# MEDICAL INSURANCE PREMIUM PREDICTION WITH MACHINE LEARNING

Prof. M. S. Patil

Assistant Professor, Artificial Intelligence & Data Science, FTC, Sangola, Maharashtra, India


Kulkarni Sanika,

Khurpe Sanjana

Student, Artificial Intelligence & Data Science, FTC, Sangola, Maharashtra, India

## Abstract

A machine learning method for predicting health insurance rates is presented in this article. With healthcare expenditures becoming more complex, it is critical for insurance companies and policyholders to accurately estimate insurance prices. Utilizing a dataset that included medical history, demographic data, and other pertinent variables, a variety of machine learning techniques, such as ensemble methods and regression, were used to create prediction models. R-Squared and mean absolute error were two measures used to assess these models' performance. According to the developed models' results, insurance premiums can be predicted with accuracy, offering useful information for insurance counteragents. This approach has the potential to optimize pricing strategies, enhance risk assessment, and improve decision-making in the healthcare insurance sector. Machine Learning-Based Prediction of Medical Insurance Premiums Make predictions about health insurance companies based on personal traits. A dataset of policyholder attributes (such as age, gender, BMI, number of children, smoking behaviors, and geography) was gathered and preprocessed .Divide the data into sets for testing and training. Create and train a model for an artificial neural network with TensorFlow and Karas. R-squared metrics and mean R-squared error were used to assess the performance of the model. created a high R-Squared predictive model that was accurate. determined the main determinants of insurance rates. Machine learning has shown promise in estimating healthcare costs. This experiment demonstrates how well machine learning predicts medical insurance rates. Insurance companies may offer more individualized insurance plans, expedite the underwriting process, and help customers make well-informed decisions about their healthcare coverage by creating these predictive models. The created model can help policyholders make educated judgments and insurance companies establish proper prices. In the long run, our research helps the insurance industry enhance data-driven techniques, which benefits insurers as well as insured individuals in general.

**Keywords** - Medical insurances, feature importance, underwriting, premium prediction, machine learning, predictive modelling, and data-driven decision-making.

**Novateur Publications**
**International Journal of Innovations in Engineering Research and Technology [IJIERT]**
**ISSN: 2394-3696 Website: ijiert.org**
**Volume 11, Issue 5, May - 2024**

## I.    INTRODUCTION

Predicting  medical insurance premiums using machine learning  involves using historical data on individuals' demographics , health factors, and insurance convergence to build models that can estimate future premiums for new customers. By leveraging algorithms like regression ,decision tress, or neural networks, insurances insurers can improve accuracy in pricing policies and better manage risk. This introduction sets the stage for exploring how machine learning can revolutionize insurance pricing, ensuring fairness and accuracy in premium assessments.

Background : Discuss the importance of accurately predicting medical insurance premiums in the healthcare industry. Highlight the challenges faced by insurance providers in determining premiums and the potential impact on policyholders.

Motivation : Explain the motivation behind using machine learning techniques for premium prediction .Discuss the  limitations of traditional actuarial methods and the potential benefits of data-driven approaches.

Research Objective : Clearly state the aim of the study, which is to develop and evaluate machine learning models for predicting medical insurance premiums based on various factors.

Significance : Highlight the significance of the research in advancing the field of healthcare analytics and its potential impact on insurance underwriting practices, policyholder satisfaction, and overall healthcare affordability.

Structure of the Paper : Provide an overview of the organization of the paper, outing the sections that will be covered in detail ,such as data collection , Methodology, experimental results and discussion.

## II. Literature Survey:

1."Predicting Health Insurance Costs Using Machine Learning Techniques" by Pratibha G. Joshi and Sunanda Dixit: This paper explores the application of machine learning algorithms such as Linear Regression, Decision Trees and Random Fore sets to predict health insurance costs. It compares the performance of these algorithms influencing premium prediction accuracy.

2. "Machine Learning Techniques For Predicting Insurance Premiums" by A. Khalfan , Hassan ,and M.S Ansari: This study investigates various machine learning models for predicting insurance premiums, Gradient Boosting and Neural Networks. It discusses feature selection methods and model evaluation techniques to optimize prediction accuracy.

3."Predicting Health Insurance Premiums : A Comparatives Study of Machine Learning Techniques" by S. Gupta and  S. Sharma: This research compares the effectiveness of machine learning algorithms such KNN, Naïve  Bayes ,  and Ensemble methods in predicting health insurance premiums. It analyzes the impact of different features sets and preprocessing techniques on prediction performance.

4."Deep Learning  Approaches for Health Insurance Premium Prediction" by R. S. Raj and S. Kumar : This paper explores the application of deep learning techniques, including Convolution Neural Network and Recurrent Neural Networks, for health insurance premium prediction. It discusses of using deep learning models in this context.

5."Feature Selection Techniques for Medical Insurance Premium Prediction" by M. A. Rahman and S.Begum: this study investigates various feature Selection methods, such as Wrapper , Filter , and Embedded approaches,

Novateur Publications
**International Journal of Innovations in Engineering Research and Technology [IJIERT]**
**ISSN: 2394-3696 Website: ijiert.org**
**Volume 11, Issue 5, May - 2024**

to identify the most relevant predictors fir medical insurance premium prediction. It compares the performance of different feature selection techniques and their impact on prediction accuracy.

6. "Fairness in Medical Insurance Premium Prediction : A Machine Learning Perspective" by L .Zhang and H. Wang : This research examines the issue of fairness in medical insurance premium prediction and discusses techniques for mitigating bias and discrimination in machine learning models. It explores methods for promoting fairness and equality in premium assessments across different demographic groups.

7."Temporal Analysis of Medical Insurance Premiums Using Machine Learning" by J. Chen and X . Li :
This paper investigates the temporal patterns in medical insurance premiums and explores how machine learning models can capture and predict fluctuations over time. It discusses the implications of temporal analysis for premium pricing and risk management strategies.

This studies collectively prove insights into the information / application of machine learning techniques for medical insurance premium prediction , converging various algorithms feature selection methods , fairness considerations , and temporal analysis approaches.

## II. EXISTING WORK AND PROPOSED WORK

### A .Existing Work:

➢ **DATA COLLECTION** : The existing system collects historical data on insured individuals , including demographics(age , gender , location) , health factors (BMI ,pre-existing conditions) , and insurance Converge details.

➢ **DATA PREPROCESSING** : The collected data undergoes preprocessing steps such as cleaning , normalization and feature engineering to prepare it for analysis.

➢ **Feature Selection :**  Relevant features affecting insurance premiums are identified using techniques like correlation analysis , feature importance ranking , or domain knowledge**.**

➢ **Model Selection :** Various machine learning algorithms such as Linear Regression , Decision Trees , Random Forest , Gradient Boosting , and Neural Networks are evaluated for their suitable in predicting insurance premiums.

➢ **Model Training** : The selected machine learning models are trained using the preprocessed data , with the aim of learning patterns and relationships between input features and insurance premiums.

➢ **Model Evaluation** :  The trained models are evaluated using metrics such as Mean Squared Error (MSE) , Mean Absolute Error(MAE) or Root Mean Squared Error (RMSE) to assess their predictive performance.

➢ **Hyperparameter Tuning** : Hyperparameters of the machine learning models are fine – tuned using techniques like grid search or Random Search to optimize model performance.

➢ **Validation** : The performance of the trained models is validated using holdout sets, cross -validation , or other validation techniques to ensure generalizability to unseen data.

➢ **Deployment :** Once validated , the best -performing model is deployed into the existing insurance system to predict premiums for new customers based on their demographic and health information.
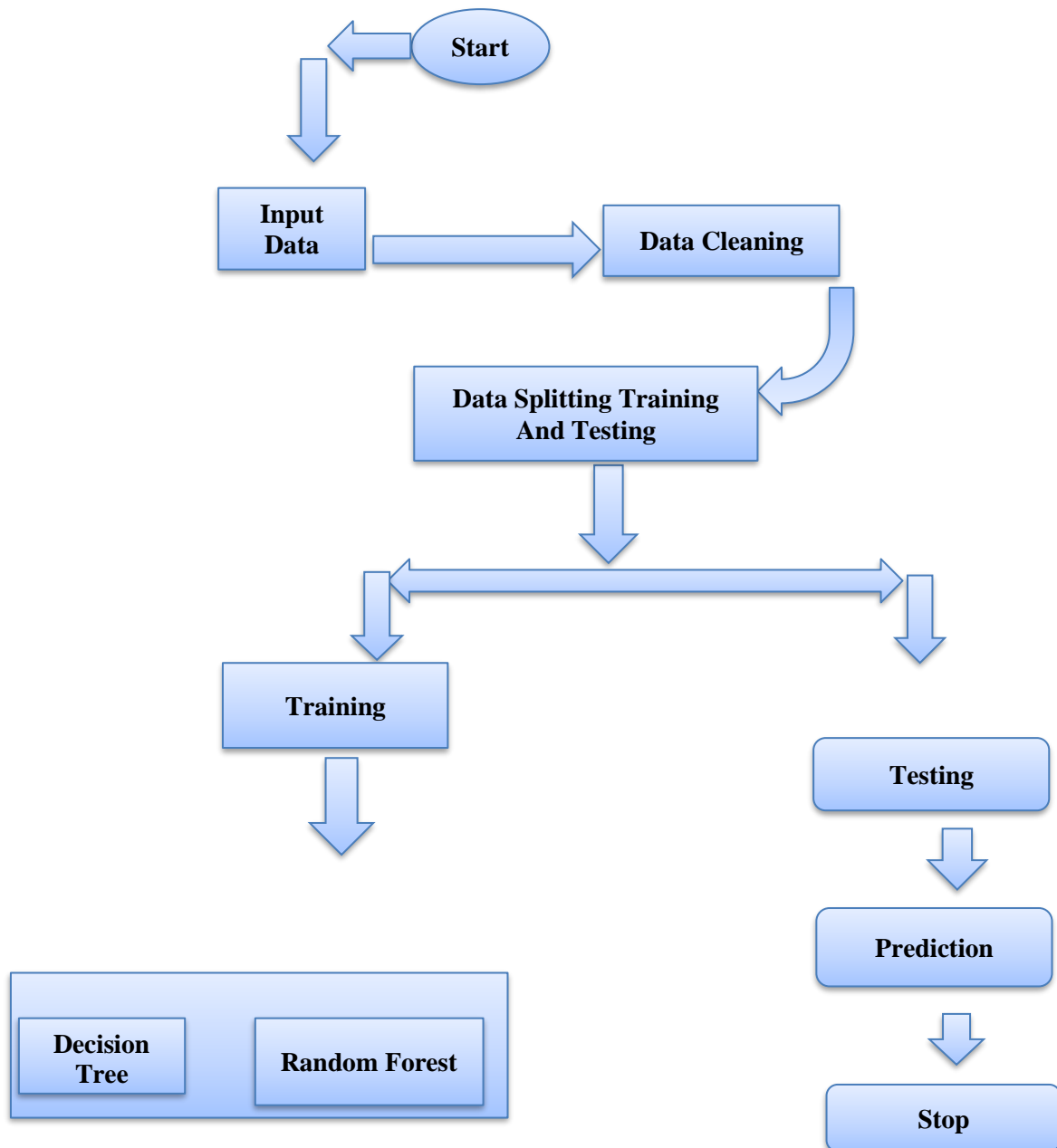
**Block Diagram :**



**Fig (a) Block Diagram he**alth Insurance

The machine learning pipeline for estimating health insurance premiums is started at the beginning of the process. A variety of sources provide pertinent information, such as policyholder demographics, medical histories, lifestyle choices, and insurance premium amounts. Preprocessing is applied to the gathered data to address missing numbers, outliers, and discrepancies. This stage makes sure the data is prepared in a way that makes it easy to analyze later. Two subsets are created from the preprocessed data: a training set that is used to train the machine learning models and a testing set that is used to assess how well they perform.About 70–80% of the data is usually used for training,

**Novateur Publications**
**International Journal of Innovations in Engineering Research and Technology [IJIERT]**
**ISSN: 2394-3696 Website: ijiert.org**
**Volume 11, Issue 5, May - 2024**

and the remaining 20–30% is used for testing. Figure trsining uses two different types of models: Random Forest and Decision Tree.

### B. Proposed Work:

➢ **Enhanced Feature Engineering** : Implementing advanced feature engineering advanced feature engineering techniques to edxtract more insightful features fro demographic , health , and insurance data.

➢ **Advanced Model Selection** : Exploring advanced machine learning Models such as ensemble methods deep learning architectures , and hybrid models for improved prediction accuracy.

➢ **Feature Selection :** Relevant features affecting insurance premiums are identified using techniques like correlation analysis , feature importance ranking , or domain knowledge**.**

➢ **Fairness and Bais Mitigation :** Integrating fairness -aware techniques to address bias and ensurance equalitable premium predictions across diverse demographics groups.

**Temporal Analysis** : Incorporating temporal analysis methods to capture dynamic patterns and trends in insurance premiums over time , enhancing predictive capabilities.

➢ **Interpretability and Transparency** : Incorporating temporal analysis methods to capture dynamic patterns and trends in insurance premiums over time , enhancing predictive capabilities.

➢ **Deployment :** Designing a scalable and robust deployment framework to seamlessly integrate the predictive model into existing insurance systems, ensuring efficient real-time premium predictions for new customers.
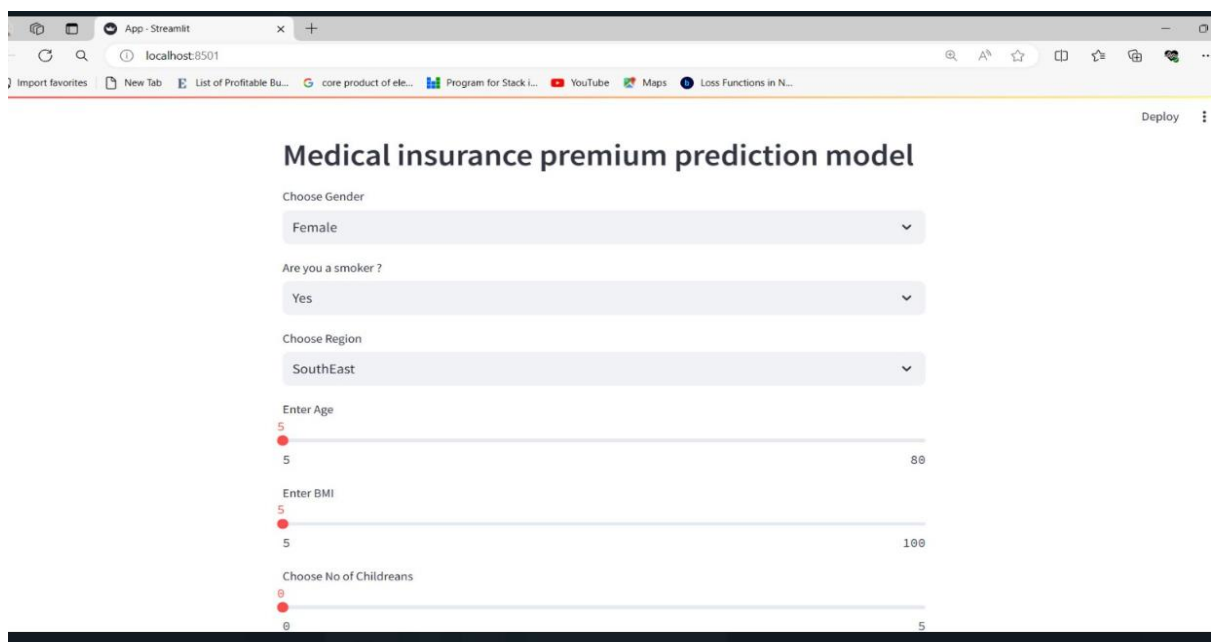
### C. Experimental Results:



Fig 1.1

**Novateur Publications**
**International Journal of Innovations in Engineering Research and Technology [IJIERT]**
**ISSN: 2394-3696 Website: ijiert.org**
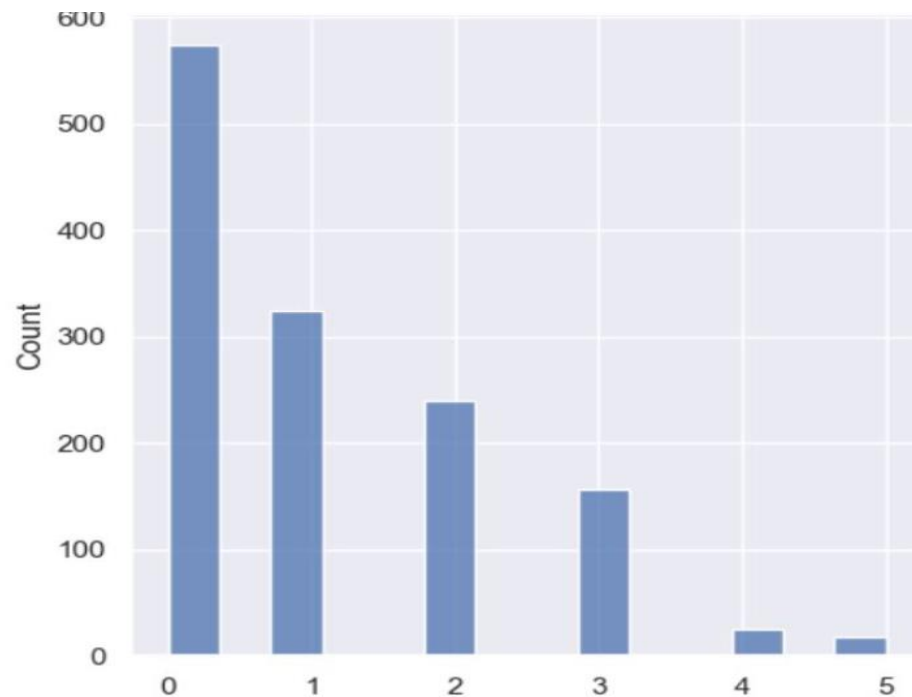**Volume 11, Issue 5, May - 2024**

Fig 1.2

### III. Conclusion :

In conclusion ,the application of machine learning to predict medical insurance premiums to predict medical insurancepremiums holds significicant promise for enhancing accuracy , fairness , and efficiency in insurance pricing.Through the analysis of historicak data and the development of predictive models , insurance can better assess risk and tailor premiums yo individual characteristics , ultimately benefiting both insurances and policyholders.

However ,while machine learning offers valuable insights and predictive capabilities , several challenges must be addressed to realize its full potential in this domain. These include ensuring fairness and transparency in premium predictions , mitigating bias , and maintaining model interpretability to foster trust among stakeholders.

Deapite these challenges , the ongoing advancements in machine learning techniques ,coupled with a deeper understanding of insurance dynamics , are paving the way for more understanding.

### REFERENCES

1. Kaushik, K., Bhardwaj, A., Dwivedi, A. D., & Singh, R. (2022). Machine learning-based regression framework to predict health insurance premiums. International Journal of Environmental Research and Public Health, 19(13), 7898.
2. Rao, V., Iswarya, M., Hamza, S. A., & Satish, B. (2023). Interpreting the Premium Prediction of Health Insurance Through Random Forest Algorithm Using Supervised Machine Learning Technology. International Journal of Innovative Science and Research Technology, 8(5), 726-731
3. Alzoubi, H. M., Sahawneh, N., AlHamad, A. Q., Malik, U., Majid, A., & Atta, A. (2022, October). Analysis Of Cost Prediction In Medical Insurance Using Modern Regression Models. In 2022

International Conference on Cyber Resilience (ICCR) (pp. 1-10). IEEE.

4. Albalawi, S., Alshahrani, L., Albalawi, N., & Alharbi, R. (2023). Prediction of healthcare insurance costs. Computers and Informatics, 3(1), 9-18.

5. Prakash, V. S., Bushra, S. N., Subramanian, N., Indumathy, D., Mary, S. A., & Thiagarajan, R. (2022). Random forest regression with hyper parameter tuning for medical insurance premium prediction. International Journal of Health Sciences, 6(S6), 7093-7101.

6. Sahai, R., Al-Ataby, A., Assi, S., Jayabalan, M., Liatsis, P., Loy, C. K., ... & Kolivand, H. (2022, December). Insurance Risk Prediction Using Machine Learning. In The International Conference on Data Science and Emerging Technologies (pp. 419-433). Singapore: Springer Nature Singapore.

7. Sahu, A., Sharma, G., Kaushik, J., Agarwal, K., & Singh, D. (2022, February). Health Insurance Cost Prediction by Using Machine Learning. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC).

8. Bhardwaj, N., & Anand, R. (2020). Health insurance amount prediction. Int. J. Eng. Res, 9, 1008-1011.

9. Goel, S., & Chaudhary, A. (2024, February). Prediction of Health Insurance Price using Machine Learning Algorithms. In 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1345-1350). IEEE.

10. Panda, S., Purkayastha, B., Das, D., Chakraborty, M., & Biswas, S. K. (2022, May). Health insurance cost prediction using regression models. In 2022 International conference on machine learning, big data, cloud and parallel computing (COM-IT-CON) (Vol. 1, pp. 168-173). IEEE.

11. Bau, Y. T., & Hanif, S. A. M. (2024). Comparative Analysis of Machine Learning Algorithms for Health Insurance Pricing. JOIV: International Journal on Informatics Visualization, 8(1), 481-491.

12. Goyal, A., Elhence, A., Chamola, V., & Sikdar, B. (2021, November). A blockchain and machine learning based framework for efficient health insurance management. In Proceedings of the 19th ACM conference on embedded networked sensor systems (pp. 511-515).

13. Ejiyi, C. J., Qin, Z., Salako, A. A., Happy, M. N., Nneji, G. U., Ukwuoma, C. C., ... & Gen, J. (2022). Comparative analysis of building insurance prediction using some machine learning algorithms.

14. Sun, J. J. (2020). Identification and Prediction of Factors Impact America Health Insurance Premium (Doctoral dissertation, Dublin, National College of Ireland).

15. Vimont, A., Leleu, H., & Durand-Zaleski, I. (2022). Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France. The European Journal of Health Economics, 23(2), 211-223.

16. Dua, P., & Bais, S. (2014). Supervised learning methods for fraud detection in healthcare insurance. Machine learning in healthcare informatics, 261-285.

17. Isa, U. A., & Fernando, A. (2022). Explainable AI and Interpretable Model for Insurance Premium Prediction.

18. Marinova, G., & Todorova, M. (2023, November). Regression Analysis for Predicting Health Insurance. In 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES) (pp. 1-4). IEEE.

19. Kose, I., Gokturk, M., & Kilic, K. (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. Applied Soft Computing, 36, 283-299.