

HEALTH INSURANCE PREDICTION USING PYTHON FOR DATASCIENCE AND MACHINE LEARNING TECHNIQUES

Mr. Avinash B Bhuriya Anshu

Assistant Professor, Assistant Prof. Dept. of MCA,D.S.C.E

Article Information

Received : 11 June 2025
Revised : 13 June 2025
Accepted : 14 June 2025
Published : 17 June 2025

Corresponding Author:

Avinash B Bhuriya Anshu

Abstract— This Wellbeing safety net providers keep up enormous databases containing data on medicinal administrations used by petitioners, regularly crossing a few human services administrations and suppliers. Appropriate utilization of these databases could encourage better clinical and regulatory choices. In these informational collections, there exists numerous inconsistent divided occasions, for example, emergency clinic visits. Be that as it may, information mining of fleeting information and point procedures is as yet a creating research territory and separating helpful data from such information arrangement is a difficult undertaking. In this paper, we built up a period arrangement information mining way to deal with anticipate the quantity of days in clinic in the coming year for people from a general protected populace dependent on their protection guarantee information.

Keywords: *Logistic regression, Machine Learning, AI*

Copyright © 2025: Mr. Avinash B Bhuriya Anshu , This is an open access distribution, and reproduction in any medium, provided Access article distributed under the Creative Commons Attribution License the original work is properly cited License, which permits unrestricted use.

Citation: Mr. Avinash B Bhuriya Anshu , “HEALTH INSURANCE PREDICTION USING PYTHON FOR DATASCIENCE AND MACHINE LEARNING TECHNIQUES”, Journal of Science, Computing and Engineering Research, 8(04), April 2025.

I. INTRODUCTION

Introduction These information contains tables of medical clinic confirmation regulatory records and emergency clinic strategy claims, just as essential demorealistic data of clients. Client socioeconomics incorporate data identified with clients, for example, sex, age, the sort of Biomedical Informatics. Medical coverage records incorporate information fields of essential client's name, age, sex, Renewal, past illnesses, client's record number and a few other data. Protection methodology cases incorporate data identified with the system conveyed during an emergency clinic confirmation, similar to date and kind of strategies just as data on the related expenses.

Each case is identified with a medical clinic confirmation. Since the source factors contain different information types, for example, dates, numeric, content, they should have been pre-prepared before they could be utilized for displaying. Numeric factors were kept in their numeric organization. Every single all out factor was specified, i.e., swapped by numbers to make the entire element lattice numeric and rationing figuring memory. For some of the unmitigated factors, paired highlights were likewise extricated by creating a different section for every one of the classes, as a classification pointer. Some extra highlights were likewise produced utilizing explicit estimations.

Body of Paper Insurance company has several models, used for analyzing, understanding and predicting the cost of health insurance. The prediction needs several statistical techniques as well. In ground analysis of our project survey, we ought to let you know [1] the use of logistic regression for health insurance claim prediction. we used data set, which consist of 1337 records and fields explained below:

1. Age: policy holder's age[3]
2. Sex: policy holder's gender(male=1,female=0)
3. BMI(Body Mass Index): It specifies the body measurements like height and weight
4. Children: Number of children or dependents of policy holder
5. Pre-disease: It will show us previous disease of policy holder
6. Region: Locality/area of policy holder
7. Charges: Medical costs of individual which is billed by health insurance
8. Insurance claim: It provides information whether insurance has been claimed for individual is valid or not(valid=1,invalid=0) Currently, we will import pandas to pursue information from a CSV document and for further use. We will likewise utilize numpy to change over out

information into a configuration appropriate to bolster our characterization model. We will utilize seaborn and matplotlib for graphical representations. Apparently we will import Logistic Regression calculation from sklearn. This calculation will enable us to assemble our characterization model. Ultimately, we will utilize joblib which is accessible in sklearn to spare our model for sometime later[1]. We have data set information spared [1] in a CSV document called insurance.csv. We previously read our dataset in a pandas dataframe called insuranceDF, and after that utilization the head() capacity to demonstrate the initial five records from our dataset.

When utilizing AI calculations we should constantly part our information into a preparation set and test set. (In the event that the quantity of analyses we are running is huge, at that point we can ought to isolate our information into 3 sections, to be specific – preparing set, advancement set and test set). For our situation, we will likewise separate certain information for manual cross checking[1]. The informational collection comprises of record of 1337 patients altogether.

To prepare our model we will utilize 1000 records. We will utilize 100 records for testing, and the last 10 records to cross check our model[1]. Next, we separate the mark and highlights (for both preparing and test dataset). Notwithstanding that, we will likewise change over them into NumPy exhibits as our AI calculation process information in NumPy cluster design[1].

As the last advance before utilizing AI, we will standardize our information sources. AI models regularly advantage generously from information standardization. It additionally [1] makes it simpler for us to comprehend the significance of each component later, when we'll be taking a gander at the model loads. We'll standardize the information with the end goal that every factor has 0 mean and standard deviation. The expanded expense of medical coverage is disturbing all through the world. These expenses are accomplished for purchasers and bosses supported medical coverage premium which has expanded by 131 percent in the course of the most recent decade. A noteworthy reason for this expansion is installment mistakes made by the insurance agencies while handling claims. Besides, due to the installment mistakes brings about re-preparing [1] of the cases which is referred to be called as re-work and records for huge bit of regulatory expense and administrations issues of wellbeing plan which have an immediate effect in the term of fiscal of the insurance agency paying pretty much than what it ought to have. The best sort of AI calculations is those that mechanize a basic leadership forms by summing up from known models. Insurance agencies apply various models for breaking down and anticipating medical coverage cost. A portion of the work explored the prescient displaying of medicinal services cost utilizing a few factual

procedures. AI approach is likewise utilized for foreseeing staggering expense consumptions in human services[1].

Logistic Regression In this venture, we will talk about the utilization of Logistic Regression to anticipate the protection guarantee ,We take an example of 1338 information which comprises of the accompanying highlights:- Presently we will import pandas to peruse our information from a CSV document and control it for further use. We will likewise utilize numpy to change over our information into an organization reasonable to bolster our order model. We'll utilize seaborn and matplotlib for perceptions.

We will at that point import Logistic Regression calculation from sklearn. This calculation will enable us to assemble our grouping model. Finally, we will utilize joblib accessible in sklearn to spare our model for sometime later[1]. `import pandas as pd` `import numpy as np` `import seaborn as sns` `import matplotlib.pyplot as plt` `from sklearn.linear_model import LogisticRegression[1]` We have our information spared in a CSV record called insurance.csv.

We originally read our dataset in a pandas dataframe called insuranceDF, and after that utilization the head() capacity to demonstrate the initial five records from our dataset[1]. When utilizing AI calculations we should constantly part our information into a preparation set and test set. (In the event that the quantity of analyses we are running is huge, at that point we can ought to isolate our information into 3 sections, in particular – preparing set, improvement [1] set and test set). For our situation, we will likewise separate out certain information for manual cross checking[1].

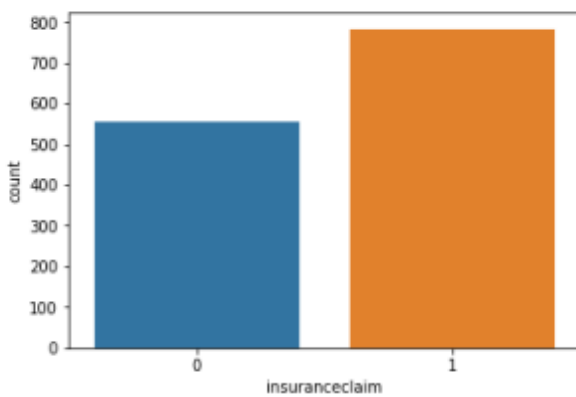
The informational collection comprises of record of 1338 patients altogether. To prepare our model we will utilize 1000 records. We will utilize 300 records for testing, and the last 38 records to cross check our model. Next, we separate the name and highlights (for both preparing and test dataset). Notwithstanding that, we will likewise change over them into NumPy clusters as our AI calculation process information in NumPy exhibit design[1].

As the last advance before utilizing AI, we will standardize our information sources. AI models frequently advantage significantly from info standardization. It additionally makes it simpler for us to comprehend the significance of each element later, when we'll be taking a gander at the model loads. We'll standardize the information with the end goal that every factor has 0 mean and standard deviation of 1[1].

We would now be able to prepare our arrangement [1] model. We'll be utilizing a machine straightforward learning model called calculated relapse. Since the model is

promptly accessible in sklearn, the preparation procedure is very simple and we can do it in few lines of code. To start with, we make an occasion called insuranceCheck and afterward utilize the fit capacity to prepare the model. `log=LogisticRegression()` `linear_regression.fit(x_train,y_train)` we need to the accuracy score of the model by importing the accuracy function given below: `fromsklearn.metrics import accuracy_score` After importing the accuracy_score we need to check the accuracy score for the both x_test and y_test which acts like an main standard leading attribute to portray the implemented logistic model projects the correctness of data set or not. `print(accuracy_score(y_test,y_pred))`

0.883177570093458



II. CONCLUSION

In this analysis, we have chosen the logistic regression model for predicting[4] the accuracy of the input data set of health insurance. This model accepts each and every regards of data set to present the actual output from expected one. By researching and consistent work through this model gave us accuracy score about 87.68% which helped to drive the successive step of this model. It led us productive barrier.

REFERENCES

- [1]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1955009, 13 pages, 2022. <https://doi.org/10.1155/2022/1955009>.
- [2]. D. Sathyanarayanan, T. S. Reddy, A. Sathish, P. Geetha, J. R. Arunkumar and S. P. K. Deepak, "American Sign Language Recognition System for Numerical and Alphabets," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10369455.
- [3]. J. R. Arunkumar, Tägele berihun Mengist, 2020" Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier" *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.
- [4]. Ashwini, S., Arunkumar, J.R., Prabu, R.T. et al. Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network. *Soft Comput* (2023). <https://doi.org/10.1007/s00500-023-09480-3>
- [5]. J.R.Arunkumar, Dr.E.Muthukumar," A Novel Method to Improve AODV Protocol for WSN" in *Journal of Engineering Sciences*" ISSN NO: 0377-9254 Volume 3, Issue 1, Jul 2012.
- [6]. R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [7]. Arunkumar, J. R. "Study Analysis of Cloud Security Challenges and Issues in Cloud Computing Technologies." *Journal of Science, Computing and Engineering Research* 6.8 (2023): 06-10.
- [8]. J. R. Arunkumar, R. Raman, S. Sivakumar and R. Pavithra, "Wearable Devices for Patient Monitoring System using IoT," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 381-385, doi: 10.1109/ICCES57224.2023.10192741.
- [9]. S. Sugumaran, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [10]. R. S. Vignesh, V. Chinnammal, Gururaj.D, A. K. Kumar, K. V. Karthikeyan and J. R. Arunkumar, "Secured Data Access and Control Abilities Management over Cloud Environment using Novel Cryptographic Principles," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199616.
- [11]. Syamala, M., Anusuya, R., Sonkar, S.K. et al. Big data analytics for dynamic network slicing in 5G and beyond with dynamic user preferences. *Opt Quant Electron* 56, 61 (2024). <https://doi.org/10.1007/s11082-023-05663-2>
- [12]. Krishna Veni, S. R., and R. Anusuya. "Design and Study Analysis Automated Recognition system of Fake Currency Notes." *Journal of Science, Computing and Engineering Research* 6.6 (2023): 16-20.
- [13]. V. RamKumar, S. Shanthi, K. S. Kumar, S. Kanageswari, S. Mahalakshmi and R. Anusuya, "Internet of Things Assisted Remote Health and Safety Monitoring Scheme Using Intelligent Sensors," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199766.
- [14]. R. S. Vignesh, R. Sankar, A. Balaji, K. S. Kumar, V. Sharmila Bhargavi and R. Anusuya, "IoT Assisted Drunk and Drive People Identification to Avoid Accidents and Ensure Road Safety Measures," 2023 International Conference on Advances in Computing, Communication and Applied

- Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10200809.
- [15]. I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856.
- [16]. G. Karthikeyan, D. T. G, R. Anusuya, K. K. G, J. T and R. T. Prabu, "Real-Time Sidewalk Crack Identification and Classification based on Convolutional Neural Network using Thermal Images," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1266-1274, doi: 10.1109/ICACRS55517.2022.10029202.
- [17]. R. Meena, T. Kavitha, A. K. S, D. M. Mathew, R. Anusuya and G. Karthik, "Extracting Behavioral Characteristics of College Students Using Data Mining on Big Data," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10084276.
- [18]. S. Bharathi, A. Balaji, D. Irene, J. C. Kalaivanan and R. Anusuya, "An Efficient Liver Disease Prediction based on Deep Convolutional Neural Network using Biopsy Images," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1141-1147, doi: 10.1109/ICOSEC54921.2022.9951870.
- [19]. I. Chandra, G. Sowmiya, G. Charulatha, S. D, S. Gomathi and R. Anusuya, "An efficient Intelligent Systems for Low-Power Consumption Zigbee-Based Wearable Device for Voice Data Transmission," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083856. I. Chandra, K. V. Karthikeyan, R. V, S. K, M. Tamilselvi and J. R. Arunkumar, "A Robust and Efficient Computational Offloading and Task Scheduling Model in Mobile Cloud Computing," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084293.
- [20]. Revathi, S., et al. "Developing an Infant Monitoring System using IoT (INMOS)." International Scientific Journal of Contemporary Research in Engineering Science and Management 6.1 (2021): 111-115.
- [21]. R. K, A. Shameem, P. Biswas, B. T. Geetha, J. R. Arunkumar and P. K. Lakineni, "Supply Chain Management Using Blockchain: Opportunities, Challenges, and Future Directions," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023, pp. 1-6, doi: 10.1109/ICI60088.2023.10421633.
- [22]. J.R.Arunkumar. "Comprehensice Analysis of Security Issues in Cloud Computing Technologies", Journal of Science, Computing and Engineering Research, 6(5), 06-10, June 2023.
- [23]. S. Sugumaran, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [24]. I. Chandra, K. V. Karthikeyan, R. V, S. K, M. Tamilselvi and J. R. Arunkumar, "A Robust and Efficient Computational Offloading and Task Scheduling Model in Mobile Cloud Computing," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084293.
- [25]. R. S. Vignesh, A. Kumar S, T. M. Amirthalakshmi, P. Delphy, J. R. Arunkumar and S. Kamatchi, "An Efficient and Intelligent Systems for Internet of Things Based Health Observance System for Covid 19 Patients," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084066.
- [26]. DC Jullie Josephine, J Sudhakar, T Helan Vidhya, R Anusuya, G Ramkumar, "An Improved Multi class Breast cancer classification and Abnormality Detection based on Modified Deep Learning Neural Network Principles", Deep Learning in Biomedical Signal and Medical Imaging, CRC Press, Taylor and Francis, 2024.
- [27]. R. Anusuya, Pragya Vashishtha, "Real Automatic Number Plate Image Detection With Yolo Algorithms", Journal of Science, Computing and Engineering Research, 7(7), July 2024.
- [28]. K. Shetty, S. Tyagi, A. Jha, D. N. M. K. Rao, J. R. Arunkumar and L. R, "Natural Language Processing in Strategic Planning Analysis," 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), Lucknow, India, 2024, pp. 1-5, doi: 10.1109/IC3TES62412.2024.10877514.
- [29]. S. Sugumaran, C. Geetha, S. S, P. C. Bharath Kumar, T. D. Subha and J. R. Arunkumar, "Energy Efficient Routing Algorithm with Mobile Sink Assistance in Wireless Sensor Networks," 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ACCAI58221.2023.10201142.
- [30]. R. S. Vignesh, A. Kumar S, T. M. Amirthalakshmi, P. Delphy, J. R. Arunkumar and S. Kamatchi, "An Efficient and Intelligent Systems for Internet of Things Based Health Observance System for Covid 19 Patients," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICECONF57129.2023.10084066.
- [31]. Jullie Josephine DC, Sudhakar J, Helan Vidhya T, Anusuya R, Ramkumar G. 15 An Improved Multi. Deep Learning in Biomedical Signal and Medical Imaging. 2024 Sep 30:237.
- [32]. Arunkumar, J.R., Anusuya, R., Chilukuri, P., Ramkumar Prabhu, M. (2024). Secure Data Transfer and Deletion Using Secure Encryption Algorithm in Cloud Computing. In: Singh, N., Bashir, A.K., Kadry, S., Hu, YC. (eds) Proceedings of the 1st International Conference on Intelligent Healthcare and Computational Neural Modelling. ICIHCNN 2022. Advanced Technologies and Societal Change. Springer, Singapore. https://doi.org/10.1007/978-981-99-2832-3_84

- [33].G. Manoharan, P. D. Sawant, J. Vanitha, M. Lourens, R. Anusuya and I. Bhati, "Cognitive Computing for HR Decision-Making," 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), Lucknow, India, 2024, pp. 1-5, doi: 10.1109/IC3TES62412.2024.10877480.
- [34].S. Sivakumar, R. Anusuya, V. Nagaraju, L. P. Narendruni and R. Thamizhamuthu, "QoS Based Efficient Link and Consistent Routing in Wireless Sensor Network," 2023 *International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bengaluru, India, 2023, pp. 1241-1246, doi: 10.1109/IITCEE57236.2023.10091080.

