2nd Intelligent Computing and Technology Conference (ICTCon)
at
Central Institute of Technology Kokrajhar, Assam, India

2024

PREDICTING INSURANCE PREMIUM FOR LIFE INSURANCE APPLICANTS USING MACHINE LEARNING

#25

AUTHOR(S)

Subhash Kumar Wary
Akher Uddin Ahmed
Birhang Borgoyary

Central Institute of Technology Kokrajhar

u20csel2110@cit.ac.in

ABOUT US

We are final-year B.Tech students from the Department of Computer Science and Engineering at the Central Institute of Technology Kokrajhar. We are passionate about leveraging technology to solve real world challenges and look forward to contributing to the ever evolving tech landscape.

2nd ICTCon at
CIT Kokrajhar, Assam

Springer

u20csel2110@cit.ac.in

AGENDA

INTRODUCTION

LITERATURE REVIEW

PROBLEM STATEMENT

METHODOLOGY

RESULTS

CONCLUSION

u20csel2110@cit.ac.in

Introduction

1. Life Insurance secures financial well-being in case of unexpected demise.

2. Premium assessment traditionally relies on human underwriters analyzing factors like age, health, and lifestyle.

3. ML revolutionizes insurance by enhancing accuracy and efficiency in premium assessment.

4. ML streamlines underwriting, reduces costs, and improves premium predictions for better coverage options.

u20csel2110@cit.ac.in

LITERATURE REVIEW

1. Transition from actuarial methods to ML techniques for precise, efficient, and data-driven underwriting.
2. ML models like RF, LR, and GB improve risk prediction and handle complex datasets effectively.
3. Automation reduces time, cost, and biases while increasing transparency and fairness.
4. Dependence on data quality, potential bias, and ethical concerns in ensuring fairness across demographics.

LITERATURE REVIEW

Table 1: Related Work

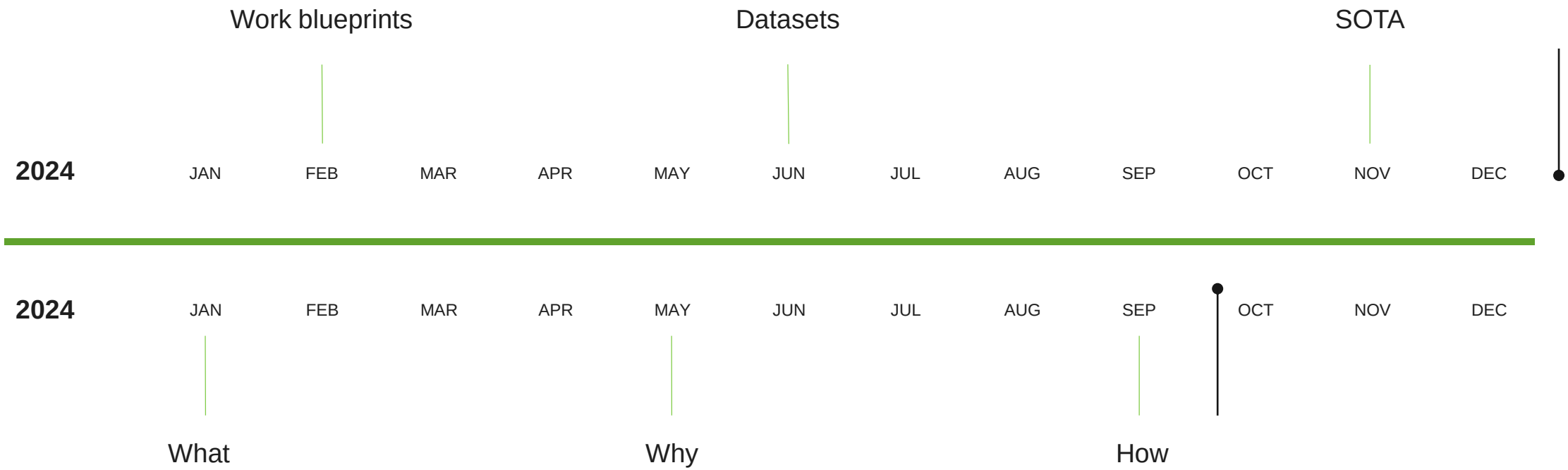
Authors & Year	Objectives	Methodology	Results
Junedi and Mauritsius (2020)	to predict the risk level of life insurance applicants using ML.	Three ML classifiers as SVM, Naive Baye's & RF are used.	Precision using RF, SVM using a linear kernel & NB are 0.85, 0.72 and 0.49 respectively.
Chang et al. (2022)	to develop an automated diabetes diagnosis system using ML models.	It discusses SMOTE and ADASYN for dealing with imbalanced datasets & used DT, RF, NB, KNN and LR.	5 types of classifiers (DT, RF, NB, KNN and LR) for predicting diabetes and RF achieved highest accuracy.
Kaushik et al. (2022)	to predict health insurance premiums by using the concept of AI and ML in healthcare.	The training and evaluation of an ANN and LR based regression model is done to predict health premiums.	ANN based regression model for predicting health insurance premium with accuracy of 92.72%.
Singh et al. (2017)	To classify and predict the different level of diabetes in the patients	Data mining approach	Comparison with association rule mining based approach
Baruah et al. (2023)	to improve risk assessment in life insurance industries of the applicants using predictive analytics	Geographical Information Systems (GIS) and ML approaches	ML approaches are applied to predict the applicants risks on both the datasets.
Baruah and Singh	Review on role of Risk prediction for customers in insurance companies.	Risk classification is done based on their risk levels by grouping of customers with ML	underwriters role is to calculate the premium based on the calculated risk of customers

PROBLEM STATEMENT

To build an efficient ML model to predict life insurance premiums and recommend suitable policies based on predictions.

u20csel2110@cit.ac.in

SOTA METHODS



u20csel2110@cit.ac.in

METHODOLOGY

1. Pre-processing of data has been done before training the model.
2. 80% (1070) data has been taken for training and 20% (268) data taken for testing.
3. The model has trained with the various algorithms such as Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor.
4. Various machine learning algorithms were compared and Gradient Boosting Regressor model was selected as the best option for prediction.
5. Predict the result using the trained model on the input data and provide the prediction.

BLOCK DIAGRAM

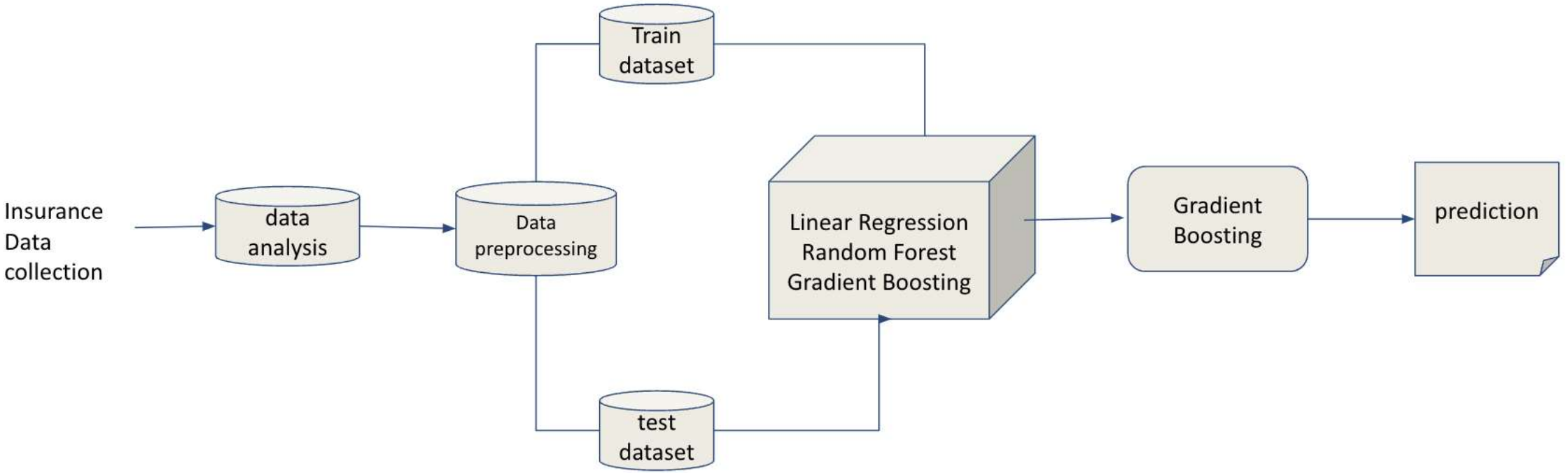


Fig.1 System Architecture

DATASET

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

Table 2: Insurance Dataset

	A	B	C	D	E	F	G
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

u20csel2110@cit.ac.in

DATASET

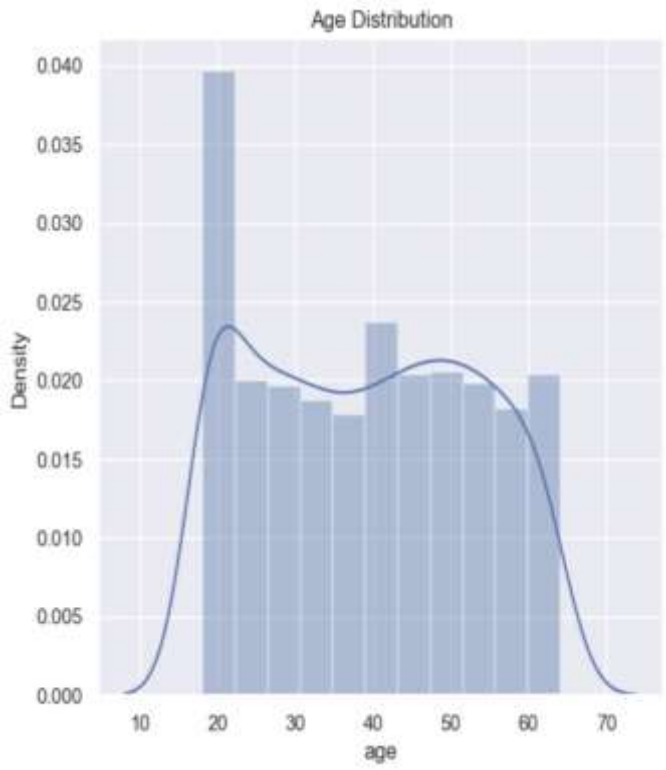


Fig. 2: Age Distribution

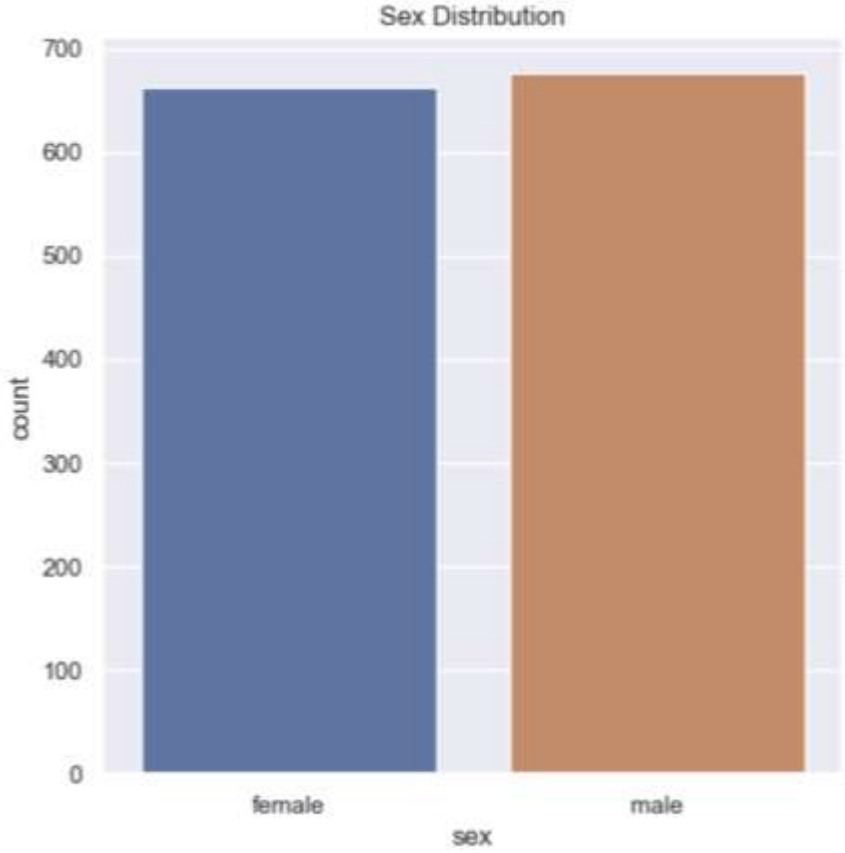


Fig. 3: Sex Distribution

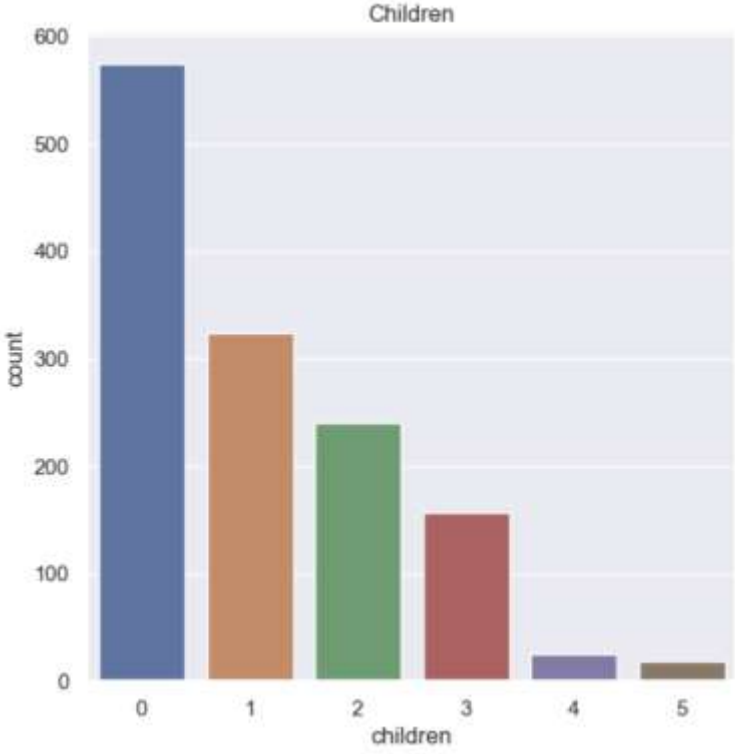


Fig. 4: Children Distribution

DATASET

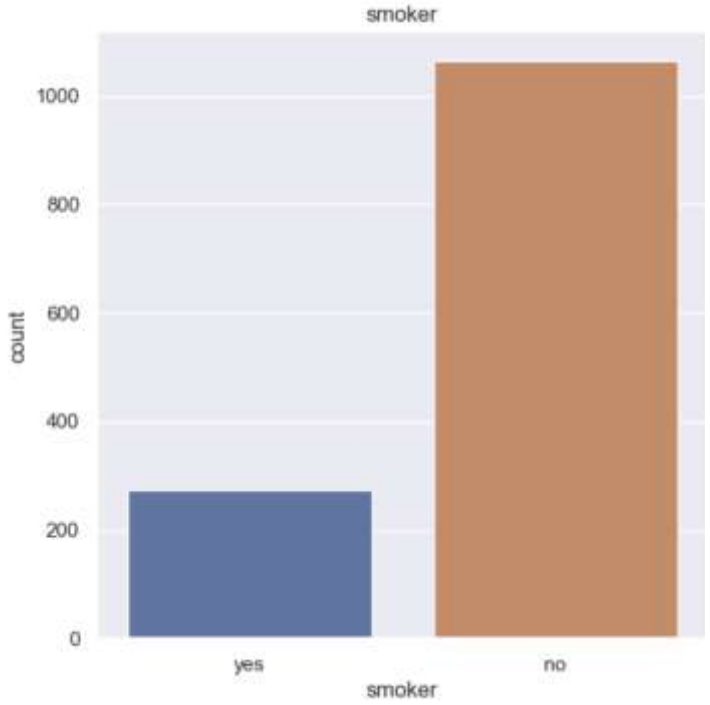


Fig. 5: Smoker Distribution

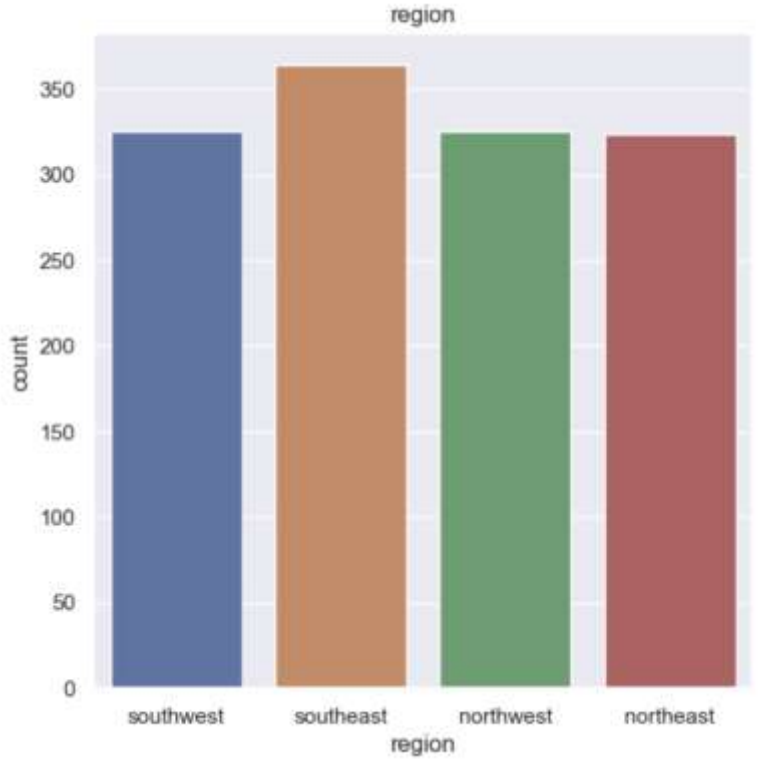


Fig. 6: Region Distribution

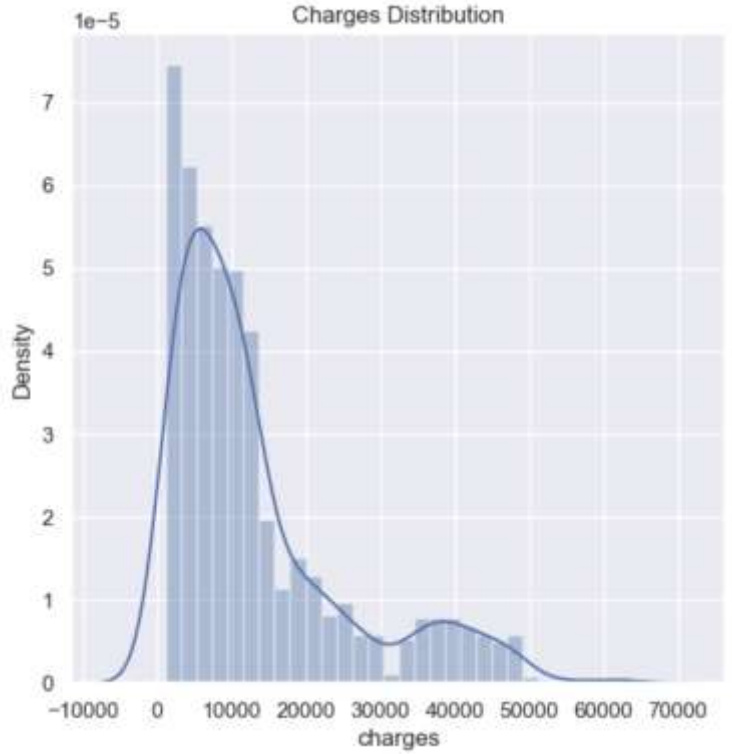


Fig. 7: Charges Distribution

u20csel2110@cit.ac.in

2nd ICTCon at
CIT Kokrajhar, Assam

ICTCon 2024

EXPERIMENTAL RESULTS

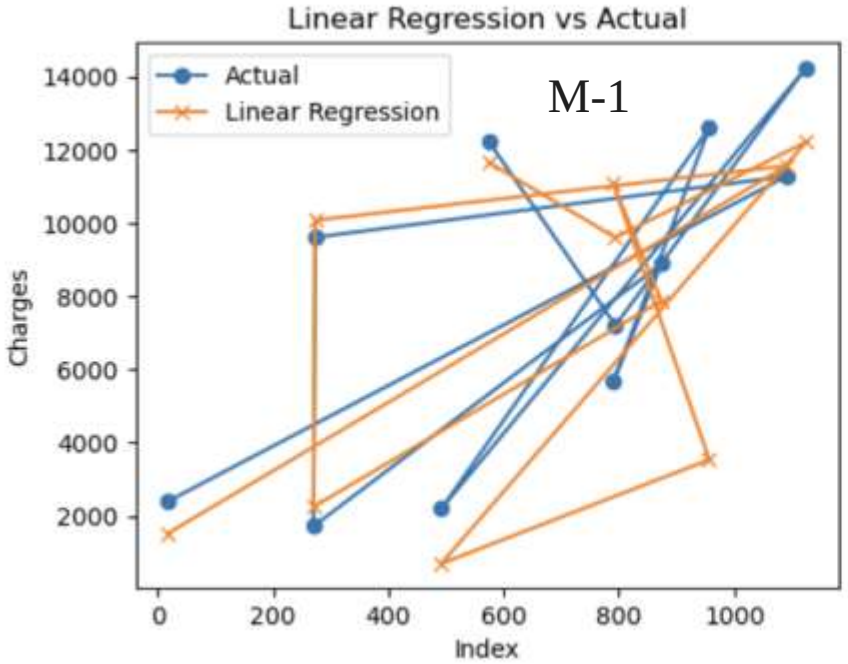
MODELS ACCURACY:

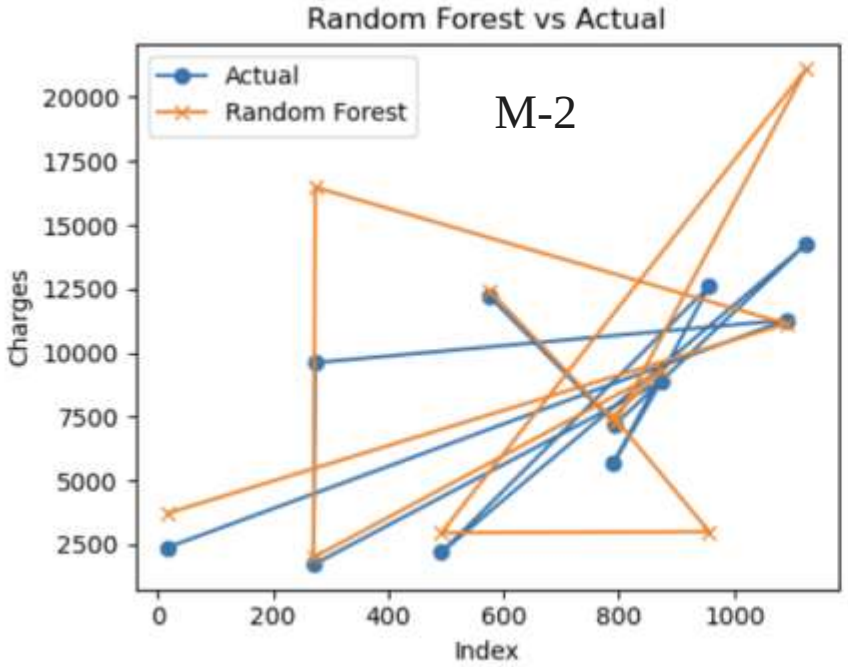
1. LINEAR REGRESSION (LR): 74%
2. RANDOM FOREST REGRESSOR (RF): 83%
3. GRADIENT BOOSTING REGRESSOR (GB): 86%

Central Institute of Technology Kokrajhar, Subhash Kumar Wary, Akher Uddin Ahmed and Birhang Borgoyary

u20csel2110@cit.ac.in

ANALYSIS





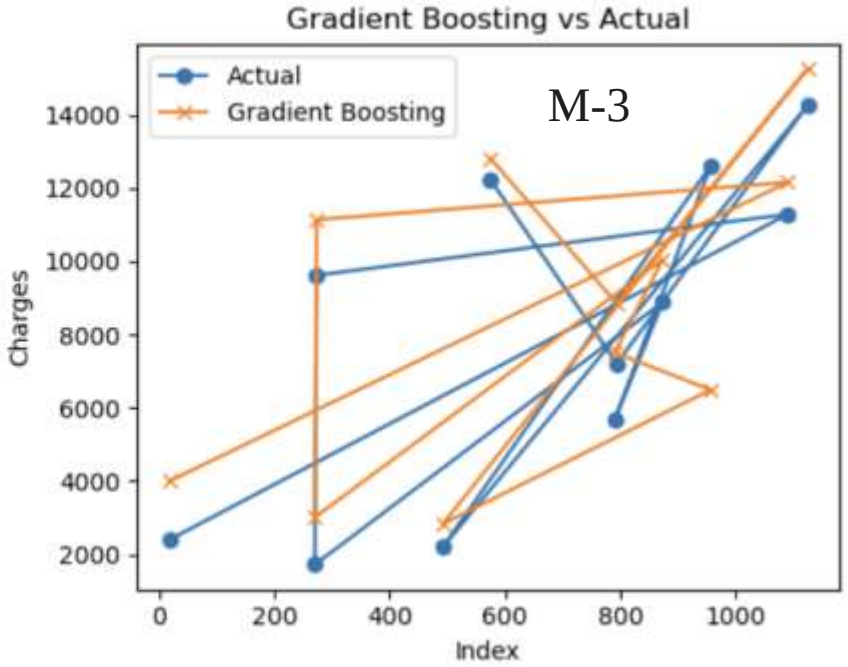


Fig.8 Compare Performance Visually

RESULTS

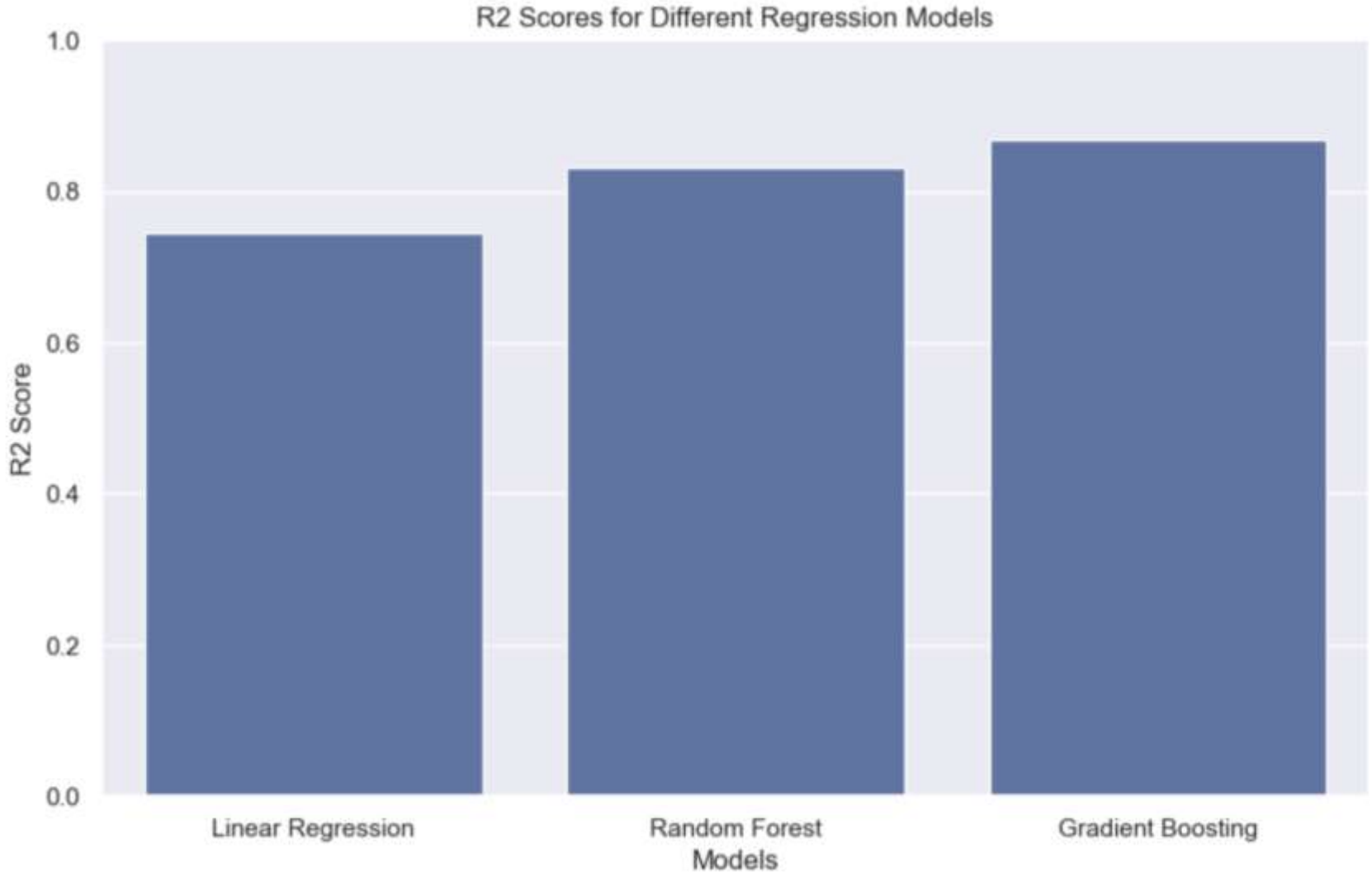


Fig. 9 R2 Scores for Different Regression Models

u20csel2110@cit.ac.in

RESULTS

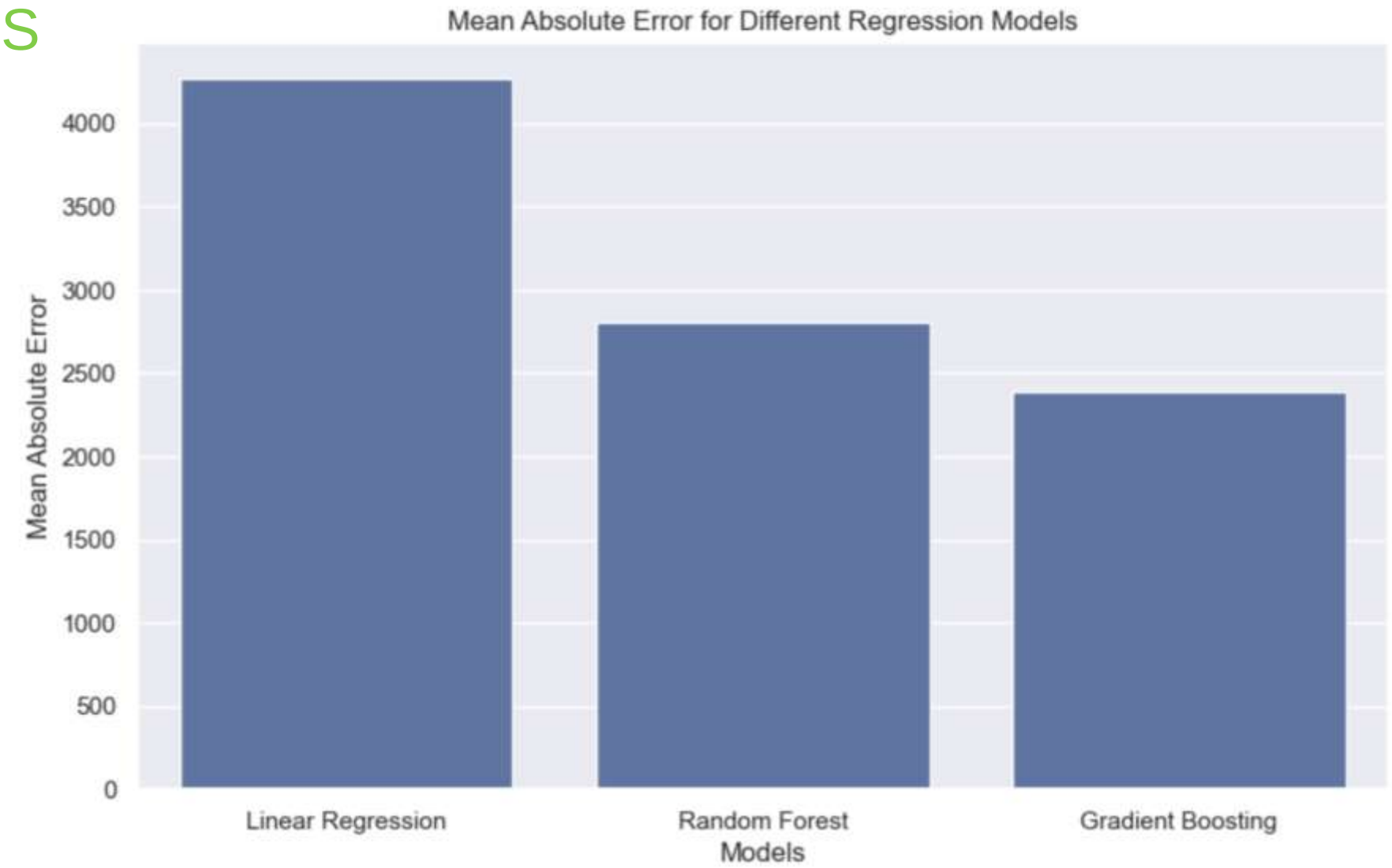
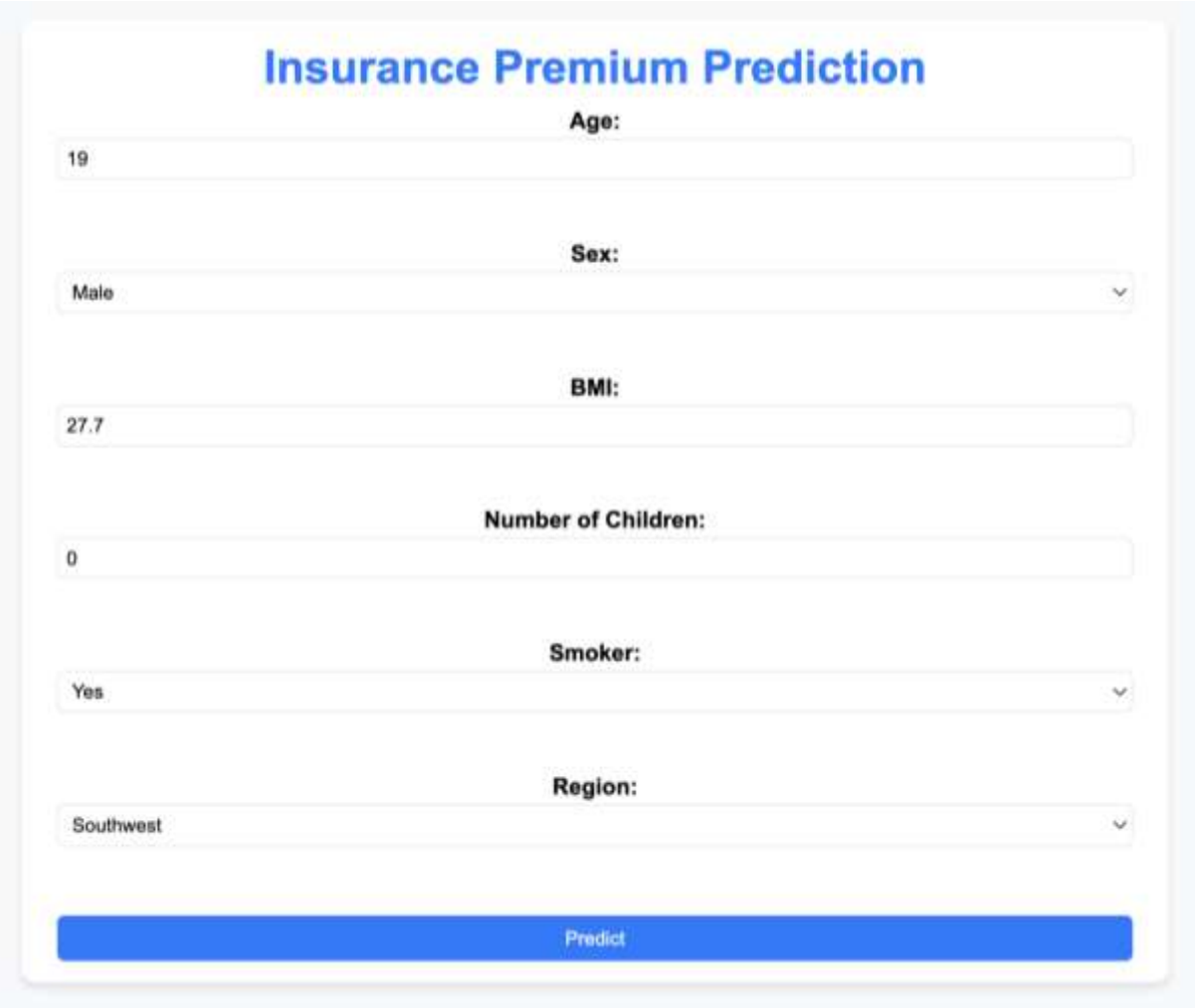


Fig. 10 Mean Absolute Error for Different Regression Models

INTERFACE



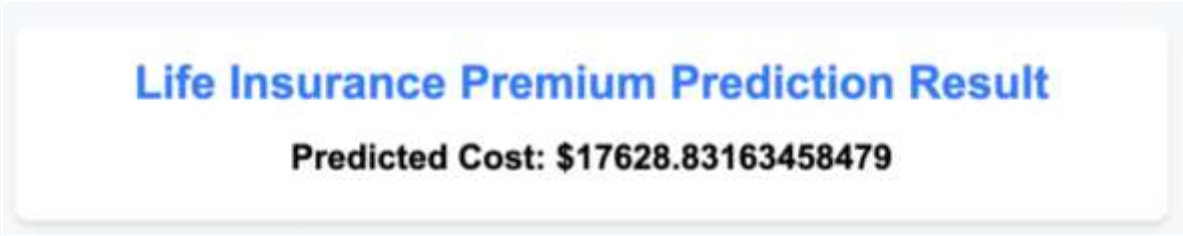


Fig. 11 Interface created using Flask and Python

u20csel2110@cit.ac.in

DISCUSSION

LINEAR REGRESSION (LR) 74%	RANDOM FOREST (RF) REGRESSOR 83%	GRADIENT BOOSTING (GB) REGRESSOR 86%

1. Our study focused on ML algorithms for predicting life insurance premiums.
2. Models used are Linear Regression (LR), Random Forest (RF) Regressor, Gradient Boosting (GB) Regressor.
3. All models effectively predicted premiums with high accuracy.
4. GB Regressor achieved the best performance with 86% accuracy.

CONCLUSION

EXISTING

Three ML classifiers as SVM, Naïve Baye's & RF are used.

Precision using RF, SVM using a linear kernel & NB are 0.85, 0.72 and 0.49 respectively.

PROPOSED

Three ML regressors as LR, RF and GB are used.

Precision using LR, RF and GB are 74%, 83% and 86% respectively.

u20csel2110@cit.ac.in

MEET THE TEAM

SUBHASH KUMAR WARY

Member 1

AKHER UDDIN AHMED

Member 2

BIRHNAG BORGOYARY

Member 3

THANK YOU

Subhash Kumar Wary

u20csel2110@cit.ac.in

+918812013021