Asian Journal of Research in Computer Science

Volume 16, Issue 2, Page 46-53, 2023; Article no.AJRCOS.101266
ISSN: 2581-8260

Health Insurance Cost Prediction Using Deep Neural Network

Md. Samiuddin [a], G. Rajender [a], K. Sai Abhiram Varma [a], A. Ravi Kumar [a++*] and Subhani Shaik [a#]

[a] Department of Information Technology, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad, T.S, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2023/v16i2338

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/101266

Original Research Article

Received: 03/04/2023
Accepted: 09/06/2023
Published: 13/06/2023

ABSTRACT

Artificial intelligence (AI) and Deep Learning (DL) are strategies for making human being's lives simpler in the healthcare enterprise through predicting and identifying ailments more fast than the general public of scientific specialists. There may be an immediate connection between the insurance organization and the policyholder while technology reduces the distance between them to zero in particular with digital medical insurance. In preference to commonplace protection, simulated intelligence and profound mastering have meaningfully impacted the way in which guarantors build health care coverage designs and empowered customers to hastily get benefits greater. With a view to provide clients with accurate, spark off, and effective medical health insurance, insurance companies use DL. Medical health insurance quotes have been expected the use of an artificial neural network (ANN) and a deep neural network (DNN) algorithm on this take a look at. Based on the traits of the individuals, the author envisioned how a good deal medical insurance might price. Age, gender, body mass index, the range of kids, smoking behavior, and place had been all used to train and examine an artificial neural network model.

[++] Assistant Professor;
[#] Associate professor;
**Corresponding author: E-mail: aravikumar@sreenidhi.edu.in;*

Keywords: Deep learning; healthcare analysis; insurance cost; web development.

1. INTRODUCTION

We live on a hazardous and eccentric world. There are many different kinds of risks that can affect people, families, businesses, buildings, and land and the risk levels can vary. These dangers include the possibility of dying, getting sick, or losing land or wealth. The most important aspects of a person's existence are life and happiness. However, risks cannot always be avoided, so the financial industry has developed numerous products to compensate for them and shield individuals and businesses from them. Consequently, insurance is a program that reduces or eliminates costs associated with various risks [1-5]. It is essential for insurance companies to be precise enough to measure or define the amount protected by this policy and the insurance costs that must be made for it when it comes to the value of insurance in people's lives. Various factors are used to estimate these costs. Each of these matters a lot. When the numbers are calculated, the strategy as a whole changes if any part is left out. Consequently, it is essential to complete these tasks precisely. Since human mistakes are potential, guarantors utilize specialists in this field. Additionally, they calculate the insurance rate using a variety of instruments. milliliter proves to be useful here. The method or work that was used to create the policy might be generalized by ml. It is possible to master these milliliter models on your own. Using data from previous insurance policies, the algorithm is trained. Insurance policy prices can be accurately predicted by the model when the necessary variables for measuring payouts are specified as model inputs. Personal effort and resources are reduced, and the company's revenue is increased. Consequently, milliliter can improve accuracy.

2. LITERATURE REVIEW

2.1 An Emerging Trend of Big Data Analytics with Health Insurance in India

Life insurance issues, which, crucially, share significant milestones with general insurance, have engulfed India's healthcare services and delivery, including funding, for the past century. There is a significant knowledge gap regarding developments in public and commercial health funding and service, according to numerous specialists, practitioners, policymakers, and academics. Using "Big Data Analytics," it is evident that the existing framework of health care services must be altered. In addition to lower healthcare costs, we will discuss the various implications and characteristics of this new era of advanced and improved data management in this paper. We will also pay close attention to the use cases that propel new technology and ultimately result in economic advancement.

2.2 Predicting Motor Insurance Claims Using Telemetric Data—XG Boost Versus Logistic Regression

XGBoost is first-rate for its exquisite gauging potential. Models for a dichotomous answer that show the presence of accident claims versus no claims may be used to determine out what causes road accidents. This study looked at how nicely the logistic regression and XGBoost procedures predicted the presence of mishap claims based on telemetry statistics. Statistics approximately human's transportation conduct, consisting of the overall annual distance traveled and the share of general distance pushed in metropolitan regions, were blanketed inside the sample from an coverage organisation. It became determined that logistic regression become suitable version because of its forecasting and interpretability capabilities.Numerous version-tuning strategies and further analysis are required for XGBoost to suit the logistic regression model’s forecast overall performance [6-10].

2.3 Automating Car Insurance Claims Using Deep Learning Techniques

There has been an increase in the number of auto insurance claims filed as a result of an adding number of people driving each day. The time- consuming life cycle of form, reviewing, and deciding on each claim includes a written evaluation from the damage report's service expert and a physical check from an insurance company assessor. We propose a comprehensive approach to reuse operation that would be profitable to both the customer and the business. This system takes filmland of the damaged vehicle as input and gives applicable information like the damaged corridor and an estimate of how important damage each part has

(no damage, moderate damage, or serious damage). This indicates a rough estimate of the restoration cost, which will be used to figure out the quantum of the insurance claim. We experimented with well- known case segmentation models like the Mask R- CNN, PANet, and an ensemble of these two, as well as a transfer literacy (1) grounded VGG16 network, to negotiate colorful tasks of localizing and relating colorful groups of factors and blights discovered in the vehicle. Also, the suggested approach receives high chart conditions for damage and element localization.0.38 and0.40 in each case).

2.4 Predicting the Customer Churn with Machine Learning Methods: Case: Private Insurance Customer Data

The maturity of businesses uses tools for customer relationship operation to outperform their rivals. This includes giving each client a value for their entire client lifecycle, which gives them a precise standing for each client in their database. This is done to concentrate plutocrat on marketing and deals, making spending further effective and focused. The problem is that it requires a comprehensive client history for each client, which is infrequently available. In practice, the kinds of records needed to calculate customer lifecycle values are extensively different from the information that's accessible in operation systems. Any accurate estimation of customer lifecycle values is insolvable due to this gap. In this study, we present a fashion for producing some of the missing factors for CLV calculations. This necessitates a flexible soothsaying and evaluation strategy as well as a specialist design for storing data [11-13].

2.5 The Accuracy of XG Boost for Coverage Declare Prediction

The development of styles for snappily and precisely filing claims is necessary due to the rising trend of claim frequency and intensity in bus insurance. Machine literacy is one of them, and it approaches the problem like guided literacy. In utmost cases, there's a lot of data on former claims. Also, there are multitudinous deficient figures for multitudinous data characteristics. Thus, it's necessary to use machine literacy algorithms that are suitable to manage both data traits. The new ensemble learning algorithm XGBoost should work well with both types of data. We apply and assess the efficacity of XGBoost for the problem of claim protuberance in this composition. XGBoost's performance is also compared to that of other ensemble literacy strategies like AdaBoost, Random Forest, and Stochastic GB, as well as online literacy- grounded strategies like Neural Network. Grounded on our computations, XGBoost performs better than other styles when it comes to regularized Gini.

3. METHODOLOGY

3.1 Preliminaries

As of now, the undertaking's reasonableness is evaluated, and an essential understanding is furnished with a vital errand plan and a few expense projections. During the framework investigation, it is necessary to investigate the reasonableness of the proposed framework. This will guarantee that the recommended arrangement will not cause problems for the company. It is essential to have a solid understanding of the primary requirements for the framework for the feasibility study.

The feasibility study is significantly influenced by the following three elements:

Take into account economic, technological, and social feasibility simultaneously.

3.2 Economical Feasibility

This examination is being finished to see what sort of cash related influence the construction will have on the affiliation. The association has a limited extent of cash to put resources into the framework's creative energy. The expenses should be kept up with by check. Thus, the made framework came in under spending plan, which was made conceivable by the way that an immense piece of the improvement were public space. Only those particular items ought to have been purchased.

3.3 Technical Feasibility

The purpose of this evaluation is to evaluate the specific requirements or reach ability of the framework. Any framework made shouldn't put a tremendous burden on the mechanical assets reachable. How much mechanical assets accessible will be vigorously stressed subsequently. Subsequently, the client will confront exclusive standards. The made

framework should have a low interest in light of the fact that its execution only requires next to zero changes [14,15].

3.4 Social Feasibility

The objective of the review is to decide how much the client acknowledges the framework. This involves the guidance required for the client to actually work the framework. The system shouldn't make the client feel split the difference; taking everything into account, they should consider it to be a need. The procedure used to illuminate and adapt the client with the framework is the chief factors that influence the degree of attestation by the clients. As the framework's legitimate client, his conviction should be stretched out with the objective that he could offer some obliging investigation, which is locked in.

3.5 Proposed Algorithm

DNN: Deep neural networks are a sort of brain organization. A deep neural network (DNN) is an artificial neural network (ANN) with numerous levels between the input and output layers. There are many different kinds of neural networks, but they all have the same parts: functions, weights, biases, and connections among neurons.

Autoencoder NN: An autonomous learning technique for neural networks known as an autoencoder instructs the network to ignore signal "noise" in order to learn efficient data models (encoding). Denoising, encoding, and, in some cases, creating image data all require auto encoders.

Linear Regression: The purpose of linear regression analysis is to predict the value of one variable from the value of another. The term "dependent variable" refers to the one you want to forecast. The independent variable is the one you're using to predict the other variable's value.

Random forest: The Random Forest a directed ML calculation that is often utilized in Order and Relapse undertakings. Decision trees are created by utilizing the majority vote for classification and the average for regression from multiple samples.

Decision tree: Non-parametric supervised learning in the form of a decision tree can be utilized for both classification and regression. A root hub, branches, inward hubs, and leaf hubs make up its various leveled tree structure.

Voting classifier: Kagglers frequently use a machine-learning method called the Voting Classifier to improve their model's performance and rise in rank. Projecting a democratic Classifier may in like manner be used to augment execution on real world datasets, despite the way that it has basic cutoff points.

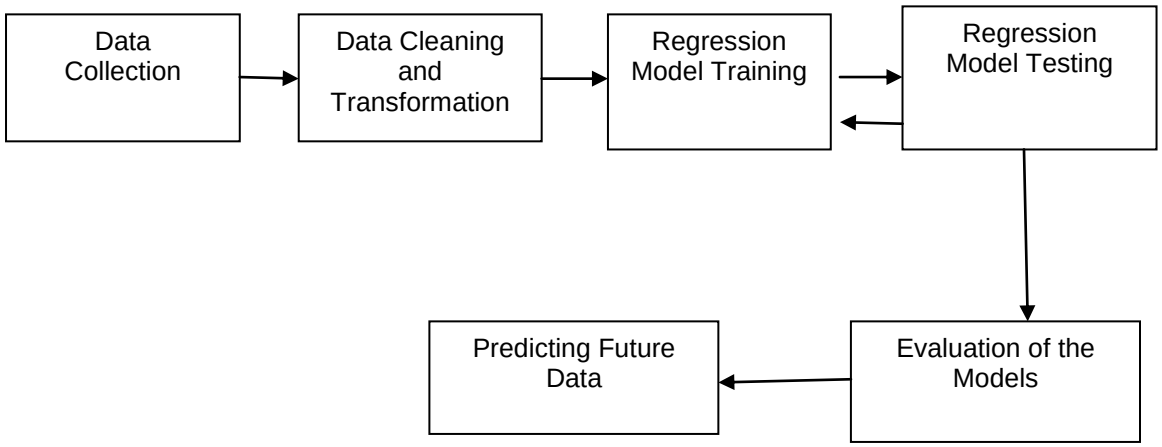

Fig. 1. System architecture

3.6 Dataset

In our project, we used data from various hospital websites and trained the data using deep learning and machine learning algorithms. This data set includes six attributes, and the set has been split into two parts: training data and testing data. For training the model, 80% of the total data is used, and the rest is used for testing. To build a predictor model of medical insurance costs, the training dataset is applied, and to evaluate the regression model, a test set is used.

4. RESULTS AND DISCUSSION

When we runs the code the above website is opened and it will display sign up button with some other widgets.

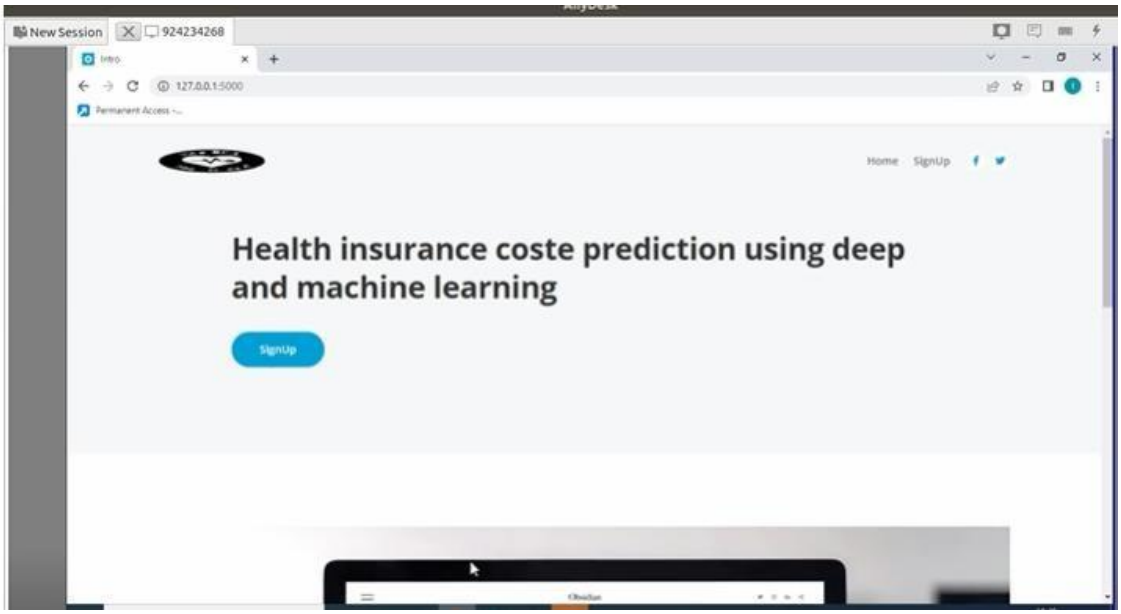

Fig. 2. Web page for health insurance

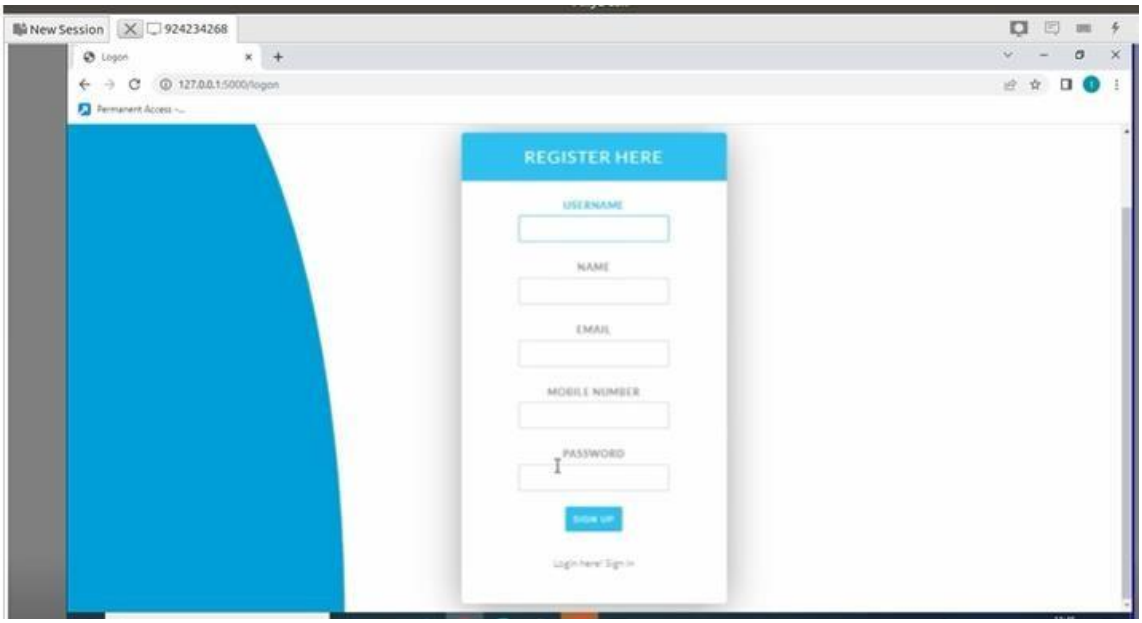

Fig. 3. Registration page for health insurance

When we click signup button it will display the registration form we need to fill the details and click submit.

We need to log into the website using username and password that we created during registration.

After logging in the website will be appear like this and notification regarding changing the password will bedisplayed.

The website displays six attributes and the user need to fill those attributes to get about best insurance cost. The attributes displayed are age, gender, bmi, children, whether user has a habit of smoking and region, using all these attributes it can predict the best insurance cost for user.

This is the final result displayed on the screen the insurance cost of the user according to the filled attributes. The website is easy to use and it uses machine learning and deep learning algorithms to find the insurance costof the user.

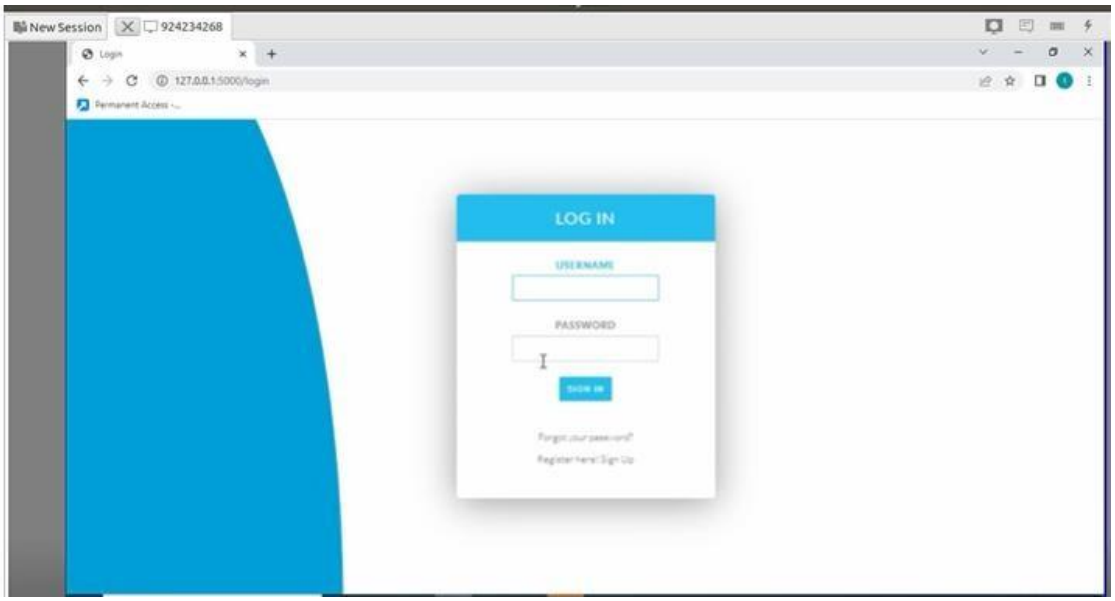

Fig. 4. Login page for health insurance

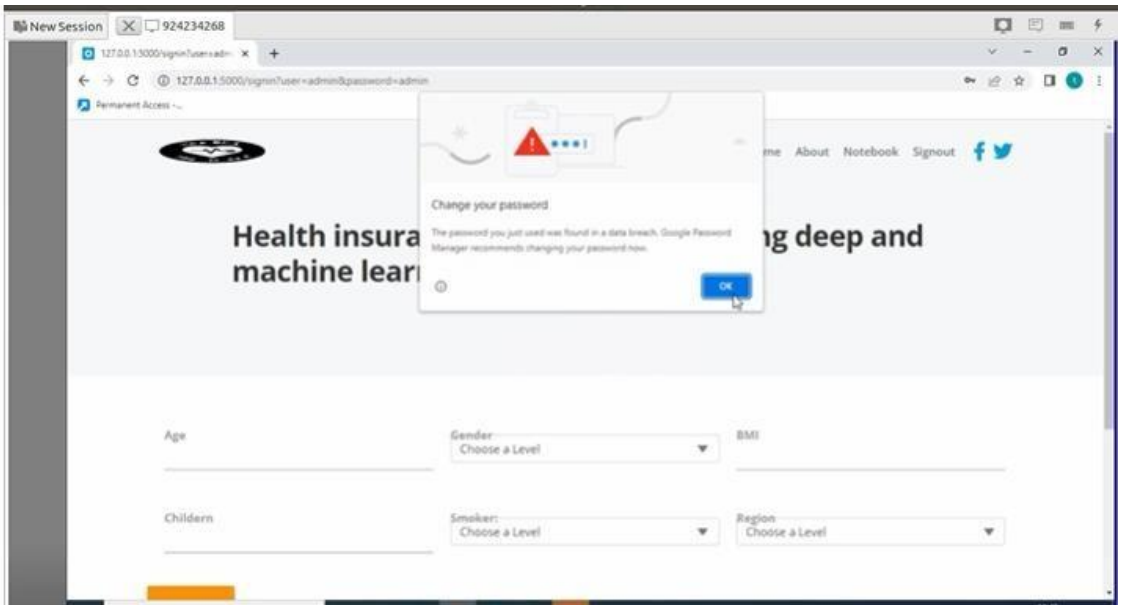

Fig. 5. Details page for health insurance

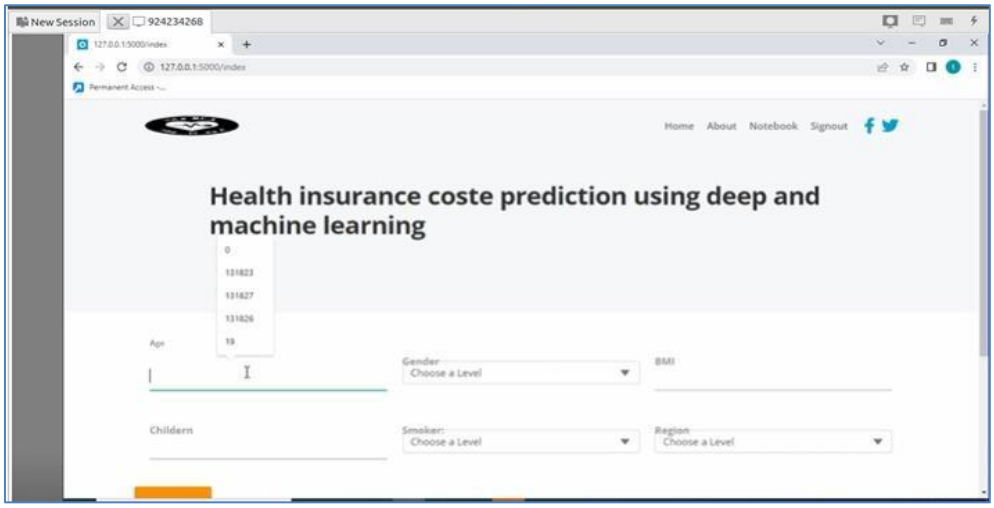

Fig. 6. Details page for health insurance

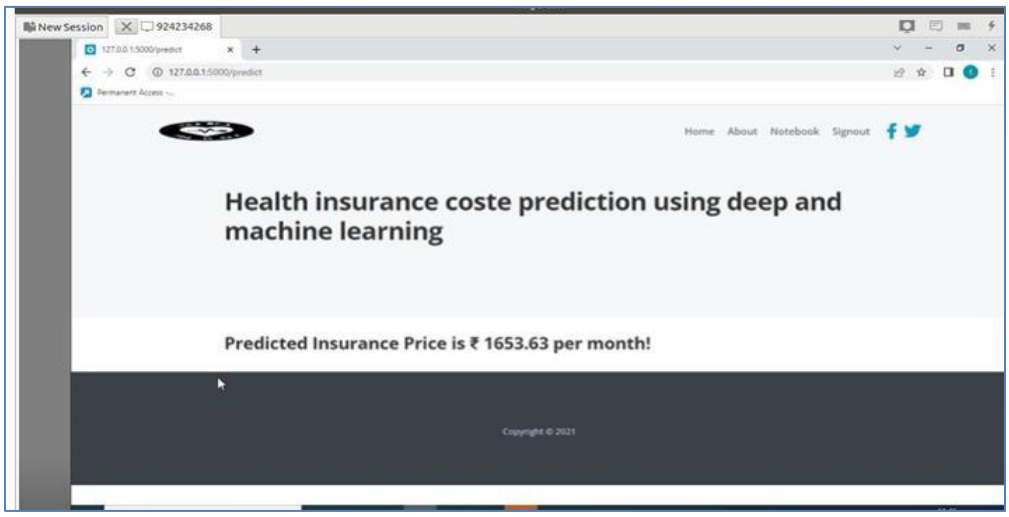

Fig. 7. Cost prediction page

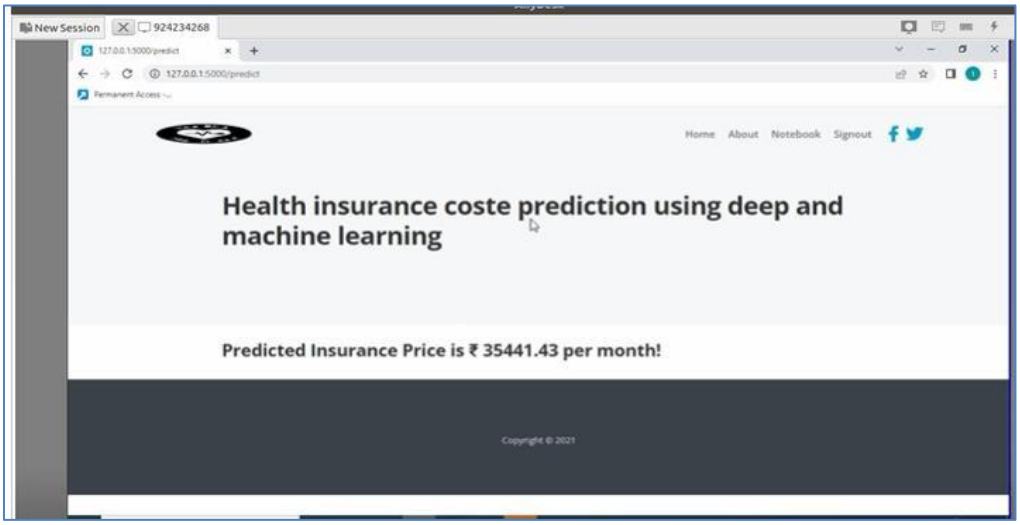

Fig. 8. Cost prediction page

5. CONCLUSION AND FUTURE WORK

In our project, we used medical samples and sets of data from websites. We used machine learning and deep learning algorithms to calculate coverage charges for human beings based on their unique characteristics. Coverage organisations can attract new clients and save time while developing male or female programmes by forecasting insurance costs based totally on particular variables. Because DL models can quickly calculate cots when compared with machine learning algorithms, deep learning can substantially reduce these authorities efforts. Companies can be able to improve their sales as a result of this. Furthermore , DL algorithms are able to deal with large quantities of facts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Gupta S, Tripathi P. An emerging trend of big data analytics with health insurance in India. In: International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH). Vol. 2016. IEEE Publications. 2016;64-9.
2. Medical Cost K. Personal datasets. Kaggle Inc. Available:https://www.kaggle.com/mirichoi0218/insurance
3. Pesantez-Narvaez J, Guillen M, Alcañiz M. Predicting motor insurance claims using telematics data—XGBoost versus logistic regression. Risks. 2019;7(2): 70.
4. Singh R, Ayyar MP, Sri Pavan TV, Gosain S, Shah RR. Automating car insurance claims using deep learning techniques. In: IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE Publications. 2019; 199-207.
5. Stucki O. Predicting the customer churn with machine learning methods [case]: private insurance customer data; 2019.
6. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; 338:b2393.
7. Van Buuren S. Flexible imputation of missing data. CRC Press; 2018.
8. Fauzan MA, Murfi H. The accuracy of XGBoost for insurance claim prediction. Int J Adv Soft Comput Appl. 2018;10(2).
9. Kowshalya G, Nandhini M. Predicting fraudulent claims in automobile insurance. In: Second International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE Publications. 2018;1338-43.
10. Kayri M, Kayri I, Gencoglu MT. The performance comparison of multiple linear regression, random forest and artificial neural network by using photovoltaic and atmospheric data. In: 14th International Conference on Engineering of Modern Electric Systems (EMES). Vol. 2017. IEEE Publications. 2017;1-4.
11. Denuit M, Hainaut D, Trufin J. Effective statistical learning methods for actuaries I: GLMs and extensions; 2019.
12. Breiman L. Random forests.‖ Machine Learning. Springer. 2001;45(1):5-32.
13. Chen T, Guestrin C. XGBoost: A scalable tree boosting system 22nd ACM SIGKDD Int. In: conference on Knowledge Discovery and Data Mining; 2016.
14. Aler R, Galván IM, Ruiz-Arias JA, Gueymard CA. Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. Sol Energy. 2017;150: 558-69.
15. Volkovs M, Yu GW, Poutanen T. Content-based neighbor models for cold start in recommender systems. In: Proceedings of the recommender systems challenge. 2017;1-6.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/101266