IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Secured Medical Records Storage and Insurance Cost Prediction System

Renu Kachhoria[1], Hitesh Kothawade[2] Amardipsinh Girase[3], Purushottam Jadhav[4], Abhijit Nathe[5]
Professor, Department of Computer Engineering[1]
UG Students, Department of Computer Engineering[2,3,4,5]
Pimpri Chinchwad College of Engineering, Nigdi, Pune

Abstract: *Medical data is very important these days, not only for doctors and medical organizations but also for patients. Every time we visit a doctor we describe our medical situation to them. However, if we consult multiple doctors about our medical condition we have to delineate our entire medical condition to them individually. Also sometimes medical files may get misplaced. We need secure storage for our medical records. Today we have digital storage for everything from educational documents to money, then why not the same for medical data, so that it can be easily available for us whenever needed? This paper focuses on developing a system to provide storage and security to medical records and also develop insurance cost prediction models for users. The Electronic Health Record (EHR) is used to store patient-centric information which is secure and real-time. Health Records will consist of text data as well as images from different pathologies, ophthalmologists, etc. Our proposed system will also predict the insurance cost by using various supervised algorithms such as Stochastic Gradient Boosting, XGBoost and Random Forest Regressor algorithms which give accurate results. These machine learning methods are used to show how different regression models are used in forecasting insurance costs. According to World Health Organization (WHO) data, the global cost of healthcare in 2016 was estimated to be $7.5 trillion USD, or 10% of the global GDP [5]. Understanding the value of health insurance is crucial, and estimating the cost of insurance can persuade customers to buy the right coverage[10]. The dataset we utilized for the Insurance cost prediction model is "Analysis and prediction of health insurance cost*

Keywords: Electronic Health Record Security, ETH storage, Insurance cost prediction, Stochastic gradient boosting

I. INTRODUCTION

People's health care cost prediction is nowadays a valuable tool to improve accountability in health care [1]. Today we see healthcare workers and doctors finding it difficult to go through long lines of patients and see them. We find doctors struggling a lot when there is an overflow of emergency care situations.

There are various Medical records management tools available in the market, some of which are free and some expect you to pay a reasonable amount in order to avail their services for storage. MRM is arranging and handling medical health information of patients so that it is convenient for medical professionals, doctors and patients itself to access this information and also carry out advancements related to it. These systems either provide you prediction on your health information or it provides you with secure storage for your information. But then also there are trust issues which also need to be considered. A person's health history can tell a lot about a person, more than he ever can tell about himself. This record history can have multiple uses. This health history can be used to predict how he or she will react or behave to a particular treatment, so health records are very important in predicting such things.

This tells us that there is a need for a system which on implementation will surely fasten this process and also provides safe storage and access to patient information [3]. One of the main reasons why the healthcare industry has a higher risk of data breaches compared to other industries is the type of data collected and stored. Healthcare providers can have highly detailed records of their patients, including names, dates of birth, addresses, social security numbers, payment account information, and more. The solution to these problems could be the use of new technologies such as blockchain and cloud. We are using security in two fields user login and security of health records.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

We are trying to implement a system where one will have the privilege to securely save their medical documents and files, where they can predict insurance costs applicable to them. Furthermore, if one knows probable future expenses earlier ahead of time can guide patients to select right insurance plans with appropriate premiums and co-payment associated with insurance schemes [4]. For the prediction of insurance cost we are using a supervised learning algorithm, which is based on a regression technique, called stochastic gradient boosting.

II. LITERATURE SURVEY

For the literature survey we went through some of the previously published papers regarding Health data storage, its security and also papers with the prediction of Insurance Cost. The research and findings from these papers are listed below along with their author name(s).

A. Hassan, J. Iqbal, S. Hussain, H. AlSalman , M. Mosleh and S. Ullah [2] presented that Stochastic Gradient Algorithm gives the highest accuracy in predicting the costs of health insurance. Here, the regression algorithm is used on a dataset which follows the steps of preprocessing, feature engineering, data splitting, regression, and evaluation. The resultant outcome revealed that Stochastic Gradient Boosting (SGB) achieved a high accuracy of 86% with an RMSE of 0.340.

M. Morid, K. Kawamoto, T. Ault, J. Dorius and S. Abdelrahman [11], In this paper, has stated that for high-cost individuals, Artificial Neural Networks which have not been previously reported for use in healthcare cost prediction had the highest performance. Here, ANN is used for the classification of Demographic information and the number of laboratory tests. The Ridge model is used for comparison with ANN to have distinguishable performance compared to ANN.

B. Pandey, B. Baloian, J. Pino, S. Peñafiel, H. Sanson and N. Bersano [25] used this paper to generate a time-series using the EVREG model, and the degradation trend is forecasted using condition monitoring data from a methane compressor. In order to speed up calculations, the method also makes use of the nearest k-neighbor algorithm. This Weighted Evidential Regression (WEVREG) Model's objective is to improve prediction accuracy and give the model the ability to carry out feature selection tasks.

J. Narvaez, M. Guillen and M. Alcañiz [9] focuses on the fundamental principle of a logistic regression model, which holds that there must exist a linear combination of risk factors that is connected to the likelihood of observing an event. An alternate technique for forecasting a response variable in the presence of specific covariates is XGBoost. This algorithm's basic concept is that it builds D classification and regression trees (or CARTs) one at a time, allowing each new model to be trained using the residuals of the preceding one.

M. Devi, P. Swathi, M. Reddy, V. Verma, A. Reddy, S, Vibekannansa and P. Moorthy [14], In this paper, the study applies linear and ensembling regression models before and after feature scaling, as well as the Anova test, to investigate the prediction of health cost insurance. According to experimental findings, polynomial regression achieves 88% R2 Score both before and after feature scaling. Before and after feature scaling, the Random Forest regression is attaining an R2 Score of 86%.

I. Duncan, M. Loginov & M. Ludkovski [6], published a research paper regarding Testing Alternative Regression Frameworks for Predictive Modeling of Health Care Costs. The study estimates the predictive modeling of future healthcare costs using a range of statistical methodologies. Their study was based on a dataset of 30,000 insured people having claims information from two consecutive years. The dataset contains more than 100 covariates for each insured person, including a detailed breakdown of earlier costs and causes tracked using coexisting disease markers. They give statistical models for the relationship between next-year expenses and medical and cost information in order to estimate the mean and quantiles of future costs, rate risks, and identify the best predictive variables. In addition to the traditional linear regression model that underpins risk adjusters, there is a comparison of numerous models, including the Lasso GLM, multivariate adaptive regression splines, random forests, decision trees, and boosted trees. This essay demonstrates the shortcomings of linear models and suggests that alternative models should be evaluated right away as the frequency of premium transfers among insurers increases.

J. Pesantez-Narvaez , M. Guillén and M. Alcañiz [7], this research paper recognises XGBoost as an algorithm with exceptional predictive potential. Models with a binary answer showing the presence of accident claims vs. no claims can be used to identify the causes of traffic accidents. This study looked at how well the XGBoost and logistic regression

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

methods predicted the existence of accident claims using telematics data. The use of XGBoost has proven to be an excellent method for creating confusion matrices, or tables in which observations and forecasts are compared, with very few false positives and false negatives, and incredibly successful for fitting real binary response data. The predictive performance of the XGBoost (tree booster) was significantly better than the logistic regression in the training sample but noticeably worse in the testing sample when a logistic regression and XGBoost compete to predict the occurrence of accident claims without model-tuning procedures.

Findings showed that the traditional logistic regression model can forecast accident claims using telematics data and that the coefficient estimates are simple to comprehend. Moreover,Considering that just two coefficients were significant at the 90% confidence level, the technique offered a comparatively strong predictive performance. The XGBoost approach does not yield superior outcomes.

M. Hanafy [8] from Assiut University, studies in his research paper about prediction of Health Insurance Cost using machine learning algorithms and DNN Regression models, demonstrates how several regression models may estimate insurance costs and assess the models' results. The study employs a variety of machine learning regression models and deep neural networks to anticipate health insurance costs based on particular characteristics using a set of individual medical cost data from Kaggle.com. Multiple linear regression models are used in this study because the dependent (target) variable is determined by a large number of independent factors. This study makes use of the health insurance costs dataset. The most effective method concluded in this study is stochastic gradient boosting, which has an accuracy of 85.82, an RMSE of 0.380189, and an MAE of 0.17448. Therefore, stochastic gradient boosting outperforms alternative regression models in the estimate of insurance costs. The inclusion of a novel method of assessing insurance costs is the main motivation behind this work. The best model, Stochastic Gradient Boosting, came in second place, followed by XGBoost, and the poorest model was the K-Nearest Neighbours algorithm.

Alexandra Dmitrienko,Christopher Liebchen,Christian Rossow,Ahmad-Reza Sadeghi published a research paper that explains how popular ISPs like Facebook, Twitter, Dropbox, Google, and Dropbox have implemented 2FA. As it continues, it provides a more comprehensive attack against mobile 2FA techniques. We looked at the security of mobile two-factor authentication (2FA) systems, which have recently drawn a lot of attention and are used in applications that require strong authentication for login and online banking. Our findings demonstrate the fundamental flaws in the present mobile 2FA systems, as an attacker can obtain the private keying information used to generate the OTP or intercept the victim's OTP. Using a second authentication token, a two-factor authentication (2FA) strategy aims to increase the security of login password-based authentication.

Nilesh A. Lal, Calendar Prasad, Mohammed Farik[20] in this research paper the author explains the management of online information, access control is crucial. User identification and authorization, permission operations, licensing agreement regulations, and digital materials processes for deciding which resources, prompts, or DRM users can access are the four components of access management. papers on the state-of-the-art in access control for digital repositories, covering user authentication, user authorisation, authentication, and methods for secure digital transmission of data.

Ayesha Shahnaz, Usman Qamar, Ayesha Khalid [22]. In this research paper the author explains Blockchain technology is useful in the healthcare sector and how it can be used for electronic medical records. Despite advances in the healthcare sector and innovation in its EHR system, it still faced some problems brought on by this new technology. The paper explains the combination of secure record storage and detailed access rules for those records. This new technology provides a secure and temperamentally robust platform for storing medical records and other health-related information. Paper discusses the implementation of security using blockchain.

Joel JPC Rodrigues, Isabel de la Torre,Gonzalo Fernández.Miguel López-Coronado [23]. In this research paper the author explains Analysis of the The cloud computing paradigm gives eHealth systems the chance to improve the functionality they provide. Security and privacy requirements of cloud-based electronic health records systems. However, transferring critical patient records' security and privacy to the cloud comes with a number of security and privacy hazards. Healthcare providers and cloud service providers must solve security and privacy issues before shifting patient records to the cloud. The common cloud service providers' security needs are examined.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

III. METHODOLOGY

The purpose of this paper is to design and develop a system that provides storage and security to critical medical health records and it also provides insurance cost prediction by taking various parameters as input into consideration. This paper aims to develop a system that will store patient demographics and diagnosis reports in a database by providing authentication to it [12]. It also aims in creating an intrinsic interpretable regression model for health insurance cost prediction.

The dataset which we have used for modeling purpose is Health_insurance.csv (available on https://www.kaggle.com/code/sainithish/analysis-and-prediction-of-health-insurance-data/data). This dataset contains 6 features in total, namely age, sex, bmi, children, smoker, region and charges. Among these available variables, charges in the output variable for our purpose and rest are input variables. we will predict amount of charges for given input parameters like age, sex, bmi, children, smoker and region.

The workflow of the proposed system is shown in Fig. 1. Firstly, to get started with the system, the user must have to sign up and create an account. The user details are then validated using proper authentication. After that, the user has to enter details such as patient demographics and diagnosis reports which will be stored in the real-time database. These health records are also provided with proper authentication. For the prediction model, the user has to enter parameters such as Age, BMI, Region, Premium, etc. which will act as input for training and testing of the model. For training purposes, various supervised regression algorithms are used such as Stochastic Gradient Boosting, XGBoost and Random Forest Regressor as these algorithms provide results which are accurate with high accuracy and cross-validation value. After that, insurance cost prediction is evaluated based on the accuracy of the stipulated algorithms

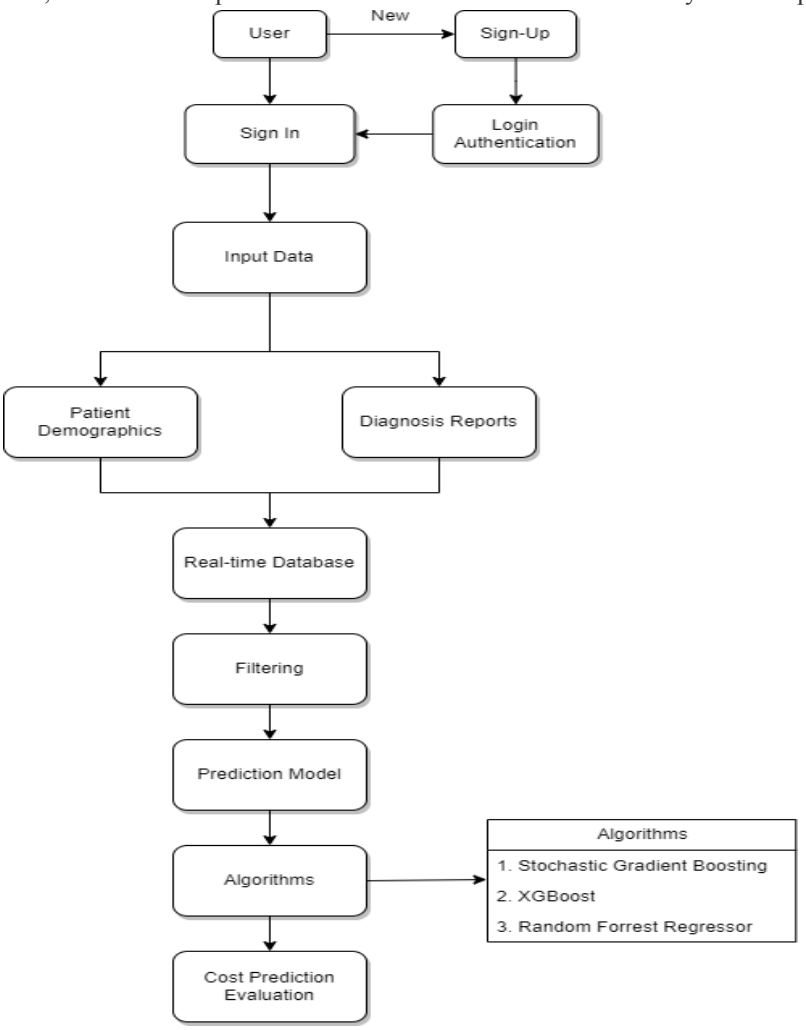


Fig. 1. Flowchart of Proposed system

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Storage of Medical Records :

Once the user logged in, if we store the password in plain form then the database administrator would be able to see every user's password which is not good. That's why we are going to use hashing to store the password in our database. Using encryption to store users' passwords is not the right approach as if someone can get a decryption key then he/she would get access to all the data and that would be a privacy issue.

In Hashing the application can simply hash the entered password and compare it with the stored hash value in the database. That's how no one can decrypt the stored values [19]. That's why we will use hashing here. We are going to use the SHA-256 algorithm.

Information to be receive from user :

- Name, blood type, birth date, and emergency contact information.
- Any chronic diseases
- Major illnesses and surgeries, with dates.
- A list of medicines and supplements, the dosages, and how long that he/she has taken them.
- Dates and results of tests and screenings.
- Any allergies.
- Any history of illnesses in family

Selection of a Database :

As this data would be different for every person and sudden changes may be there so we will use **MongoDB** database as it is faster than MySQL.

We will use MongoDB schemas to handle the data and it's a non-relational database so it will be fast. If we use RDBMS then categorizing each disease into a particular column would be a difficult task, as one can have many different types of diseases so using RDBMS won't be an efficient method. That's why we are going to use the NoSQL database MongoDB.

Security :

Security to Login into system :

Two-factor Authentication :

The user is prompted to log in from the application or website.Users enter what they know Usually username and password. The site's server then finds a match and recognizes the user. For processes that do not require a password, the website will generate a unique security key for the user[20]. An authenticator processes the key and the site's server verifies it. The site then asks the user to initiate a second login sequence. This procedure takes many forms, but users must prove that they own something that belongs only to them. Biometric Data, Security Tokens, ID Cards, Smartphones or Other Mobile Devices. This is an inherent or proprietary factor. Next, the user may need to enter the one-time code generated. If both factors are provided, the user is authenticated and granted access to the application or her website.

Security for Health-Records :

Blockchain Based security :

A new transaction submitted by a user on the blockchain network indicates the creation of a new block. Blockchain blocks are used to hold transactions and these blocks are distributed to all connected nodes in the network. This transaction, placed in blocks, will be sent to all nodes in the network.[23]

This entire process of adding blocks to the blockchain is performed by nodes reaching a consensus that determines which blocks are valid and which are not to be added to the blockchain. This verification is performed by connected nodes using several well-known algorithms to validate transactions and ensure that the sender is an authenticated part of the network.

After validation change is added to the blockchain.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Algorithms used : PoW(Proof of work),SHA(secure hashing algorithm)

Cloud based security :

Providing cloud-based healthcare solutions is an important step in the evolution of eHealth. Our cloud-based system allows you to build a scalable environment that fits your needs. This full customization is complemented by the savings offered by pay-as-you-go systems like cloud computing. Another big advantage comes from the fact that when an EHR is hosted in the cloud, medical staff and patients can access the information anytime, anywhere with an internet connection.With a private cloud, your trusted cloud provider can handle your privacy and data security.[24]

IV. MODEL PREDICTION

We have used a variety of machine learning approaches on information related to insurance cost prediction. The dataset was retrieved from the KAGGLE repository. First, the dataset is subjected to data preprocessing. After that, feature engineering is carried out to choose the features that contribute to the model's accuracy. The dataset is then divided into a train dataset and a test dataset [13]. Training datasets are used to estimate health insurance costs, whereas testing datasets are used to assess regression models.

Data Pre-processing :

There are seven variables in the dataset used to forecast the price of health insurance as shown in the table below. We evaluated the dataset to see if there were any missing values before estimating the cost of health insurance. We discovered that the BMI and charges columns both contain missing values that were assigned to their mean values.

Categorical columns like sex, smoker, and region were transformed into numerical values because regression models only require numerical data. We utilize Label Encoding to convert categorical values to numerical values.

Name	Description
Age	Age of the user
BMI	Body Mass Index of the user
Children	Number of children of the user
Gender	Male/Female
Smoker	Whether the user is a smoker or not
Region	Where the user lives
Charges (Target value)	Medical expenditure the user has to pay

Feature Engineering :

In feature engineering, features were extracted out of raw data to improve how well machine learning algorithms work. Age, BMI, and Smoker are the parameters that influence charges in our health insurance cost prediction dataset. Contrarily, determining the charges is not at all influenced by elements like sex, children, or religion. Therefore, we can also remove these columns. We can use a heap map to find the features that are most connected to other features or the goal value.

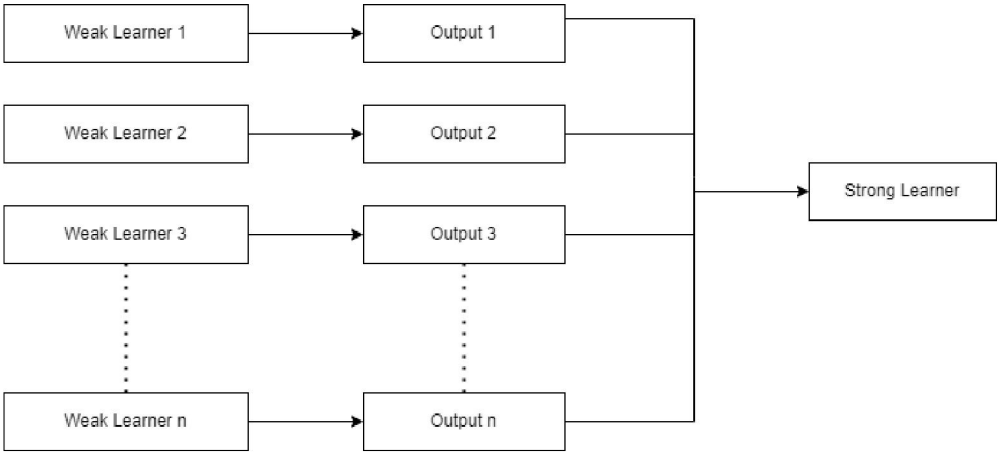
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Data Splitting :

Overall in our dataset we have a total of 6 features, out of which "charges" is output variable and first 5 are input variables. The dataset is splitted in 80:20 proportion i.e 80% of dataset will be used to train model and rest 20% of data will be used as test data to test model.

Regression :

In order to solve complex and convoluted problems, we need more advanced techniques. Boosting is a process that uses a set of machine Learning algorithms to combine weak learners to form strong learners in order to increase the accuracy of the model. Ensemble Learning is a method that enhances the performance of a Machine Learning model by combining several learners.



Accurately estimating a person's expected future healthcare costs is a crucial informatics tool for managing healthcare costs. In order to meet this crucial demand, we carried out a thorough literature analysis and found three techniques for forecasting healthcare expenses [14]. We empirically assessed the prediction performance of each described approach, allowing for a direct comparison of these various approaches.

These approaches are as follows :

A. Stochastic Gradient Boosting :

A machine learning technique called ensemble learning combines multiple base models to create a single, ideal predictive model. An ensemble method for building a set of predictors is boosting. This method involves teaching learners sequentially, starting with early learners fitting basic models to the data and moving on to later learners checking the data for flaws. In other words, we fit successive trees (random sample) with the aim of resolving the net error from the previous tree at each stage.

Input: training set $\{(x_i, y_i)\}_{i=1}^{n}$, a differentiable loss function $L(y, F(x))$, number of iterations M.

Algorithm:

1. Initialize model with a constant value:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma).$$

2. For $m = 1$ to M:
   1. Compute so-called *pseudo-residuals*:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \ldots, n.$$

   2. Fit a base learner (e.g. tree) $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $\{(x_i, r_{im})\}_{i=1}^{n}$.
   3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\right).$$

   4. Update the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. Output $F_M(x)$.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

Here, M is the number of the base model, and L(y,F(x)) is the loss function (log-loss, linear loss, hinge loss, etc.). Since everyone completed their tasks simultaneously, the GBDT was not parallelized, requiring more time to train than the RF.

A fundamental learner with strong bias and low variation is gradient boosting. Here, we'll employ a decision tree with gradient boosting and very little depth.

Because we don't use it in RF, we are unable to minimize loss. However, with gradient boosting, we can reduce any losses.

B. XGBoost Algorithm

An enhanced library for gradient boosting in machine learning is called XGBoost.

This algorithm's basic concept is that it builds D classification and regression trees (or CARTs) one at a time, allowing each new model to be trained using the residuals of the preceding one. In other words, the new model predicts the outcome after correcting the mistakes caused by the earlier trained tree.

The sum of D functions is used by each ensemble model in the XGBoost to predict the result

$$\hat{Y}_i = \text{₣}(X_i) = \sum_{d=1}^{D} f_d(X_i),\ f_d \in \text{F},\ i = 1, \ldots, n$$

where, F = Function space;

fd = Independent CART Structure n = Number of observations

X = Covariates Wq(X) = Score.

C. Random Forest Regressor

A method called Random Forest uses ensemble learning to aggregate numerous weak classifiers to offer solutions to challenging issues. Bagging is a technique used in random forests. The overfitting issue is resolved since it constructs a subset of the original dataset and bases the outcome on the majority ranking.

The fact that this fantastic method can also be applied to feature selection is another outstanding feature. It can be used to determine the value of the feature. We must first comprehend how feature importance is calculated using Decision Trees in order to comprehend how it is calculated in Random Forests.

To calculate Feature importance, the following formula is used :

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k}$$

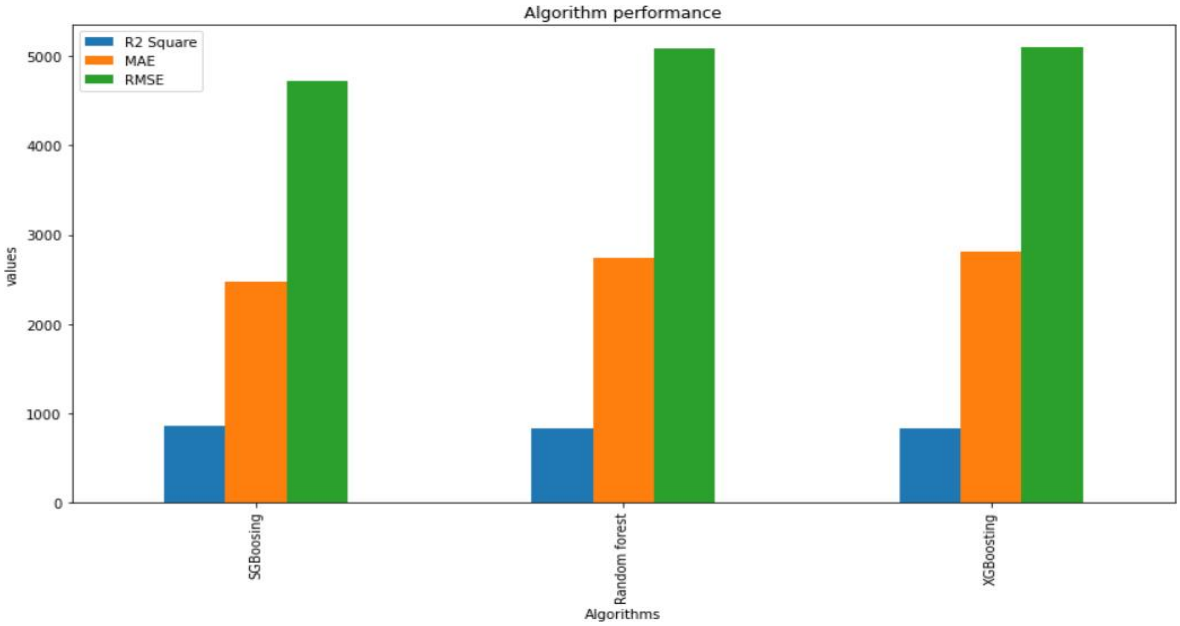
V. RESULTS

Following are the comparative results of all the three algorithmic models which we used to predict Insurance cost charges with the dataset mentioned above

Algorithms Used	R-Squared Value	MAE Value	RMSE Value
Stochastic Gradient Boosting	0.8523	2471.5328	4722.9231
Random Forest Regression	0.8294	2742.5423	5085.5100
XGBoosting	0.8278	2806.6131	5098.7481

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

As one can clearly see, Stochastic Gradient Boosting is having 4722.9231 as RMSE value and 2471.5328 as MAE value, which are lowest among all three methods, then it is followed by Random Forest Regression and last of all XGBoosting method. This results tells us that Stochastic Gradient Boosting outperforms among all other methods. Following is the bar graph which graphically represents performance of all three algorithms.



In the above bar graph "R2 Square" value is scaled up (by 1000 times) to be visible in the graph, because due to very low scale as compared to MME and RMSE values it was not visible on the graph as if it does not exist.

VI. CONCLUSIONS

Medical health records are a very critical and personal thing, and it is undoubtedly most useful today. We can't even imagine in what way it can be used. Today medical based organizations need this data, so that they can make predictions based on it. In this research paper we have tried to lighten the most ignorant and yet immensely useful subject that is, secure storage of individuals' medical history and prediction on insurance cost using gradient algorithms. The KAGGLE repository's medical insurance dataset was used to train and test the Stochastic Gradient Boosting, XGBoost, and Random Forest Regression ML algorithms. Preprocessing, feature engineering, data slicing, regression, and evaluation were the procedures that this dataset underwent. Stochastic Gradient Boosting (SGB) had a high accuracy of 86% and an R-Squared value of 0.8523, according to the final results

REFERENCES

[1]. Dutta, Madan. (2020). Health insurance sector in India: an analysis of its performance. Vilakshan - XIMB Journal of Management. ahead-of-print. 10.1108/XJM-07-2020-0021.

[2]. Anwar Ul Hassan, Ch & Iqbal, Jawaid & Hussain, Saddam & Mosleh, Mogeeb & Ullah, Syed Sajid.(2021). A Computational Intelligence Approach for Predicting Medical Insurance Cost. Mathematical Problems in Engineering. 2021. 1-13. 10.1155/2021/1162553.

[3]. Singh, Rana & Singh, Abhishek. (2020). A Study of Health Insurance in India. 10. 2249-0558.

[4]. Bhardwaj, Akashdeep. (2020). Health Insurance Claim Prediction Using Artificial Neural Networks.

[5]. International Journal of System Dynamics Applications. 9. 40-57. 10.4018/IJSDA.2020070103.

[6]. Geeta, Shetty Deepa & Subramanian, Sp Mathiraj & Muthu, M. Vinoth. (2018). A Study on Health Insurance Premium, Claims, Commission and its Growth of Select Companies in India. 7. 109-121.

[7]. Duncan, M. Loginov & M. Ludkovski (2016) Testing Alternative Regression Frameworks for Predictive Modeling of Health Care Costs, North American Actuarial Journal, 20:1, 65-87, DOI: 10.1080/10920277.2015.1110491

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, May 2023

[8]. Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. Risks 2019, 7, 70. https://doi.org/10.3390/risks7020070

[9]. Hanafy, Mohamed. (2021). Predict Health Insurance Cost by using Machine Learning and DNN Regression Models. International Journal of Innovative Technology and Exploring Engineering. Volume-10. 137. 10.35940/ijitee.C8364.0110321.

[10]. Slaveykov, Kiril, Kalina Trifonova, Valentin Stoyanov, and Ljubima Despotova-Toleva. "Electronic health records–benefits, savings and costs." Medicine 3, no. 1 (2013): 97-100.

[11]. Hindawi, et al. "A Computational Intelligence Approach for Predicting Medical Insurance Cost." A Computational Intelligence Approach for Predicting Medical Insurance Cost, 28 Dec. 2021, www.hindawi.com/journals/mpe/2021/1162553.

[12]. Morid, Mohammad Amin, et al. "Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation." PubMed Central (PMC), 16 Apr. 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC5977561.

[13]. Panay, Belisario, et al. "Feature Selection for Health Care Costs Prediction Using Weighted Evidential Regression." PubMed Central (PMC), 6 Aug. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7472302.

[14]. Kaushik, Keshav, et al. "Machine Learning-Based Regression Framework to Predict Health Insurance Premiums - PubMed." PubMed, 28 June 2022, pubmed.ncbi.nlm.nih.gov/35805557.

[15]. Devi, M. Shyamala, et al. "Linear and Ensembling Regression Based Health Cost Insurance Prediction Using Machine Learning." Linear and Ensembling Regression Based Health Cost Insurance Prediction Using Machine Learning SpringerLink, 14 July 2021, link.springer.com/chapter/10.1007/978-981-16- 1502-3_49.

[16]. Mathioudakis A, Rousalova I, Gagnat AA, et al. How to keep good clinical records. Breathe 2016; 12: 371–375.

[17]. Innocent Mapanga, Prudence Kadebu, "Database Management Systems: A NoSQL Analysis", International Journal of Modern Communication Technologies & Research (IJMCTR) , ISSN: 2321-0850, Volume-1, Issue-7, September 2013.

[18]. Mr.Bhojaraju G. Dr.M.M.Koganurmath, "Database System : Concepts and Design" XXIV All India Conference of IASLIC, 15-18 December, 2003, Survey of India.

[19]. Dr. Rajasulochana , Jayalakshmi, "HOSPITAL MANAGEMENT SYSTEM AND ITS DATABASE" Dept. of Genetic Engineering BIHER, BIST, Bharath University Chennai- 600073 International Journal of Pure and Applied Mathematics Volume 119 No. 12 2018, 3047-3057

[20]. Kotapati Saimanoj, Grandhi Poojitha, Khushbu Devendra Dixit, Laxmi Jayannavar "Hospital Management System using Web Technology" ICAIT-2020 , Bengaluru Article Info Volume 83 Page Number: 4493-4496

[21]. Alexandra Dmitrienko,Christopher Liebchen,Christian Rossow,Ahmad-Reza Sadeghi ."Security Analysis of Mobile Two-Factor Authentication Schemes", Intel® Technology Journal | Volume 18, Issue 4, 2014

[22]. Nilesh A. Lal,Salendra Prasad,Mohammed Farik. "A Review Of Authentication Methods",International Journal Of Scientific & Technology Research Volume 5, Issue 11, November 2016

[23]. Md. Zahid Hossain Shoeb. "Access Management for Digital Repository", DESIDOC Journal of Library & Information Technology, Vol. 29, No. 4, July 2009, pp. 21-27

[24]. Ayesha Shahnaz, Usman Qamar, Ayesha Khalid "Using Blockchain for Electronic Health Records", Digital Object Identifier 10.1109/ACCESS.2019.2946373,October 9 2019.

[25]. Joel JPC Rodrigues, Isabel de la Torre,Gonzalo Fernández.Miguel López-Coronado. "Analysis of the Security and Privacy Requirements of Cloud-Based Electronic Health Records Systems", JMIR publicationPublished on 21.8.2013 in Vol 15, No 8 (2013).

[26]. Pandey, Belisario & Baloian, Nelson & Pino, José & Peñafiel, Sergio & Sanson, Horacio & Bersano-Méndez, Nicolás. (2019). Predicting Health Care Costs Using Evidence Regression. Proceedings. 31. 74. 10.3390/proceedings2019031074.