

Prediction of Health Insurance Emergency using Multiple Linear Regression Technique

¹Dilip Kumar Sharma, ²Ashish Sharma,

¹Dilip Kumar Sharma, ²Ashish Sharma,

Department of Computer Engineering, GLA University, Mathura.
Department of Computer Engineering and Application, GLA University,
Mathura

E-Mail: dilip.sharma@gla.ac.in, ashishs.sharma@gla.ac.in

Introduction

Health insurance is a crucial market because one-third of Gross Domestic Product is spent on health insurance in the United States, and everyone needs some level of health care. Health insurance is one of the most significant investment an individual makes every year. This study is an effort to find mathematical models to predict future premiums and verify results using regression models. Medical costs that occur due to illness, accidents or any other health reasons are considerably expensive, by having health insurance, an individual is not liable for paying the entire medical costs of the procedure. According to the Office of Health and Human Services, the total health service budget for the fiscal year 2015 is around 1100 billion dollars.

There are several health care systems around the world. For example, single payer system followed by Canada where premiums are paid by taxes, government health care system followed by the United Kingdom where healthcare is the responsibility of central government. In country like the United states the insurance is funded through both government and private organizations. For example, 65% of insurance is through private insurances, 25 % through Medicare, Medicaid and veterinary programs which are funded by government and the remaining population is 2 uninsured. Kaiser Family Foundation analyzed health insurance premiums over the years from 2007-2016 and found out that health care premiums have been going up by 17% each year. It is also predicted that premiums are further going to increase. The question of finding the trends in the health care premiums is profoundly important as it tends to be the core of all presidential campaigns and other social movements. The healthcare premiums keep changing every year because of various factors such as medical trends, pharmaceutical trends, and political factors etc. over which the customer has no control.

ABSTRACT

The objective of proposed work is to predict the insurance charges of a person and identify those patients with health insurance policy and medical details whether they have any health issues or not. There are some types of health insurances, which are required to be predicted for a patient. The level of treatment in crisis department vary drastically depending the type of health insurance a person has by this we predict the insurance charges of a person In this paper, a multiple linear regression model for health insurance prediction is proposed. Some factors like age, gender, bmi, smoker, and children were input for developing the linear regression model. The model is very accurate considering that it works with real data from practice, with MAPE (Mean Absolute Percentage Error) around 3%, and coefficient of determination $R^2=0.7615896$. This is significant improvement in comparison to traditional models in some recent investigations. The proposed model can be utilized for future health care purpose that process to facilitate the decision making process.

Keywords: regression analysis, multiple linear regression.

Copyright

© 2020 The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

The only option an individual can have is to plan carefully for future expenses. There are no existing tools to the best of our knowledge that can predict future premiums based on historic data. Therefore, there is a need to conduct research to find the premiums across the United States. This research is focused on predicting the health insurance premiums for 2017 based on the individual health insurance market place data made public since 2014.

Survival of Health Care Insurance: Although there is great deal of uncertainty around healthcare system in recent years, health care system is not completely doomed because of the following reasons:

1) Premiums fluctuate according to the market place demand. For example, Aetna, United Health marketers lowered their premiums for the year 2015 as compared to 2014 and increased in 2016. Other Companies have affordable prices which target certain section of the society.

2) The government provides insurance from an increase of hikes. Consumers receive a tax credit for their insurance premiums. Advancements in Big data and Machine Learning algorithms Big data is being generated all around us by various digital, electronic devices such as smart phones, IOT devices, and social media such as Facebook, Twitter, Instagram etc. To take out significant worth from big data, optimal processing power, analytics capabilities, and skills are needed. Health insurance marketplace data provides premiums of health plans of several insurance companies participating in the marketplace. These rates are based on various complex factors such as

age, geographic location, the level of health coverage etc. and vary across states and markets. Individual plans data considered for this research may not be considered as big data by industrial standards but still, the conventional methods will not be able to detect patterns in data. The new data arrives every year. The methodology developed should be able to predict the trends in health care premiums based on the data collected in future. Machine Learning in calculations Tasks: machine Learning is utilized for an assortment for assignments in distinctive fields. An expansive amount of provisions utilize machine learning, And that amount will be expanding consistently. The machine Learning in errands could make comprehensively gathered under the accompanying categories:

- order.
- relapse.
- Clustering.

Machine learning algorithms use information with train An model. Those transform for preparing a model is Likewise alluded with By fitting a model for information. In other words, a machine Learning in algorithm fits An model ahead An preparation information situated. Relying upon the kind of the preparing data, machine Learning in calculations need aid comprehensively assembled under two categories: regulated and unsupervised machine Learning in. A regulated machine Learning in algorithm trains a model for a marked dataset. It might a chance to be utilized best with marked preparation information sets. Each perception in the preparing dataset need An set about offers Furthermore An name. Those indigent variable, otherwise called the reaction variable, speaks to those mark. Those free variables, otherwise called logical or predictor variables, represent able the features. A regulated machine Learning in algorithm takes in from information with gauge or estimated the relationship the middle of a reaction variable Also one or additional predictor variables. Those labels to An preparation information situated might be created manually alternately sourced from another framework. For example, to spam filtering, an extensive example of messages would gathered And manually marked Likewise spam alternately not. On the different hand, for offers forecasting, those mark will make chronicled sales, which might a chance to be sourced from An offers database. Regulated machine Learning in calculations might be comprehensively gathered under two categories: relapse And order calculations.

An unsupervised machine Learning in algorithm will be utilized when An information set may be unlabeled. It draws inferences starting with unlabeled datasets. Generally, those objective is should discover stowed away structure On unlabeled information.

Unsupervised machine Learning in calculations would by and large utilized 5 to clustering, aberrance detection, And dimensionality decrease. Those rundown for regularly utilized unsupervised machine Learning in calculations incorporates k-means, vital part analysis,

and independent esteem decay. Machine Learning in calculations utilize information should train a model. Those transform about preparing An model will be Likewise alluded to Similarly as fitting An model for information. In different words, An machine Learning in algorithm fits a model looking into a preparation information set. Contingent upon those sort of the preparation data, machine Learning in calculations need aid comprehensively aggregated under two categories: regulated Also unsupervised machine learning.

Literature review

Health insurance is a sort from claiming insurance result that particularly certifications the health costs or consideration of the insurance operator parts In the fall sick alternately bring a mischance. Comprehensively speaking, there need aid two sorts about medicines advertised By insurance companies, in particular inpatient (in-patient treatment) And outpatient (out-patient treatment).

Health insurance may be insurance that blankets the entirety or a and only the hazard of a individual causing restorative expenses, spreading those hazard In an expansive amount from claiming persons. Toward estimating those in general danger from claiming health awareness Also health framework liabilities again those hazard pool, an rebellion could create a schedule fund structure, for example, An month to month premium alternately payroll tax, with provide those cash to pay for the human services reductions specified in the insurance operator concurred upon. That profit will be administered by a focal association for example, a administration agency, private business, or not-for-profit substance. [1]. as stated by those health insurance Acquaintanceship for America, health insurance may be characterized as “coverage that gives to the installments of profits as an aftereffect from claiming disorder alternately harm. It incorporates insurance for misfortunes from accident, medicinal expense, disability, alternately unintentional demise Furthermore dismemberment”.

We led a dissection of analysis insurance information in the us. Those information that i utilization may be optional information that i get starting with kaggle[2]. This information holds 1338 lines comprising of a few variables, including age, sex, children, smoker, region, Furthermore charges. Yet the variables that we utilize would not all, best a couple would required. The reason for this Investigation may be to figure out that impact of a few variables once insurance operator of several factors on the expense of health insurance dependent upon factors that impact. Those Investigation utilized will be different straight relapse.

Each advanced procedure Also Online networking trade produces it. Systems, sensors Furthermore versatile apparatuses transmit it. Huge information may be arriving starting with numerous sources In a disturbing

velocity, volume, Furthermore mixed bag. With extricate serious quality from enormous data, ideal preparing power, analytics capabilities, and aptitudes are needed. Health insurance marketplace data provides premiums of health plans of several insurance companies participating in the marketplace. These rates are based on various complex factors such as age, geographic location, the level of health coverage etc. and vary across states and markets. Individual plans data considered for this research may not be considered as big data by industrial standards but still, the conventional methods will not be able to detect patterns in data. The new data arrives every year. The methodology developed should be able to predict the trends in health care premiums based on the data collected in future. [3]

However after analyzing the data for four- years from 2014-2017, it has been observed that future premiums incorporates the complex algorithmic details and market place data of the prior years. For example for determining the premiums for year 2015, insurance companies consider the total number of customers enrolled and customers expenditures, market competition etc. based on the previous year 2014 market experience. Like-wise 2016 year data includes the 2015 year details. In this paper, future rates for the year 2017 were predicted based on the combined historical data from years 2014 through 2016 using machine learning algorithms.

Various machine learning algorithms showed a great deal of significance in detecting patterns from variety of data. However, there is need to find the most suitable algorithms to predict the future 8 premiums from many existing algorithms. For this various machine learning algorithms are explored, tested with previous data. In this thesis, various regression models were used to predict future premiums. Flow chart of the entire process. In the data storage or data ingestion part, the initial data for the learning process is collected and stored. Typically, the data can be in different formats, including structured and unstructured formats. These data can be collected from Streaming APIs, web logs of businesses stored in cloud in distributed manner or in .csv, .xls or .json formats.[4]

The raw data is obtained from health insurance market place for the years 2014, 2015, 2016 and 2017. The data was preprocessed and transformed into relevant formats to be utilized by models. Preprocessing includes removing null values, inconsistent values etc. The transformation includes converting string and categorical variables into numerical values and finally scaling the variables to interval [0, 1]. In Regression model training and Regression model testing parts, the transformed data is trained and tested on different models such as simple Multiple Linear Regression, Decision tree regression, AdaBoost decision tree regression, and Gradient Boosting Regression.[5]

The process of training and testing is iterated several

times by changing parameters, till the best accuracy is achieved. Regression metrics are measured in evaluation model. The best model with satisfied metrics can be used to predict future values.

The overview of the health insurance market in the United States was discussed. According to Kaiser Family Foundation , health insurance coverage breakdown is as follows: 49% employee provided insurance, 20% Medicaid (which includes low-income households and Children's Health Insurance Program, 14 % Medicare(over age of 65), 2% other public (for veterans), 7% non-group such as individuals and finally 9% uninsured. Before discussing the features of the actual data set; in the next section, basic terminologies of health insurance are defined. These definitions are taken from healthcare.gov official website.

The amount paid for the health insurance each month. In calculation to the best, other costs are also paid for health care, including a deductible, copayments, and coinsurance. If users have a marketplace health plan, costs can be lowered with a premium tax credit. While choosing a plan the lowest monthly premium may not be the best match for a person. If a person requires a lot of medical assistance, a plan with a slightly higher premium but a lower deductible might save a lot of money. After enrolling in a plan, the first premium directly paid to the insurance company, not to the Health Insurance Marketplace.

Variables That Affect The Medical Cost:

We made this density analysis to find the variables which affect the medical costs. By the following analysis (you need to see the density chart):

1. Sex is not aggravating a effect on medical cost protection on male Also female need a same distribution/density against charges.
2. Locale may be not settled on a huge sway to medical expenses protection a result 4 areas very nearly need An same distribution/density against charges.
3. Smokescreen Also non-smoker influence the medical cost protection a result those distribution/density need An huge diverse.
4. Downright amount about children/dependents have An same thickness against charges, but zero amount about know youngsters. So, Assuming that you're not bring a tyke it will foundation an impact should medical fetches protection.

According to those insights, we conclude that **Smoke, children, age, sex and BMI** are the variables to predict the medical costs insurance.

According to a yearly issue brief released by American Health Insurance Program, there are several factors that influence individual marketplace premiums. These

include individuals' income, geography, age, and other factors. Individual coverage preferences and benefits also influence premiums. In addition to these, other market factors for example, such that underlying human services cosset development eliminate from claiming Brief premium adjustment programs, expanded use of services, business competition, Also danger pool impacts On certain states Also businesses climb premiums. Administration contribution for example, such that single person scope requirement, premium subsidies, What's more expanded familiarity with scope choices Also enrollment initiatives reduces the premiums. Premiums in Ohio are very different from premiums in New Jersey. Insurance companies with a population of customers of both healthy and unhealthy individuals are likely to survive in insurance markets and can provide effective costs to customers. Kaiser Family Foundation analyzed premium changes based on insurance companies exit from the market place. Kaiser Family Foundation showed that overall health insurance market place is less impacted by exiting of a health insurance company when there are many players in the market place. However, the states with fewer insurance companies that 18 provide health plans to customers had a higher impact on the insurance premiums because the remaining companies in the market place can hike the premiums. Other papers by KFF mentioned medical costs for people without insurance are more compared to people with insurance. This is partially due to billing methods of hospitals and negotiation between insurance companies and hospitals. Henry et al. proposed a model based on the data from HMO network using multi variety analysis to predict, the people more likely to utilize health services. Health Care Payment Learning & Action Network (HCP LAN) suggested substitute compensation plan with customer centric model based on services utilized [6].

Methodology

This section should contain information about Algorithms applied to the model and its Implementation.

Machine learning be those field of study that provides for workstations those ability on take in without continuously unequivocally programmed. ML will be a standout amongst the A large portion energizing innovations that you quit offering on that one might need at any point run into. Similarly as it may be apparent starting with that name, it provides for the PC that makes it all the more comparative on humans: the capacity on figure out. Machine taking in will be actively constantly utilized today, maybe for a lot of people more puts over person might hope. We likely utilize a Taking in calculation handfals about run through without indeed knowing it.

Today, organizations are utilizing machine Taking in on enhance benefits of the business decisions, expansion productivity, recognize disease, conjecture weather,

Furthermore do a lot of people more things. With those exponential Growth about technology, we not best compelling reason superior devices to comprehend those information we at present have, Anyhow we additionally necessity on get ready ourselves to those information we will need. Should accomplish this objective we need to manufacture canny machines. We could compose a project should do basic things. Be that for the vast majority of times Hardwiring brainpower clinched alongside it is challenging. Most ideal approach on do it is should have sixth review somehow to machines with gain things themselves. An instrument to Taking in – whether An machine can figure out starting with enter that point it does those difficult worth of effort to us.

For our model we utilize various straight relapse Also ANN Around them the numerous straight relapse provides for the best exactness thus we precede for various straight relapse. There would two elementary stages in the system:. 1. Preparing phase: That framework is prepared by utilizing those information in the information situated and fits An model dependent upon those algorithm decided Appropriately. 2. Trying phase: That framework will be Gave with the inputs What's more will be tried to its working.

The precision may be checked. Also therefore, the information that is used to prepare the model alternately test it, need with a chance to be suitable. The system is designed to predict health insurance charges.

Multiple linear regression:

Multiple linear regression (MLR), likewise referred to basically Likewise numerous regression, will be An measurable procedure that utilization a few logical variables should anticipate those result of a reaction variable. The objective of multiple linear regression (MLR) may be to model the straight relationship the middle of those logical (independent) variables Furthermore reaction (dependent) variable.

By essence, numerous relapse will be those development of ordinary least-squares (OLS) relapse that includes more than particular case logical variable.

The Formula for Multiple Linear Regression Is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y – intercept (constant term)

β_p

= slope coefficients for each explanatory variable

ϵ

= the model's error term (also known as the residuals)

Clarifying Multiple Linear Regression

A straightforward direct relapse is a capacity that permits an examiner or analyst to make forecasts around one variable dependent on the data that is thought about another variable. Direct relapse must be

utilized when one has two consistent factors—an autonomous variable and a needy variable. The free factor is the boundary that is utilized to compute the reliant variable or result. A numerous relapse model reaches out to a few illustrative factors.

The numerous relapse model depends on the accompanying suppositions:

There is a direct connection between the needy factors and the autonomous factors.

The free factors are not very exceptionally corresponded with one another.

yi perceptions are chosen freely and arbitrarily from the populace.

Residuals ought to be ordinarily disseminated with a mean of 0 and change σ .

The coefficient of assurance (R-squared) is a factual metric that is utilized to gauge the amount of the variety in result can be clarified by the variety in the autonomous factors. R2 consistently increments as more indicators are added to the MLR model despite the fact that the indicators may not be identified with the result variable.

R2 without anyone else can't in this way be utilized to recognize which indicators ought to be remembered for a model and which ought to be avoided. R2 must be somewhere in the range of 0 and 1, where 0 demonstrates that the result can't be anticipated by any of the free factors and 1 shows that the result can be anticipated without blunder from the autonomous factors.

When deciphering the consequences of a numerous relapse, beta coefficients are substantial while holding every single other variable steady ("all else equivalent"). The yield from a numerous relapse can be shown on a level plane as a condition, or vertically in table structure.

Steps Involved in any Multiple Linear Regression Model

Stage 1: Data Pre Processing.

1. Bringing in the libraries.
2. Bringing in the information set.
3. Encoding the straight out information.
4. Maintaining a strategic distance from the fake variable snare.
5. Parting the informational collection into preparing set and test set.

Stage 2: Fitting Multiple Linear Regression to the Training set.

Stage 3: Predicting the test set outcome.

The methodology, depicted in the following figure, has been adopted for conducting health insurance prediction dataset classification experiment.

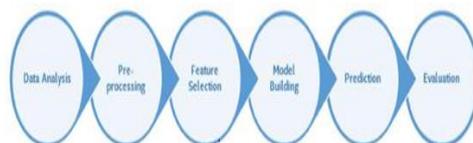


Figure.1. Data set classification

Health Insurance Data

Data from market place shows that 30 million plans are available each year for purchases in different states across the United States. And these plans are selected according to users anticipated service usage. It can be seen prices of premiums are affected by various factors such as age, plan type, geographic location etc. Therefore there is a need to develop a mathematical model to predict premiums based on various parameters which impact the premiums. The premium rates are set by insurance providers using complex algorithms based on previous years health care utilization and total number of enrollments.

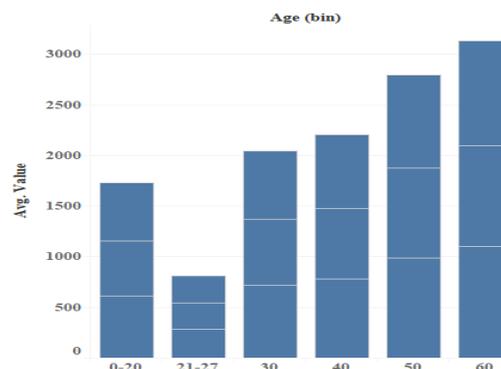


Figure.2. Average premium prices plotted against years for different ages

However after analyzing the data for four- years from 2014-2017, it has been observed that future premiums incorporates the complex algorithmic details and market place data of the prior years. For example for determining the premiums for year 2015, insurance companies consider the total number of customers enrolled and customers expenditures, market competition etc. based on the previous year 2014 market experience. Like-wise 2016 year data includes the 2015 year details. In this paper, future rates for the year 2017 were predicted based on the combined historical data from years 2014 through 2016 using machine learning algorithms. Various machine learning algorithms showed a great deal of significance in detecting patterns from variety of data. However, there is need to find the most suitable algorithms to predict the future premiums from many existing algorithms. For this various machine learning algorithms are explored, tested with previous data.

In this thesis, various regression models were used to predict future premiums. Flow chart of the entire

process is in Figure. In the data storage or data ingestion part, the initial data for the learning process is collected and stored. Typically, the data can be in different formats, including structured and unstructured formats. These data can be collected from Streaming APIs, web logs of businesses stored in cloud in distributed manner or in .csv, .xls or .json formats.

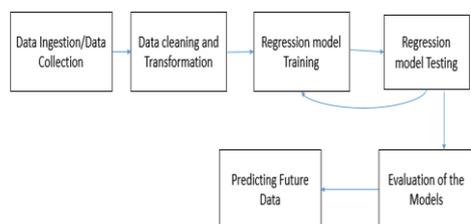


Figure.3. Flowchart of the Regression Process

The raw data is obtained from health insurance market place for the years 2014, 2015, 2016 and 2017. The data was preprocessed and transformed into relevant formats to be utilized by models. Preprocessing includes removing null values, inconsistent values etc. The transformation includes converting string and categorical variables into numerical values and finally scaling the variables to interval [0, 1]. In Regression model training and Regression model testing parts, the transformed data is trained and tested on different models such as simple Multiple Linear Regression, Decision tree regression. The process of training and testing is iterated several times by changing parameters, till the best accuracy is achieved. Regression metrics are measured in evaluation model. The best model with satisfied metrics can be used to predict future values

IMPLEMENTATION

To implement the project we used jupyter notebook editor.

Step 1:

Importing the required packages those are numpy, pandas and matplotlib which is used in further their alias names np, pd, plt respectively.\

NumPy is a usually utilized Python information examination bundle. By utilizing NumPy, you can accelerate your work process, and interface with different bundles in the Python biological system, as scikit-learn, that utilization NumPy in the engine. NumPy was initially evolved in the mid 2000s, and emerged from a considerably more established bundle called Numeric.

NumPy is an open source numerical Python library. NumPy contains a multi-dimensional cluster and network information structures. It tends to be used to play out various numerical procedures on clusters, for example, trigonometric, measurable and mathematical schedules. NumPy is an expansion of Numeric and Num array.

Pandas is a library composed for the Python programming language for information control and examination. Specifically, it offers information structures and activities for controlling numerical tables and time arrangement, which is a Panel Data. In this way, the library is named is Pandas. Pandas library is based on NumPy bundle.

Matplotlib is a plotting library for the Python programming language and its numerical science expansion NumPy. It gives an article arranged API to implanting plots into applications utilizing broadly useful GUI toolboxes.

matplotlib.pyplot is an assortment of order style works that make matplotlib work like MATLAB. Each pyplot work rolls out some improvement to a figure: e.g., makes a figure, makes a plotting zone in a figure, plots a few lines in a plotting region, enlivens the plot with names, etc.

- `import numpy as np`
- `import pandas as pd`
- `import matplotlib.pyplot as plt`

Step 2:

Importing dataset which was collected.

CSV is a basic document position used to store plain information, for example, a spreadsheet or database.

`read_csv` is a significant pandas capacity to peruse csv records and do procedure on it. One can peruse a csv record locally, however from a URL through `read_csv` or one can pick what segments expected to trade with the goal that we don't need to alter the cluster later In this we read the file from local disk. Here 'r' indicates the raw data file.

- `df = pd.read_csv(r"C:\Users\hp\Desktop\insurance.csv")`

Step 3:

Data Preprocessing is a process of converting the raw data into cleaned text. In

general preprocessing refers to the transformations applied to the data before feeding into the algorithm. Therefore to avoid misleading results, data has to screened before giving as input to the ML algorithm.

The following instruction checks the dataset that it contains any null values or not. If it contains null values then that attribute value returns as True otherwise it returns False.

- `df.isnull().any()`

The following instruction replaces the null values with the median value of that column.

- `df.fillna(df.median(),inplace=True)`

This is for checking first five records in the dataset.

- `df.head()`

Step 4:

In this step we are checking for the data correlation.

Correlation means When independent variable and dependent variable are emphatically connected together we state they have a High Correlation. The word Correlation is made of Co-(signifying "together"), and Relation. Connection is Positive when the qualities

increment together, and. Connection is Negative when one worth abatements as different increments.
df.corr()

Based on the correlation factor we ignore the weak independent variables in order to get the accurate prediction.

Step 5:

Pandas give a remarkable strategy to recover lines from a Data outline. Dataframe. iloc[] technique is utilized when the file mark of an information outline is some different option from numeric arrangement of 0, 1, 2, 3... n or in the event that the client doesn't have the foggiest idea about the list mark.

In this we are separating independent variable as y and dependent variable as x from the dataset in order to process the data.

- `x = df.iloc[:,0:5].values`
- `y = df.iloc[:,6].values`
checking the variable and examining its dimensions.
- x
- y

Step 6:

Parting the information as train and test for that we use `train_test_split` is a capacity in Sklearn model determination for parting information clusters into two subsets: for preparing information and for testing information. With this capacity, you don't have to isolate the dataset physically. As a matter of course, Sklearn `train_test_split` will make irregular parcels for the two subsets.

In this we have given 80 percent of information for preparing and 20 percent of information for testing. This can be achieved through `test_size` parameter which assigns them records sequentially by making `random_state = 0`.

The result of the `train_test_split` method assigns four values as shown in the instruction.

- `from sklearn.model_selection import train_test_split`
- `x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)`

We checked the train values and test values for examining their dimensions and values.

- `x_train`
- `y_train`
- `x_test`
- `y_test`

Step 7:

In the following instructions we create an object to the model for multilinear regression. From that object using `fit` method we train the model with our dataset. For predicting the test values `predict` method is used.

- `from sklearn.linear_model import LinearRegression`

- `mlr = LinearRegression()`
- `mlr.fit(x_train, y_train)`
- `y_pred = mlr.predict(x_test)`

Step 8:

To get precision, R-squared is a decency of-fit measure for straight relapse models. R-squared estimates the quality of the connection between your model and the reliant variable on an advantageous 0 – 100% scale. In the wake of fitting a direct relapse model, you have to decide how well the model fits the information.

- `from sklearn.metrics import r2_score`
- `r2_score(y_test, y_pred)`

Conclusion

In this task, we have projected strategies for anticipating medical coverage. Medical coverage is a kind of protection item that explicitly ensures the health expenses or care of the protection individuals in the event that they become sick or have a mishap. Comprehensively, there are two kinds of medicines offered by insurance agencies and a sort of protection inclusion that takes care of the expense of a guaranteed person's clinical and careful costs. The degree of treatment in crisis offices shifts altogether relying upon what kind of medical coverage an individual has by this we foresee the protection charges of an individual.

References

1. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikitlearn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, Oct 2011.
2. E. Wang; and G. Gee, "Larger Issuers, Larger Premium Increases: Health insurance issuer competition post-ACA," 2015.
3. C. C. a. A. Semanskee, "Analysis of UnitedHealth Group's Premiums and Participation in ACA Marketplaces," 2016.
4. Ng, "Machine learning for Housing Price Prediction Mobile Application," Masters, Department of Computing, Imperial College London, 2015
5. World Health Organization. Public spending on health: a closer look at global trends. No. WHO/HIS/HGF/HFWorkingPaper/18.3. World Health Organization, 2018.
6. Sturm, Roland. The effects of obesity, smoking, and drinking on medical problems and costs. Health affairs 21, no. 2 (2002): 245-253.
7. Sturm, Roland, Ruopeng An, Josiase Maroba, and Deepak Patel. The effects of obesity, smoking, and excessive alcohol intake on healthcare expenditure in a comprehensive medical scheme. South African Medical Journal 103, no. 11 (2013): 840-844.
8. An, Ruopeng. Health care expenses in relation to obesity and smoking among US adults by gender,

- race/ethnicity, and age group: 19982011. Public Health 129, no. 1 (2015): 29-36.
9. Kim, David D., and Anirban Basu. Estimating the medical care costs of obesity in the United States: systematic review, meta-analysis, and empirical analysis. Value in Health 19, no. 5 (2016): 602-613.
 10. Han, Kimyoung, Minho Cho, and Kihong Chun. Determinants of health care expenditures and the contribution of associated factors: 16 cities and provinces in Korea, 2003-2010 Journal of Preventive Medicine and Public Health 46, no. 6 (2013): 300.
 11. Kumar, Manoj, and Ashish Sharma. "Mining of data stream using "DDenStream" clustering algorithm." 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE). IEEE, 2013.
 12. Sharma, Ashish, Anant Ram, and Archit Bansal. "Feature Extraction Mining for Student Performance Analysis." Proceedings of ICETIT 2019. Springer, Cham, 2020. 785-797.
 13. Sharma, Ashish, Ashish Sharma, and Anand Singh Jalal. "Distance-based facility location problem for fuzzy demand with simultaneous opening of two facilities." International Journal of Computing Science and Mathematics 9.6 (2018): 590-601.
 14. Ram, Anant, et al. "A density based algorithm for discovering density varied clusters in large spatial databases." International Journal of Computer Applications 3.6 (2010): 1-4.
 15. Mishra, Ayushi, et al. "A robust approach for palmprint biometric recognition." International Journal of Biometrics 11.4 (2019): 389-408.
 16. Singh, Anshy, Shashi Shekhar, and Anand Singh Jalal. "Semantic based image retrieval using multi-agent model by searching and filtering replicated web images." 2012 World Congress on Information and Communication Technologies. IEEE, 2012.
 17. Shekhar, Shashi, et al. "A WEBIR crawling framework for retrieving highly relevant web documents: evaluation based on rank aggregation and result merging algorithms." 2011 International Conference on Computational Intelligence and Communication Networks. IEEE, 2011.
 18. Varun K L Srivastava, N. Chandra Sekhar Reddy, Dr. Anubha Shrivastava, "An Effective Code Metrics for Evaluation of Protected Parameters in Database Applications", International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.1.3, 2019. doi.org/10.30534/ijatcse/2019/1681.32019
 19. Prasad, K.S., Reddy, N.C.S. & Puneeth, B.N. A Framework for Diagnosing Kidney Disease in Diabetes Patients Using Classification Algorithms. SN COMPUT. SCI. 1, 101 (2020).