

MEDICAL INSURANCE COST PREDICTION USING MACHINE LEARNING ALGORITHM

1.ARUNPRAKASH - 20140102
2.MOHAMMED KAIF - 20140103
3.ANBU SELVAN - 20140111

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 healthcare 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.

Existing System

- Use various data sources :

Demographic information, health history, lifestyle factors, and claims data are some common sources used in insurance cost prediction systems.

- Apply machine learning algorithm :

Algorithms such as linear regression, decision trees, random forests, and neural networks are used to estimate insurance costs.

- Provide accurate and reliable estimates :

The goal of these systems is to provide individuals and families with accurate and reliable estimates of their medical insurance costs to help them make informed decisions about their health coverage.

Proposed System

- **Data Collection** : Gather demographic information, medical history, and current health status Collect data through electronic health records, questionnaires and health assessments.
- **Risk Assessment** : Evaluate individual's heart health, diabetes, kidney function, cancer history, and nervous system function Determine the likelihood of developing certain health conditions.
- **Statistical Modeling** : Use statistical methods such as regression analysis or decision trees to estimate the cost of the insurance policy . Incorporate individual's risk profile and other relevant factors into the prediction.
- **Machine Learning Algorithms** : Use machine learning algorithms to make more accurate predictions . Algorithms can learn from historical data and adjust predictions based on new information.
- **Final Prediction** : Present the final prediction of the cost of the individual's medical insurance policy . Take into account all relevant factors, including heart health, diabetes, kidney function, cancer history, and nervous system function.
- **Review and Update** : Regularly review and update the system to ensure that predictions are accurate and relevant . Take into account changes in demographic information, medical history, and health status

Tools Used

- Hardware Requirements: Laptop , Storage devices
- Software Requirements: Google colab, Jupyter notebook, Python
- Other requirements: kaggle dataset

Phases

- Data collection
- Data importing
- Data Analyzing
- Data Cleansing
- Data preprocessing
- Machine learning algorithms
- Prediction
- Conclusion