



## PREDICTION OF PROSPECTIVE MEDICAL INSURANCE COST

**S. VENKATA LAKSHMI SAI SUBRAHMANYAM**, MCA, DCA, DVR & Dr. Hima Shekar MIC

College of Technology, A.P., India.

**SK. UDDANDU SAHEB**, Assistant Professor, Dept. Of AI & IT, DVR & Dr. Hima Shekar MIC college of Technology, A.P., India.

**Abstract**—In health insurance many factors such as pre-existing body condition, family medical history, Body Mass Index (BMI), marital status, location, past insurances etc affects the amount. proposed research approach uses Linear Regression, Decision Tree Regression and Gradient Boosting Regression and also streamlit as a framework. We had used a medical insurance cost dataset that was acquired from the KAGGLE repository for the cost prediction purpose, and machine learning methods are used to show the forecasting of insurance costs by regression model comparing their accuracies.

### INTRODUCTION

Health is a leading provider of information for professionals and students in medicine, nursing, allied health, pharmacy and the pharmaceutical industry. Health insurance is a necessity nowadays, and almost every individual is linked with a government or private health insurance company. Factors determining the amount of insurance vary from company to company. Also people in rural

areas are unaware of the fact that the government of India provide free health insurance to those below poverty line. It is very complex method and some rural people either buy some private health insurance or do not invest money in health insurance at all. Apart from this people can be fooled easily about the amount of the insurance and may unnecessarily buy some expensive health insurance. The value of insurance in the lives of individuals. That's why it becomes important for insurance companies to be sufficiently precise to measure the amount covered by this specific policy and the insurance charges which must be paid for it.

In this project our goal is to predict medical prices based on the data we have in hand. In the first few chapters of this report, we will compare the work of various authors in the area of price prediction and we will also provide the information in detail, about some of the techniques used in health care domain to predict the health care prices. Later, we will propose the design of a new system which will use Medicare payment datasets. The proposed system



can be called as a medical price prediction system. Such a system will be useful for patients, and government officials alike.

## LITERATURE REVIEW

Elish et al. (2008) [5] applied SVM on the four NASA data sets [12] against other machine learning algorithms and realized SVM's performance was superior to that of others. In this experiment, the study of Malhotra et. al. (2015) [10] have been guiding us while deciding to select which machine learning algorithms we have used for defect prediction in software systems. They categorized the machine learning algorithms based on distinct learners such as Ensemble Learners, Bayesian Learners, Neural Networks and SVM. According to these categories, we selected seven different machine learning algorithms to estimate software defect.

## RELATED WORK

Insurance is a policy that helps to cover up all loss or decrease loss in terms of expenses incurred by various risks. A number of variables affect how much insurance costs. Machine Learning ( ML) in the insurance sector can make insurance more effective. ML is one of the computational intelligence aspects when it comes to exploitation of historical data in a wide range of applications

and systems. There are some limitations in ML so; Predicting medical insurance costs using ML approaches is still a problem in the healthcare industry and thus it requires few more investigation and improvement. Using the machine learning algorithms, this study provides a computational intelligence approach for predicting healthcare insurance costs.

## PROPOSED WORK

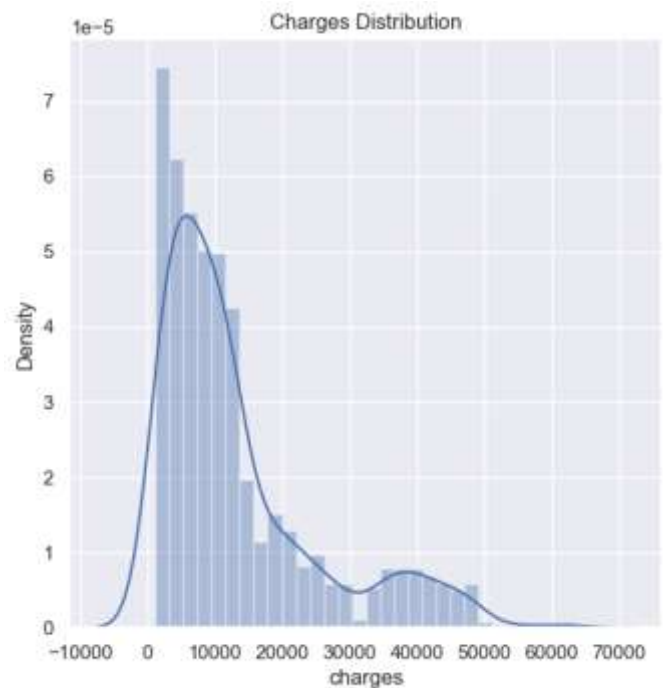
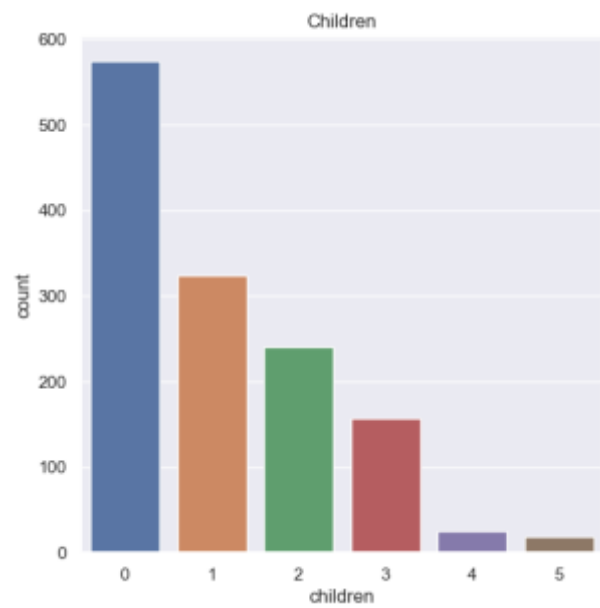
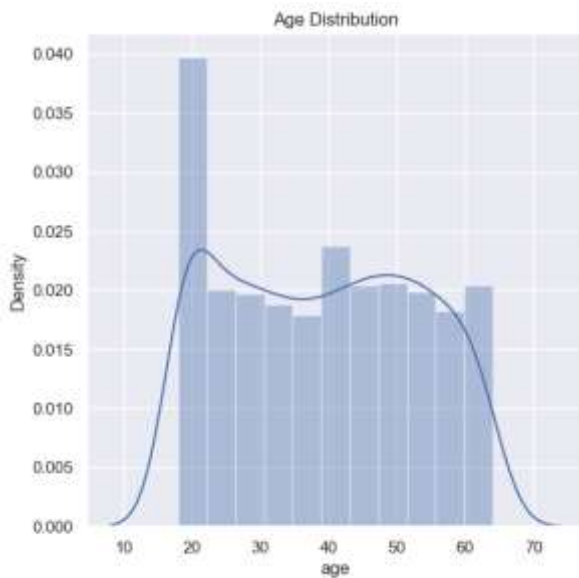
Health-Care insurance prediction using linear regression analysis: Linear regression is used to predict data. Linear regression is used for modeling the relationship between a scalar dependent variable  $y$  and one or more explanatory variables denoted  $x$ . There are advances in this field but the limitations remain the same. Simple Linear Regression is the one where only one explanatory variable is used.

**Proposed System:** Random Forest is a Supervised Learning algorithm that uses ensemble learning method for classification and regression. Random forest is a bagging technique and not a boosting technique. The trees in random forests are run in parallel. There is no interaction between these trees while building the trees. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. A



random forest is a meta-estimator (i.e. it combines the result of multiple prediction) which aggregates many decision trees, with some helpful modifications.

### SAMPLE SCREENSHOTS



### CONCLUSION

In this paper, we propose a In this project, we have successfully implemented



Regression Trees and Random Forest Regression algorithms from scratch to predict the medical prices from the input dataset. The results of Regression Trees, Random Forest Regression, Gradient Boosted Regression Trees and Linear Regression for the same dataset. From the results, we cannot conclude which model performed the best because the model performance can vary depending upon the configuration tried while testing. Hence, the model performing best for some configurations can give unsatisfactory results for some other configurations. Overall for the test configuration parameters, the order of performance of each model from the best to worst is Gradient Boosted Regression Trees, Random Forest Regression, Regression Trees and Linear Regression.

#### References

- [1] A.Tike and S.Tavarageri.(2017).A Medical Price Prediction System using Hierarchical Decision Trees. In: IEEE Big Data Conference 2017. IEEE.
- [2] “National Health Expenditures 2015 Highlights,” CMS.gov, 2015.

[Online].Available:<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/National Health Expend Data/downloads/highlights.pdf>

- [3] J. Cubanski, C. Swoop and T. Neuman,“How Much Is Enough? Out-of-Pocket Spending Among Medicare Beneficiaries: A Chartbook,” The Henry J. Kaiser Family Foundation, Menlo Park, CA, 2014. [Online]. Available: <http://files.kff.org/attachment/how-much-is-enough-out-of-pocket-spending-among-medicare-beneficiaries-a-chartbook->
- [4] J. Cubanski and T. Neuman,“The Facts on Medicare Spending and Financing,” The Henry J.Kaiser Family Foundation, Menlo Park, CA, 2017. [Online]. Available:<http://files.kff.org/attachment/Issue-Brief-The-Facts-on-Medicare-Spending-and-Financing>. [Accessed Nov. 2, 2017].



[5] The HenryJ. Kaiser Family Foundation. (2017).Total Number of Medicare Beneficiaries.[Online].Available:https://www.kff.org/medicare/state-indicator/total-medicare-beneficiaries/. [Accessed Nov. 2, 2017].

[6] Scikit-learn.org.(2018).About—scikit-learn0.19.1documentation.[online] Availableat:http://scikit-learn.org/stable/about.html#people[ Accessed23Apr.2018].

[7] R. Hafezi, J. Shahrabi, and E. Hadavandi, “A bat-neural network multi-agent system(bnnmas) for stock price prediction: Case study of dax stock price,” Applied SoftComputing,vol. 29, pp. 196–210, 2015.