PROJECT REPORT
"PROFIT PREDICTION MODEL"

Submitted by:

      SUVRANEEL DAS

Project Overview:

      The aim of this project is to develop a profit prediction model using data of 50 startup companies. The model will use various machine learning techniques to analyse the data and predict future profits, enabling the company to make informed decisions and optimize its business strategies.

Project Objectives:

1.Collect and preprocess historical financial data.

2.Perform exploratory data analysis to gain insights into the data.

3.Select appropriate machine learning algorithms for profit prediction.

4.Train and fine-tune the selected model.

5.Evaluate the model's performance using relevant metrics.

6.Deploy the model for real-time profit predictions.

Data Collection and Preprocessing:

1.Obtain historical financial data including R & D spend, administration cost and marketing spend.

2.Handle missing values, outliers, and ensure data consistency.

3.Normalize or standardize numerical features to ensure equal weight during modeling.

Exploratory Data Analysis (EDA):

1.Visualize the distribution of profit over time.

2.Analyze correlations between profit and other financial indicators.

3.Identify seasonal trends and patterns in the data.

Model Selection:

Choose appropriate regression algorithms for profit prediction (e.g., Linear Regression, Decision Trees, Random Forest, XGBoost, Ridge Regression, Lasso Regression, SVR).

Model Training:

1.Split the data into training and testing sets.

2.Train the selected models using the training data.

Model Evaluation:

1.Evaluate the models using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R2) score. The lower value of MAE & MSE implies higher accuracy of a regression model. However, a higher value of R square is considered desirable.

2.Compare the performance of different models and select the best-performing one.

Deployment:

Deploy the selected model to a production environment.

Set up a mechanism for real-time profit prediction using new data.

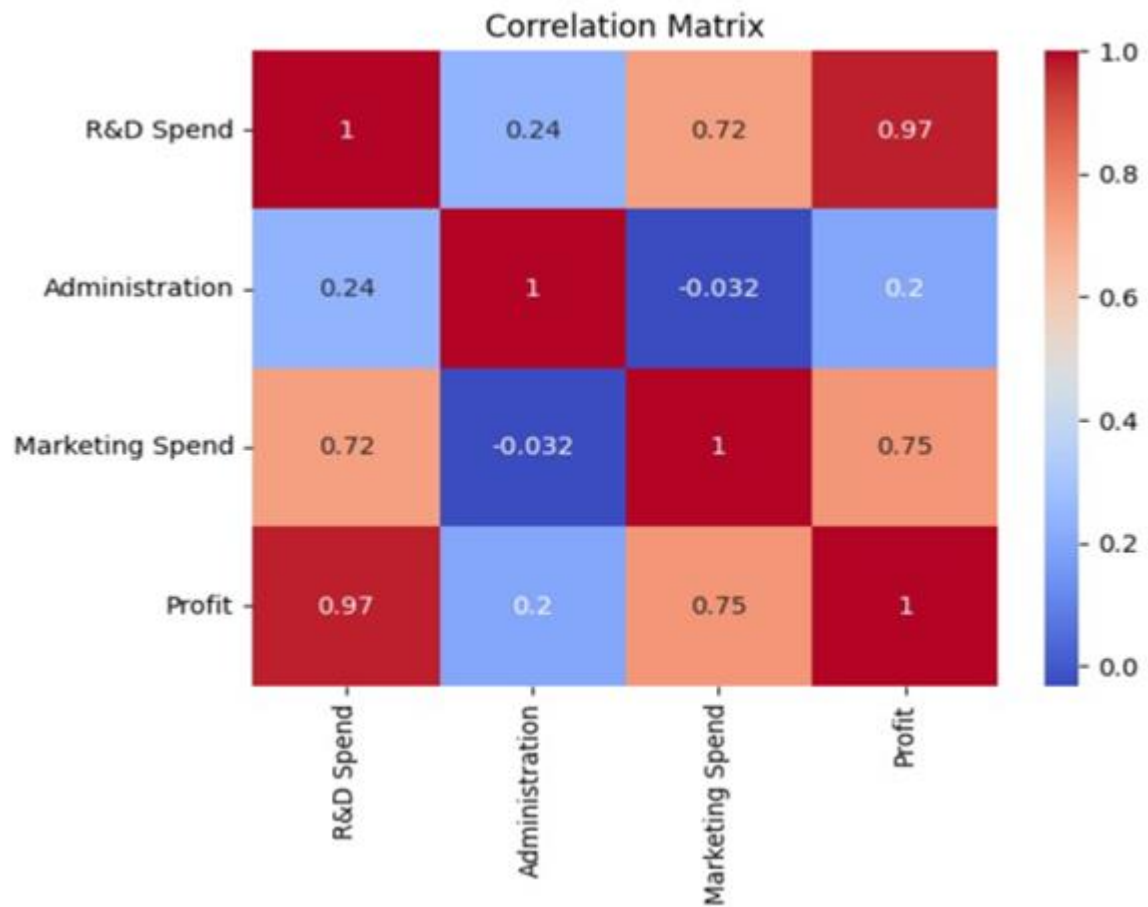Implement monitoring and alerting systems to detect any anomalies in profit predictions.

Project Title: PREPARING AN ML MODEL TO PREDICT PROFIT

➢ A profit prediction project is a machine learning project that aims to predict the profit of a company based on various factors such as R&D spend, administration cost and marketing spend, generated by the company.

➢ This project involves training a machine learning model on a dataset that contains information about the various factors that affect the profit of a company.

➢ Three input(independent) variables are given: R&D spend, Administration cost and Marketing spend

➢ One output(dependent) variable is: Profit

➢ For this, I used various regression models and comparing the values (R-squared-score, Mean-absolute-error & Mean-squared-error) I chose the best model for my project.

➢ The model is then used to predict the profit of a new company based on its characteristics.

Project Outputs:

X variables are : R&D spend, Administration cost and Marketing spend.
Y variable is : Profit

Correlation Matrix

	R&D Spend	Administration	Marketing Spend	Profit
R&D Spend	1	0.24	0.72	0.97
Administration	0.24	1	-0.032	0.2
Marketing Spend	0.72	-0.032	1	0.75
Profit	0.97	0.2	0.75	1

Models used with their scores are below.

Out[57]:

	Model	R2_score	MSE	MAE
0	Linear	0.900065	8.092632e+07	6979.152252
1	Decision Tree	0.872684	1.030996e+08	7445.785000
2	Random Forest	0.908542	7.406215e+07	6514.085000
3	Ridge	0.900065	8.092632e+07	6979.152252
4	Lasso	0.900065	8.092632e+07	6979.152251
5	Elasticnet	0.900065	8.092632e+07	6979.152252
6	SVR	0.871779	1.038321e+08	7702.623216
7	XGB	0.904580	7.727013e+07	7779.489250

Here I have calculated R-squared value, Mean Squared Error, Mean Absolute Error to predict the best model for my project. The lower value of MAE & MSE implies higher accuracy of a regression model. However, a higher value of R square is considered desirable.

According to the algorithm which is used here, Final score is calculated. Lowest final score implies the best model.

	Model	R2_score	MSE	MAE	R2_rank	MSE_rank	MAE_rank	Final_rank
2	Random Forest	0.908542	7.406215e+07	6514.085000	1.0	1.0	1.0	3.0
4	Lasso	0.900065	8.092632e+07	6979.152251	3.0	3.0	2.0	8.0
5	Elasticnet	0.900065	8.092632e+07	6979.152252	4.0	4.0	3.0	11.0
7	XGB	0.904580	7.727013e+07	7779.489250	2.0	2.0	8.0	12.0
0	Linear	0.900065	8.092632e+07	6979.152252	5.0	5.0	4.0	14.0
3	Ridge	0.900065	8.092632e+07	6979.152252	6.0	6.0	5.0	17.0
1	Decision Tree	0.872684	1.030996e+08	7445.785000	7.0	7.0	6.0	20.0
6	SVR	0.871779	1.038321e+08	7702.623216	8.0	8.0	7.0	23.0

Result Evaluation

➤ After training different models and testing output data, I came to a conclusion that **Random Forest Regression** is the best regression model for calculating profit based on given data.

➤ Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees. It has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

Conclusion:

In conclusion, this project aimed to develop a profit prediction model using historical financial data of 50 startup companies. The process involved data collection, preprocessing, exploratory data analysis, model selection, training, tuning, evaluation, and deployment. The developed model provides the company with the capability to make informed decisions, optimize strategies, and plan for the future based on accurate profit predictions. Ongoing monitoring and updates to the model will ensure its continued accuracy and relevance over time.