.

Smart Health Predictions Using Data Mining

Name: Guy-Erne Assika

Institution: Griffith College Dublin

Table of Contents

1. Introduction……………………………………………………………….4

1.1. Aims and goals of the project proposal……………………………5

1.2. Reasons for conducting the research………………………………5

1.3. Literature Review…………………………………………………..5

1.4. Original value introduced by the project……………………………6

2. Evidence……………………………………………………………………….7

2.1 One survey……………………………………………..…………….7

2.2Brainstorming Session……………….…………………………….…8

3. Proposed System……………………………………………………………….8

4. Methodology………………………………………………………………….9

5. Ethical considerations…………………………………………………………10

6. Business plan…………………………………………………………………..10

6.1Costing and Resource estimation………………………………………10

6.2 Initial Investment……………………………………………………..11

6.3Return of investment…………………………………………………..11

7. Conclusion and possible future directions………………………………………12

**Abstract**

One of the motivating areas of research that is becoming famous in all health organizations is data mining. Several data mining techniques are presented in this project concerning medicinal and instructive parts of Clinical predictions, educational fields, and other various aspects. Due to the regulation and availability of computers in medical and health areas, there is the presence of large amounts of data. According to modern technology, there are noticeable improvements that have been made in the computer field, making it possible not to deal with large amounts of data at the same time. The primary objectives are to evaluate data mining techniques in clinical and healthcare applications aiming at the development of accurate decisions. In the computer world, technology is believed to be robust and famous. The technology is a subfield of computer science whereby through different databases, the already existing data is used whereby it is transformed into new research and results. It makes use of database management, artificial intelligence, and machine learning, where new patterns of data are extracted from the knowledge associated with the database and all large databases available. Data is obtained automatically using this technology. Different parameters are used in data mining, which includes forecasting, clustering, predictive analysis, and path analysis.

Smart Health Prediction Using Data Mining

**1. Introduction**

Data mining is an essential and motivating area of research to find useful information from a large set of data. The technology is becoming popular in healthcare sectors because there is a need for an analytical methodology that is efficient in detecting valuable information and detection of unknown information in robust data. Different benefits are associated with data mining, including fraud detection in health centers, especially the insurance sector (Kelati, Plosila, & Tenhunen, 2019). It also ensures there is the availability of a solution in the medical industry where patients will get medicine at a low cost. Organizations always consider having a powerful tool that helps in analyzing and extracting meaningful information from the massive set of data. Another benefit associated with data mining is the detection of different causes of diseases and identifying the best medical treatment for any disease. Data mining helps the researchers and healthcare managers to reconstruct systems for drugs that are recommended and provide the health care policies required by the health industry. The researchers also aim at giving out the best health profiles for the patients for future reference.

Additionally, data provided by the health organization is complicated in such a way that it is difficult to analyze it. The hard analysis makes it difficult to make a crucial decision concerning the patients' health. Much information in the data consists of medical claims, the patients available, costs of treatment, and various hospitals that are available. The extensive knowledge pushes for the urge to develop a powerful tool for extracting and analyzing critical information from a vast data set. The analysis of health data improves the excellent performance of different healthcare organizations concerning patient management tasks. Data mining technology has excellent outcomes that benefit the healthcare organizers in grouping patients that are affected by the same disease or have similar health issues so that it could be more comfortable in providing health services and treatment (Kelati, Plosila, & Tenhunen, 2019). It also helps the organizers to determine the length of stay of the patients and making plans for effective information system management together with a medical diagnosis.

Cost-effectively, the current technology helps in improving medical services in the medical field. Various factors that are responsible for diseases, especially food and working areas, are also analyzed using data mining technology. Some of the health prediction systems recognize the automation of therapeutic information that helps in upgrading the administration of health services, clinical care, training, and medical analysis (Kelati, Plosila, & Tenhunen, 2019). With the highest significance of data, privacy, and respectability are provided and employed by the doctors through the planned system. For instance, it is easier for a doctor who is examining several patients to operate efficiently since the data and information he or she has is precise. Conscience and there is a clear and quantity of records.

* 1. **Aims and goals of the project proposal**

The main objective and aim are to evaluate data mining techniques in all clinical and healthcare applications which helps in the development of accurate decisions.

* 1. **Reasons for conducting the research**

Since data mining is famous and influential, the information is useful for offering a close exchange of medicinal information and handling strategies that helps in improving various parts of clinical predictions. This kind of technology is the latest technology of interest in the world of computers, which uses the existing data in different databases to give out new research and results.

* 1. **Literature review**

Apriori Algorithm is presented by researchers using different machine learning tools, for instance, Weka. According to researchers, the new improvised method called an improved algorithm that is used to the advantages of the Apriori Algorithm. Different angles have been presented by researchers that help in studying substantial data sets and obtain extensive and useful information. The method helps in detecting diseases and giving out a remedy for ailments. Other researchers have the objective of determining various data mining methods that provide decisions and a piece of precise information concerning medicine. Multiple angles of clinical prediction are improved through the data mining process. The association rule is presented by researchers, which give knowledge on data sets that provide the results and help in the calculation of frequent itemsets.

The technology is a concept that studies large amounts of data and extracts patterns that can be converted to necessary knowledge. An efficient algorithm is identified for the project, which provides data mining result. Different applications for the medical sector are created to fulfill the required information by using all the predictive data mining techniques. An example of the algorithms is the Naïve Byes algorithm that helps in getting the various data items and compares them with the present algorithm. There is enough scope used for improving the health solutions applied to medical data. If used for the right purpose, electronic health records and other historical medical data work much and well. The complex and massive data generated by the health care sector consists of details of the patients, methods used for diagnosis, diseases that affect the patients, and all the electronic information concerning the patients.

* 1. **Original value introduced by the project**

The trends of all patients are introduced to the healthcare organization through data mining. The privacy of every patient is valued, and through this, one can analyze the different healthcare organizations to attend. The tendency of data mining applications brings about great information and becomes a great necessity for healthcare organizations. Through various perspectives and connections, the healthcare organizers can give predictions on the trends presented by several patients according to the conditions and behavior. Different contagious diseases have been discovered with the advance of data collection. Data concerning the patients is collected and stored in organized forms, which help in forming an information system for the hospitals. Hidden pattern investigation is provided from the datasets obtained through data mining. The patterns are used by physicians to get a diagnosis of several diseases and their medication.

1. **Evidence**
   1. **One survey**

A research was conducted by two researchers Ranjani and Durairaj, that aimed at making a detailed study report of two different types of data mining applications in healthcare organizations while reducing the study of the data transactions in healthcare. Different mining techniques and methodologies are applied for extracting knowledge from the database generated in the healthcare industry. They also presented the existing methods of mining data and the different algorithms used in mining. Various tools that were more valuable for mining the data which are more relevant to the healthcare industry were presented. Different techniques of mining data were performed by two researchers, which include clustering, classification, association, and regression in a strong domain, together with their pros and cons. The survey had highlights concerning applications, challenges, and future issues that would be brought by data mining in healthcare organizations.

Through the use of the Naïve Byes algorithm, different models are developed that are used to assign class labels to problem instances. From a finite set, the class labels are made. The algorithm is guided by a specific principle that states that the value of each future of all naïve byes classifiers. For instance, to identify that fruit is an orange, it has to be yellow, round, and measures up to ten centimeters, which is the principle of the naïve byes algorithm. Through supervised learning training, the naïve algorithm is efficiently trained, adding to the probability models used.

**2.2 Brainstorming session**

Naïve Byes algorithm has several steps on how it conducted. First, the texts are divided into different segments when the first one starts from the element presented at the zero indexes, the second at index one, and the process continues. Secondly, the first character is compared with the segment taken correspondingly. If there happens to be a match, the next step is made, and if there is a mismatch, the next portion is taken, and step two is repeated. Thirdly, the last character of the pattern is compared with the previous figure of the corresponding segment, and if they match, the next step is taken. If they do not match, the next section is taken, and step two is repeated. A comparison of characters is made for the rest of the characters with the patterns and the characteristics of the segments taken. A comparison is stopped when mismatch has been detected, and the next section is made, which is passed through process two. If the patterns match, then the process continues with other segments through process two to achieve another design.

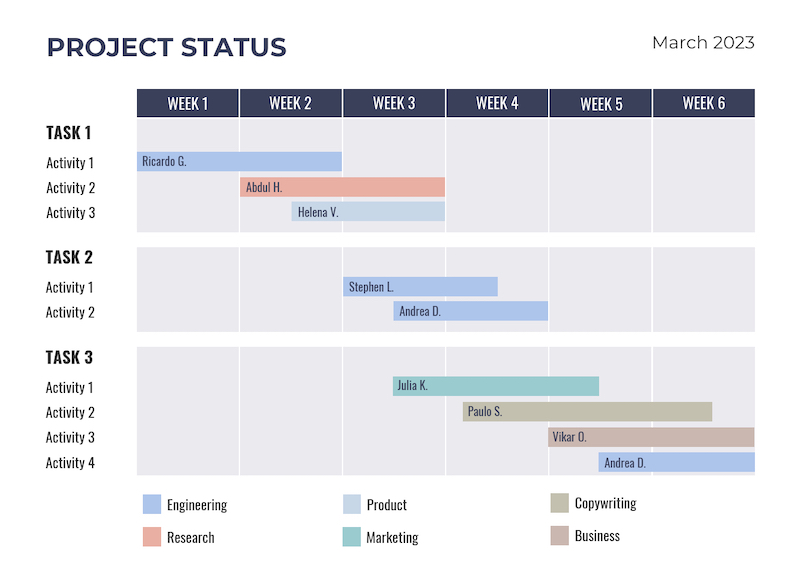
1. **Proposed System**

A smart health prediction system needs to be created, which helps in building a specialist network of smart health prediction framework. The system majorly aims at improving the tasks of all specialists in the healthcare units. It helps at examining a patient at the start level and records the disease present. Information concerning the patient is immediately recorded, and the framework distinguishes the sickness and allocates the required specialist for the patient. Few questions are asked to the patient when the structure is not sure about, and several tests are done to the patient, and later the results are presented. Depending on the database of the patients, the precise disorder associated with the patient is performed, and the required records are made. The framework majorly helps the patients and specialists in receiving the quickest solution possible.

1. **Methodology**

The main objective of this project is to develop a web application using data mining concepts through the use of Java Server Pages. This is done to predict the disease through the information and knowledge acquired. There are steps carried out to predict a disease for inputted patient symptoms. Data is gathered and selected from the World Wide Web, where the necessary information is acquired for processing. Data is then transformed into Xls file, forming a standard base where it is stored in a database after extraction (Tan, Ji, Lim & Tseng, 2017). From Xls file, only the required information is read that is associated with the fundamental objective of the necessary application. Apriori and FP growth algorithms are used for the prediction of the patient's disease for the required symptoms, which are provided by the user as input. After inputs are accepted, and execution of the algorithm takes place through the access of the data created during the stage of reprocessing. The bottom-up approach allows for an extension of one item at a time (Tan, Ji, Lim & Tseng, 2017). It is designed to operate on a database that contains transaction. The Apriori algorithm is implemented and generated by the use of one candidate set. The implementation aims at creating a motive that predicts only one database management system that is relational.

Additionally, another method of mining data for health prediction is by use of the FP Growth Algorithm, which stands for constant pattern growth. By use of a Gantt chart, one is able to assign resources to the tasks and projects. Setting up a Gantt chat helps one to think through all the tasks involved in the project, whereby they are divided into manageable components. The resources here are people and the budget amounts. The technique is scalable and helps in frequent mining information in a database. The algorithm has got two procedure steps on how it should be handled. First, one should build a data structure called FP- a tree that is compact, and secondly, frequent items should be extracted directly from the FP-Tree. The following Gantt chart shows an example of how events are planned for the purpose of a project.



**5 Ethical considerations**

In this regard, an emphasis on the uniqueness and specialty of medical data mining which is one of the best and challenging areas in the application of data mining and discovery of knowledge. Challenges available are due to the hue and complexity of the data, which has got varying quality and time series. The datasets of healthcare organizations are fragmented and distributed in nature, making it difficult for data integration (Ienca, Ferretti, Hurst, Puhan, Lovis & Vayena, 2018). There are ethical issues that are related to data mining, whereby the specialists should be ethically fit to handle all the hospital processes. They should have ethical and legal clearance from a public and specialized hospital. While conducting the research, the issue of confidentiality should be ethically ft since it controls the stringent and regulations. The best ethical rules should be made to avoid discouraging of study and encourage good results from the research process. A significant prevalence of cross-field evaluation of health big and related data is a clear indication of a big data approach from multiple medical data approaches. Research indicates that clinical outcomes can be produced from medical data sources that are non-strict (Ienca, Ferretti, Hurst, Puhan, Lovis & Vayena, 2018). The big data retrieved makes it difficult for separation between different medical specialties and clinical intervention levels, which gives more chances for inter-specialty exchange. The high numbers of technical challenges that are addressed when assessing the big data that is health-related show the presence of several technical weaknesses and limitations, which mostly depend on the social and technological trend of the specific healthcare organization. Thus, to handle the technical problems that arise after research, there should be keen considerations made to the security, integrity, and quality of data that allows secure transmission and storage. There should be better security for the data infrastructure. Data cleaning tools should be available to give a guarantee of quality data.

1. **Business plan**

**6.1 Cost and resource estimation**

Data mining has got a different application in different sectors, including the field of the communication industry, financial data analysis, biological data analysis, and many others. A lot of data is being produced in the field of health informatics and organizations. It is hard to analyze big data in such sectors, and therefore, extensive knowledge is needed. For health diagnosis, the healthcare organization applies data mining and telecommunication techniques. To the patients who require a continuous check-up and needs the doctor's service immediately need to have stored data for retrieving their previous medical information. In the business world today, data mining technology is used to form several blogs for patients and physicians where monitored data such as sensor attached to patients is used for better diagnosis of the patients (Mohapatra, Patra, Mohanty & Pati, 2018). It also helps in continuous check-up of sensitive patients and frequently confirming their progress through data taken and stored for them.

**6.2 Initial investment**

Other business plans related to data mining and healthcare is educating students, tracking diseases, monitoring public health, and pursuing research. The system is also used as an inappropriate area to improve work efficiency and enhance decision making quality. The technology was also working by providing computer theories and tools that assist in extracting useful information from the growing number of digital data. The daily decisions are made concerning the needs of the doctors in their daily decision-making activities. Different patterns and models have identified that help in providing support for prediction and decision-making process for diagnoses and treatment plans.

**6.3 Return of investment**

The use of technology allows proper management of medical knowledge in the healthcare organization. Through the technology, the specialists can provide the trends of the customers and their behavior hence creating success in the business. The information is obtained through data analysis from different findings, connections, and various perspectives. Through the historical information collected from the stored files, unknown trends and patterns are into essential business solutions. Knowledge discovery involves the use of algorithmic means through which models are extracted and enumerated from the data (Mohapatra, Patra, Mohanty & Pati, 2018). Through online analytical processing and traditional statistical methods, data is easily mined through database management, statistics, and computer science, including machine learning and artificial intelligence.

**7 Conclusion and Possible future directions**

In the field of therapeutic space, data mining is believed to be crucial. The technology portrays the preposition of a crossover information mining model that separates arrangement learning for the guide of different situations in the clinical choice framework. A structure of the apparatus presents different devices utilized for investigation. The condition occurs when one needs assistance from the specialists, but they end up not being accessible due to unknown reasons. An online framework has been planned and organized to avoid problems, and several patients from different areas can access specialists. The online structure consists of essential parts such as quiet login, enter side effect in the system, recommended medication, and among others.

Different manifestations and complaints from the patients are presented through the application, and the introduced side effect is examined, giving expectations of fitting sickness. The indication given by the clients in the framework are considered, giving the symptoms they are experiencing. However, when specialists are needed more, they may not be available due to unknown reasons. The future scope explains that learning will be extracted from the correct information that is available in the framework. This is achieved by getting the datasets ready through the application of calculations of apriori. In case of failure of the structure, there should be adjustable changes that should be done. Through the Association Rule Mining, a database is contrasted, and the approaching questions and last report are produced. The exact and productive information is given concerning the patients making them get their solution in a second. More work should be done later after using a more informational index that is identified with any sickness. For better precision and expectation, the dataset should be of quality and well organized. It should be free from individual cases and missing characters that might cause inconveniences to the specialists. The system can also be accessed by other people to make android apps where the locally available doctors could be referred to the patients. The software created for data mining may be extended using other links where medicines may be purchased online. Other features may be added, like detecting the causes of some diseases that can be added to the website. The software owns features that give more concern to the patients and mostly concerned with the things that the patients should avoid during the illness period.

References

Anaya, L. S., Alsadoon, A., Costadopoulos, N., & Prasad, P. W. C. (2018). Ethical implications of user perceptions of wearable devices. *Science and engineering ethics*, *24*(1), 1-28.

Ienca, M., Ferretti, A., Hurst, S., Puhan, M., Lovis, C., & Vayena, E. (2018). Considerations for ethics review of big data health research: A scoping review. *PloS one*, *13*(10), e0204937.

Kelati, A., Plosila, J., & Tenhunen, H. (2019, August). Smart Meter Load Profiling for e-Health Monitoring System. In *2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE)* (pp. 97-102). IEEE.

Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. *Kidney research and clinical practice*, *36*(1), 3.

Mohapatra, S., Patra, P. K., Mohanty, S., & Pati, B. (2018, December). Smart Health Care System using Data Mining. In *2018 International Conference on Information Technology (ICIT)* (pp. 44-49). IEEE.

Pandey, S. C. (2016, October). Data mining techniques for medical data: a review. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)* (pp. 972-982). IEEE.

Rajasekar, M. R. (2019). Cloud-Centric IOT based Disease Diagnosis in Smart E-Health Prediction System. *Journal of the Gujarat Research Society*, *21*(16), 2214-2220.

Sujatha, R., & Nithya, A. (2016). A Survey of Health Care Prediction Using Data Mining. *International Journal of Innovative Research in Science, Engineering and Technology*, *5*(8), 14538.

Tan, K. H., Ji, G., Lim, C. P., & Tseng, M. L. (2017). Using big data to make better decisions in the digital economy.

Watson, K., & Payne, D. M. (2020). Ethical practice in sharing and mining medical data. *Journal of Information, Communication, and Ethics in Society*.