

Disease Prediction Using Machine Learning

Akash C. Jamgade, Prof. S. D. Zade

Student, Dept. of Computer Science and Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur, Maharashtra, India

Professor, Dept. of Computer Science and Engineering, Priyadarshini Institute of Engineering & Technology, Nagpur, Maharashtra, India

Abstract - One such implementation of machine learning algorithms is in the field of healthcare. Medical facilities need to be advanced so that better decisions for patient diagnosis and treatment options can be made. Machine learning in healthcare aids the humans to process huge and complex medical datasets and then analyze them into clinical insights. This then can further be used by physicians in providing medical care. Hence machine learning when implemented in healthcare can lead to increased patient satisfaction. In this paper, we try to implement functionalities of machine learning in healthcare in a single system. Instead of diagnosis, when a disease prediction is implemented using certain machine learning predictive algorithms then healthcare can be made smart. Some cases can occur when early diagnosis of a disease is not within reach. Hence disease prediction can be effectively implemented. As widely said "Prevention is better than cure", prediction of diseases and epidemic outbreak would lead to an early prevention of an occurrence of a disease. This paper mainly focus on the development of a system or we could say an immediate medical provision which would incorporate the symptoms collected from multisensory devices and other medical data and store them into a healthcare dataset. This dataset would then be analyzed using K-mean machine learning algorithms to deliver results with maximum accuracy.

Key Words: Big Data, healthcare, Machine learning, K-mean algorithm, etc.

1. INTRODUCTION

Disease prediction using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. Many works have been applied data mining techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of disease. Also, some approaches try to do prediction on control and progression of disease. The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little pre processing and produce more accurate results. With the development of big data technology, more attention has been paid to disease prediction from the perspective of big data analysis; various researches have been conducted

by selecting the characteristics automatically from a large number of data to improve the accuracy of risk classification rather than the previously selected characteristics.

The main focus is on to use machine learning in healthcare to supplement patient care for better results. Machine learning has made easier to identify different diseases and diagnosis correctly. Predictive analysis with the help of efficient multiple machine learning algorithms helps to predict the disease more correctly and help treat patients.

The healthcare industry produces large amounts of healthcare data daily that can be used to extract information for predicting disease that can happen to a patient in future while using the treatment history and health data. This hidden information in the healthcare data will be later used for affective decision making for patient's health. Also, this areas need improvement by using the informative data in healthcare.

One such implementation of machine learning algorithms is in the field of healthcare. Medical facilities need to be advanced so that better decisions for patient diagnosis and treatment options can be made. Machine learning in healthcare aids the humans to process huge and complex medical datasets and then analyze them into clinical insights. This then can further be used by physicians in providing medical care. Hence machine learning when implemented in healthcare can lead to increased patient satisfaction. The k-mean algorithm is used to predict diseases using patient treatment history and health data.

2. EXISTING SYSTEM

Prediction using traditional disease risk model usually involves a machine learning and supervised learning algorithm which uses training data with the labels for the training of the models. High-risk and Low-risk patient classification is done in groups test sets. But these models are only valuable in clinical situations and are widely studied. A system for sustainable health monitoring using smart clothing by Chen et.al. He thoroughly studied heterogeneous systems and was able to achieve the best results for cost minimization on the tree and simple path cases for heterogeneous systems.

The information of patient’s statistics, test results, and disease history is recorded in EHR which enables to identify potential data-centric solutions which reduce the cost of medical case studies. Bates et al. propose six applications of big data in the healthcare field. Existing systems can predict the diseases but not the subtype of diseases. It fails to predict the condition of people.

The predictions of diseases have been non-specific and indefinite

3. PROPOSED SYSTEM

In this paper, we have combined the structure and unstructured data in healthcare fields that let us assess the risk of disease. The approach of the latent factor model for reconstructing the missing data in medical records which are collected from the hospital. And by using statistical knowledge, we could determine the major chronic diseases in a particular region and in particular community. To handle structured data, we consult hospital experts to know useful features.

In the case of unstructured text data, we select the features automatically with the help of k-mean algorithm. We propose a k-mean algorithm for both structured and unstructured data.

3.1 The k-means algorithm

The k-means algorithm is a simple iterative method to partition a given dataset into a specified number of clusters, k. This algorithm has been discovered by several researchers across different disciplines. The algorithm operates on a set of d-dimensional vectors, $D = \{x_i \mid i = 1, \dots, N\}$, where $x_i \in R^d$ denotes the *i*th data point. The algorithm is initialized by picking k points in R^d as the initial k cluster. Techniques for selecting these initial seeds include sampling at random from the dataset, setting them as the solution of clustering a small subset of the data or perturbing the global mean of the data k times.

4. SYSTEM ARCHITECTURE

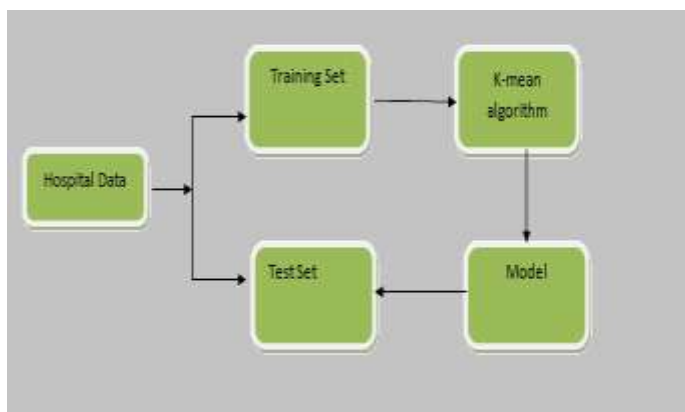


Fig -1: System Architecture

5. CONCLUSION

With the proposed system, higher accuracy can be achieved. We not only use structured data, but also the text data of the patient based on the proposed k-mean algorithm. To find that out, we combine both data, and the accuracy rate can be reached up to 95%. None of the existing system and work is focused on using both the data types in the field of medical big data analytics. We propose a K-Mean clustering algorithm for both structured and unstructured data. The disease risk model is obtained by combining both structured and unstructured features.

ACKNOWLEDGEMENT

I express my sincere gratitude towards my guide of Prof. S. D. Zade for their constant help, encouragement and inspiration throughout the project work. Also I would like to thank the Head of Computer Science and Engineering Department Dr. P. S. Prasad for him valuable guidance , ability to motive me and even willingness to solve difficulty made it possible to make my project unique and made task easier. My sincere thanks to Principal, Dr. V. M. Nanoti for providing me necessary facility to carry out the work.

REFERENCES

- [1] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [2] K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna, "Performance comparison of three data mining techniques for predicting kidney disease survivability", *International Journal of Advances in Engineering & Technology*, Mar. 2014.
- [3] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", *International Journal of Pure and Applied Mathematics*, 2018.
- [4] Boshra Brahmi, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", *Journals of Multidisciplinary Engineering Science and Technology*, vol.2, 2 February 2015, pp.164-168.
- [5] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, "Incorporating temporal ehr data in predictive models for risk stratification of renal function deterioration," *Journal of biomedical informatics*, vol. 53, pp. 220–228, 2015.
- [6] S. Patel and H. Patel, "Survey of data mining techniques used in healthcare domain," *Int. J. of Inform. Sci. and Tech.*, Vol. 6, pp. 53-60, March 2016.