**A Project report on**

**Medical Insurance Premium Prediction**

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

Submitted by

T. Bhagya Laxmi

(20H51A05F7)

Yasmeen

(20H51A05G1)

B. Sree Harsha

(20H51A05M9)

Under the esteemed guidance of

Mr. B. SIVAIAH

(Associate Professor)



**Department of Computer Science and Engineering**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

(UGC Autonomous)

\*Approved by AICTE \*Affiliated to JNTUH \*NAAC Accredited with A+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

**2020- 2024**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**CERTIFICATE**

This is to certify that the Major Project Phase I report entitled

**“Medical Insurance Premium Prediction ”** being submitted by T. Bhagya Laxmi (20H51A05F7),Yasmeen (20H51A05G1), B.Sree Harsha (20H51A05M9) in partial fulfillment for the award of **Bachelor of Technology in Computer** **Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Mr.B.SIVAIAH**

**Dr. Siva Skandha Sanagala**

**Associate Professor**

**Dept. of CSE**

**Associate Professor and HOD**

**Dept. of CSE**

**ACKNOWLEDGEMENT**

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Mr.B.SIVAIAH, Associate Professor** , Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

|  |  |
| --- | --- |
| T. Bhagya Laxmi | 20H51A05F7 |
| Yasmeen | 20H51A05G1 |
| B. Sree Harsha | 20H51A05M9 |

Medical Insurance Premium Prediction



**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **CHAPTER** |  |  |  |
| **NO.** |  | **TITLE** | **PAGE NO.** |
|  | ABSTRACT | | iii |
| **1** | **INTRODUCTION** | | 1 |
|  | 1.1 | Problem Statement | 2 |
|  | 1.2 | Research Objective | 2 |
|  | 1.3 | Project Scope and Limitations | 3 |
| **2** | **BACKGROUND WORK** | | 4 |
|  | 2.1. Health Insurance Cost Prediction by Using Machine Learning | |  |
|  |  | 2.1.1.Introduction | 5 |
|  |  | 2.1.2.Merits and Demerits | 6 |
|  |  | 2.1.3.Implementation | 7 |
|  | 2.2. Medical Insurance Cost Prediction using Machine Learning | |  |
|  |  | 2.2.1.Introduction | 7-8 |
|  |  | 2.2.2.Merits and Demerits | 8-9 |
|  |  | 2.2.3.Implementation | 9-10 |
| **3** | **RESULTS AND DISCUSSION** | | 11 |
|  | 3.1. | Performance metrics | 12-13 |
| **4** | **CONCLUSION** | | 14-15 |
| **5** | **REFERENCES** | | 16-17 |



CMRCET B. Tech (CSE) Page No i

Medical Insurance Premium Prediction



**List of Figures**

**FIGURE**

**NO.** **TITLE** **PAGE NO.**

3.1 table 12

3.2 flowchart 13



CMRCET B. Tech (CSE) Page No ii

Medical Insurance Premium Prediction



**ABSTRACT**

Insurance is a policy that helps to cover up all loss or decrease loss in terms of expenses incurred by various risks. A number of variables affect how much insurance costs. These considerations of different factors contribute to the insurance policy cost expression. Machine Learning (ML) in the insurance sector can make insurance more effective. In the domains of computational and applied mathematics the machine learning (ML) is a well-known research area. ML is one of the computational intelligence aspects when it comes to exploitation of historical data that may be addressed in a wide range of applications and systems. There are some limitations in ML so; Predicting medical insurance costs using ML approaches is still a problem in the healthcare industry and thus it requires few more investigation and improvement. Using the machine learning algorithms, this study provides a computational intelligence approach for predicting medical insurance costs. The proposed research approach uses Linear Regression, Decision Tree Regression and Gradient Boosting Regression and also streamlit as a framework. We had used a medical insurance cost dataset that was acquired from the KAGGLE repository for the cost prediction purpose, and machine learning methods are used to show the forecasting of insurance costs by regression model comparing their accuracies.



CMRCET B. Tech (CSE) Page No iii

Medical Insurance Premium Prediction



**CHAPTER 1**

**INTRODUCTION**



CMRCET B. Tech (CSE) Page No 1

Medical Insurance Premium Prediction



**CHAPTER 1**

**INTRODUCTION**

**1.1. Problem Statement**

In the modern world, it is essential to have health insurance, The goal of this project is to help individuals understand the amount of money they may need for health insurance based on their personal health status. This can assist individuals in focusing more on the health related aspects of insurance rather than the unnecessary ones. Most people have a relationship with a public or private health insurance provider. The factors that influence insurance costs vary from company to company.

**1.2 Research Objective**

* To develop a predictive model that accurately estimates medical insurance costs for individuals based on relevant demographic, health-related, and lifestyle factors. The goal is to provide insurance companies and policyholders with a reliable tool for pricing and planning, ultimately improving the transparency and fairness of the insurance market.
* This objective outlines the main aim of your research, which is to create a predictive model for medical insurance costs, and highlights the potential benefits of such a model for both insurance providers and individuals seeking coverage

CMRCET B. Tech (CSE) Page No 2



Medical Insurance Premium Prediction



**1.3 Project Scope and Limitations**

**Scope:-**

**Personalization:** Advanced prediction models can consider individual characteristics, such as age, gender, medical history, and lifestyle factors, to personalize premium rates.

**Cost Control:** Accurate prediction models help insurance companies manage costs by aligning premiums with expected claims, reducing underwriting losses, and minimizing adverse selection.

**Customer Retention:** Personalized premium rates can enhance customer satisfaction and retention, as policyholders feel their premiums are fair and tailored to their needs.

**Portfolio Management:** Insurers can use premium prediction to effectively manage their portfolios, balancing risk and revenue to optimize their business.

**Regulatory Compliance:** Premium prediction must comply with local and national insurance regulations, ensuring fairness and non-discrimination.

**Limitation :**

1. **Data Availability and Quality**: Premium prediction relies heavily on data, and the accuracy of the predictions is limited by the quality and availability of data. Inaccurate or incomplete data can lead to unreliable predictions.
2. **Privacy Concerns:** Collecting and using personal data for premium prediction may raise privacy concerns. Insurers need to handle customer data responsibly and in accordance with relevant privacy regulations.
3. **Regulatory Constraints**: Insurance companies must adhere to regulatory guidelines when determining premium rates. This can limit the flexibility in setting rates based on predictive models.



CMRCET B. Tech (CSE) Page No 3

Medical Insurance Premium Prediction



**CHAPTER 2**

**BACKGROUND**

**WORK**



CMRCET B. Tech (CSE) Page No 4

Medical Insurance Premium Prediction



**CHAPTER 2**

**BACKGROUND WORK**

**2.1 Health Insurance Cost Prediction by Using Machine Learning**

**2.1.1. Introduction**

The goal of this research is to help individuals understand the amount of money they may need for health insurance based on their personal health status. This can assist individuals in focusing more on the healthrelated aspects of insurance rather than the unnecessary ones. In the modern world, it is essential to have health insurance, and most people have a relationship with a public or private health insurance provider. The factors that influence insurance costs vary from company to company. Additionally, some people in rural areas may not be aware that the Indian government offers free health insurance to those who are below the poverty line. However, the process can be complex, and some rural residents either get private health insurance or make no investment at all. Additionally, people may be vulnerable to being misled into paying for expensive health insurance that they don't need. Our research does not provide an exact amount required by any specific health insurance provider, but it does give a general sense of the cost a person may incur for their own health insurance. This is a preliminary estimate and does not adhere to any particular company, so it should not be the only factor considered when choosing health insurance. Early estimation of health insurance costs can help individuals consider the required amount more thoughtfully.

CMRCET B. Tech (CSE) Page No 5



Medical Insurance Premium Prediction



**2.1.2 ADVANTAGES**

Risk Management: Predictive modeling helps insurance companies assess and manage risk effectively.

Personalization: Premiums can be tailored to individual health and lifestyle factors, making insurance more affordable for many.

Cost Control: Predictive models enable the identification of high-risk policyholders and the implementation of preventive measures to reduce claims.

Competitive Edge: Accurate premium predictions can give insurance companies a competitive advantage in the market.

Customer Satisfaction: Tailored premiums can lead to higher customer satisfaction as policyholders feel their premiums are more fairly priced

**Disadvantages**

Privacy Concerns: Extensive data collection for predictive modeling can raise privacy concerns and potential data breaches.

Bias and Fairness: Predictive models may introduce bias if training data contains historical disparities, leading to unfair premium pricing.

Complexity: Developing and maintaining accurate predictive models requires substantial computational resources and expertise.

CMRCET B. Tech (CSE) Page No 6



Medical Insurance Premium Prediction



**2.1.3.Implementation**

**Training:** After the necessary data has been formatted and prepared, the model can begin its training and testing phases. A key focus during the training phase is choosing the appropriate model for the task at hand. This may involve deciding on the optimal modelling strategy or determining the best parameter values for a particular model (V. Roth, 2014) In some cases, this process is referred to as model selection because various models may be tested and the one that performs the best, is ultimately chosen , which is created using extreme vectors and points called support vectors.

**Prediction:** The model used for predicting the insurance sum for health was based on the relationship between certain features and the label. The accuracy of this prediction was determined by the extent to which the expected value matched the actual amount. In order to improve the accuracy, the model employed various characteristics, methods, and train-test split sizes. It was found that the amount of data used for training had a significant impact on the accuracy, with a larger train size leading to better results. The model also employed multiple algorithms in order to forecast the premium amount, and showed how each attribute affected the outcome(Kaggle, Regression data)

**2.2 Medical Insurance Cost Prediction using Machine Learning**

**2.2.1. Introduction**

We live on a planet full of threats and uncertainty. Including People, households, durables, properties are exposed to different risks and the risk levels can vary. These risks range from risk of health diseases to death if not get protection, and loss in property or assets[1]. But, risks cannot usually be avoided, so the world of finance has developed numerous products to shield individuals and organizations from these risks by using financial capital to shield them.



CMRCET B. Tech (CSE) Page No 7

Medical Insurance Premium Prediction



Therefore Insurance is one of the policies that either decreases or removes loss costs incurred by various risks. The value of insurance in the lives of individuals. That's why it becomes important for insurance companies to be sufficiently precise to measure the amount covered by this specific policy and the insurance charges which must be paid for it. Various parameters or factors play an important role in estimating the insurance charges and Each of these is important. If any factor is omitted or changed when the amounts are computed then, the overall policy cost changes. It is therefore very critical to carry out these tasks with high accuracy. So, the possibility of human mistakes are high so insurance agents also use different tools to calculate the insurance premium. And thus ML is beneficial here. ML may generalize the effort or method to formulate the policy. These ML models can be learned by themselves. The model is trained on insurance data from the past. The model can then accurately predict insurance policy costs by using the necessary elements to measure the payments as its inputs. This decreases human effort and resources and improves the company's profitability. Thus the accuracy can be improved with ML. Our goal is to predict insurance costs. The value of insurance fees is based on different variables. As a result, insurance fees are continuous. Regression is the best choice available to fulfill our needs. We use multiple linear regression in this analysis since there are many independent variables used to calculate the dependent(target) variable. For this study, the dataset for cost of health insurance is used .[2] Preprocessing of the dataset done first. Then we trained regression models with training data and finally evaluated these models based on testing data. In this article, we used several models of regression, for example, multiple linear regression, Decision Tree Regression and Gradient Boosting Regression.

**2.2.2. Advantages**

Customer Satisfaction: Tailored premiums can lead to higher customer satisfaction as policyholders feel their premiums are more fairly priced.

Fraud Detection: Predictive models can help detect fraudulent claims by identifying unusual patterns and behaviors.

Efficiency: Automation of premium prediction processes can improve efficiency within insurance companies.



CMRCET B. Tech (CSE) Page No 8

Medical Insurance Premium Prediction



Data-Driven Decision Making: Predictive modeling encourages data-driven decision-making, improving underwriting processes.Long-Term Planning: Better premium predictions allow for more accurate long-term planning for insurance companies

.Financial Stability: Improved risk assessment leads to better financial stability for insurance providers

**Disadvantages**

Ethical Concerns: Predictive modeling raises ethical questions about fairness, transparency, and the potential for discrimination.

Resistance from Policyholders: Some may be resistant to premium predictions using models, as they may not fully understand the process or have concerns about its accuracy.

Limited Data for Some Factors: Not all relevant factors affecting an individual's health and insurance risk are easy to capture in data, limiting prediction accuracy.

Maintenance Costs: Continuous maintenance and updates of predictive models can be costly.

Overreliance on Models: Overreliance on predictive models may lead to a lack of human judgment and expertise in underwriting and risk assessment.

**2.2.3.Implementation**

The objective of the study is to prophetic the insurance cost supported age, BMI, kid number, the region of the person living, sex, and whether or not a shopper is smoking or not, drinks alcohol or not, having diabeties or not . These options contribute to our target variable prediction of insurance costs. For the measuring of the value of insurance, many regression models are applied during this study. The dataset is split into 2 sections. One half for model training and also the other part for model analysis or testing.



CMRCET B. Tech (CSE) Page No 9

Medical Insurance Premium Prediction



During this study, the info set is separated into twopart the first half is termed coaching knowledge and also the second called take a look at data, training data makes up for eighty percent of the whole data used, and the rest for test data. all of those models are trained with the training data part and so evaluated with the test data. The accuracy is checked with the assistance of r2 score.



CMRCET B. Tech (CSE) Page No 10

Medical Insurance Premium Prediction



**CHAPTER 3**

**RESULTS AND**

**DISCUSSION**



CMRCET B. Tech (CSE) Page No 11

Medical Insurance Premium Prediction



**CHAPTER 3**

**RESULTS AND DISCUSSION**

The data set we're using is Kaggle's medical cost personal dataset, which contains anonymous information about persons. Age and BMI are both continuous variables, whereas gender and smoking status are all categorical.

The following are the characteristics in this dataset

Age - Age of primary beneficiary person

Sex - Gender of insurance contractor (female/male)

BMI -Body mass index (BMI) is a metric that compares the size of a person's body to the health of that person's body Smoker -If your primary insurance company smokes

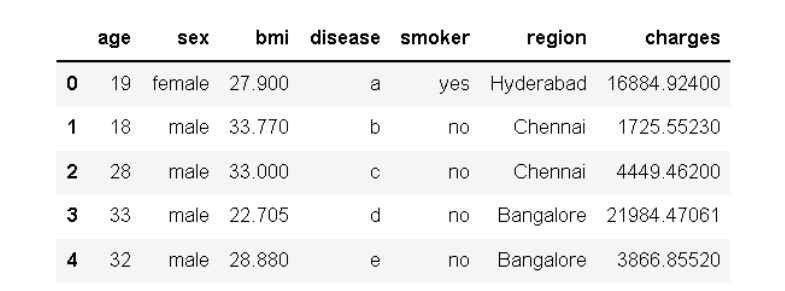


Fig-3.1-prediction table



CMRCET B. Tech (CSE) Page No 12

Medical Insurance Premium Prediction

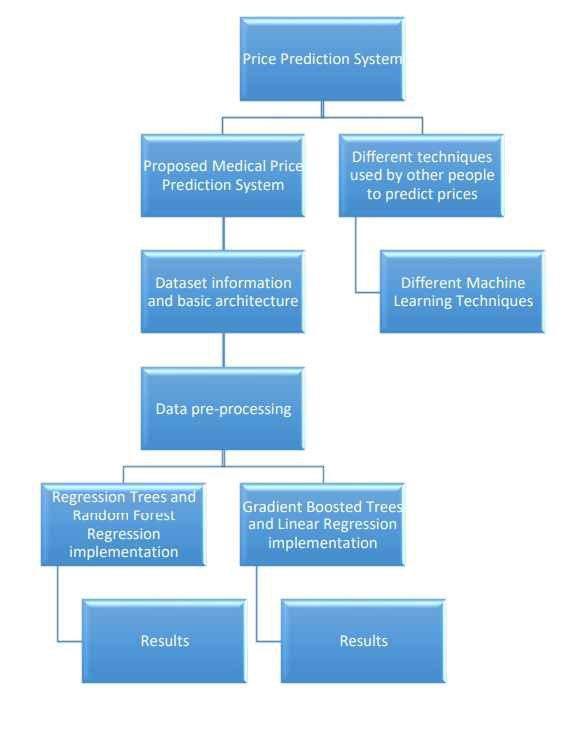


Fig-3.2-flowchart



CMRCET B. Tech (CSE) Page No 13

Medical Insurance Premium Prediction



CHAPTER 4

**CONCLUSION**



CMRCET B. Tech (CSE) Page No 14

Medical Insurance Premium Prediction

**CHAPTER 4**



**CONCLUSION**

* Our project has practical implications for healthcare providers, insurers, and individuals seeking insurance coverage.
* It can guide fairer pricing strategies and informed decision-making in the healthcare finance sector.
* Continued data collection and model refinement could enhance prediction accuracy.
* Exploring the impact of policy changes and new data sources can further enrich our insights.



CMRCET B. Tech (CSE) Page No 15

Medical Insurance Premium Prediction



**REFERENCES**



CMRCET B. Tech (CSE) Page No 16

Medical Insurance Premium Prediction



**REFERENCES**

* [Gupta, S., & Tripathi, P. (2016, February). A leading trend of data analytics with](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_) health insurance in India. In 2016 International Conference on Innovation and [Challenges in CS (ICICCS-INBUSH) (pp. 64-69). IEEE.](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_)
* [[Yerpude, P., Gudur, V.: Prediction modeling of crime dataset using data mining. Int. J.](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_) Data Min. [Knowl. Manage. Process (IJDKP) 7(4) (2017)](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_)
* [Grosan, C., Abraham, A.: Intelligent Systems: A Modern Approach, Intelligent](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_) Systems Reference Library Series. Springer, Cham (2011)
* [https://www.ijraset.com/research-paper/medical-insurance-cost-prediction-using-](https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.researchgate.net/publication/334535923_Management_Analysis_of_Uber&ved=2ahUKEwj-huqB2oD7AhWcaGwGHagiAK8QFnoECA8QAQ&usg=AOvVaw1mTOsWE4ktHZ4NGxFU-R0_)machine-learning



CMRCET B. Tech (CSE) Page No 17