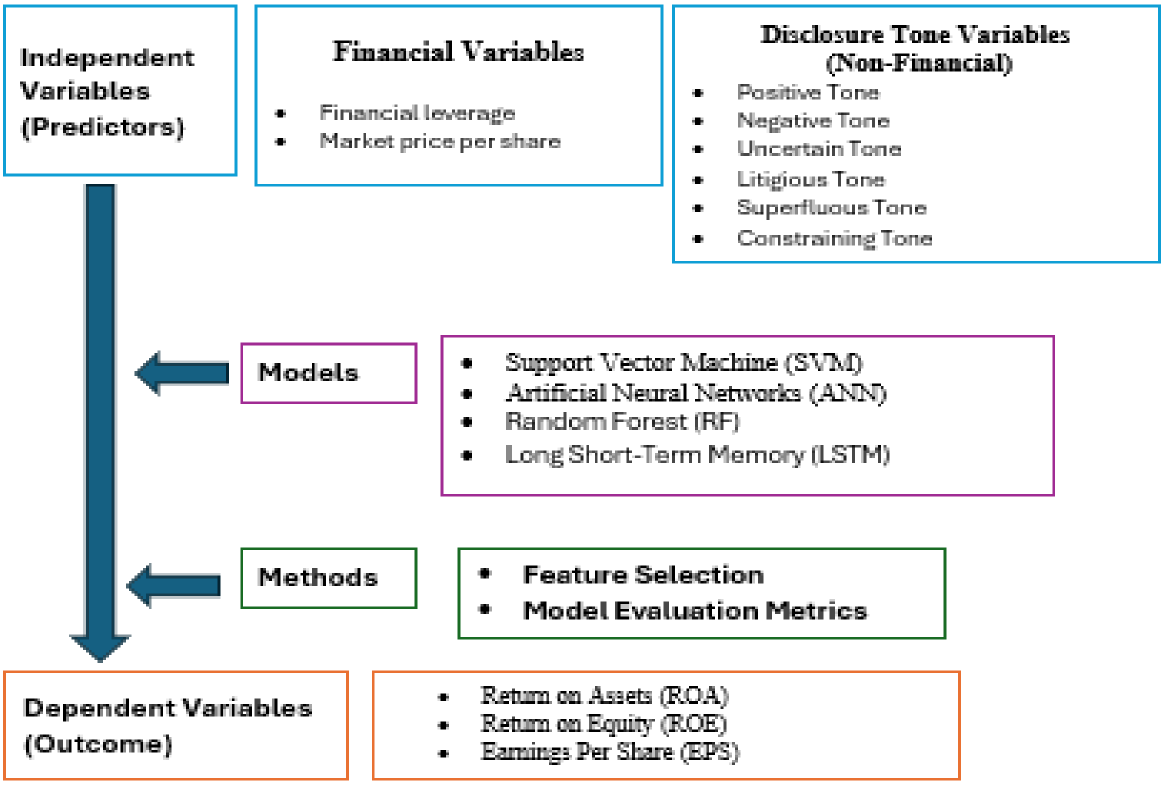
The study provides investors and analysts with a more nuanced instrument to decipher a company's financial condition and make more informed investment decisions. The models are weighted by the language management selects, thus providing a glimpse of the firm's strategic agenda and the reporting approach used, and thereby contributing to the evaluation of the firm's risk. The results also hold implications for corporate management, reflecting the impact of disclosure decisions on investor expectations and, in turn, market values.

Also, deep learning techniques, including long short-term memory (LSTM) networks, have been used to model trends in financial data to predict stock prices, generate revenue growth, and analyze corporate profitability (Sharifani, 2023; Ashraf et al., 2020). Previous studies show that deep learning models typically outperform traditional financial forecasting models, such as ARIMA because they can capture both short-term and long-term dependencies in data (Bola et al., 2021; Liu et al., 2023). For the Pakistani market in particular, financial datasets are likely to be sparse and exhibit disparities, making these advanced models invaluable, as they can make predictions using incomplete and noisy datasets (Koosha Sharifani, 2023; Fedorova, 2022).

2.1 Conceptual Framework

The conceptual framework of this study is based on three dimensions: quantitative financial ratios, qualitative disclosure tone variables, and machine-learning predictive modeling. In the first stage, financial performance is measured using three proxies: Return on Assets (ROA), Return on Equity (ROE), and Earnings per Share (EPS). At the second stage, the independent variables are categorized into two groups. The first group comprises traditional financial ratios, such as financial leverage and market price per share, that represent the quantitative dimension of firm performance. The second group consists of disclosure tone variables extracted from the annual reports of non-financial firms listed on the Pakistan Stock Exchange. These include positive tone ratio, negative tone ratio, uncertainty tone ratio, litigious tone ratio, superfluous tone ratio, and constraining tone ratio, all measured using Natural Language Processing techniques applied to textual data. The framework explicitly relies on the notions of signaling theory (Pouryousof et al., 2023; Spence, 1973) and impression management theory (Al Lawati et al., 2023; Beretta et al., 2019) that explain the managerial choice of words in narrative disclosures and the strategic dimension of tone, respectively. These two theoretical pillars explain why some ML models can leverage both quantitative ratios and NLP-derived tone features to extract economically meaningful signals that linear financial-ratio models miss.

At the third level, the features are trained using four machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM) network. The conceptual linkage is based on the idea that combining narrative sentiment and quantitative financial information yields a more complete information set for predictive modeling, especially in emerging-market settings with inconsistent disclosure practices and information asymmetry. This suggests that the tone and wording of corporate disclosures convey forward-looking information about management's expectations and strategic direction, thereby adding to the predictive power of conventional financial statements. Importantly, the use of machine learning is described here as a computational tool that embodies constructs derived from these theories, but not as an alternative to theoretical reasoning. The theoretical contribution of the paper is to show that there is an economic content (signaling) in the narrative tone that remains preserved in the face of managerial bias (impression management) and that can be recovered more efficiently using non-linear ensemble learners than linear econometric estimators.



3. Research Methodology

This study adopted a longitudinal secondary-data research design and applied the post-data-acquisition (post-collection, cleaned, and pre-processed) data to firm-year observations of 40 non-financial firms listed on the Pakistan Stock Exchange (PSX) for 2019–2022, resulting in 164 firm-year observations. The sampled firms cover the key non-financial industries of the PSX, such as textile composite, cement, pharmaceuticals, oil & gas exploration and marketing, power generation and distribution, chemicals and fertilizers, automobile assemblers and parts, food and personal care, and technology and communication. The composition of the sectors involved in the analysis accounts for sector heterogeneity in disclosure quality and profitability drivers, thereby enhancing the external validity of the predictive models. The financial predictors (revenue ratios, profitability ratios, asset turnover, financial leverage, and market price per share) and disclosure-tone variables (tone ratios measuring positive, negative, uncertainty, litigious, superfluous, and constraining) derived from the NLP are included on the left-hand side (X), while the performance outcomes (ROA, ROE, and EPS) are placed on the right-hand side (Y).

The study employed purposive sampling, and only non-financial companies listed on the PSX with available financial disclosure information were retained. Because of this restriction, it was possible to maintain uniformity of the training data and to restrict it to firms that did not report with significant time gaps between reports. The financial institutions sector has been excluded because the corporate financial structures of these firms, the level and scope of regulatory scrutiny, and the disclosure requirements of these institutions differ from those of the rest of the market. There are also different accounting treatments for these entities, such as IFRS

9 on financial instruments, making ratios and disclosures difficult to compare with those of non-financial companies. However, the panel is short, with only 40 firms, so standard dynamic panel estimators (such as the system GMM, which requires N > T and additional moment conditions) are not applicable. As a robustness check, static panel approaches (fixed-effects and random-effects regression) were assessed, and their suitability for the tone features derived from NLP to act as a proxy for informational content proved to be inappropriate, as the information content of these tone features interacts non-linearly with the financial ratios (Mousa et al., 2022; Sufi et al., 2024).

Machine-learning algorithms, specifically Random Forest, Support Vector Machine, Artificial Neural Network, and Long Short-Term Memory, were therefore preferred because they (i) natively capture non-linearities and interactions among high-dimensional tone and ratio features, (ii) are robust to moderate sample sizes through ensembling and regularisation, and (iii) provide variable-importance diagnostics that correspond conceptually to the signaling content of each tone dimension. Financial-ratio predictors (financial leverage and market price per share) were included because prior evidence shows that leverage conditions managerial incentives to disclose risk-related language (Hassan et al., 2022), and MPS reflects real-time market pricing of firm information that can correlate with disclosure tone (Pouryousof et al., 2023). It is also important to note that machine learning inherits certain assumptions; for example, a support vector machine with a linear kernel still fits a linear boundary and treats firm-year observations as independent, much as OLS does. A radial-basis kernel was therefore used for the SVM, and the tree-based and recurrent learners were included precisely because they relax the linearity and independence assumptions that constrain the simpler specifications.

3.1 Measurement of Disclosure Tone:

The sentiment and language of the reports' disclosures were analyzed using Natural Language Processing (NLP). Data for non-financial companies listed on the Pakistan Stock Exchange (PSX) were obtained from their financial statements, earnings announcements, and regulatory filings via web scraping and data repositories to maximize coverage. Each year of the annual report (MD&A, Directors' Report, and the Chairman's Statement) was broken down and presented in clean text for each firm-year. Tone was then compared with the financial dictionary of Loughran and McDonald (2011, 2020), which provides a taxonomy of words into categories of positive, negative, uncertainty, litigious, superfluous, and constraining. The six categories correspond to six tone ratios, which are machine-learning features: Category $\text{Ratio}_{i,t}$ is the number of words in a category for firm i in year t, divided by the number of words in firm i's disclosure in year t. For the construction, tone is a leading rather than a contemporaneous variable, with ROA, ROE, and EPS in the year t+1 modeled as functions of tone ratios in year t and financial ratios in year t.

Next, the collected texts underwent preprocessing. This involved breaking down the text into individual words or phrases (tokenization), removing common non-informative words such as "the," "is," and "and" (stopword removal), and reducing words to their basic root forms, such as converting "profits" to "profit" (stemming and lemmatization). Specifically, PDFs of annual reports were converted to text using PDFMiner, de-hyphenated, lower-cased, and stripped of tables, headers, and page numbers. Tokenization used the NLTK word tokenizer, stop words were removed using a finance-aware stop-list (excluding negation terms such as "not" and "no" which carry sentiment), and the Porter stemmer was applied. Numeric strings and punctuation were discarded. Token counts per tone category and total token counts were computed per firm-year to construct the six tone ratios. A sample of Loughran–McDonald words matched in the corpus is provided in Appendix A.

Tone itself was measured with the lexicon-based approach described above. Each word was matched against the Loughran and McDonald (2011, 2020) finance dictionary, and the six category ratios were computed directly from those counts. A separate machine-learning or transformer-based sentiment classifier (such as BERT) was not trained because the lexicon provides a transparent, replicable mapping from words to tone categories and avoids the labeled-data requirement imposed by supervised classifiers. The machine-learning models introduced later in this paper, including Random Forest and LSTM, are predictive models that take lexicon-derived tone ratios and financial ratios as inputs to forecast ROA, ROE, and EPS; they are not used to score sentiment. The net disclosure tone reported in the descriptive statistics is the count of positive words minus the count of negative words, divided by the total number of words.

The analysis, therefore, drew on several tone measures: positive disclosure tone, the proportion of positive words; negative disclosure tone, the proportion of negative words; Net Disclosure Tone, the difference between the positive and negative proportions; and an Uncertainty Score counting words that signal uncertainty, such as "may," "might," or "possibly." Each of these entered the forecasting models as a predictor of financial performance.

3.2 Techniques and Procedures: Data Collection, Preprocessing, Model Training, and Evaluation:

Primary data could not be collected directly from the listed companies due to confidentiality and privacy restrictions, so the study relies primarily on secondary sources. Firm-level financial statements and related information came from the PSX. A second source, the Securities and Exchange Commission of Pakistan (SECP), supplied regulatory filings, compliance reports, and annual disclosures, all of which bear on governance and on the accuracy of the data. To situate each firm within the wider economy, macroeconomic indicators, including inflation, interest rates, exchange rates, and GDP growth, were taken from the State Bank of Pakistan (SBP); these place firm performance in relation to prevailing conditions. Finally, established financial outlets, Brecorder.com, Pakistan Gulf Economist, Yahoo Finance, and The News, were used to collect supplementary information on fiscal policy, tax reform, political events, and market sentiment, any of which can move corporate profits.

Data collection was thus completed; hence, pre-processing was an important step to ensure the accuracy and usability of the data set for predictive modeling. If financial data contained missing values, duplicate values, or variables with dissimilar scales, they were considered unsuitable. The models' performance was impaired when the variable scales differed. Imputation schemes were used to deal with missing data, primarily due to lack of reporting and/or corporate restructuring (mean substitution or k-nearest neighbors (KNN)). Conversely, records with a long string of missing values would be rejected to avoid affecting the selected estimate. This prevented data manipulation, and data redundancy was gradually eliminated, primarily during data merging. Standardization and normalization were used to address issues where financial variables were on different scales, enabling comparison between features. For example, the revenue was rescaled to the range [0, 1] or standardized to have a mean of 0 and a variance of 1 for the models used. To identify the most predictive variables and eliminate multicollinearity by removing highly correlated features, feature selection was performed using methods such as random forest importance rankings. Given the time-series nature of financial data, lagged variables were created to capture temporal dependencies and historical performance trends, which are often influential in forecasting future outcomes. To address the high dispersion evident in the descriptive statistics (notably financial leverage, MPS, EPS, and ROE, which contain extreme negative and positive tails), all continuous financial variables were winsorized at the 1st and 99th percentiles prior to model training. Winsorization was preferred to deletion because it retains sample size while attenuating the leverage of outlying firm-years on non-linear learners such as Random Forest and LSTM. Comparability across firms of different scales was further ensured through feature-wise z-score standardization for SVM and ANN, and min-max scaling for LSTM, while Random Forest was trained on the winsorized but unscaled features (since tree-based learners are scale-invariant). Regarding lagged variables, one-period lags of ROA, ROE, EPS and the tone ratios were introduced to model temporal dependence: the choice of a single lag was based on (i) the short panel length ($T = 4$), which precludes longer lag structures, (ii) the typical annual reporting cycle of Pakistani non-financial firms, and (iii) partial autocorrelation function (PACF) inspection, which indicated negligible second-order autocorrelation after controlling for the first-order lag. The inclusion of lagged outcomes also mitigates omitted-variable bias in the ML specification.

Financial ratios and macroeconomic variables were used to predict the firm's performance by using trained models. Advanced machine learning algorithms, such as Random Forest and Long Short-Term Memory (LSTM) networks, were employed in the analyses. Random Forest is well-suited to the high-dimensional feature set: as an ensemble of decision trees, it mitigates overfitting and captures the complex interactions between financial ratios and performance. LSTM, a form of Recurrent Neural Network (RNN), was included for its complementary strength: its ability to handle sequential structure in time-series data. The LSTM models were also well-suited to extracting the long-term historical patterns from the data and making out-of-sample predictions for other finance performance measures, such as net profit margins. Both models were exhaustively trained, and hyperparameters were optimized to maximize prediction accuracy and minimize the risk of overfitting.

Given the small sample size, the models were not assessed with a single train–test split, as this would leave too few observations for reliable evaluation. Instead, they were evaluated with five-fold cross-validation: the 164 firm-year observations were partitioned into five folds, each model was trained on four folds and tested on the held-out fold, and the procedure was repeated so that every observation served once for testing. The accuracy, precision, recall, and F1 scores reported below are averages across the five folds. Folds were drawn so that high and low performers appeared in each fold, and all preprocessing, including winsorization and scaling, was applied to the training folds only to avoid leakage into the test fold. Hyperparameters were tuned with an inner cross-validation loop on the training folds.

One caveat applies specifically to the LSTM. Recurrent networks are built to learn from long sequences, and with only four annual observations per firm, the time dimension here is far shorter than such models normally need. The LSTM was kept in the comparison as a benchmark rather than a serious contender, precisely to show what happens when a sequence model is asked to work without sufficient sequence data. Its weaker results below should be read in that light, and not as evidence that recurrent models have nothing to offer on richer panels.

Model evaluation consisted of a set of extensive performance metrics specific to the nature of predictions in the tasks. Where the forecast was framed as a regression, fit was judged mainly by R^2, the share of variance the model explains, alongside Mean Squared Error (MSE), the mean of the squared gaps between predicted and actual values. Where the task was framed as classification, sorting firms into financial-health tiers, the relevant measures were precision, recall, and F1, with balanced accuracy added because the classes were uneven. Taken together, these metrics tested whether the predictions held up across different market conditions rather than in a single sample. The resulting framework, built from secondary data, careful preprocessing, and the machine-learning models above, improved predictive accuracy and, just as usefully, gave a clearer reading of the financial dynamics of firms in Pakistan's emerging market.

4 Result analysis

4.1 Descriptive Statistics

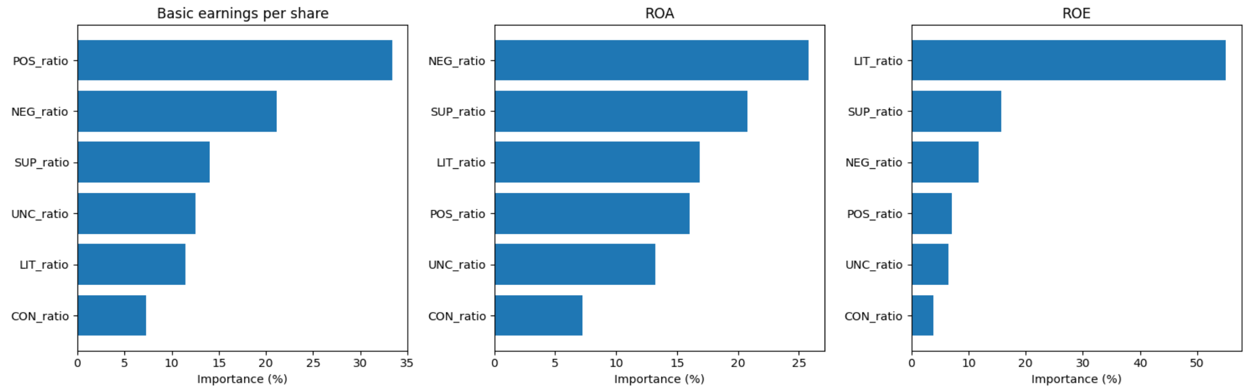
Table 1: Descriptive Statistics

Ratios	Mean	SD	Min	25%	Median	75%	Max
Financial leverage	1.3047	10.6448	-131.81	1.468	1.918	2.5522	22.552
Market price per share	895.95	3429.36	0	20.58	74.635	269.26	25000
NEG Ratio	0.0067	0.0032	0	0.004	0.0072	0.0089	0.0175
POS Ratio	0.0079	0.0032	0	0.007	0.0082	0.0097	0.0253
UNC Ratio	0.0037	0.0015	0	0.003	0.0038	0.0047	0.0079
LIT Ratio	0.0037	0.0019	0	0.003	0.0036	0.0041	0.0175
SUP Ratio	0.0027	0.0018	0	0.002	0.0024	0.0032	0.0125
CON Ratio	0.0002	0.0002	0	0	0.0001	0.0002	0.0012
Basic earnings per share	69.273	204.949	-31.639	0	7.5914	32.097	1529.7
ROA	6.9311	12.8869	-31.724	0	4.8769	12.385	57.966
ROE	12.046	52.4818	-402.43	0.326	11.056	27.376	180.84

The descriptive statistics for the financial and textual variables in Table 1 provide an initial understanding of the distribution of observations across firm-years. Financial leverage varies widely across firms, with a mean of 1.30 and a very high standard deviation of 10.64. The minimum value of -131.81 is a red flag for extreme outliers or accounting errors and should be winsorized before modeling. Market price per share (MPS) has a similar distribution, with a mean of 895.95, a standard deviation of 3,429.36, and a range of 0 to 25,000 shares. It is difficult to make cross-firm comparisons because firms vary widely in size and valuation. The tone variables are average, leaning towards optimism. The positive ratio is the most frequent, at 0.0079, slightly ahead of the negative ratio at 0.0067. The uncertainty ratio and litigious (LIT) ratios are around 0.0037, indicating some moderate occurrence of ambiguous or defensive language, while the superfluous (SUP) and constraining (CON) ratios are correspondingly lower, at 0.0027 and 0.0002, indicating that the use of filler and restrictive language is comparatively rare. Although the tone remains positive, these subtle changes in tone frequency could convey messages about future performance. The profitability is not uniform for the sample: the average basic earnings per share (EPS) is 69.27, but the standard deviation is 204.95 – the ROA and ROE have noticeable skewness. ROA averages 6.93 and runs from -31.72 to 57.96, while ROE averages 12.05 over a much wider span, from -402.43 to 180.84. The variance in ROE suggests significant disparities in efficiency and the creation of shareholder wealth, likely related to management, industry, or disclosure quality.

4.1.1 Importance of Tone-Based Textual Variables

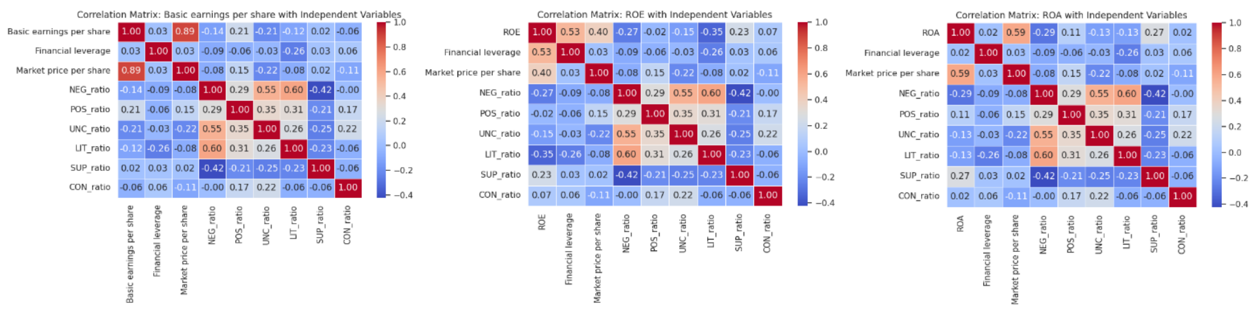
Figure 1: Importance of Tone-Based Textual Variables in Predicting EPS, ROA, and ROE Using Machine Learning Models



The Variable Importance plots visualize the relative importance of tone features for predicting Basic Earnings per Share (EPS), Return on Assets (ROA), and Return on Equity (ROE). The positive tone ratio (POS_ratio) accounts for more than 30% of the model's weight, consistent with the notion that positive language carries more weight in earnings. The next most important is the negative tone ratio (NEG_ratio), which indicates the ratio of negative to positive sentences in terms of sentiment polarity. Other factors, such as verbosity, uncertainty, and litigious wording, make smaller contributions, suggesting that these elements are present and still signal, but not decisively. For ROA, the top predictor is NEG_ratio, which is associated with lower asset efficiency, with the ranking changing around 24% of the time. Next, the SUP_ratio and LIT_ratio suggest that wordiness and legalistic language may indicate inefficient use of assets, and POS_ratio is not as important here as it was in EPS. ROE is the most extreme case, with the model dominated by a single factor, LIT_ratio, which alone accounts for more than 50% of the importance and is the most important driver of expected returns for shareholders.

4.2 Correlation Analysis:

Figure 2: ***Correlation Matrix Showing Relationships Between Independent Variables: Basic EPS, ROA, ROE***



The correlation matrix shows the relationship between the basic earnings per share (EPS) and financial and tone measures. The most significant correlation is between EPS and market price per share (r = 0.89): positive EPS is associated with a higher market price per share, which is not surprising, as the market reflects the firm's performance. A moderate positive correlation (r = 0.21) exists between the words on the tone (POS_ratio) and the earnings (EPS). For negative and uncertainty tones, this is reversed: NEG_ratio and UNC_ratio are negatively related to EPS (r = -0.14 and r = -0.21), suggesting pessimistic or hedged language and poorer results. Litigious tone (LIT_ratio) has a lower negative correlation (r = -0.12), but it is again the relationship between defensive and risk-driven language and lower profitability. The other tone features of financial leverage, CON_ratio, and SUP_ratio are only moderately or weakly correlated with EPS and therefore do not have a direct impact on the earnings in this sample. Combined, the matrix indicates that it contains information about earnings, either in the form of a quantitative signal (market price) or a qualitative signal (disclosure tone).

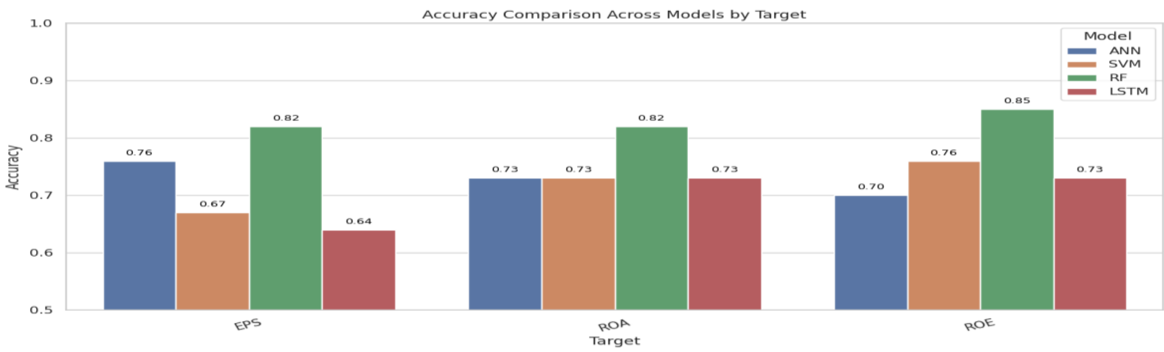
4.3 Precision Comparison of ANN, SVM, RF and LSTM Models

Figure 3: ***Precision Comparison of ANN, SVM, RF, and LSTM Models Across EPS, ROA, and ROE Targets***

To compare the precision of the models, the ANN, SVM, RF, and LSTM models were compared for the three targets: EPS, ROA, and ROE (Figure 3). Random Forest outperforms across all targets, achieving precisions of 0.89 (for EPS), 0.81 (for ROA), and 0.93 (for ROE), and it rarely misclassifies a firm as a strong performer. ANN is neck and neck on EPS, but less reliable on the tougher targets, at 0.88, 0.71 for ROA, and 0.70 for ROE, as its positive calls get weaker on more volatile results. Overall, SVM falls in the middle at 0.74 for EPS, 0.68 for ROA, and 0.78 for ROE, resulting in more false positives than RF. LSTM is the weakest in precision, worse at ROA (0.68), and slightly better at EPS (0.75) and ROE (0.74). It can process sequential data, although it is not as competitive as the other models in financial classification precision.

4.4 Accuracy Comparison of ANN, SVM, RF, and LSTM Models

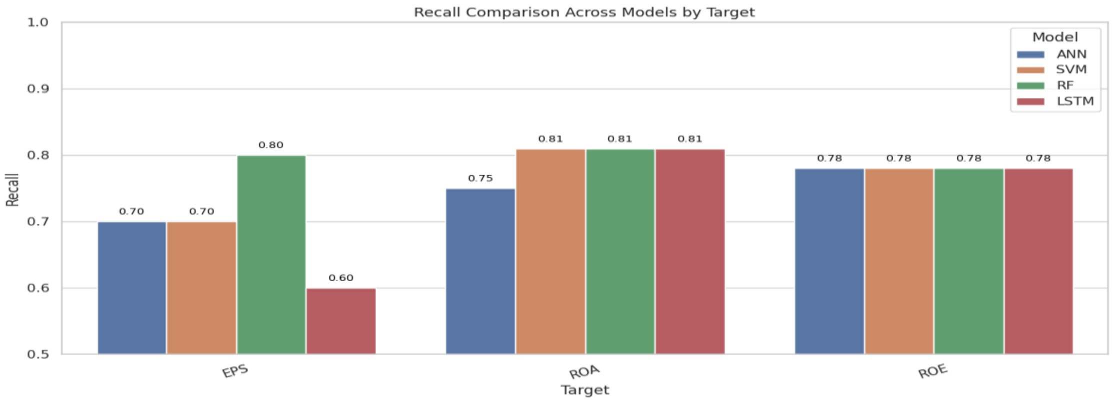
Figure 4: ***Accuracy Comparison of ANN, SVM, RF, and LSTM Models Across EPS, ROA, and ROE Targets***



In Figure 4, the four models are compared with respect to overall accuracy for EPS, ROA, and ROE. Random Forest is the clear winner on EPS (0.82), ROA (0.82), and ROE (0.85), and is more consistent than the other models in distinguishing the weak from the strong models. ANN is in line with the respectable values on EPS (0.76) and ROA (0.73), but falls short on ROE (0.70), which is useful for returns and asset performance and less reliable for equity. When it comes to the tougher targets, SVM and LSTM are even, with both having ROA and ROE of 0.73, but both fail on EPS at 0.67 and 0.64, respectively.

4.4.1 Recall Comparison of ANN, SVM, RF, and LSTM Models

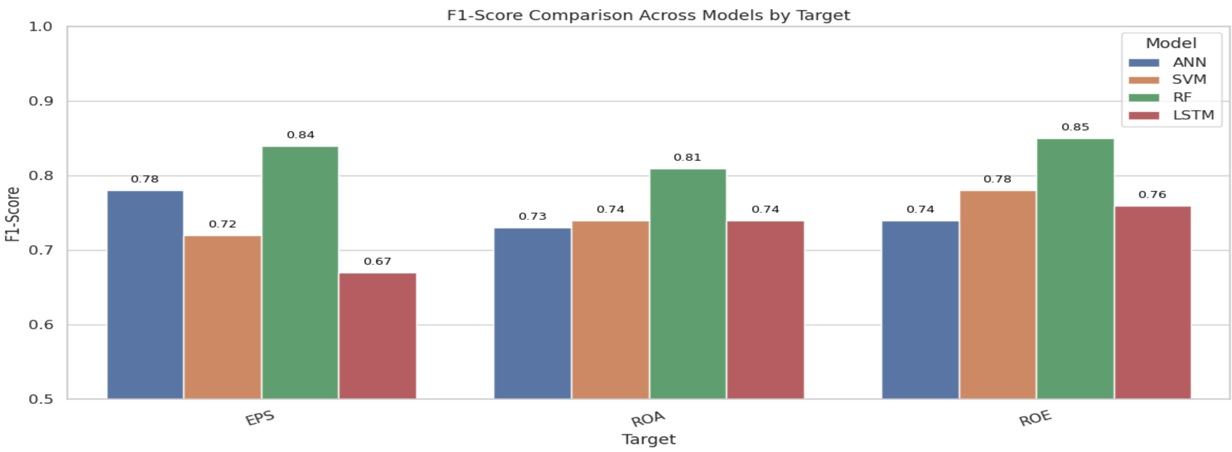
Figure 5: ***Recall Comparison of ANN, SVM, RF, and LSTM Models Across EPS, ROA, and ROE Targets***



Recall measures the percentage of the true strong performers each model can capture, again across EPS, ROA, and ROE. On EPS, Random Forest achieves the highest recall for true high performers at 0.80, while ANN and SVM are tied at 0.70, and LSTM lags at 0.60. The difference is smaller for ROA, with scores of 0.81 for both RF and SVM and 0.75 for LSTM, whereas ANN remains lower. The four are not distinguishable on ROE, with each at 0.78, indicating similar sensitivity in their perception of what they are after: shareholder returns.

4.5 F1-Score Comparison of ANN, SVM, RF, and LSTM Models

Figure 6: ***F1-Score Comparison of ANN, SVM, RF, and LSTM Models Across EPS, ROA, and ROE Targets***



The F1-score consolidates precision and recall into a single value, providing a fair assessment of each model across EPS, ROA, and ROE. For EPS, the best model is the Random

Forest model with a score of 0.84, followed by the ANN model with a score of 0.78 and the SVM model with a score of 0.72, the LSTM model with a score of 0.67, and the remaining models are less balanced in performance. The SVM, ANN, and LSTM models are tightly grouped around 0.73–0.74, which are competitive but less effective in identifying their asset-based performers. RF is still ahead of the pack on ROA at 0.81. The SVM, ANN, and LSTM models are grouped together around 0.73-0.74, close but not quite as effective at identifying their asset-based performers as RF's 0.81. On ROE, RF achieves its highest score of 0.85 and is the most reliable model for equity classification, while SVM (0.78), LSTM (0.76), and ANN (0.74) are somewhat similar but differ in their trade-offs between precision and recall.

The findings are consistent with previous studies on the forecastability of disclosure tone. According to Li (2010) and Loughran and McDonald (2011), narrative sentiment conveys forward-looking information that goes beyond what is available in financial ratios. The strong positive association with the prediction of Earnings per Share (EPS) is similar to that of Alalwani & Mousa (2020), who associated positive disclosure with increased profitability and investor confidence. The negative tone–Return on Assets (ROA) relationship identified here is consistent with Mousa et al. (2022), in which risky or protective language serves as a pre-warning of a decline in efficiency. The strong pull of positive tone on EPS and of negative tone on ROA is consistent with managers' word choices conveying private information about future profitability and asset use (Pouryousof et al., 2023; Iqbal & Riaz, 2021). The other direction is toward impression management, as the legalistic language can be used to cue the risk on decisions that involve equity financing; this type of language is picked up as a usable cue by the machine learning ensembles that are able to model it (Al Lawati et al., 2023; Beretta et al., 2019). Overall, the evidence supports the paper's main argument, namely, that tone carries economic significance, has a theoretical basis, and that non-linear learners are the appropriate tool for extracting it from emerging-market disclosures (Sufi et al., 2024; Tran et al., 2021).

The Random Forest model's performance relative to the other models further supports research on sentiment-based financial prediction ensembles (Feldman et al., 2010; Nassirtoussi et al., 2014). SVM and ANN achieved only moderate accuracy, and their lack of success with richer linguistic features is like the results reported by Kogan et al. (2009): Linear classifiers are not very successful in handling the semantic complexity of financial narratives. Interestingly, LSTM performed worse, which aligns with the work of Ranta et al. (2023), who advocated hybrid models to better leverage sentiment in financial modeling. Broadly speaking, the message is that disclosure tone, in addition to the financial information, enhances predictive accuracy and helps reduce the gaps in predictive power created by quantitative-only forecasting, particularly in emerging markets like Pakistan.

The LSTM's poor performance is primarily because it has only four annual observations per firm; the network can only learn from the limited number of observations it is designed to learn, and if it has too few, it will overfit the examples it does see. There are two other similar factors that are likely to exacerbate this. The LSTM is the most parameterized of the four models, which is a poor fit for 164 observations, and recurrent networks are very sensitive to hyperparameters like the number of hidden units and the learning rate – there is not much room to tune them with a short panel. The reason the LSTM doesn't work here is the same as why ensembles of shallow trees work better: they are relatively immune to small samples and don't require time depth. In practice, the choice of algorithm should depend on the nature of the data: if the data is short and structured as a yearly panel, a recurrent network is the wrong choice, even if it works well with long data.

5. Conclusion and Implications

This study aimed to test whether the disclosure tone of the positive, negative, and uncertain tone can predict the financial performance (ROA, ROE, and EPS) of non-financial firms of Pakistan in comparison with traditional financial models, and whether the positive, negative, and uncertain tone of the disclosure has predictive value in the context of an emerging market. It does. The incorporation of tone improved predictive accuracy for several machine-learning algorithms, with Random Forest being the most consistent. Positive forecasts were tracked by sharper forecasts, and negative and uncertain tones were expressed more contextually. The pattern reinforces the notion that a story in a report may lead to a firm's financial results before they are officially recognized. The implications extend beyond accuracy in predicting. Their call to mind the importance of transparency and accountability within companies, and to promote the harmonization and monitoring of disclosure practices. In Pakistan, regulators may mandate improved, more balanced tone to reduce information asymmetry; lenders and investors may incorporate tone into credit assessment and investment decisions; and lenders and investors could even incorporate tone into macroeconomic forecasting.

The results are relevant for a few market groups in emerging markets. The resulting increase in information, when disclosure tone is added to the machine learning model, provides another dimension that investors and analysts can use to better assess a company's financial health. Based on the results of tone-augmented models, portfolio managers and institutional investors might consider tightening the filters they impose on non-financial firms trading on the Pakistan Stock Exchange when using Random Forest. The findings reinforce that — for managers and boards — the quality of an annual report is not a veneer: it has been demonstrated here that Return on Assets, Return on Equity, and Earnings per Share can all be tracked by the tone of the annual report, so a more clearly worded, stable report can win investors more confidence — and a higher valuation. Firmer guidance on narrative reporting is supported by the evidence, as stated by the Securities and Exchange Commission of Pakistan (SECP). For fintech and analytics companies, the hybrid method described here—a natural-language approach combined with traditional financial analysis—can serve as a template for developing commercial tools that enhance risk and credit assessment in situations where information is limited.

There are definite boundaries on the study. It included only non-financial companies and excluded other industries such as manufacturing and telecom. It also relied on English-language stories, which might not capture the cultural and contextual nuances of disclosure in a linguistically diverse country such as Pakistan, and the absence of a localized sentiment dictionary further complicates interpretation. A broader range of industries should be used in the future to observe the generalizability of tone-based forecasting. In addition, there is a need for a sentiment lexicon for Urdu and regional languages, as these are more likely to reflect the context of Pakistani disclosures, and a bilingual or hybrid approach would facilitate interpretation and enhance accuracy. To sum up, tone analysis is a valuable tool for conventional forecasting, especially in situations where information is limited. Combining linguistic analysis with quantitative information provides a more comprehensive view of firms' performance, and more transparent disclosure, supplemented by sentiment indicators designed in the local language, would aid financial decision-making and regulatory monitoring in markets such as Pakistan.

The size and timing of the sample are the most important limitations, and they bound what can be claimed. The 40 firms and 164 firm-year observations from 2019 to 2022 are a small number for training machine learning models, particularly the LSTM. The four-year period is too brief to distinguish a true tone–performance relationship from a pattern specific to the period. The tone of the disclosure and the realized returns are unusual because the window spans both the COVID-19 period and an unusually long period of acute macroeconomic stress in Pakistan, which may not be applicable in calmer times. The Pakistan Stock Exchange lists over 500 non-financial firms, and a sample of 40 does not represent the population. Thus, the screening and credit-evaluation applications presented above are considered promising starting points rather than tools already available. They need to be confirmed by enlarging the panel, in terms of the number of firms, the number of years, and the inclusion of a non-crisis period.

The sample is also subject to survivorship bias. Only firms that consistently reported through 2019-2022 remained, and these tend to be weaker firms, as they were not included because they were delisted, suspended, or ceased reporting. The estimated association between tone and performance may appear cleaner, as there are underrepresented, struggling firms. A design that records the firm's entries and exits over the years would provide a more realistic and less rosy picture.

Firms were also grouped by industry rather than by industry model. What is considered acceptable disclosure in one company, such as a textile manufacturer, can differ from that in another, such as a technology or communications company, and the words expected to be used in one company in that manner may be accepted there without comment. The relationships reported are an average across industries because the number of observations is too small to fit reliable sector-specific models (164 firm-years). If manufacturing and services were divided in a larger sample, the finding that tone predicts performance differently by sector would be replicated.

Another constraint is disclosure credibility. If readers don't believe it, then tone carries no information, and a company with a shoddy track record can report positive performance with no actual change. In fact, this is what impression management theory predicts, given the capacity for using positive language in a non-genuine manner. It is important to note that the models learn an average tone-performance mapping across the sample and do not disambiguate credible signals from cheap talk, meaning that a positive-tone score from a firm with a track record of missed targets must be interpreted with caution. Later work should consider weighting tone based on disclosure quality and/or on the accuracy of the past forecasts.

Appendix A.

Illustrative Loughran–McDonald (2011, 2020) Finance Dictionary Words Matched in the Pakistani Non-Financial Annual-Report Corpus

Category | Illustrative matched words (partial list)

Positive | achieve, advance, beneficial, boost, delightful, efficient, enhance, excellent, favourable, gain, great, grew, improve, outperform, profitable, progress, strength, strong, succeed, surpass, upturn

Negative | adverse, bankruptcy, breach, decline, default, deficit, deteriorate, difficult, discontinue, disruption, downturn, erosion, failure, litigation, loss, shortfall, suffer, weakness, write-down

Uncertainty | almost, appear, approximate, assume, believe, conditional, contingent, depend, fluctuate, may, might, possibly, potentially, probable, risk, seem, tentative, uncertain, variable

Litigious | adjudicate, allegation, appeal, arbitration, claim, complaint, compliance, damages, enforcement, injunction, jurisdiction, law, lawsuit, liability, regulation, statute, verdict

Superfluous | additionally, absolutely, actually, basically, certainly, consequently, furthermore, indeed, obviously, particularly, primarily, relatively, respectively, substantially, undoubtedly

Constraining | bound, compel, constrain, covenant, limit, mandate, must, necessitate, oblige, prohibit, required, restrict, shall

Note: Word lists are indicative samples from the Loughran–McDonald dictionary; the complete classification was applied in the tone-ratio computation.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Fabrication/Falsification Statement

The author(s) declare that no data have been fabricated, falsified, or manipulated in this study.

Participant Consent

This study is based on secondary data obtained from publicly available sources and did not involve any human participants. Therefore, no participant consent was required, and all data were used in accordance with ethical standards for secondary data research.

Copyright and Licensing

References

Abimbola, B., de La Cal Marin, E., & Tan, Q. (2024). Enhancing legal sentiment analysis: A convolutional neural network–long short-term memory document-level model. Machine Learning and Knowledge Extraction, 6(2), 877–897. https://doi.org/10.3390/make6020050

Acheampong, A., & Elshandidy, T. (2025). Does sustainability disclosure improve analysts' forecast accuracy? Evidence from European banks. Financial Innovation, 11(1), 25. https://doi.org/10.1186/s40854-024-00693-5

Ahmed, S., Majeed, M. E., Thalassinos, E., & Thalassinos, Y. (2021). The impact of bank-specific and macro-economic factors on non-performing loans in the banking sector: Evidence from an emerging economy. Journal of Risk and Financial Management, 14(5), 217. https://doi.org/10.3390/jrfm14050217

Akinsulire, A. A., Idemudia, C., Okwandu, A. C., & Iwuanyanwu, O. (2024). Strategic planning and investment analysis for affordable housing: Enhancing viability and growth. Magna Scientia Advanced Research and Reviews, 11(2), 119–131. https://doi.org/10.26870/msarr.11.2.119

Al Lawati, H., Hussainey, K., & Sagitova, R. (2023). Forward-looking disclosure tone in the chairman's statement: obfuscation or truthful explanations. International Journal of

Accounting and Information Management, 31(5), 838–863. https://doi.org/10.1108/IJAIM-03-2023-0060

Al-Okaily, M., & Al-Okaily, A. (2024). Financial data modeling: An analysis of factors influencing big data analytics-driven financial decision quality. Journal of Modelling in Management. https://doi.org/10.1108/JM2-01-2024-0261

Alalwani, Z., & Mousa, G. A. (2020). Optimistic disclosure tone in corporate annual reporting and financial performance. 2020 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/DASA49378.2020.9162287

Aljohani, A. (2023). Predictive analytics and machine learning for real-time supply chain risk mitigation and agility. Sustainability, 15(20), 15088. https://doi.org/10.3390/su152015088

Almagtome, A., & Abbas, Z. (2020). Value relevance of financial performance measures: An empirical study. International Journal of Psychological Rehabilitation, 24(7), 6777–6791. https://doi.org/10.37200/IJPR/V24I7/PR2020/9623

Alshorman, S. A. A., & Shanahan, M. (2022). The voice of profit: Exploring the tone of Australian CEO's letters to shareholders after the global financial crisis. Corporate Communications: An International Journal, 27(1), 127–147. https://doi.org/10.1108/CCIJ-10-2021-0197

Alvi, J., Arif, I., & Nizam, K. (2024). Advancing financial resilience: A systematic review of default prediction models and future directions in credit risk management. Heliyon. https://doi.org/10.1016/j.heliyon.2024.e12655

Anwar, R. (2021). Determinants of corporate social responsibility disclosure and its impact on investment behavior of firms: The case of China and Pakistan. Capital University. https://doi.org/10.2139/ssrn.3834826

Ashraf, S., Félix, E. G., & Serrasqueiro, Z. (2020). Development and testing of an augmented distress prediction model: A comparative study on a developed and an emerging market. Journal of Multinational Financial Management, 57, 100659. https://doi.org/10.1016/j.mulfin.2020.100659

Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decision Analytics Journal, 3, 100071. https://doi.org/10.1016/j.decan.2022.100071

Bello, O. A. (2023). Machine learning algorithms for credit risk assessment: An economic and financial analysis. International Journal of Management, 10(1), 109–133. https://doi.org/10.22259/ijm.01002012

Beretta, V., Demartini, C., & Trucco, S. (2019). Does environmental, social and governance performance influence intellectual capital disclosure tone in integrated reporting? Journal of Intellectual Capital, 20(1), 100–124. https://doi.org/10.1108/JIC-02-2018-0049

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-K text to gauge financial constraints. Journal of Financial and Quantitative Analysis, 50(4), 623–646. https://doi.org/10.1017/S0022109015000033

Brunner, D., Legat, C., & Seebacher, U. (2024). Towards next-generation data-driven management: Leveraging predictive swarm intelligence to reason and predict market dynamics. In Collective Intelligence (pp. 152–203). CRC Press.

Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A systematic review. Expert Systems with Applications, 156, 113464. https://doi.org/10.1016/j.eswa.2020.113464

Chakraborty, B., & Bhattacharjee, T. (2020). A review on textual analysis of corporate disclosure according to the evolution of different automated methods. Journal of Financial Reporting and Accounting, 18(4), 757–777. https://doi.org/10.1108/JFRA-01-2020-0026

Dana, L.-P., Salamzadeh, A., Mortazavi, S., Hadizadeh, M., & Zolfaghari, M. (2022). Strategic futures studies and entrepreneurial resiliency: A focus on digital technology trends and emerging markets. Tec Empresarial, 16(1), 87–100. https://doi.org/10.14483/22487638.2022.1.87

Diouf, D., & Boiral, O. (2021). CSR or social impression management? Tone management in CSR reports. Journal of Financial Reporting and Accounting, 19(1), 23–42. https://doi.org/10.1108/JFRA-04-2020-0115

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. Review of Accounting Studies, 15, 915–953. https://doi.org/10.1007/s11142-009-9118-z

Fisher, R., Van Staden, C. J., & Richards, G. (2020). Watch that tone: An investigation of the use and stylistic consequences of tone in corporate accountability disclosures. Accounting, Auditing & Accountability Journal, 33(1), 77–105. https://doi.org/10.1108/AAAJ-10-2016-2745

Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. The North American Journal of Economics and Finance, 54, 101251. https://doi.org/10.1016/j.najef.2020.101251

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577. https://doi.org/10.1016/j.jbef.2021.100577

Hajek, P., & Munk, M. (2024). Corporate financial distress prediction using the risk-related information content of annual reports. Information Processing & Management, 61(5), 103820. https://doi.org/10.1016/j.ipm.2024.103820

Hassan, M. K., Abu-Abbas, B., & Kamel, H. (2022). Tone, readability and financial risk: The case of GCC banks. Journal of Accounting in Emerging Economies, 12(4), 716–740. https://doi.org/10.1108/JAEE-09-2021-0200

He, X., Gokmenoglu, K. K., Kirikkaleli, D., & Rizvi, S. K. A. (2023). Co-movement of foreign exchange rate returns and stock market returns in an emerging market: Evidence from the wavelet coherence approach. International Journal of Finance & Economics, 28(2), 1994–2005. https://doi.org/10.1002/ijfe.2359

Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. Frontiers of Business Research in China, 14(1), 13. https://doi.org/10.1007/s11782-020-00074-1

Iqbal, J., & Riaz, K. (2021). Predicting future financial performance of banks from management's tone in the textual disclosures. Quality & Quantity, 1–31. https://doi.org/10.1007/s11135-021-01156-x

Jain, P. K., Saravanan, V., & Pamula, R. (2021). A hybrid CNN-LSTM: A deep learning approach for consumer sentiment analysis using qualitative user-generated contents.

Transactions on Asian and Low-Resource Language Information Processing, 20(5), 1–15. http://dl.acm.org/doi/abs/10.1145/3457206
Kiani, M. M., Raisani, S. T., Nabeel, M., Khan, S. Z., & Lodhi, K. (2023). Impact of international trade and globalization on financial markets: A case of Pakistan. Bulletin of Business and Economics, 12(3), 493–500.
Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., & Smith, N. A. (2009). Predicting risk from financial reports with regression. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 272–280). Association for Computational Linguistics.
Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. The Accounting Review, 84(5), 1639–1670. https://doi.org/10.2308/accr.2009.84.5.1639
Li, F. (2006). Do stock market investors understand the risk sentiment of corporate annual reports? SSRN Working Paper No. 898181. https://doi.org/10.2139/ssrn.898181
Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. Journal of Accounting Research, 48(5), 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x
Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x
Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187–1230. https://doi.org/10.1111/1475-679X.12123
Loughran, T., & McDonald, B. (2020). Textual analysis in finance. Annual Review of Financial Economics, 12(1), 357–375. https://doi.org/10.1146/annurev-financial-080219-051109
Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives: Incremental information or impression management? Journal of Accounting Literature, 26, 116–196.
Molnár, P., Suta, A., Lukács, B., & Tóth, Á. (2024). Towards predicting business activity classes from European digital corporate reports. Engineering Proceedings, 79(1), 50. https://doi.org/10.3390/engproc20240790150
Mousa, G. A., Elamir, E. A., & Hussainey, K. (2022). Using machine learning methods to predict financial performance: Does disclosure tone matter? International Journal of Disclosure and Governance, 19(1), 93–112. https://doi.org/10.1057/s41310-021-00129-x
Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16), 7653–7670. https://doi.org/10.1016/j.eswa.2014.06.009
Nguyen, D. K., Sermpinis, G., & Stasinakis, C. (2023). Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology. European Financial Management, 29(2), 517–548. https://doi.org/10.1111/eufm.12367
Patil, D., Rane, N., Desai, P., & Rane, J. (2024). Machine learning and deep learning: Methods, techniques, applications, challenges, and future research opportunities. In Trustworthy Artificial Intelligence in Industry and Society (pp. 28–81). https://doi.org/10.1007/978-3-030-63879-7_2

Pirayesh, H., & Sallan, J. M. (2024). Advancements in AI for predicting price movements: An in-depth literature survey. Journal of Financial Data Science, 6(4). https://doi.org/10.21314/JFDS.2024.234

Pouryousof, A., Nassirzadeh, F., & Askarany, D. (2023). Inconsistency in managers' disclosure tone: The signalling perspective. Risks, 11(12), 205. https://doi.org/10.3390/risks11120205

Pouryousof, A., Nassirzadeh, F., Hesarzadeh, R., & Askarany, D. (2022). The relationship between managers' disclosure tone and the trading volume of investors. Journal of Risk and Financial Management, 15(12), 618. https://doi.org/10.3390/jrfm15120618

Ranta, M., Ylinen, M., & Järvenpää, M. (2023). Machine learning in management accounting research: Literature review and pathways for the future. European Accounting Review, 32(3), 607–636. https://doi.org/10.1080/09638180.2023.2168121

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 160. https://doi.org/10.1007/s42979-021-00731-9

Shah, M. H., Wang, N., De Chun, H., Zhang, K., Ullah, I., Ullah, A., & Iqbal, K. (2024). Exploring the interwoven relationship: Property rights, financial freedom, government regulation, and stock market fluctuations in emerging economies-A novel system GMM perspective. Heliyon, 10(1). https://doi.org/10.1016/j.heliyon.2024.e11999

Singh, V., Chen, S.-S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. International Journal of Information Management Data Insights, 2(2), 100094. https://doi.org/10.1016/j.joim.2022.100094

Spence, M. (1973). Job market signaling. The Quarterly Journal of Economics, 87(3), 355–374. https://doi.org/10.2307/1882010

Sufi, U., Hasan, A., & Hussainey, K. (2024). Improving the prediction of firm performance using nonfinancial disclosures: A machine learning approach. Journal of Accounting in Emerging Economies, 14(5), 1223–1251. https://doi.org/10.1108/JAEE-07-2023-0205

Teles, G., Rodrigues, J. J., Rabelo, R. A., & Kozlov, S. A. (2021). Comparative study of support vector machines and random forests machine learning algorithms on credit operations. Software: Practice and Experience, 51(12), 2492–2500. https://doi.org/10.1002/spe.3033

Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. Information Fusion, 65, 95–107. https://doi.org/10.1016/j.inffus.2020.11.002

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The computational limits of deep learning. arXiv preprint arXiv:2007.05558, 10. https://doi.org/10.48550/arXiv.2007.05558

Tran, L. T. H., Tu, T. T. K., & Nguyen, T. T. H. (2021). Annual report narrative disclosures, information asymmetry and future firm performance: Evidence from Vietnam. International Journal of Emerging Markets, 18(2), 351–375. https://doi.org/10.1108/IJOEM-08-2020-0925

Wu, D. X., Yao, X., & Guo, J. L. (2021). Is textual tone informative or inflated for firm's future value? Evidence from Chinese listed firms. Economic Modelling, 94, 513–525. https://doi.org/10.1016/j.econmod.2020.12.004

Xia, B., Rawte, V., Gupta, A., & Zaki, M. (2024). FETILDA: Evaluation framework for effective representations of long financial documents. ACM Transactions on Knowledge Discovery from Data, 18(7), 1–27. https://doi.org/10.1145/3657299

Zhang, A. Y., & Zhang, J. H. (2024). Renovation in environmental, social and governance (ESG) research: The application of machine learning. Asian Review of Accounting, 32(4), 554–572.

Zhang, Z., Luo, M., Hu, Z., & Niu, H. (2022). Textual emotional tone and financial crisis identification in Chinese companies: A multi-source data analysis based on machine learning. Applied Sciences, 12(13), 6662.

Zheng, H., Wu, J., Song, R., Guo, L., & Xu, Z. (2024). Predicting financial enterprise stocks and economic data trends using machine learning time series analysis. Applied and Computational Engineering, 87, 26–32.