**Predicting Company Profits Using Regression Models**

**ABSTRACT**

This research project will support business planning. Thus, there is a gap in the order in which the literature objective can be achieved by creating an effective study profit from the prediction of the following models variables; that can be utilized Regression analysis was performed on 50 companies' R&D, administration, and marketing expenditures in order to forecast their profits. The Mean Squared Error (MSE), Mean and Absolute R2 Error score, and Decision Tree Regression and Random Forest Regression were the three regression techniques that were evaluated in this context. The experiment's results showed that the models that produced the best model R2 with the highest score of 0.91 were Random Forest models. precision The study offers a framework for the application of machine learning in financial decision-making because it has the lowest error rate when compared to other methodologies, implementations, and outcomes.

**TABLE OF CONTENTS**

1. Abstract
2. Table of Contents
3. Introduction
4. Existing Method
5. Proposed Method
6. Methodology
7. Implementation
8. Conclusion

**INTRODUCTION**

Precise profit planning is important. predictions Many companies spend a significant portion of their strategic resources on marketing, management, and research and development. For this reason, it's critical to understand how these expenditures impact the company's profitability in order to effectively manage the costs. Using the spending data of fifty businesses, the following analysis aims to create a machine learning model that can forecast a company's profit. It will be possible to identify trends through data analysis that will likely aid in predicting a company's profits, making wise decisions, and even allocating resources appropriately. Traditional methods have relied on static and traditional statistical analysis, which can be rigid and imprecise. Numerous machine learning models, including Random Forest, Decision Linear Tree, and Regression, are effective and efficient in their own right. In an attempt to forecast business profits, the authors test these models here to see which one works best.

**EXISTING METHOD**

The use of financial modelling and descriptive analysis are prerequisites for the discovery of profit prediction. These methods usually use basic statistical measures to compile historical data and make inferences. Their dependence on a priori assumptions and incapacity to model interactions of any degree of nonlinearity are significant drawbacks, though. They frequently struggle when used in a market that is changing or evolving because they do not lend themselves well to scalability and flexibility. On the other hand, machine learning techniques create a better paradigm that can examine large amounts of data, identify complex patterns, and provide more accurate predictions. In order to increase the predictiveness of the models, this work evaluates the shortcomings of conventional methods and applies machine learning techniques.

**PROPOSED METHOD**

We suggest putting three regression methods into practice: Random Forest, Decision Tree, and Linear Regression. Data preparation, dataset separation into training and testing data, and model evaluation in relation to standard parameters are the suggested procedures. Because of its straightforward descriptive character and ability to provide a straightforward starting point for comparison with more sophisticated models, linear regression is included. Random Forest is an ensemble learning model that provides excellent accuracy, and Decision Tree Regression is utilized because it can handle non-linear interactions. According to the required architecture, 80% of the data is used for model training, with the remaining 20% being used for model testing. The analyzed models are evaluated by evaluating MSE, MAE, and R² score to determine the best model for driving profit prediction.

**METHODOLOGY**

R&D Spend, Administration, Marketing Spend, and Profit variables are all included in the dataset. Data preparation is the first step in the process, which cleans and sharpens the data. The dataset is divided into two parts: the test set is 20% and the training set is 80%. Among them, models like the Random Forest, Decision Tree, and Linear Regression models are trained using the training data.   
• Mean Squared Error (MSE) is the squared difference between the actual value and the anticipated value within the predetermined average.  
• The mean absolute error, or MAE, provides the absolute error average.

• The model's usage of the R2 Score indicates the degree of pattern variance.Make a profit. Data preparation is the first step in the process to guarantee consistency and quality. Training (80%) and testing (20%) portions of the dataset are separated. Using the training data, models such as Random Forest, Decision Tree, and Linear Regression are trained. On the test set, predictions are formed, and metrics are used to assess how well they perform:   
When comparing expected and actual numbers, the mean squared error, or MSE, calculates the average squared difference.   
•The mean absolute difference is given by the Mean Absolute Error (MAE).   
The R2 Score measures how effectively the model accounts for data variation.   
Out of all the models, Random Forest performed the best because it can effectively handle non-linear interactions and avoid overfitting brought on by ensemble learning.

**IMPLEMENTATION**

The implementation phase of this study utilized three powerful regression models: Linear regression, Trees and forests which include decision tree and random forest respectively. To this aim, it was aimed to forecast the company profits from the spending data and to find the most suitable model for this.• Made sure the data set was clean for modeling.powerful regression models: Linear Regression, Decision Tree, and Random Forest. The goal was to predict company profits based on spending data and to identify the most effective model. Here’s a step-by-step overview:

Step 1: Data Preprocessing

•Ensured the dataset was clean and ready for modeling.

•Features used: R&D Spend, Administration, Marketing Spend.

•Target variable: Profit.

Step 2: Data Splitting

•Divided the dataset into 80% training data and 20% testing data to prevent overfitting and ensure robust evaluation.

Step 3: Model Training

1.Linear Regression: A simplistic method of regression that presumes a kind of direct relation. This one was the basic one:

2.Decision Tree: This model catered for complicated, and non-linear dependencies but lacked robustness in the sense that it fit the data too tightly.

3.Random Forest: This forest rationalized decision trees and included frequency of different choices in order to lessen over-fitting with improved prediction capability.

Step 4: Model Evaluation

The models were evaluated using the following metrics:

• Mean Squared Error (MSE): Estimates the average of squared prediction errors.

• Mean Absolute Error (MAE): Represents the average size of chief and minor errors.

• R² Score: Points out the extent of the proportion of variation in the dependent variable which can be explained by the independent variable(s).

**Results**:

| **Model** | **MSE** | **MAE** | **R² Score** |
| --- | --- | --- | --- |
| Linear Regression | 80,926,321.2 | 6,979.15 | 0.90 |
| Decision Tree | 400,026,479.3 | 13,755.66 | 0.51 |
| **Random Forest** | **72,625,008.6** | **6,437.50** | **0.91** |

**Step 5: Best Model Selection**  
The **Random Forest Regression** model outperformed others, achieving the lowest errors and the highest R² score. Its ensemble learning capability made it robust and accurate for predicting profits.

**CONCLUSION**

The application of machine learning to financial prediction is demonstrated in this paper. With the best accuracy among all the models evaluated, the Random Forest model proved appropriate for predicting profits using expenditure data. It decreased inaccuracy and became extremely stiff by using an ensemble learning approach. Thus, this research again emphasizes how advanced analytics may help firms enhance their strategy. Future research should examine the effects of hyperparameter optimization techniques and incorporate additional features, such as market trends, to improve the model's effectiveness and applicability.