Text

Description automatically generated with low confidence

Module 12 Assignment: Capstone Group Project Report

INSTRUCTOR NAME: WADA, ROY

Group 7:

Vaidehi Patole

Chaitang Shah

Bhanoji Reddy Meka

NORTHEASTERN UNIVERSITY

MPS ANALYTICS

CLASS: ALY6980

CRN: 80841

Capstone Project

DATE: 07/02/23

**INTRODUCTION**

The objective of this report is to present our research project titled "An Empirical Examination of Board Types in Companies Based in the United States." The primary aim of this project is to investigate the structural implications of different board categories within companies and develop predictive models to classify them into specific categories, namely "ARISTOCRATIC," "DEMOCRATIC," "MONARCHY," "OLIGARCHY," "SOCIALIST," "STANDARD," and "TOTALITARIAN." This comprehensive report will encompass diverse facets of our study, including literature reviews, research methodologies, data processing, descriptive analysis, modeling techniques, and more.

This project is supported by Free Float Media as its sponsor. Currently, we have access to several data sources that can be categorized into three main aspects:

1. Individual directors: This includes information such as their names, educational backgrounds, estimated ages, genders, board tenures, roles within their respective companies, committee positions, and various shareholder percentages.
2. Firms: We gather data related to the company's board category, corporate parents, ownership category, sector, and other relevant details.
3. Director connectivity: We analyze the connections between directors, considering factors such as existing connections between any two directors, overlapping time periods of their engagements, and the number of third-degree connections they have.

These data sources provide us with a comprehensive understanding of the project's scope and allow us to conduct a thorough analysis of board dynamics within the context of the study.

The primary objectives of this project involve examining the potential changes in board categories using predictive variables and forecasting the likelihood of board categories, which will be publicly disclosed. Both tasks revolve around a central question: "When were their companies involved in events?" To address this question, we utilize precise information regarding the entry and exit dates of each director from the board. By comparing this data with the timing of board decisions and connections, we can readily tackle the first task.

However, the second task poses a challenge. There is no concrete evidence or proof to suggest that if a director was previously connected to a company's board, they would continue making decisions in the same manner for subsequent companies. Consequently, relying solely on historical behavior is not suitable for our research.

Nevertheless, directors who have a significant influence on board decisions and company outcomes may exhibit similar characteristics. For instance, a director with abundant resources that are valuable to their company (such as expertise, network connections, or stock shares) may differ in their ability to predict the company's category compared to a director with limited resources. Hence, our research aims to analyze how the characteristics of directors influence the board's category within their respective companies.

**LITERATURE REVIEW**

**Literature Review on the Relationship between Board Structure and Firm Performance**

**Introduction** The relationship between board structure and firm performance has been an area of intense research interest in the field of corporate governance. This literature review summarizes major theories, findings, and arguments from various studies, and provides a foundation for determining project goals in examining this relationship.

**Agency Theory** Many studies use agency theory as their theoretical basis. According to agency theory, the separation of ownership and control in firms can lead to conflicts of interest between shareholders and managers. The board of directors plays a crucial role in mitigating these conflicts and ensuring managers act in the best interests of shareholders, potentially improving firm performance.

**Board Composition** Board composition, particularly the proportion of independent directors, is a key element of board structure that has received significant attention. While some studies (e.g., Fama and Jensen, 1983) suggest that independent directors can enhance performance by providing effective oversight and reducing agency costs, others (e.g., Bhagat and Black, 2002) find little or no relationship between board independence and firm performance.

**Board Size** The relationship between board size and firm performance has also been a central focus. While some scholars argue that larger boards are more competent due to diversity (Dalton et al., 1999), others contend that they can lead to inefficiencies and communication problems, negatively affecting firm performance (Yermack, 1996).

**CEO Duality** CEO duality (when the CEO also serves as the board chair) is another element of board structure studied. While some researchers argue that CEO duality can facilitate strategic decisions and firm performance (Finkelstein and D'Aveni, 1994), others find that it can entrench management, reducing board independence and negatively impacting firm performance (Brickley et al., 1997).

**Gender Diversity** More recent studies focus on gender diversity on boards. While the moral and ethical argument for gender diversity is clear, its impact on firm performance is mixed. Some studies (e.g., Carter et al., 2010) show a positive correlation, others find no significant effect (Adams and Ferreira, 2009).

**Implications for Project Goals** This literature review provides valuable insights for my project. I can formulate my project goals based on gaps in the existing literature or contentious areas. For example, I could further explore the impact of gender diversity on firm performance or examine the effects of board structure on firm performance in a specific industry or geographical context. Additionally, I can investigate the role of other board characteristics (like resume power, tenure, or education level) in shaping firm performance.

**Conclusion** The literature review indicates that the relationship between board structure and firm performance is complex and depends on various factors. It highlights the importance of conducting further research to better understand these dynamics and to help companies design effective board structures.

**RESEARCH METHODOLOGY**

To accomplish our project goal, our research will encompass four key phases:

Phase 1: Assumptions and Hypotheses During this initial phase, we will establish a set of assumptions to overcome any limitations present in the original data sets. Additionally, we will formulate hypotheses based on the existing features related to individual directors, as well as new features that we will create. These hypotheses will be tested in Phase 3 and Phase 4.

Phase 2: Data Processing and Feature Engineering Given that the data provided by our sponsor originates from various sources and is structured at different levels (company and director levels), thorough data processing is essential prior to analysis. In this phase, we will integrate the data, clean it, and filter it based on the scope of our research. Moreover, we will generate new features from the processed data to suit the subsequent research process. These new features will include characteristics of directors, such as board categories, multiple directorships across companies, and connectivity. Additionally, we will utilize a secondary data structure to address historical features, such as influence percentage scores, in the modeling stage.

Phase 3: Descriptive Analysis Once Phase 2 is complete, we will conduct statistical analyses and employ visualizations to compare and showcase the distributions of various board categories, including "ARISTOCRATIC," "DEMOCRATIC," "MONARCHY," "OLIGARCHY," "SOCIALIST," "STANDARD," and "TOTALITARIAN." We will analyze the historical changes associated with these categories in relation to the characteristics identified, thereby gaining valuable insights for the subsequent phase.

Phase 4: Modeling Analysis In this final phase, we will develop predictive models to address the primary objective outlined by our sponsor: predicting the likelihood of a company's board category. Additionally, we will employ a logistic regression model to quantitatively evaluate the influence of specific characteristics on these categories. By undertaking this modeling analysis, we will provide valuable insights into the relationship between directors' characteristics and board categorization.

**RESEARCH ASSUMPTIONS**

The original data we have at hand poses limitations as it solely provides the yearly board category for each company. Consequently, we are unable to discern which variables significantly impact a company's board category, nor can we determine who should be held responsible for the indicated influence percentages. Moreover, the influence percentages themselves do not shed light on the subsequent consequences or outcomes. The intricate details and outcomes of board categories spanning multiple years may remain elusive since the effects of these variables can endure over an extended period. Furthermore, corporate boards involve confidential business secrets and strategic considerations.

For instance, while it may seem logical for the individual who bears greater importance to the company (possessing more resources that the company relies upon, etc.) to have a more significant influence on the board, it is plausible that a less critical director can exert influence over the company's board. This scenario may arise when the less critical director plays a crucial role in showcasing the company's commitment to embracing organizational development through tough decisions. Taking these factors into account, we have formulated the following assumptions:

Assumption 1: The highest influencers, as indicated by the yearly total counts across the board of directors for all companies, are considered influential factors.

Assumption 2: If a director is an insider and maintains a relationship with the CEO, their influence is classified as a categorical order from the company.

Assumption 3: If a director departs from the company during the year in which their former company's records are maintained, their active director status is regarded as a categorical order from the company.

These assumptions serve as foundational premises for our research, allowing us to explore the dynamics of board categories and the potential influences of directors within companies.

**EXPLORATORY DATA ANALYSIS**

This segment will outline the preliminary Exploratory Data Analysis (EDA) performed on the raw data using Tableau. This encompasses data modifications enacted to concentrate the analysis, which were executed using Excel formulas.

During the early stage of analysis, a decision was made to home in on all companies, since historical data regarding directors' influence is exclusively available for this group. Consequently, the EDA segment portrays an analysis strictly related to companies to define board categories. To facilitate this, we procured a spreadsheet encompassing the catalog of all companies listed by Free Float Media, which was subsequently utilized to categorize companies within the primary document, retaining only the relevant ones. This categorization was accomplished by merging files using the Inner Join feature, centered on the companies' ticker symbols.

In addition, the remaining files were fused together using the Inner Join feature, directed by companies' IDs (ISSUERID) and directors' IDs (INDIVIDUAL\_ID) depending on the document, with the intent to retain only the records of company directors who possessed historical records of share influence. Ultimately, this led to the consolidation of 9,877 records of directors and companies. Required data purging and corrections to data types were undertaken, including the substitution of zero values in the age column with the median value of 63.

In our study, the aim is to probe and discern the impact of directors' identities and their performance on the constitution of a company's board. We intend to comprehend how specific directorial attributes can mold and sway the assembly of various board types. By scrutinizing these attributes, we can acquire a clearer understanding of their role in the creation of board formations within a corporation. Further, a close examination of the performance of directors from diverse backgrounds enables us to understand the profound influence of their unique experiences and traits on the board.

The revelations from this research will offer meaningful perspectives to organizations, shareholders, and other stakeholders, elucidating the effect of directors' identities on the efficiency of the board, strategic decision making, risk mitigation, and the comprehensive performance of the corporation. The practical implications of this research for organizations lie in the arenas of board constitution, diversity, leadership cultivation, and succession planning, ultimately contributing to the formation of stronger, well-regulated companies.

A diagram of a company

Description automatically generated with low confidence

Subsequently, we employed visualization techniques to better comprehend the critical features in the primary document related to the board of directors, thereby gaining deeper insights. Figure 1 showcases a dashboard report that delivers a comprehensive view of the companies and directors. The number of companies that remain post-data transformation primarily belong to the financial, industrial, and IT sectors. The prevalent form of ownership is principal shareholder, trailed by widely held, and then controlled ownership.

In terms of educational background, the directors hailing from non-elite schools constitute 86.85% across all the evaluated companies, while a mere 13.15% are from elite schools. The maximum influx and exodus of board directors were observed in the industrial sector, with the least activity witnessed in the utilities sector. This was subsequently followed by an upsurge in the count of directors entering the board, thereby influencing its dynamics.

A screenshot of a graph

Description automatically generated with low confidence

The table shows that the number of board members varies between 4 and 34, with an average of 10 to 11 directors. When considering the share of influence, it evidently spans from 0 to 100%, with directors typically holding about 10% of the influence share within the board.

*Summary Statistics Table of Directors' Influence Share in US based Companies*

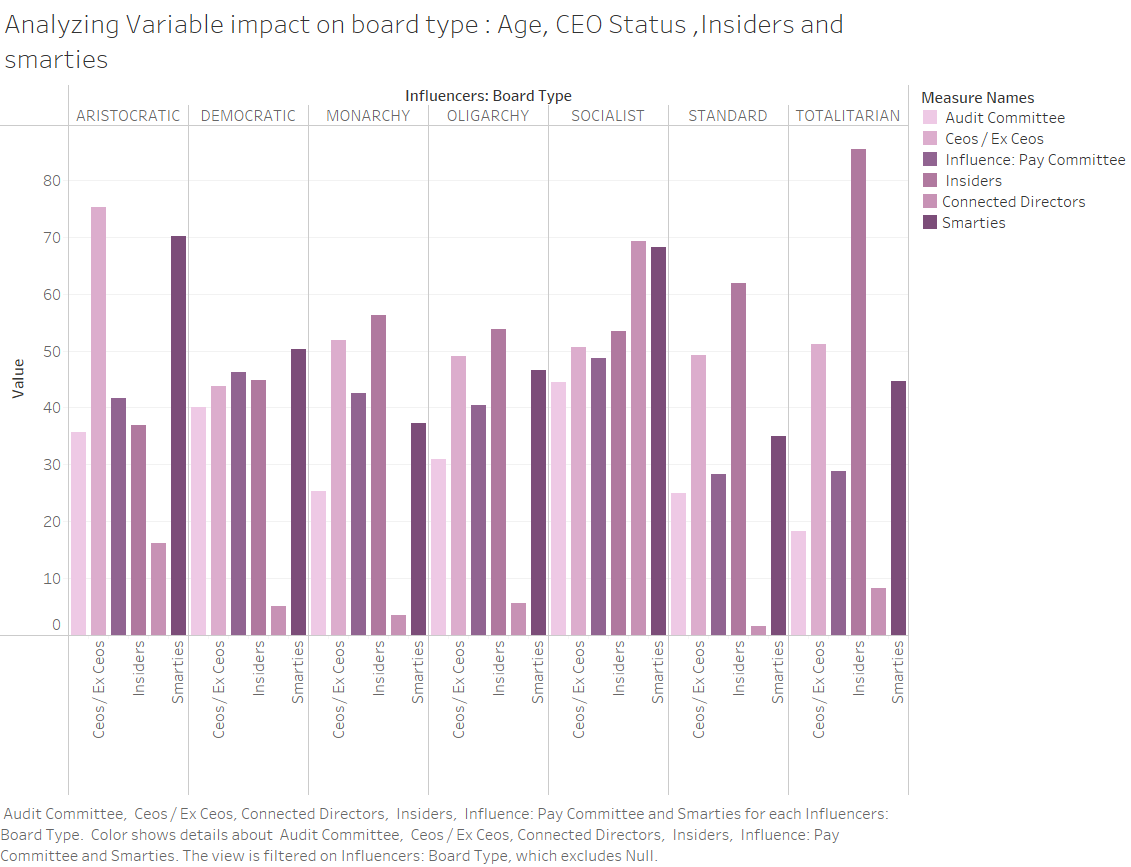
A picture containing text, number, font, screenshot

Description automatically generated

The graph unveils a clear trend concerning the impact of connected directors on varying board types. In the absence of connected directors, the standard or monarchy board type is predominant, whereas the socialistic board relies heavily on-board interconnections. Boards dominated by insiders typically manifest a totalitarian board type. Furthermore, in the aristocratic board type, CEOs and insiders wield substantial influence.

As the title indicates, the democratic board type displays a more equitable distribution across all variables, suggesting that numerous factors influence decision-making within this board type.

The socialistic board type stands out due to the high number of connected directors and individuals known as "smarties". These elements command a significant influence in determining the decision-making procedure within this board type.



**DATA PREPROCESSING**

This segment will illustrate the procedures we implemented for data manipulation, encompassing data transformation, amalgamation, feature engineering, and the refining of scope.

**Data transformation**

While the formats of the provided data sets are largely analogous—each director or company possesses a single record in the original data, which includes statistics on the board of directors, performance details, and so forth—the data concerning influence and interconnections amongst directors follows distinct structures. Each corporation has multiple entries with varying director statistics or from diverse time periods. Therefore, prior to the integration of these two data sets, we must adapt their structures to align with the other data sources.

Initially, we restructure the data pertaining to the highest influence percentage by transposing the values from separate rows for individual directors into distinct columns. This process yields data where each observation encapsulates the annual average influence score for a singular director. Given that we lack precise information on share price percentages or company rankings, and in line with our assumptions, we only require the directors' average scores over various years.

**Figure: Modified structure of Organization statistics**

A screenshot of a computer

Description automatically generated with low confidence

Upon restructuring the data, our next step involves merging all the requisite data sources to generate a comprehensive dataset. This dataset encapsulates all relevant information pertaining to directors, companies, influence percentages, connectivity, among other aspects (it's only through this process that we can execute feature engineering and scope refinement in subsequent sections).

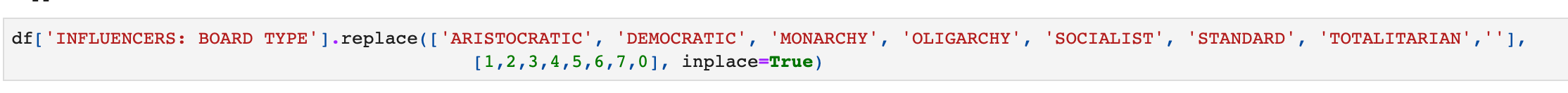
Following all integration procedures, the dataset now comprises 9587 observations of individual directors. Among these observations, 9342 records pertain to directors with historical influence values, 9582 directors possess weighted connectivity values, and 9467 directors exhibit third-degree connection values (here, the same directors serving on different companies' boards are considered distinct). The rationale for not employing the 'inner join' merging approach is to avoid losing data regarding directors who hold positions on other companies' boards. Applying 'inner join' would only yield overlapping observations from all data sources, thereby potentially causing us to forfeit valuable insights related to our hypothesis.

**Feature Engineering**

The feature engineering phase of our study is threefold: 1) crafting the target variable (categorizing the Board of Directors based on directors' statistics); 2) pinpointing directors' concurrent directorships spanning multiple companies; 3) recognizing the historical and current roles (CEO, lead director, and chairman) held by individual directors across various companies.

Holding the position of CEO or Chairman typically signifies substantial influence within a company. Yet, individuals who have previously held these roles could also wield considerable sway. Consequently, instead of solely focusing on current indicators of significant roles within the company, we also aim to identify if directors have previously or currently occupied these positions. By inspecting the non-null tenure values of these roles, we can effectively determine the indicators of such identities that directors have held or currently hold.

Converting Categorical variables in to numerical for target variable



Conversely, we also opted to employ an alternative strategy for handling the feature to gauge its exact impact on the Board Category, a characteristic previously outlined by our sponsor to determine "who should be held accountable?". At this juncture, we have identified from the target variable who bore the brunt of Directors' actions on the board and utilized the average value to assess individual directors at various temporal points. However, while this method proves efficient, it could potentially overlook certain insights concerning when the board of directors influences the type of board. Hence, we proposed a second data structure for managing the data.

**Data Cleaning**

Upon merging all necessary data sources, we currently possess over 87,842 observations of individual directors along with their associated companies. Nevertheless, owing to the limitations in the scope of controversy records, only 9,854 company boards have been categorized as Aristocratic, Democratic, Monarchy, Oligarchy, Socialist, Standard, Totalitarian. Additionally, numerous features, such as age, genders, identities, among others, are riddled with missing values and some categorical features exhibit extreme imbalance. Therefore, this segment will outline the process of data cleaning.

We merely need to filter out all records with NaN values in our target variable, as they were not factored into the target-generating process. This implies that these observations belong to directors whose companies do not have records of controversial events. Conversely, the second condition presents more complexity than the preceding one. Missing values are primarily found in categorical variables such as "CURRENT\_BOARD\_MEMBER," "CEO (or not)," "COMMITTEE\_MEMBER," among others. These values are absent because the respective directors do not occupy these roles, leading to blank entries in the original data. Therefore, we substitute these missing values with designated categorical values. Conversely, missing values in "GENDER" may occur when directors prefer not to disclose their gender. Thus, to simplify our research and avoid manual verification of missing genders, we will designate the missing genders with the categorical value "Unknown".

The attribute "AGE" includes numerous zero entries, which are clearly erroneous records. As such, we treat these zero entries as missing values as well. Given that different genders might present varying age distributions, we examine the age distribution sorted by gender. Due to the presence of outliers and skewness in all distributions, we use median values to individually replace the missing entries in "AGE".

A screenshot of a computer screen

Description automatically generated with low confidence

**Predictive Models**

This section details the initial modeling strategy employing the dataset consistent with the analysis carried out so far. The approach includes two methods: creating multiple models using the Auto Machine Learning (ML) technique and performing regression analysis via logistic regression models. Unnecessary features were excised from both methods, including elements tied to IDs, names, education, and other company-level details. Furthermore, models were constructed using 33 features, the target variable, and 9857 records. Of the 32 features, all were numerical.

**Performance Evaluation**

In response to our sponsor's request for the most accurate model to predict potential departures due to fraudulent activities, we are tasked with constructing multiple models. Rather than manually assembling each model, we opted to employ the Auto ML technique, which automates the process by swiftly constructing various ML models. In this scenario, we utilized numpy and scikit libraries for the classification problem, facilitating the creation of multiple models and proceeding with the most efficient one, showcasing the optimal parameters applicable to the data. This not only accelerates the process but also enhances performance accuracy by mitigating potential human errors or biases, thereby concentrating on the actual problem at hand. Moreover, it incorporates a setup function that permits the setting and modification of a wide range of parameters as needed, such as eradicating multicollinearity, normalizing, imputing, encoding, rectifying class imbalance, and even stipulating the method.

To begin with, we partitioned our dataset of 4,155 records into an 80/20 split, where 80% constitutes the seen data for Auto ML application, and 20% represents the unseen data for prediction purposes. This division aids us in evaluating the predictive prowess of the Auto ML technique. Consequently, we established models with 80% of the original dataset, amounting to 7576 records.

A picture containing text, screenshot, font, number

Description automatically generated

Subsequently, we established the environment within the Jupyter notebook by identifying the parameters required, leaving the remaining parameters at their default settings. By default, the setup function distributes the data into testing and training sets at an 80/20 ratio. In this instance, we set parameters to eliminate multicollinearity with a threshold of 0.9, rectify imbalance, and designated the target variable. Moreover, features will be selected based on their importance score using sklearn's SelectFromModel. Regarding the numeric features, missing values will be substituted with 0 as this method represents the most accurate portrayal of data. The output of the code displays all parameters employed for modeling, the target variable, and the size of the dataset.

Utilizing numpy, we first generated four distinct models for classification problems, excluding the four models with an AUC of 1. Subsequently, a table was produced, sorted by Accuracy score. This allows for a comparative analysis of different models based on their performance metrics. Consequently, the Random Forest model emerged as the superior model with the highest accuracy of 82%, while the Extra Gradient model excelled in differentiating Board categories, boasting the highest AUC of 83%. From the standpoint of an optimal balance between recall and precision, the Gradient Boost model performed the best. Therefore, each model carries its own set of strengths and weaknesses, and the selection should be influenced by the project's objectives and the problem that needs resolution.

**Regression analysis**

Subsequently, we utilized the statsmodels library to construct a Logistic Regression (LR) model, owing to its capability to generate a summary table that explicates the constant effect of each feature on a director's departure. The LR model, despite its comparably lower accuracy against other models such as Random Forest and Gradient Boosting, is more straightforward in its interpretation. Its application allows for a quantitative examination of the influence exerted by characteristics (features) on the target variable within our model. This will assist in fulfilling our sponsor's request to identify those characteristics of directors that amplify the likelihood of their departure from the board in the event of controversies.

**Data Preparation**

Upon distinguishing between the independent features (X) and target variables (y), we proceeded to transform the categorical variables in the former into dummy variables employing the get\_dummies() function and subsequently dropped the first category. This means that if a categorical variable possessed seven potential values, it would retain a column with values from 0 to 6, effectively converting the seven categories into numerical equivalents. This procedure aids in eliminating superfluous columns, thereby preventing multicollinearity. Moreover, we scrutinized the correlation and scores of the features to further reduce multicollinearity. The outcomes indicated that the variables "insiders" and "Highest influencers" registered the highest scores, exceeding the set cutoff score. Consequently, the exclusion of this variable contributed to the reduction in the scores of the remaining features.

A screenshot of a computer program

Description automatically generated with medium confidence

Subsequently, we constructed a Logistic Regression Model utilizing both sklearn and statsmodels, the latter of which facilitates a summary table that aids in the analysis of feature importance. The performance results from both were nearly identical. The accompanying figure presents the Confusion Matrix and Performance metrics of the model, which achieved an accuracy rate of roughly 74%. While this may not be the pinnacle of accuracy we aspire to attain, it is nonetheless a reasonably good performance given the limited data available at the director's level. Moreover, we observe that the model accurately predicted 73.77% of all data, with an error rate of 26.22%, yielding more False Positive values than True Positives.

*Confusion Matrix and Performance Metrics of Logistic Regression*

A screenshot of a computer

Description automatically generated with medium confidence

Next, we crafted an ROC curve to visually convey a summary of classifier performance. In an ideal scenario, if the classifier's curve skews closer to the top-left, beginning at one, then the model is regarded as having executed a perfect prediction. As seen in Figure 6-10, our plot appears reasonably satisfactory, albeit not perfect. The Area Under the ROC curve is a useful tool in assessing the efficacy of model classification. A perfect classifier would have an AUC = 1. In this instance, the AUC is 0.97, implying that 97% of the area under the ROC curve - depicted as a blue line in the plot - and therefore may be deemed a proficient classifier. Consequently, our model accurately classifies nearly 97% of values for the Aristocratic board type.

A picture containing text, line, screenshot, plot

Description automatically generated

**Random Forest Model:**

**A screenshot of a computer program

Description automatically generated with low confidence**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a graph

Description automatically generated with low confidence**

**Gradient Boosting Model:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A picture containing text, screenshot, font, software

Description automatically generated**

Primarily, we revisited the logistic regression model. The precision of this model stood at 72%, while the average prediction accuracy for the board category was 68%. There were merely 1894 data points supporting the target variable, constituting 20% of the entire dataset. In this scenario, the aggregate model accuracy underwent a subtle enhancement from 82% to 83% after the application of gradient boosting. Conversely, the precision in predicting target variables exhibited a significant surge from regression to the random forest model, a remarkable advancement. Given these alterations, we believe that gradient boosting constitutes a suitable technique for this situation.

In conclusion, among all evaluated models, the Random Forest exhibits superior performance. We propose the utilization of Random Forest, in conjunction with Gradient Boosting, for this data set. The latter method demonstrates an overall superior accuracy rate. Particularly in logistic regression, the proportional representation of insider influence bears a high coefficient. This secondary model imparts a unique perspective on the data. Ultimately, despite the numerous parallels drawn with the initial approach, it introduces an alternative thought process and data analysis strategy that corroborates our model construction and investigative efforts.

**Conclusion**

As the CEO or Chairman of an organization, holding the highest influence yields substantial positive effects on the Board type. The likelihood of their Board type being determined is 2 to 4 times greater when considering insiders, influential individuals, intelligent individuals, and audit committees.

On the other hand, the performance category or win rates of former CEOs, Chairmen, or lead directors have notably adverse consequences on departure outcomes. The accuracy of predicting their Board types is less than 10% compared to other variables considered.

Furthermore, serving as a member of nominating, pay, or audit committees moderately affects the Board type categories. Their predictions align with the actual Board types less than 40% of the time, as opposed to other variable insights.

Despite utilizing a duo of distinct prognostic methodologies, each has demonstrated impressive accuracy in the binary departure outcome prediction. Evaluating the multiple facets of model efficacy, encompassing precision, recall, AUC, and prediction duration, our endorsement leans towards the Random Forest classifier. This choice is not without reason. Firstly, the Random Forest classifier exhibits superior performance in terms of predictive accuracy. Secondly, it proficiently mitigates the issue of overfitting via the incorporation of a Gradient Boost model, a model that is already recognized for its excellent accuracy and harmonized performance in precision and recall.

The application of Gradient Boosting, a classification method, demonstrated the highest accuracy in predicting the board type of companies based on multiple variables and the influence of data structures. To further improve the predictive model, the following aspects should be considered:

**Addressing Imbalanced Distributions:** Overcoming the imbalance in independent variables' distributions.

**Variable Control:** Controlling variables related to sectors, capital levels, and other relevant factors. In our data manipulation process, we've confined our application to directors associated with companies incorporated in the US to prevent data attrition and potential discrepancies in board management strategies across diverse nations. However, due to data scope limitations, we did not segregate observations by varying company capital levels. It's plausible that billion-dollar companies may adopt different strategies or behaviors in director expulsion during contentious events compared to their million-dollar counterparts. This also applies to varying industries, such as finance and manufacturing. Essentially, this predicament should be viewed as an 'outlier' issue within our study. With access to more expansive data or observations, the introduction of variables controlling capital levels and industry sectors should be considered in the model construction phase.

**Combination of Processing Strategies:** Exploring the combination of two different strategies for processing IPS (Independent Variable Set) in various structures.

**Refining Assumptions and Target-Identifying Algorithm:** Modifying assumptions and the algorithm used to identify the target, incorporating additional and more detailed data sources.

**REFERENCES**

Barnhart, S. W., Marr, M. W., & Rosenstein, S. (1994, August). *Firm performance and board composition: Some new evidence*. Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1002/mde.4090150407>

A. A., D., & F. V. , B. (2010, August 19). Endogeneity and the relationship between board structure and firm performance: a simultaneous equation analysis for the Athens Stock Exchange. Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1002/mde.1492>

Martin, K. G. (2022, October 14). Why use odds ratios in logistic regression? The Analysis Factor. Retrieved December 5, 2022, from [https://www.theanalysisfactor.com/why- use-odds-ratios/](https://www.theanalysisfactor.com/why-%20use-odds-ratios/)

Srinivasan, S. (2005). Consequences of Financial Reporting Failure for Outside Directors: Evidence from Accounting Restatements and Audit Committee Members. Journal of Accounting Research, 2, 291–334. <https://doi.org/10.1111/j.1475-679x.2005.00172.x>