



Daffodil
International
University

Project Report

Course Code: CSE-316.

Course Title: Artificial Intelligence Lab.

Submitted By

Group: Mind_Optimizer

Name:

ID:

Hafizur Rahman Shadhin

191-15-12666

Mst. Shuchana Kabir

191-15-12769

Md. Omur Faruk Redoy

191-15-13023

Fazlerabbi Antor

191-15-12688

Shukdeb Bhowmik Shuvo

191-15-12832

Section: 'F'

Department of CSE

Daffodil International University

Submitted To

Faisal Arafat

Lecturer

Department of CSE

Daffodil International University

Index

1. Project Name
2. Abstract
3. Introduction
4. Model
5. Conclusion

Project Name:

Medical Cost Prediction

Abstract:

Rising medical costs are one of the world's most important problems. This is a major economic and public health issue worldwide. The health care costs constitute a significant fraction of the U.S. economy. Nearly 20% of the Gross Domestic Product (GDP) is spent on health care. The health spending in the US is the highest among all developed nations in absolute numbers as well as a percentage of the economy. The U.S. government bears a large portion of seniors' health expenditure through its Medicare program. The growing health related expenses combined with the fact that the babyboomer generation is retiring, and hence they will be eligible for Medicare, puts a great burden on the U.S. exchequer. Therefore, it is essential to contain health related payments through all means possible. In this work, we will develop a medical price prediction system using machine learning algorithms which will aid in steering patients to cost effective providers and thereby curb health spending. The policymakers can also use the tool to better understand which providers are relatively expensive and take punitive actions if necessary. The prediction of the medical price will be done using implementing Random Forest Regression algorithm in machine learning. Additionally, we plan to include the experiments on the same data with other machine learning models such as Gradient Boosted Trees and Linear Regression and compare results. The findings from these experiments will also be included.

Introduction:

This project report is about Medical cost Prediction. We have done our project by using data analysis, data preprocessing. A health insurance company can only make money if it collects more than it spends on the medical care of its beneficiaries. On the other hand, even though some conditions are more prevalent for certain segments of the population, medical costs are difficult to predict since most money comes from rare conditions of the patients. The objective of this article is to accurately predict insurance costs based on people's data, including age, Body Mass Index, smoking or not, etc. Additionally, we will also determine what the most important variable influencing insurance costs is. These estimates could be used to create actuarial tables that set the price of yearly premiums higher or lower according to the expected treatment costs. This is a regression problem.

Model:

Our project is an automatic system that can predict what the medical insurance cost of a person is will be using some methods. The first step is to collect the data. Here we need insurance medical cost data. We were working a dataset which contains, age, sex, bmi, children, smoker, region, charges. We need to do some data analysis. Then we need to do data preprocessing. Then we have to check our model. In this way, we can predict medical cost.

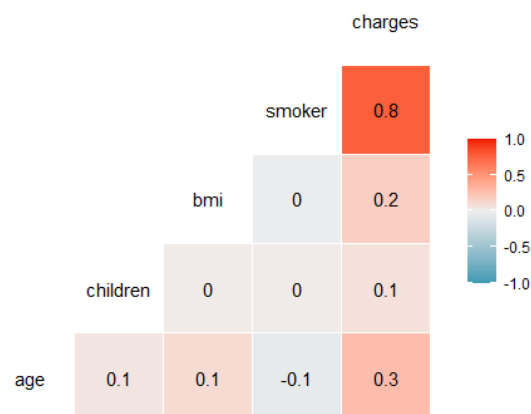
We will build and train Linear Regression model for this problem. For starters, let's use all available features in the model.

Linear Regression:

Linear Regression will be implemented with automatic feature selection using backward elimination. Starting from using all features, the *backward elimination* process will iteratively discard some and evaluate the model until it finds one with the lowest Akaike Information Criterion (AIC). Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models based on information loss. Lower AIC means better model. We'll use `step()` function to apply *backward elimination* in a greedy manner.

Linearity of the Data:

We need to make sure that there is a linear relationship between predictors and target variable. This can be done by visually looking at the correlation between each pair of predictor and target variable.

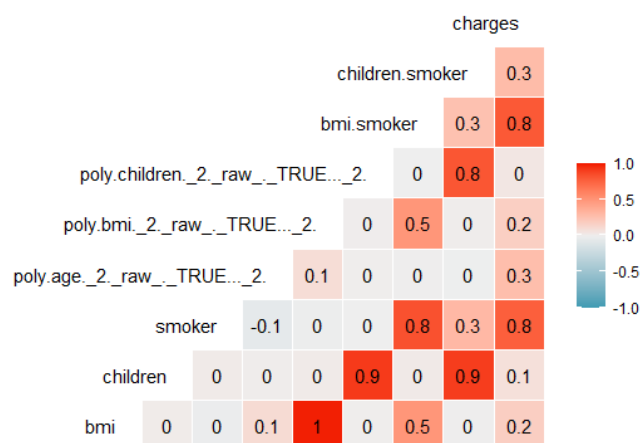


Another way is to use hypothesis testing with Pearson's product-moment correlation.

- H0: the predictor does not correlate with charge
- H1: the predictor correlates with charge

Since the p-value for each predictor-target pair is below alpha (0.05), reject H0. We can safely say that the predictors correlate with target variable.

Now, for the Polynomial Regression.



We can say that the predictors correlate with target variable, except for children.

Motivation: The treatment is basic need of human life. Predictive analytics can be described as a branch of advanced analytics that is utilized in the making of predictions about unknown future events or activities that lead to decisions. From our project, we can predict medical cost. Dataset consist age, sex, bmi, children, smoker, region, charges. We can predict from the dataset which type of patients need more cost or less cost. According to data analysis, we can see that, people who are smoker have higher medical cost. Those who have lower and higher bmi, their medical cost is high. Medical cost is higher for girls who have had more children at younger age. On the other hand, Medical costs are much lower for those who have accurate bmi and do not smoke. Data analytics can help providers react as quickly as possible to changes in a patient's vitals, and may be able to identify an upcoming deterioration before symptoms clearly manifest themselves to the naked eye.

Conclusion: Medical is very important part of our life. In recent years, with overall improvements in living standards, health care is becoming increasingly important worldwide. The total cost of medical is also gradually increasing. Correspondingly, predicting such costs with accuracy is a significant first step in addressing this problem. Thus, accurately predicting future costs and understanding which factors contribute to increases in health care expenditures are important. The objective of this project was to predict patient's medical costs development in the subsequent year and to identify factors contributing to this prediction. We have used data analysis and data preprocessing. we have explored the basics of the linear regression model and applied it to predict charges. we predict medical-cost of individual people by using Medical Insurance Data.