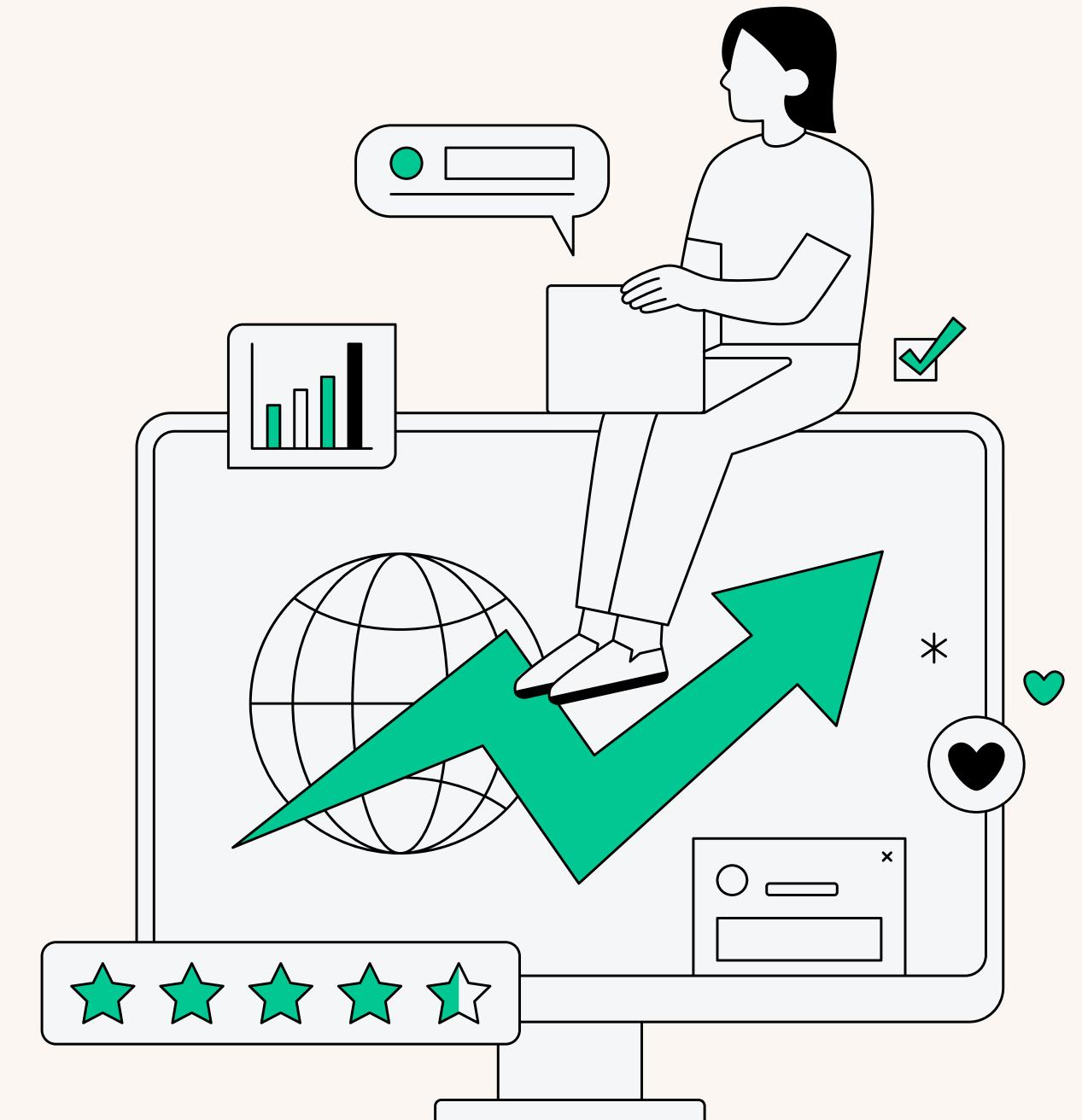


Presented by Arisha Akhtar

Financial Analysis and Profit Prediction of Startup Companies

Analyzing Key Financial Metrics and Implementing Machine Learning Models



ABSTRACT

This study analyzes the financial dynamics of 50 startups, focusing on R&D spending, administration costs, marketing expenditures, and resulting profits. The dataset highlights the importance of these variables in resource allocation and profitability.

Key findings include statistical summaries and correlation analyses that reveal the relationships between investments and profit. Visualizations, like scatter plots, illustrate these dynamics and aid in strategic decision-making.

The study evaluates several regression models, including Linear Regression, Decision Tree, Random Forest, Ridge Regression, Support Vector Machine, Lasso Regression, and XGBoost. Among these, Random Forest emerges as the most accurate for predicting profit.

In conclusion, the study provides valuable insights for optimizing resource allocation and enhancing profitability in startups, leveraging predictive models to guide strategic planning and investment decisions.

Introduction

Startup companies play a vital role in driving innovation, fostering economic growth, and shaping industry landscapes. As these ventures navigate the competitive business environment, effective financial management becomes paramount for sustainable growth and success. Understanding the financial dynamics of startup companies, including their investments in research and development (R&D), administration, marketing, and resulting profitability, is essential for stakeholders seeking to optimize resource allocation and enhance financial performance.

This study talks about the financial dynamics of 50 startup companies, focusing on key financial metrics such as R&D spending, administration costs, marketing expenditures, and profits. The dataset provides a comprehensive overview of the financial landscape within which these startups operate, offering valuable insights into the allocation of resources and the relationship between investment and profitability.

This study embarks on a comprehensive evaluation of various regression algorithms to predict profit in startup companies. Linear Regression, Decision Tree, Random Forest, Ridge Regression, Support Vector Machine, Lasso Regression, and XGBoost are among the algorithms scrutinized for their predictive capabilities. By comparing accuracy, R-squared score, Mean Squared Error, and Mean Absolute Error, the study aims to identify the optimal model for profit prediction, providing actionable insights for stakeholders.

DATASET

In the given dataset, R&D Spend, Administration Cost and Marketing Spend of 50 Companies are given along with the profit earned.

R&D Spend	Administration	Marketing Spend	Profit
165349.2	136897.8	471784.1	192261.83
162597.7	151377.59	443898.53	191792.06
153441.51	101145.55	407934.54	191050.39
144372.41	118671.85	383199.62	182901.99
142107.34	91391.77	366168.42	166187.94
131876.9	99814.71	362861.36	156991.12
134615.46	147198.87	127716.82	156122.51

Methodology used in the analysis

1. Data Collection

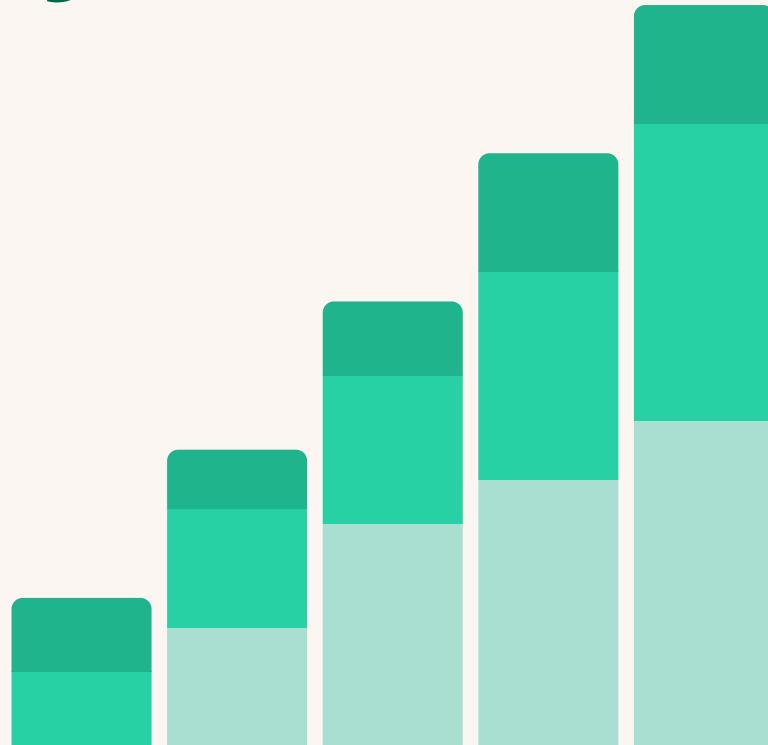
- Source: Financial data of 50 startup companies.
- Variables: R&D Spend, Administration Cost, Marketing Spend, Profit.

2. Data Preprocessing

- Handling Missing Values: Ensure there are no missing values in the dataset.
- Feature Scaling: Standardize the data to bring all features to a similar scale.

3. Exploratory Data Analysis (EDA)

- Summary Statistics: Calculate mean, standard deviation, minimum, and maximum for each variable.
- Correlation Analysis: Determine the correlation coefficients to understand relationships between variables.
- Scatter Plots: Visualize the relationships between each independent variable (R&D Spend, Administration, Marketing Spend) and the dependent variable (Profit).



4. Model Selection

- Linear Regression: Simple linear approach to model the relationship between dependent and independent variables.
- Decision Tree: A tree-based model for capturing non-linear relationships.
- Random Forest: An ensemble method that improves the accuracy by using multiple decision trees.
- Ridge Regression: A regularized version of linear regression to handle multicollinearity.
- Support Vector Machine (SVM): A model that finds the hyperplane which best separates the data.
- Lasso Regression: Similar to Ridge, but with L1 regularization to enforce sparsity.
- XGBoost: An advanced ensemble method for high predictive performance.

5. Model Training and Validation

- Train-Test Split: Divide the dataset into training and testing sets to evaluate model performance.

6. Model Evaluation

- Accuracy (ACC): Measure of the number of correct predictions.
- R-squared score (R₂_score): Proportion of variance in the dependent variable explained by the independent variables.
- Mean Squared Error (MSE): Average of the squared differences between predicted and actual values.
- Mean Absolute Error (MAE): Average of the absolute differences between predicted and actual values.

Results and Findings

Based on the provided evaluation metrics for each model, we can draw the following conclusions:

Random Forest (RF) appears to be the best-performing model among those evaluated, as it achieves high accuracy, high R-squared score, and relatively low MSE and MAE values. XGBoost also performs well in terms of accuracy but has higher MSE and MAE compared to RF. Decision Tree (DT) has the highest accuracy but seems to overfit the data, as indicated by its lower R-squared score and higher MSE and MAE values compared to other models.

Models	ACC	R2_score	MSE	MAE
LR	95.359278	90.006531	8.092632e+09	697915.225237
DT	96.862937	80.011501	1.618653e+10	991469.745833
RF	99.093530	91.031647	7.262501e+09	643749.774000
Ridge	95.359278	90.006531	8.092632e+09	697915.225243
SVM	94.931355	87.177927	1.038321e+10	770262.321589
LASSO	95.359278	90.006531	8.092632e+09	697915.223548
XgBoost	100.000000	87.301689	1.028299e+10	775766.048438

Insights and Recommendations

The analysis provides valuable insights into the financial dynamics of startup companies, highlighting the importance of investment in R&D and marketing for enhancing profitability.

Recommendations may include prioritizing R&D and marketing expenditures to drive profitability, based on the observed correlations and model interpretations.

Thank
you very
much!

