

## Health Insurance Prediction Using Machine Learning

Mr.M.A.Ram Prasad,M.Tech ,Assistant Professor ,Dept of CSE,KHIT,Guntur

T.Pavan Kalyan, Kallam Haranadhareddy Institute of Technology,Guntur

D.Hari babU, Kallam Haranadhareddy Institute of Technology,Guntur

SD.Zoyaferdose, Kallam Haranadhareddy Institute of Technology,Guntur

S.Narasimha Rao, Kallam Haranadhareddy Institute of Technology,Guntur

### ABSTRACT

Health care costs increases day by day. As there are a greater number of new viruses entering into people, there is a need to predict health charges. This type of prediction helps governments to make a decision regarding health issues. People also know the importance of health care costs. Machine Learning is a field that has an impact on every field. The health care system also uses machine learning models for several health-related applications. In this paper, we have done a predicate analysis on medical health insurance charges. We build a model to predict the medical insurance cost of a person based on gender. We collect the data set from Kaggle, which contains 1338 rows of data with the features age, gender, smoker, BMI, children, region, insurance charges. The data contains medical information and costs billed by health insurance companies. We applied various regression algorithms to this dataset to predict medical costs. For implementation, we used the Python programming language.

### INTRODUCTION

As indicated by the World Bank, the absolute use on medicinal services as an extent of GDP in 2015 was 3.89%. Out of 3.89%, the legislative well being extent of the present well being use was 65.06% in 2015. Throughout the consumption as an extent of GDP is simply 1%, and the cash-based use as an most recent couple of decades, the progression in clinical innovation has made it conceivable to fix illnesses that were once viewed as serious. In any case, the expense of their treatment is so high, it is practically incomprehensible for a white-collar class individual to manage the cost of them. As indicated by insights, Rs 5 lakh family floater strategy will cover self, mate and one kid will cost anyplace between Rs 10,000 and Rs 17,000

on a yearly premise though Rs. 5 lakh singular well being plan will cost a multi-year-old Rs. 4,000-7,000 per year.

The goal of this project is to allows a person to get an idea about the necessary amount required according to their own health status. Later they can comply with any health insurance company and their schemes & benefits keeping in mind the predicted amount from our project. This can help a person in focusing more on the health aspect of an insurance rather than the futile part.

### **LITERATURE SURVEY**

Machine Learning is a technology where machines can learn from previous data and predict new samples. Machine Learning models are applicable in all fields. Medical files also do not have any exclusion to machine learning. Medical field using models in different situations for the last several years. Many of the researchers applied machine learning techniques to medical-related cost prediction. B. Nithya et.al applied machine learning models in Predictive Analytics in Health Care. They applied variously supervised and unsupervised models for predictive analysis. They also suggested machine learning tools and techniques are decisive in health care provinces and exclusively used in the diagnosis and predictions of various types of cancers. Anuja Tike et.al applied hierarchical decision trees for the medical price prediction systems. Their experiments showed that the price prediction system achieves high accuracy. Moran et al. utilized linear regression techniques to anticipate Intensive Care Unit (ICU) expenses and utilize and a couple of others as highlights. Gregori et.al applied various regression models for analyzing medical costs in the health care system. They mainly concentrated on reducing the bias in the cost estimates to achieve good results. Dimitris Bertsimas et.al applied different data mining techniques which provided an accurate prediction of medical costs and represent a powerful tool for the prediction of healthcare costs.

## PROPOSED

This method is used when all the attribute values are not on the same scale. The data contains medical information and costs billed by health insurance companies. It contains 1338 rows of data and the following columns: age, gender, BMI, diseases, smoker, insurance charges. In these features, insurance charge is a dependent variable and the remaining features are called independent variables. In regression analysis, we need to predict the value of the dependent variable using independent variables. First, we collected the data set and applied various data preprocessing methods. Data preprocessing is a technique in which we can remove missing values in the data. Because of these missing values, it is not possible to apply machine learning algorithms. After removal of missing values, we need to apply label encoding, one hot encoding data to the categorical features. Categorical features are the features whose values are labels instead of values. After that, apply standardization or normalization techniques to our data

## RESULTS

```
Please enter your Age: 40
Please enter your gender(m/f): m
Please enter your Body Mass Index: 300
Please enter the disease:
0-Ischemic heart disease, or coronary artery disease
1-Brain Stroke
2-Lower respiratory infections
3-Chronic obstructive pulmonary disease
4-Trachea, bronchus, and lung cancers
5-Diabetes mellitus
6-Alzheimer's disease and other dementias
7-Dehydration due to diarrheal diseases
8-Tuberculosis
9-Cirrhosis
Select the disease appropriately: 2
Do you smoke?
Yes- 1
No -0
Please enter your answer: 0
Please select the region where you wanted to get medicated:
0-Hyderabad
1-Banglore
2-Chennai
3-Delhi
Please enter an appropriate option: 0
The average cost of medication is 1042343.381629851
The estimated insurance that can be claimed is 7928845.202207747
```

S.NO	INPUT	OUTPUT	RESULT
Test Case 1	The user gives the input in the form of attributes when Gender=Male and Smoking is 0	The average cost and the estimated insurance is executed successfully	The result is that the dataset is trained. Therefore the test case1 is passed successfully.
Test Case 2	The user gives the input in the form of attributes when Gender =Male and Smoking is 1 .	The average cost and the estimated insurance is executed successfully	The result is that the dataset is trained. Therefore the testcase2 passed successfully.
Test Case 3	The user gives the input in the form of attributes when Gender=Female and Smoking=0	The average cost and the estimated insurance is executed successfully	The result is that the dataset is trained. Therefore the test case3 passed successfully.
Test Case 4	The user gives the input in the form of attributes when Gender=Female and Smoking=1	The average cost and the estimated insurance is executed successfully	The result is that the dataset is trained. Therefore the test case4 passed successfully

### CONCLUSIONS

we have explored the basics of the linear regression model and applied it to predict charges and seen the correlation between predicted and actual results. I hope you find this post useful and get some basic idea of a linear regression model. we proposed a machine learning model for predicting medical costs. We applied regression techniques Linear Regression and observed that age, BMI are features that decide the dependent variable. Out of all experiments, this model gave a better result.

### REFERENCES

- 1) B. Nithya, Dr. V. Ilango, "Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems ICICCS 2017, 978-1-5386-2745-7/17/\$31.00 ©2017 IEEE.
- 2) A. Tike and S. Tavarageri. (2017). A Medical Price Prediction System using Hierarchical Decision Trees. In: IEEE Big Data Conference 2017. IEEE, 978-1-5386- 2715-0/17/\$31.00 ©2017 IEEE.

3) Lahiri and N. Agarwal, "Predicting healthcare expenditure increase for an individual from medicare data," in Proceedings of the ACM SIGKDD Workshop on Health Informatics, 2014.

4) Gregori, M. Petrinco, S. Bo, A. Desideri, F. Merletti, and E. Pagano, "Regression models for analyzing costs and their determinants in health care: an introductory review," *International Journal for Quality in Health Care*, vol. 23, no. 3, pp. 331–341, 2011.

5) Bertsimas, M. V. Bjarnadóttir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala, and G. Wang, "Algorithmic prediction of health-care costs," *Operations Research*, vol. 56, no. 6, pp. 1382–1392, 2008.