International Journal of Recent Advances in Multidisciplinary Topics
Volume 2, Issue 7, July 2021
https://www.ijramt.com | ISSN (Online): 2582-7839

Hospital and Doctor Recommendation Using Machine Learning and NLP

K. S. Sindhu[1*], H. B. Pramod[2]
[1]*Student, Department of Computer Science, Rajeev Institute of Technology, Hassan, India*
[2]*Associate Professor, Department of Computer Science, Rajeev Institute of Technology, Hassan, India*

***Abstract*: The health care industry collects data from the patient's database by applying data mining and machine learning techniques is used to predict the disease where disease is the leading cause for the human's death from the past years. Medical profiles or pathological data where it has been extracted from the data mining technique has been used to predict the specific disease like heart disease, cancer disease, diabetes disease and tried this approach to predict the recurrence of disease. Nowadays these three diseases are common in people's health where they are suffering badly in a critical condition to live their life.**

***Keywords*: Hospital, doctor recommendation, machine learning and NLP**

1. Introduction

Health plays a major role in a person's life leading a peaceful life but people get into many diseases due to lack of nutrients and food. In modern technology we are developing an app where we will be predicting diseases and recommending hospitals and leading physicians based on patient reviews. Patient satisfaction is one of the most relevant indicators of doctors and hospitals where they care for quality and individual patient reviews are critical to delivering the best results. Many health care providers will download patient inputs and analyze patient updates and will personally collect data from doctors 'offices, clinics, and hospitals and will record patient experiences to evaluate doctors' performance and services and hospital management. Data collection is analyzed using a random forest algorithm with neighbors near K (KNN) where a problem approaches with a specific question of analysis and finding a solution between two or more independent variables and dependent variables. They will do research and count the responses they received from patients and change the percentage depending on the hospital services or management.

Disease creates serious health problems in people's lives. To overcome this problem, we are developing new ideas in this app for predicting diseases based on sets of data collected from patients. We are focused on three diseases such as heart disease, cancer and diabetes where these three diseases are commonly found in human health and present major challenges with a critical situation. Often, scientists do research on these diseases to find a solution. We also try to solve this problem by using new ideas through a machine learning process. We will be collecting data on a patient about a particular disease and giving recommendations to hospitals and doctors to consult. For example, it is necessary to record details of the most important symptoms and health practices that catch heart disease when it leads to heart attack and stroke. Various tests should be performed before heart disease is diagnosed, including auscultation, ECG, blood pressure, cholesterol and blood sugar.

According to cancer, various tests will be done depending on the type of cancer and the most important symptoms will be recorded. Lastly, it is diabetes, and there are many tests to be taken such as fasting plasma glucose (FPG) and random plasma glucose (RPG). Based on this study we will guess, what disease does the patient have? And according to that particular disease we will guess the best and most appropriate hospital for the patients we help easily to admit to it.

2. Literature Survey

[1]Title: Hospital Consumer Assessment of Healthcare Providers and Systems Author: Elliot et al [HCAHPS] is a standard diagnostic tool used by many hospitals to evaluate patients' experiences. This data is provided by the Hospital Consumer Assessment of Healthcare Provider and Systems database, sponsored by the U.S. health research agency. Medicaid and Medicare service centers use points from the Hospital Consumer Assessment for health care providers and hospital rehabilitation programs. Providing high quality care is directly related to hospital revenue and many hospitals are looking for ways to improve the patient experience and achieve high HCAHPS scores. This study provides a brief description of the satisfaction questions in the HCAHPS testing tool and the categories they enter. The research questions are divided into six sections where each section has multiple choice questions. For example, the section "physician care" measures patient satisfaction and physician care (s) using three questions about physician respect, listening, and explanation. Each question has four options (Never, Sometimes, Often, and Always).

[2] Title: Investigated the relationship between postoperative morbidity and mortality and patients' perspectives of care in

*Corresponding author: sindhu.ks76@gmail.com

surgical patients. Author: Sheetz et al In their article, the satisfaction points used were used along with the registered Registry of the Michigan Collaborative Clinic as a measure of patient care ideas. A few studies have examined the relationship between one satisfaction question and one or more patient information.

[3] Title: conducted an analysis of racial/ethnicity in patients' perceptions of inpatient care. Author: Goldstein et al. Using retrospect, they concluded that non-Spanish whites on average tend to go to hospitals that offer a better patient experience for all patients compared to hospitals commonly used by African American, Hispanic, Asian / Pacific Islander, or multiracial patients..

[4].Title: explored the relation between the willingness to recommend the hospital and other satisfaction identifiers. Author : Klink Enberg This paper finds that hospitals that focus on resources to improve aspects of care such as nurses and physician respect, respect, obedience, room hygiene, etc. This paper does not look at patient data. The literature in the HCAHPS database analysis is largely driven by hypothesis and only looks at specific aspects of patient satisfaction or census. On the contrary, the approach expressed in this paper does not imply that you have any ideas. Instead, we conduct data-driven experimental analyzes that examine all aspects of patient satisfaction and demographic and patient statistics.

3. System Design and Architecture

System Architecture - indicates the total hypermedia structure of the Web App. The building design is tied to the objectives established by the Web App, the content to be presented, the users to visit, and the established travel philosophy. Content design, focuses on content and design to present and navigate.

W Design, it talks about how the system is a structured system for managing user communications, managing internal processing tasks, result navigation, and content presentation. WebApp development is defined within the context of the development environment in which the app will be used.

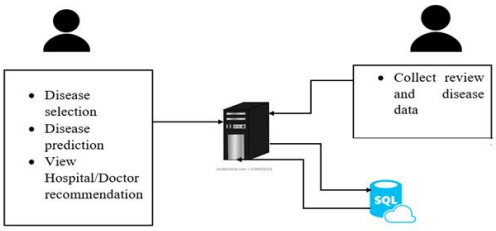

Fig. 1. System architecture

1) Data prepossessing

Data preprocessing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn; therefore, it is extremely important that we preprocess our data before feeding it into our model.

2) K-Nearest Neighbor (KNN)

The K-Nearest Neighborhood is one of the easiest ways to learn mechanically based on a directed learning process. The K-NN algorithm matches the similarity between a new case / data and available cases and puts a new case in the same category as the available categories. The K-NN algorithm stores all available data and separates the new data point accordingly. This means that where new data emerges it can then be easily categorized into a functional category using the K-NN.K-NN algorithm can be used for decomposition and fragmentation but mainly for partition problems. K-NN is a non-parameter algorithm, which means it makes no sense in basic data thinking. It is also called an algorithm for lazy learner because it does not learn from set-up training faster than it retains a database and during classification, performs an action on the database. The KNN algorithm in the training phase simply stores the data again when it receives new data, and then separates that data into a section very similar to the new data.

Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category. Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset.

3) Natural Language Processing

The NLP field is back in a few decades and has grown significantly over the years. Originally banned from collecting data from a limited set of computerized documents, the arrival of the World Wide Web saw an explosion of information in many different languages. Much work has been done in the Retrieval (IR) field of knowledge which is considered the Natural Language Processing domain application. Before discussing IR strategies, for several, let us examine the theoretical and practical aspects of NLP. Cultural Approach - Key Concepts Initially, the NLP approach followed these different steps. 1. Pre-Processing Document / Token Softening 2. Clinical Analysis 3. System Analysis 4. Semantic Analysis 2.1.1. Pre-creation / token production the first challenge is to divide the given document into words and sentences.

The word token - originally limited to teaching linguistic theory - is now the same as dividing text into words. Many languages use white space as a delimiter but it can be deceptive in some languages. While it seems straightforward, the challenges include separating words like 'I am' to 'I am' and deciding whether I should split a token like 'high-impact' into two words. Continuing to create a problem becomes the language of the text. The unicode level is very helpful as each character is given a unique value and therefore makes it possible to decide on the original language. Another concept used by NLP experts is often "Normal Disclosure (RE)". And to get its roots in the concept of computer programming language, RE

specifies the cable format that needs to be viewed. For example, a password string (token) that can contain capital letters will be specified as [A-Z] while string counting numbers will be specified as [0-9].

4. Conclusion

In the early days in hospitals they had to do various tests and had to wait one day for all related laboratory reports where it leads to abortion and efficiency, accuracy is low. We have tried the program effectively to reduce size, eliminate unnecessary data, increase learning accuracy and improve understanding of outcomes and focus on the three most common diseases in human health and we also predict hospitals and leading physicians for a different and more effective way for patients to get hospitals with good medical care. All in all we are using our new ideas to provide benefits to people suffering from health problems and can use this app where they will get all the best options at once for one complaint. The opinions expressed by people in hospitals and doctors play an important role and they can easily make a decision. The aim was to use such organizations to build patient satisfaction in accordance with the hospital's recommendations system

References

[1] UCI Machine Learning Repository." [Accessed: 21-Apr2018].
[2] W. Bergerud, "Introduction to logistic regression models with worked forestry examples: biometrics information handbook no. 7," no. 7, pp. 147, 1996.
[3] S. Sperandei, "Lessons in biostatistics Understanding logistic regression analysis," Biochem. Medica, vol. 24, no. 1, pp. 12– 18, 2014.
[4] J. R. Quinlan, "Induction of Decision Trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.
[5] T. M. Mitchell, "Decision Tree Learning," Machine Learning. pp. 52–80, 1997.
[6] L. Breiman, "Random Forest," pp. 1–33, 2001.
[7] M. Denil, D. Matheson, and N. De Freitas, "Narrowing the Gap: Random Forests InTherein, M., Matheson, D., & De Freitas, N. (2014). Narrowing the Gap: Random Forests in Theory and In Practice. Proceedings of The 31st International Conference on Machine Learning, (1998), 665–673. Retrieved from ht," Proc. 31st Int. Conf. Mach. Learn., no. 1998, pp. 665–673, 2014.
[8] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," Sch. EECS, Washingt. State Univ., pp. 1–13, 2006.
[9] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods," vol. 22, no. 2, pp. 103–104, 2000.
[10] Mohammad Soleymani, Member, IEEE, SadjadAsghari-Esfeden, Student member, IEEE Yun Fu, Senior Member, IEEE, MajaPantic, Fellow, IEEE, "Analysis of EEG signals and facial expressions for continuous emotion detection" IEEE Transactions on Affective Computing, volume 7, Issue :1, pg 17- 28, March 2016.