Proceedings of the Third International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT-2025)
IEEE Xplore Part Number: CFP25CV1-ART; ISBN: 979-8-3315-2754-9

2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 979-8-3315-2754-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/IDCIOT64235.2025.10914851

Prediction of Insurance Claims for Health Sector using Machine Learning Techniques

Pragatheeswari E
Assistant Professor
Department of Computer Technology
Kongu Engineering College
Perundurai, Erode, India
epragatheeswari@gmail.com

Kavitha M N
Assistant Professor
Department of Computer Science and Design
Kongu Engineering College
Perundurai, Erode, India
kavithafeb1@gmail.com

Saranya S S
Assistant Professor
Department of Computer Science and Engineering
PSG Institute of Technology and Applied Research
Coimbatore, India
saranya.sowndarajan@gmail.com

Dhanushree D
Department of Computer Technology
Kongu Engineering College
Perundurai, Erode, India
dhanushree20052015@gmail.com

Jasmitha S
Department of Computer Technology
Kongu Engineering College
Perundurai, Erode, India
jasmithasaravanaraj@gmail.com

Abiradhie P V
Department of Computer Technology
Kongu Engineering College
Perundurai, Erode, India
abiradhieanitha@gmail.com

***Abstract*— Nowadays, health issues play a tremendous role in day-to-day life and the medical expenditure to get treatment becomes more difficult for the ordinary people. Health insurance has become a vital aspect of people's lives. In this massive community, to access healthcare services such as insurance policies, LIC, ICICI, HDFC ERGO, Star Healthcare are benefits for claiming an amount for their medical expenses. The dataset encompasses three categories, such as generalized data, hospitalized data and claim data. Each record in the dataset represents an individual's health insurance charges along with corresponding demographic and health-related characteristics. In this paper, Machine Learning techniques are used to group the claim amount that has been adopted by the individuals based on the hospitalized expenditure and insurance charges paid by the individuals. In the health sector, machine learning techniques are used to forecast insurance claims by building predictive models from historical data on claims, patient demographics, treatment types, and results. By predicting future claims based on trends seen in the data, these models help insurance firms better manage risk and allocate resources. In case of a fresher to this concept, it feels challenging to understand whereas generating as a pattern can be more flexible to acquire knowledge about how much amount has been claimed.**

***Keywords* — K-Means Clustering, Agglomerative Clustering, DBSCAN Clustering, Gaussian Mixture Clustering, Mean Shift Clustering, Machine Learning.**

I. Introduction

Healthcare is one of the prominent aspects for ordinary people. People face more problems for claiming the insurance amount as well as they don't have knowledge about how much money has been provided by the insurance company. Health insurance claims fall into two categories: reimbursement and cashless. A health insurance claim is the exchange of money between an insurance company and a medical provider with the enthusiasm of acquiring medical expenditure for a patient to acquire surgery, pharmaceutical, or other medical services. Utilizing machine learning strategies to forecast the amount of claims that an individual will receive based on their hospital expenses and the insurance they have accounted towards healthcare. By incorporating scientific expertise onto the illicit activity that raises accuracy, lowers false alarms, and strengthens general prevention functionalities in health insurance claims. [8] The research optimizes insurance claim to examine and use the resources by utilizing data mining strategies like model regression. With a Mean Absolute Error (MAE) indicating its durability in evaluating the quantities of health insurance claims and helping insurance companies with aggressive fiscal strategy and risk analysis. Gradient Boosting Regressor (GBR) manage intricate data relationships, evaluate the significance of features, and improve regress task completion [5].

Since everyone demands access to healthcare and the COVID-19 pandemic has become rigorous, a robust and transparent medical insurance system is necessary. Fewer research has been done about the claim amount to be provided based on medical expenditure and the charges. In early studies, researchers perform minimal machine learning algorithms and apply block

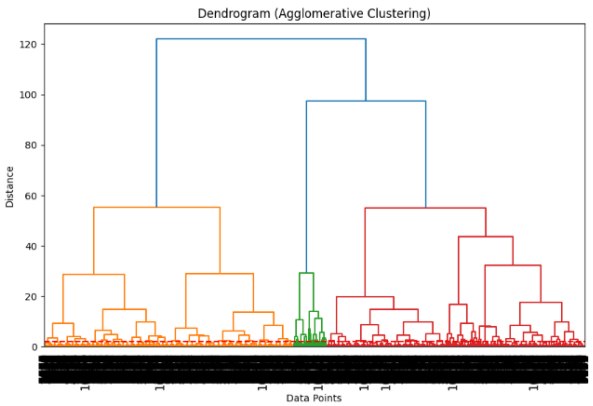

Fig.4. Dendogram for Agglomerative

Fig.5. Agglomerative Clustering

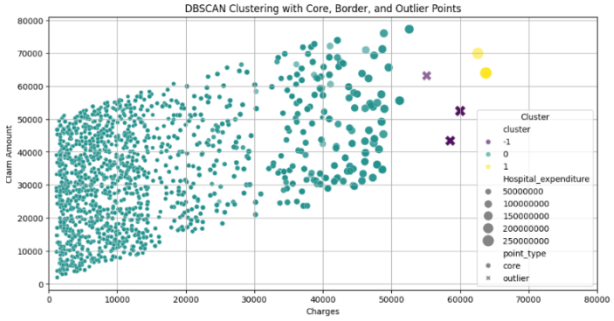

Fig.6. DBSCAN Clustering with outliers

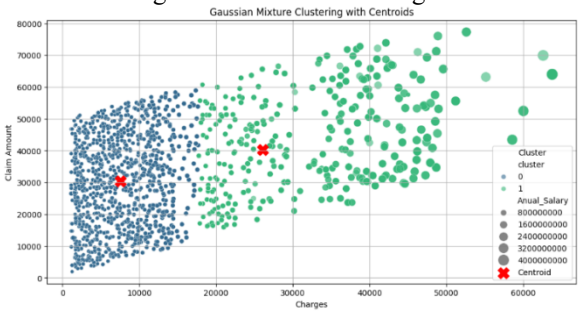

Fig.7. Gaussian Mixture with Centroid

Fig.8. Mean Shift Clustering

VI. CONCLUSION & FUTURE SCOPE

Hybridized algorithms in medical insurance can enhance risk assessment, fraud detection, and policy recommendations by merging machine learning techniques like clustering with deep learning and ensemble methods. These approaches offer more accurate, robust solutions and improve decision-making and client satisfaction. Future developments may involve real-time analytics and advanced feature selection to further optimize results. Subsequent studies may investigate the creation of more complex and hybrid models that incorporate several machine learning methods for even higher accuracy. User feedback loops and ongoing advancements in real-time learning and adaptation will further optimize.

REFERENCES

[1] A. Alnuaimi, A. Alshehhi, K. Salah, R. Jayaraman, I. A. Omar, and A. Battah, "Blockchain-Based Processing of Health Insurance Claims for Prescription Drugs," *IEEE Access*, vol. 10, pp. 118093–118107, 2022, doi: 10.1109/ACCESS.2022.3219837.

[2] K. Bhatia, S. S. Gill, N. Kamboj, M. Kumar, and R. K. Bhatia, "Health Insurance Cost Prediction using Machine Learning," in *2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India: IEEE, May 2022, pp. 1–5. doi: 10.1109/INCET54531.2022.9824201.

[3] S. Chowdhury, P. Mayilvahanan, and R. Govindaraj, "Optimal feature extraction and classification-oriented medical insurance prediction model: machine learning integrated with the internet of things," *Int. J. Comput. Appl.*, vol. 44, no. 3, pp. 278–290, Mar. 2022, doi: 10.1080/1206212X.2020.1733307.

[4] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement," *IEEE Access*, vol. 8, pp. 58546–58558, 2020, doi: 10.1109/ACCESS.2020.2983300.

[5] K. Dutta, S. Chandra, M. K. Gourisaria, and H. Gm, "A Data Mining based Target Regression-Oriented Approach to Modelling of Health Insurance Claims," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India: IEEE, Apr. 2021, pp. 1168–1175. doi: 10.1109/ICCMC51019.2021.9418038.

[6] I. Fursov *et al.*, "Sequence Embeddings Help Detect Insurance Fraud," *IEEE Access*, vol. 10, pp. 32060–32074, 2022, doi: 10.1109/ACCESS.2022.3149480.

[7] R. Y. Gupta, S. S. Mudigonda, and P. K. Baruah, "A Comparative Study of Using Various Machine Learning and Deep Learning-Based Fraud Detection Models For Universal Health Coverage Schemes," *Int. J. Eng. Trends Technol.*, vol. 69, no. 3, pp. 96–102, Mar. 2021, doi: 10.14445/22315381/IJETT-V69I3P216.

[8] M. E. Haque and M. E. Tozal, "Identifying Health Insurance Claim Frauds Using Mixture of Clinical Concepts," *IEEE Trans. Serv. Comput.*, vol. 15, no. 4, pp. 2356–2367, Jul. 2022, doi: 10.1109/TSC.2021.3051165.

[9] N. Jajam, N. P. Challa, K. S. L. Prasanna, and C. H. V. S. Deepthi, "Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction," *IEEE Access*, vol. 11, pp. 93111–93128, 2023, doi: 10.1109/ACCESS.2023.3304669.

[10] K. Kapadiya *et al.*, "Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: an Analysis, Architecture, and Future Prospects," *IEEE Access*, vol. 10, pp. 79606–79627, 2022, doi: 10.1109/ACCESS.2022.3194569.

[11] K. Kaushik, A. Bhardwaj, A. D. Dwivedi, and R. Singh, "Machine Learning-Based Regression Framework to Predict Health Insurance Premiums," *Int. J. Environ. Res. Public. Health*, vol. 19, no. 13, p. 7898, Jun. 2022, doi: 10.3390/ijerph19137898.

[12] A. Kumar, R. Krishnamurthi, A. Nayyar, K. Sharma, V. Grover, and E. Hossain, "A Novel Smart Healthcare Design, Simulation, and Implementation Using Healthcare 4.0 Processes," *IEEE Access*, vol. 8, pp. 118433–118471, 2020, doi: 10.1109/ACCESS.2020.3004790.

[13] I. Matloob, S. A. Khan, and H. U. Rahman, "Sequence Mining and Prediction-Based Healthcare Fraud Detection Methodology," *IEEE Access*, vol. 8, pp. 143256–143273, 2020, doi: 10.1109/ACCESS.2020.3013962.

[14] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, "A Sequence Mining-Based Novel Architecture for Detecting Fraudulent Transactions in Healthcare Systems," *IEEE Access*, vol. 10, pp. 48447–48463, 2022, doi: 10.1109/ACCESS.2022.3170888.

[15] E. Nabrawi and A. Alanazi, "Fraud Detection in Healthcare Insurance Claims Using Machine Learning," *Risks*, vol. 11, no. 9, p. 160, Sep. 2023, doi: 10.3390/risks11090160.