MEDICAL EXPENSE PREDICTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

E. PUNEETH KUMAR (18113028)

P. PAVAN KRISHNA (18113054)

M. RAJA VARDHAN (18113057)

Under the guidance of

DR. P. SELVI RAJENDRAN

Professor

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE

CHENNAI - 603 103

MAY 2022

HINDUSTAN
INSTITUTE OF TECHNOLOGY & SCIENCE
(DEEMED TO BE UNIVERSITY)

BONAFIDE CERTIFICATE

Certified that this project report **Medical Expense Prediction** is the bonafide work of E. Puneeth Kumar (18113028), P. Pavan krishna (18113054), M. Raja Vardhan (18113057) who carried out the project work under my supervision during the academic year 2**018-2022**.

Dr. J. THANGAKUMAR,

HoD,

Department of CSE.

Dr. P. SELVIRAJENDRAN,

SUPERVISOR

Professor

INTERNAL EXAMINER

Name:_____

Designation: _____

EXTERNAL EXAMINER

Name: _____

Designation:_____

Project Viva - voce conducted on _____

TABLE OF CONTENTS

APPENDIX A: SAMPLE SCREEN

APPENDIX B: SAMPLE CODE

APPENDIX C: PLAGIARISM REPORT

APPENDIX D: FUNDING / PATENT / PUBLICATION DETAILS

APPENDIX E: TEAM DETAILS

ACKNOWLEDGEMENT

First and foremost we would like to thank **ALMIGHTY** who has provided us the strength to do justice to our work and contribute our best to it

We wish to express our deep sense of gratitude from the bottom of our heart to our guide **DR.P.SELVI RAJENRAN Professor,Computer Science and Engineering,** for her motivating discussion, overwhelming suggestions, ingenious encouragement, invaluable supervision and exemplary guidance throughout this project work

We would like to extend our heartfelt gratitude to **Dr. J. THANGAKUMAR, PhD. ,Professor & HEAD, Department of Computer Science and Engineering** for his valuable suggestions and support in successfully completion of project

We thank the management of **HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE** for providing us the necessary facilities and support required for the successful completion of the project

As a final word, we would like to thank each and every individual who have been a source of support and encouragement and helped us to achieve our goal and complete our project work successfully

DEDICATION (optional)

This project is dedicated to my beloved parents, for their love,

endless support, encouragement and sacrifices.

ABSTRACT

Medical costs are one of the most common reoccurring expenses in a person's life. It is general known that a person's lifestyle and numerous physical factors determine the diseases or disorders they may get, and that these conditions determine medical expenses. According to several research, there are several significant reasons that lead to greater expenditures. smoking, age, and BMI are all factors in personal medical care. The goal of this study is to examine and identify a link between personal medical costs and other characteristics. Then, by generating linear regression models and comparing them using ANOVA, we use the significant traits as predictors to forecast medical expenditures. In our research, we discovered that smoking, age, and a higher BMI all have a significant connection with higher medical expenditures, showing that they are key contributors to the charges, and that the regression can predict the charges with more than 75% accuracy. According to the World Health Organization, personal medical and healthcare spending is growing faster than the global economy This rise in spending has been related to a variety of factors, the most prominent of which are smoking, ageing, and higher BMI. Using insurance data from diverse persons with variables such as smoking, age, number of children, area, and BMI, we hope to uncover a link between medical expenditures and other parameters.

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Expansion
LR	Linear regression
ML	Machine learning
BMI	Body mass index
DFD	Data flow diagram

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

The expense of health care is rising every day. There is a need to forecast health costs as the number of novel viruses infecting humans grows. This form of forecasting aids governments in making health-related decisions. People are also aware of the significance of health-care spending. Machine Learning is a field that touches all aspect of life. Machine learning models are also used in the health-care system for a variety of health-related applications. We conducted a predicate analysis on medical health insurance expenses in this study. We create a model to forecast a person's medical insurance costs depending on gender. The dataset comes from Kaggle and comprises 1338 rows of data with the following attributes: age, gender, smoker, BMI, children, region, and insurance charges. Medical information and expenditures billed by health insurance companies are included in the data. To forecast medical expenses, we used a variety of regression techniques on this dataset. The Python programming language was utilized to implement the project.

In 2015, the absolute consumption of medical services as a proportion of GDP was 3.89 percent, according to the World Bank. Legal medical usage amounts for barely 1% of GDP, down from 3.89 percent in 2015, and cash-based medical use makes for 65.06 percent of current medical use. Over the previous few decades, advances in clinical innovation have made it feasible to cure ailments. It deals with a subject that was formerly supposed to be fatal. In any event, the expense of her therapy is so high that it is unaffordable for someone in the white-collar class. A 5,000 rupee floater plan for yourself, your spouse, and your children is anticipated to cost 10,000 to 17,000 rupees per year, while a 5,000 rupee sickness plan will cost 4,000 rupees for multiple people.. In any case, the expense of their therapy is so high that it is practically impossible for a white-collar worker to pay it. A 5,000 rupee floater plan for a family

will cost between 10,000 and 17,000 rupees per year, whereas a 5,000 rupee health insurance plan for multiple years will cost 4,000 rupees. It ranges in price from a few rupees to 7,000 rupees. It'll take a year. According to the calculations.

It then generates linear regression models and compares them using ANOVA to employ an important function as a predictor for forecasting medical expenditures. Our research discovered that smoking, being older, and having a higher BMI were all linked to higher medical expenditures. This demonstrates that these are the primary sources of cost, and that regression can accurately estimate expenses with a 75% accuracy rate.

1.2 Motivation for the Project

To project will predict the outmost medical expenses that will help the candidate or patient to claim the medical insurance and can know the estimated medication cost. Finding the estimated medicine expenses for a treatment is difficult, so that we want to them by giving estimation for medication. That will know the approximately. In this project our objective in this research is to forecast medical prices using the data we have. We will examine the work of many writers in the field of price prediction in the first few chapters of this study, as well as offer detailed information on some of the strategies utilized in the health care pricing. Later, we'll suggest a new system's design based on Medicare payments statistics.

1.3 Problem Definition and Scenarios

In the current world, the medical expenses plays the vital role in life. sometimes we need to pay huge amount for treatments. We are going to build a system, which predicts the medical insurance amount from the company based on certain parameters and which also provides cost of treatment for specific disease.

With help of machine learning we are going to build a system, which takes inputs from the user and predicts the medical treatment cost of the disease and insurance claim amount. for this we are using linear regression.

Linear regression: Predict the response variable's result using some explanatory factors.

1.4 Organization of the report

The chapter 1is details about Overview of the motivation for the project and problem statement purpose of the project, this tell us how the project is described

The chapter 2 is about the literature review on the various paper that are studied and understanding of the project in more detailed way and analysis of the system that are described

The chapter 3 is mainly discussed on the Project goals, current systems, existing system flaws, prospective systems, and suggested system benefits.

The chapter 4 is discussed on the project description in which the architecture diagram, sequence diagram use case diagram, activity diagram.

The chapter 5 is all about the requirements which is hardware and software requirements specification and technologies used in the project.

The chapter 6 is all about the module description used in which the testing stage is classified into stages such as unit test, integrated testing and coding of the parts that are implement in the system and the result part of the system and also the bread bord testing which helps for correct working of the model.

The chapter 7 is all about user interface and system that will explain detailed about the project.

The chapter 8 is all about which we got results obtained and testing.

The chapter 9 is all about the conclusion of the project in which how the project is concluded and what we have done in this project and also about the future enhancement of the project what will be the extension of this project for later development This the organization of thesis what we have discussed in the following chapter this brief understanding on each and every chapter

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Machine learning is a technique that allows computer to learn from past data and anticipate fresh samples. Machine Learning models may be used in any sector. Medical records are likewise not exempt from machine learning. For numerous years, the medical industry has used models in various settings. Many of the studies used machine learning approaches to forecast medical costs B. Nithya [1] et.al In Predictive Analytics in Health Care, machine learning models were used. For predictive analysis, they used a variety of supervised and unsupervised models. They also claimed that machine learning tools and techniques are crucial in health-care sectors, and that they are exclusively employed in the detection and prognosis of various malignancies. Ahuja Tike[2] 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. [3] utilized linear regression techniques to anticipate Intensive Care Unit (ICU) expenses and utilize understanding socioeconomics, DRG (Diagnostic Related Group), length of stay in the clinic, and a couple of others as highlights. Gregory [4] 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[5] 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.

2.2 Medical Expense Prediction System using Machine learning Techniques and Intelligent Fuzzy Approach (2020, H. Chen Jonathan, M. Asch Steven)

Prediction isn't a new concept in medicine. Clinical predictions based on data are becoming commonplace in medicine., ranging Risk categorization of patients in the critical care unit ranges from risk scores to anticoagulant treatment (CHADS2) and cholesterol medicine usage (ASCVD) (APACHE). You may easily develop prediction models for hundreds of similar clinical questions using clinical data sources and contemporary machine learning. These approaches might be used for everything from sepsis early warning systems to superhuman diagnostic imaging.

The real data source, on the other hand, has an issue. Unlike traditional techniques, which rely on data from cohorts that have been thoroughly prepared to prevent bias, new data sources are sometimes unstructured due to the fact that they were developed for various purposes (clinical care, billing, etc.). Patient self-selection, indication misunderstanding, and inconsistent outcome data can all contribute to unintentional biases and even racist programming in machine prediction. As a result of this understanding, discussing the potential of data analysis to aid medical decision-making isn't just wishful thinking.

2.3 Predicting Days in hospital using health Insurance Claims (2020, Yang Xie, C.W. Chang, Sandra Neubauer)

Identifying and managing patients most at risk within the health care system is vital for governments, hospitals, and health insurers but they use different metrics for identifying the patients they perceive to be at most risk[1]. Hospitals focus on re-admission rate [2], [3] and cumulative risk of death during hospitalization [4]. Accurately predicting these indicators could assist in allocating limited resources and thus improve the hospital’s operational efficiency. Health insurers are mostly concerned with insurance risk, because they agree to reimburse health-related services in exchange for a fixed monthly premium. Poor risk measure could result in exceeding a financial budget. Therefore, one of the most obvious goals for health insurers is to

Various predictive models have been developed to identify high-risk customers by predicting health-care expenses [5]. Traditional prediction models used demographic information and prior costs to predict future costs [1]. More sophisticated models that incorporate diagnoses [6], [7], drug claims [6], [7] and self-reported health status data [8], have been shown to improve prediction performance. Zhao et al. [6] reported a coefficient of determination (R2) of 0.168 when both drug and diagnosis were used to estimate costs for the coming year. Bertsimas et al. proposed two models: a decision tree model and a clustering model [7], to improve the performance of an earlier model based on classical regression models [6], [9]. Since hospitalization is usually the largest component of health expenditure [10], a separate identification of subpopulations at higher hospitalization risk could improve current underwriting processes and pricing methodology. Moreover, insurance companies also manage insurance risk by using specific interventions in different sub-populations (especially high-risk groups) to minimize the resources they require [11]. Programs that use case and disease management have been developed, which target different sub-groups of customers, such as aged care programs, chronic disease programs, and more recently telehealth programs, all of which have been shown to improve health care outcomes. identify high risk customers by predicting their health care expenditures.

2.4 Summary

This paper says about the various projects that are related to the Medical price expense prediction using machine learning. Each one paper has discovered new thing to the existing system and they implemented various types of relations in them on this surveys and the other thing we made new design to the existed system this is all about the literature survey

CHAPTER 3
PROJECT DESCRIPTION

3.1 Objective of the Project work

This device will help people to save time. As there will be no wastage of time, the user will be satisfied. The WATERFALL MODEL, which asserts that the stages are structured in a linear manner, Is essentially being foll0wed First and foremost, Aa feasibility assessment has been c0mpleted. After that the requirements analisis and project planning may c0mmence. If an existing system requires changes or the installation of a new module, an analysis of the existing system can be utilized as a starting point. After the requirements analysis is completes . the design phase begins . followed by the coding phase. Tesing is a accomplished after the progrsmming is completed. Requirement anlaisis, project planning, system interated and testing are the activities conducted in this approach in a software development project. The linear sequence of these action is crucial in this case. The phase ends, and the output of one phase becomes the input of the next.

3.2 Existing System

- In sample sizes ranging from small to large, statistical approaches (E.g., Lenear regresion) suffer due to the zsero point spike and skewed distibutionof health care expenses with a strong right-hand tail.
- To address this issue, advanced methods have been proposed, such as The data source, on the other hand, has a flaw. Unlike traditional techniques, which rely on data from cohorts that have been thoroughly prepared to eliminate bias, new data sources are sometimes unstructured since they were developed for various purposes (clinical care, billing, etc.). A range of issues, ranging from patient self-selection to indication uncertainty and inconsistent outcome data, can induce unintentional biases and even racist programming in computer prediction. As a result of this understanding, speculating on the potential of data analysis to aid medical decision-making isn't far-fetched.

3.3 Proposed System

- The Medical information and expenditures billed by health insurance companies are included in the data. It has 1338 rows of information with thw following columns: age, gender, BMI, illnesses, smokers, and insurance costs.
- 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 dataset 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.

CHAPTER 4

SYSTEM DESIGN

4.1 Architecture Diagram

The diagram represents the implementation of the project that has done step by step which has data preprocessing, cleaning, training and testing of the data as well. Our projects procedure was as follows: first, we had to open the files with the Jupiter notebook.

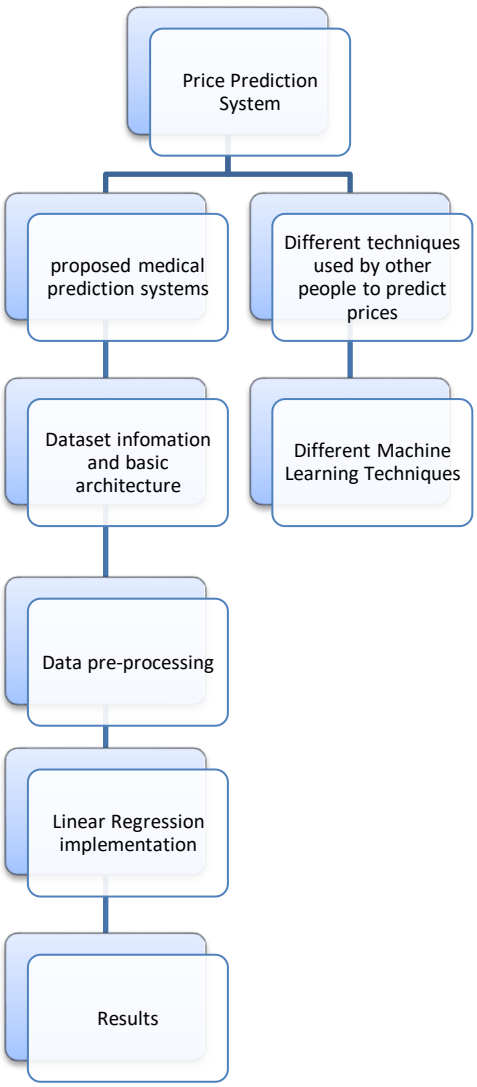


Fig 4.1 Architecture Diagram

4.2 Sequence Diagram

A sequence diagram depicts item interactions in chronological order. It illustrates the scenario's objects and classes, as well as the sequence of messages sent between them in order to carry out the scenario's functionality. In the Logical View of the system under development, sequence diagrams are often related with use case realizations.

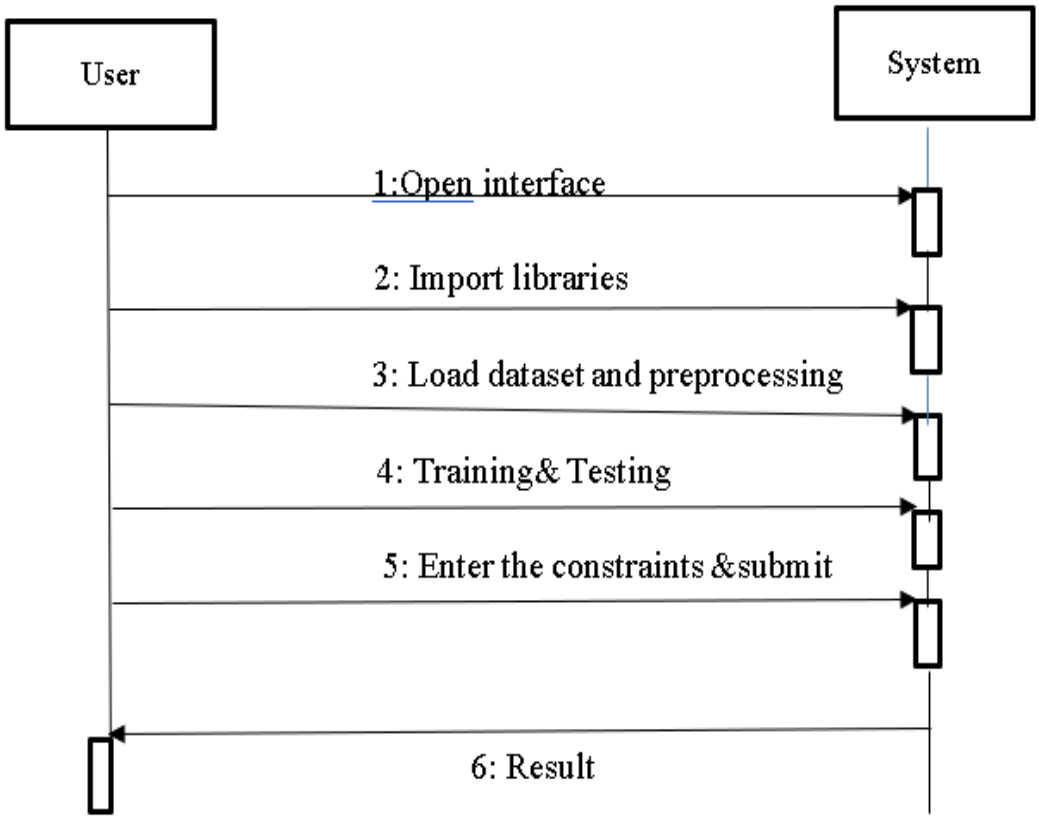


Fig 4.2 Sequence Diagram

4.3 Use case Diagram

Use case diagrams reflect the user's engagement with the system at the most basic level by depicting the relationships between the user and the many use cases in which the user participates. Use case diagrams may be used to depict a variety of system users and use cases, and are frequently accompanied by other diagrams. External entities that engage with the system are referred to as actors. Circles or ellipses are used to depict the use cases.

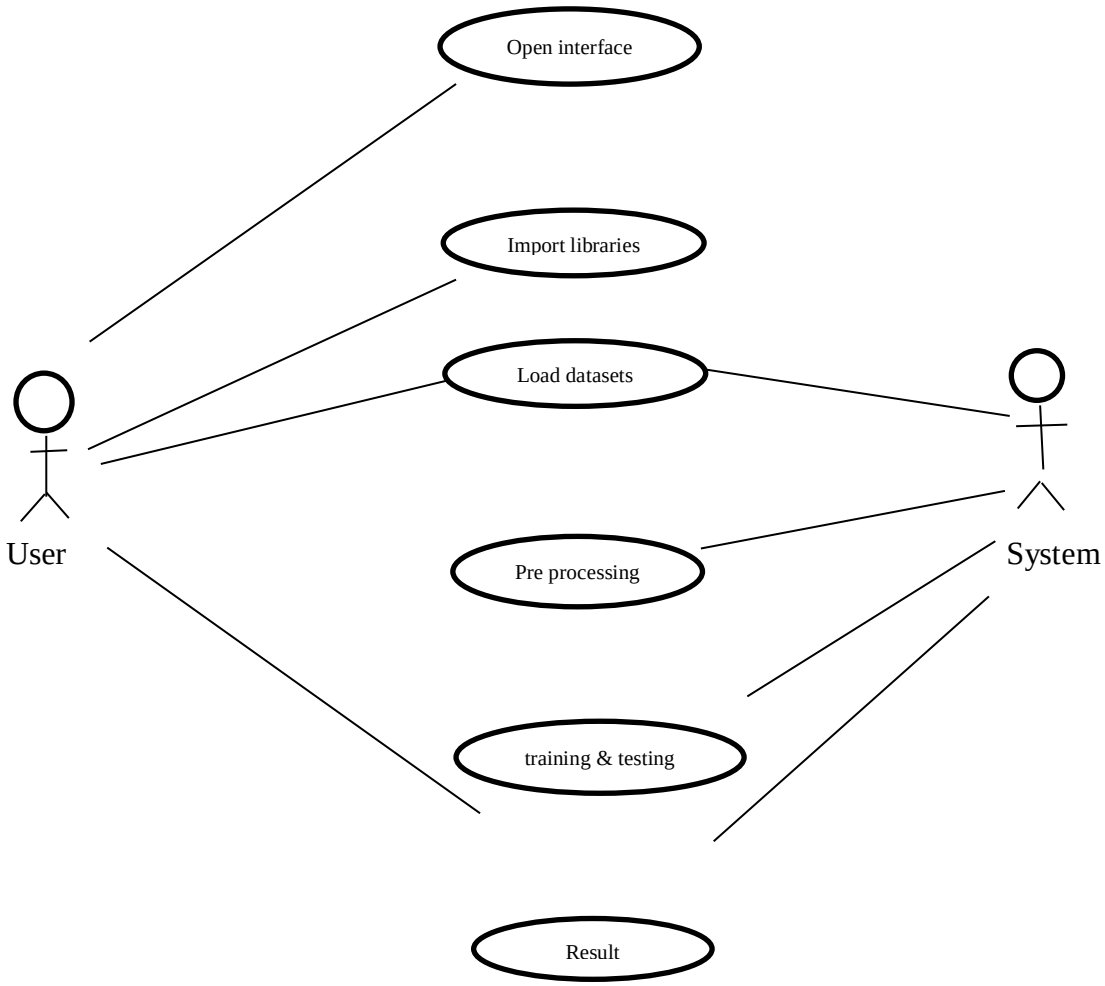


Fig 4.3 Use case Diagram

CHAPTER 5
PROJECT REQUIREMENTS

5.1 Hardware and Software Specification

5.1.2 Hardware Requirements:

- System: Pentium IV 2.4 GHz
- Hard Disk:40GB
- Floppy Drive: 1.44Mb
- Monitor: 15VGA Color
- Mouse: Logitech
- Ram:512Mb

5.1.3 Software Requirements:

- Operating system: Windows XP/7
- Coding Language: python
- IDE: Anaconda Navigator

5.2 Technologies Used

Software requirements specifications (SRS), often known as software specifications, are a precise description of the behaviour of the system in development. It provides a collection of scenarios that define all of the software's interactions with the user. SRS also covers non-functional requirements in addition to use cases. Non-functional requirements are constraints on a system's design or execution. B. System requirements specs & # 41; Performance requirements, quality standards, or design limitations This is a set of data that contains all of the system's needs. Business analysts, sometimes known as systems analysts, are in charge of researching customer and stakeholder business demands in order to uncover and provide solutions to business problems. There are three components that must be included in the project. Business requirements define what must be supplied in order to generate value.

CHAPTER 6
MODULE DESCRIPTION

6.1 Modules

There are 5 modules in this project that have detailed explanation below:

1) Dataset Information
2) Selection of features from the dataset
3) Data preprocessing

6.2 Module 1

Data set Information:

Our suggested system's input dataset will be a dataset. That combines two different datasets. A series of inpatient Medicare payment data and a column of Zillow data will be displayed. be included in our final input dataset. In the following part, we'll look at these columns in further depth. The first dataset, Hospital-level payments to about 30,000 hospitals in the 100 most often billed Diagnosis Related Groups are included in the Medicare payment data collection (DRGS). The top one-hundred DRGS account for 60% of total inpatient Medicare payments. expenditures and accounts for 7 million discharges. Each row in the payment dataset comprises 10 columns, as seen in the preceding section. Each of these columns denotes a characteristic of machine learning. A feature is a significant attribute that influences the prediction variable under consideration. Every problem under investigation has a collection of independent characteristics that aid in the construction of an accurate machine learning

6.3 Module 2

Feature Selection from the dataset

Selecting the key features from a large range of features that are more relevant and building a strong model is a critical effort. As a result, we will choose just those variables from our Medicare dataset that will independently assist us in predicting

medical prices. The location of the provider is represented by columns such as provider address, ZIP code, state, city, and hospital area referral description. As a result, rather than examining all of them, we will just include one of them in our feature set. We'll go with 'hospital region referral description' because it's not as particular as a provider's location or city, but it's also not as wide as a state. DRG Definition and Total Discharges are two further independent characteristics we'll pick from medical payment data. In addition to these features, the Medicare payment dataset now includes a new feature: real estate values. These are the prices of real estate in the hospital's immediate vicinity for the same year as the medical data. This element, in our opinion, can potentially be a prominent component in price prediction. The notion is that a hospital's operating costs are factored into the price it charges. Among the several expenditures that a hospital must face, one of the most significant is the cost of real estate - the cost of owning or renting a facility. Real estate expenses are also a proxy for other costs, in the sense that if a location has high real estate costs, it is likely to have higher costs elsewhere 19 Salary given to physicians, personnel, and other categories are also included.

6.4 Module 3

Data pre-processing:

The data that will be utilized to answer the problem is one of the most significant aspects of machine learning difficulties. Data preparation accounts for around sixty to seventy percent of the overall time spent on a typical machine learning project. In order to get successful outcomes, it is critical to have the proper data for the situation at hand. In general, data preparation consists of selecting characteristics and pre-processing those features. As a result, after selecting features from a vast quantity of data, the following step is to pre-process those features. Because the data is useless in its raw form. The goal of pre-processing in this case is to make features appropriate for the machine learning model we'll use. If the characteristics are set up correctly, the model can produce better results. In addition, the data formats for various models varies. There are machine learning models

The data pre-processing task includes following steps:

1. Data Integration
2. Data Cleaning
3. Data Transformation

We will describe each of these steps in details.

Data Integration

Data integration is identifying the many data sources that will be needed for processing and combining their information into one. This stage is critical for any system that requires a large amount of data processing to tackle the challenge at hand. In the sector, there are several tools for integrating and combining data from various sources. Such technologies can be utilized in situations when the amount of data is large and there are several data sources.

Data Cleaning:

Many contaminants can be found in raw data. These contaminants can have an impact on the ultimate result, especially when it comes to machine learning difficulties. As a result, after the data has been incorporated, it must be cleansed. Impurities such as incorrect entries, irrelevant data, and inconsistent data are detected during data cleaning. Eliminating these contaminants from records Data cleansing may be accomplished using a variety of methods. Using automated tools, personal interaction, and building scripts to programmatically clean the data according to our needs are some of the data cleaning strategies.

Transformation Data:

After cleaning data and integration data, the following is step to convert the Clean data that has been incorporated into the system's format. In most cases, data conversion entails transforming the target data into the format needed by the source data.

CHAPTER 7
IMPLEMENTATION

7.1 ALGORITHM USED: LINEAR REGRESSION

Linear regression is a kind of supervised machine learning. Carry do the regression analysis. Regression models anticipate the target value based on the independent variable. Its primary purpose is to anticipate and identify relationships between variables. The sort of relationship that is evaluated between the dependent and independent variables, as well as the collection of independent variables that are used, differs amongst regression models.

Linear regression is a statistical technique for predicting the value of the dependent variable (y) based on the value of the independent variable (y). A linear relationship between x (input) and y (output) is identified as a result of this regression technique. The term "linear regression" was coined as a result.

In the graph above, denotes personal salary and indicates job experience. The passenger line is the finest option for our vehicle.

The following model is returned.

X: Schooling records (one variable enter parameter)

Y: Data that has been labelled (supervised learning)

While schooling the version, it suits the fine line to estimate the cost of Y for a certain cost of. The version obtains the fine regression match lane 1: by determining the fine 1 and a few numbers. Intersept 2: x's C0-green

We get the fit lane after we get the best one and two values. So when we use our model to predict the value of y for the input value of x, it will predict the value of y.

How can I change the values of 1 and 2 to achieve the greatest fit line?

Cost Function (J):

By establishing an ideal regression trace, the model attempts to predict the value in such a way that the error difference between the projected and actual value is as little as feasible. As a result, it's critical to update the values between 1 and to discover the best value that minimizes the gap between the predicted y value (pred) and the actual y value (y).

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

A linear regression cost function y is the root mean square error (RMSE) between the predicted y value (pred) and the actual y value tre.

Machine learning in regression in linear

Linear regression is one of the most widely used and straightforward Machine Learning techniques. It's a statical predictive analytics technique. Linear regression is used to predict sales, salary, age, product price, and other continuous, real, or numeric variables.

As the name indicates, the linear regression process reveals a linear relationship between a dependent and one or more independent variables. Because linear regression displays a linear relationship, it identifies how the dependent variable's value varies as the independent variable's value changes. In a linear regression model, the relationships between variables are represented by sloping straight lines. Consider the diagram below.

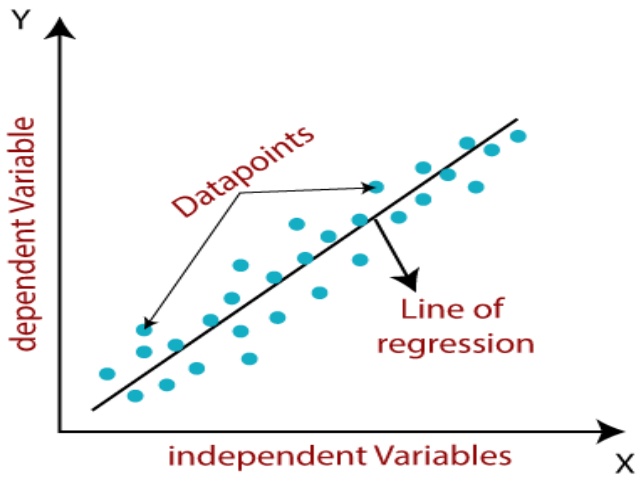


Fig 7.1 A slop straight line representing the relationship between the variables

Mathematically, we can represent a Linearr regression as:

$$y= a_0+a_1x+ \varepsilon$$

Y = Dependent variable in this case (target variable)

X stands for independent variable (predictive variable)

a0 is the intersection of two lines (additional degrees of freedom)

a1 = Coefficient of linear regression (scaling factor for each input value).

= Error at Random

Training datasets for a linear regression model representation are the values of the x and y variables.

Linear regression types

There are two different types of linear regression methods.

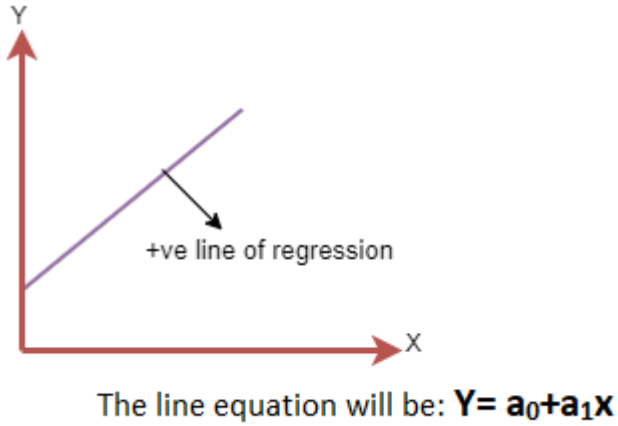
Simple linear regression is a kind of linear regression that uses a single independent variable to predict the value of a numerically dependent variable.

Numerous regression is a linear regression technique that predicts the value of a numerically dependent variable using multiple independent variables.

Line of linear regression

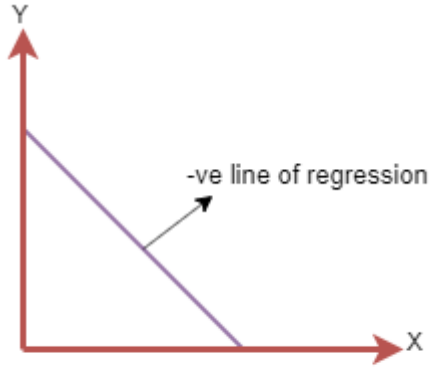
Linear graphs that depict the connection between dependent and independent variables are known as regression lines. There are two sorts of associations that regression lines can represent:

Positive linear connection: If the dependent variable rises on the y-axis while the independent variable increases on the x-axis, the relationship is said to be positive linear.



- **Negative Linear Relationship:**
  If the dependent variable falls on the Y-axis while the independent variable grows on

the X-axis, the connection is called a negative linear relationship**.**



The line of equation will be: $Y= -a_0+a_1x$

Assumptions of Linear Regression

Linear regression requires the following requirements. These are some formal tests that you should run while developing your linear regression model to guarantee that you receive the best results possible from your dataset.

A linear relationship between the features and target:

In linear regression, the dependent and independent variables are supposed to have a linear relationship.

Small or no multicollinearity between the features:

Multicollinearity refers to a significant connection between independent variables. Due to multicollinearity, determining the true relationship between the predictor and the target variables can be challenging. As a consequence, the model denotes the absence of minimum or multicollinearity in the feature or independent variable.

Homoscedasticity Assumption:

When the error interval for all independent variable values is the same, homoscadasty arises. There should be no discernible pattern distribution of the data in a scatter plot using homoscefasciyu.

Normal distribution of error terms:

A normal distribution pattern is predicted for this linear regression error. The confidence intervals will be too big or too narrow if the error terms are not normally distributed, making estimating the coefficient of determination difficult.

The qq chart may be used to verify this. The mistakes are evenly distributed if the

figure depicts a straight line with no deviations.

No autocorrelations:

The linear regression model, by error, does not imply autocorrelation. The model's accuracy will be greatly diminished if the error terms are correlated. If there is confidence between the residual errors, autocorrelation might arise.

7.2 Implementation steps:

Step 1: Start

Step 2: Open interface

Step 3: Upload data sets

Step 4: Pre-processing

Step 5: Training and testing

Step 6: Enter constraints and submit

Step 7: result

Step 8: stop

7.3 Implementation procedure:

The process of our project is as follows, Frist we had to open the files by using the jupyter notebook. After opening we had to import the libraries, now the user had to load the datasets. Now the system will calculate the average charges for the each disease. The preprocessing process will be started so now the data set will be preprocessed. The dataset now will be trained first, next the testing will be performed. Now the user need to enter the disease, smoker or non smoker, age. etc. After submitting the constraints the result will be generated. The result is the average cost and the insurance that can be claimed.

CHAPTER 8
RESULT ANALYSIS

8.1 Results Obtained

The data set we're using is Kaggle's medical cost personal dataset, which contains anonymous information about persons as well as yearly insurance premiums. Age and BMI are both continuous variables, whereas gender, smoking status, and geographic location are all categorical. The dataset has no missing values, as can be seen from the summary statistics below. The following are the characteristics in this dataset Age - Age of primary beneficiary person at the time treatment Sex -Gender of insurance contractor (female/male)

BMI -Body mass index (BMI) is a metric that compares the size of a person's body to the health of that person's body.

Children -Number of children for insurance coverage.

Smoker -If your primary insurance company smokes.

Region -The beneficiary’s current residential address.

Charges -Health insurance companies bill for individual medical expenses

	age	sex	bmi	disease	smoker	region	charges
0	19	female	27.900	a	yes	Hyderabad	16884.92400
1	18	male	33.770	b	no	Chennai	1725.55230
2	28	male	33.000	c	no	Chennai	4449.46200
3	33	male	22.705	d	no	Bangalore	21984.47061
4	32	male	28.880	e	no	Bangalore	3866.85520

Fig 8.1 Dataset charges

This section deals with the result analysis of the project performance of the model in the which we used the Linear regression algorithms. In exploratory data analysis we have found the correlation between the features and heat map to show correlation which we can see the following observation. String correlation between charges and smoker yes. Weak correlation between charges and age. Weak correlation between charges and BMI. Weak correlation between BMI and region _ southeast. Since the values for the weak correlations are less than 0.5 we can term them as insignificant and drop them. Which is shown in below figure.

```
# finding out the correlation between the features
data.corr()
```

	age	bmi	charges	disease_a	disease_b	disease_c	disease_d	disease_e	disease_f	disease_g	...	disease_i	dise
age	1.000000	0.109272	0.299008	-0.013072	0.005364	-0.009704	-0.005449	-0.004740	0.028231	0.022027	...	0.014845	-0.0
bmi	0.109272	1.000000	0.198341	-0.005528	0.004556	-0.015752	0.017370	-0.024842	-0.020347	-0.000754	...	0.034010	0.0
charges	0.299008	0.198341	1.000000	-0.002656	0.012770	-0.006645	0.014940	-0.024600	-0.034886	0.009268	...	0.023430	0.0
disease_a	-0.013072	-0.005528	-0.002656	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_b	0.005364	0.004556	0.012770	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_c	-0.009704	-0.015752	-0.006645	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_d	-0.005449	0.017370	0.014940	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_e	-0.004740	-0.024842	-0.024600	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	...	-0.110834	-0.1
disease_f	0.028231	-0.020347	-0.034886	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	...	-0.110834	-0.1
disease_g	0.022027	-0.000754	0.009268	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	...	-0.110834	-0.1
disease_h	-0.026721	0.010621	-0.003778	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_i	0.014845	0.034010	0.023430	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	...	1.000000	-0.1
disease_j	-0.010766	0.000782	0.012276	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	...	-0.110373	1.0
sex_female	0.020856	-0.046371	-0.057292	-0.001489	0.003491	-0.026386	0.008470	-0.026386	0.013450	0.038348	...	0.000978	0.0
sex_male	-0.020856	0.046371	0.057292	0.001489	-0.003491	0.026386	-0.008470	0.026386	-0.013450	-0.038348	...	-0.000978	-0.0
smoker_no	0.025019	-0.003750	-0.787251	-0.015788	-0.021957	0.002720	-0.003449	-0.003449	0.033568	0.002720	...	0.001462	0.0
smoker_yes	-0.025019	0.003750	0.787251	0.015788	0.021957	-0.002720	0.003449	0.003449	-0.033568	-0.002720	...	-0.001462	-0.0
region_Bangalore	-0.000407	-0.135996	-0.039905	-0.026407	-0.038018	-0.067046	-0.003185	0.037454	0.014232	-0.020601	...	-0.054205	0.0
region_Chennai	-0.011642	0.270025	0.073982	-0.019326	0.019836	0.036619	0.014241	-0.019326	-0.008137	-0.008137	...	-0.023476	0.0
region_Delhi	0.002475	-0.138156	0.006349	0.020640	0.003206	0.009017	0.003206	-0.020041	0.003206	0.026452	...	0.068770	-0.0
region_Hyderabad	0.010016	-0.006205	-0.043210	0.025843	0.014232	0.020037	-0.014796	0.002621	-0.008990	0.002621	...	0.009869	-0.0

21 rows × 21 columns

Fig 8.2 correlation between features

We begin to predict the charges with the help of the other features. In Model prediction our basic linear regression model predicting model predicting the cost of treatment looks good. And the closely matching results between training and test data means that our model is accurate.

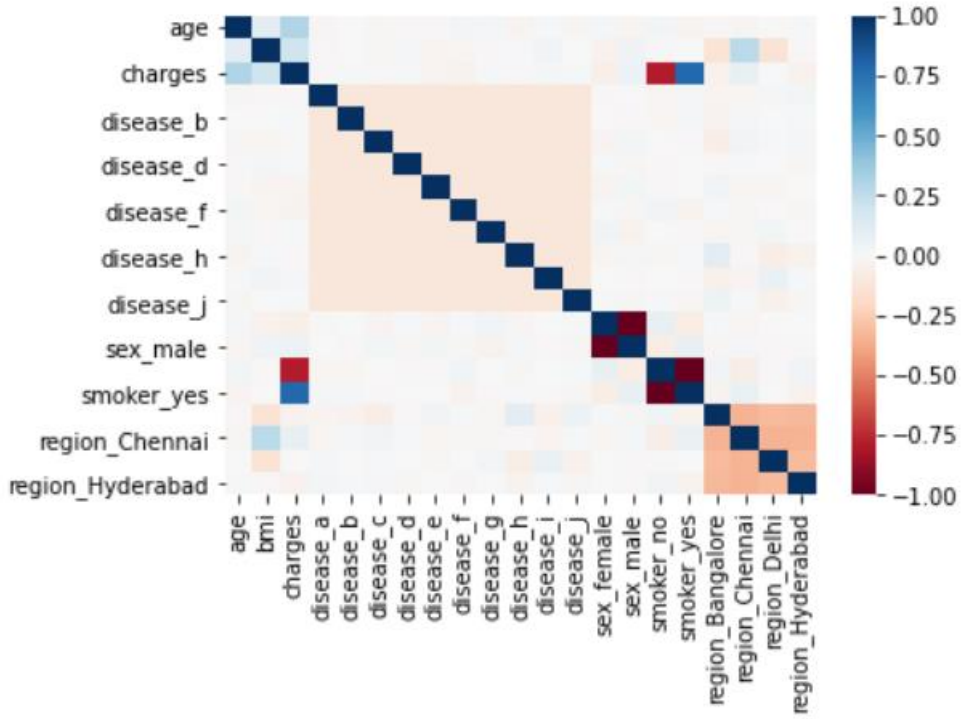


Fig 8.3 Heatmap

In [17]:
```
x = data.drop(['charges'], axis = 1)
y = data['charges']

x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.3, random_state = 0)
lr = LinearRegression().fit(x_train,y_train)
```

Model prediction

In [18]:
```
y_train_pred = lr.predict(x_train)
y_test_pred = lr.predict(x_test)

print(lr.score(x_test,y_test))
```

```
0.7835543337310373
```

our basic linear regression model predicting the cost of treatment looks good.

Model Evaluation ¶

In [19]:
```
from sklearn.metrics import r2_score,mean_squared_error
```

In [20]:
```
print('MSE train data:' , mean_squared_error(y_train,y_train_pred))
print('MSE test data:' , mean_squared_error(y_test,y_test_pred))

print('R2 train data:', r2_score(y_train,y_train_pred))
print('R2 test data:' , r2_score(y_test,y_test_pred))
```

```
MSE train data: 37814334.66529843
MSE test data: 34516474.62520199
R2 train data: 0.7317870418739104
R2 test data: 0.7835543337310373
```

The closely matching results between training and test data means that our model is accurate.

In [21]:
```
print('Please enter the disease: \n0-Ischemic heart disease, or coronary artery disease\n1-Brain Stroke \n2-Lower respirato
```

Fig 8.4 Training and testing accuracy

```
Please enter your Age: 36
Please enter your gender(m/f): m
Please enter your Body Mass Index: 39
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: 3
Do you smoke?
Yes- 1
No -0
Please enter your answer: 1
Please select the region where you wanted to get medicated:
0-Hyderabad
1-Banglore
2-Chennai
3-Delhi
Please enter an appppropriate option: 3
 The average cost of medication is  3409371.7261767276
The estimated insurance that can be claimed is  2386560.208323709
```

Fig 8.5 Output

The above figure explains the graph and medical expenses that are confirmed for the disease and also it also explains the insurance that can be claimed by the patient. The patient can determine the cost that is going to expense him for the disease from which he/she is suffering and also it explains the patient that how much amount can be claimed for the following disease based on the graph and expenditures.

CHAPTER 9
CONCLUSION AND FUTURE WORK

9.1 Conclusion

We've looked at the fundamentals of the linear regression model, how to use it o forecast charges, and how to compare anticipated and real outcomes. I hope you found this post helpful and that you now have a basic understaning of how a linear regression model works. For estimating medical expenditures, we suggested a machine learning approach.. 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.

9.2 Future Work

Work can be done in the future , making the system even more scalable. We're currently training and testing the algorithm with a few thousand records. We csan attempt scaling the method for a larger dataset with at least a million records in the future and see how it performs . We can leverage distributed frameworks like spark and Hadoop tomake the system scalable. These frameworks are capable of effectively handling large amounts of data.

CHAPTER 10

INDIVIDUAL TEAM MEMBERS REPORT

10.1 Individual Objective

E. Puneeth kumar – Building a model training and testing

P. Pavan krishna – Searching data of charges from hospitals and relatives and testing.

M. Raja Vardhan – Testing different algorithms and getting output

10.2 Contribution of Team Members

In this project all 3 team members have contributed equally under the guidance of Dr. P. Selvirajendran

E. Puneeth kumar – Building the he models and searching for suitable algorithms and paperwork.

P. Pavan krishna– Coordination with teammates and developing project and paper and testing.

M. Raja Vardhan – Testing, output and pictures of paperwork.

REFERENCES

[1] A. Ravishankar Rao, Subrata Gardaí, Coumarate Dey, Hang Peng, “Building predictive models of healthcare costs with open healthcare data”,2020 IEEE International Conference on Healthcare Informatics (ICHI) | 978-1-7281-5382-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICHI48887.2020.9374348

[2] Pei Shen, “Factor Analysis of Medical Expenses of the Hepatitis A patients in Guangdong”, 2016 8th International Conference on Information Technology in Medicine and Education.

[3] Ker-Tahj Shantung-Ming Yan and Pei-Wen Liu, “A Study on the Annualized Medical Expense Prediction Model of the Bureau of National Health Insurance --The Application of the Grey Prediction Theory”, 2006 IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan.

[4] Sheng Yao Zhou, Run tong Zhang*, “A Novel Method for Mining Abnormal Expenses in Social Medical Insurance” Auckland University of Technology. Downloaded on November 07,2020 at 17:01:50 UTC from IEEE Xplore.

[5] Li Cheng, Sino Jalin Pan, “Semi-supervised Domain Adaptation on Manifolds”, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 25, NO. 12, DECEMBER 2014.

[6] Yang Xia, Gunter Schreier, David C.W. Chang, Sandra Neubauer, Ying Liu, Stephen J. Redmond, Nigel H. Lovell,” Predicting Days in Hospital Using Health Insurance Claims”, IEEE Journal of Biomedical and Health Informatics - DOI :10.1109/JBHI.2015.2402692.

[7] Shruti Kaushik, Abhinav Choudhury, Sayed Natarajan, Larry A. Pickett, Varun Dutti,” Medicine Expenditure Prediction via a Variance Based Generative Adversarial Network”, 2018 IEEE International Conference volume

[8] Anuja Tike, Sanket Tavarageri,”A Medical Price Prediction System using Hierarchical Decision Trees”,2017 IEEE International Conference on Big Data (BIGDATA

APPENDEX A

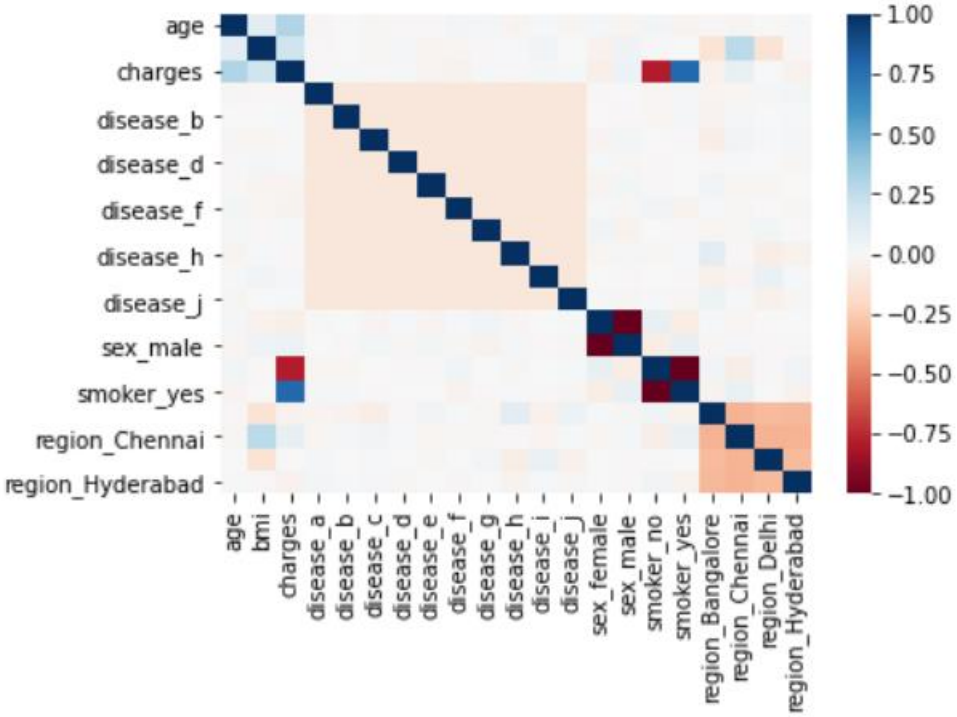
SAMPLE SCREEN SHOT

	age	sex	bmi	disease	smoker	region	charges
0	19	female	27.900	a	yes	Hyderabad	16884.92400
1	18	male	33.770	b	no	Chennai	1725.55230
2	28	male	33.000	c	no	Chennai	4449.46200
3	33	male	22.705	d	no	Bangalore	21984.47061
4	32	male	28.880	e	no	Bangalore	3866.85520

```
# finding out the correlation between the features
data.corr()
```

	age	bmi	charges	disease_a	disease_b	disease_c	disease_d	disease_e	disease_f	disease_g	...	disease_i	dise
age	1.000000	0.109272	0.299008	-0.013072	0.005364	-0.009704	-0.005449	-0.004740	0.028231	0.022027	...	0.014845	-0.0
bmi	0.109272	1.000000	0.198341	-0.005528	0.004556	-0.015752	0.017370	-0.024842	-0.020347	-0.000754	...	0.034010	0.0
charges	0.299008	0.198341	1.000000	-0.002656	0.012770	-0.006645	0.014940	-0.024600	-0.034886	0.009268	...	0.023430	0.0
disease_a	-0.013072	-0.005528	-0.002656	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_b	0.005364	0.004556	0.012770	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_c	-0.009704	-0.015752	-0.006645	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_d	-0.005449	0.017370	0.014940	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_e	-0.004740	-0.024842	-0.024600	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	-0.111296	...	-0.110834	-0.1
disease_f	0.028231	-0.020347	-0.034886	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	-0.111296	...	-0.110834	-0.1
disease_g	0.022027	-0.000754	0.009268	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	1.000000	...	-0.110834	-0.1
disease_h	-0.026721	0.010621	-0.003778	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	-0.111296	...	-0.110834	-0.1
disease_i	0.014845	0.034010	0.023430	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	...	1.000000	-0.1
disease_j	-0.010766	0.000782	0.012276	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	-0.110834	...	-0.110373	1.0
sex_female	0.020856	-0.046371	-0.057292	-0.001489	0.003491	-0.026386	0.008470	-0.026386	0.013450	0.038348	...	0.000978	0.0
sex_male	-0.020856	0.046371	0.057292	0.001489	-0.003491	0.026386	-0.008470	0.026386	-0.013450	-0.038348	...	-0.000978	-0.0
smoker_no	0.025019	-0.003750	-0.787251	-0.015788	-0.021957	0.002720	-0.003449	-0.003449	0.033568	0.002720	...	0.001462	0.0
smoker_yes	-0.025019	0.003750	0.787251	0.015788	0.021957	-0.002720	0.003449	0.003449	-0.033568	-0.002720	...	-0.001462	-0.0
region_Bangalore	-0.000407	-0.135996	-0.039905	-0.026407	-0.038018	-0.067046	-0.003185	0.037454	0.014232	-0.020601	...	-0.054205	0.0
region_Chennai	-0.011642	0.270025	0.073982	-0.019326	0.019836	0.036619	0.014241	-0.019326	-0.008137	-0.008137	...	-0.023476	0.0
region_Delhi	0.002475	-0.138156	0.006349	0.020640	0.003206	0.009017	0.003206	-0.020041	0.003206	0.026452	...	0.068770	-0.0
region_Hyderabad	0.010016	-0.006205	-0.043210	0.025843	0.014232	0.020037	-0.014796	0.002621	-0.008990	0.002621	...	0.009869	-0.0

21 rows × 21 columns



In [17]:
```
x = data.drop(['charges'], axis = 1)
y = data['charges']

x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.3, random_state = 0)
lr = LinearRegression().fit(x_train,y_train)
```

Model prediction

In [18]:
```
y_train_pred = lr.predict(x_train)
y_test_pred = lr.predict(x_test)

print(lr.score(x_test,y_test))
```

```
0.7835543337310373
```

our basic linear regression model predicting the cost of treatment looks good.

Model Evaluation ¶

In [19]:
```
from sklearn.metrics import r2_score,mean_squared_error
```

In [20]:
```
print('MSE train data:' , mean_squared_error(y_train,y_train_pred))
print('MSE test data:' , mean_squared_error(y_test,y_test_pred))

print('R2 train data:', r2_score(y_train,y_train_pred))
print('R2 test data:' , r2_score(y_test,y_test_pred))
```

```
MSE train data: 37814334.66529843
MSE test data: 34516474.62520199
R2 train data: 0.7317870418739104
R2 test data: 0.7835543337310373
```

The closely matching results between training and test data means that our model is accurate.

In [21]:
```
print('Please enter the disease: \n0-Ischemic heart disease, or coronary artery disease\n1-Brain Stroke \n2-Lower respirato
```

```
Please enter your Age: 36
Please enter your gender(m/f): m
Please enter your Body Mass Index: 39
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: 3
Do you smoke?
Yes- 1
No -0
Please enter your answer: 1
Please select the region where you wanted to get medicated:
0-Hyderabad
1-Banglore
2-Chennai
3-Delhi
Please enter an apppropriate option: 3
 The average cost of medication is  3409371.7261767276
The estimated insurance that can be claimed is  2386560.208323709
```

APPENDIX B

SAMPLE CODE

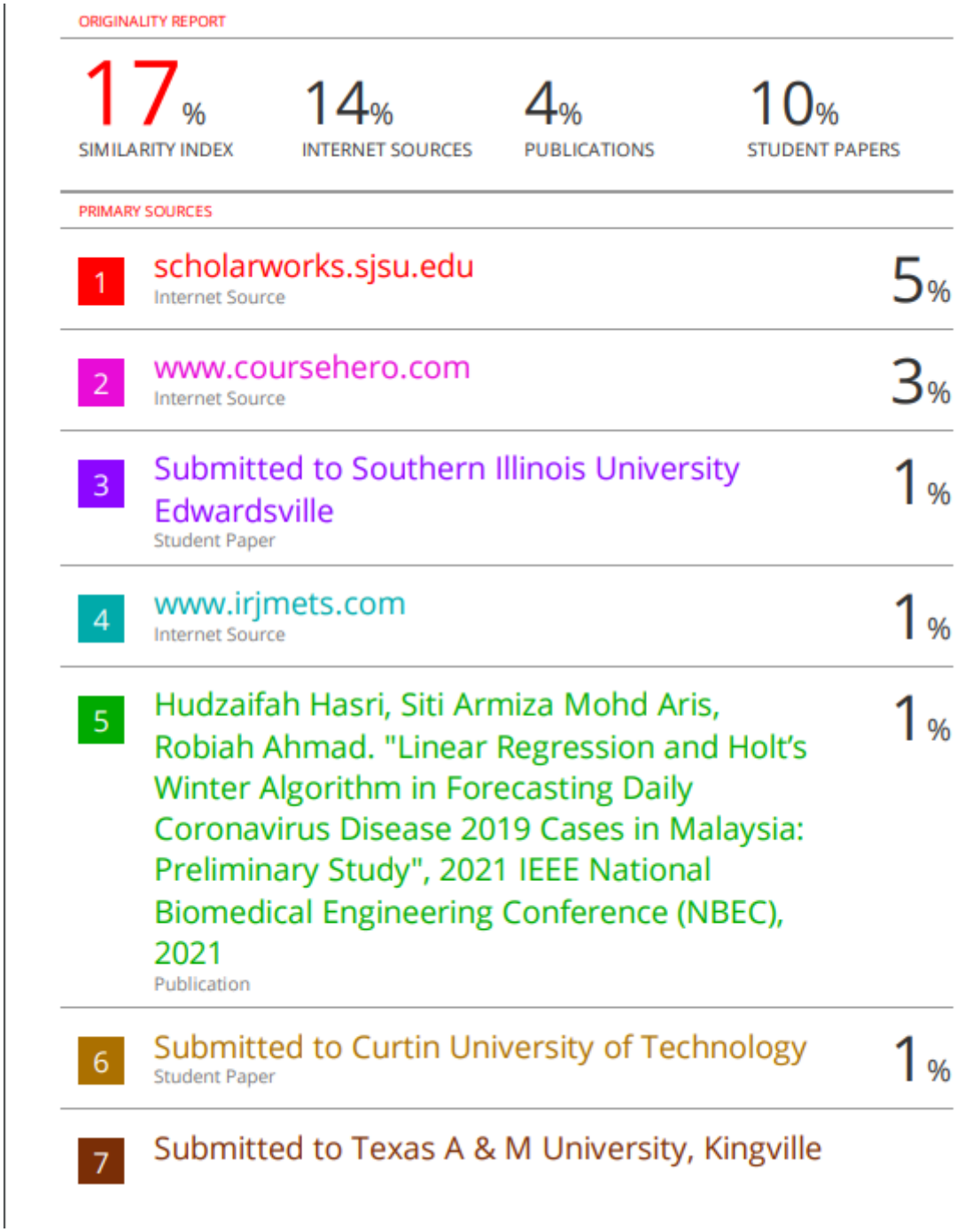
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
data_df=pd.read_csv("insurance.csv")

# See sample data
data_df.head()
data_df.info()
data_df.isnull().sum()
data_df.describe()
data_df.describe().columns
df_num = data_df[['age','bmi','charges']]
df_cat = data_df[['disease','sex','smoker','region']]
df1 = pd.get_dummies(df_cat)
df1
data.corr()
sns.heatmap(data.corr(), cmap='RdBu')
data.corr()['charges'].sort_values()
count, bin_edges = np.histogram(data['charges'])
data['charges'].plot(kind='hist', xticks=bin_edges, figsize=(20,12))
plt.title("Patient Charges")
plt.show()

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score,mean_squared_error
```

```
print('Please enter the disease: \n0-Ischemic heart disease, or coronary artery disease\n1-Brain Stroke \n2-Lower respiratory infections \n3-Chronic obstructive pulmonary disease\n4-Trachea, bronchus, and lung cancers \n 5-Diabetes mellitus \n6-Alzheimer’s disease and other dementias \n7-Dehydration due to diarrheal diseases \n8-Tuberculosis \n9-Cirrhosis\n')
age, sex, b, disease, smoker, region=0,0,0,0,0,0
age=int(input('Please enter your Age: '))
sex=input('Please enter your gender(m/f): ')
bmi=float(input('Please enter your Body Mass Index: '))
print('Please enter the disease: \n0-Ischemic heart disease, or coronary artery disease\n1-Brain Stroke \n2-Lower respiratory infections \n3-Chronic obstructive pulmonary disease\n4-Trachea, bronchus, and lung cancers \n5-Diabetes mellitus \n6-Alzheimer’s disease and other dementias \n7-Dehydration due to diarrheal diseases \n8-Tuberculosis \n9-Cirrhosis')
disease=int(input('Select the disease appropriately: '))
smoker=int(input('Do you smoke? \nYes- 1\nNo -0\nPlease enter your answer: '))
region=input('Please select the region where you wanted to get medicated: \n0-Hyderabad\n1-Banglore\n2-Chennai\n3-Delhi\nPlease enter an apppropriate option: ')
x_new_predict={
    "age":age,
    "bmi":bmi,
    "disease_a":1 if disease==0 else 0,
    "disease_b":1 if disease==1 else 0,
    "disease_c":1 if disease==2 else 0,
    "disease_d":1 if disease==3 else 0,
    "disease_e":1 if disease==4 else 0,
    "disease_f":1 if disease==5 else 0,
    "disease_g":1 if disease==6 else 0,
    "disease_h":1 if disease==7 else 0,
    "disease_i":1 if disease==8 else 0,
    "disease_j":1 if disease==9 else 0,
    "sex_male":0 if sex=='f' else 1,
```

```
"sex_female":0 if sex=='m' else 1,
"smoker_no":1 if smoker!=1 else 0,
"smoker_yes":1 if smoker==1 else 0,
"region_Hyderabad":1 if region==0 else 0,
"region_Bangalore":1 if region==1 else 0,
"region_Chennai":1 if region==2 else 0,
"region_Delhi":1 if region==3 else 0,
}
y_new_predict=abs(lr.predict([list(x_ new_ predict. values())]))
print(' The average cost of medication is ',y_new_predict[0]*100)
print('The estimated insurance that can be claimed is ',y_new_predict[0]*70)
```

APPENDEX C

PLAGARISM REPORT



ORIGINALITY REPORT

17%	14%	4%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	scholarworks.sjsu.edu Internet Source	5%
2	www.coursehero.com Internet Source	3%
3	Submitted to Southern Illinois University Edwardsville Student Paper	1%
4	www.irjmets.com Internet Source	1%
5	Hudzaifah Hasri, Siti Armiza Mohd Aris, Robiah Ahmad. "Linear Regression and Holt's Winter Algorithm in Forecasting Daily Coronavirus Disease 2019 Cases in Malaysia: Preliminary Study", 2021 IEEE National Biomedical Engineering Conference (NBEC), 2021 Publication	1%
6	Submitted to Curtin University of Technology Student Paper	1%
7	Submitted to Texas A & M University, Kingville	

APPENDEX D

PUBLICATION DETAILS

7th International Conference on Communication and Electronics Systems (ICCES 2022)

22-24, June 2022 | icoecs.org/2022/ | icces2022@gmail.com

Acceptance Letter

International Conference on Communication and Electronics Systems
Coimbatore, India
22-24, June 2022

Paper ID : **ICCES192**

Manuscript Title : **MEDICAL EXPENSE PREDICTION USING MACHINE LEARNING**

Author/s : **PUNEETH KUMAR E ,PAVAN KRISHNA P ,RAJA VARDHAN M,P SELVI RAJENDRAN**

On behalf of the Conference committee, I would like to congratulate you on your to the ICCES 2022 IEEE Conference, which will be held from 22nd to 24th June 2022 at PPG Institute of Technology, Coimbatore, India. You have been selected to deliver your oral presentation at the International Conference on Communication and Electronics Systems.

ICCES 2022 is an internationally-recognized IEEE Xplore IEEE conference, which dedicated solely for publication in the IEEE Xplore. Please visit the conference website for further updates [http://icoecs.org/2022/].

As a result of the review and results, we are pleased inform that you can now submit the full-length paper for inclusion into the IEEE Xplore ICCES proceedings. We appreciate if you could send the final version of your research paper at your earliest convenience, in order to ensure the timely publication. When submitting your final paper, please highlight the changes made according to the review comments.

Thank you for your contribution to the ICCES 2022 conference.

Yours sincerely,

V. Bindhu

Dr. V. Bindhu
[ICCES 2022 – Conference Chair]
Professor and Head
ECE Dept, PPG Institute of Technology
Coimbatore, India

APPENDEX E

TEAM DETAILS

NAME	Contact no	Mail ID	Role
Dr.P.SELVI RAJENDRAN	9445257280	selvir@hindustanuniv.ac.in	Supervisor

NAME	ROLL NUMBER	Contact no	Mail ID	Role
E.PUNEETH KUMAR	18113028	9390040550	18113028@student.hindustanuniv.ac.in	Team member

NAME	ROLL NUMBER	Contact no	Mail ID	Role
P.PAVAN KRISHNA	18113054	8074421642	18113054@student.hindustanuniv.ac.in	Team member

NAME	ROLL NUMBER	Contact no	Mail ID	Role
M.RAJA VARDHAN	18113057	9100191553	18113057@student.hindustanuniv.ac.in	Team member