FINAL PRESENTATION OF MINI PROJECT

On the Topic

Predicting Insurance Premium for Life Insurance Applicants using Machine Learning

Central Institute of Technology Kokrajhar
Computer Science & Engineering

Project Members:

Akher Uddin Ahmed
(202102022100)
Birhang Borgoyary
(202102023126)
Subhash Kumar Wary
(202002022110)

Project Supervisor:

Mr. Prasanta Baruah
Assistant Professor
Dept. of Computer Science & Engineering

CONTENTS

- Introduction
- Objectives
- Problem Statement
- Block Diagram
- Methodology
- Dataset
- Results and Analysis
- Conclusion
- Future Scope
- References

INTRODUCTION

➢Life Insurance plays a crucial role in securing the financial well-being of loved ones in the event of an unexpected demise.

➢However, determining the appropriate premium for each policyholder involves a complex premium assessment process. Traditionally, this process relies on human underwriters who analyze various factors like age, health history and lifestyle etc.

➢Machine learning (ML) techniques has revolutionized the insurance industry, offering new opportunities to enhance the accuracy and efficiency of premium assessment processes.

➢By leveraging ML, insurers can streamline the underwriting process, reduce costs and improve premium prediction accuracy, ultimately providing better coverage options for policyholders

OBJECTIVES

➢To develop a machine learning model that can accurately predict the premium based on various features.

➢To provide accurate predictions which can be used to provide early analysis for Life Insurance Company.

➢The proposed solution for ML based premium prediction presents cost-effective and reliable solution.

➢This approach believes that using the Machine Learning, Life Insurance Company can be benefited.

PROBLEM STATEMENT

- To build an efficient model which can predict premium of a Life Insurance Applicant.
- Implementing latest technological advancements such as Machine learning.
- Specific Life Insurance Policy can be suggested to applicants based on predicted result.

METHODOLOGY

- Pre-processing of data has been done before training the model.
- 80% (1070) data has been taken for training and 20% (268) data taken for testing.
- The model has trained with the various algorithms such as Linear Regression, Random Forest Regressor, Support Vector Regressor and Gradient Boosting Regressor.
- Various machine learning algorithms were compared and Gradient Boosting Regressor model was selected as the best option for prediction.
- Predict the result using the trained model on the input data and provide the prediction.

BLOCK DIAGRAM

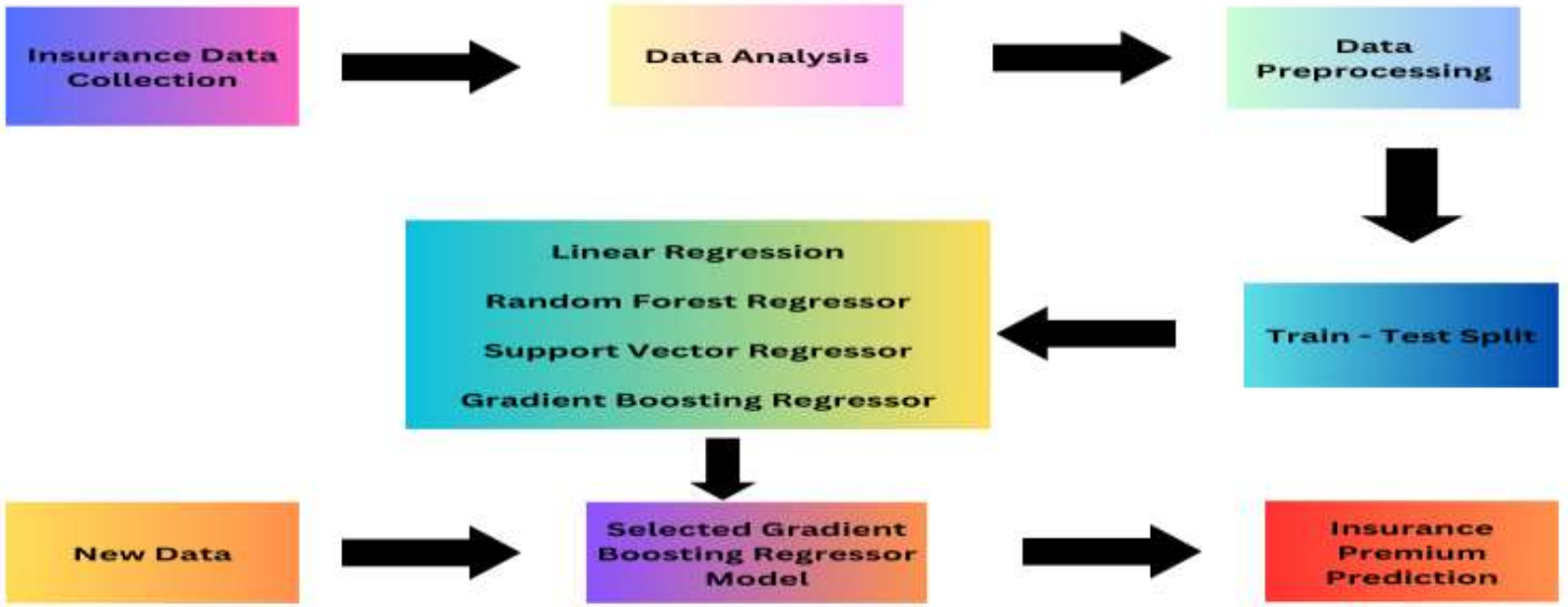

Fig. 1: Block Diagram

DATASET

➢Dataset has been taken from Kaggle containing the features Age, Sex, BMI, Children, Smoker, Region and Charges.

Table 1: Insurance Dataset

	A	B	C	D	E	F	G	H
1	age	sex	bmi	children	smoker	region	charges	
2	19	female	27.9	0	yes	southwest	16884.92	
3	18	male	33.77	1	no	southeast	1725.552	
4	28	male	33	3	no	southeast	4449.462	
5	33	male	22.705	0	no	northwest	21984.47	
6	32	male	28.88	0	no	northwest	3866.855	
7	31	female	25.74	0	no	southeast	3756.622	
8	46	female	33.44	1	no	southeast	8240.59	
9	37	female	27.74	3	no	northwest	7281.506	
10	37	male	29.83	2	no	northeast	6406.411	
11	60	female	25.84	0	no	northwest	28923.14	
12	25	male	26.22	0	no	northeast	2721.321	
13	62	female	26.29	0	yes	southeast	27808.73	
14	23	male	34.4	0	no	southwest	1826.843	
15	56	female	39.82	0	no	southeast	11090.72	

Total Number of Rows and Columns in our Dataset are 1338 and 7 respectively

DATASET

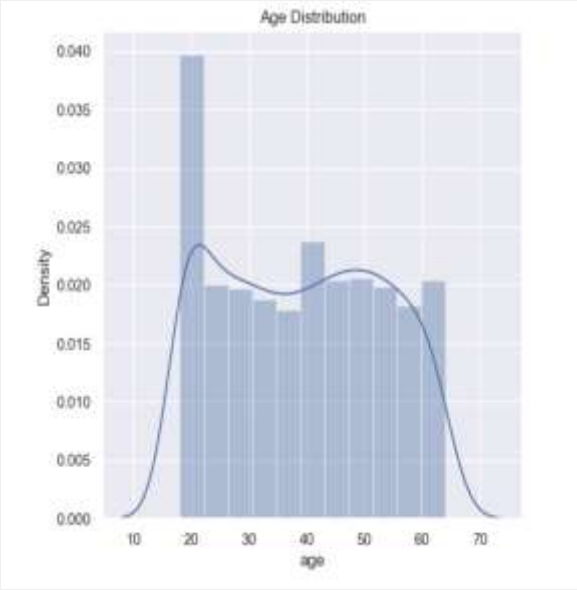

Fig. 2: Age Distribution

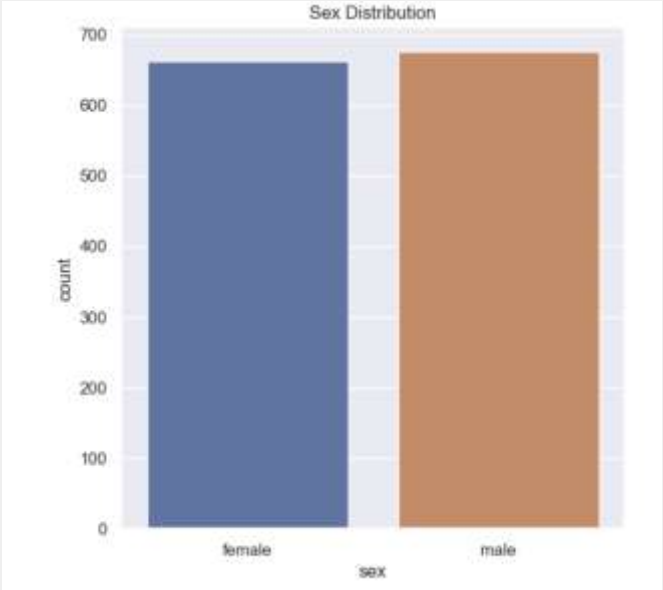

Fig. 3: Sex Distribution

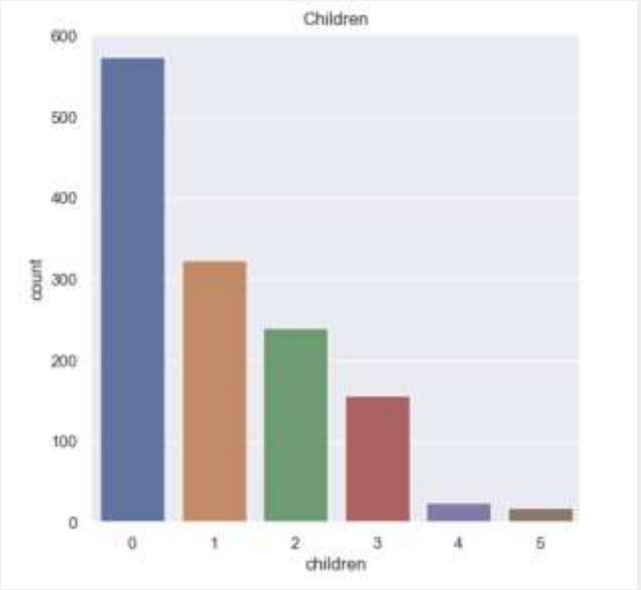

Fig. 4: Children Distribution

DATASET

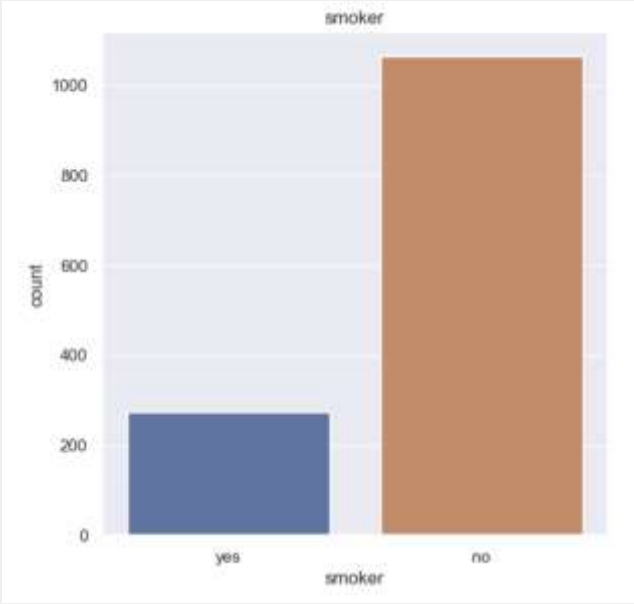

Fig. 5: Smoker Distribution

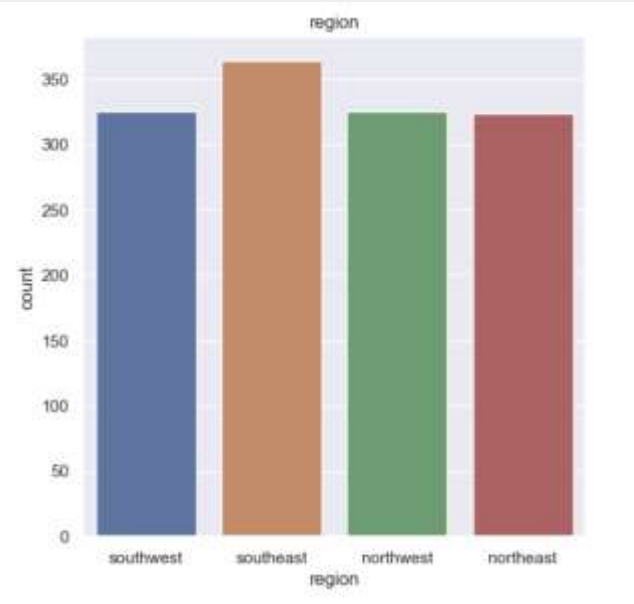

Fig. 6: Region Distribution

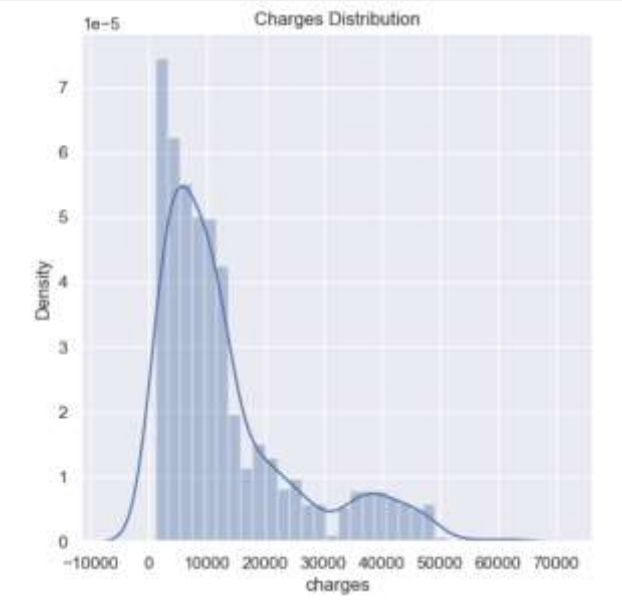

Fig. 7: Charges Distribution

RESULTS and ANALYSIS

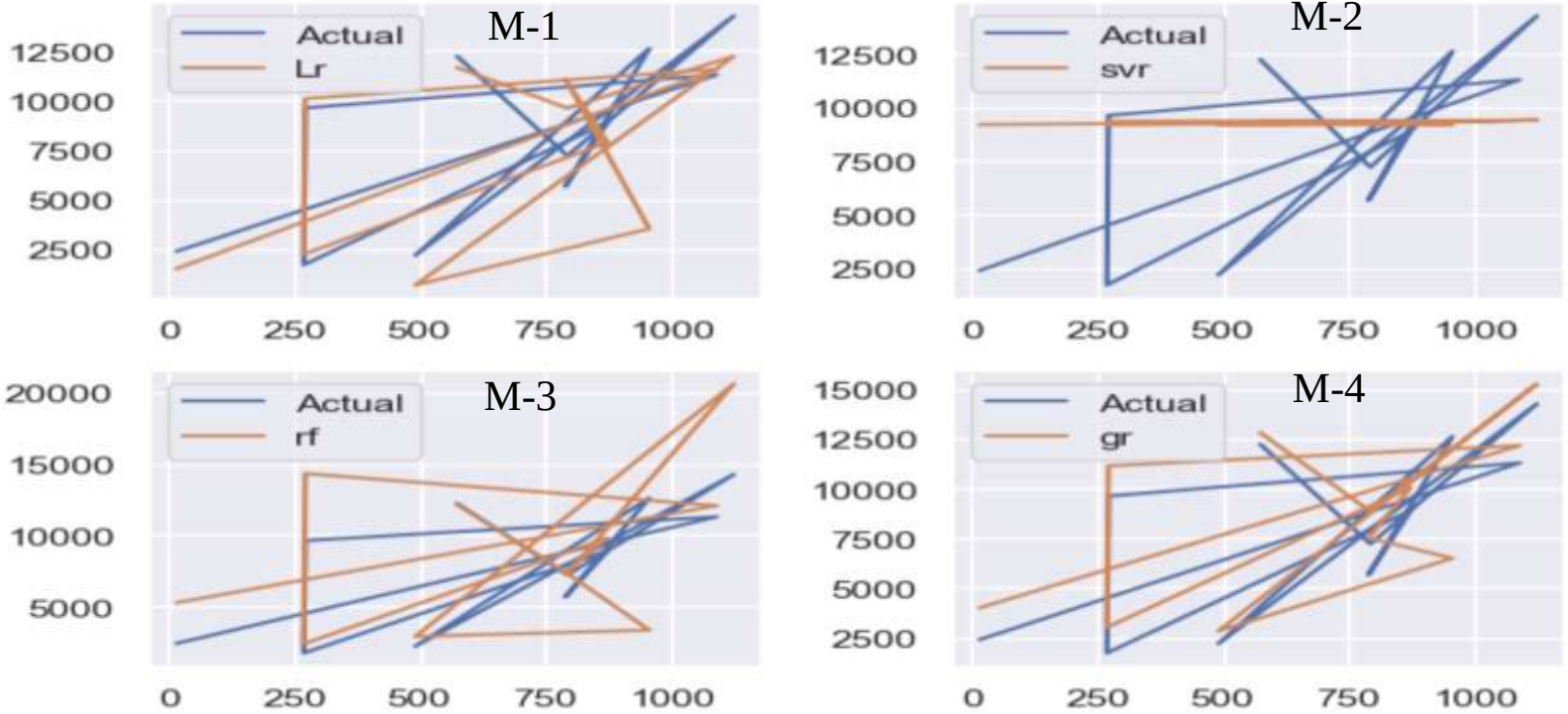

Fig. 8: Compare Performance Visually

RESULTS and ANALYSIS

Table 2: Accuracy Score of Different ML Algorithms

Algorithms Name	Accuracy Score
Linear Regression	74%
Random Forest Regressor	83%
Support Vector Regressor	8%
Gradient Boosting Regressor	86%

RESULTS and ANALYSIS

:

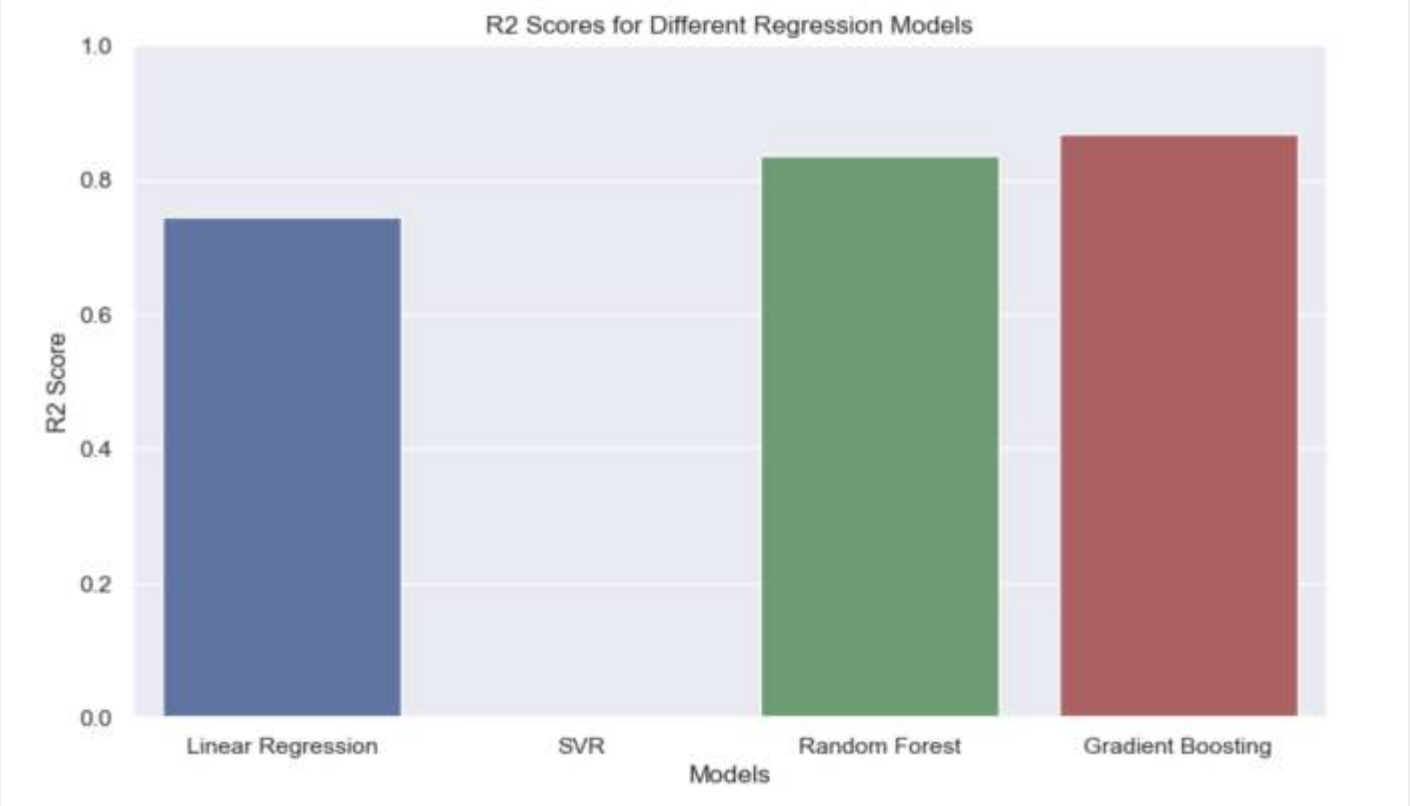

Fig. 9: R2 Scores for Different Regression Models

RESULTS and ANALYSIS

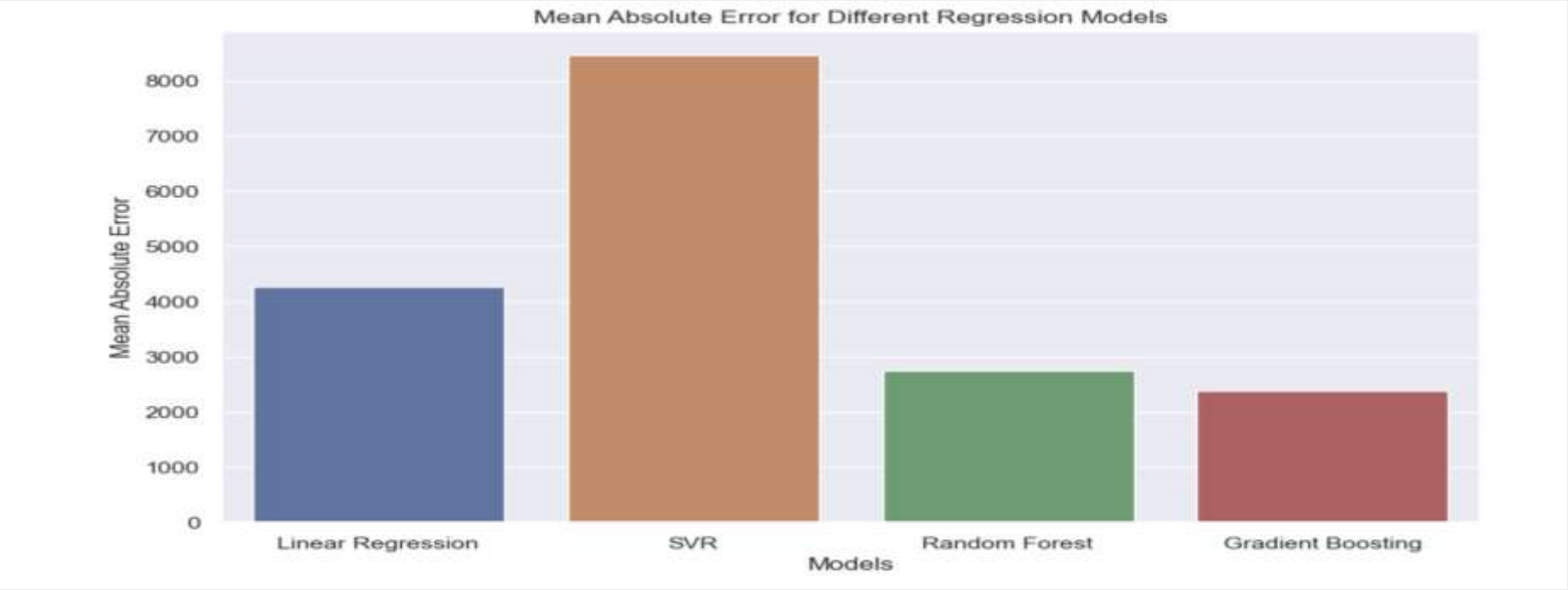

Fig. 10: Mean Absolute Error for Different Regression Models

CONCLUSION

➢ Based on the analysis among the algorithms Linear Regression, Random Forest Regressor, Support Vector Regressor and Gradient Boosting Regressor. The Gradient Boosting Regressor algorithm is giving the accuracy of 86% which is the highest among all.

➢Gradient Boosting Regressor model is used to predict the premium of the life insurance applicant and can be suggested as better model to the Life Insurance Company.

FUTURE SCOPE

➢Continuous improvements in machine learning algorithms enhance accuracy and efficiency in assessing premium profiles.

➢Availability of large datasets enables the development of more sophisticated models for predicting individual premium.

➢Machine learning facilitates better understanding and management of premiums, contributing to overall improvements in the Life Insurance industry.

➢Implementation of GIS (Geographical Information System) and finding the best prediction.

REFERENCES

1. Rahul Sahai, Ali Al-Ataby, Sulaf Assi, Manoj Jayabalan, Panagiotis Liatsis, Chong Kim Loy, Abdullah Al-Hamid, Sahar Al-Sudani, Maitham Alamran and Hoshang Kolivand, “Insurance Risk Prediction Using Machine Learning”, April 2023
2. Siti Nurasyikin Shamsuddin, Noriszura Ismail and R. Nur-Firyal, “Life Insurance Prediction and Its Sustainability Using Machine Learning Approach”, July 2023
3. Keshav Kaushik, Akashdeep Bhardwaj, Ashutosh Dhar Dwivedi, and Rajani Singh, “Machine Learning-Based Regression Framework to Predict Health Insurance Premiums”, June 2022
4. Noorhannah Boodhun & Manoj Jayabalan, “Risk prediction in life insurance industry using supervised learning algorithms” April 2018
5. Prasanta Baruah & Pankaj Pratap Singh, “Risk Prediction in Life Insurance Industry Using Machine Learning Techniques — A Review”, September 2023

Thank You