International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Intelligent Medical Expense Prediction

Mrs.K.Surekha[1], Das Neha[2], Geddada Darshitha[3], Jillela Anoosha Deepthika[4], Alapati Lokesh[5], Manabottula Srinivasu Rao[6]

[1]Associate Professor, Cse Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh – 533255
[2,3,4,5,6]Student, Cse Department, Vsm College Of Engineering, Ramachandrapuram, Dr.B.R.Ambhedkar Konaseema Dt, Andhra Pradesh – 533255

ABSTRACT
Medical costs are one of the most common reoccurring expenses in a personal life. It is general known that a person's lifestyle and numerous physical factors determine medical expenses. By physical research several significant reasons like smoking, age, accidents, age and BMI causes personal medical care. This project is valuable because it can help healthcare providers, insurance companies and policy makers forecast future cost, optimize resource allocation and maintaining organizational budgets, identify high risk patients for proactive care and reducing overall budget. This project is to examine and identify a link between personal medical cost and other characteristics. Then by generating linear regression model and comparing them using ANOVA. We use the significant traits as predictors to forecast medical expenditures. The regression model can predict the charges with more than 75% accuracy. Using insurance data from diverse persons variables such as Smoking, age, number of children, area and BMI can predict the data.To predict medical expenses, the data should be collected and preprocess. Perform Exploratory Data Analysis (EDA) to identify correlations and outliers. Train a linear regression model using Mean Squared Error and Rsquared. Apply ANOVA to analysis the impact of categorial factors on medical cost. Finally optimize the model using features selection ,polynomial regression or regularization techniques to improve accuracy

INTRODUCTION
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

EXISTING SYSTEM
In sample sizes ranging from small to large, statistical approaches (E.g., Linear regression) suffer due to the zero point spike and skewed distribution of 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.).

PROBLEM STATEMENT

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

PROPOSED SYSTEM

The medical information and expenditures billed by health insurance companies are included in the data. It has 1338 rows of information with the 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.

IMPLEMENTED ALGORITHMS

a) Linear Regression

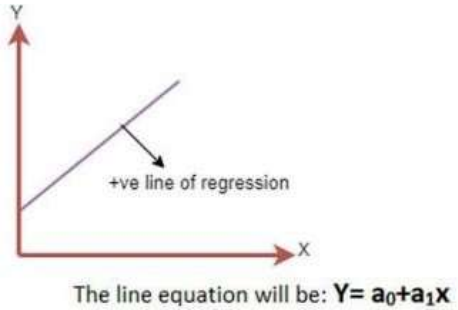
Linear regression is a kind of supervised machine learning. Carry do the regression analysis. Regression models anticipate the target value based on the independent variable. Its primary purpose is to anticipate and identify relationships between variables. The sort of relationship that is evaluated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

a) Positive 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:

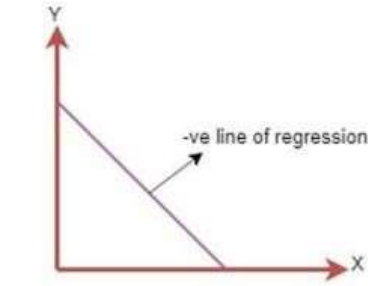
Positive linear connection:

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.



c)Negative linear regression

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.

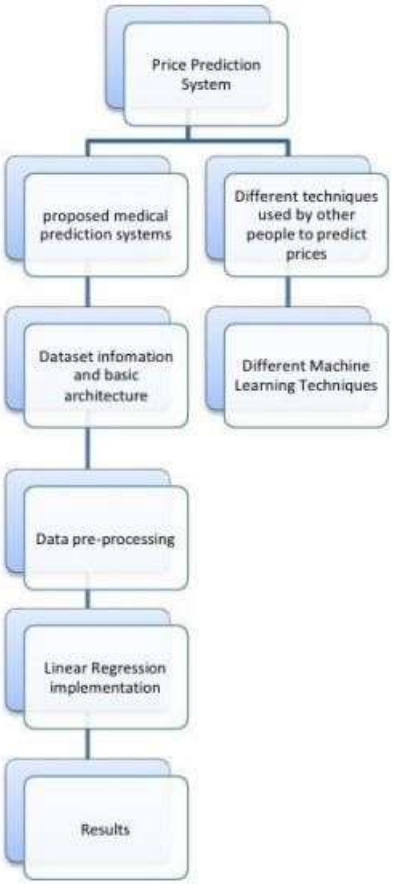


The line of equation will be: $Y= -a_0+a_1x$

SYSTEM ARCHITECTURE

The Intelligent medical expense prediction system architecture typically consists of different components working together. The project implementation, which includes data preprocessing, cleaning, training, and testing, is depicted in the diagram step-by-step. The steps involved in our project were as follows: first, we used the Jupiter notebook to open the files.

System Architecture Diagram



TESTING

Testing is the process of detecting errors. Testing is a process of executing a program with the indent of finding an error. Testing is a crucial element of software quality assurance and presents ultimate review of specification, design and coding. Testing performs a very critical role for quality assurance and for ensuring the reliability of software. The results of testing are used later on during maintenance also.

Psychology of testing

The aim of testing is often to demonstrate that a program works by showing that it has no errors. The basic purpose of testing is to detect the errors that may be presenting the program. Hence one should not start testing with the intent of showing that a program works, but the intent should be to show that a program doesn't work. Testing is the process of executing a program with the intent of finding errors

TESTING METHODOLOGIES

a) System testing

Testing has become an integral part of any system or project especially in the field of information technology. The importance of testing is a method of justifying, if one is ready to move further, beit to be check if one is capable to with stand the rigors of a particular situation cannot be underplayed and that is why testing before development is so critical.

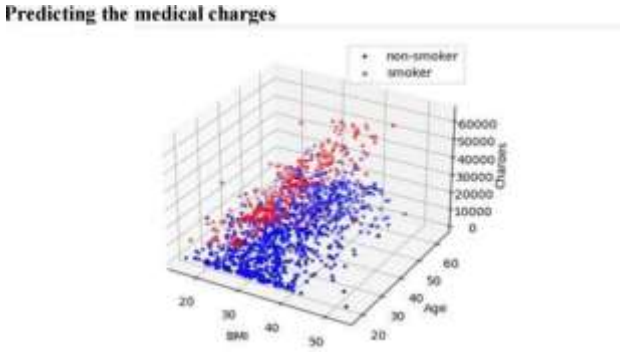
b) Module testing

To locate errors, each module is tested individually. This enables us to detect error and correct it without affecting any other modules. All the modules are individually tested from bottom up starting with the smallest and lowest modules and proceeding to the next level User Interface, Crop Recommendation, Fertilizer Recommendation Modules**.**

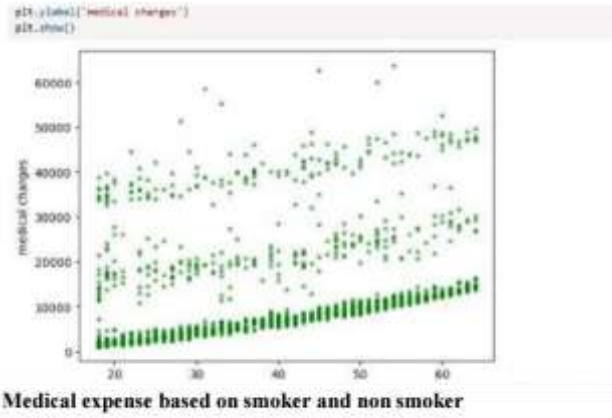
c) Integration testing

After the module testing, the integration testing is applied. When linking the modules there may be chance for errors to occur, these errors are corrected by using this. taking into account user-specific data, and return the recommendations to the user through the User Interface Module

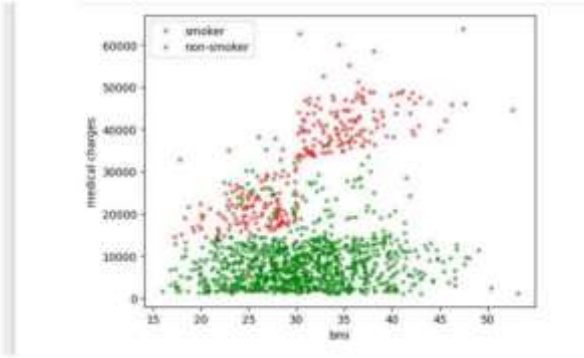
RESULT ANALYSIS



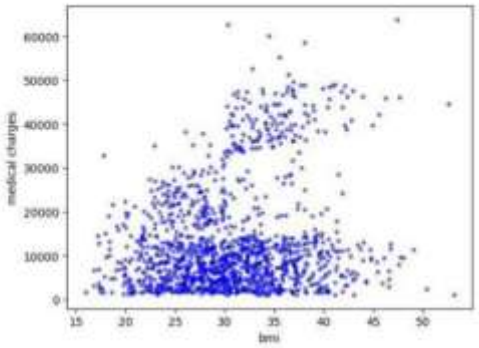
Plotting medical charges



Medical expense based on smoker and non smoker



Through BMI predictions



Binerizing smoker and sex columns

CONCLUSION

We've covered the basics of the linear regression model, how to forecast charges using it, and how to compare expected and actual results. I hope this post helped you understand the fundamentals of how a

linear regression model operates. We proposed a machine learning method for estimating medical expenses. We used linear regression techniques and found that the dependent variable is determined by features such as age and BMI. This model produced a better result than any other experiment. 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 from3.89 percent in 2015, and cash-based medical use makes for 65.06 percent of current medicaluse. 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. 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.

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 ©2020IEEE|DOI:10.1109/ICHI48887.2020.937434 8
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 Anuja Tike, Sanket Tavarageri,”A Medical Price Prediction System using Hierarchical Decision Trees”,2017 IEEE International Conference on Big Data (BIGDATA engineers)