REPORT ON

FOUR WEEKS OF INTERNSHIP

Carried out at

Exposys Data Labs

Submitted to

NMAM INSTITUTE OF TECHNOLOGY, NITTE

(An Autonomous Institution under VTU, Belagavi)

In partial fulfillment of the requirements for the award of the

Degree of Bachelor of Engineering
in
Computer Science and Engineering

by

Ryan Alexander D'Souza
4NM20CS148

Under the guidance of

Ms. Arvind Kumar
R & D Engineer

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

CERTIFICATE

This is to certify that the "Internship report" submitted by Mr. Ryan Alexander D'Souza bearing USN 4NM20CS148 of 8th semester B.E., a bonafide student of NMAM Institute of Technology, Nitte, has undergone four weeks of internship at Exposys Data Labs during month July year 2023 fulfilling the partial requirements for the award of degree of Bachelor of Engineering in Computer Science and Engineering at NMAM Institute of Technology, Nitte.

Name and Signature of Mentor

Signature of HOD

COMPANY CERTIFICATE

Exposys
Data Labs

Certificate of Internship

TO WHOM IT MAY CONCERN

This is to certify that **Mr. RYAN ALEXANDER D'SOUZA** has completed internship programme on **"Data Science"** from 22.06.2023 to 21.07.2023.

He took keen interest in the work assigned and successfully completed it. During the period of internship, we found him to be punctual, hardworking and inquisitive.

We wish him luck and success in all his future endeavours.

Y Vishnuvardhan

Chief Director

EXPOSYS DATA LABS
BENGALURU

hr@exposysdata.com
www.exposysdata.com

ACKNOWLEDGMENT

First and foremost, I express my sincere gratitude to my institute for giving me the chance to complete the internship. Our professional development has been dramatically influenced by the institution's persistent dedication to supporting real world learning opportunities.

My heartfelt thanks to my esteemed guide and mentor, Mr. Arvind Kumar from Exposys Data Labs, for his valuable advice, endless support and motivation, constantly throughout. I record my indebtedness to Exposys Data Labs for giving me a platform to learn during my internship.

I would also like to thank Dr. Niranjan Chiplunkar, Principal, NMAMIT for his consistent and providing me this opportunity to do the internship.

I would like to thank Dr. Jyothi Shetty, Head of Department, Computer Science and Engineering for her constant support and providing me this opportunity to do the internship.

I would also like to acknowledge the support and assistance of our faculty members who have provided me with valuable insights and knowledge throughout my academic journey. Their expertise and guidance have laid a strong foundation for my internship and have helped me navigate the challenges and complexities of the chosen field.

Additionally, I would like to thank all the professionals who took the time to mentor me during my internship. Their patience, willingness to share their knowledge, and commitment to my growth have been truly inspiring. Their guidance has not only enhanced my technical skills but also instilled in a sense of professionalism and work ethic that will undoubtedly shape my future endeavours.

TABLE OF CONTENTS

ABSTRACT

The aim of this project is to develop a machine learning model that predicts the profit value of a company based on its R&D Spend, Administration Cost, and Marketing Spend. The accurate prediction of profits can help businesses make informed decisions, optimize resource allocation, and enhance overall performance. In this report, we explore different regression algorithms, perform data preprocessing, train the models, evaluate their performance, and choose the best model for profit prediction. The report covers the entire process from data preparation to model implementation and concludes with the results and findings.

In today's highly competitive business landscape, predicting profits accurately is crucial for making informed strategic decisions. The Profit Prediction Report Project aims to develop a comprehensive predictive model that assists businesses in forecasting their future profits with greater precision. This abstract provides an overview of the project's objectives, methodology, key findings, and potential impact.

Through meticulous data collection, preprocessing, and feature selection, the project identifies key factors influencing profit margins, such as sales data, operating expenses, and market trends. Leveraging machine learning algorithms and rigorous evaluation techniques, the developed predictive model provides actionable insights and recommendations to assist businesses in making informed strategic decisions. By quantifying the impact of various internal and external factors on profit forecasts, the project empowers organizations to optimize resource allocation, pricing strategies, and operational efficiency, thereby enhancing long-term sustainability and competitive advantage.

The Profit Prediction Report Project aims to develop a comprehensive predictive model that assists businesses in forecasting their future profits with greater precision. This abstract provides an overview of the project's objectives, methodology, key findings, and potential impact.

CHAPTER 1

INTRODUCTION TO THE INDUSTRY

Exposys Data Labs aims to solve real world business problems like Automation, Big Data and data Science. There are core team of experts in various technologies help businesses to identify issues, opportunities and prototype solutions using trending technologies like AI, ML, Deep Learning and Data Science. A human focussed and not technology driven approach is followed to achieve success in client’s endeavours.

It is a prominent organization involved in rendering IT services. Services are to cater to different sectors at cost-competitive prices as per the suitability and requirements of the clients. High quality services that match the expectations of our customers are rendered. Best possible solution is given keeping in mind the clients' strategic business interests and risks involved.

The main mission is to tap and train best brainpower to give solutions for real challenges of the world. Exposys Data Labs helps you to assess, develop and test your product idea and deliver a software that will satisfy your clients, sustain competition and achieve highest returns.

CHAPTER 2

DETAILS OF THE TRAINING UNDERGONE

Introduction

The introduction provides an overview of the project's objective, explaining the significance of predicting company profits and the potential benefits it can offer to businesses. It highlights the importance of accurate profit prediction for financial planning, budgeting, and investment decisions. Additionally, the introduction briefly describes the dataset used, the target variable for prediction, and the variables considered as predictors.

Profit prediction plays a crucial role in the business world, enabling organizations to make informed decisions, devise effective strategies, and allocate resources wisely.

Project Objectives

- Develop a machine learning model for profit prediction based on R&D Spend, Administration Cost, and Marketing Spend.

Existing Method

This section discusses any existing methods or traditional approaches that companies might use for profit prediction. It explores manual techniques, simple linear regression models, or rule-based methods that may have been employed before the advent of machine learning algorithms. The limitations of these methods are discussed, emphasizing the need for a more sophisticated approach that machine learning can offer.

Proposed Method

Here, we delve into the proposed method for profit prediction using machine learning. We discuss the choice of regression algorithms considered in the project, such as Linear Regression, Decision Tree Regression, and Random Forest Regression. The architecture of each model is explained, highlighting their strengths and weaknesses. We also discuss the ensemble learning approach to combine multiple models for improved prediction accuracy.

Methodology

The methodology section provides detailed insights into the steps followed during the project:

1.Data Collection:

- Information on R&D Spend, Administration Cost, Marketing Spend, and corresponding profit data was obtained.

2. Data Preprocessing:

- Data normalization techniques such as Min-Max scaling or Z-score normalization were applied to ensure all features were on a similar scale.

3.Feature Engineering:

- Creation of new features and handling categorical variables. New features were created by combining existing features or extracting relevant information.

4. Model Selection:

- Algorithms such as Linear Regression, Ridge Regression, and Lasso Regression were selected based on their ability to handle continuous target variables.

5. Model Training and Tuning:

- Details of the model training process, including hyperparameter tuning. The selected models were trained using the pre-processed data.

6. Evaluation Metrics:

- Explanation of the chosen evaluation metrics and their interpretation.
  Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared were chosen to evaluate model performance.

CHAPTER 3

Implementation

Implementation of the project is as follows

This section presents the actual implementation of the project. It provides a step-by-step walkthrough of the code used to read the dataset, preprocess the data, build and train the regression models, and evaluate their performance using metrics like Mean Squared Error, Mean Absolute Error, and R-squared.

In addition to the model evaluation metrics, we can conduct cross-validation to assess the models' generalizability and reduce overfitting.

Regression algorithms such as Linear Regression, Ridge Regression, and Lasso Regression were selected based on their ability to handle continuous target variables. Ensemble techniques like Random Forest Regression or Gradient Boosting Regression were considered for their ability to capture complex relationships in the data.

Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared were chosen to evaluate model performance. MAE represents the average magnitude of errors, MSE quantifies the average squared difference between predicted and actual values, and R-squared measures the proportion of variance explained by the model.

By following this methodology, a predictive model for cost prediction can be developed and implemented effectively, providing valuable insights for businesses to optimize their resource allocation and decision-making processes.

Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared were chosen to evaluate model performance. MAE represents the average magnitude of errors, MSE quantifies the average squared difference between predicted and actual values, and R-squared measures the proportion of variance explained by the model.

CONCLUSION

The conclusion provides a comprehensive summary of the project's findings and outcomes. It highlights the achievements, including the successful development of a profit prediction model. The report summarizes the best model selected for profit prediction and its potential benefits for businesses. We discuss the practical applications of the model in real-world scenarios and provide suggestions for future work, such as improving data collection or exploring more advanced machine learning techniques.

The Profit Prediction project has been a remarkable endeavour in leveraging the power of machine learning to forecast company profits accurately. The successful implementation of various regression algorithms and thorough evaluation of their performance have yielded valuable insights that can significantly impact business decision-making processes.

Through meticulous data processing, feature engineering, and model selection, this project has showcased the potential of machine learning algorithms to accurately forecast costs based on various factors. By providing actionable insights and facilitating informed decision-making, such models offer a pathway towards optimizing resource allocation, mitigating risks, and ultimately enhancing financial performance. As businesses strive for agility and resilience in dynamic market environments, the adoption of predictive analytics for cost prediction emerges as a strategic imperative for sustainable growth and competitive advantage.

REFERENCES

- https://www.w3schools.com/js
- https://www.emailjs.com/
- https://www.w3schools.com/html/html_forms.asp
- https://www.w3schools.com/whatis/
- https://en.wikipedia.org/wiki/Wikipedia:Mass_message_senders
- https://www.w3schools.com/css/