Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140612

Statistical Analysis of CEO Sentiment in Indian Annual Reports for Predicting Bank Performance Using Machine Learning

Sunil Sharma[1], *S. Prasanna*[1,*], *Murtaza M. Junaid Farooque*[2] *and Khaliquzzaman Khan*[3]

[1] Department of Management Studies, B S Abdur Rahman Crescent Institute of Science and Technology, Chennai, India
[2] College of Commerce and Business Administration, Dhofar university, Salalah, Oman
[3] School of Business, American University in Dubai, Dubai, UAE

Received: 2 Sep. 2025, Revised: 2 Oct. 2025, Accepted: 23 Oct. 2025
Published online: 1 Nov. 2025

Abstract: Leadership communication, especially through the use of media such as leaflets and letters, is increasingly being considered as one of the main direct avenues through which CEO relationships with external stakeholders are fostered. In other words, such communication by the management leadership helps to sway public opinion and eventually impacts the organization's financial viability/performance success, the present work intends to establish this dynamic for the banking sector in India (the top five banks). The research focuses on uncovering the concealed linkage between the bank CEO's tone in their annual shareholder letters and the bank's financial performance measured via Return on Equity (ROE). The study covers the seven years annual letters to shareholders of the five leading Indian banks (Axis Bank, HDFC Bank, City Union Bank, ICICI Bank, and Yes Bank) and employs a lexicon-based sentiment analysis method to quantify the sentiment or the mood prevailing in the CEO letters. The study develops the bank sentiment measures from the shareholder letters, in particular the Sentiment Score, the Positive Ratio, and the Negative Ratio, through a lexicon-based sentiment analysis method. These indicators from the letters are then juxtaposed with the corresponding ROE data using both linear regression and Random Forest models. The findings of the research are contrary to what may have been expected as they reveal a mildly negative association between ROE and high positive mood levels. Thus, it reveals the possibility that the tone used in the letters might not always reflect the reality on the ground. The inference to be drawn from this is that positive communication may be utilized to mask or offset the negative performance perception occasioned by bad performing situations. Moreover, the bank-wise regression models show better results as compared to global ones which indicate that factors related to the institution and context play a significant role. The random forest model is better in terms of prediction performance than the standard regression for very low ROE values as it has a lower Root Mean Squared Error (RMSE). The research offers a new innovative model that merges sentiment data with financial indicators to forecast financial performance and measure executive effectiveness. This is very valuable and informative to the said groups of people- the analysts, the investors, and the regulators, that depend heavily on the qualitative as well as the quantitative cues in the rapidly changing data-driven financial industry landscape.

Keywords: ROE, RMSE, regression models, Sentiment Score, Positive Ratio, Negative Ratio

1 Introduction

Chief Executive Officers (CEOs) in global finance operate within a continually evolving environment and have been compelled to augment their managerial responsibilities to encompass broader dimensions of their roles, including strategic vision, stakeholder engagement, leadership communication, and the expression of organizational sentiment and values. Contemporary CEOs are required not only to address complex financial and operational challenges but also to communicate their vision effectively to stakeholders and, through both formal and informal channels, influence stakeholder perceptions. One of the most significant manifestations of this influence is CEO mood, which refers to the emotional orientation (optimism or pessimism) or the tone reflected in corporate communications, particularly annual letters to shareholders [1].

* Corresponding author e-mail: prasanna.mba@crescent.education

One of the main topics of discussion among an increasing number of scholars and industry professionals is CEO sentiment as a potential leading indicator of financial performance [2]. These communications are not merely words; they often contain implicit references to internal assessments of the company's success, managerial confidence, and future expectations. This focus on top management communication is particularly pronounced in the banking industry, which is highly regulated, capital-intensive, and sensitive to macroeconomic volatility. Consequently, even minor changes in the leadership's communication style can serve as important information for the market, potentially influencing market behavior and corporate valuation [2].

The literature on the subject of CEO sentiment and financial performance that has been existing can be categorized into three strands. The former one is an investigation into the connection between managerial tone and the firm-level financial performance. The second one is on how sentiment analysis has been used in the financial markets and institutional settings. The third strand examines CEO attributes, form of governance, and performance especially in the banking industry. The current research adds to these strands by combining the measures of textual sentiment with machine learning methods to forecast bank performance in the Indian perspective [3-5].

The Indian banking industry has been experiencing revolutionary transformations that are driven by government initiatives such as the Digital India campaign, financial inclusion, and adoption of Unified Payments Interface (UPI) technology, among other initiatives, over the last decade. At the same time, the industry has struggled through ongoing challenges, such as non-performing assets (NPAs), issues in corporate governance, as well as liquidity crisis in certain markets. Besides influencing performance at the workplace, these dynamics have brought doubts on the executive leadership and its strategic rationalizations [6]. Here, the tone and the content of the CEO messages, which can normally be found in shareholder letters and annual reports, have become possibly informative indicators of the well-being of the company, the self-confidence of the management and the future of the company finances. Return on equity or ROE is one of the most significant ratios used to determine the performance of banks with respect to the efficiency with which they utilize shareholder capital. It is the best criterion through which financial performance should be measured as it is efficiency and profitability at the same time. Ascertaining whether or not the mood of CEOs can predict or determine the change in ROE is, thus, a very important first step in learning more about the psychological and behavioral factors of executive leadership in Indian banking. Besides academic importance, these findings also have practical implications to the analysts, investors and policymakers who rely on quantitative and qualitative indicators to make prudent decisions. The correlation between CEO feelings and financial performance is theoretically grounded in the Upper Echelons Theory of management that suggests that management personalities and experience can predict the organizational performance to a certain degree [7]. Based on this concept, the cognitive resources and values of CEOs influence the way they perceive situations and, consequently, the way they make strategic decisions. The sentiment of the CEO in the banking sector may provide a leeway in decisions made on risk-taking, investments, and resource allocation and subsequently financial performance indicators such as ROE. Another aspect through which agency theory helps to understand this interaction is the potential conflict of interest between managers and shareholders [8]. The creation of values might be achieved because of the mood of the CEO that aligns with the interests of shareholders, or it might not and deliver less than optimal financial outcomes.

The sentiment of CEOs can thus be assessed to provide a lens through which management efficacy and its impact on financial performance can be considered. Empirical research has been conducted in a variety of settings and by adopting numerous approaches to investigate how sentiment of the CEO is linked to financial performance. As an example, Li (2010) evaluated the managerial tone of annual reports conversing by text and found that positive language was related to higher earnings in the future [9]. Huang et al. (2014) who studied transcripts of conference calls found that positive stock returns were anticipated under positive CEO emotion [10]. Research has begun to take a closer examination of this relationship in the background of Indian banking. In a study by Kumar et al. (2024), the monetary policy communication of the Reserve Bank of India was analyzed with the help of natural language processing techniques [11]. The research established that dovishness led to a fall in the equity markets, which means that the market participants will perceive such language as reflecting an economy that is weak and not as one that is running a policy. Such an outcome highlights the complexity of the influence of sentiment on financial markets. In another study, Rehman et al. (2022), applied transformer architecture and transfer learning to financial sentiment analysis and demonstrated that sentiment analysis could enhance credit scoring model of financial institutions [12]. These articles indicate that sentiment analysis is gaining growing significance in financial decision making and how it can be used to inform the assessment of CEO sentiment and its impact on financial performance. The Indian banking industry is a unique backdrop that makes it possible to examine the effect of the CEO attitude on financial performance. The change in leadership and enforcement measures has increased the emphasis of the emotions and behavior of the CEO, as of late. The decision by the Securities and Exchange Board of India (SEBI) to prohibit former IndusInd Bank CEO Sumant Kathpalia and other leaders to engage in stock trading due to alleged insider trading is one example of how leadership activities can have a significant impact on financial performance [13]. Also, investment in technology, talent, and Environmental, Social, and Governance (ESG) activities are becoming increasingly significant to Indian banking leaders, according to the KPMG 2024 Banking CEO Outlook report. Such a change in the strategic focus may affect financial performance [14]. It is also important to learn the impact of attitude of the CEO on organizational priorities and, by extension, financial measures such as ROE in this dynamic environment. Empirical studies have shown that the characteristics and personality of the CEO have significant influence on the performance of banks. The results of the analysis of the correlation between the performance of the banks with the characteristics of the CEOs in India as presented by Gupta and Mahakud (2020) indicate that the age, tenure, and educational level of the CEOs have significant effects on the financial performance of the banks [15]. Their study suggests that these characteristics can influence the CEO mood that subsequently can influence financial performance and strategy decisions. Board independence, as well as, CEO duality are facets of corporate governance, which have been shown to affect the financial performance of Indian banks. These research results [16] suggest that the impact of CEO duality, where the CEO also serves as the chair of the board, depends on the ownership structure of the bank. The importance of the governance mechanisms in reducing the impact of the CEO mood on financial performance is demonstrated by their study.

Besides, sentiment analysis of CEO communications is a helpful approach to predicting financial performance. Jena et al. (2022) conducted a study to examine the sentiment expressed in the remarks of the chairperson of the annual report of NIFTY 50 companies including banks [17]. The researchers found that there exists a good association between better financial performance and positive sentiment in such communications, which manifests that CEO emotion as communicated to stakeholders can be a predictor of future financial performance.

The reason why the Indian banking industry is a good place to investigate the impact of CEO sentiment on financial performance is recent leadership transition and strategic activities as well as the idiosyncrasy of this industry. Empirical research reveals the importance of CEO traits and structures, systems of governance, and opinions expressed in terms of their effectiveness in affecting financial performance which necessitates additional researches in the area.

Although an increased amount of literature has been conducted on CEO sentiment and financial performance, the lack of research on this relationship in the Indian banking industry has been noted. Since the sector is critical to the economy and has its own set of challenges including nonperforming assets and changes in the regulations, it is essential to investigate the impact of the attitude of the CEO on the financial results.

The paper will discuss the association between financial performance of major Indian banks according to ROE and attitude of CEO. It quantifies sentiment on the Sentiment Score, Positive Ratio and Negative Ratio, in terms of lexicon-based sentiment analysis of CEO letters. The predictive value of these indicators is estimated on the basis of Random Forest and linear regression models, and they are related to ROE. The analysis identifies the sentiment indicators that produce the largest effect on ROE and also analyses how contextually correct a model can be by comparing global models with bank-specific models. Ultimately, it examines how the tone of the CEO is a reliable indicator of financial performance, and information is given on the validity and relevance of executive communication in the banking sector.

2 Methodology

The annual letters to shareholders sent out by the CEOs of five major Indian banks Axis Bank, HDFC Bank, City Union Bank, ICICI Bank and Yes Bank are the main source of data to be used in the study. A sample of 35 letters (seven consecutive years of each bank) was gathered and the sample covers the period 2017-2023. The letters were sourced out of the annual reports of the different banks that can be found on their respective official websites. The ROE of each bank was also gathered during the same time. ROE has been selected as a meaningful measure of financial performance because it is popularly applied in the evaluation of the banking performance. It is considered best practice to calculate ROE based on average equity over a period because of the mismatch between the income statement and the balance sheet. ROE is calculated by comparing the proportion of net income against the amount of shareholder equity. The equation is:

$$\text{ROE} = \frac{\text{Net Income}}{\text{Shareholders' Equity}}$$

The main measure of financial performance was chosen as ROE, since it is the direct ratio of how a bank is able to produce returns on the capital of the shareholders, so it is specifically applicable in terms of CEO communication and leadership performance. ROE portrays both profitability and capital efficiency to the shareholder as opposed to Return on Assets (ROA) which focuses on the utilization of assets. Other banking ratios like Net Interest Margin (NIM), Capital Adequacy Ratio (CAR), and Non-Performing Asset (NPA) ratios are more functional or regulatory in style and give attention to particular aspects of banking performance as opposed to the creation of value in general. The fact that CEO shareholder letters are mainly sent to investors and equity holders means that ROE would be a better and more compatible metric to determine whether executive sentiment is in line with shareholder-oriented financial results. The sentiment analysis of the letters of the CEOs was preprocessed before the actual analysis. It was changed to lowercase, punctuation marks, and numbers were deleted, and the common English stopwords were deleted. Lemmatization was used to bring the words down to their base forms so that the various inflexions were handled uniformly. This preprocessing pipeline was useful in standardizing the textual information and cutting off noise prior to sentiment extraction. In the sentiment analysis, a lexicon-based method was employed to measure the tone of every letter. An established financial sentiment lexicon which is specifically designed to reflect domain-specific language typically used in corporate and financial disclosures was used in a lexicon-based sentiment analysis approach. The financial text analysis of such lexicons is preferred due to the fact that the general-purpose sentiment dictionaries can incorrectly label neutral financial words (e.g., ‘liability’ or ‘depreciation’ as negative words). Nevertheless, methods based on lexicon are limited per se. They do not entirely explain contextual subtleties, sarcasm, negation, or prospective utterances, and they suppose that sentimental emotion in words does not vary in different contexts. In spite of such constraints, lexicon-based approaches are transparent, interpretable, and consistent and are therefore appropriate in structured corporate text like the shareholder letter of CEOs.

A set of known emotional words of finance was applied to identify the frequency of positive, negative, and total words per letter. Sentiment extraction was performed using Python. A lexicon-based approach was implemented whereby each word in the text was matched against predefined lists of positive and negative terms. The occurrences of these words were then computed separately for each bank. The resulting counts of positive and negative words are summarized in Table 1. After counting words, three derived sentiment measures were calculated of each CEO letter:

$$\text{Positive Ratio} = \frac{\text{Positive Word Count}}{\text{Total Word Count}}$$

$$\text{Negative Ratio} = \frac{\text{Negative Word Count}}{\text{Total Word Count}}$$

$$\text{Sentiment Ratio} = \frac{\text{Positive Word Count} - \text{Negative Word Count}}{\text{Total Word Count}}$$

These standardized measurements give one the same picture of the strength of sentiment in long and short letters.

In a bid to examine the potential relationship that exists between CEO communication tone and financial performance, correlation analysis was carried out on each of the sentiment measures (Positive Ratio, Negative Ratio and Sentiment Score) and corresponding ROE values.

The relationship between sentiment and financial performance were assessed using a multiple linear regression analysis, where ROE is taken as the dependent variable, and Positive Ratio and Negative Ratio are taken as independent variables. This was a test of the explanatory power of CEO mood in predicting return on equity of a bank using this model. The performance of the models as regards to prediction accuracy was measured using standard measurements such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) [18-22]

Besides the regression analysis, numerous visualizing methods were employed to supplement the exploratory and inferential results. They comprised: Linear trend Scatter plots depict the inter-relationship of ROE and each sentiment indicator (Sentiment Score, Positive Ratio, and Negative Ratio). A boxplot of ROE between high and low sentiment with a median value of Sentiment Score. The figure of pairwise correlation (pairplot) is used to visualize the relationships and possible multicollinearity between the significant variables: ROE, Positive Ratio, Negative Ratio, and Sentiment Score. Those visual tools complemented the findings in the form of data, providing intuitive information on how tone of CEO as assessed by measures of sentiment can be associated with financial performance.

Along with the general regression analysis, a global and bank-specific linear regression analysis was developed to determine whether the relationship between CEO sentiment and ROE varied across banks. In the initial step, the entire dataset was employed and a global regression model was trained Positive Ratio and Negative Ratio are predictors of ROE. The model reached a point of benchmarking the total correlation between sentiment and financial performance of all the banks. Separate linear models of each of the five banks were then fitted to obtain bank-wise regression models. This allowed a deeper analysis of whether the effect of the attitude of the CEO on ROE can be institution-specific. The forecasts of the ROE of each bank were calculated and compared to those of global models. To evaluate the model performance, root mean Squared error (RMSE) and the mean absolute percent error (MAPE) were calculated on the global and bank specific models. Comparison of these variables enabled us to conclude whether individualized models performed better than the aggregated model with regard to predicted accuracy.

In order to further improve the predictive analysis, a random forest regression model was developed with Positive Ratio, Negative Ratio and Sentiment Score as input variables in order to forecast ROE. The model chosen was the Random Forest model, which is a non-parametric ensemble learning algorithm due to its capacity to not only model complex and nonlinear relationships and interactions between variables. The model was run using 500 decision trees, mean squared error as the split criteria and fixed random seed to make it reproducible. Other options such as the maximum tree depth, minimum samples per leaf were left at the default value in order to avoid overfitting since the sample size was not too large. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used to evaluate model performance, which allows it to be compared directly with the linear regression model.

In order to compare the model efficacy, both linear regression and random forest methodology were tested using their individual prediction errors. The purpose of this comparison was to determine whether the implementation of machine learning methods caused a high predicted accuracy. Moreover, a variable importance plot was plotted in order to describe the contribution of each component of sentiment to those predictions made by the model. In this analysis the sentiment indicators that were most significant in predicting the variation in financial performance were identified.

3 Result and Analysis

3.1 Sentiment Analysis of CEO Letters

Textual sentiment analysis was employed in assessing the tone and sentiment conveyed through the chosen Indian banks' CEO letters in their annual reports. The word count information helped obtain three sentiment measures: Positive Ratio, Negative Ratio, and Sentiment Ratio. These variables were further contrasted among banks and interpreted in light of their ROE, which is an important indicator of their financial performance.

The results from the sentiment analysis of CEO letters for five major Indian banks—Axis Bank, HDFC Bank, City Union Bank, ICICI Bank, and Yes Bank are shown in Table 1.

Table 1: Sentiment Metrics and ROE for Selected Indian Banks

Bank	Positive	Negative	Total	ROE	Positive Ratio	Negative Ratio	Sentiment Ratio
Axis	126	21	2523	0.61	0.04994	0.00832	0.04162
Axis	123	30	2706	-0.12462	0.04545	0.01109	0.03437
Axis	179	33	3325	0.376861	0.05383	0.00992	0.04391
Axis	191	49	3495	0.005074	0.05465	0.01402	0.04063
Axis	165	27	3512	0.557083	0.04698	0.00769	0.03929
Axis	37	9	1802	0.976313	0.02053	0.00499	0.01554
Axis	150	19	4077	0.457809	0.03679	0.00466	0.03213
HDFC	416	217	12780	1.7	0.03255	0.01698	0.01557
HDFC	458	306	16081	1.700273	0.02848	0.01903	0.00945
HDFC	613	330	19571	1.69928	0.03132	0.01686	0.01446
HDFC	79	28	1697	1.890526	0.04655	0.01650	0.03005
HDFC	121	43	2903	1.764671	0.04168	0.01481	0.02687
HDFC	140	33	3582	1.788523	0.03908	0.00921	0.02987
HDFC	144	22	4012	1.779302	0.03589	0.00548	0.03041
City Union	130	64	5600	1.48	0.02321	0.01143	0.01179
City Union	127	52	6354	1.572226	0.01999	0.00818	0.01180
City Union	160	67	6793	1.583241	0.02355	0.00986	0.01369
City Union	137	63	6345	1.002543	0.02159	0.00993	0.01166
City Union	159	91	7377	1.150116	0.02155	0.01234	0.00922
City Union	184	94	8076	1.299657	0.02278	0.01164	0.01114
City Union	169	97	8142	1.385923	0.02076	0.01191	0.00884
ICICI	38	5	980	1.2	0.03878	0.00510	0.03367
ICICI	46	6	826	0.783995	0.05569	0.00726	0.04843
ICICI	53	14	935	0.315999	0.05668	0.01497	0.04171
ICICI	90	31	1925	0.742463	0.04675	0.01610	0.03065
ICICI	77	22	1689	1.368984	0.04559	0.01303	0.03256
ICICI	82	23	1792	1.747707	0.04576	0.01283	0.03292
ICICI	83	25	1802	2.07581	0.04606	0.01387	0.03219
Yes Bank	85	11	1869	1.56	0.04548	0.00589	0.03959
Yes Bank	51	12	1156	1.602201	0.04412	0.01038	0.03374
Yes Bank	60	15	1092	0.496166	0.05495	0.01374	0.04121
Yes Bank	74	24	1419	-5.74943	0.05215	0.01691	0.03524
Yes Bank	105	35	1720	-1.42771	0.06105	0.02035	0.04070
Yes Bank	85	16	1955	0.354868	0.04348	0.00818	0.03529
Yes Bank	126	13	2547	0.154991	0.04947	0.00510	0.04437

Sentiment analysis is an analytical tool that could also help evaluate the future orientations of bank managers and communication strategies vis-à-vis stakeholders; it must also incorporate financial quantitative measures in an overall assessment of bank performance.

3.2 Correlation Analysis

The data about sentiment expressed in relation to the letters written by the CEOs, and financial data (Table 2), was employed in carrying out a Pearson correlation test to establish thev relationship between the two sets of data. It was necessary to establish the relationship between the ROE and the individual metrics of sentiment, namely Positive Ratio, Negative Ratio, and Sentiment Score.

Table 2: Correlation Matrix

	ROE	Positive Ratio	Negative Ratio	Sentiment Score
ROE	1.000	-0.422	-0.196	-0.355
Positive Ratio	-0.422	1.000	0.182	0.940
Negative Ratio	-0.196	0.182	1.000	-0.166
Sentiment Score	-0.355	0.940	-0.166	1.000

A heatmap of the correlation matrix is presented below to visually illustrate these relationships.

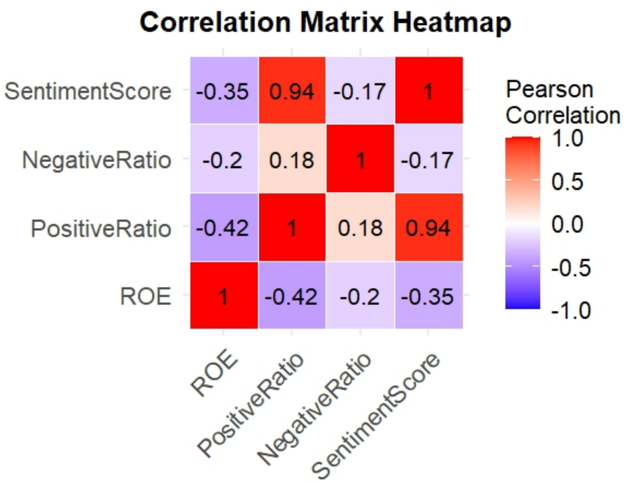

Fig. 1: Correlation heatmap of ROE and sentiment measures

The Fig. 1 presents the Pearson correlation coefficients between financial performance, measured by ROE, and three textual sentiment indicators calculated from CEO shareholder letters: Positive Ratio, Negative Ratio, and Sentiment Score. The correlation heatmap provides a concise visual summary of the interrelationships among financial performance (ROE) and sentiment-based textual indicators derived from CEO letters. As shown in Fig. 1, ROE exhibits a moderately negative correlation with both the Positive Ratio (-0.422) and the Sentiment Score (-0.355). This suggests that a more positive tone in CEO letters does not necessarily correspond to improved financial performance and may, in some cases, reflect optimistic communication during periods of weaker results. The Negative Ratio shows a weak negative relationship with ROE (-0.196), indicating a limited association. Additionally, as shown in Fig. 1, a strong positive correlation is observed between the Positive Ratio and the Sentiment Score (0.94), which is expected given that the sentiment score is derived from positive and negative word counts.

3.3 Regression Analysis and Visualization of Sentiment Impact on ROE

In testing the predictive potential of the sentiment indicators on financial performance, a linear regression model was developed where ROE is the dependent variable and Positive Ratio and Negative Ratio are independent variables. This approach aims to measure the impact of mood embodied in CEO letters on bank performance.

The results of the linear regression analysis are given in Table 3. The model estimates an intercept and coefficients for the two sentiment predictors, along with the estimated standard errors, t-values, and p-values.

Table 3: Linear Regression Results

Predictor	Estimate	Std. Error	t-value	p-value
Intercept	3.0654	0.8641	3.547	0.00122
Positive Ratio	-44.0549	17.8334	-2.470	0.01901
Negative Ratio	-39.1108	51.3438	-0.762	0.45179

The overall fit for the model was given by the R^2 value, which was 0.1924, and this indicated that about 19.24% of ROE variance was explained using the metrics for sentiment. The adjusted R^2 for the model was given as 0.1419, and the F-statistic with a p-value of 0.03277 was 3.811. The model was therefore significant at a 5% level.

Accuracy measures for predictive accuracy were also obtained, and these produced a Root Mean Squared Error (RMSE) of 1.2129 and a Mean Absolute Percentage Error (MAPE) of 1.3634, indicating a moderate degree of predictive error.

3.3.1 Visualization

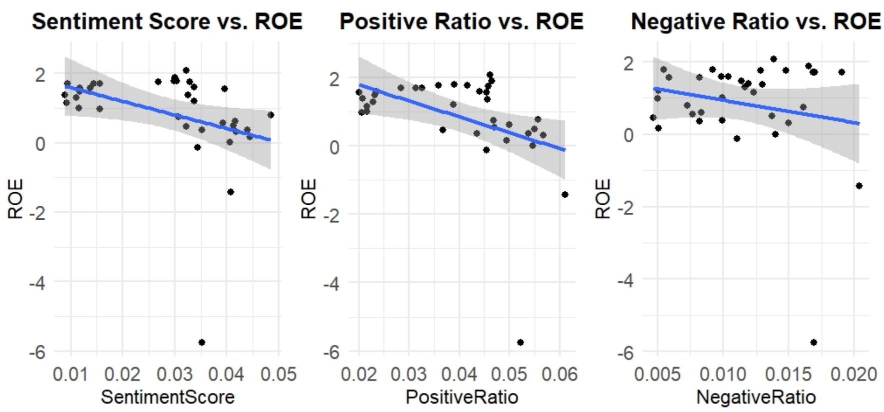

Fig. 2: Scatter plots illustrating the relationship between ROE and CEO sentiment measures

Figure 2 presents three scatter plots showing the association between ROE and (i) Sentiment Score, (ii) Positive Ratio, and (iii) Negative Ratio, derived from sentiment analysis of CEO shareholder letters. Scatter diagrams for ROE and the different measures of sentiment (Positive Ratio, Negative Ratio, and Sentiment Score) showed mostly negative trends, although the trend for Sentiment Score vs. ROE and Positive Ratio vs. ROE had a more pronounced negative trend, as shown in Fig. 2.

The Fig. 2 descriptions are as follows:

–**Sentiment Score vs. ROE:** There is a negative relationship. The higher the sentiment score, the lower the ROE. The regression line slopes downward.

–**Positive Ratio vs. ROE:** This plot also reflects a moderately negative trend, stronger than the plot of Sentiment Score. A higher ratio of positive words in sentiment correlates with lower ROE.

–**Negative Ratio vs. ROE:** The relationship is less obvious and consequently appears weaker, with greater scatter and a lower-gradient regression line.

The relationship between ROE and sentiment was further examined by dividing the banks into "High" and "Low" groups according to the median of their Sentiment Score. This boxplot showed that banks with higher sentimental scores are not necessarily yielding a higher ROE; instead, lower-sentiment groups could perform much better, underlining the complicated linkage between language tone and actual financial success.

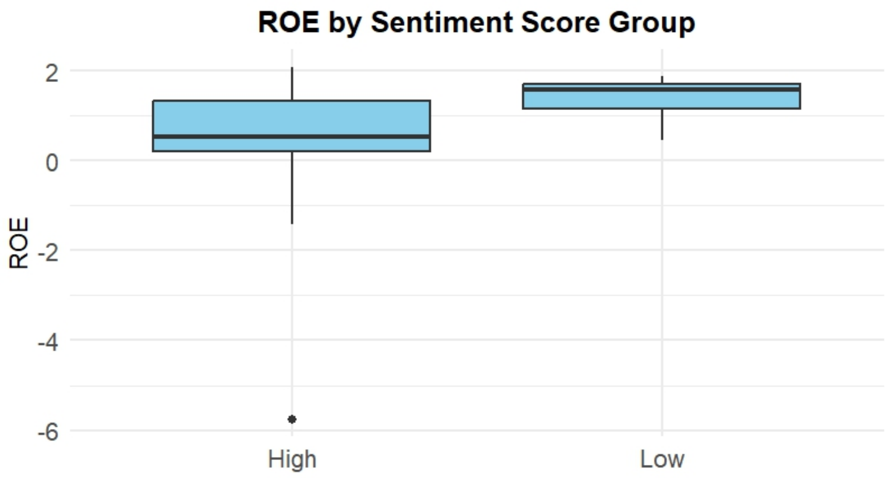

Fig. 3: Boxplot of ROE grouped by high- and low-sentiment CEO communications

Fig. 3 presents boxplots of ROE for two groups of bank-year observations, classified according to whether the Sentiment Score derived from CEO shareholder letters is above or below the sample median. The data cover five major Indian banks over the period 2017–2023. The results indicate that banks with lower sentiment levels tend to report higher ROE values than those with higher sentiment levels. This suggests that more optimistic CEO tone does not necessarily correspond to stronger financial performance and may, in some cases, reflect periods of weaker results. Pairwise correlation visualization further clarifies the relationships between ROE and the sentiment indicators, revealing generally modest linear associations among the variables.

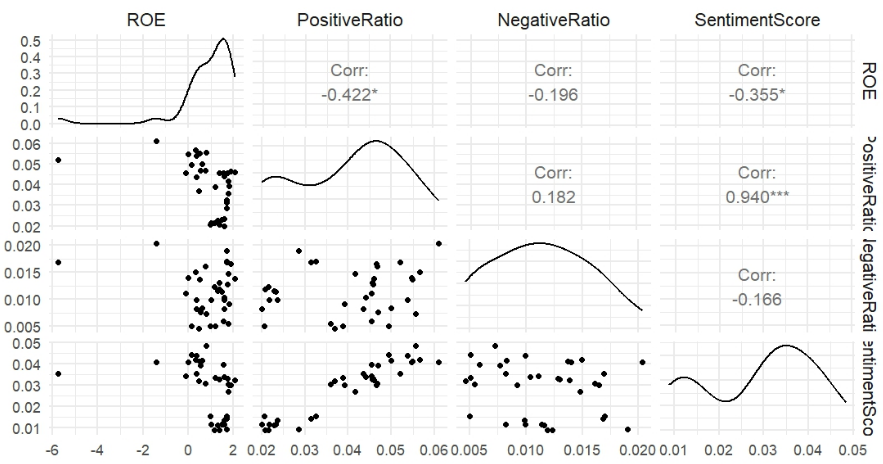

Fig. 4: Pairplot of Sentiment Measures and ROE

As shown in Fig. 4, the pairplot illustrates pairwise relationships and marginal distributions for ROE and the sentiment variables. ROE exhibits a negative correlation with both Sentiment Score and Positive Ratio, with the association appearing stronger for the Positive Ratio. In contrast, the Negative Ratio shows a weak positive correlation with ROE. These visual patterns are consistent with the statistical analysis, which indicates that higher Positive Ratio values and higher overall Sentiment Scores are significantly associated with lower ROE. Additionally, the figure highlights a strong linear relationship between Positive Ratio and Sentiment Score, reflecting their closely related construction.

3.4 Comparative Performance of Overall vs. Bank-wise Linear Regression Models

This section examines the predictive effectiveness of linear regression models designed to estimate ROE with sentiment variables-Positive Ratio and Negative Ratio-as predictors. Two approaches were used:

1. Overall Linear Regression Model: One model was trained on the entire dataset, including all the banks.
2. Bank-specific Linear Regression Models: Each model was trained independently for each bank to be able to capture the different sentiment-to-ROE connections across institutions.

This thus allows us to determine if modelling banks individually increases prediction accuracy over a generalised model.

The performance prediction by both the modeling approaches was done using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Table 4: Performance Comparison of Overall and Bank-wise Regression Models

Model Type	RMSE	MAPE
Overall Model	1.2129	1.3634
Bank-wise Models	0.8499	1.0609

The bank-wise models are more accurate than the overall model, with smaller RMSE and MAPE values as shown in Table 4. This is an indication that the correlation between the sentiment and ROE varies for the various banks, and hence bank-wise modelling improves the estimates. Although the world model is simpler, the tailored bank-wise modelling approach is more accurate in financial prediction tasks involving multiple entities.

3.5 Linear Regression vs. Random Forest for Predicting ROE

To enhance the forecast for the value of ROE, a machine learning approach was adopted using the Random Forest method on the results of the bank-wise and overall linear regression model. Two predictor variables based on the concept of sentiment analysis, Positive Ratio and Negative Ratio, were adopted for the model. To determine whether a more flexible model like Random Forest would handle non-linear relationships between the variables better than the Linear Regression model, both the Random Forest model and the Linear Regression model were adopted for the predictions, and the Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) were adopted for comparing the performance of both the models. A comparative table has also been prepared for investigation on the predictive ability for different banks.

3.5.1 Model Performance Comparison

The Root Mean Squared Error for the Random Forest model is significantly lower compared to the Linear Regression model with values of 0.7306 and 1.2129, respectively as shown in Table 5. This is an indication that the Random Forest model's estimates are closer to the actual values for the ROE.

However, the MAPE for the Random Forest model (1.7476) is slightly higher compared to the Linear Regression model (1.3634) possibly due to the relative errors made by the model in estimating the low ROE's as shown in Table 5. This is an indication that the Linear Regression model may perform better than the Random Forest model for the specific data points, although the Random Forest model is more accurate.

Table 5: Model Performance Comparison

Model	RMSE	MAPE
Linear Regression	1.2129	1.3634
Random Forest	0.7306	1.7476

3.5.2 Future Prediction

From the prediction table 6, it can be seen that in most instances, both models forecast the value of the Random Forest between a reasonable range. In most extreme instances, for example, negative values of ROE, the Random Forest forecasters approximate the value of Random Forest much better than Linear Regression forecasters. This shows that Random Forest performs better than Linear Regression in handling instances of outliers and non-linear instances.

Table 6: Prediction Table

Actual ROE	Predicted ROE (Linear Regression)	Predicted ROE (Random Forest)
0.6100	0.5397	0.3491
-0.1246	0.6293	0.7109
0.3769	0.3055	0.2990
0.0051	0.1094	0.2294
0.5571	0.6949	0.6708
0.9763	1.9655	1.0517
0.4578	1.2622	0.8226
1.7000	0.9672	0.8408
1.7003	1.0664	0.7651
1.6993	1.0260	1.3174
-5.7494	0.1064	-1.9954
-1.4277	-0.4199	-1.6194

A scatterplot reveals a number of points clustering fairly around the ideal 45-degree reference line as shown in Fig. 5, but with noticeable deviations at higher and lower ROE values. The scatterplot reflects tighter clustering of the predicted values around the reference line, reflecting improved fit and reduced bias in prediction across the ROE spectrum.

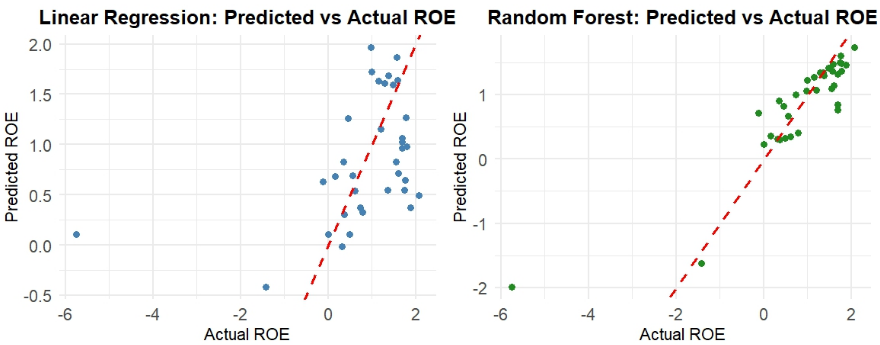

Fig. 5: Comparison of actual and predicted ROE values using Linear Regression and Random Forest models

As shown in Fig. 5, the comparison of actual and predicted ROE values demonstrates that the Random Forest model outperforms the Linear Regression model in terms of predictive accuracy when sentiment variables are used as inputs. The Random Forest model achieves a lower root mean squared error (RMSE) and more effectively captures non-linear relationships and outliers, resulting in predictions that align more closely with observed ROE values. Although the Random Forest model exhibits a slightly higher mean absolute percentage error (MAPE), indicating relatively larger proportional errors for some observations, the overall visual and statistical evidence suggests superior robustness and forecasting performance compared to the Linear Regression approach.

4 Discussion

The study also reveals that, in a major paradox of CEO communications, lower ROE values tend to be associated with more affirmative tones of shareholder letters. Good executive communication usually comes with good financial performance which is not the case with the findings of this study. To preserve investor confidence and manage outer impressions, the CEOs can strategically manipulate the tone to sound more optimistic in case of bad performance. The Positive Ratio in which the negative correlation with ROE is significantly higher than the Negative Ratio indicates that the role of CEO optimism in financial communication is quite complicated. Thus, excessive optimism emotion can signify overcompensation as opposed to operational strength, particularly when a person is in a stressful financial situation.

The importance of institutional setting is also supported by the fact that the bank-specific regression model proves to be better than the global model. The CEO sentiment-financial outcome relationship appears to be mediated by the past performance differences, regulatory exposure, governance structure, and leadership style. The implication of this finding is that major institution level processes could be masked by pooling banks in one forecasting model. In that respect, bank-specific, tailored models provide more accurate and context specific, insights into the sentiment-performance relationship.

To increase the interpretability, the measures of feature importance were obtained of the Random Forest model. The findings show that the Positive Ratio is the strongest predictor of ROE, then Sentiment Score and finally Negative Ratio has a relatively smaller contribution towards the model prediction. This rank is consistent with the regression and correlation results and supports the conclusion that positive language is the major factor in portraying the relationship detected between CEO sentiment and financial performance. A comparatively less significance of negative sentiment could be related to regulatory provisions and reputational issues that inhibit bare negative disclosures during banking communications. The negative correlation between positive sentiment on the one hand and ROE on the other can be viewed in the context of the theory of tone management as well as impression management. CEOs could also use more positive language to placate investors and reduce bad impressions and ensure stakeholder confidence during times of poor financial results. Prior accounting and finance literature has recorded this type of smoothing in their narratives whereby optimism on behalf of managers increases when the underlying performance is declining. In its turn, higher positive sentiment, therefore, might not pass as an indicator of good performance, but as a remedial form of communication, oriented at shaping an external perception.

Generally, the findings indicate the importance of integrating quantitative financial metrics with qualitative emotion scales in an attempt to have a more comprehensive understanding of bank performance and CEO effectiveness. The study has led to improved predictability and interpretability of leadership communication in the banking sector that can be utilized by investors, regulators, and analysts, since they combine classical regression analysis with features importance in machine learning.

5 Conclusion

This research establishes that there is a complex and intense relationship between CEO sentiment and financial performance for Indian banks. This research revealed that it is not always the case that positive communication translates to successful financial performance using a lexicon-based sentiment analysis approach to analyze the CEOs' shareholder letters and then attribute these measures of sentiment to ROE. However, it is established that positive sentiment is, at times, related to low ROE, suggesting that it is a form of tone management, where executives actually aim to dampen the impact of adverse performance through overly positive communication. Conversely, moderate or earthy tone could mean that they are being transparent and realistic about performance that mirrors organizational realities.Random Forest Modelling brings in a very effective, non-linear machine learning technique that enhances predictive efficiency, specifically in the context of complex and complexly patterned information in textual form. In particular, it can be seen that the regression models based on each bank perform better than the general model, indicating that characteristics of each financial institution, such as leadership style and stakeholder expectations, serve as moderators in managing the relationship between mood and financial variables. On a very applied level, implications are sought among investors, financial regulators, financial analysts, and politicians in recognizing that all financial decision-making must now incorporate non-financial communication patterns. CEO communication, often overlooked in standard financial analysis, seems very much a crucial variable that can supplement the standard list of financial variables. Future research studies can certainly build on this platform through attempts to focus upon current CEO sentiment through earnings call transcript analyses, quarterly reporting, and public speeches. In general, extending sentiment analyses into multilingual datasets and other financial variables such as Return on Assets (ROA), market capitalization, and stock market performance could lead to a much larger paradigm shift on the applications and development of sentiment in financial intelligence analyses.

References

[1] Lee, Y. (2022). Dynamics of symmetrical communication within organizations: The impacts of channel usage of CEO, managers, and peers. *International Journal of Business Communication*, 59(1), 3–21.

[2] Gao, Z., Hua, M., Liu, H., and Yan, C. (2023). News sentiment and CEO retirement: The impact on firm performance and risk. *Research in International Business and Finance*, 66, 102031.

[3] Alta'any, M., Kayed, S., Meqbel, R., and Albitar, K. (2025). Speaking success: managerial tone in earnings conference calls and financial performance. *Corporate Governance: The International Journal of Business in Society*, 25(2), 311–336.

[4] Du, K., Xing, F., Mao, R., and Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 56(9), 1–42.

[5] Saha, R., Kabir, M. N., and Chowdhury, A. H. (2023). The impact of CEO attributes on sustainability performance: evidence from an emerging economy. *Accounting Research Journal*, 36(6), 539–557.

[6] Vermani, R., and Arora, N. (2025). Unravelling the dynamics: a theoretical exploration of the role of unified payment interface (UPI) in digital inclusion. In *From Digital Disruption to Dominance* (pp. 75–102). Emerald Publishing Limited.

[7] Hambrick, D. C., and Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2), 193–206.

[8] Jensen, M. C., and Meckling, W. H. (2019). Theory of the firm: Managerial behavior, agency costs and ownership structure. In *Corporate Governance* (pp. 77–132). Gower.

[9] 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.

[10] Huang, X., Teoh, S. H., and Zhang, Y. (2014). Tone management. *The Accounting Review*, 89(3), 1083–1113.

[11] Kumar, R., Paul, S. B., and Singh, N. (2024). Words that move markets—Quantifying the impact of RBI's monetary policy communications on Indian financial market. *arXiv preprint* arXiv:2411.04808.

[12] Rehman, T., Bose, R., Chattopadhyay, S., and Sanyal, D. K. (2022). Transfer learning and transformer architecture for financial sentiment analysis. In *Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing (IEM-ICDC 2021)* (pp. 17–27). Springer Nature Singapore.

[13] Ghosh, S. (2025). Tone at the top: leadership makeover and ESG focus in Indian private banks. *International Journal of Disclosure and Governance*, 1–14.

[14] KPMG. (2024). *KPMG 2024 Banking CEO Outlook*. Retrieved from https://kpmg.com/in/en/insights/2024/11/kpmg-2024-banking-ceo-outlook.html

[15] Gupta, N., and Mahakud, J. (2020). CEO characteristics and bank performance: evidence from India. *Managerial Auditing Journal*, 35(8), 1057–1093.

[16] Sharma, D., and Dey, S. K. (2022). Impact of corporate governance facets on financial performance of Indian banks. *Journal of Commerce and Accounting Research*, 11(3), 34.

[17] Jena, M. K., Das, S., and Kar, B. (2024). Sentiment analysis of chairperson's message and its influence on financial performance: Study on NIFTY 50 companies. *Indian Journal of Finance*, 18(11), 8–23.

[18] Hisam, M. W., Khan, K., Khan, M. S., Arora, V., and Ahmad, A. (2026). GDP forecasting for policy: Evaluating ARIMA, exponential smoothing, and XGBoost models for fiscal, monetary, and welfare planning.

[19] Dar, A. A., and Shaik, A. J. (2025). Spatiotemporal clustering and multivariate forecasting of air quality index across Indian cities using machine learning and deep learning models. *Franklin Open*, 100435.

[20] Rehman, M. Z., Dar, A. A., and Wangmo, T. A. (2024). Forecasting CO_2 emissions in India: A time series analysis using ARIMA. *Processes*, 12(12), 2699.

[21] Dar, A. A., Jain, A., Malhotra, M., Farooqi, A. R., Albalawi, O., Khan, M. S., and Hiba. (2024). Time series analysis with ARIMA for historical stock data and future projections. *Soft Computing*, 28(21), 12531–12542.