



Machine Learning Implementation for Health insurance

Sujith Thota¹, Kotha Vishnu Sai², P Swarnalatha³

Undergraduate student, Vellore Institute of Technology, India, thota.sujith2018@vitstudent.ac.in

Undergraduate student, Vellore Institute of Technology, India, vishnusai.kotha2018@vitstudent.ac.in

Faculty of Computer Science and Engineering, Vellore Institute of Technology, pswarnalatha@vit.ac.in

ABSTRACT

Health insurance is considered as one of the most important things to possess in recent times. Medical treatment expenses are ever growing even though technology is advanced. Many people are not so wise while choosing the health insurance plans. Our aim is to design an interface which helps the people to be more benefited. Our project helps in predicting the amount that can be claimed by an individual based on their basic information as parameters provided such as BMI, Number of children, smoking habit etc. so based on these attributes a machine learning model which we made will be predicting. This will be beneficial to both the customer for their clarification on charges to be claimed and for organization in developing new schemes and plans.

Key words : Health insurance, Prediction, Machine learning, Middle ware

1. INTRODUCTION

“Good health is not something we can buy. However, it can be an extremely valuable savings account” by Anne Wilson Schaefer. During this pandemic, being healthy has become a challenging task for any individual. no matter how much care you take there will be some inevitable circumstances which may lead to seeking treatment from the hospitals. Health insurance has become a life saving asset which helps in reducing the treatment expenses to a great extent. Every organization assesses many parameters while calculating the amount to be claimed to a client. A loss may occur when a client purchases a plan which may be claimed a less amount than their eligibility. This project provides a basic insight regarding the amount that can be claimed on the basis of simple parameters provided by the client. Machine learning is one of the advanced technologies which has a potential in finding patterns and developing functions based on the given information. Applications of machine learning are vast. Machine learning can act as a backend technology which helps in developing many interfaces. We developed a web interface for our machine learning paradigm. The basic information given by the client will be fed to the machine learning model which gives the amount to be claimed based on the data it trained.

This project is developed using HTML, CSS, BOOTSTRAP in the front end and PHP, MySQL, python in the backend. The purpose of this website is to provide complete

information about the health insurance policies of a company. This website will have three pages: Home, Service, Predictor.

Home page is the first page the user will see when they enter the website. This page consists about the details of the insurance policies the company can provide. The importance of insurance in life. This page was developed using HTML, CSS, Bootstrap. Then comes the service page, This page shows the services provided by the company. This website supports two types of services. Requesting a call from the company, to get the latest updates of the company by subscription. Both of them would be completed by taking user data using a form. We store the data in the SQL server in the “customers” dataset using PHP in the backend. The user will enter their name, email and mobile number. The admin will see the details and make further proceedings. The last page is the predictor page, This page helps the company to increase their market. This page shows the insurance charges by taking basic information from the user like age, BMI, number of children, gender, smoking habit, and living area. The charge will be predicted based on the inserted values. These values will be inserted into a machine-learning model created using python language by a predefined dataset. The values will be given as input and the output will be displayed on the page. PHP is used to collect the values and to make a connection between model and website. As the website is a promotional website, this predictor page helps the sales to boost up.

2. LITERATURE REVIEW

There are significant variations in the attitudes and responses of the customers towards the insurance around the globe. Primarily, the online mode of buying is another challenge for the insurance organization. The main aim is to make the customers buy online insurance with confidence. this can be done by providing awareness among people about efficient usage of online services in a secure way and make people know the benefits of this kind of practices which saves time, makes the process comfortable and secure. The practice should bring into usage first and let people know the things by themselves that there is no threat and the benefits of the practice[1].As in the near future the time constraint will be a serious problem to deal with in one's life the online mode will fulfill this problem [1].Websites at the backend can be embedded with many

features. Cryptojacking is one of the methods where the website owners get benefit with the one who visits websites and can indulge their cpu power in crypto mining of others without knowing.

There are significant variations in the attitudes and responses of the customers towards the insurance around the globe. Primarily, the online mode of buying is another challenge for the insurance organization. The main aim is to make the customers buy online insurance with confidence this can be done by providing awareness among people about efficient usage of online services in a secure way and make people know the benefits of this kind of practices which saves time, makes the process comfortable and secure. The practice should bring into usage first and let people know the things by themselves that there is no threat and the benefits of the practice[1].As in the near future the time constraint will be a serious problem to deal with in one's life the online mode will fulfill this problem [1].Websites at the backend can be embedded with many features. Cryptojacking is one of the method where the website owners get benefit with the one who visits websites can indulge their cpu power in crypto mining of others without knowing. We can use the machine learning the same way as a backend function which can be used to obtain solutions to various problems. The machine learning model is trained, tested and saved. Then the model is embedded into the website. Flask can be used to embed the model into the website[2][3].

Life Insurance Risk Prediction using Machine Learning - Specialists have analysed a life Insurance Risk Prediction using Machine Learning - Specialists have applied AI procedures to perform predictive and automate the life insurance application evaluation process. The fundamental idea here is that rating insurance applications is a supervised learning problem. Here, an application is treated as a data point, the information segments given by a candidate are the highlights of this information point and the risk rating of the candidate is the yield that we are attempting to predict[4].This study conveys the approach of using machine learning for the identification of images and to classify them to respective disease or illness. It is the comparative study of machine learning models based on the samples of images. models such as decision tree, SVM, ANN and Naive Bayes[5]. Now-a-days data is playing a major role and is carrying big assets in the insurance industry. In today's journey insurance industry has a vital role. Insurance transporters have access to more information than ever before. From the past 700+ years in the insurance industry we can consider the three major eras Starting from 15th century to 1960,industry followed the manual era, from 1960s to 2000 we are in the systems era, now we are in digital era i.e. 2001-2020.The aim of applying data science analytics in the insurance is the same as other industries , to find the optimal strategies to get solutions to a given problem. the performance of various algorithms should be tested with the data given and compare the performances respectively and analyse which model suits best for your data[6].Here analysis is done taking many algorithms into play for a dataset[6]. Ensemble methods of

machine learning aim at improving the predictive analysis and the problem of overfitting for a given data. The main principle of ensemble methods is to construct a combination of models instead using one model alone. thereby increasing performance. For very large data these ensemble methods work very fine, so this can be used taking this point into consideration[7].

3. METHODOLOGY

Our aim is also to make an interface for the trained machine learning model which can be accessed by a large proportion of the population and the interface should be interactive. So we have chosen the web interface. One of the major constraints here is to develop the machine learning model with best accuracy. As the target variable for this project is a numerical value we have chosen regression algorithms such as regression, xgboostregressor ,decision tree, svr , random forest and a stacked model of all the four algorithms except svr.We have used 70% of data as a training set and 30% as a testing set.

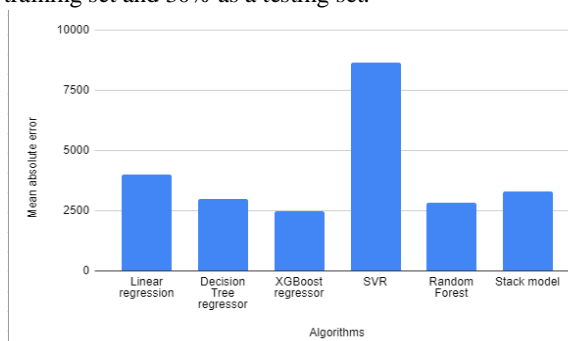


Figure 1: Comparison of mean squared errors of different machine learning algorithms.

According to figure 1, out of all the algorithms we chose for comparative study xgboost and decision tree has shown less “mean absolute error”. Based on this metric we have taken out these 2 algorithms and we tuned the hyperparameters for those algorithms to improve the performance.xgboost algorithm has been tuned using random search cv to find the best parameters and decision tree algorithm is tuned using grid search cv. Tuning of hyperparameters helps us to find the best version of the given algorithm.

After tuning we have witnessed a decrease in mean absolute error and the best model we have confirmed is the decision tree after the hyperparameter tuning.Now we have saved that model to our local PC using pickle module in python. Interface is developed with three pages Home, Services and Predictor. We have used PHP as the middleware. Using PHP we connected the saving model and the predictor page. In the services page we have given two options for the customer to get an assist from the management. Basic details have been asked from the viewer which then will be stored in the database. The passage of this data will be done using the PHP as well. SQL has been used to manipulate the data in the database.

4. WORKING

Once the visitor/customer opens this web interface the first page will be the homepage. Then the visitor will have two options: service and predictor. In the service page again there

“subscribe”. Both of these options will be asking basic contact details.

The Predictor page is the main component of this interface. This page asks the visitor to enter the basic details such as age, BMI, number of children, region of living, gender and smoking habit. These details will be then sent to the machine learning model with the post method through PHP.

The predicted amount then will be displayed on the same page. Both of these pages have the validation which ensures the data entered by the visitor will be legit. The following image figure 2 will show a basic flow chart of the project. In the flow chart we can witness what are the user choices he can make.

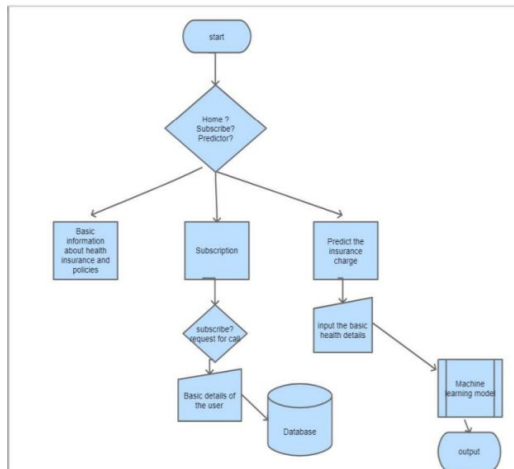


Figure 2: Flow diagram of proposed methodology

As we see in figure 2 the flow of data from one module to another module we can see the data flow. First the data is being filled in the HTML form from the visitor with the post method this data will be then handled in the middleware by the php in the both pages.

For the service page it will be directly stored to the database. But for the predictor page it will be taken by the machine learning model.

The following figures (figure 3 and figure 4) will show the flow of data briefly.

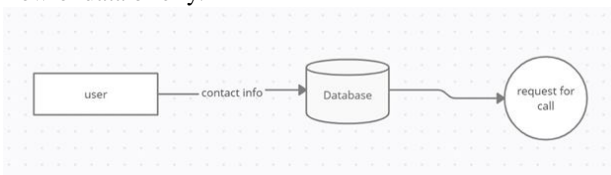


Figure 3: Data flow diagram of submission page

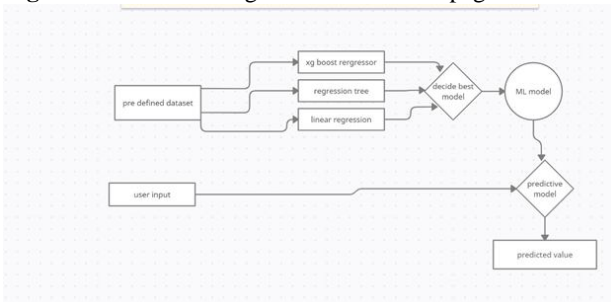


Figure 4: Data flow diagram of predictor page

5.CONCLUSION

As said by the officials, data will become new oil to the world. In this project we tried to incorporate two technologies i.e., web and machine learning, we have succeeded in embedding them. Various machine learning models have been trained with different algorithms which helped in finding the best algorithm. This project can be taken as a prototype and can be used by the organization's own data while training the machine learning model. We have used so many scripting and programming languages to get this project in shape.

6.FUTURE WORK

This project can be further improved by using advanced javascript scripting languages such as react and node instead of HTML,CSS ,Javascript and PHP. This helps in providing many options while developing the web interface for the machine learning model. As of now the data which we used is a predefined data. We can collect the data real time which helps in constantly updating data and adding new observations to the data. Machine learning algorithms can also be developed and modified to be more precise while training with the data. As of now this project is hosted in the local server but this can be hosted in any of the cloud platforms. This helps in the accessibility o

7.ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our supervisor Prof.Swamalatha, Senior Lecturer of the Department of Computer Science and Technology, Vellore Institute of Technology University of India.

REFERENCES

1. Prof. M. Devmane, Sujitkumar Jagtap “**Online Insurance Management System** ”International Journal of Engineering Research & Technology (IJERT)
2. ChaminNalinda ”**Embedding Machine Learning Models to Web Apps** “ towards data science
3. Grinberg, Miguel. *Flask web development: developing web applications with python.*" O'Reilly Media, Inc.", 2018.
4. Bharathsethuramansharman **Life Insurance Risk Prediction using Machine Learning Algorithms- Part I: Data Preprocessing and Dimensionality Reduction** towards data science.
5. Arasi, Munya A., and SangitaBabu. "**Survey of machine learning techniques in medical imaging.**" International Journal of Advanced Trends in Computer Science and Engineering 8.5 (2019): 2107-2116.
6. Burri, Rama Devi, et al. "**Insurance claim analysis using machine learning algorithms.**" *International Journal of Advanced Science and Technology* 127.1 (2019): 147-155.
7. Bühlmann, Peter. "**Bagging, boosting and ensemble methods.**" *Handbook of computational statistics.* Springer, Berlin, Heidelberg, 2012. 985-1022.