ENHANCING THE ACCURACY FOR MEDICAL COST TO PREDICT THE HEALTH INSURANCE USING POLYNOMIAL REGRESSION ALGORITHM OVER RIDGE REGRESSION ALGORITHM

Vengala Rashmika[1], M. Amanullah[2]*

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The main objective of the research study is to improve the accuracy for Medical costs for Health Insurance using Polynomial Regression Algorithm compared with Ridge Regression Algorithm.

Materials and Methods: The dataset needed for the Medical cost prediction for health insurance is acquired from Google's Kaggle Website. The data set columns have the columns patient name, age, sex, bmi, smoker, children, region. In these features insurance charges are dependent variables and the remaining features are called independent variables. In regression analysis to predict the values of dependent variables. The data sets are imported and Novel Polynomial regression Algorithm and Ridge regression Algorithms are tested. The number of groups are 2 for two algorithms with the G-power value of 80%. The sample size is 10 per group.

Results: The results are acquired in the form of accuracy for the inputs provided. The IBM SPSS tool is used in order to obtain the results. From these results the author has obtained, statistical significance difference was observed between the Novel Polynomial Regression and has an accuracy of 84.07%, Ridge Regression Algorithm 75.02%, which is more accurate than the value. The independent sample T-Test was performed to find the mean, standard Deviation, standard Error Mean significance between the groups. The study has obtained a significance value of $p=0.001$ ($p<0.05$).

Conclusion: In this article, based on the results obtained, the Polynomial Regression Algorithm has more accuracy than Ridge Regression Algorithm.

Keywords: Medical Cost, Health Insurance, RMSLE, Ridge Regression algorithm, Novel Polynomial Regression algorithm, Machine learning.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and technical Science, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

Accurate prediction of healthcare costs is important for optimally managing health costs. Health care expenditures are one of the biggest expenses in Germany and optimally managing these costs has great economical importance (Miner et al. 2014; Vecchio et al. 2018). Therefore, methods for accurate patient-level prediction of future health care cost are needed to provide the basis for decision making (Rifai 2022). As medical costs reflect the development of health over time and health in turn is influenced by many factors such as social demographics by RMSLE (B, Manjulatha, and Pabboju, n.d.) previous medical history, environmental influences, genetics but also by random events such as accidents, predicting the future health is inherently challenging (A and Jenita 2020). Consequently, predicting health costs is a challenging problem.

Existing work on prediction of health cost can be divided into two categories Rule based prediction methods, in which decision rules of an algorithm to predict future costs are manually defined (K. S. Kumar, Sambath Kumar, and Rajendran 2021). The disadvantage of this approach is that it requires deep domain knowledge and that the capability of resulting models to reflect complex relations in the data is limited. Supervised learning based methods (e.g. linear regression models, random forests or support vector methods) that learn to predict future cost from the data (Jain and Chatterjee 2020). These methods have the advantage that they are not limited in their expressiveness as rule based methods. However, they typically require large datasets for training. For training of these methods, health insurance claims records are an appealing data source. They cover most of the health care expenditures of the patients and have the advantage of having sample sizes that allow fitting rich models (Panesar 2019). Additionally, they contain detailed information on patients, such as the medical history and social demographic information. It calculates the error rate using the RMSLE. The challenges of this data is that it is high dimensional, that there are many hidden interactions between variables, and that the data is often not normally distributed. The aforementioned supervised learning methods are believed to typically not leverage the potential of population scale data to detect complex patterns (Gupta and Sedamkar 2020). Recent developments in deep learning techniques (Srivastava et al. 2020), such 2 as novel deep neural network architectures and numerical approaches to fit the networks, promise to address some of these challenges (Srivastava et al. 2020; Y.-W. Chen and Jain 2019)(Gupta and Sedamkar 2020)(Srivastava et al. 2020; Y.-W. Chen and Jain 2019). Machine learning has been successfully applied in the medical domain to tasks such as prediction of various clinical outcomes from electronic health records, or the detection from retinal fundus photographs, showing the potential of this technology. Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; P. G. Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

The problem in the existing research of Medical cost prediction for Health Insurance is less accuracy. There are certain algorithms with more accuracy when comparing it with existing ones. The main aim of the study is to improve the accuracy of Health insurance by implementing a Polynomial Algorithm.

2. Materials and Methods

The proposed work is done in the Machine Learning lab, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The number of groups is 2 for two algorithms. The sample size is 10 per group in total 20.

The dataset named ‘Medical Insurance cost’ ((Srivastava et al. 2020; Y.-W. Chen and Jain 2019)) is downloaded from Google Kaggle Website. The data predicted the log-transformed actual values by RMSLE. The data in this dataset explains about the Health Insurance provided for each patient varying with the factors (Takeshima et al. 2018). Many factors that affect how much you pay for health insurance are not within your control. Health insurance is calculated based on the patient's expenses (Severino and Peng 2021). In this dataset, information about the patients and the analytics about the patients with different factors.

The Health Insurance is an important eye-opener for emergency needs during accidents and disease pandemic situations (H. Chen et al. 2021). Many of the people will lag to hit financially and to bear the operational censoring expenses during treatment. The need for health insurance changes from youth to old age depending on your lifestyle and genetics.

Polynomial Regression

Polynomial Regression is a Machine learning based algorithm (Srinivas, Sucharitha, and Matta 2021) which is also a form of Linear regression model. Can improve our model by features of the prediction, specifically, by making new features that capture the interactions between existing features. This is called Polynomial regression

(Larson and Halfon 2010). The idea is to generate a new feature matrix consisting of all polynomial combinations of the features with degree less than or equal to the specified degree.

Step 1: Import the Packages required.
Step 2: Import the dataset into the code environment
Step 3: Assign the features of dataset such as Age, Sex, Bmi, Children, Smoking, Region. The next process is checking the data for correction. After the corrections store the data into dataframes. Since predicting the insurance costs, charges will be our target feature.
Step 4: Once after importing the data, processes such as encoding are to be performed. The dataset should be chosen and start pre-processing the input so that the model can be used.
Step 5: In these Polynomial and Ridge regression algorithms the person uses the Backward Elimination method in order to work their way down.
Step 6: The determination of required parameters are done so that the model is good to fit. The parameters taken are predicted and performed.
Step 7: Further analysis is performed and the measurement of accuracy is done successfully.

Ridge Regression

Ridge regression (RR) is another particular case of linear regression model that helps shrink the coefficients and reduce the model's complexity (Larson and Halfon 2010; Mahlich and Sruamsiri 2019). It also helps in reducing multicollinearity. Unlike the LR model, the RR model does not provide absolute shrinkage of the coefficient (Drewe-Boss et al. 2022). However, the RR model makes some of the coefficient values very low or close to zero. Therefore, the features which are not contributing much to the model will have very low coefficients. As a result, the RR model helps in reducing overfitting, which appears from the MLR model.

Step 1 : Import the packages required.
Step 2 : Import the dataset taken from kaggle into the code environment.
Step 3 : Assign the features of dataset such as Age, Sex, Bmi, Children, Smoking, Region. The next process is checking the data for correction. After the corrections store the data into dataframes. Since predicting the insurance costs, charges will be our target feature.
Step 4 : Once after importing the data, processes such as encoding are to be performed. The dataset should be chosen and start pre-processing the input so that the model can be used .
Step 5 : Import the Ridge regression algorithm dataset from the kaggle and predict the output for the testing datasets.
Step 6 : The determination of required parameters are done so that the model is good to fit. The parameters taken are predicted and performed.
Step 7 : Further analysis is performed and the measurement of accuracy is done successfully.

For Polynomial Regression Algorithm, the test size is 49% of the total dataset and the remaining of 51% is used for training the datasets. Accuracy of both the algorithms are tested from sample sizes of 45 to 95. The dataset used for this article on Machine Learning based Algorithms are obtained from Google's official dataset website Kaggle.

Statistical Analysis

The statistical software used for performing analysis is IBM SPSS version 21.0. IBM SPSS is a statistical software tool used for the analysis of data. The datasets are normalized and then the data is converted into arrays. The number of clusters needed are visualized and analyzed and the existing algorithms are obtained in accountability of RMSLE. For the Novel Polynomial Regression algorithm, it is observed that if the number of censoring iterations increased, then the error rate decreased and accuracy increased. It is declared that the Polynomial Regression Algorithm shows higher value compared with the Ridge Regression Algorithm.

3. Results

In Table 1, a file collection of people with the charges obtained and the details of the people is given. The dataset is taken from Kaggle and it contains Age, Sex, BMI, Children, Smoking, Region and Charges obtained by this bases using Machine Learning. The age column is the respective number of each patient with different ages, bmi of the patients and the smoking cause for charges. These the bmi and the charges are represented in numerical values which are taken from the dataset collected on the first format. The statistical comparison of the charges with respect to the region and the smoking cause using two sample groups was done through SPSS version 21. Analysis was done for mean, standard deviation, independent T-Test.

The Outcome of the Novel Polynomial Regression and Ridge Regression algorithms which are predicted values are compared to the values and these outcomes are shown as tables and bar graphs.

In Table 2, statistical results of Polynomial Regression and Ridg Regression algorithms.

Mean accuracy value, standard deviation and standard error mean for PRA and RRA algorithms are obtained for 10 iterations. It is observed that the PRA (84.07%) algorithm performed better than the RRA (75.02%) algorithm.

In Table 3, independent sample evaluation of the P Regression Algorithm value
and Ridge Regression algorithm results with significant two-tailed values (p = <. 001). So both, the PRA and RRA algorithms have a value of less than 0.05 with a confidence interval of 95%.
In Figure 1, a graph is shown with respective medical charges in the smoking category. They are plotted with the accuracy of smoking and non-smoking .
In Figure 2, the graph is shown with respective medical charges per region. Here consider 4 regions such as Northeast, Northwest, Southeast and Southwest. It is plotted with the accuracies obtained from different regions. Also it has been noticed that the sex and region dont have noticeble differences for each category terms of charges given. Error rate given by RMSLE was 39. Can see that there is an increasing trend in the charges as the number of children increases. Lastly, smokers seem to make a significant change to charges given by Health Insurance.
In Figure 3, The bar chart plotted with the accuracies of both the algorithms for different sample sizes is represented. The bar chart is plotted by taking algorithms as x-axis and accuracy as y-axis. From the bar chart, it can be seen that the Novel Polynomial Regression Algorithm is more accurate than the Ridge Regression Algorithm. The last row shows the average of accountability of accuracy of both the algorithms. At sample size 10, the average accuracy of Polynomial Regression Algorithm is 84.07% and Ridge Regression Algorithm is 75.02%.

4. Discussion

The results of the study shows that the Novel Polynomial Regression Algorithm performs better than the Ridge Regression Algorithm. Polynomial Regression has obtained an accuracy of 84.07% compared to Ridge Regression which has an accuracy of 75.02%.
The Ridge regression (RR) model showed an RMSLE of 39 and an R-square score of 0.37 for the training data (Zhang, Ding, and Peng 2021). However, it showed an MSE of 38.49, and the R-square score was 0.3711 for the test data. Since these results were also far from ideal, RR model performance was also low (Cai and Zaslavsky 2019). Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors (Cai and Zaslavsky 2019; Erim et al. 2021). This means the RMSE should be more useful than MAE when large errors are particularly undesirable.
Knowing the metrics and accountability, it validates the performance of our model simply by applying new data insurance.csv to it and checks the metric score. K-fold cross validation is not done since the data is small. MLR, RR could not fit this hospital data properly and performed poorly on the data. No multicollinearity found in the Ridge Regression model, but many found in the Polynomial Regression model. This makes sense since some features in Novel Polynomial Regression were created by multiplying two features from the Ridge Regression model.

5. Conclusion

In this article, the results obtained in executing several Algorithms based on the various data samples using the Polynomial Regression Algorithm (84.07%) and Ridge Regression Algorithm (75.02%) are presented. The Polynomial Regression Algorithm was used to test the accuracy of medical charges for Health Insurance and was shown to be more accurate than the Ridge Regression Algorithm. It is recommended that future work involve more variables in the given dataset to build a more accurate model that could predict hospital cost insurance more accurately.

Declarations

Conflicts of Interests

No conflict of interest in this manuscript.

Authors Contributions

Author VR was involved in data collection, data analysis, and manuscript writing. Author SS was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgement

We would like to thank our management, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing facilities and opportunities for our research study.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.
1. The Big Event
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

A, Jenita Mary, and Mary A. Jenita. 2020. "Empirical Review on Fraudulence Detection in Health Care Insurance Industries Using Machine Learning Techniques." *Journal of*

Advanced Research in Dynamical and Control Systems. https://doi.org/10.5373/jardcs/v12i7/20202042.

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

B, Manjulatha, B. Manjulatha, and Suresh Pabboju. n.d. "A Survey on Health Care Data Using Machine Learning Algorithms." https://doi.org/10.21203/rs.3.rs-16736/v1.

Cai, Tianyi, and Alan M. Zaslavsky. 2019. "Bayesian Hierarchical Modeling of Substate Area Estimates from the Medicare CAHPS Survey." *Statistics in Medicine* 38 (9): 1662–77.

Chen, Hua, Jianing Xing, Xiaoxu Yang, and Kai Zhan. 2021. "Heterogeneous Effects of Health Insurance on Rural Children's Health in China: A Causal Machine Learning Approach." *International Journal of Environmental Research and Public Health* 18 (18). https://doi.org/10.3390/ijerph18189616.

Chen, Yen-Wei, and Lakhmi C. Jain. 2019. *Deep Learning in Healthcare: Paradigms and Applications*. Springer Nature.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Drewe-Boss, Philipp, Dirk Enders, Jochen Walker, and Uwe Ohler. 2022. "Deep Learning for Prediction of Population Health Costs." *BMC Medical Informatics and Decision Making* 22 (1): 32.

Erim, Daniel O., Antonia V. Bennett, Bradley N. Gaynes, Ram Sankar Basak, Deborah Usinger, and Ronald C. Chen. 2021. "Mapping the Memorial Anxiety Scale for Prostate Cancer to the SF-6D." *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation* 30 (10): 2919–28.

Gupta, Shiwani, and R. R. Sedamkar. 2020. "Machine Learning for Healthcare: Introduction." *Learning and Analytics in Intelligent Systems*. https://doi.org/10.1007/978-3-030-40850-3_1.

Jain, Vishal, and Jyotir Moy Chatterjee. 2020. *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*. Springer Nature.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, K. Sambath, K. Sambath Kumar, and A. Rajendran. 2021. "Machine Learning Classifiers in Health Care." *Machine Learning and Analytics in Healthcare Systems*. https://doi.org/10.1201/9781003185246-5.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Larson, Kandyce, and Neal Halfon. 2010. "Family Income Gradients in the Health and Health Care Access of US Children." *Maternal and Child Health Journal* 14 (3): 332–42.

Mahlich, Jörg, and Rosarin Sruamsiri. 2019. "Co-Insurance and Health Care Utilization in Japanese Patients with Rheumatoid Arthritis: A Discontinuity Regression Approach." *International Journal for Equity in Health*. https://doi.org/10.1186/s12939-019-0920-7.

Miner, Linda, Pat Bolding, Joseph Hilbe, Mitchell Goldstein, Thomas Hill, Robert Nisbet, Nephi Walton, and Gary Miner. 2014. *Practical Predictive Analytics and Decisioning Systems for Medicine: Informatics Accuracy and Cost-Effectiveness for Healthcare Administration and Delivery Including Medical Research*. Academic Press.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency

Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Panesar, Arjun. 2019. *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Apress.

Rifai, Nader. 2022. *Tietz Textbook of Laboratory Medicine - E-Book*. Elsevier Health Sciences.

Severino, Matheus Kempa, and Yaohao Peng. 2021. "Machine Learning Algorithms for Fraud Prediction in Property Insurance: Empirical Evidence Using Real-World Microdata." *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2021.100074.

Srinivas, Mettu, G. Sucharitha, and Anjanna Matta. 2021. *Machine Learning Algorithms and Applications*. John Wiley & Sons.

Srivastava, Rajshree, Pradeep Kumar Mallick, Siddharth Swarup Rautaray, and Manjusha Pandey. 2020. *Computational Intelligence for Machine Learning and Healthcare Informatics*. Walter de Gruyter GmbH & Co KG.

Takeshima, T., S. Keino, R. Aoki, T. Matsui, and K. Iwasaki. 2018. "Development of Medical Cost Prediction Model Based on Statistical Machine Learning Using Health Insurance Claims Data." *Value in Health*. https://doi.org/10.1016/j.jval.2018.07.738.

Vecchio, Fabrizio, Francesca Miraglia, Francesco Iberite, Giordano Lacidogna, Valeria Guglielmi, Camillo Marra, Patrizio Pasqualetti, Francesco Danilo Tiziano, and Paolo Maria Rossini. 2018. "Sustainable Method for Alzheimer Dementia Prediction in Mild Cognitive Impairment: Electroencephalographic Connectivity and Graph Theory Combined with Apolipoprotein E." *Annals of Neurology* 84 (2): 302–14.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Zhang, Chengcheng, Yujia Ding, and Qidi Peng. 2021. "Who Determines United States Healthcare out-of-Pocket Costs? Factor Ranking and Selection Using Ensemble Learning." *Health Information Science and Systems* 9 (1): 22.

Tables and Figures

Table 1. Represents the File containing details about the patients with factors as age, sex, bmi, children, smoker, region and the charges based on it.

S.No	Age	Sex		Bmi	Children	Smoker	Region	Charges
0	19	female		27.900	0	yes	southwest	16884.92400
1	18	male		33.770	1	no	southeast	1725.55230

2	28	male		33.000	3	no	southeast	4449.46200
3	33	male		22.705	0	no	northwest	21984.47061
4	32	male		28.880	0	no	northwest	3866.85520

Table 2. Statistical results of Polynomial Regression and Ridge Regression algorithms. Mean accuracy value, standard deviation and standard error mean for PRA and RRA algorithms are obtained for 10 iterations. It is observed that the PRA (84.07%) algorithm performed better than the RRA (75.02%) algorithm.

Algorithms (Accuracy)	**Sample (N)**	**Mean**	**Std Deviation**	**Std Error Mean**
Polynomial Regression	10	84.0700	3.08951	.97699
Ridge Regression	10	75.0170	3.07301	.97177

Table 3. Independent sample t-test of the significance level Polynomial Regression Algorithm and Ridge Regression algorithm results with two tailed significant values (p=<.001). Therefore both the PRA and RRA algorithms have a significance level less than 0.05 with a 95% confidence interval.

Accuracy	**Levene's Test for Equality of Variances**		**T-test of Equality of Means**					**95% of the confidence interval of the Difference**	
	F	**Sig.**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std Error Difference**	**Lower**	**Upper**
Equal Variance Assumed	.004	.953	6.570	18	<.001	9.05300	1.37798	6.15796	11.94804
Equal Variance Not Assumed			6.570	17.999	<.001	9.05300	1.37798	6.15796	11.94804

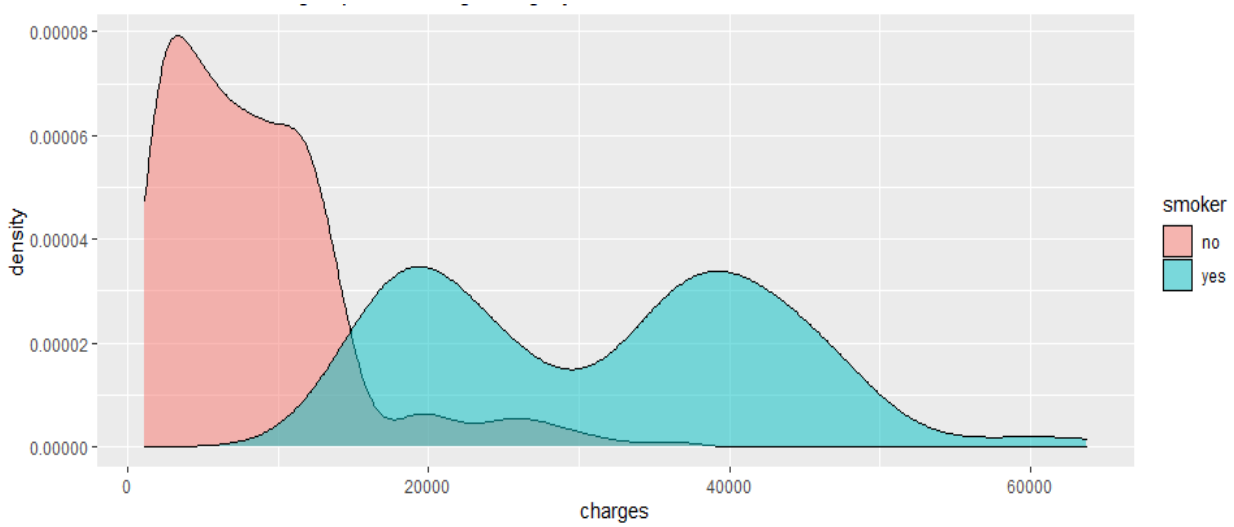

Fig. 1. Displaying the graphs varying the charges in smoking category

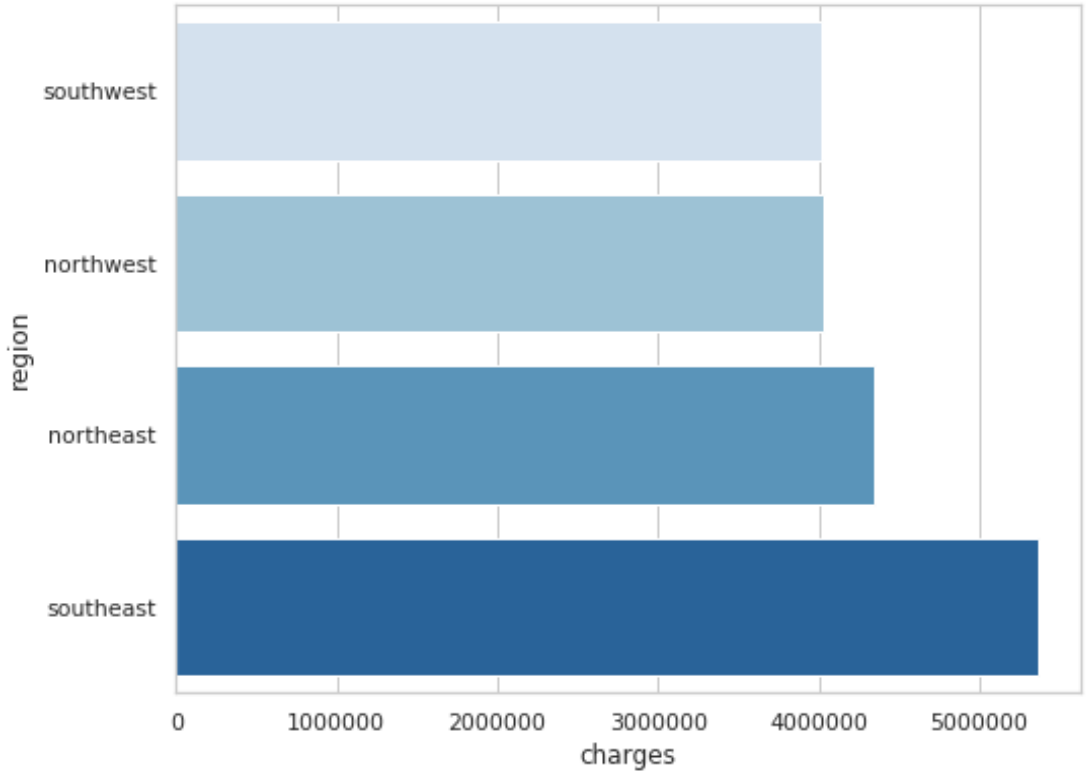

Fig. 2. Displaying the medical charges varying with different regions

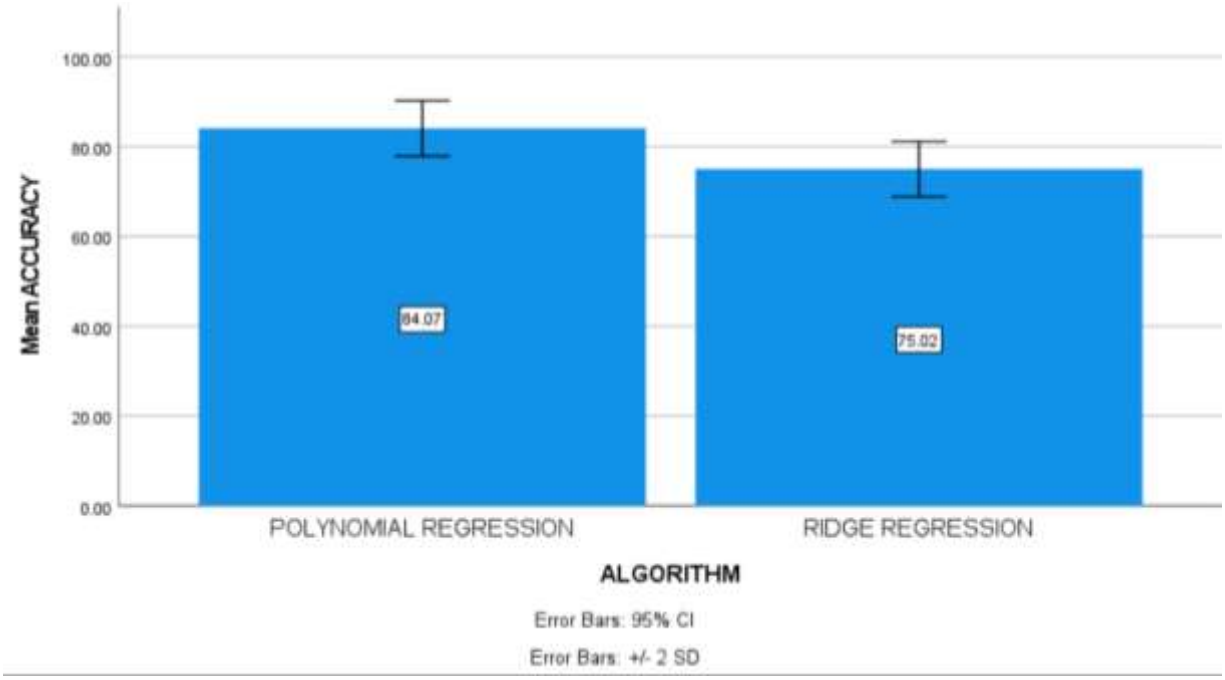

Fig. 3. Comparison of Polynomial Regression algorithm and Ridge Regression Algorithm in termsof mean accuracy. The accuracy of PRA is better than RRA and the standard deviation of PRA isslightly better than the RRA algorithm. X-axis: (GROUPS) PRA vs RRA algorithm and Y - axis :Mean accuracy of prediction ±2 SD