

# **Health Insurance Price Prediction using Machine Learning**

## **IEEE BASE PAPER TITLE:**

### **Prediction of Health Insurance Price using Machine Learning Algorithms**

#### **IEEE BASE PAPER ABSTRACT:**

In the realm of health insurance pricing prediction, this research leverages advanced machine learning techniques, including linear regression and neural networks, to uncover pivotal insights. Smoking habits emerged as a major driver of elevated insurance charges, emphasizing the financial impact of this behavior on policyholders. Age also played a critical role, highlighting the progressive nature of healthcare costs as individual's age. These findings offer invaluable guidance for insurers and individuals. Additionally, by providing transparent linear regression models and the predictive power of neural networks this research equips the industry with practical tools to enhance pricing accuracy in a dynamic healthcare landscape.

#### **OUR PROPOSED PROJECT ABSTRACT:**

Health insurance is a critical component of financial planning, providing individuals and families with a safety net against unexpected medical expenses. The rising cost of healthcare has made it essential for individuals to predict potential insurance charges accurately, enabling them to choose suitable insurance plans. Given the complexity of calculating health insurance premiums, which depend on multiple factors such as age, lifestyle, and health status, machine learning offers a promising solution for automating and optimizing this process.

This project, titled "Health Insurance Price Prediction using Machine Learning," aims to develop a system that can accurately predict health insurance premiums based on personal and lifestyle attributes. The system is built using Python as the coding language, with a user-friendly web interface developed using HTML, CSS, and JavaScript. The Flask web framework integrates the front end and backend, allowing seamless interactions for users. The prediction process leverages two machine learning models: Random Forest Regressor and Stacking Regressor, both well-suited for regression tasks due to their ability to handle complex data patterns and interactions.

The dataset used for this project comprises 1,338 records and includes seven essential features that influence insurance costs: age, sex, body mass index (BMI), number of children, smoker status, region, and insurance charges. Each of these attributes plays a significant role in determining the premium, making it crucial to account for them in predictive modeling. The dataset serves as the foundation for training and testing the machine learning models.

The Random Forest Regressor model, which builds multiple decision trees and merges their outputs for better accuracy, achieved a Training Set Mean Absolute Error (MAE) of 0.609 and a Test Set MAE of 1.443. This performance demonstrates the model's ability to generalize well to new data while maintaining a low error rate. On the other hand, the Stacking Regressor, a model that combines multiple regression models to improve prediction performance, attained a Training Set MAE of 0.663 and a Test Set MAE of 1.442. The marginal difference between the two models highlights the strength of both approaches in handling this specific dataset.

In summary, this system offers a reliable and efficient way to predict health insurance charges based on various factors, thus providing value to both insurance

companies and individuals. By automating the price prediction process, it reduces the reliance on manual calculations and helps users make informed decisions when selecting health insurance plans. The combination of Random Forest Regressor and Stacking Regressor ensures that the model delivers accurate predictions, reinforcing the role of machine learning in enhancing decision-making processes within the insurance industry.

## **SYSTEM REQUIREMENTS:**

### **HARDWARE REQUIREMENTS:**

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15” LED.
- Input Devices : Keyboard, Mouse.
- Ram : 8 GB.

### **SOFTWARE REQUIREMENTS:**

- Operating System : Windows 10 / 11.
- Coding Language : Python 3.12.0.
- Web Framework : Flask.
- Frontend : HTML, CSS, JavaScript.

## **REFERENCE:**

S. Goel and A. Chaudhary, "Prediction of Health Insurance Price using Machine Learning Algorithms," 2024 11th International Conference on Computing for Sustainable Global Development, IEEE Conference, 2024.