**Gender Diversity and Economic Growth**

**<Rachel Amey, rameyamey, ramey@udel.edu, data and literature manager>**

**<Priyanka Mondal, pmondal0402, pmondal@udel.edu, Methodology Manager >**

**<Alima Nurlan,alimanur,** [**alima@udel.edu**](mailto:alima@udel.edu)**, Visualisation manager>**

**<Abhin Suresh, abhinsuresh,** [**abhins@udel.edu**](mailto:abhins@udel.edu)**, Communication manager>**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Abstract:**

In corporate domains, gender discrepancies have ramifications for both company diversity and stock performance. Past literature demonstrates how companies run by female CEOs perform better over time; however, both the factors that determine whether companies are more likely to have a female CEO and the later company performance changes that result from these decisions have not been explicitly documented. Using machine learning algorithms, the present study classifies stocks that are run by male or female CEOs isolating factors that may be indicative of whether companies are more likely to employ a female CEO and their potential economic benefits.

**Introduction:**

Gender discrepancies have implications for both stock diversity and economic growth. For example, gender diversity in corporations has been shown to boost productivity, efficiency, and profit [10,16,20,21]. Companies led by underrepresented female CEOs make more money on average [12]. Thus, it is possible that increasing gender diversity in S&P 500 companies by hiring female CEOs may promote global economic growth. However, to tap into these potential economic benefits, research needs to identify key factors that may be unique to male or female owned CEO companies. Understanding these factors can help determine if there are stock characteristics that are more likely to predict a more gender diverse company and its ramifications on stock performance.

To predict whether companies are run by a male or female CEO and the resulting economic growth from that decision, factors that contribute to gender diversity and corporate profit need to be examined. Gender diversity within companies is contingent on numerous factors. Education level, the number of children the CEO has, and the general number of employees in the corporation play meaningful roles in determining the likelihood of having a female CEO [11]. Gender diversity can also be promoted by one’s social environment. According to social role theory, individuals are more likely to act according to the behaviors of those around them [13,6]. Companies that have headquarters located in areas with populations where there may be more women than men in the business sector and fewer stereotypical attitudes may be more prone to having a female CEO [14, 21]. Predicting economic growth also depends on myriad factors. For example, companies are inherently tied to overall market trends and trading strategies [27]. Thus, in order to model 1) the unique features of companies that decide to hire a female CEO, increasing company gender diversity and 2) overall economic growth as a result of that decision, all factors that predict both gender diversity and economic growth need to be included in the analyses.

The goal of this proposal is to use machine learning algorithms to classify which companies have male or female CEOs and whether or not that company’s value can increase or decrease as a result of this change. Previous studies have employed machine learning algorithms to predict economic growth specifically random forest models and an analysis of cumulative return value of stocks [9, 3, 23].

**Data collection.**

Data was collected from multiple sources and merged into a single data set. Factors include time, CEO gender, implicit and explicit attitudes towards women in corporate domains, and overall gender diversity in the company’s regional headquarters (S&P 500 websites, Project Implicit, U.S. Census Bureau). In order to ensure the variance in the present models was due to the main independent variables, other factors about the CEO and global market were included. In the models age, number of children, and education of the CEO were included in addition to company industry type and overall market price (S&P 500 websites, Kaggle). The dependent variable of interest in the classification model was whether or not a company was owned by a male or female CEO; the analysis of cumulative return value of each stock model requires difference of mean of two time series with dependent variable being the closing price of each stock every business day (Figure 1).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset name** | **URL** | **Number of rows** | **Number of columns** | **Number of relevant columns** | **Number of valid rows (not NaN on relevant columns)** | **Data type for each relevant column** |
| S&P500 Company Descriptions | https://docs.google.com/spreadsheets/d/1GORjuC8LVWmrXdmfmnBW6nujQ1rNzm3Q6osVtkwyjmY/edit#gid=0 | 506 | 9 | 4 | 506 | descriptive text |
| S&P 500 Price Time Series | Y finance | 2378325 | 5 | 3 | 2378325 | descriptive text (2) numerical (1) |
| Gender-Career\_IAT.public.2018 | https://drive.google.com/drive/u/0/folders/1gKncTYkiJKGdoMJ1cEc0DuLh\_0t\_qHZr | 346584 | 90 | 9 | 158643 | descriptive text(1) numerical (8) |
| Female CEO Timeline Expanded | https://github.com/rameyamey/MLTSAProject/tree/master/CEOdata | 66 | 11 | 7 | 66 | descriptive text(2) numerical (5) |
| Malefemalepercents and census data MSA | https://github.com/rameyamey/MLTSAProject/tree/master/censusdata | 66 | 17 | 10 | 57 | descriptive text(2) numerical (8) |
| RF2finalcoded.csv | https://raw.githubusercontent.com/rameyamey/MLTSAProject/master/RF2finalcoded.csv | 221 | 16 | 16 | 221 | numerical(16) |

**Methodology:**

First, to ensure the overall market activity is controlled for in the time series, principal component analysis (PCA) was conducted on the time-series stock data prior to any analyses. PCA allows the principle components of each stock that explain the most variance to be identified. By running PCA on all time series stock data, the overall nature of the market can be controlled for in the later analyses [8] (Figure 1). In order to classify which stocks may decide to hire a female CEO and to model the resulting economic growth from that decision, a random forest model was implemented in addition to an analysis of cumulative return value of each stock hiring new CEOs. Analyses revealed that the 1st PCA component corresponds to market trend driving all stocks in the data and therefore showing high correlation among all stocks of interest. The data cleaning process involved getting rid of contribution from market trend which also removed major correlation between two different stocks and the result is shown in Figure 1.

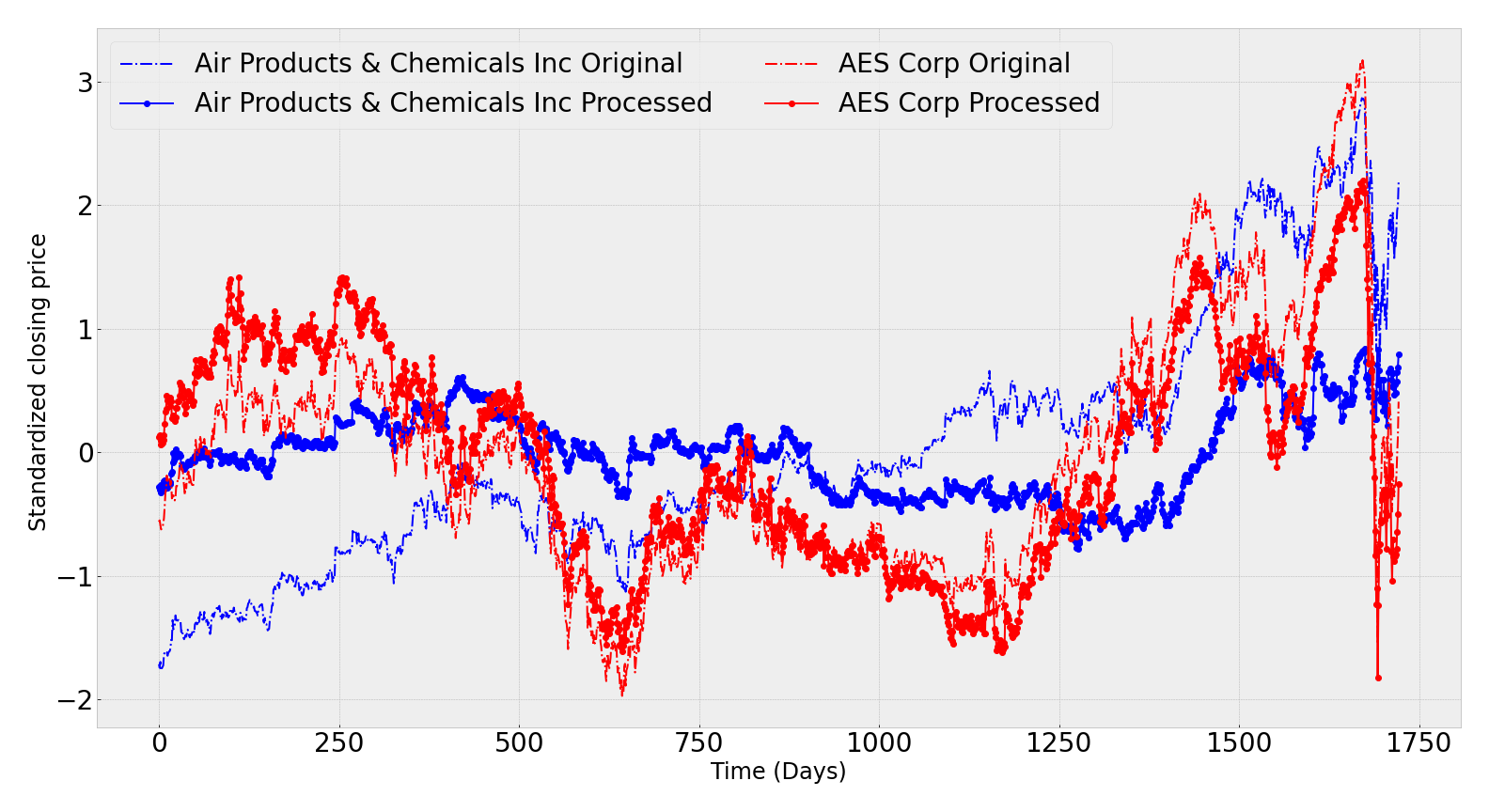


Figure 1. The figure above shows the time series of the first two stocks in the current database. The dashed lines indicate the original standardized stock price and the solid lines indicate the same once 1st PCA component contribution ( with variance ratio of 61 % ) is removed.

A random forest classifier was implemented to classify stocks based on whether they were run by a male or female CEO. This method was utilized because the features of interest varied across scaling and type. Random forest models are apt at handling this [30]. Random Forest models allow feature importance to be extracted. Thus, features that are most indicative of whether or not a stock is run by a male or female CEO can be extracted.

Prior to running the random forest model the present data had to be transformed. Utilizing real world national samples like implicit and explicit attitudes and specific details about individuals such as education levels and the number of children they have often lead to missing data as these values are not always readily available. To circumvent this KNN was utilized to fill in the missing data [18]. This method was used over mean or median replacement because it replaces the data with the most appropriate value instead of just a standard number, which often leads to a better model fit. The present model also had to account for the discrepancy between the number of male and female CEOs (there are only 32 stocks with female CEOs in the entire S&P 500 sample). Thus a bootstrap was implemented on training and testing datasets such that for each iteration an equal number of male CEO stocks was randomly sampled and classified against the female stocks [17]. The bootstrap ran for 1,000 iterations to obtain a training and test set distribution. The average of that distribution was extracted to see how well overall the given model features could classify whether or not the company was owned by a male or female CEO. The mean training accuracy of the training distribution was 88% while the mean test accuracy of the test distribution was 70% (Figure 2).

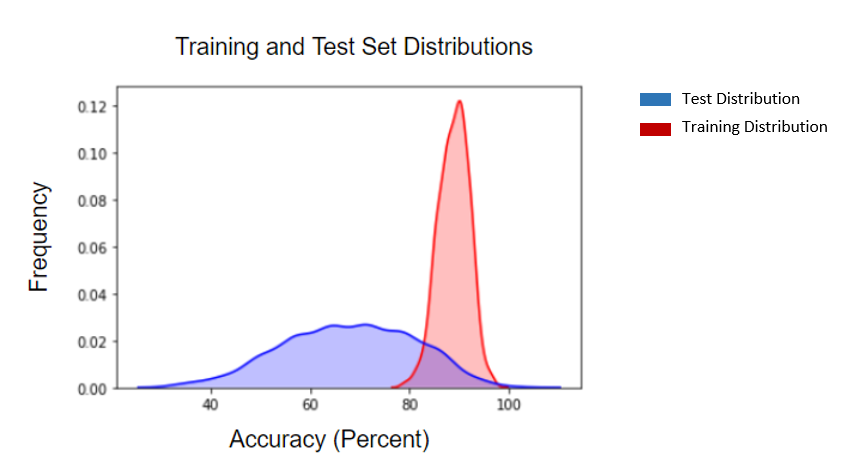


Figure 2. Training data is labeled in red while the testing data is labeled in blue. The training data reached an accuracy of 88% while the testing data reached an accuracy of 70%. One can notice much more variance in the testing distribution which is likely caused by features that were not implemented in the current model, i.e., there may be other features that were not accounted for that can influence whether a male or female owns a specific stock.

Feature classification revealed that the explicit stereotypic attitude towards women in corporate environments was the most meaningful feature in classifying whether a stock was owned by a male or female CEO (Figure 3). This is in contrast to previous work which highlights CEO age, education, and number of children highlighting how other features should be taken into account when understanding the features that dictate whether a stock may be owned by a male or female CEO [11].

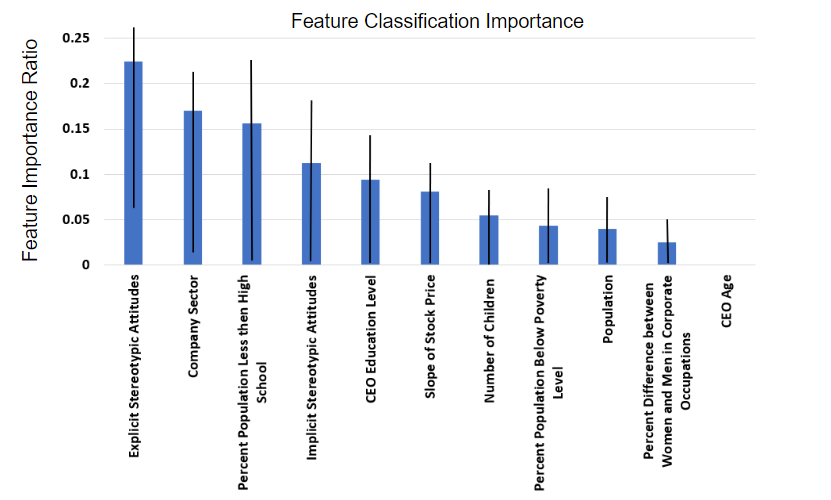
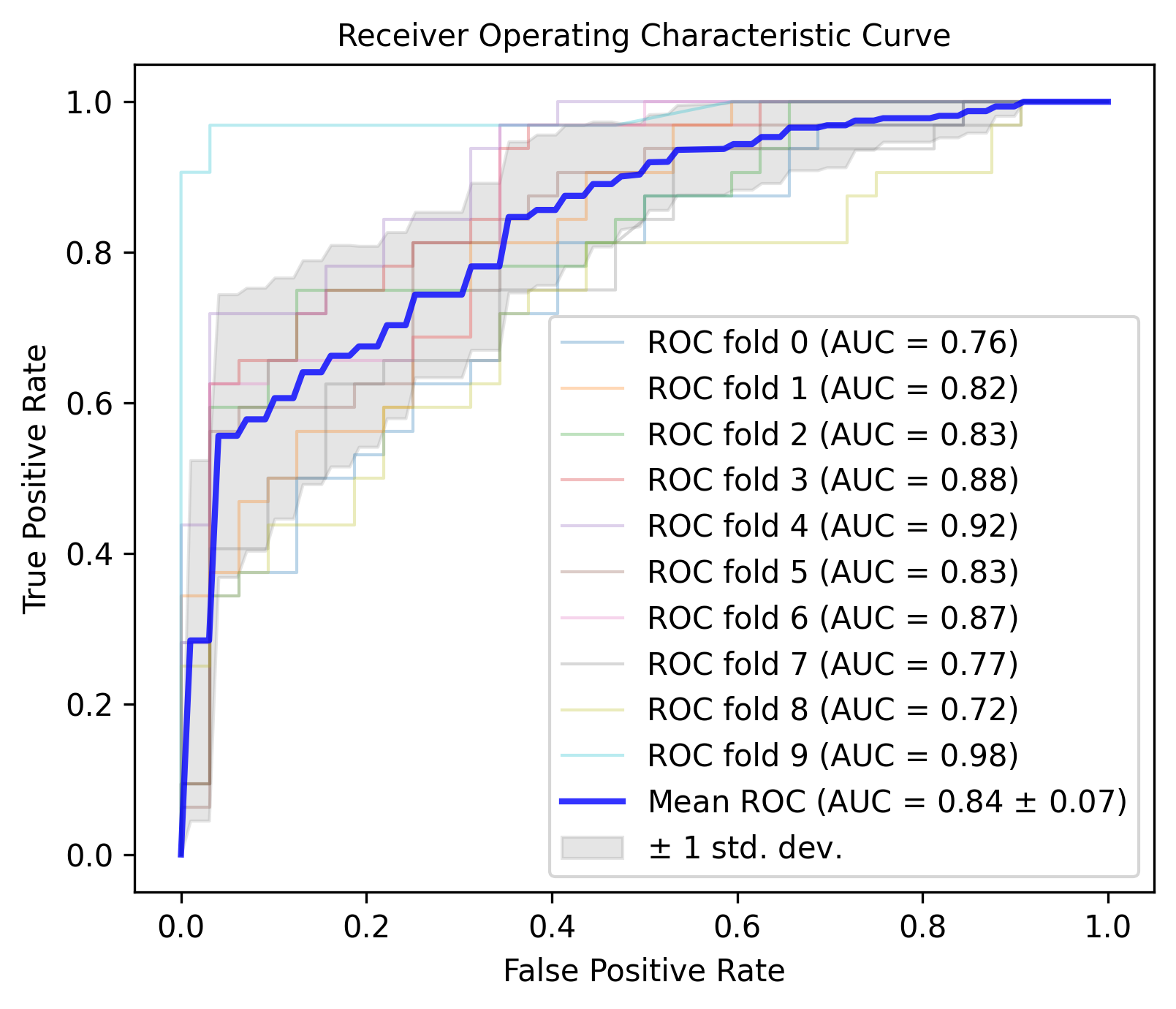


Figure 3. Feature importance is standardized such that it adds up to one and that the individual features are located on the X axis. Taking error into account, explicit stereotypical attitudes of company headquarter MSA has the most influence on whether or not a company is owned by a male or female CEO. All other features have too much error to accurately predict. Future models may add additional features to further specify feature importance. 

After running each classifier, the plot of true positive rate versus false positive rate at different thresholds, the receiver operating characteristic curves (ROC) for out of box samples are produced. The votes represent the probabilistic measure. The area under the ROC curve (AUC) gives an aggregate measure of performance at different thresholds. This value is averaged over several classifier runs (folds) over male-female sampling bootstrap to get the overall performance of our classification, which we found to be about 80%.

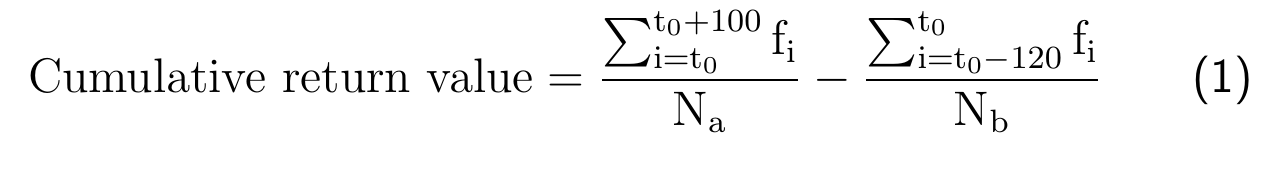
Figure 4. The ROC curve plotted using scikit learn metric object for several folds.

Another analysis was run to predict the economic benefit of hiring a female CEO [1][5]. Leeth et al. [1] [5] demonstrated when women were appointed as CEO companies had better financial health measured in terms of cumulative raw and market and risk adjusted return. For this particular study, a database of stock closing prices was prepared with starting dates in july, 2013 until April, 2020 and analyzed the performance of the companies that had their new CEO appointed during this time period. Companies who had a CEO who was set to join later in 2020 were excluded.

For some companies in the current database that did not have a newly appointed CEO within the chosen timeframe were computed using polynomial interpolation of order 1 (linear).

The performance of a newly appointed CEO is calculated by measuring the cumulative return value (the difference between the mean of the closing stock value 120 days before and 100 days after the appointment of the CEO). For example, if a female CEO of a company was appointed on 04/12/2019, fi can be denoted as the value of closing stock price on that day t = ti *,*

thus,



where t0 = 0 for 04/12/2019, Nais 100, Nb is 120.

Total number of days included = 100 + 1+ 120 = 221 days

As a further step, linear interpolation was used to predict the cumulative return value of some companies in the database that had hired their current CEO before July, 2013 since the stock closing price data covers only July, 2013 to April, 2020.

Since the cumulative return value is the difference of means of two time series, t-test was performed to check whether the average value of each time series for each company varies significantly. Figure 5 shows the result of the analysis. A comparison of cumulative return value of female CEOs based companies (panel (a)) with all companies with newly appointed CEOs (panel (b)) is shown. Ignoring the fact that the female company database contributes to only 10.47 % of the data,, it can be seen that both panel (a), (b) show that some companies have positive cumulative return indicating “Profit” and some have negative cumulative return indicating “Loss”. If the cumulative return is zero, it is assumed that the company was “Stable” over the period of the present analysis. A t-test was performed on each company time series data to check whether variation of average is significant or not. Panel ( c), (d) show the t-value for female CEO based companies and overall and panel (e), (f) show the p-value for the same. There are a few p-values that are above 10% for female CEO companies but most are below that threshold. Similar pattern is seen in p-value on overall data (panel (f)), which indicates that the averages are not random and are reliable.

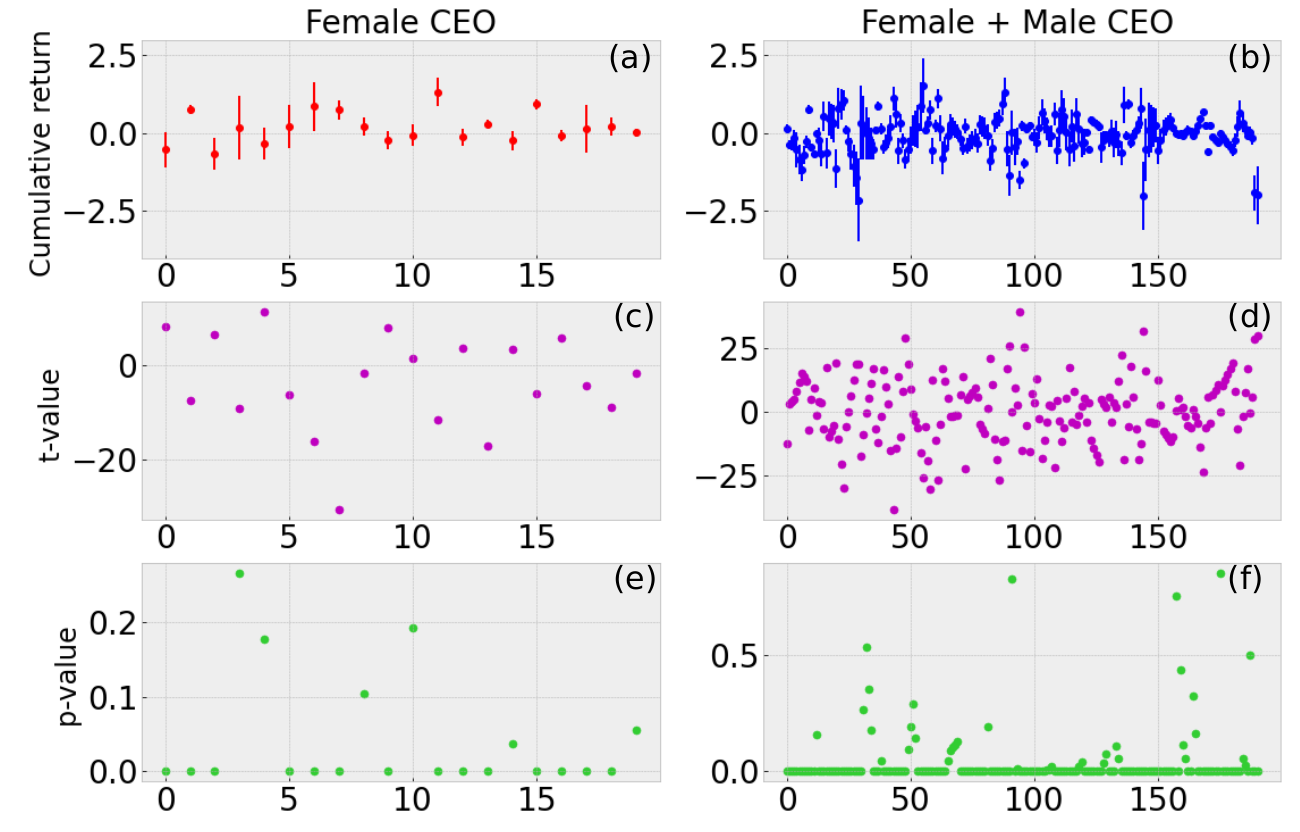


Figure 5. The figure above shows a comparison of the performance of newly appointed female CEOs with all newly appointed CEOs. Panel (a) and (b) show the cumulative return of companies with female CEOs and overall. T-test was performed for each stock with time series of 120 days before the CEO was appointed and 100 days after the appointment date. Panel ( c), (d) show the t-value of female CEO based companies and all companies in the database and panel (e), (f) correspond to the p-value i.e. probability of random number producing the same result.

The t-test revealed that 15 out of 20 stocks (75%) with female CEOs have p-value below 5% and out of those 10 had “Profit”. For stocks with male CEOs, 148 out of 171 (86%) met the threshold out of which, 80 had “Profit” and 91 went through “Loss”. This makes 85% of total newly appointed CEO companies meeting the threshold, and 78 of them making “Profit”.

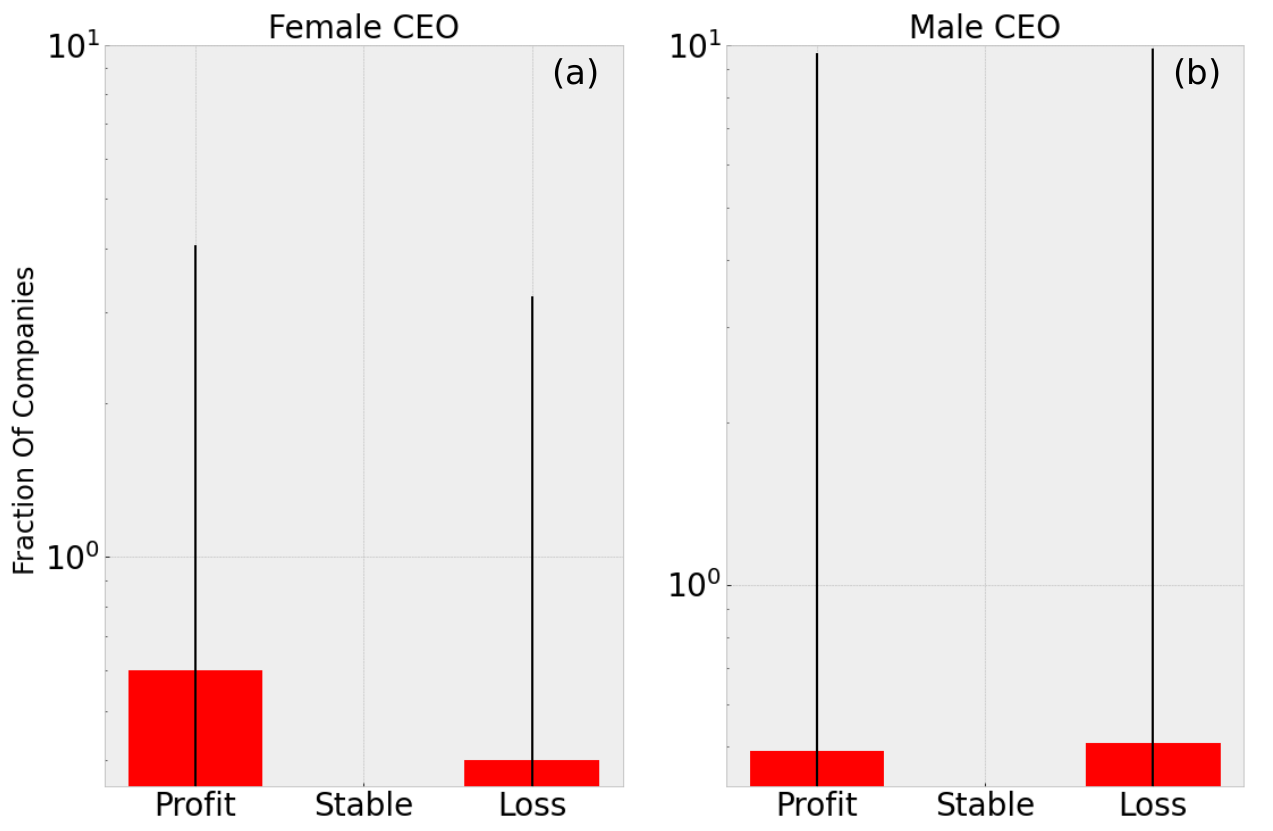


Figure 6. The figure above shows how the companies performed due to the change of CEO based on gender. Panel (a) shows what fraction of companies with newly appointed female CEOs had Profit (60%) and Loss (40%). Panel (b) shows the same for newly appointed male CEOs (Profit : 46.7%, Loss : 53.2%).

Now to get an overall idea of performance of newly appointed female (male) CEOs, fractions of female (male) CEO based companies with “Profit” , “Stable”, “Loss” was computed as shown in Figure 6. Panel (a) of Figure 6 shows that the dominating fraction of female CEO based companies is “Profit” (60 %) whereas the dominating fraction of male CEO based companies is “Loss” (53 %). There is no “Stable” fraction in either case which would mean that the company had no gain or loss over the chosen period. The error in each fraction is shown as where is the number of companies in each histogram bar. It can be clearly seen that newly appointed female CEO based companies performed better than newly appointed male CEO based companies.

**Conclusions and future work.**

Results suggest that male and female CEO run companies can be differentiated from one another. Moreover, this analysis suggested that the explicit stereotypic attitudes of individuals who live in the MSA of the company’s headquarters had the biggest influence on whether a company was likely to have a male or female CEO. Results contrast previous literature that suggest age, education level, and the number of children a CEO has are the most influential predictors for whether a company hires a male or female CEO[11]. The analysis of performance evaluation for newly appointed male and female CEO based companies using cumulative return value revealed some interesting results. 12 out of 20 companies with new female CEOs had “Profit” (60%) compared to male CEOs with “Profit” of 46%. It must also be noted that female CEO based companies had a dominating fraction of “Profit” compared to male CEO based companies where “Loss” was dominating fraction indicating that hiring a female CEO is economically more beneficial for a company. It will be interesting to check whether data imbalance had any effect towards this current conclusion. Together results suggest both that characteristics of stocks are indicative of whether companies are run by male or female CEOs and that companies run by a female CEO, which promotes gender diversity in S&P 500 companies, benefit economically from this choice.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**References:**

[1] Adams, S.M., Gupta, A. and Leeth, J.D.; Are female executives over‐represented in precarious leadership positions?(2009). British Journal of Management, 20(1), pp.1-12.

[2] Alagidede, P. Month of the year and pre-holiday effects in African Stock Markets.(2013) South African Journal of Economic and Management Sciences, 16(1), 64–74.

[3] Bang, J. T., Basuchoudhary, A., & Sen, T. (n.d.). New Tools for Predicting Economic Growth Using Machine Learning: A Guide for Theory and Policy. (2017) 44.

[4] Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (n.d.). Catastrophic Natural Disasters and Economic Growth. THE REVIEW OF ECONOMICS AND STATISTICS, 23.

[5] Cook, A. and Glass, C.; Above the glass ceiling: When are women and racial/ethnic minorities promoted to CEO?(2014). Strategic Management Journal, 35(7), pp.1080-1089.

[6] Eagly, A. H., & Wood, W. Social Role Theory. In P. Van Lange, A. Kruglanski, & E. Higgins, Handbook of Theories of Social Psychology(2012). (pp. 458–476). SAGE Publications Ltd. <https://doi.org/10.4135/9781446249222.n49>

[7] Fankhauser, S., & S.J. Tol, R. On climate change and economic growth. Resource and Energy Economics, (2005) 27(1), 1–17. <https://doi.org/10.1016/j.reseneeco.2004.03.003>

[8] Ghorbani, M., & Chong, E. K. Stock price prediction using principal components. (2020). PloS one, 15(3), e0230124.

[9] Gründler, K., & Krieger, T. Democracy and growth: Evidence from a machine learning indicator. (2020) European Journal of Political Economy, 45, 85–107. <https://doi.org/10.1016/j.ejpoleco.2016.05.005>

[10] Hamilton, B. H., Nickerson, J. A., & Owan, H. Diversity and Productivity in Production Teams. In A. Bryson (Ed.), Advances in the Economic Analysis of Participatory & Labor-Managed Firms (2012) (Vol. 13, pp. 99–138). Emerald Group Publishing Limited. <https://doi.org/10.1108/S0885-3339(2012)0000013009>

[11] Hurley, D., & Choudhary, A. Factors influencing attainment of CEO position for women. Gender in Management: (2016).An International Journal.

[12] Khan, W. A., & Vieito, J. P. CEO gender and firm performance. (2013).Journal of Economics and Business, 67, 55-66.

[13] Koenig, A. M., & Eagly, A. H. Evidence for the social role theory of stereotype content: Observations of groups’ roles shape stereotypes. (2014). Journal of Personality and Social Psychology, 107(3), 371. <https://doi.org/10.1037/a0037215>

[14] Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. Are leader stereotypes masculine? A meta-analysis of three research paradigms. (2011). Psychological Bulletin, 137(4), 616–642. <https://doi.org/10.1037/a0023557>

[15] Lahmiri, S. (n.d.). A Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information.

[16] Laudicina, E. V. Diversity and Productivity: Lessons from the Corporate Sector. (1993). Public Productivity & Management Review, 16(4), 457–463. JSTOR. <https://doi.org/10.2307/3381025>

[17] Liang, G., & Cohn, A. G. (2013, June). An effective approach for imbalanced classification: unevenly balanced bagging. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

[18] Malarvizhi, R., & Thanamani, A. S. (2012). K-nearest neighbor in missing data imputation. *International Journal of Engineering Research and Development*, *5*(1), 5-7.

[19] Miller, D. I., Eagly, A. H., & Linn, M. C. Women’s representation in science predicts national gender-science stereotypes: Evidence from 66 nations. (2015). Journal of Educational

Psychology, 107(3), 631–644. <https://doi.org/10.1037/edu0000005>

[20] Page, S. E. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition. (2008). Princeton University Press.

[21] Reagans, R., & Zuckerman, E. W. Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. (2001). Organization Science, 12(4), 502–517. <https://doi.org/10.1287/orsc.12.4.502.10637>

[22] Shabnam, N. Natural Disasters and Economic Growth: A Review. (2014) International Journal of Disaster Risk Science, 5(2), 157–163. <https://doi.org/10.1007/s13753-014-0022-5>

[23] Sokolov-Mladenović, S., Milovančević, M., Mladenović, I., & Alizamir, M. Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. Computers in Human Behavior,(2016), 65, 43-45.

[24] Thaler, R. Anomalies: Seasonal Movements in Security Prices II: Weekend, Holiday, Turn of the Month, and Intraday Effects. (1987).Journal of Economic Perspectives, 1(2), 169–177. <https://doi.org/10.1257/jep.1.2.169>

[25] Xue-shen, S., Zhong-ying, Q., Da-Ren, Y., Qing-hua, H., & Hui, Z. (2007, August). A novel feature selection approach using classification complexity for SVM of stock market trend prediction. In 2007 International Conference on Management Science and Engineering (pp. 1654-1659). IEEE.

[26] Yu, H., Chen, R., & Zhang, G. (2014). A SVM Stock Selection Model within PCA. Procedia Computer Science, 31, 406–412. <https://doi.org/10.1016/j.procs.2014.05.284>

[27] Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007, June). Combining news and technical indicators in daily stock price trends prediction. In International symposium on neural networks (pp. 1087-1096). Springer, Berlin, Heidelberg.

[28] Taylor SJ, Letham B. Forecasting at scale (2017). PeerJ Preprints 5:e3190v2 <https://doi.org/10.7287/peerj.preprints.3190v2>

[29] Sapankevych, N.I. and Sankar, R.,. Time series prediction using support vector machines: a survey (2009). *IEEE Computational Intelligence Magazine*, *4*(2), pp.24-38.

[30] Singh, A., & Lakshmiganthan, R. (2018). Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms.

[31] Krollner, B., Vanstone, B.J. and Finnie, G.R., April. Financial time series forecasting with machine learning techniques: a survey(2010). In *ESANN*.

[32] McNally, S., Roche, J. and Caton, S., March. Predicting the price of bitcoin using machine learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) (*2018*)* (pp. 339-343). IEEE.

[33] Teräsvirta, T., Tjøstheim, D. and Granger, C.W.J., *Modelling nonlinear economic time series (*2010*)*. Oxford: Oxford University Press.

[34] Bentler, P.M., Comparative fit indexes in structural models.(1990) *Psychological bulletin*, *107*(2), p.238.

[35] Andrae, R., Schulze-Hartung, T. and Melchior, P.,. Dos and don'ts of reduced chi-squared (2010). *arXiv preprint arXiv:1012.3754*.