Working Paper No. 10-2023

Improving the Prediction of Firm Performance using Nonfinancial Disclosures: A Machine Learning Approach

Usman Sufi
Arshad Hasan
Khaled Hussainey

Innovation and Technology Centre
Lahore School of Economics

Innovation and Technology Centre (ITC)

Dr. Azam Chaudhry
Professor, Dean & Head of Department (Economics)
Pro-Rector
Co-Director

Dr. Theresa Chaudhry
Professor (Economics)
Co-Director

Innovation and Technology Centre,
The Lahore School of Economics,
Intersection of Main Boulevard, Burki Road,
Phase VI, DHA, Lahore 53200, Pakistan
Tel: +92-(0)42-3656-0969
URL: https://itc.lahoreschool.edu.pk
Email: ayeshakh@lahoreschool.edu.pk

Working Paper No. 10-2023

Improving the Prediction of Firm Performance using Nonfinancial Disclosures: A Machine Learning Approach

Usman Sufi
Teaching and Research Fellow
Innovation & Technology Center
Lahore School of Economics
usman.sufi@lahoreschool.edu.pk
ORCID ID: 0000-0003-1701-4962 (*Corresponding Author*)

Arshad Hasan
Professor
Lahore School of Economics
arshad@lahoreschool.edu.pk
ORCID ID: 0000-0002-3914-3901

Khaled Hussainey
Professor
University of Portsmouth
khaled.hussainey@port.ac.uk
ORCID ID: 0000-0003-3641-1031

Abstract: *The purpose of the study is to test if the prediction of firm performance can be improved by incorporating nonfinancial disclosures into financial predictive models. The study utilizes narrative disclosure tone and corporate governance mechanisms as nonfinancial disclosures. Accordingly, three predictive models with each carrying a different set of predictors are developed. The study utilizes two widely popular machine learning techniques, random forest and stochastic gradient boosting, for prediction using the three models. Data are collected from a sample of 1250 annual reports of 125 nonfinancial firms in Pakistan for the period 2011-2020. As a proxy for firm performance, two market-based and two accounting-based measures are used. Our results indicate that both narrative disclosure tone and corporate governance mechanisms significantly improve the accuracy of financial predictive models of firm performance, especially market-based firm performance. The study contributes by addressing the neglect of nonfinancial disclosures in the prediction of firm performance. Furthermore, it addresses the scarcity of corporate governance literature relevant to the use of machine learning techniques. Finally, the study also contributes by exploring the contrasting role of market and accounting-based performance in this context, rendering market-based performance more relevant to prediction by nonfinancial disclosures.*

Keywords: Firm performance, machine learning, random forest, stochastic gradient boosting, disclosure tone, corporate governance, artificial intelligence.

JEL Classification: M41, G10, O31, O32.

First Printing: November, 2023

Funding: There is no funding for this research.

Compliance with ethical standards: The authors have complied with ethical standards.

Conflict of interest: The authors declare no conflict of interest.

Data availability statement: The data is available on request.

Improving the Prediction of Firm Performance using Nonfinancial Disclosures: A Machine Learning Approach

1. Introduction

The importance of being able to predict a firm´s performance with ever improved accuracy is unmatchable, especially in a post global financial crisis period (Hunt, J. N. Myers & L. A. Myers, 2022; Yameen, Farhan & Tabash, 2019; Yang, Bento & Akbar, 2019). This has recently become especially relevant with the advent of new predictive technologies such as machine learning and artificial intelligence (Moll & Yigitbasioglu, 2019; van der Heijden, 2022). Despite this, most empirical research relevant to firm performance is embedded in the use of traditional regression techniques (Chang, Yu & Hung, 2015, Florio & Leoni, 2017; Ibhagui & Olokoyo, 2018). Interestingly, most of these discussions about firm performance investigate financial indicators as its predictors (Chang et al., 2015; Ibhagui & Olokoyo, 2018; Wang, 2016). In contrast, there have been recent suggestions that nonfinancial disclosures provide incremental information in this context (Mousa, Elamir & Hussainey, 2022). Despite this, Hunt et al. (2022) contend that empirical research regarding the prediction of firm performance using nonfinancial disclosures is scarce.

However limited, nonfinancial disclosures have received theoretical and empirical importance in their association with firm performance (Jensen & Meckling, 1976; Mousa et al., 2022). For instance, Beretta, Demartini, Lico & Trucco (2021) elaborate on incremental information theory´s stance that narrative disclosure tone provides additional information relative to financial disclosures. Moreover, they establish an empirical association between disclosure tone and ESG performance. Nevertheless, there is a dearth of studies that utilize nonfinancial disclosures as predictors of performance (Hunt et al., 2022). Accordingly, Hunt et al. (2022) suggest that the negligence of narrative disclosure tone in this regard is an evident gap in the literature. Another form of nonfinancial disclosures that can be used to determine firm performance are corporate governance mechanisms (Ahmed, Alabdullah, Thottoli & Maryanti, 2020). Accordingly, agency theory posits that better corporate governance mechanisms enhance performance (Azeez, 2015; Jensen & Meckling, 1976). While this notion has received sufficient empirical attention, most of it is embedded in the use of traditional regression techniques (Di Vito & Trottier, 2022). Interestingly, Moll & Yigitbasioglu (2019) contend that the world has improved with technological advances in the field of statistical analysis. Despite this, the use of machine learning algorithms in the corporate governance literature is scant (Di Vito & Trottier, 2022). Accordingly, Mousa et al. (2022) suggest utilizing machine learning methods to test if the prediction of firm performance can be improved by incorporating corporate governance mechanisms as predictors. Furthermore, an additional source of inconsistency stems from the contrasting roles of market and accounting-based firm performance (Yang et al., 2019). Accordingly, Yang et al. (2019) posit that the unique role of these facets of performance is relevant to its prediction. Finally, Mousa et al. (2022) deem that identifying predictors of firm performance is more vital in an emerging economy, as most research in this regard is concentrated on developed markets.

Stemming from these concerns, the purpose of the study is to fill the aforementioned gaps by using machine learning algorithms to investigate whether nonfinancial disclosures, such as narrative disclosure tone and corporate governance mechanisms, can improve the prediction of firm performance in Pakistan, an emerging economy. Furthermore, the study also explores the role of market and accounting-based measures of performance in its prediction.

The study utilizes two widely popular machine learning techniques for prediction purposes, namely, random forest and stochastic gradient boosting. The data are collected from the annual reports of 125 nonfinancial firms in Pakistan spanning a period of 10 years from 2011-2020. A sentiment analysis is performed for the operationalization of narrative disclosure tone, while corporate governance mechanisms are taken directly from the annual reports. Firm performance is proxied by two accounting-based (ROA and ROE) and two market-based estimates (Tobin´s Q and MTB). Finally, three predictive models with each containing a different set of predictors are developed. Model 1 contains a set of financial disclosures as predictors, while model 2 contains narrative disclosure tones and financial disclosures as predictors. Finally, model 3 contains corporate governance mechanisms and financial disclosures as predictors. Prediction using each of these models is then performed by utilizing the two machine learning techniques. Accordingly, the performance of models 2 and 3 is compared with model 1 to test if the addition of nonfinancial disclosures in a financial predictive model of firm performance improves accuracy.

By conducting this research, we contribute to the literature in five ways. First, we address the scarcity in research relevant to the use of disclosure tone as a predictive tool by establishing its imperativeness in the prediction of firm performance (Hunt et al., 2022; Mousa et al., 2022). Second, our results also contribute by establishing the importance of corporate governance mechanisms in the prediction of firm performance by utilizing machine learning algorithms. In doing so, we respond to the call of Di Vito & Trottier (2022), as they contend that corporate governance needs to be amalgamated with machine learning literature. Third, as suggested by van der Heijden (2022), we contribute to the limited machine learning literature in the realm of finance and accounting. By doing this, we respond to the call of Moll & Yigitbasioglu (2019) by adding to the reliability of machine learning techniques as predictive tools of financial outcomes. Fourth, responding to Mousa et al. (2022), the study fills a gap by utilizing machine learning to examine the predictive ability of nonfinancial disclosures in an emerging economy. As suggested by Ullah & Saqib (2018), Pakistan provides a unique setting with heightened economic uncertainty. Therefore, identifying ways to improve the predictability of firm performance can be crucial to all stakeholders in such a setting. Finally, the study also answers the call of Yang et al. (2019) and explores the conflicting role of market and accounting-based firm performance. To this end, our results suggest that nonfinancial disclosures in particular are more relevant to the prediction of market-based firm performance and thus reflect market´s reaction.

2. Literature Review and Hypothesis

2.1 Theoretical Framework

The theoretical foundation of the current study is primarily based on incremental information and agency theories (Ahmed et al., 2020; Beretta et al., 2021). The association between disclosure tone and the prediction of firm performance in the study is explained by incremental information theory, while agency theory justifies the use of corporate governance mechanisms in this context.

The main premise of incremental information theory is that companies utilize the tone of additional narrative disclosures to reduce informational asymmetry (Beretta et al., 2021). Elaborating further, Arena, Bozzolon & Michelon, (2015) contend that managers specifically utilize the tone of narrative disclosures to provide value-relevant information to future investors. Moreover, they explain this by suggesting that managers have no financial incentives to distort information, as investors are easily able to assess the market. Therefore, they suggest that managers signal the future performance of the firm accurately through narrative disclosures and their tone, beyond which can be explained by financial statements. While the notion that disclosure tone significantly improves the prediction of firm performance is linked to incremental information theory, a similar suggestion supported by agency theory is made about corporate governance mechanisms (Ahmed et al., 2020).

Agency theory states that the separation of ownership and control in a firm enhances agency conflicts (Jensen & Meckling, 1976). While managers are focused on short-run performance and their own compensation, shareholders are more concerned about the long-term value of the firm. These conflicts of interest cause a subsequent rise in agency costs, which results in decreased performance (Azeez, 2015). A possible solution grounded in agency theory is to employ a better corporate governance framework, as it can potentially eliminate agency conflicts and result in enhanced performance (Azeez, 2015; Jensen & Meckling, 1976).

Below, we utilize the overarching theoretical framework of the study grounded in agency and incremental information theories to discern relevant literature regarding the prediction of firm performance.

2.2 Review of relevant literature

Most literature regarding the prediction of outcomes such as financial distress, performance and bankruptcy is embedded in the use of financial disclosures as predictors (Delen, Kuzey & Uyar, 2013; Wang, 2016) More specific to firm performance, Delen et al. (2013) employ financial ratios and indicators in their study. Accordingly, they suggest that financial ratios such as liquidity and leverage have predictive ability in regard to forecasting future firm performance. Interestingly, firm size also plays a critical role in this regard (Ibhagui & Olokoyo, 2018). As such, Ibhagui & Olokoyo (2018) predict that the negative effect of leverage on performance is more imminent in smaller firms. Furthermore, Lambey, Tewal, Sondakh & Manganta (2021) advocate a signaling theory perspective, as they suggest that older firms signal more experience and are associated with high performance. Consequently, they suggest that firm age improves the forecast of firm performance. In addition, financial indicators such as operating cash flows and firm risk are also deemed crucial in the prediction of firm performance (Chang, Yu & Hung, 2015; Florio & Leoni, 2017). While it is clear that financial disclosures have a degree of predictive ability relevant to firm performance, the utilization of nonfinancial disclosures such as narrative disclosure tone in this regard is scarce (Arena et al., 2015; Beretta et al., 2021; Mousa et al., 2022).

According to Aerts (2015), narrative disclosures offer a detailed description of how the management views the firm and usually accompanies financial statements. In a comprehensive analysis, Beretta et al. (2021) utilize narrative disclosure tone as a determinant of future firm performance. However, their research was limited to the automotive industry. Similarly, Aly, El-Halaby & Hussainey (2018) utilize disclosure tone as a predictor of financial performance. Their study is limited to the use of traditional regression techniques, whereas more advanced methods of analysis, such as machine learning are recommended (Moll & Yigitbasioglu, 2019; Mousa et al., 2022). In addition, Caserio, Panaro & Trucco (2020) investigate the ability of narrative disclosure tone to predict financial performance relevant to U.S. banks. However, their analysis is also limited to the use of traditional regression techniques and is specific to a developed economy. In contrast, Mousa et al. (2022) utilize three machine learning algorithms to determine if the performance of financial predictive models can be improved by incorporating narrative disclosure tone in them. Despite their use of advanced statistical techniques such as machine learning, their study is limited to banking institutions and a smaller sample size. Accordingly, they suggest extending their study by incorporating a larger and more diverse sample comprising various sectors within nonfinancial firms. Furthermore, they suggest using other types of nonfinancial disclosures, such as certain corporate governance mechanisms, in the prediction of performance via machine learning algorithms.

Corporate governance mechanisms constitute an important part of overall nonfinancial disclosures (Ahmed et al., 2020). Consequently, the notion that better corporate governance mechanisms enhance firm performance has received ample empirical support (Adjaoud, Zeghal & Andaleeb, 2007; Kakanda, Salim & Chandren, 2017). While most of these studies that focus on analyzing the corporate governance-firm performance nexus employ traditional regression techniques, the use of machine learning algorithms in this context is severely limited (Di Vito & Trottier, 2022). Consequently, Mousa et al. (2022) and Di Vito & Trottier (2022) advocate the use of machine learning algorithms to test whether adding corporate governance mechanisms into financial predictive models of performance improves its prediction. Despite their being ample research relevant to the prediction of firm performance, its multifaceted nature is largely ignored in this regard (Arena et al., 2015; Beretta et al., 2021).

It is clear from recent literature that firm performance is a multidimensional construct and that its different measures capture distinctive dimensions (Florio & Leoni, 2017; Yang et al., 2019). Despite this, most studies relevant to the prediction of performance proxy it via a singular dimension (Arena et al., 2015; Beretta et al., 2021; Mousa et al., 2022). For instance, the study of Beretta et al. (2021) is limited to the prediction of ESG performance. In addition, most other studies in this regard are restricted to the prediction of financial performance (Mousa et al., 2022). However, Yameen et al. (2019) posit that market-based measures of firm performance take into account a number of factors that cannot be captured by financial or accounting-based measures of performance. Consistent with this, there have been suggestions to explore further into the distinctive roles of these different facets of firm performance with regards to its prediction (Arena et al., 2015; Elvin & Bt Abdul Hamid, 2016; Florio & Leoni, 2017; Yameen et al., 2019; Yang et al., 2019). Accordingly, research regarding the prediction of firm performance is subject to prominent gaps.

First, most research in this regard is restricted to financial disclosures as predictors, while the use of nonfinancial disclosures is limited (Hunt et al., 2022; Mousa et al., 2022). The current study responds to this by employing narrative disclosure tone and corporate governance mechanisms as predictors. Second, the limited literature that focuses on narrative disclosure tone as a predictor of performance is restricted to the use of traditional regression techniques and is mostly limited to a specific type of institution or industry (Aly et al., 2018; Beretta et al., 2021; Caserio et al., 2020; Mousa et al., 2022). The current study addresses this gap by employing machine learning algorithms for prediction purposes and a sample comprising of nonfinancial firms representing various sectors. Third, the existing corporate governance literature is also severely limited to the use of traditional regression techniques, whereas the use of machine learning algorithms in this context has been specifically advocated (Di Vito & Trottier, 2022). The current study answers the call of Di Vito & Trottier (2022) by amalgamating the corporate governance literature with machine learning. Finally, most literature regarding the prediction of firm performance ignores the distinctive role of its multiple facets (Yameen et al., 2019; Yang et al., 2019). The current study fills that gap by exploring the contrasting role of accounting and market-based measures of performance in its prediction.

Below, we delineate empirical literature backed by theoretical justifications and the context of Pakistan specific to the building of our hypotheses.

2.2.1 Narrative disclosure tone

Narrative disclosure tone as a significant predictor of firm performance has only recently achieved empirical significance (Aly et al., 2018; Beretta et al., 2021; Caserio et al., 2020; Mousa et al., 2022). For instance, Aly et al. (2018), in an analysis of disclosure tone and financial performance, empirically prove that narrative disclosure tone influences performance. Consistently, Caserio et al. (2020) also provide empirical evidence that narrative disclosure predicts future firm performance in U.S. banks. In addition, Beretta et al. (2021) empirically demonstrate that the positive tone in narrative disclosures of the top 10 automotive companies worldwide predicts ESG performance. Finally, Mousa et al. (2022) provide empirical evidence supported by machine learning algorithms that narrative disclosure tone incrementally improves the prediction of firm performance. This notion in most empirical literature is grounded in incremental information theory (Arena et al., 2015; Beretta et al., 2021).

As mentioned above, incremental information theory posits that narrative disclosures have value-relevant incremental information about the future of firm performance that cannot be captured by financial disclosures alone (Arena et al., 2015; Beretta et al., 2021). Despite its theoretical and empirical importance, the predictive ability of narrative disclosures is an underresearched area especially relevant to developing economies (Mousa et al., 2022).

In this context, Pakistan provides the unique setting of an emerging economy constituting a weak regulatory framework coupled with heightened economic uncertainty (Ullah & Saqib, 2018). Consequently, the accurate prediction of firm performance in such a setting can be extremely challenging, shattering the confidence of investors as a result. Consequently, most investors in such an environment rely on narrative disclosures for decision-making (Aly et al., 2018). Accordingly, it is important to test whether narrative disclosures provide incremental information about a firm's performance, especially in the Pakistani context. Therefore, we hypothesize the following:

H1: Narrative disclosure tone improves the ability of financial disclosures to predict firm performance.

2.2.2 Corporate governance mechanisms

Corporate governance mechanisms have always been relevant to firm performance in terms of empirical support (Adjaoud et al., 2007; Kakanda et al., 2017; Yameen et al., 2019). For instance, Adjaoud et al. (2007) found a positive and significant relationship between board quality and long-term value-based firm performance. In addition, Ciftci, Tatoglu, Wood, Demirbag & Zaim (2019) add to this claim by empirically proving that measures such as board size, foreign and concentrated ownership of the firm have a positive impact on performance in the Turkish context. Consistent with this, Kakanda et al. (2017) contend that board-related variables in particular significantly improve performance. A similar result came from the panel data analysis of the Bombay Stock Exchange, where Yameen et al. (2019) find a significant impact of corporate governance variables on firm performance. Corporate governance and its impact on firm performance is grounded in agency theory (Jensen & Meckling, 1976).

As previously mentioned, the main premise of agency theory is that agency costs and informational asymmetries caused by internal agency conflicts hinder performance. Agency theory further deems that better governance mechanisms can curb agency costs and eventually enhance performance. Interestingly, this becomes more relevant in the context of emerging economies with a weak regulatory framework and extreme economic uncertainty (Azeez, 2015; Ciftci et al., 2019).

Accordingly, the economic setting of Pakistan is subject to severe economic instability and weak enforcement of governance regulations (Saeed, Ali, Riaz & Khan, 2022; Ullah & Saqib, 2018). As previously mentioned, it is imperative for regulators and policymakers to restore the confidence of investors in such a setting. Therefore, establishing the reliability of corporate governance mechanisms as indicators of firm performance is more relevant in settings such as that of Pakistan. As suggested by Di Vito & Trottier (2022) and Mousa et al. (2022), this can be achieved by testing the predictive ability of corporate governance mechanisms via machine learning algorithms. The use of artificial intelligence is especially relevant, as most literature in this context is limited to the use of regression techniques. Therefore, we hypothesize the following based on the empirical and theoretical significance of corporate govenance mechanisms in the Pakistani context:

H2: Corporate governance disclosures improve the ability of financial disclosures to predict firm performance.

2.2.3 Market and Accounting-based performance measures

Literature relevant to firm performance is subjected to inconsistencies between its multiple dimensions, especially market and accounting-based performance (Yameen et al., 2019; Yang et al., 2019). Consequently, the literature has made several attempts to explain these differences (Adjaoud et al., 2007; Yang et al., 2019).

One major difference where the literature converges is that accounting-based estimates of firm performance reflect the past, whereas market-based measures reflect the future (Yang et al., 2019). Furthermore, Yang et al. (2019) contend that market-based measures take market factors into account and therefore reflect investors' expectations of future growth or the market´s reaction. Interestingly, Davis, Piger & Sedor (2012) find that nonfinancial disclosures such as the tone of earnings press releases also reflect market reactions and expectations of future growth. Consistent with this, they empirically test this association and deem that net optimistic tone in earnings press releases is positively associated with market-based future performance. The contradictory role of market-based and accounting-based measures is also evident through their association with corporate governance variables (Adjaoud et al., 2007; Yameen et al., 2019).

For instance, Adjaoud et al. (2007) find that corporate governance variables can differentiate between high- and low-performing firms better when firm performance is proxied by more holistic market-based estimates, as opposed to accounting-based estimates. In addition, Elvin & Bt Abdul Hamid (2016) empirically provide evidence that corporate governance and ownership structure variables are more relevant to market-based measures.

They explain this by suggesting that governance mechanisms have evolved to be more market oriented and are focused on futuristic value creation (Elvin & Bt Abdul Hamid, 2016). The notion that nonfinancial disclosures reflect market-based performance relatively more than accounting-based performance is especially pertinent in the Pakistani context.

Given the rising investor uncertainty in Pakistan, it would be interesting to test whether the market responds relatively more to nonfinancial disclosures, as suggested by the relevant literature. This distinction has important implications for policymakers and regulators in this context, as it would help establish the importance of nonfinancial disclosures to investors in the market. This is important for the restoration of investor confidence in a setting that is plagued by economic instability (Ullah & Saqib, 2018). Following from that, we form the following hypothesis:

H3: Narrative disclosure tone improves the ability of financial disclosures to predict market-based measures of firm performance relatively more than accounting-based measures.

H4: Corporate governance disclosures improve the ability of financial disclosures to predict market-based measures of firm performance relatively more than accounting-based measures.

3. Data and Methodology

3.1 Data collection and sample

The data are extracted through the annual reports of 125 nonfinancial firms in Pakistan, which are downloaded from the firm websites. As suggested by Mousa et al. (2022), most studies relevant to the prediction of firm performance utilize a sample limited to financial companies or are limited to one sector. Therefore, they suggest studying a sample of firms covering a diverse set of sectors. Accordingly, our sample covers firms from several different sectors. For the operationalization of some variables such as firm risk, data are also obtained from the Pakistan Stock Exchange. Furthermore, the data span a timeframe of 10 years from 2011-2020. This time period is suitable because it marks the beginning of the post-global financial crisis era. As suggested by Harakeh, Leventis, El Masri & Tsileponis, 2022, global financial markets suffered a significant loss of investor confidence in the market during this time due to heightened economic uncertainty. Therefore, identifying ways to improve the prediction of firm performance during such a time is crucial to the restoration of investor confidence in financial markets. Furthermore, relevant data at the time of collection were publicly available until 2020. Therefore, the final sample consists of a total of 1250 annual reports, which constitutes a large sample. Mousa et al. (2022) specifically recommended using a larger sample size when conducting a study based on machine learning algorithms, and this study fits that criterion.

The predictor variables used in the proposed study are divided into two categories: nonfinancial disclosures and financial disclosures.

3.2 Nonfinancial disclosures as predictor variables

3.2.1 Disclosure tone

The first category of nonfinancial disclosures used as predictor variables in the study are represented by different disclosure tones operationalized through a sentiment analysis of the annual reports. For this purpose, we utilize the Loughran & McDonald (2011) dictionary, which is a widely popular resource to perform such an analysis in the accounting literature (Del Gaudio, Megaravalli, Sampagnaro & Verdoliva, 2020; Mućko, 2021; Mousa et al., 2022). Moreover, it is specific to research related to business areas, whereas other alternatives are more generalized (Mousa et al., 2022). In addition, Loughran & McDonald (2011) developed this dictionary by analyzing a comprehensive sample of both forms of annual reports (annual and quarterly) from 1994-2008. For these reasons, the LM dictionary is an appropriate tool to perform sentiment analysis.

According to Mućko (2021), sentiment analysis is a technique that classifies and quantifies emotional content within textual content into categories of emotions. It is also reliable in regard to our study, as it has been prevalent in any discussion relevant to disclosure tone (Del Gaudio et al., 2020; Mousa et al., 2022). Accordingly, we perform a sentiment analysis on our sample of annual reports based on the six categories of the LM dictionary and their respective list of words. In this way, the six categories of the LM dictionary are quantified to form our first category of six nonfinancial predictor variables, namely, positive, negative, uncertainty, litigious, superfluous and constraining.

3.2.2 Corporate governance mechanisms

The corporate governance mechanisms used as predictor variables in this study are chosen after a thorough analysis of the literature, as previously discussed. In total, we utilize twelve corporate governance variables associated with firm performance in the empirical literature, namely, board size, board independence, board gender diversity, board meetings, audit committee size, audit committee independence, audit committee gender diversity, audit committee meetings, institutional, foreign, managerial and concentrated ownership. The data for the operationalization of these variables are taken directly from the annual reports of the firm. These twelve corporate governance variables form our second category of nonfinancial predictor variables.

Therefore, the final list of nonfinancial predictor variables that are used in the study constitutes 6 disclosure tone variables and 12 corporate governance variables.

3.3 Financial disclosures as predictor variables

We utilize a total of six financial disclosures as predictor variables in this study, namely, firm age, firm size, leverage, firm risk, cash flow from operating activities and liquidity. All of these variables, barring firm risk, are sourced from the annual reports of the firms.

In summary, we have a total of 24 predictor variables used in the study for the prediction of firm performance. Of these, 18 are nonfinancial and 6 are financial. We use these variables as predictors of our target variables, which are represented by market and accounting-based proxies of firm performance.

3.4 Target Variables

For the target variables, two accounting-based and two market-based estimates of firm performance are utilized. Return on assets and return on equity represent the accounting-based estimates, while Tobin´s Q and market-to-book value are the market-based estimates. Once we have operationalized these variables, the next step is to form classes of each of these target variables. Specifically, we follow Mousa et al. (2022), where they classified a single target variable into three classes based on the upper quartile, the interquartile range and the lower quartile for prediction.

For instance, the data points lying within the upper quartile of a particular target variable are labeled high performing for that particular variable. The observations lying within the interquartile range are labeled mid-performing, and similarly, observations lying in the lower quartile are labeled low performing. This process is repeated for all four target variables separately. Consequently, we have these three classes for each of the four target variables. The target variables are operationalized and sourced through the annual reports.

All variables used in the study, their operationalization and source are summarized in Table 1.

Table 1: Variables, Operationalization and Source

Symbol	Definition	Operationalisation	Source
Panel A: Target variables			
ROA	Return on Assets	Net Income/Total Assets	Annual Report
ROE	Return on Equity	Net Income/Total Equity	Annual Report
Tobin´s Q	Tobin´s Q	Market Value of Total Assets/Total Assets Replacement Cost	Annual Report
MTB	Market to Book Value	Market Value/Book Value	Annual Report
Panel B: Financial predictor variables			
SIZE	Firm Size	Natural logarithm of Total Assets	Annual Report
LIQ	Liquidity	Current Assets/Current Liabilities	Annual Report
LEV	Leverage	Total Liabilities/Total Assets	Annual Report
AGE	Firm Age	The number of years the since the firm was formed	Annual Report
CFO	Cash flow from operations	Net cash flow generated from operating activities	Annual Report
BETA	Firm Risk	Covariance of the stock´s returns with the market return/ Variance of the Market return	Pakistan Stock Exchange
Panel C; Non-financial predictor variables			
Disclosure tone			
POS	Positive sentiment	The number of positive words in the annual reports	Annual Report
NEG	Negative sentiment	The number of negative words in the annual reports	Annual Report
UNC	Uncertain sentiment	The number of uncertain words in the annual reports	Annual Report
LIT	Litigious sentiment	The number of litigious words in the annual reports	Annual Report
SUP	Superfluous sentiment	The number of superfluous words in the annual reports	Annual Report

Symbol	Definition	Operationalisation	Source
CON	Constraining sentiment	The number of constraining words in the annual reports	Annual Report
Corporate governance			
BSIZE	Board Size	The number of directors on the board	Annual Report
BI	Board Independence	The proportion of independent directors on the board	Annual Report
BGD	Board Gender Diversity	The proportion of female directors on the board	Annual Report
BM	Board Meetings	The number of times the board meets in a year	Annual Report
ACSIZE	Audit Committee Size	The number of directors on the audit committee	Annual Report
ACI	Audit Committee Independence	The proportion of independent directors on the audit committee	Annual Report
ACM	Audit Committee Meetings	The number of times the audit committee meets in a year	Annual Report
ACGD	Audit Committee Gender Diversity	The proportion of female directors on the audit committee	Annual Report
IOWN	Institutional Ownership	The percentage of shares owned by institutions	Annual Report
MOWN	Managerial Ownership	The percentage of shares owned by managers	Annual Report
FOWN	Foreign Ownership	The percentage of shares owned by foreigners	Annual Report
COWN	Concentrated Ownership	The percentage of shared owned by shareholders having 5% or more shares	Annual Report

3.5 Preparing the best fit model using optimal feature selection

One of the most important steps of machine learning classification techniques is feature selection (Xiaomao, Xudao & Yuanfang, 2019). Features are another word for independent or predictor variables in the machine learning literature. Therefore, predictor variables are henceforth referred to as features. Feature selection works by filtering out irrelevant features for a particular target variable (Yeh & Chen, 2020). According to Yeh & Chen (2020), this avoids overfitting. In addition, this not only improves the simplicity of the model but also helps in its interpretation (Xiaomao et al., 2019). Simply put, feature selection is a process aimed at improving the accuracy of predictive models by identifying the most relevant predictor features for a particular target variable. After checking all the features in the study for multicollinearity via a VIF test and generating a correlation matrix, we conduct feature selection via the Boruta algorithm.

The Boruta algorithm works by using the random forest classifier and performing several iterations on the overall features (Mousa et al., 2022). It eliminates features that are relatively inconsequential for classification of the target variable at every iteration. As we have four target variables, we run the Boruta algorithm for each target variable separately. At the end of this process, we have four sub-datasets, with each sub-dataset containing a specific target variable and its most relevant features for prediction as identified by the Boruta algorithm. Finally, we further split each dataset into a training and testing dataset.

3.6 Splitting the dataset into two – Training and Testing

After performing the Boruta algorithm and optimally selecting our features, we split each of our final four sub-datasets into training and testing data. As in all machine learning prediction problems, the splitting of data into training and testing data is crucial (Yeh & Chen, 2020). Training data are a subset of the entire dataset that the machine learning algorithm uses to learn patterns and applies them to the test dataset for the purpose of prediction. For analysis specific to the prediction of future data from past historical data, the training data always precede the testing data with respect to time. Many have followed this course while building their training data, especially in research where forecasting is concerned (A. H. Moghaddam, M. H. Moghaddam & Esfandyari, 2016; Mousa et al., 2022). Therefore, we follow Mousa et al. (2022) in splitting our 10-year datasets into a 2011-2019 subsample as training data and data within 2020 as test data. This constitutes a total of 1250 observations split into 1125 observations for training and 125 observations for testing. This process is performed for each of the four datasets relevant to each target variable.

Finally, we apply suitable machine learning algorithms on the training data to train them and then accordingly on the testing data for prediction. Two machine learning algorithms are utilized for both training and testing in this study. They are described in the empirical framework below along with different models for the testing of hypotheses.

4. Empirical Framework

4.1 Algorithms

4.1.1 Random forest (RF)

The first supervised and ensemble learning method that we employ is random forest (hereafter referred to as RF). Chen, Li & Sun (2020) contend that RF is a popular technique for classification and maximizing purity. In addition, they state that RF builds a myriad of randomized decision trees using the training data. Accordingly, it works by partitioning the feature space of a decision tree at each node using various tests. This process is continued until all nodes of the decision tree contain samples of a single class (Chen et al., 2020). This is how the RF algorithm learns. In terms of predicting the test data, it can identify the output class given a set of inputs by utilizing what it has learned (Chen et al., 2020). Kim, Ku, Chang & Song (2020) state that it is able to predict the outputs in the test data by identifying the most commonly predicted class for a given set of inputs across decision trees during the training phase. According to van der Heijden (2022), RF is useful tool in accounting research. Moreover, it can be used for both classification and regression, it prevents overfitting and it is robust to missing data. The RF algorithm is run by utilizing the *randomForest* package and library in R.

4.1.2 Stochastic gradient boosting (SGB)

Stochastic gradient boosting (SGB) represents a powerful ensemble prediction method (Halteh, Kumar & Gepp, 2018). Unlike the RF method, it generates numerous decision trees in a more sequential manner. It then aggregates them to produce the most accurate model. Furthermore, as suggested by Halteh et al. (2018), SGB is robust to measures of the target variable being relatively inaccurate. The sequential nature of tree building in SGB allows it to learn extra information with the addition of each new tree (Sadorsky, 2021). Consequently, it helps SGB build an aggregate model with the highest accuracy. Several tuning parameters can be adjusted in an SGB model to find the optimal model. The SGB algorithm is run using *xgBoost* and the *caret* packages and libraries in R.

4.2 Models

To test our hypothesis, we form different models for prediction with each model distinguished by the set of features in it. Consequently, each of the below models is used to predict each target variable via both RF and SGB separately.

4.2.1 Model 1 (Financial features)

Model 1 only contains a set of financial features used for the prediction of firm performance and is utilized to predict each of the four target variables via both algorithms separately.

4.2.2 Model 2 (Financial and disclosure tone features)

Model 2 contains both disclosure tone and financial features utilized for the prediction of firm performance. Consequently, each of the four target variables is predicted utilizing the features in model 2 via both algorithms separately.

The comparison of model 2 and model 1 is utilized as a means to test H1. By comparing model 2´s prediction with model 1´s prediction of a particular target variable using a particular algorithm, we test whether disclosure tone features add to the predictive ability of financial features in regards to the prediction of firm performance. In total, we have a total of 8 comparisons of model 2 and model 1 as we predict four target variables using each of the two algorithms. If model 2 is a better predictive model than model 1 for a particular comparison, H1 is supported for that comparison. Note that model 2 and model 1 run using the same algorithm and for a particular target variable must be significantly different for the comparison between them to be valid. For that reason, following Mousa et al. (2022), a t test is employed.

4.2.3 Model 3 (Financial and corporate governance features)

Model 3 contains corporate governance features, in addition to financial variables used for the prediction of firm performance. Similar to previous models, each of the four target variables is predicted with respect to the features in model 3 utilizing both algorithms separately.

The comparison of model 3 and model 1 is utilized as a means to test H2. By comparing model 3´s prediction with model 1´s prediction of a particular target variable using a particular algorithm, we test whether corporate governance predictor features add to the predictive ability of financial features in the prediction of firm performance. If model 3 is a better predictive model than model 1 for a particular target variable using a particular algorithm, H2 is supported for that comparison. Note that model 3 and model 1 run using the same algorithm and for a particular target variable must be significantly different for the comparison to be valid. For that reason, following Mousa et al. (2022), a t test is employed.

4.2.4 Comparison of market-based and accounting-based measures of performance

For the testing of hypothesis 3, a comparison of model 2´s performance with respect to the prediction of market and accounting-based target variables is performed. As mentioned above, model 2 contains disclosure tone and financial features. In addition, model 2 predicts each target variable separately. The prediction of each market-based target variable is compared to the prediction of each accounting-based target variable using the features in model 2. These comparisons are repeated with respect to each algorithm separately. At the end of this process, we have a total of four comparisons between market-based and accounting-based target variables for each algorithm. For a particular comparison within the same algorithm, model 2´s performance with the two target variables being compared must be significantly different in order for the comparison to be valid. This is identified using a t test. Finally, H3 is supported if model 2 shows a greater improvement in the prediction of market-based measures of performance relative to accounting-based measures, for a particular comparison. This process is repeated with the features in model 3 for the testing of H4.

4.3 Parameters for comparing the predictive models

To determine the predictive power of these algorithms, we utilize some parameters commonly used in the literature (Mousa et al., 2022; Petropoulos, Siakoulis, Stavroulakis & Vlachogiannakis, 2020). Below, we briefly describe each of these metrics used in the study to evaluate these predictive models.

To evaluate a single model as a whole, we employ accuracy and the Kappa coefficient (Mousa et al., 2022). Specifically, accuracy measures the proportion of correct classifications and predictions, while the Kappa coefficient measures how frequently the model is performing when it is compared with itself by chance. These two measures are employed by Mousa et al. (2022) when they assess the performance of banking institutions through machine learning algorithms. In addition, the model´s significance is also monitored using a statistical test. The null hypothesis for the said test is that accuracy is equal to the no information rate (the highest proportion of the observed classes) while the alternate hypothesis is that accuracy is greater than the no information rate. Accordingly, if the null hypothesis of this statistical test is rejected, the model is significant. In addition, there are certain class-specific metrics that we utilize to analyze the performance of the classes individually.

The confusion matrices, as Mousa et al. (2022) suggest, are imperative to interpret classification problems, especially where there are more than two classes. The confusion matrices are utilized to visualize the classification process for each class. In addition, sensitivity and specificity are especially used for classes (Mousa et al., 2022; Petropoulos et al., 2020). According to Petropoulos et al. (2020), these measures remove any doubt of misinterpretation of model performance. Accordingly, these measures are utilized by Petropoulos et al. (2020) to evaluate the performance of machine learning algorithms in the prediction of bank insolvency. As explained by Mousa et al. (2022), for a given class, sensitivity reflects the percentage of acceptance of a correct classification, while specificity reflects the percentage of rejection of an incorrect classification. Having covered specific measures for classification, we also utilize measures for the prediction performance of each class. Consequently, we gauge performance by employing the measures Positive predicted value or PPV and Negative predictive value or NPV. PPV measures the percentage of acceptance of a correct prediction, while NPV measures the percentage of rejection of an incorrect prediction (Mousa et al., 2022). Finally, as suggested by Petropoulos et al. (2020), balanced accuracy, which is the mean of sensitivity and specificity, is also employed. With the help of all of these metrics, especially accuracy and the Kappa coefficient, we compare the models described in the study for the testing of hypotheses.

Finally, to determine the most important variables in each model for the prediction of each target variable using a particular algorithm, certain variable importance measures are utilized. Following Sadorsky (2021), we employ the mean decreased Gini metric for random forest. This metric can be generated using the *VarImp* function in the random forest package. Specifically, for stochastic gradient boosting, we use the same function in the Caret package. However, for SGB, the metric generated is one of relative importance, as it identifies the variables that contribute the most to the prediction across all trees (Halteh et al., 2018).

5. Results and Discussion

The results of the study are presented below in the following order. First, we present the descriptive statistics of all features used in the study. Second, we present the results pertaining to feature selection using the Boruta algorithm. Third, we compare the results of all three models pertaining to each target variable using both the RF and SGB algorithms for the testing of H1 and H2. Fourth, we compare the predictions of accounting-based and market-based estimates of performance for the testing of H3 and H4. Finally, we present a discussion of the findings.

5.1 Descriptive Statistics

The descriptive statistics of the overall sample are presented in table 2. In addition, we generate a correlation matrix and run a VIF test on all predictor variables. All correlation coefficients are generally low, and the mean VIF is 3.63, indicating no problems of multicollinearity.

Table 2: Descriptive Statistics

Variable	Mean	Median	SD*	Min*	Max*	1st Quartile	3rd Quartile	VIF*
SIZE	16.46	16.52	1.8	0	20.68	15.44	17.47	1.68
LIQ	1.16	0.9	1.17	0.01	14.29	0.51	1.38	1.18
LEV	0.97	0.47	2.93	0.00	25.05	0.04	1.10	1.07
AGE	40.46	37	18.9	4	107	25	56	1.13
CFO	4.03	0.72	1.54	-37.3	37.9	0.02	3.2	1.30
BETA	0.88	0.94	1.11	-25.38	8.79	0.45	1.35	1.04
NEG	426.9	337	288.7	0	2093	211	570.8	12.5
POS	290.8	225.5	214.7	0	1704	132	398.2	4.01
UNC	253	205	154.4	0	893	133.2	336	12.98
LIT	248.4	193	169.9	0	1267	128	324.8	5.9
SUP	5.875	4	8.18	0	89	1	7	1.64
CON	183.3	148	112.4	0	759	101.2	237	12.2
BSIZE	8.38	8	8.38	5	17	7	9	1.41
BI	0.18	0.14	0.13	0	0.86	0.11	0.28	1.62

Variable	Mean	Median	SD*	Min*	Max*	1st Quartile	3rd Quartile	VIF*
BGD	0.08	0	0.12	0	0.71	0	0.14	2.22
BM	5.56	5	2.34	2	22	4	6	1.3
ACSIZE	3.6	3	0.86	3	9	3	4	1.5
ACI	0.31	0.33	0.21	0	1.33	0.25	0.33	2.5
ACM	4.38	4	0.85	0	10	4	5	1.16
ACGD	0.08	0	0.15	0	0.75	0	0	2.02
IOWN	0.59	0.69	0.30	0.0	0.99	0.31	0.84	7
MOWN	0.18	0.06	0.24	0.0	0.89	0.0	0.29	5.27
FOWN	0.17	0.02	0.27	0.0	0.98	0.0	0.23	1.46
COWN	0.64	0.68	0.20	0.0	0.98	0.50	0.78	2.13

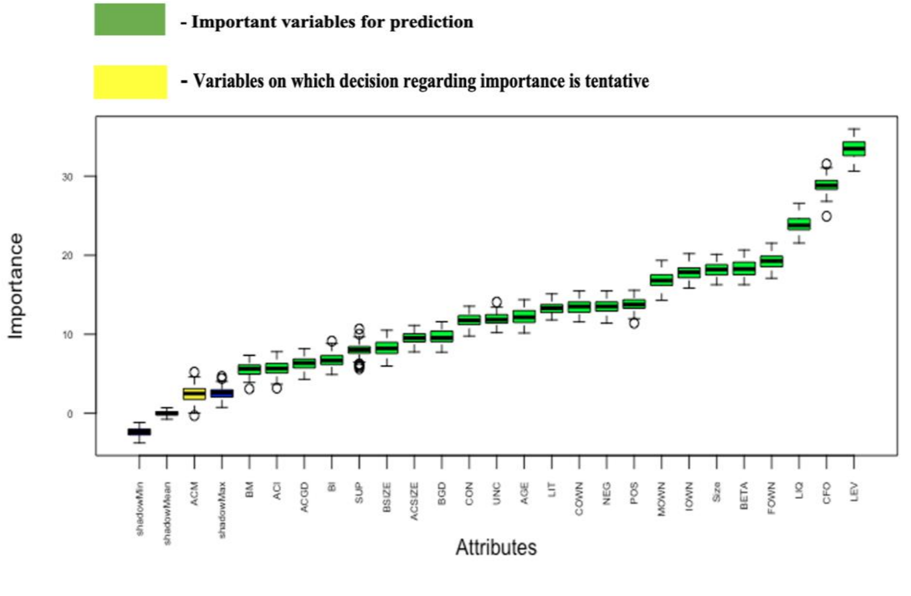
*SD: Standard Deviation. Min: Minimum. Max: Maximum. VIF: Variance Inflation Factor.

5.2 Feature selection using the Boruta Algorithm

5.2.1 ROA

The results of the Boruta algorithm for the prediction of ROA are shown in Figure 1.

Figure 1: Feature Selection using the Boruta Algorithm for the Prediction of ROA



As depicted by the color green, 23 attributes are confirmed as important predictors of ROA. However, ACM is depicted in yellow. This means that the Boruta algorithm does not have the desired confidence on this feature with the number of runs used, and its decision is tentative (Kursa & Rudnicki, 2010). Therefore, we follow Kursa & Rudnicki (2010) and use the Boruta package´s *TentativeRoughFix* function to make a decision on this feature. After performing the said function, ACM is deemed unimportant for the prediction of ROA and is accordingly eliminated. The remaining 23 features barring ACM are confirmed as important predictors of ROA.

5.2.2 ROE

Figure 2: Feature Selection using the Boruta Algorithm for the prediction of ROE

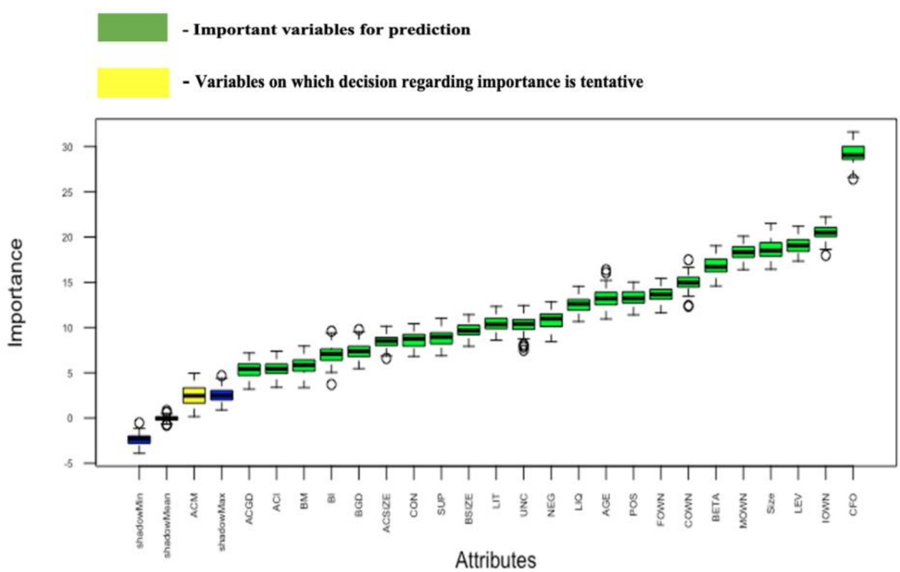
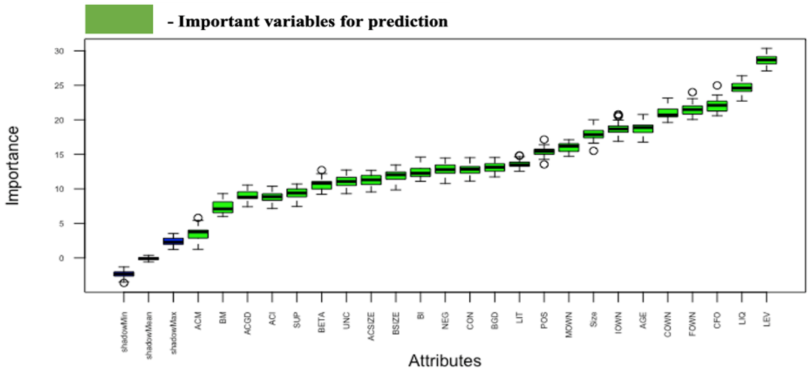


Figure 2 shows the results of the Boruta algorithm pertaining to the prediction of ROE. Similar to the results of ROA, all features used in the study are confirmed as important predictors of ROE (as depicted by the color green), except for ACM, which is tentative (as depicted by the color yellow). After using the *TentativeRoughFix* function to make a decision on ACM, it is deemed unimportant to the prediction of ROE and is accordingly removed. The remaining 23 features are confirmed as important predictors of ROE by the Boruta algorithm.

5.2.3 Tobin´s Q

Figure 3: Feature Selection using the Boruta Algorithm for the prediction of Tobin´s Q

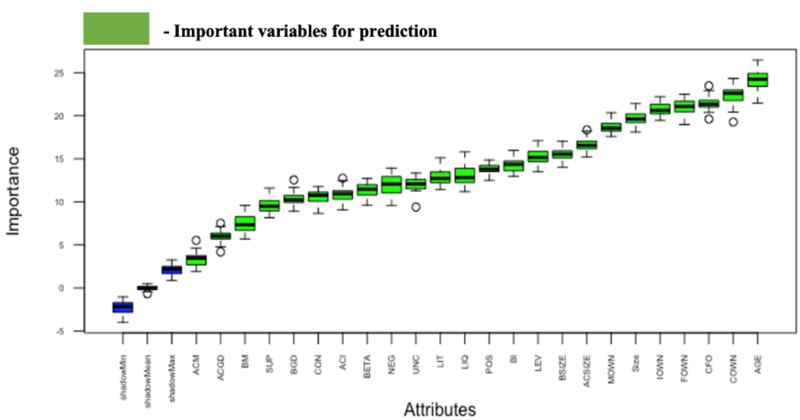


The results of the Boruta algorithm for the prediction of Tobin´s Q are shown in Figure 3, and as shown, all features are depicted in green and are confirmed as important predictors of Tobin´s Q.

5.2.4 MTB

Finally, we perform the Boruta algorithm to identify important predictors of MTB. The results are shown in Figure 4. All 24 features are depicted in green and are confirmed as important predictors of MTB.

Figure 4: Feature Selection using the Boruta Algorithm for the prediction of MTB



Therefore, after performing the Boruta algorithm on all four of our target variables, we proceed to their prediction using random forest and stochastic gradient boosting.

5.3 Performance comparison of models using RF and SGB algorithms

5.3.1 ROA

First, the confusion matrices for the prediction of ROA are shown in Table 3.

Table 3: Confusion matrices for the prediction of ROA

	Model 1			**Model 2**			**Model 3**		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
LOW	12	4	0	7	6	0	15	5	0
MID	17	63	6	23	64	6	15	67	8
TOP	1	12	10	0	9	10	0	7	8
Panel B: Stochastic Gradient Boosting									
LOW	10	4	0	11	7	0	16	11	0
MID	19	60	5	18	62	7	14	63	7
TOP	1	15	11	1	10	9	0	5	9

Confusion matrices of all three models for the prediction of ROA using the RF and SGB algorithms.

The results show that for all three models, classes Low and Mid represent relatively more correct classifications when we predict ROA. This is true for both predictions using both RF and SGB.

Table 4 shows the overall performance of each model with metrics relevant to the testing of hypotheses. RF´s prediction of ROA using model 1 achieve 68% accuracy and a kappa coefficient of 37%. However, the p-value for model 1 is 0.15, which indicates no difference between accuracy and the no information rate, rendering the model insignificant. Model 2 performs worse relative to model 1, as it achieves 65% accuracy and is also insignificant with a p value of 0.39. Interestingly, model 3 performs best as it is 72% accurate with a kappa coefficient of 43% and a p-value of 0.02. As model 3 performs better than model 1 and is statistically significant, H2 is supported. This implies that corporate governance features improve the prediction of firm performance when proxied by ROA and predicted using RF. The results regarding SGB´s prediction of ROA show follow a similar trend as predictions using both models 1 and 2 are insignificant with a p-value above 0.1. However, model 3 achieves 70% accuracy, a kappa coefficient of 42% and is significant with a p-value of 0.06. Therefore, our results regarding the prediction of ROA using SGB confirm our results using RF and H2 is supported for both algorithms. This implies that corporate governance features significantly improve the prediction of ROA using financial features alone.

Table 4: Overall metrics for the prediction of ROA

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.68	0.65	0.72	0.65	0.66	0.70
95% CI*	(0.59, 0.76)	(0.56, 0.73)	(0.63, 0.80)	(0.56, 0.73)	(0.57, 0.74)	(0.62, 0.78)
NIR*	0.632	0.632	0.632	0.632	0.632	0.632
p-value	0.15	0.39	0.02	0.39	0.32	0.06
Kappa coefficient	0.37	0.27	0.43	0.32	0.32	0.42
McNemar's p-value	0.01	NA	NA	0.00	0.09	NA

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of ROA using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

The class-specific characteristics are shown in Table 5. In model 1´s prediction using RF, the Mid class performs best with regard to sensitivity alone, while the Low class performs best in terms of specificity and PPV. Finally, the top class performs best in terms of NPV and balanced accuracy. Furthermore, models 2 and 3 using the RF algorithm achieve parallel results, barring a few exceptions. The class-specific metrics using the SGB for all three models also follow a similar pattern.

Table 5: Class specific metrics for the prediction of ROA

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.4	0.8	0.63	0.23	0.81	0.63	0.5	0.85	0.5
Specificity	0.96	0.5	0.88	0.94	0.37	0.92	0.94	0.5	0.93
PPV*	0.75	0.73	0.43	0.54	0.69	0.53	0.71	0.74	0.5

	Model 1			Model 2			Model 3		
NPV*	0.84	0.59	0.94	0.79	0.53	0.94	0.86	0.62	0.93
Balanced Accuracy	0.7	0.67	0.76	0.59	0.59	0.77	0.72	0.66	0.72
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.33	0.78	0.69	0.37	0.78	0.56	0.53	0.80	0.56
Specificity	0.94	0.5	0.87	0.93	0.46	0.90	0.95	0.50	0.94
PPV*	0.67	0.73	0.44	0.61	0.71	0.45	0.75	0.74	0.53
NPV*	0.82	0.58	0.95	0.82	0.55	0.93	0.86	0.66	0.93
Balanced Accuracy	0.68	0.65	0.75	0.65	0.62	0.73	0.72	0.67	0.72

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of ROA using RF and SGB algorithms.

The most important variables for the prediction of ROA in all three models using the RF algorithm are shown in Figure 5. LEV is the most important variable in all three models, followed by CFO and LIQ. However, in model 2, POS outranks AGE, while in model 3, all ownership structure variables outrank AGE, while IOWN outranks SIZE. This implies the importance of disclosure tone and corporate governance variables over certain financial variables. These results are similar for the SGB algorithm, barring certain exceptions, as CFO is the most important predictor of ROA in all three models, while both NEG outranks Age as important predictors in model 2, as shown in Figure 6.

Figure 5: Variable Importance - Random Forest – ROA

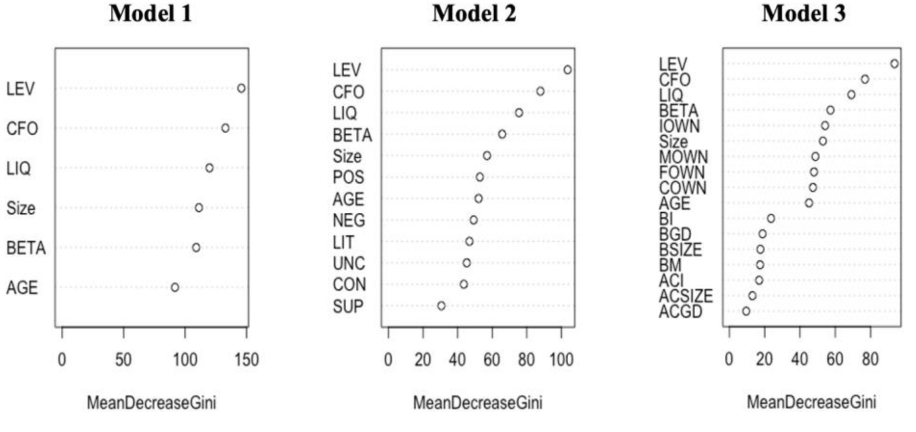
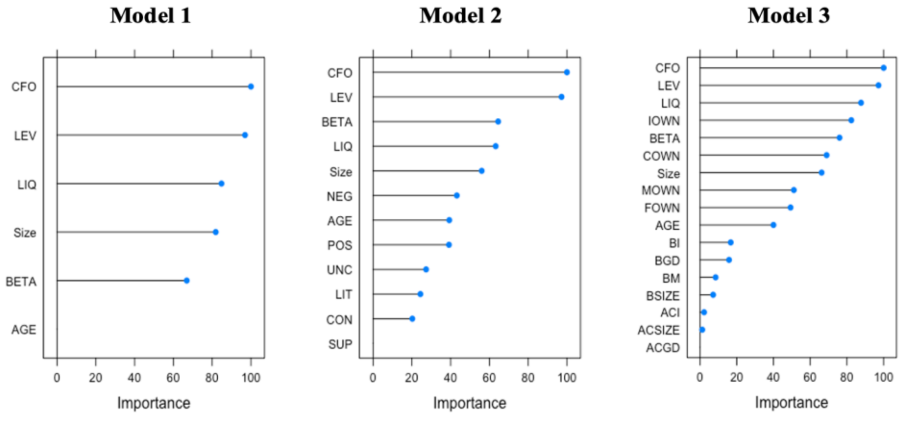


Figure 6: Variable Importance – Stochastic Gradient Boosting – ROA



5.3.2 ROE

Table 6: Confusion Matrices for the prediction of ROE

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
LOW	11	8	3	13	8	3	13	9	3
MID	13	54	12	11	55	12	13	53	7
TOP	2	9	13	2	8	13	0	9	18
Panel B: Stochastic Gradient Boosting									
LOW	11	7	5	11	10	3	16	13	5
MID	12	53	11	13	53	12	9	48	7
TOP	3	11	12	2	8	13	1	10	16

Confusion matrices of all three models for the prediction of ROE using the RF and SGB algorithms.

For ROE, each model´s confusion matrices generated using both algorithms are shown in Table 6. In all three models, all three classes have relatively more correctly classified predictions. This is specific to the RF algorithm. For the SGB algorithm, the Mid class alone has a relatively higher proportion of correct classifications in all three models, while the Top class also achieves the same in models 2 and 3. These results imply that both these algorithms perform relatively better when predicting ROE as most respective classes have a higher percentage of correct classifications using features in all three models.

Table 7: Overall metrics for the prediction of ROE

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.62	0.65	0.67	0.61	0.62	0.64
95% CI*	(0.53, 0.71)	(0.56, 0.73)	(0.58, 0.75)	(0.52, 0.69)	(0.52, 0.70)	(0.55, 0.72)
NIR*	0.568	0.568	568	0.568	0.568	0.568
p-value	0.12	0.04	0.01	0.21	0.16	0.06
Kappa coefficient	0.33	0.38	0.43	0.31	0.32	0.40
McNemar's p-value	0.61	0.69	0.26	0.61	0.71	0.27

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of ROE using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

As is evident from the overall metrics presented in Table 7, models 2 and 3 both significantly outperform model 1 in terms of accuracy, kappa coefficient and significance when predicted using the RF algorithm. This lends support to both H1 and H2. Specifically, adding disclosure tone to financial variables improves accuracy from 62% to 65% and the kappa coefficient from 33% to 38%, as is evident from the comparison between models 1 and 2. Moreover, adding corporate governance to financial variables improves accuracy from 62% to 67% and the kappa coefficient from 33% to 43%, as is evident from the comparison between models 1 and 3. Finally, both models 2 and 3 are significant while model 1 is insignificant. Therefore, it is clear from our results using the RF algorithm that both disclosure tone and corporate governance significantly improve the prediction of firm performance as proxied by ROE. However, when the prediction of ROE is performed using the SGB algorithm, both models 1 and 2 are insignificant, as their p values are greater than 0.1. However, model 3 is significant with a p-value of 0.06 and performs relatively better than both models 1 and 2 with respect to accuracy and the kappa coefficient. This lends support to H3. Therefore, the prediction of ROE using the RF algorithm supports both H1 and H2, while its prediction using the SGB algorithm supports only H2. Finally, the class-specific characteristics are shown in Table 8.

Table 8: Class specific metrics for the prediction of ROE

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.42	0.76	0.46	0.5	0.78	0.46	0.50	0.75	0.64
Specificity	0.89	0.54	0.89	0.89	0.57	0.90	0.88	0.63	0.91
PPV*	0.5	0.68	0.54	0.54	0.71	0.57	0.52	0.73	0.67
NPV*	0.85	0.63	0.85	0.87	0.66	0.85	0.87	0.65	0.90
Balanced Accuracy	0.66	0.65	0.68	0.69	0.68	0.68	0.69	0.69	0.78
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.35	0.77	0.39	0.42	0.75	0.46	0.62	0.68	0.57
Specificity	0.91	0.5	0.86	0.87	0.54	0.9	0.82	0.70	0.89
PPV*	0.5	0.67	0.44	0.46	0.68	0.57	0.47	0.75	0.59
NPV*	0.84	0.63	0.83	0.85	0.62	0.85	0.89	0.62	0.88
Balanced Accuracy	0.63	0.64	0.62	0.65	0.64	0.68	0.72	0.69	0.73

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of ROE using RF and SGB algorithms.

As far as the variable importance in these models is concerned, Figures 7 and 8 show that CFO is the most important feature in the prediction of ROE with regard to all three models using both algorithms. However, ownership structure variables show their importance in the prediction of ROE using both algorithms, as they outrank certain financial features in model 3. In addition, POS outranks AGE using the RF algorithm, while both NEG and POS outrank AGE using SGB. This implies that in the prediction of ROE, certain disclosure tone and corporate governance features are more important than certain financial features.

Figure 7: Variable Importance – Random Forest – ROE

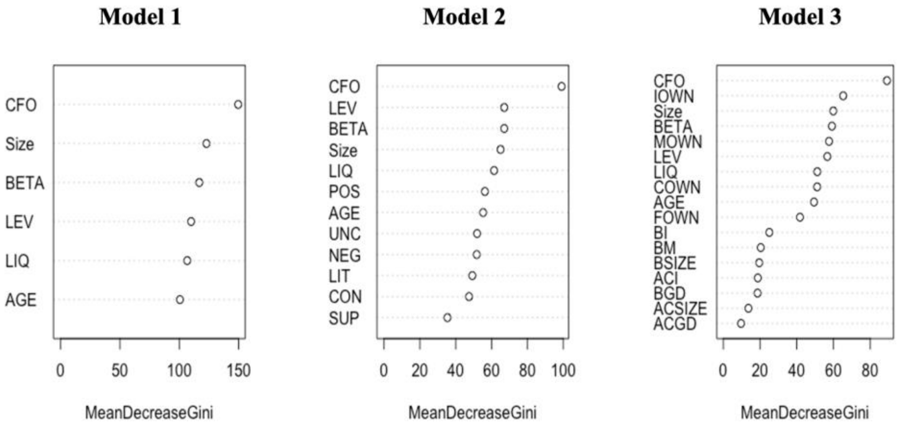
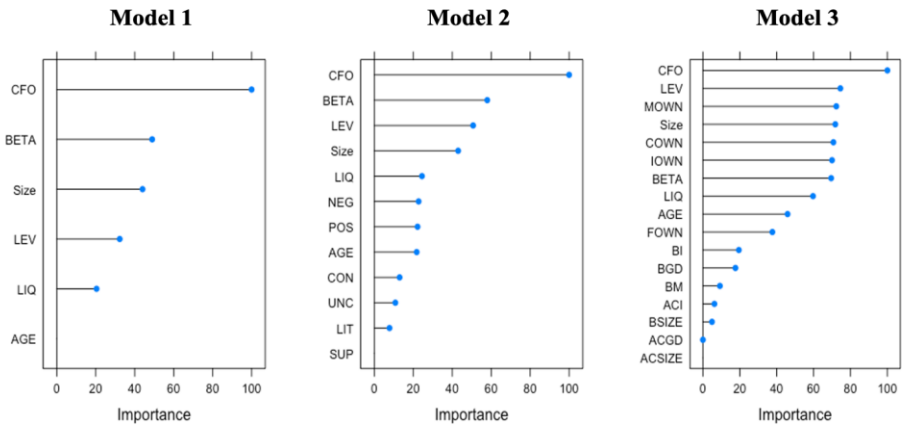


Figure 8: Variable Importance – Stochastic Gradient Boosting - ROE



5.3.3 Tobin´s Q

The confusion matrices of all three models for the prediction of Tobin´s Q generated using both algorithms are shown in Table 9. For model 1 to the RF algorithm, the Low and Mid classes have relatively more correct classifications. In models 2 and 3, all three classes achieve a relatively higher number of correct classifications. The confusion matrices generated for all three models using the SGB algorithm mirror these results.

Table 9: Confusion matrices for the prediction of Tobin´s Q

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
LOW	14	3	2	16	4	1	17	3	1
MID	23	47	13	21	51	14	23	51	10
TOP	3	10	10	3	5	10	0	6	14
Panel B: Stochastic Gradient Boosting									
LOW	10	2	2	17	6	3	24	3	3
MID	26	49	11	21	48	12	15	49	8
TOP	4	9	12	2	6	10	1	8	14

Confusion matrices of all three models for the prediction of Tobin´s Q using the RF and SGB algorithms.

Table 10 shows the overall metrics pertaining to all three models for the prediction of Tobin´s Q. Using the RF algorithm, model 1 achieves 57% accuracy versus model 2 and 3´s 62% and 66%, respectively. This trend of improvement can also be observed in the results of the kappa coefficient. All three models are significant with a p value of less than 0.05. Consistent with this, the results of the SGB algorithm show that models 2 and 3 are 60% and 70% accurate, respectively, while model 1´s 57% accurate. Similar to the prediction of the RF algorithm, this pattern is also evident for the kappa coefficient. All three models using the SGB algorithm are also significant with p values of less than 0.05. Therefore, the results using both algorithms pertaining to the prediction of Tobin´s Q lend support to H1 and H2.

Table 10: Overall metrics for the prediction of Tobin´s Q

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.57	0.62	0.66	0.57	0.60	0.70
95% CI*	(0.48, 0.66)	(0.52, 0.70)	(0.57, 0.74)	(0.48, 0.66)	(0.51, 0.69)	(0.61, 0.78)
NIR*	0.48	0.48	0.48	0.48	0.48	0.48
p-value	0.03	0.00	0.00	0.03	0.00	0.00
Kappa coefficient	0.28	0.35	0.42	0.27	0.33	0.50
McNemar's p-value	0.00	0.00	0.00	0.00	0.01	0.03

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of Tobin´s Q using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

The class-specific characteristics are presented in Table 11.

Table 11: Class specific metrics for the prediction of Tobin´s Q

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.35	0.78	0.4	0.40	0.85	0.40	0.43	0.85	0.56
Specificity	0.94	0.45	0.87	0.94	0.46	0.92	0.95	0.49	0.94
PPV*	0.74	0.57	0.43	0.76	0.59	0.56	0.81	0.61	0.71
NPV*	0.75	0.69	0.85	0.77	0.77	0.86	0.78	0.78	0.9
Balanced Accuracy	0.65	0.61	0.64	0.66	0.66	0.7	0.69	0.67	0.75
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.25	0.82	0.48	0.43	0.80	0.4	0.60	0.82	0.56
Specificity	0.95	0.43	0.87	0.89	0.49	0.92	0.93	0.65	0.91
PPV*	0.71	0.57	0.48	0.69	0.59	0.56	0.80	0.68	0.61
NPV*	0.73	0.72	0.87	0.77	0.73	0.86	0.83	0.79	0.89
Balanced Accuracy	0.6	0.62	0.68	0.66	0.65	0.66	0.76	0.73	0.74

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of Tobin´s Q using RF and SGB algorithms.

Finally, the most important variables for these models are shown in Figures 9 and 10. LEV is consistently the most important feature in all models predicting Tobin´s Q using the RF algorithm. Interestingly, POS outranks BETA in model 2, while all ownership structure variables barring MOWN overlap certain financial variables in model 3. The results for the SGB algorithm are similar, barring model 2 where LIQ outranks LEV as the most important predictor.

Figure 9: Variable Importance – Random Forest – TOBIN´S Q

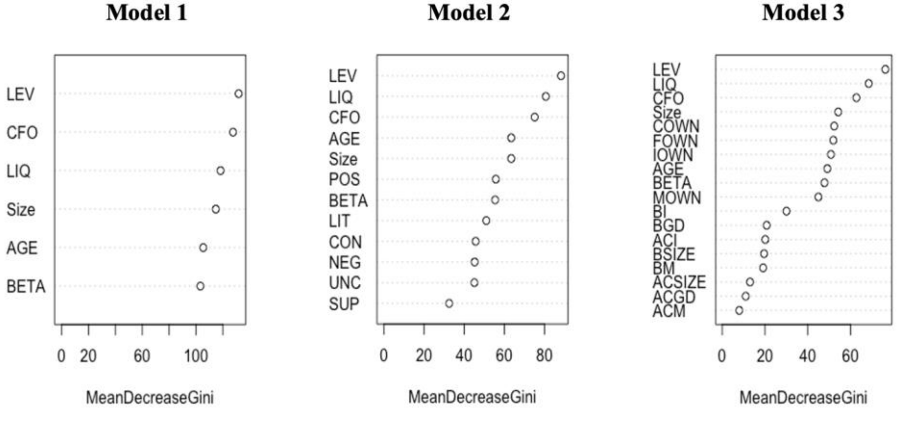
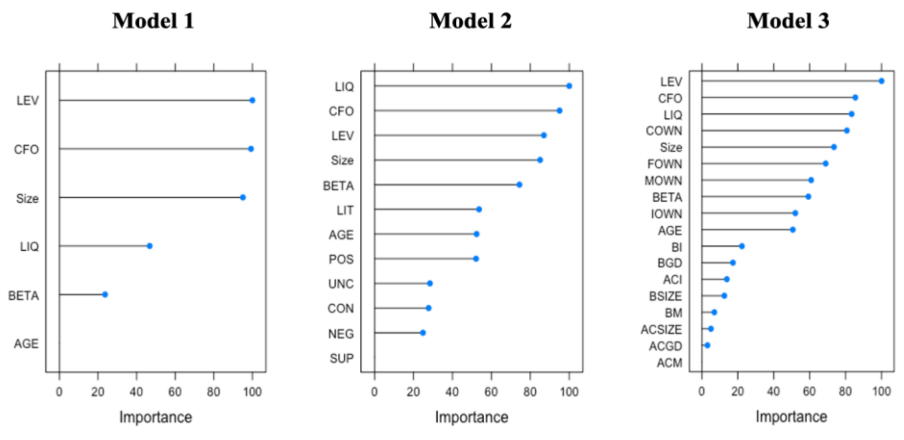


Figure 10: Variable Importance – Stochastic Gradient Boosting – TOBIN´S Q



5.3.4 MTB

Table 12: Confusion matrices for the prediction of MTB

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
LOW	15	5	2	16	3	2	21	4	0
MID	17	59	13	16	61	14	12	59	12
TOP	2	1	11	2	1	10	1	2	14
Panel B: Stochastic Gradient Boosting									
LOW	12	7	3	20	8	4	23	8	0
MID	18	54	11	13	54	10	9	55	9
TOP	4	4	12	1	3	12	2	2	17

Confusion matrices of all three models for the prediction of MTB using the RF and SGB algorithms.

In terms of MTB, the confusion matrices are presented in Table 12. Using both algorithms and all three models, all classes have a higher number of correct classifications than incorrect classifications. Consistent with this, all models achieve relatively accurate results and are also highly significant at 1%, as shown in Table 13.

Table 13: Overall metrics for the prediction of MTB

	Random Forest			Stochastic Gradient Boosting		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Accuracy	0.68	0.70	0.75	0.62	0.69	0.76
95% CI*	(0.59, 0.76)	(0.61, 0.78)	(0.67, 0.82)	(0.53, 0.71)	(0.60, 0.77)	(0.68, 0.83)
NIR*	0.52	0.52	0.52	0.52	0.52	0.52
p-value	0.00	0.00	0.00	0.01	0.00	0.00
Kappa coefficient	0.43	0.45	0.57	0.34	0.47	0.60
McNemar's p-value	0.00	0.00	0.00	0.04	0.08	0.09

*CI: Confidence interval. NIR: No information rate.
Overall metrics of all three models for the prediction of MTB using RF and SGB algorithms.
The comparison of models 1 and 2 tests H1 and the comparison of models 1 and 3 tests H2.

Using the RF algorithm, models 2 and 3 achieve 70% and 75% accuracy, respectively, relative to model 1´s 68%. Moreover, models 2 and 3 achieve kappa coefficients of 45% and 57%, respectively, relative to model 1´s 43%. Similarly, using the SGB algorithm, models 2 and 3 achieve 69% and 76% accuracy, respectively, relative to model 1´s 62%. In terms of the kappa coefficient, models 2 and 3 achieve 47% and 60%, respectively, relative to model 1´s 34%. Therefore, the results of both algorithms support H1 and H2, as both models 2 and 3 perform significantly better than model 1. The class-specific characteristics for the prediction of MTB using both algorithms are shown in Table 14.

Table 14: Class specific metrics for the prediction of MTB

	Model 1			Model 2			Model 3		
Panel A: Random Forest									
	LOW	MID	TOP	LOW	MID	TOP	LOW	MID	TOP
Sensitivity	0.44	0.91	0.42	0.47	0.94	0.38	0.62	0.91	0.54
Specificity	0.92	0.5	0.97	0.95	0.50	0.97	0.96	0.60	0.97
PPV*	0.68	0.66	0.79	0.76	0.67	0.77	0.84	0.71	0.82
NPV*	0.82	0.83	0.86	0.83	0.88	0.86	0.87	0.86	0.89
Balanced Accuracy	0.68	0.7	0.7	0.71	0.72	0.68	0.79	0.75	0.75
Panel B: Stochastic Gradient Boosting									
Sensitivity	0.35	0.93	0.46	0.59	0.83	0.46	0.68	0.85	0.65
Specificity	0.89	0.52	0.92	0.87	0.62	0.96	0.91	0.70	0.96
PPV*	0.55	0.65	0.6	0.63	0.70	0.75	0.74	0.75	0.81
NPV*	0.79	0.74	0.87	0.85	0.77	0.87	0.88	0.81	0.91
Balanced Accuracy	0.62	0.67	0.69	0.73	0.72	0.72	0.79	0.77	0.81

*PPV: Positive Predicted Value. NPV: Negative Predicted Value.
Class specific metrics of all three models for the prediction of MTB using RF and SGB algorithms.

Finally, the most important variables for all three models relevant to predictions of MTB utilizing both algorithms are shown in figures 11 and 12, respectively. AGE is the most important variable for the prediction of MTB in models 1 and 3 using the RF algorithm, while CFO outranks AGE in model 2. Consistent with previous results, however, POS outranks a financial variable (LIQ) as an important predictor of MTB, while all ownership structure variables appear to be the important corporate governance features, as they outrank certain financial features. However, using the SGB algorithm, as shown in figure 12, CFO is the most important variable for the prediction of MTB in models 1 and 2. In addition, COWN is the most important predictor of MTB in model 3, followed by FOWN and IOWN.

Figure 11: Variable Importance – Random Forest – MTB

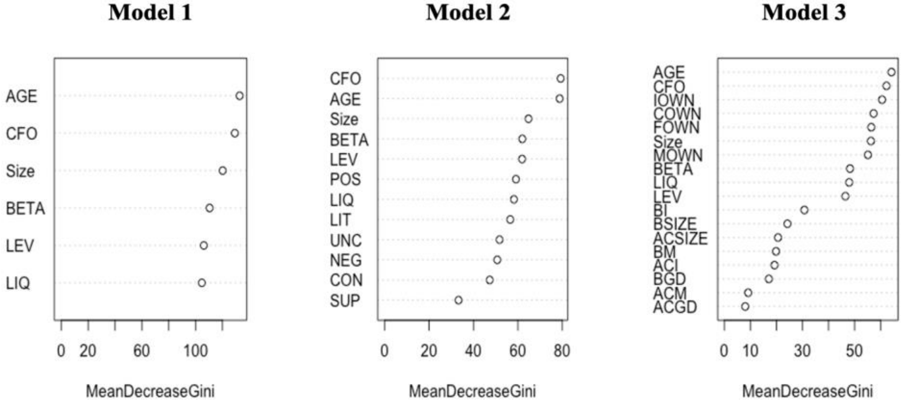
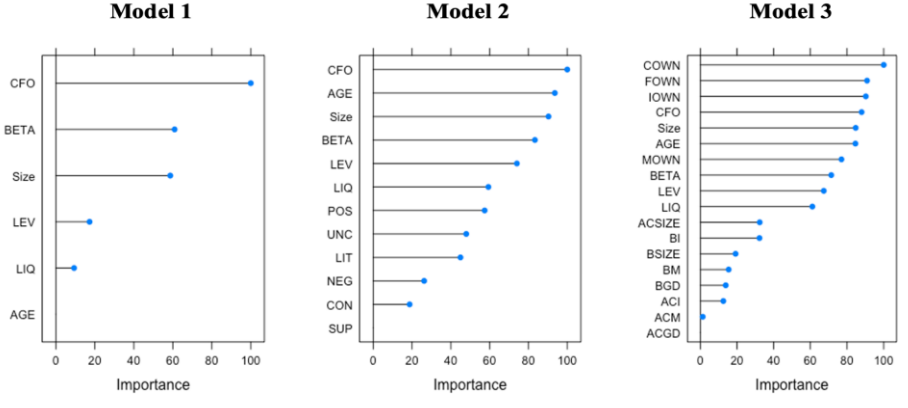


Figure 12: Variable Importance – Stochastic Gradient Boosting – MTB



5.4 Comparison of accounting and market-based performance estimates

For the testing of H3, model 2´s predictions of accounting-based estimates are compared with its predictions of market-based estimates using both RF and SGB algorithms. The comparisons are summarized in Table 15.

Table 15: Model 2´s prediction of market and accounting-based estimates

Panel A: Random Forest – Model 2				
	ROA	**ROE**	**Tobin´s Q**	**MTB**
Accuracy	0.65	0.65	0.62	0.70
95% CI*	(0.56, 0.73)	(0.56, 0.73)	(0.52, 0.70)	(0.61, 0.78)
NIR*	0.632	0.568	0.48	0.52
p-value	0.39	0.04	0.00	0.00
Kappa coefficient	0.27	0.38	0.35	0.45
McNemar's p-value	NA	0.69	0.00	0.00
Panel B: Stochastic Gradient Boosting – Model 2				
	ROA	**ROE**	**Tobin´s Q**	**MTB**
Accuracy	0.66	0.62	0.60	0.69
95% CI*	(0.57, 0.74)	(0.52, 0.70)	(0.51, 0.69)	(0.60, 0.77)
NIR*	0.632	0.568	0.48	0.52
p-value	0.32	0.16	0.00	0.00
Kappa coefficient	0.32	0.32	0.33	0.47
McNemar's p-value	0.09	0.71	0.01	0.08

*CI: Confidence interval. NIR: No information rate.
Comparison of model 2´s predictions of accounting and market-based estimates tests H3.

Specific to the RF algorithm, model 2 is insignificant when it predicts ROA, while model 2´s predictions of both Tobin´s Q and MTB are highly significant and achieve an accuracy of 62% and 70%, respectively. Therefore, this lends support to H3 that market-based estimates of firm performance are predicted better with the addition of disclosure tone to financial predictive models. Interestingly, model 2´s prediction of ROE is significant with 65% accuracy and consequently outperforms its prediction of Tobin´s Q. This contradicts H3. However, model 2´s prediction of MTB performs best when compared to its prediction of both ROE, as it is highly significant with an accuracy of 70%, lending further support to H3. This pattern is also evident in our results using the SGB algorithm as model 2´s prediction of both ROA and ROE are insignificant, while its predictions of both Tobin´s Q and MTB are highly significant.

Accordingly, these results provide support to H3 and consequently imply that narrative disclosure tones improve the prediction of market-based estimates relatively more than accounting-based estimates.

Similarly, for H4, we compare model 3´s prediction of both accounting-based estimates with that of both market-based estimates. The results are summarized in Table 16.

Table 16: Model 3´s prediction of market and accounting-based estimates

Panel A: Random Forest – Model 3				
	ROA	**ROE**	**Tobin´s Q**	**MTB**
Accuracy	0.72	0.67	0.66	0.75
95% CI*	(0.63, 0.80)	(0.58, 0.75)	(0.57, 0.74)	(0.67, 0.82)
NIR*	0.632	568	0.48	0.52
p-value	0.02	0.01	0.00	0.00
Kappa coefficient	0.43	0.43	0.42	0.57
McNemar's p-value	NA	0.26	0.00	0.00
Panel B: Stochastic Gradient Boosting – Model 3				
	ROA	**ROE**	**Tobin´s Q**	**MTB**
Accuracy	0.70	0.64	0.70	0.76
95% CI*	(0.62, 0.78)	(0.55, 0.72)	(0.61, 0.78)	(0.68, 0.83)
NIR*	0.632	0.568	0.48	0.52
p-value	0.06	0.06	0.00	0.00
Kappa coefficient	0.42	0.40	0.50	0.60
McNemar's p-value	NA	0.27	0.03	0.09

*CI: Confidence interval. NIR: No information rate.
Comparison of model 3´s predictions of accounting and market-based estimates tests H4.

Using the RF algorithm, comparing model 3´s prediction of ROA and Tobin´s Q provides evidence against H4. Specifically, the prediction of ROA is 72% accurate, significant at 5% and has a kappa coefficient of 43%. However, the prediction of Tobin´s Q is only 66% accurate with a kappa coefficient of 42%. A similar pattern is evident when the prediction of Tobin´s Q is compared with that of ROE, providing further evidence in contradiction to H4. Furthermore, model 3´s prediction of MTB using the RF algorithm performs best in terms of all overall metrics relative to its predictions of both ROA and ROE. These results lend support to H4. The results using the SGB algorithm also provide support for H4, as model 3´s predictions of both market-based estimates clearly outperform its predictions of both accounting-based estimates in terms of accuracy, the kappa coefficient and significance.

5.5 Summary and discussion of findings

In summary, our results via random forest indicate that corporate governance mechanisms improve the prediction of firm performance when proxied by both accounting and market-based measures. However, narrative disclosure tone improves the prediction of firm performance when proxied by ROE, Tobin´s Q and MTB only. For our results using stochastic gradient boosting, narrative disclosure tone significantly improves the prediction of market-based firm performance measures alone, while corporate governance mechanisms improve the prediction of both accounting and market-based estimates of performance. These results provide some valuable insights.

First, in terms of narrative disclosure tone, the results of the study are consistent with the limited literature in this regard (Beretta et al., 2021; Mousa et al., 2022). For instance, Beretta et al. (2021) empirically prove that disclosure tone captures incremental information about a firm´s ESG performance in the context of the global automotive industry. They explain this by suggesting that firms are now more aware that misreporting can have negative consequences. In addition, Mousa et al. (2022) also empirically provide evidence of disclosure tone improves the prediction of a firms' future performance specific to banking institutions in emerging markets. They justify this by suggesting that narrative disclosure tone contains incremental information regarding firm performance and stress the importance of narrative disclosures in the prediction of firm performance. Therefore, our results regarding disclosure tone contribute to the limited literature by providing further evidence of its increased importance in a developing economy. Moreover, our results are not limited to specific types of institutions, thereby indicating that narrative disclosure tone is indicative of firm performance in a diverse set of firms. In terms of theory, this result is explained by incremental information theory, which posits that managers signal value-relevant information about a firm's future performance through narrative disclosures (Arena et al., 2015; Beretta et al., 2021).

Second, relevant to corporate governance, there is an apt amount of empirical literature suggesting their significance to firm performance (Ciftci et al., 2019; Yameen et al., 2019). However, as Di Vito & Trottier (2022) and Mousa et al. (2022) point out, there is an increasing need to establish the reliability of corporate governance mechanisms as predictive tools of firm performance by utilizing machine learning algorithms in this context. To this end, our results successfully contribute, as they identify that corporate governance mechanisms disclosed in annual reports are imperative to the prediction of firm performance. Our results regarding corporate governance can be explained by agency theory (Azeez, 2015; Jensen & Meckling, 1976). As mentioned above, agency theory posits that better corporate governance mechanisms enhance firm performance. Accordingly, the theory suggests their role as important predictors of firm performance.

Finally, the comparison of market-based and accounting-based measures of performance in our study yields thought-provoking results. Interestingly, the results are skewed towards the notion that prediction using both narrative disclosure tone and corporate governance mechanisms is significantly better when they predict market-based performance. This suggests that the market responds well to nonfinancial disclosures. This is in synchronization with the results of Davis et al. (2012), who empirically prove that nonfinancial disclosures, such as the tone of earning press releases, are more market-oriented. Specific to corporate governance, our results are explained by Elvin & Bt Abdul Hamid (2016), who suggest that corporate governance and ownership structure variables have also evolved to be more market oriented. Furthermore, our overall results are also explained by the Pakistani context and provide practical implications for investors, regulators and policymakers.

As suggested above, the Pakistani financial market is plagued with heightened economic and political uncertainty (Ullah & Saqib, 2018). Therefore, investors rely on nonfinancial information in annual reports for any decision-making regarding investment in a firm (Aly et al., 2018). This is in synchronization with our results, as they suggest that market-based firm performance is better predicted with nonfinancial disclosures. These results represent an encouraging insight for the Pakistani market, and they provide several implications for investors, policy-makers and regulators.

First, our results provide empirical evidence that investors can safely use both narrative disclosures and corporate governance mechanisms as reliable predictive tools of firm performance. This is true for both accounting-based and market-based performance. This is especially important for the restoration of investor confidence in a setting with weak governance regulations and heightened uncertainty (Harakeh et al., 2022; Saeed et al., 2022; Ullah & Saqib, 2018). Second, our results also provide implications for regulators and policymakers, as they suggest that the market is actively responsive toward nonfinancial information contained in annual reports. Therefore, it is imperative for policymakers to focus on effectively strengthening the regulation of nonfinancial disclosures by firms. This is especially relevant, as the Pakistani market is characterized by a weak regulatory framework where these disclosure requirements are seen as a mere formality (Saeed et al., 2022; Ullah & Saqib, 2018).

6. Conclusion

The present study utilizes two widely popular machine learning algorithms, namely, random forest and stochastic gradient boosting, to test whether nonfinancial disclosures such as corporate governance mechanisms and disclosure tone improve the prediction of firm performance. In addition to nonfinancial variables, financial variables are also used as predictors of firm performance. Firm performance is proxied by two accounting-based measures (ROA and ROE) and two market-based measures (Tobin´s Q and MTB). Data are collected from the annual reports of 1250 nonfinancial firms in the emerging economy of Pakistan. Different predictive models are created and compared for hypothesis testing. Model 1 contains financial variables only, model 2 contains both financial and narrative disclosure tone variables, and model 3 contains both financial and corporate governance variables as predictors. Our results indicate that both narrative disclosure tone and corporate governance disclosures significantly improve the prediction of firm performance, especially market-based firm performance. The study contributes to the literature by first addressing the neglect of narrative disclosure tone in relation to the prediction of firm performance (Mousa et al., 2022). Second, the study contributes by amalgamating corporate governance with machine learning literature, which is a rarity in the literature (Di Vito & Trottier, 2022). In doing so, we establish the importance of corporate governance mechanisms to the prediction of firm performance. Third, by using machine learning algorithms, we contribute to the scant machine learning literature in the realm of accounting and finance, consequently adding to the reliability of these techniques (Mousa et al., 2022). Fourth, the study contributes by employing machine learning techniques to identify ways to improve the prediction of nonfinancial firms, especially in an emerging economy (Mousa et al., 2022). Finally, the study contributes by exploring the contradictory role of market and accounting-based performance in its prediction (Yang et al., 2019).

Our results provide some valuable insights and important implications for investors, managers and policymakers of Pakistani firms. First, the study´s results can be useful to investors, regulators and policymakers alike. The results suggest that investors can use the narrative disclosures and corporate governance mechanisms disclosed in annual reports as important information to gauge where the firm is headed. Therefore, the study outlines the imperativeness of nonfinancial disclosures in making better investment decisions. Similarly, the study has implications for managers to focus on improving the disclosure of nonfinancial information in annual reports. Furthermore, these results offer insights for regulators and policymakers to strengthen the disclosure requirements relevant to nonfinancial disclosures as these are deemed important to investors. This is especially important for Pakistan and other emerging economies with heightened economic uncertainty and a weak regulatory framework. Furthermore, this will help policymakers prevent adverse financial meltdowns by accurately anticipating them and consequently restoring investor confidence.

In addition, the results provide some implications for research, as they add to the reliability of machine learning algorithms as predictive tools of firm performance. Therefore, researchers are encouraged to use these algorithms and the study´s framework for improving the prediction of other financial outcomes such as bankruptcies, insolvencies, and crises. Furthermore, the study´s results strongly validate incremental information and agency theory perspectives. By doing so, they especially add to the reliability of incremental information theory, which has limited empirical significance. Despite having strong implications for both research and practice, the study is not without its limitations.

First, the study is limited to only one emerging economy due to a lack of available data. Future studies could incorporate more emerging economies into their analysis. Second, the study is restricted to narrative disclosures that are found in annual reports alone, whereas they are not the only mediums through which firms disclose textual information. Future studies could use other sources of content, such as earnings press releases, for the operationalization of disclosure tone. Finally, the study is limited to board and audit committee characteristics, along with some ownership structure variables as corporate governance disclosures. Future studies could employ other corporate governance characteristics, such as that of the risk committee in their analysis.

References

Adjaoud, F., Zeghal, D., & Andaleeb, S. (2007). The effect of board's quality on performance: A study of Canadian firms. *Corporate Governance: An International Review*, *15*(4), 623–635. https://doi.org/10.1111/j.1467-8683.2007.00592.x

Aerts, W. (2015). Narrative Accounting Disclosures. *Wiley Encyclopedia of Management*, 1–7. https://doi.org/10.1002/9781118785317.weom010055

Ahmed, E., Alabdullah, T., Thottoli, M., & Maryanti, E. (2020). Does Corporate Governance Predict Firm Profitability? an Empirical Study in Oman. *The International Journal of Accounting and Business Society*, *28*(1), 161–177. https://doi.org/10.21776/ub.ijabs.2020.28.1.7

Aly, D., El-Halaby, S., & Hussainey, K. (2018). Tone disclosure and financial performance: evidence from Egypt. *Accounting Research Journal*, *31*(1), 63–74. https://doi.org/10.1108/ARJ-09-2016-0123

Arena, C., Bozzolan, S., & Michelon, G. (2015). Environmental reporting: Transparency to stakeholders or stakeholder manipulation? An analysis of disclosure tone and the role of the board of directors. *Corporate Social Responsibility and Environmental Management*, *22*(6), 346–361.

Azeez, A. A. (2015). Corporate governance and firm performance: evidence from Sri Lanka. *Journal of Finance*, *3*(1), 180–189.

Beretta, V., Demartini, M. C., Lico, L., & Trucco, S. (2021). A tone analysis of the non-financial disclosure in the automotive industry. *Sustainability (Switzerland)*, *13*(4), 1–16. https://doi.org/10.3390/su13042132

Caserio, C., Panaro, D., & Trucco, S. (2020). Management discussion and analysis: a tone analysis on US financial listed companies. *Management Decision*, *58*(3), 510–525. https://doi.org/10.1108/MD-10-2018-1155

Chang, C. S., Yu, S. W., & Hung, C. H. (2015). Firm risk and performance: the role of corporate governance. *Review of Managerial Science*, *9*(1), 141–173. https://doi.org/10.1007/s11846-014-0132-x

Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, *365*, 112395.

Ciftci, I., Tatoglu, E., Wood, G., Demirbag, M., & Zaim, S. (2019). Corporate governance and firm performance in emerging markets: Evidence from Turkey. *International Business Review*, *28*(1), 90–103. https://doi.org/10.1016/j.ibusrev.2018.08.004

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research*, *29*(3), 845–868. https://doi.org/10.1111/j.1911-3846.2011.01130.x

Del Gaudio, B. L., Megaravalli, A. V, Sampagnaro, G., & Verdoliva, V. (2020). Mandatory disclosure tone and bank risk-taking: Evidence from Europe. *Economics Letters*, *186*, 108531.

Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, *40*(10), 3970–3983. https://doi.org/10.1016/j.eswa.2013.01.012

Di Vito, J., & Trottier, K. (2022). A Literature Review on Corporate Governance Mechanisms: Past, Present, and Future*. *Accounting Perspectives*, *21*(2), 207–235. https://doi.org/10.1111/1911-3838.12279

Elvin, P., & Bt Abdul Hamid, N. I. N. (2016). Ownership structure, corporate governance and firm performance. *International Journal of Economics and Financial Issues*, *6*(3), 99–108. https://doi.org/10.1108/10867371211266937

Florio, C., & Leoni, G. (2017). Enterprise risk management and firm performance: The Italian case. *The British Accounting Review*, *49*(1), 56–74. https://doi.org/10.1016/j.bar.2016.08.003

Halteh, K., Kumar, K., & Gepp, A. (2018). Financial distress prediction of Islamic banks using tree-based stochastic techniques. *Managerial Finance*, *44*(6), 759–773. https://doi.org/10.1108/MF-12-2016-0372

Harakeh, M., Leventis, S., El Masri, T., & Tsileponis, N. (2022). The moderating role of board gender diversity on the relationship between firm opacity and stock returns. *The British Accounting Review*, 101145. https://doi.org/10.1016/j.bar.2022.101145

Hunt, J. O. S., Myers, J. N., & Myers, L. A. (2022). Improving Earnings Predictions and Abnormal Returns with Machine Learning. *Accounting Horizons*, *36*(1), 131–149. https://doi.org/10.2308/HORIZONS-19-125

Ibhagui, O. W., & Olokoyo, F. O. (2018). Leverage and firm performance: New evidence on the role of firm size. *North American Journal of Economics and Finance*, *45*(March), 57–82. https://doi.org/10.1016/j.najef.2018.02.002

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360.

Kakanda, M. M., Salim, B., & Chandren, S. a/p. (2017). Corporate Governance, Risk Management Disclosure, and Firm Performance: A Theoretical and Empirical Review Perspective. *Asian Economic and Financial Review*, *7*(9), 836–845. https://doi.org/10.18488/journal.aefr.2017.79.836.845

Kim, S., Ku, S., Chang, W., & Song, J. W. (2020). Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. *IEEE Access*, *8*, 111660–111682.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, *36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Lambey, R., Tewal, B., Sondakh, J. J., & Manganta, M. (2021). The effect of profitability, firm size, equity ownership and firm age on firm value (leverage basis): evidence from the Indonesian manufacturer companies. *Archives of Business Review–Vol*, *9*(1).

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.

Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, *21*(41), 89–93. https://doi.org/https://doi.org/10.1016/j.jefas.2016.07.002

Moll, J., & Yigitbasioglu, O. (2019). The role of internet-related technologies in shaping the work of accountants: New directions for accounting research. *The British Accounting Review*, *51*(6). https://doi.org/10.1016/j.bar.2019.04.002

Mousa, G. A., Elamir, E. A. H., & 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.

Mućko, P. (2021). Sentiment analysis of CSR disclosures in annual reports of EU companies. *Procedia Computer Science*, *192*, 3351–3359.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, *36*(3), 1092–1113.

Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management*, *14*(2), 48.

Saeed, A., Ali, Q., Riaz, H., & Khan, M. A. (2022). Audit Committee Independence and Auditor Reporting for Financially Distressed Companies: Evidence From an Emerging Economy. *SAGE Open*, *12*(2), 21582440221089950.

Ullah, A., & Saqib, R. (2018). Doing Business in Pakistan: Managemental Challenges. *Journal of Management and Training for Industries*, *5*(2), 23–36. https://doi.org/10.12792/jmti.5.2.23

van der Heijden, H. (2022). Predicting industry sectors from financial statements: An illustration of machine learning in accounting research. *The British Accounting Review*, *54*(5), 101096.

Wang, J. (2016). Literature Review on the Impression Management in Corporate Information Disclosure. *Modern Economy*, *07*(06), 725–731. https://doi.org/10.4236/me.2016.76076

Xiaomao, X., Xudong, Z., & Yuanfang, W. (2019). A comparison of feature selection methodology for solving classification problems in finance. *Journal of Physics: Conference Series*, *1284*(1), 12026.

Yameen, M., Farhan, N. H., & Tabash, M. I. (2019). The impact of corporate governance practices on firm's performance: An empirical evidence from Indian tourism sector. *Journal of International Studies*, *12*(1), 208–228. https://doi.org/10.14254/2071-8330.2019/12-1/14

Yang, M., Bento, P., & Akbar, A. (2019). Does CSR influence firm performance indicators? Evidence from Chinese pharmaceutical enterprises. *Sustainability (Switzerland)*, *11*(20), 1–18. https://doi.org/10.3390/su11205656

Yeh, J.-Y., & Chen, C.-H. (2020). A machine learning approach to predict the success of crowdfunding fintech project. *Journal of Enterprise Information Management*.

The Lahore School of Economics was established in 1993 as a private, non-profit university with the goal of developing world class teaching and research in Pakistan. The objectives of the LSE are to prepare young Pakistanis to undertake research in economics, finance, banking, business management, industry, and development, in order to deepen their understanding of, and be able to productively contribute to, the major issues and policies that impact Pakistan and Asia at large.

The Innovation and Technology Centre (ITC) was established in April 2015 at the Lahore School of Economics with an aim to promote innovation, a key to growth in Pakistan. The ITC is a platform for academics, the business community and the public sector to collaborate in areas of economic and social importance including innovation and technology, macroeconomic and microeconomic constraints facing firms, productivity growth, manufacturing, export promotion, and environment sustainability. In addition to the internationally recognized academic output it produces every year, the ITC conducts annual surveys of manufacturers, exporters and policymakers on business confidence, technology adoption, innovation, and export competitiveness. The Centre enjoys a wide range of connections with top-level policymakers, the Chambers of Commerce of various major cities of Pakistan and manufacturers.

The ITC produces consumer reports, working papers and other outputs as part of the LSE's overall publication programme, which also comprises of the Lahore Journal of Economics, Lahore Journal of Policy Studies, Lahore Journal of Business, a textbook series, Lahore School Case Study Journal, the CREB Working Paper Series, and CREB Policy Paper Series. The LSE strongly encourages both in-house and external contributors.

Innovation and Technology Centre,
The Lahore School of Economics,
Intersection of Main Boulevard, Burki Road,
Phase VI, DHA, Lahore 53200, Pakistan
Tel: +92-(0)42-3656-0969
URL: https://itc.lahoreschool.edu.pk
Email: ayeshakh@lahoreschool.edu.pk