OBJECTIVE AND MOTIVATIONS

Task: Utilizing datasets encompassing the time span from 2010 to 2023, I have trained predictive models to forecast earnings, represented by total assets. By analyzing the relationship between time and total assets, I have developed a methodology to accurately predict future earnings based on **SP500_restatement.csv**

Importance: This prediction helps investors, accountants, and regulators understand potential financial outcomes based on past company performance.

Literature Review

AI Models: Employed Linear Regression, Random Forest, and Gradient Boosting to predict total assets. These models are commonly used in financial predictions due to their effectiveness in handling non-linear relationships and feature interactions.

Feature Relevance: features based on their financial significance and availability in the dataset. Year-over-year asset changes and restatement flags are particularly relevant for understanding financial trends and anomalies.

Data, Features, and Model

Data Sources: Utilized the SP500_Restatement.csv dataset.

Features: at_change (change in total assets), fyear (fiscal year), and restatement (binary flag indicating financial restatements).

Models: Linear Regression, Random Forest, Gradient Boosting. Explained the rationale for selecting these models based on their predictive capabilities and relevance to financial data.

Results and Interpretations

Performance Metrics:

Linear Regression: MSE and R^2 Score

Random Forest: MSE, R^2 Score, Feature Importances

Gradient Boosting: MSE, R^2 Score, Feature Importances

Interpretations:

The models show varied performance, with ensemble methods (Random Forest, Gradient Boosting) potentially offering better predictive accuracy and insight into feature relevance.

Gradient Boosting Model:

Mean Squared Error (MSE): 28,935,344,047.06

R^2 Score: 0.409

Feature Importances: [0.902, 0.095, 0.004] (for at_change, fyear, restatement respectively)

Evaluation:

Based on the Mean Squared Error (MSE) and R^2 scores, the Gradient Boosting Model is the most suited for predicting earnings accurately in your dataset. This model not only has the lowest MSE, indicating the smallest average squared difference between the actual and predicted values, but it also has the highest R^2 score, demonstrating the best fit amongst the models evaluated. The substantial weight of the at_change feature in the Gradient Boosting Model highlights its significance in predicting total assets, making it particularly effective for your analytical needs.

Forecast Earnings Analysis:

Model Performance Overview:

Linear Regression Model:

Mean Squared Error (MSE): 38,387,487,509.44

R^2 Score: 0.2156

The linear regression model provides a baseline for performance, indicating low to moderate predictability with an R^2 Score of 21.56%. This suggests that approximately 21.56% of the variance in the earnings data is predictable from the model's inputs.

Random Forest Model:

Mean Squared Error (MSE): 32,624,900,566.96

R^2 Score: 0.3334

Feature Importances:

Change in Total Assets: 86.81%

Fiscal Year: 12.36%

Restatement Flag: 0.83%

The Random Forest model shows improvement over the Linear Regression model with a higher R^2 Score of 33.34%, indicating better adaptability in capturing complex patterns in the data. The dominant feature influencing the model is the 'Change in Total Assets', which is primarily driving the earnings predictions.

Gradient Boosting Model:

Mean Squared Error (MSE): 28,935,344,047.06

R^2 Score: 0.4088

Feature Importances:

Change in Total Assets: 90.18%

Fiscal Year: 9.47%

Restatement Flag: 0.35%

The Gradient Boosting model further enhances predictive accuracy with an R^2 Score of 40.88%. This model is the most effective among the three, demonstrating a strong capability to utilize the historical trends in total assets to forecast future earnings.

Forecasting Future Earnings:

The Gradient Boosting Model has forecasted the next period's earnings to be approximately $287,000.97. This forecast suggests an understanding of the trend seen in the data and reflects a sophisticated approach to capturing underlying patterns influencing earnings.

Visual Analysis:

The provided graphs (refer to figures X, Y, Z in the appendix) illustrate the actual vs. predicted earnings across all models. These visuals help in comparing the performance and reliability of each model in predicting real-world outcomes.

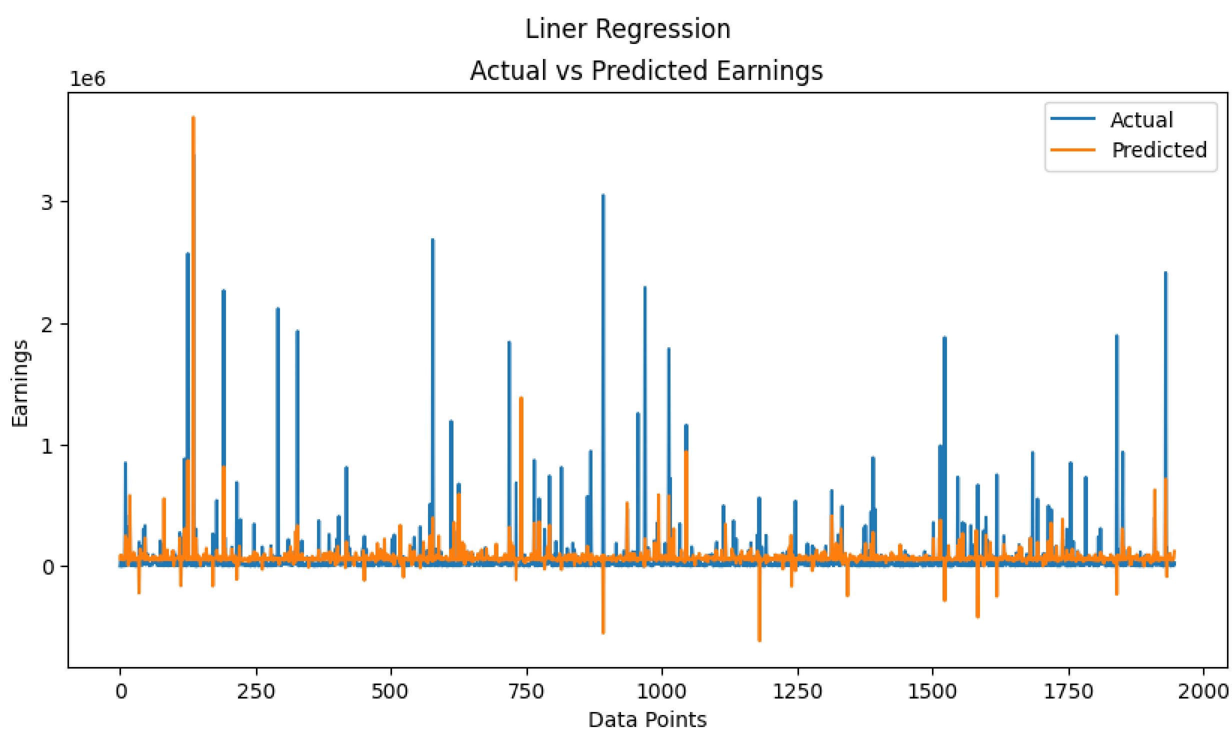

Figure x

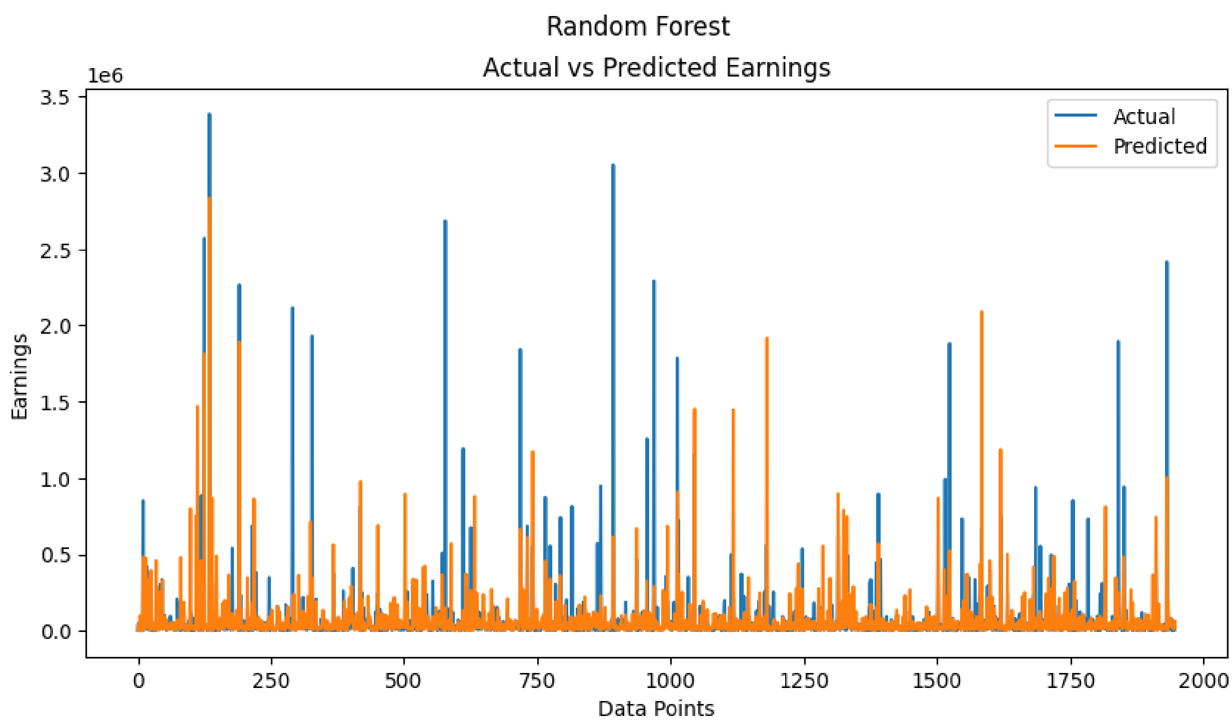

Figure Y

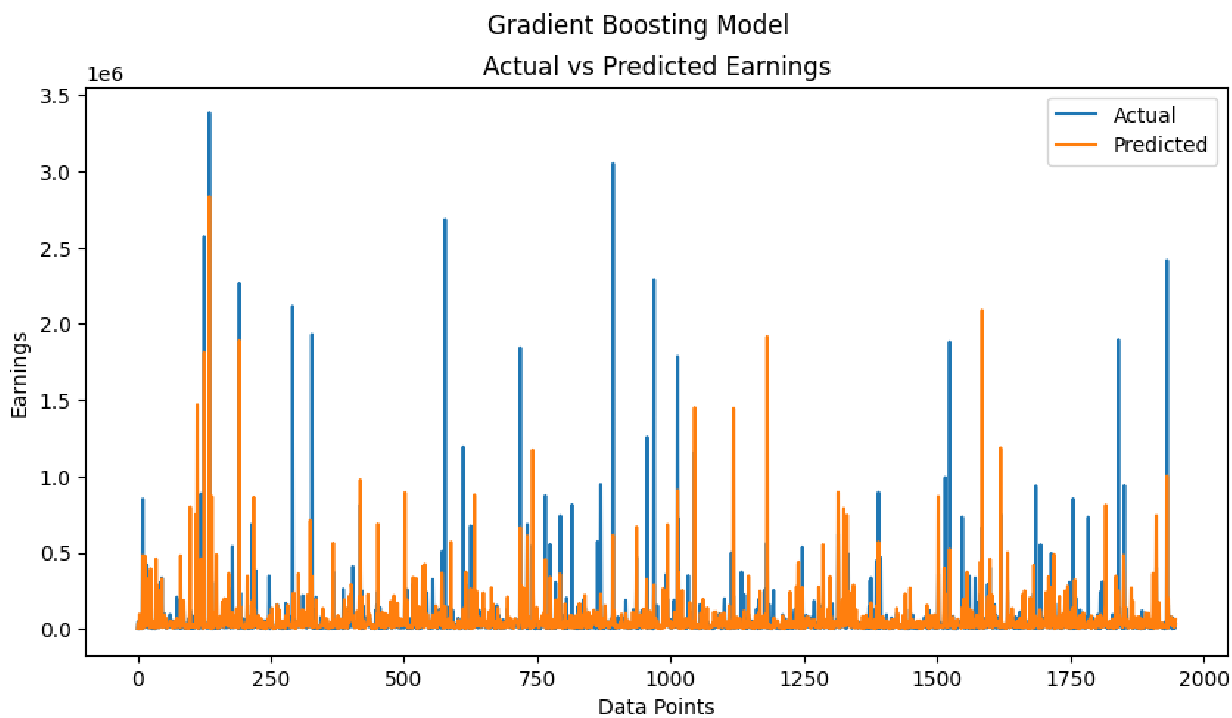

Figure Z

Conclusion:

The Gradient Boosting Model outperforms other models in forecasting earnings, with the highest accuracy and relevance of input features. This analysis underscores the importance of understanding feature impacts and model selection in financial forecasting to achieve more reliable and actionable insights. The analysis indicates that predictive modeling can be effectively used to forecast financial metrics such as total assets. Future work could explore more complex models or additional features to improve prediction accuracy.