ICEST-2K24 [International Conference on Engineering, Science and Technology]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Medical Insurance Premium Prediction with Machine Learning

Prof. Manoj S. Patil[1], Miss. Kulkarni Sanika[2], Miss. Khurape Sanjana[3]

[1]Assistant Professor, Artificial Intelligence & Data Science, Fabtech Technical Campus College of Engineering & Research, Sangola, Maharashtra, India

[2]Student, T.Y. B. Tech, Artificial Intelligence &Data Science, Fabtech Technical Campus College of Engineering & Research, Sangola, Maharashtra, India

ABSTRACT

This study presents a machine learning approach for predicting medical insurance premiums. With the increasing complexity of healthcare costs, accurate prediction of insurance premiums is crucial for both insurances and policyholders. Leveraging a dataset containing demographic information, medical history, and other relevant factors, various machine learning algorithms including regression and ensemble methods were employed to build predictive models. The performance of these models was evaluated using metrics such as mean absolute error and R-Squared. Results indicate that the developed models can effectively predict insurancepremiums,providingvaluableinsights forbothinsurancecounteragentsapproachoffersthepotential to optimizing pricing strategies, improve risk assessment, and enhanced decision-making processes in the healthcare insurance domain. Medical Insurance Premium Prediction using Machine Learning Predict health insurance cists based on individual characteristics. Collected and Preprocessed dataset of policyholders' attribute(age,gender,BMI,Number of children, smoking habits, location).Split data into training and testingsets. Build and trained artificial neural network model using Karas and TensorFlow. Evaluated model performance using mean R- squared error and R-Squared metrics. Developed an accurate predictive modelwith high R-Squared value. Identified keyfactors influencing insurance premiums. Demonstrated potential for machine learning in healthcare cost estimation. This project showcases the effectiveness of machine learningin predicting medical insurance premiums. By developing these predictive models, insurance companies can streamline their underwriting process, offers more personalized insurance plans, and assists policyholders in making informed decision about their healthcare coverage. The developed model can aid insurances in setting accurate premiums and policyholders in making informed decisions. This research contributes to the advancement of data driven approaches in the insurance industry, ultimately benefiting both insurers and ultimately benefitting both insurances and insured individuals a like.

Keywords-Medical insurance, Premium prediction, Machine Learning, predictive modeling, Underwriting, Feature importance, Data-driven decision-making.

I. INTRODUCTION

Predicting medical insurance premiums using machine learning involves using historical data on individuals' demographics, health facts, 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 accuracyin pricing policies and better manage risk. This introduction sets thestage 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: Explainthemotivationbehindusingmachinelearningtechniquesforpremiumprediction .Discussthelimitationsoftraditionalactuarialmethodsandthepotential benefitsofdata-drivenapproaches.

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 advancingthe field of healthcareanalytics 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, outingthe sections that will be covered in detail, such as data collection, Methodology, experimental results and discussion.

II. LITERATURE SURVEY

2.1 “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.

2.2 “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.

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

2.4 “Deep LearningApproaches 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.

2.5 “Feature Selection Techniques for Medical Insurance Premium Prediction” by M. A. Rahman and S.Begum:thisstudyinvestigatesvariousfeatureSelectionmethods,suchasWrapper,Filter,andEmbedded approaches , to identifythe most relevant predictors fir medical insurance premium prediction. It compares the performance of different feature selection techniques and their impact on prediction accuracy.

2.6 “FairnessinMedicalInsurancePremiumPrediction:AMachineLearningPerspective”byL.Zhangand.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

2.7 "Temporal Analysis of Medical Insurance Premiums Using Machine Learning" by J. Chen and X . Li : Thispaperinvestigatesthetemporalpatternsinmedicalinsurancepremiumsandexploreshowmachine learningmodelscancaptureandpredictfluctuationsovertime.Itdiscussestheimplicationsoftemporal 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, converting various algorithms feature selection methods , fairness considerations , and temporal analysis approaches

III.EXISTING WORK AND PROPOSED WORK

3.1 Existing Work:

3.1.1 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.

3.1.2 Data Preprocessing :Thecollecteddataundergoespreprocessingstepssuchascleaning,normalizationandfeatureengineeringtoprep areitforanalysis.

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

3.1.4 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.

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

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

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

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

3.1.9 Deployment : Once validated , the best -performing model is deployed into the existing insurancesystem to predict premiums fornewcustomers based on theirdemographicand health information.

Block Diagram:

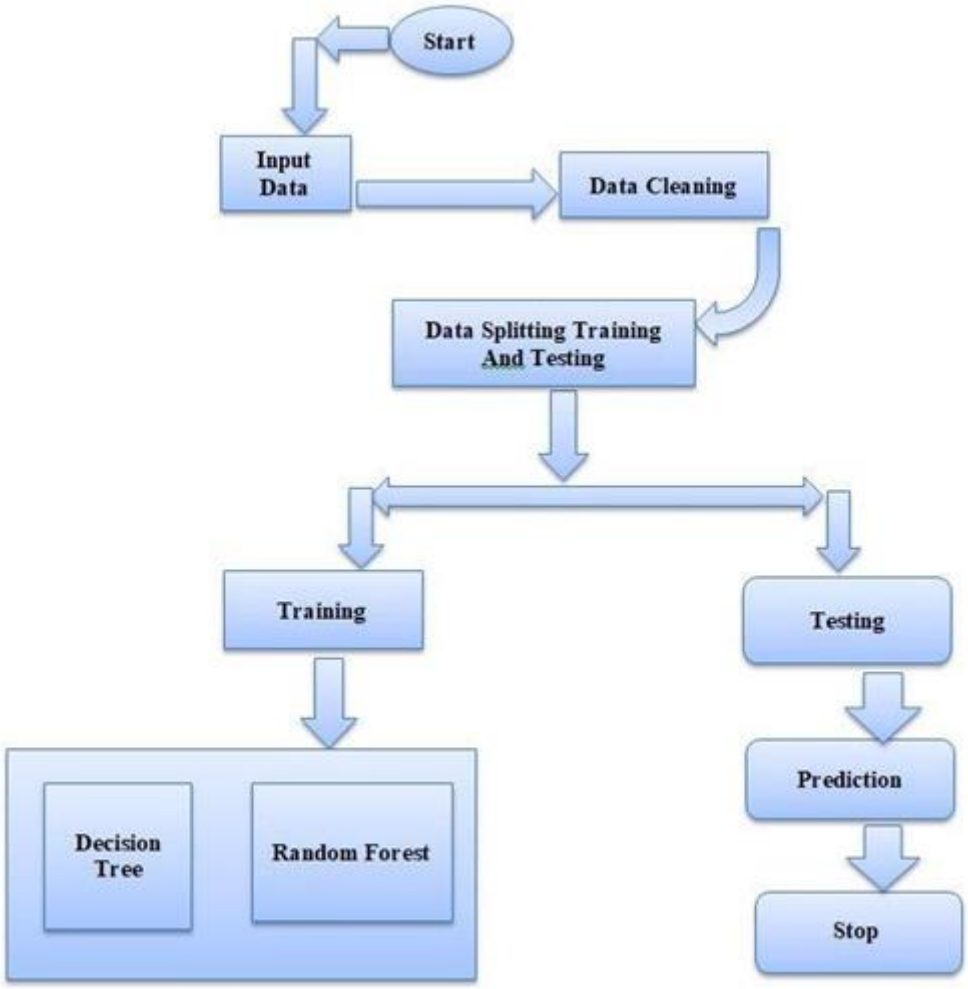

Fig.1.1BlockDiagramhealth Insurance

The Process begins with the initiation of the machine learning pipeline for predicting medical insurance premiums. Relevant data is collected from various sources , including policyholders' demographics , medical history , lifestyle factors and insurance premium amounts. The collected data undergoes preprocessingtohandlemissingvalues,outliers, andinconsistencies.Thisstepensuresthatthedataisina suitableformatfor furtheranalysis.Thepreprocesseddataisdividedintotwosubsets:atrainingsetusedto train the machine learning models and a testing set used to evaluate their performance .Typically , around 70-80% of the data is allocated to training and the remaining 20-30% to test. Two types of Models,decisiontree andRandomForest, areselectedfiguretraining.

3.2 Proposed Work:

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

3.2.2 Advanced Model Selection : Exploring advancedmachine learning Models such as ensemble methods deep learning architectures , and hybrid models for improved prediction accuracy.

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

3.2.4 Fairness and Bais Mitigation : Integrating fairness -aware techniques to address bias and insurance equali table premium predictions across diverse demographics groups.

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

3.2.6 Interpretability and Transparency :Incorporating temporal analysis methods to capturedynamic patterns and trends in insurance premiums over time , enhancing predictivecapabilities.

3.2.7 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.

3.3 Experimental Results:

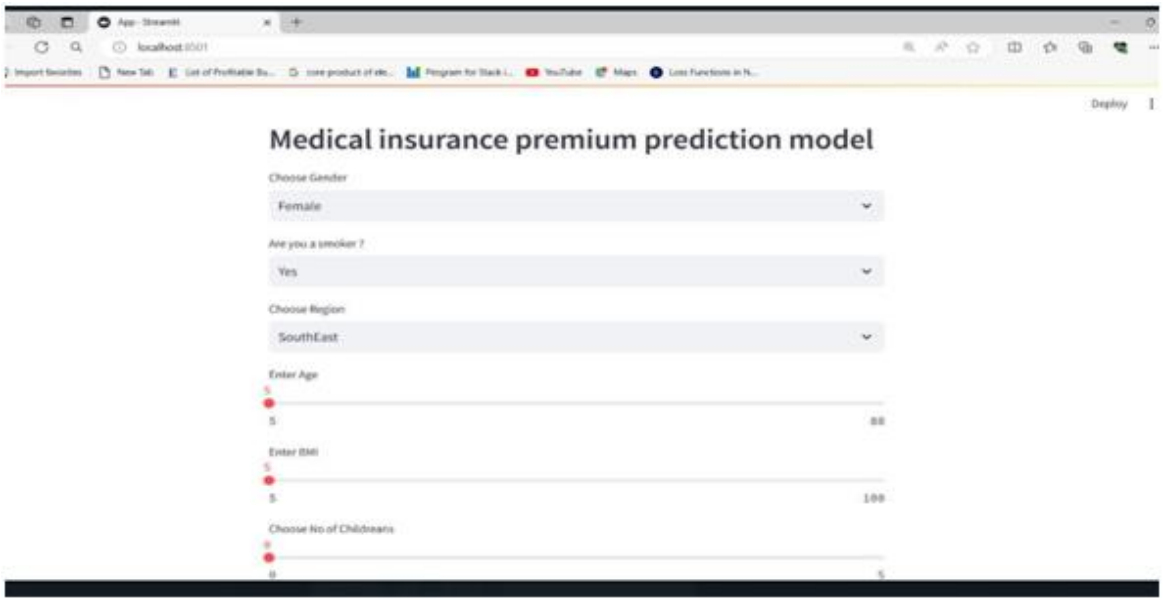

Fig1.2

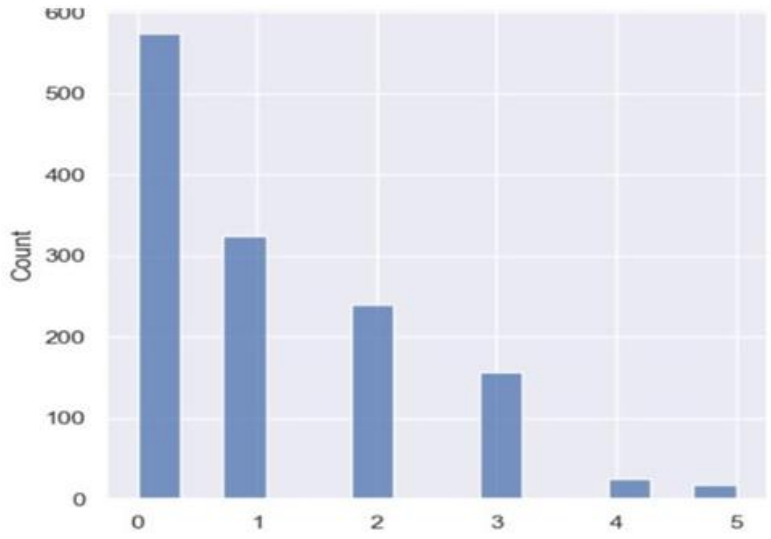

Fig1.3 Prediction

IV. CONCLUSION

In conclusion ,the application of machine learning to predict medical insurance premiums to predict medical insurance premiums holds significant promise for enhancing accuracy, fairness , and efficiencyin insurance pricing. Through the analysis of historical data and the development of predictive models , insurance can better assess risk and tailor premiums you 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.

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

V. 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 InsuranceThroughRandomForestAlgorithmUsingSupervisedMachineLearning 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 insurancecosts. Computers and Informatics, 3(1), 9-18.

[5]. Prakash, V. S., Bushra, S. N., Subramanian, N., Indumathy, D., Mary, S. A., & Thiagarajan, R. (2022). Randomforestregressionwithhyperparametertuningformedicalinsurancepremiumprediction. 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).Healthinsuranceamountprediction.Int.J.Eng.Res,9,1008-1011.

[9]. Goel, S., & Chaudhary, A. (2024, February). Prediction of Health Insurance Price using MachineLearning 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).Healthinsurancecost prediction usingregression models. In 2022 International conferenceon machinelearning, 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 machinelearning 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).IdentificationandPredictionofFactorsImpactAmericaHealthInsurance 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).Supervisedlearningmethodsforfrauddetectioninhealthcare insurance. Machine learning in healthcare informatics, 261-285.