IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

Multiple Disease Prediction System

Yash Gavande, Komal Mahale, Devendra Pitaliya
Department of Computer Science
Suman Ramesh Tulsiani Technical Campus-Faculty of Engineering, Pune, Maharashtra, India

Abstract: *For diagnosis and treatment for a patient, the medical facilities need to get advanced. So that a patient can make a better and wise decision. Machine learning algorithms are already implemented in this field of healthcare. Machine learning in the healthcare industry sucks humans to process huge, complex, and compound datasets. Then this can further be used by physicians in providing medical care. This will increase human satisfaction. In this, we try to implement different functionalities of machine learning in healthcare in a single system and also implement an appointment booking system. Instead of diagnosis, when a disease prediction is implemented using a certain machine learning predictive algorithm then healthcare can be made smart. As widely "Prevention is better than cure", Prediction of an epidemic outbreak would lead to early prevention of an occurrence of a disease*

Keywords: Disease Prediction, Appointment Booking, Decision Tree Algorithm, Random Forest Algorithm, XG-Boost Classifier

I. INTRODUCTION

1.1 Overview

For the past decades by applying data mining and machine learning techniques has been an ongoing struggle for disease prediction using patient history and health data. From medical data and profile, prediction of specific diseases has been done by applying data mining techniques. This approach has tried to predict the reappearance of the 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. Disease prediction has gotten more attention with the development of big data technology. The main focus is on using machine learning in healthcare to supplement patient care for better results [1]. Machine learning makes it easier to predict and identify different diseases and recognize them correctly. Large amounts of healthcare data is produced by the healthcare industries and it can be used for extracting information for prediction of disease. Hence machine learning techniques have been a successful implementation towards the healthcare system and it leads to patients' satisfaction.

1.2 Motivation

A major challenge facing healthcare organizations (hospitals, medical centres) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data [12]. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?"

1.3 Problem Definition and Objective

Any disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

is to predict with few attributes and faster efficiency, the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affect the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. healthcare.

1.4. Project Motivation and Limitation

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modelling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions Limitations. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial. Clinical decisions are often made based on doctor's intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affect the quality of service provided to patients. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

All these algorithms were found suitable for our system. We got higher accuracy with (Decision Tree classifier, Random Forest, XGB-classifier). We trained this model on datasets that we found on Kaggle. We found 300+ records of patients of that particular disease. We consider many statistical calculations and come to some good accuracy. The system has the ability to find out a serious patient which is infected by disease.

Disease Name	Algorithms	Test Accuracy	Data sets
Kidney Disease	Decision Tree Classifier	100%	Kaggle
Breast Cancer	Random Forest	96.49%	Kaggle
Liver Disease	XGB Classifier	88%	Kaggle

II. METHODOLOGIES FOR PROBLEM SOLVING

The system aims to ensure that no patient should have to wait for a doctor at hospital if he has minor symptoms of any diseases. It can be used to prioritise patients based on Symptoms Effects. Our system helps patients to predict diseases as early as possible, even if there is no doctor available. Patient does not need to worry at all.

This system is used to diagnose diseases from the user's Symptoms. The system has a trained machine learning model that will help patients to find out whether they have any symptoms of that particular disease (0-100 in present). To do so, the system takes the data from the user about the symptoms, this data will be analysed by system and mined through dataset model algorithms such as:

1. Decision Tree Cclassifier,
2. Random Forest,
3. XGB classifire.

Moreover, In advance our system has finality to book appointments online through a website.

The project uses Python, Flask Framework, HTML, CSS, JavaScript And jQuery. We had performed an extensive study on many of the ML algorithms such as:

1. Linear Regression.
2. Logistic Regression.
3. Decision Tree Algorithm.
4. Support Vector Machine Algorithm.
5. XGboost.
6. Randon Forest.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

III. ALGORITHM

A. Support Vector Machine:

"Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it ismostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot). Support Vectors are simply the coordinates of individual observation. The SVM classifier is a frontier that best segregates the two classes (hyper-plane/ line).

B. XG Boost Algorithm:

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree-based algorithms are considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years.

C. Linear Regression:

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

IV. CONCLUSION

Understanding the buying intention of customers on e-commerce sites. Under- standing the satisfaction of the customer with specific products through online reviews on e-commerce sites. Recommending products and places to customers based on collective opinions.

REFERENCES

[1]. A. Lawrence, "Five Customer Retention Tips for Entrepreneurs," Forbes, 2012. [Online]. Available: http://www.forbes.com/sites/alexlawrence/2012/11/01/five-customerretention-tips-for-entrepreneurs/210fc5e1 7b0b. [Accessed: 12-Apr2016].

[2]. C. Shaw, "15 Statistics That Should Change the Business World – But Haven't," Linked in, 2013. [Online]. Available:https://www.linkedin.com/pulse/20130604134550-284615-15-statisticsthat-should-change-the-busines s-world-but-haven-t. [Accessed:12-Apr2016].

[3]. "The case for customer service training," White House Office of Consumer Aft- fairs, 2009. [Online]. Available: http://aspiremarketing.com/the-case-for-customer- satisfaction/. [Accessed: 12-Apr-2016].

[4]. J. Hauser and E. Daham, "New Product Development," in Marketing Manage- mint: Essential Marketing Knowledge and Practice, R. Grover and N. K. Mahlo- tar, Eds. McGraw Hill, Inc., Columbus Ohio, 2007.

[5]. M. D. Earle and R. L. Earle, Creating New Foods. The Product Developer's Guide. The New Zealand Institute of Food Science Technology (Inc.), 2009.

[6]. J. McAuley, et al.," Image-based recommendations on styles and substitutes," in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 43-52, 2015.

[7]. X. Ning, et al.," A comprehensive survey of neighborhood-based recommend- dation methods," in Recommender systems handbook, ed: Springer, pp. 37-76, 2015.

[8]. R. He and J. McAuley," Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in proceedings of the 25th international conference on world wide web, pp. 507-517, 2016.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

[9]. Bao, Jia, et al." Recommendations in location-based social networks: a survey." Geoinformatics 19.3, pp. 525-565, 2015.

[10]. Wang, Hao, Nahyan Wang, and Did-Yan Yeung. " Collaborative deep learning for recommender systems." Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015.