Health Insurance Cost Prediction Using DeepLearning and Machine Learning

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 faster 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.

Index Terms – Deep Learning, Healthcare Analysis, Insurance cost, and web development.

I. 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]. 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.

II. LITERATURE REVIEW

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.

XGBoost is first-rate for its exquisite guaging potential. Models for adichotomous 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.

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 film land 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).

III.METHODOLOGY

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 autoencoders.

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.

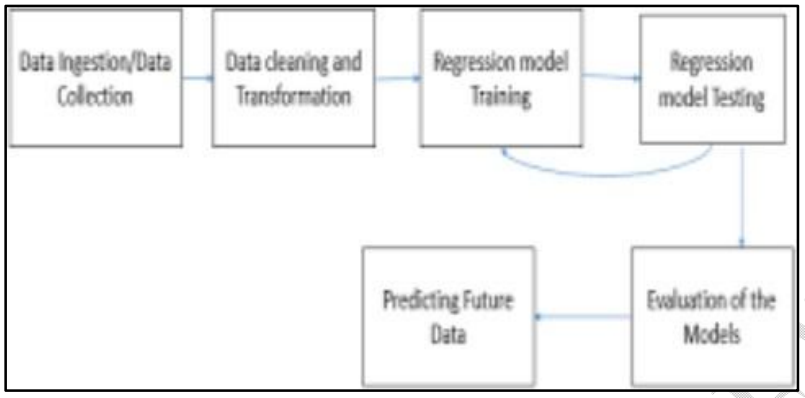

Figure 1: Phases of project implementation

3.1 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.

IV.RESULTS AND DISCUSSION

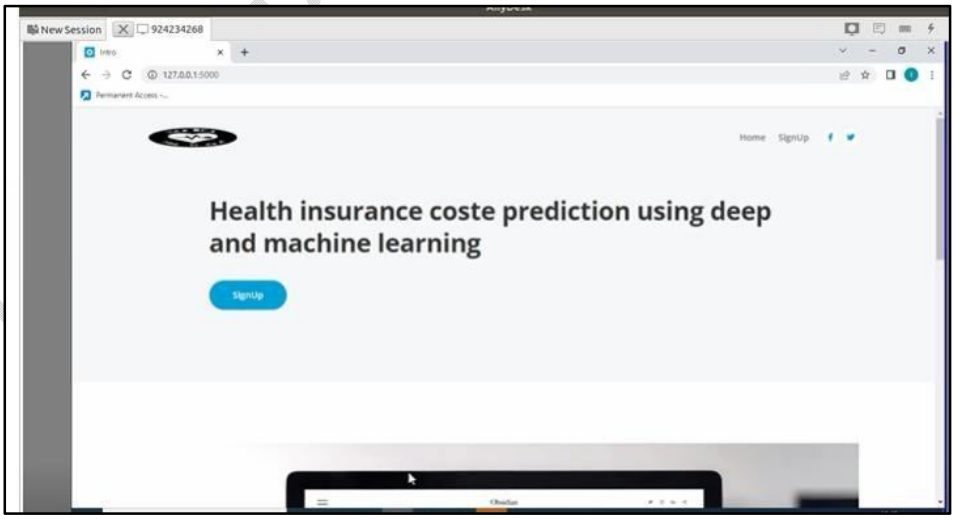

Figure 2: Web Page for health insurance

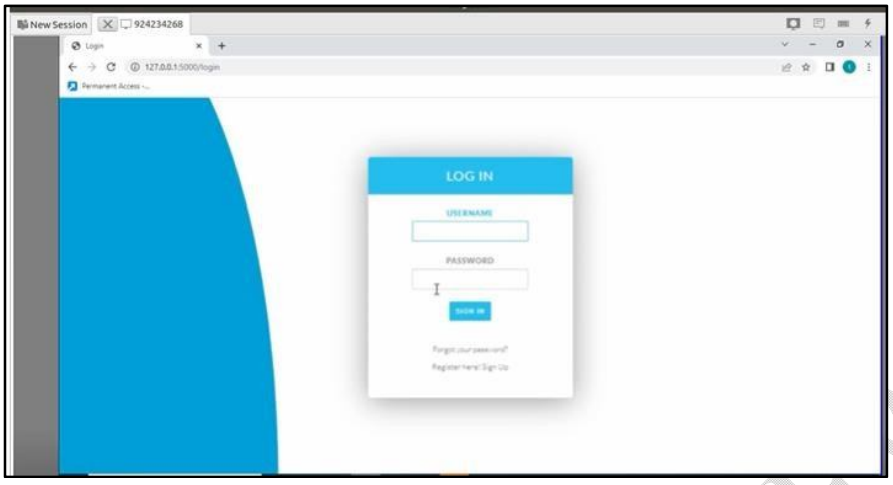

Figure 3: Login Page for health insurance

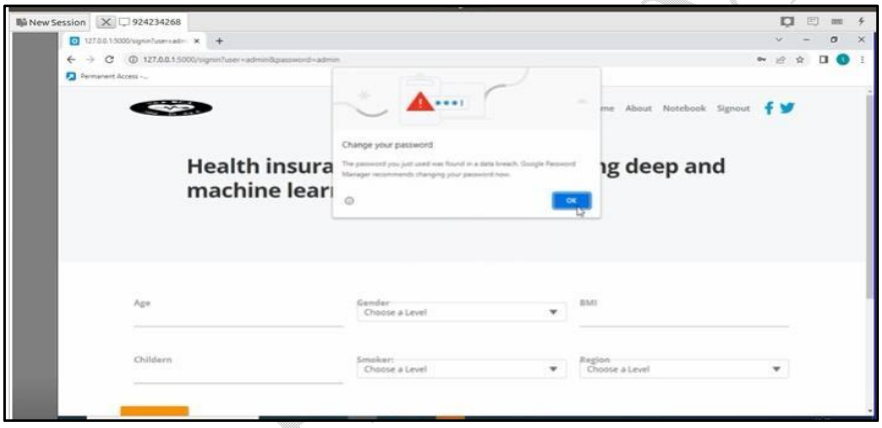

Figure 4: Details Page for health insurance

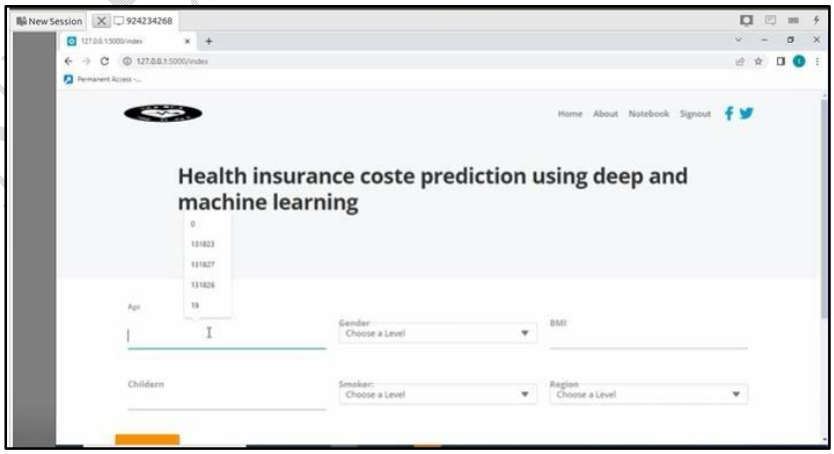

Figure 5: Details Page for health insurance

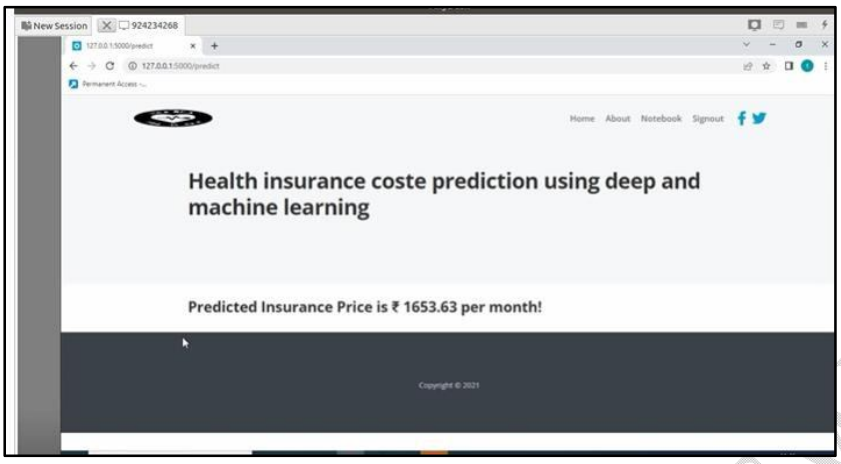

Figure 6: Details Page for health insurance

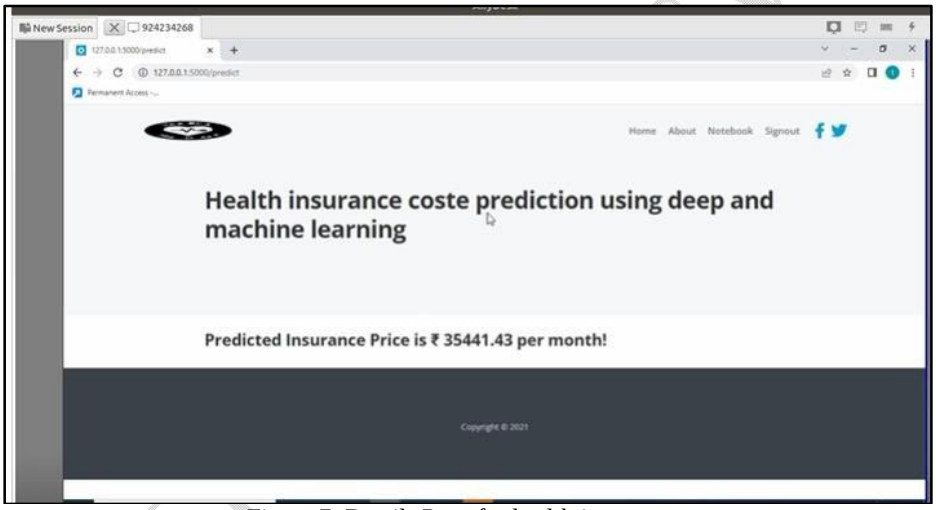

Figure 7: Details Page for health insurance

V.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 organizations can attract new clients and save time while developing male or female programmers 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.

References

1. Gupta, S., & Tripathi, P. (2016, February). An emerging trend of big data analytics with health insurance in India. In 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH) (pp. 64-69). IEEE.

2. Kaggle Medical Cost Personal Datasets. Kaggle Inc. https://www.kaggle.com/mirichoi0218/insurance.

3. Pesantez-Narvaez, J., Guillen, M., & Alcañiz, M. (2019). Predicting motor insurance claimsusing telematics data—XGBoost versus logistic regression. Risks, 7(2), 70

4. Singh, R., Ayyar, M. P., Pavan, T. S., Gosain, S., & Shah, R. R. (2019, September). Automating Car Insurance Claims Using Deep Learning Techniques. In 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM) (pp. 199-207). IEEE.

5. Stucki, O. (2019). Predicting the customer churn with machine learning methods: case: private insurance customer data.

6. Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., ... & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj, 338.

7. Van Buuren, S. (2018). Flexible imputation of missing data. CRC press.

8. Fauzan, M. A., & Murfi, H. (2018). The accuracy of XGBoost for insurance claim prediction. Int. J. Adv. Soft Comput. Appl, 10(2).

9. Kowshalya, G., & Nandhini, M. (2018, April). Predicting fraudulent claims in automobile insurance. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1338-1343). IEEE.

10. Kayri, M., Kayri, I., & Gencoglu, M. T. (2017, June). The performance comparison of multiple linear regression, random forest and artificial neural network by using photovoltaic and atmospheric data. In 2017 14th International Conference on Engineering of Modern Electric Systems (EMES) (pp. 1-4). IEEE.

11. Denuit, Michel & Hainaut, Donatien & Trufin, Julien. (2019). Effective Statistical Learning Methods for Actuaries I: GLMs and Extensions. 10.1007/978-3-030-25820-7.

12. Breiman, Leo. 2001. ―Random Forests.‖ Machine Learning 45 (1). Springer: 5–32.

13. Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system 22nd ACM SIGKDD Int. In Conf. on Knowledge Discovery and Data Mining.

14. Aler, R., Galván, I.M., Ruiz-Arias, J.A., Gueymard, C.A. (2017). Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. In Solar Energy vol. 150, pp. 558-569.

15. Volkovs, M., Yu, G. W., & Poutanen, T. (2017). Content-based neighbor models for cold start in recommender systems. In Proceedings of the Recommender Systems Challenge 2017 (pp. 1-6).