A Prediction Algorithm to Reduce Queuing Time using Parallel Patient Treatment Methodology

John[1], Vivia Mary[1], Gunda[2], Rishik Reddy[2], Mullapudi[2], Aravind[2], Yeleti[2], Naveen[2]

[1]Assistant Professor, [2]Student,
[1,2]Department of Information Technology, SRM University

Abstract

Inefficient management of the patients' queues is one of the major issues faced in medical institutions like clinics and hospitals which end up in creating large crowds at the hospital lobbies and an extended waiting time in the patients' treatments. Waiting unnecessarily for a long period of time, ends only in loss or wastage of time, human life and hospital resources. It also increases the number or frustrated patients that are waiting to get treatment required. Every single patient has to undergo a diagnosis and then be forwarded to other departments or medical personnel for further procedures. Therefore, each patient's waiting time is the time taken by the system to diagnose all the patients before him/her in the queue. In such a condition, the most practical decision would be to give out an efficient treatment plan to each patient. This can be implemented as a mobile application, wherein a predictable waiting time according to the diagnosis of the patient is uploaded, which then updates itself in real-time. Taking this into consideration, this paper proposes a Patient Treatment Time Prediction (PTTP) algorithm that can predict the time taken by a procedure for a particular patient. This algorithm can be applied to real-world scenarios and can be implemented in a large-scale environment. After predicting a treatment time necessary, the Hospital Queuing Recommendation (HQR) system can be developed. The job of calculating and predicting a convenient and an efficient treatment time for a particular patient can be done by the HQR system. The necessary input data for this is taken from a real world scenario like an actual doctor estimating time for a procedure at a particular hospital. This algorithm and system should work hand-in-hand generating responses of the utmost efficiency and very low latency. Once the model goes through extensive experimentation and simulation, an efficient model that demonstrates the effectiveness of this system can be recommended to other hospitals or medical institutions thus reducing waiting time per patient.

Keywords: Patient Treatment Time Prediction, Hospital Queuing Recommendation, Parallel Patient Treatment

1. Introduction

In the present day scenario, most of the medical institutions like hospitals and clinics are over-crowded because of the lack of medical resources and an inefficient management system of patient queues or waiting times. While the lack of resources cannot be solved with a singular method as resources vary from hospital to hospital, the management of the patient queues using the available resources in a hospital or a clinic is of the utmost importance. This process is, nevertheless, a demanding and a complicated task as each patient, after the initial diagnosis, undergoes various tests and procedures that are particular to their medical conditions – like an x-ray, a blood test, a scan, a biopsy, etc. Depending on these individual tasks, the total time required for a patient's treatment varies. This result in making the prediction of time required a very complicated task. There are also cases where the number of tasks undergone varies from patient to patient. Some procedures are dependent on the results of the previous procedures. Some procedures are independent meaning they can be completed in any random order or all can be done in parallel. The focus of this paper is to create an efficient and predictable time frame for a patient to complete all of their procedures without having to wait in long queues thereby reducing the crowds in a hospital lobby. This can allow hospitals to plan a treatment time frame for each patient based on their diagnosis.

a. Data Mining:

When certain details of data analysis are done by tools that discover patterns and relationships that were previously unknown and valid in substantial data sets, this technique is known as data mining. These analysis tools have the ability to encompass statistical models, analytical data, algorithms involving mathematical analysis and certain methods of machine learning like the decision trees or neural networks. The data mining techniques are strictly not restricted to collection of and managing the input data. It also includes the techniques of analysis of data and prediction of results. The reason data mining is used very often in today's advancements in the field of data analysis is for the identification of simple, valid, useful, novel and understandable interrelations and patterns in the existent data. The finding of such favorable relationships and patterns in the input data is put forth by different names such as knowledge extraction, information harvesting, information discovery, data archeology and pattern processing. In the database process, the process of Knowledge Discovery include many steps such as, data selection, data preparation, data cleaning and correct result interpretation obtained from the process of input data mining. These steps can be vital part of the information from which useful and profitable knowledge is derived.

b. Data Mining System Classification:

There are many criteria by which Data mining system can be categorized.

1. The type of data input that are mined: The input data received, managed and analysed forms this classification in the system. Textual data, Spatial input, time-series input data, multimedia data, etc are a few types which can be involved in the classification of the data mined.

2. The database of the system that is included: This type of classification has its base in the model of the database that is generated for the system. This can be relational database, the data warehouse, object-oriented database, transactional database, etc.

3. The discovered knowledge from the data: The criteria that forms this type of classification is derived from the functionalities that are a part of the data mining or the knowledge discovery process like classification, characterization, association, discrimination, clustering, etc.

4. The techniques involved in the mining process: The base of this type of classification is generated by the techniques or approaches used in the data analysis process. These approaches include neural networks, machine learning, genetic algorithms, visualization, statistics, datawarehouse oriented or database-oriented, etc.

c. The Process of Knowledge Discovery

The method of knowledge discovery technique from the input database is definitely one of the tasks that is involved in the process of Data Mining.
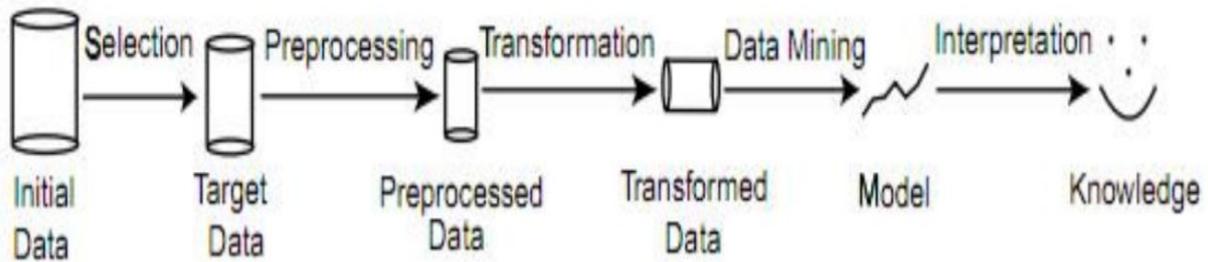
Fig. 1: Knowledge Discovery in Databases

The following are the steps taken in the knowledge discovery process:

1. Cleaning of Data: Cleansing of data is also another way of referring to the method of Datta Cleaning. The process involves the removal of noisy data, redundant data, irrelevant data, etc from the input data received.

2. Integration of Data: In this stage, the data that is collected from various sources, mostly heterogeneous, are united or put together to form a common source.

3. Selecting Data: Here, the setting on and the retrieval od the data that is pertaining to the analysis in the data collection is performed.

4. Transforming Data: Data strengthening or consolidation of data is also one way of referring to Data Transformation. This phase involves the process of transforming the chosen or selected data into different forms that are relevant for the procedure of data mining.

5. Data Mining: This is one of the very crucial steps in the knowledge discovery process in which versatile techniques are applied to result in potentially pragmatic patterns.

6. Pattern Evaluation: This is a step in which certain patterns which elicit interest and ones personifying knowledge are found with certain measures already mentioned.

7. Representation of knowledge: This is the last stage of Knowledge Discovery. here, the discovered knowledge is depicted to the user in the form of a visual output. This is an important step that portrays the results of data mining like originality and visualization techniques thereby helping user to bear in mind.

d. Life Cycle of the Data mining Process

There are six critical steps involved in the development life cycle of a data mining project.
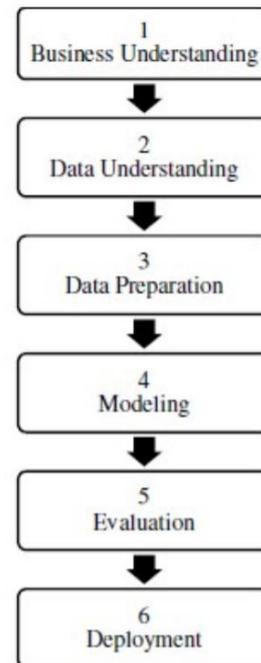
Fig. 2: Data mining life cycle

1. Understanding the Business: The focus of this aspect is based on the understanding of the objectives of the project involved and the requirements associated with it originating from a business perspective. This observation and the resultant knowledge obtained is then converted into the definition of a data mining problem and a preliminary plan is the designed through which the objectives are attained.

2. Understanding the Input Data: This process takes a start by collection of the initial and fundamental data. It can help the used to get familiar with the gist of the data, discover or identify any problems related to the quality of data, find out the initial insights

gained into the data or come across subsets that help in planning the hypothesis for the hidden information.

3. Preparing the data: This phase takes the raw data and finds the ultimate dataset by covering all activities leading to it.

4. Modeling the project: This is the phase where various techniques of modeling are selected, decided and applied to the project and the various parameters are subjected to optimum values.

5. Model Evaluation: This stage is one where the model generated in the previous phase is evaluated and reviewed. The business objectives are achieved by building the model accordingly by the steps executed.

6. Deployment of the evaluated model: The model has the purpose of amplifying the knowledge of the input data. The customer or client can use this knowledge gained only if it is formulated anf presented in a way the client can understand. This aspect of deployment can be a simple one as a report generation or it can also be a complicated result such as the implementation of a repeatable data mining process across a company enterprise.

e. Functionalities of Data Mining

The tasks involved in data mining are of two sorts:

1. Data mining tasks that are capable of identifying detailed property and characterizing the overall properties of the data that is existent.

2. Data mining tasks that is predictive or portending on try predictions according to the assumptions kept as the data that is available or accessible.

The functionalities of data mining and the knowledge range they discover are briefly conferred in the list as given below:

1. Characterization: Here, characteristic rules are produced by the summarization of the generalist attributes that we see of objects inside a target class

2. Discrimination: The phase of data discrimination is one that results are what we can call as discriminant rules. This is widely the inter-relation of the total attributes of objects betwixt by two classes very popularly known by target class and the contrasting class.

3. Association Analysis: The transactional database has the combinatorial occurrences of the prevalence of items that is studied and reported by the Association Analysis. Support is the threshold in accordance with which this occurrence happens and results in the frequent item sets being identified. Confidence is yet another threshold which can be used in pin-pointing association rules. This is also the conditional or codicillary probability of the appearance of an item inside a transaction just after the appearance of an alternative item.

4. Classification: This is a phase where the classes given are used for the organization of the data. This method makes use of addicted class labels for the ordering of the objects inside a data collection.

5. Prediction: This is usually used to refer to the forecasting or portending of certain numerical values that are missing or the increasing/decreasing of the time related data trends.

6. Clustering: This is the way of data management in classes. Here, uncharted class labels are put up and clustering algorithm accepts the responsibility of finding out classes that are acceptable. Unsupervised classification is another way of naming clustering as no definite class labels are dictated for the classification.

7. Outlier Analysis: When certain data elements are found not capable of being grouped within a given class or cluster, they are known as outliers. Exceptions or surprises are another way of referring to these outliers. Outlier's identification is one of the crucial steps in data mining.

8. Analysis of Evolution and Deviation: The time related data, also known as transient data, that changes with time at a rather quick rate can be studied with the regard of Evolution and Deviation Analysis. The evolutionary trends to data can be modified by evolution analysis. This consents to comparison, characterization, classification or clustering of the transient analogous of time-related data.

2. Literature Survey

Zhang, et al says that there have been many advances in the field of information technology related to healthcare technologies and procedures and have made great breakthroughs in many domains. However, these technologies particularly developed for the healthcare system accepts not only very large manufactured and medical data but also data that is beyond capability to handle or process. Moreover, the data that is used is collected from multiple devices in a short period of time. The characteristics or attributes of the data are collected and stored in multiple different formats. These formats are then treated simultaneously in parallel which can be regarded as a big data issue. To enable a situation that is convenient and efficient in healthcare, this paper brings forth a cyber-physical process for medical technology that is purely centered on patients, called Health-CPS. These products and services are imposed on technologies dealing with cloud and storage and massive data analytics. This technology contains a data collection layer restricted by a unified standard, a data management layer that is used for dispersed storage as well as a data oriented service layer. Zhang also mentions that as a result of the study, the technological advances on big data and cloud can help in the strengthening of the performance efficiency of the healthcare structure so that various useful healthcare applications and resources can be made available to patients in addition to services.

Kumar, et al says that a major part of human life is dominated hospitals and medical institutions. The increased health consciousness amongst people have raised the frequency and variety of patients visiting hospitals. In today's scenario the patients face dissatisfaction in the treatment provided by hospitals. The long queues and waiting time problem faced at the OPD services and products at most of the hospitals are a huge concern. For every patient that stands in the queue the waiting time each patient must face is the cumulative of the waiting time of every single patient before them. The consultation and extra tests following the diagnosis is the waiting time and a crucial reason behind patients in unavailing the products and services at the particular Hospital. More convenient and desirable move would be if every patient can be given a predictable consultation and waiting time along with a most productive and efficient treatment plan that is accessible on the mobile phone in the form of an application that updates itself in real time. This technique would enable a doctor to upload a patient's prescriptions and treatment on to a server database where the data is encrypted by the use of and AFS algorithm for high security. The patient can then view all relevant information related to their diagnosis on their mobile applications.

Patel and Sheshadri talk about requirement proactive health care and and care and care and and wholeness which results in the ever increasing cost for Healthcare and expanded Medical Health Insurance premiums. An archetype shuffle in the the Healthcare industry has been witnessed witnessed due to a recent Twist of digitising medical records. When talking about data intricacy, diversity and timeliness, a mega rise in the volume of data is witnessed by the Healthcare industry. Each and every possible way for lowering cost is being sought after by the medical experts. But this lowering of costs should not affect the improving care, process, delivery and management. In this condition, the only plausible or logical quick fix comes in the forms of big data with a promising approach to transform the health care e-commerce. An overall decrease decrease in the costs of Healthcare services and a positive positive services and a positive result in economic growth can be produced by a paradigm shift to proactive Healthcare Healthcare proactive Healthcare from a reactive Healthcare services services services. Vulnerabilities keep on growing and

security and privacy issues are at the limelight as emerging threats while the big data flexibility is harnessed by the Healthcare industry. This paper can depict Healthcare industry that focuses on the cutting edge issues issues of privacy and security in big data.

Vidyashree says that in today's scenario various domains of Healthcare Technologies have witnessed great progressions with the advancement of Information Technology. However a negative result faced is that these advancements in Technologies has made the medical data much bigger and very painstaking to handle and analyse. Moreover a big data problem that can be taken into regard is that since data requirement was created in a short period of time from a variety of devices, various formats are used to store the characteristics of the data. This paper proposes a cyber physical model for applications and services that are patient centric called health CPS to provide a more convenient service and Healthcare environment built on Big Data Analytics and cloud Technologies. This model of the system contains a layer of Data Collection built with a unified standard. The distributed storage and parallel computing is done by the data management layer. The model also consists of a data oriented service layer. Humans can enjoy a variety of smart Healthcare applications and services if the results of the conducted study so that the cloud and big data Technologies can be used in enhancing the Healthcare system performance.

Fidalgo - Merino at all says that in order for the incremental construction of binary regression trees, a new algorithm has been presented. SAIRT is this new algorithm which is when facing data streams that involve unknown dynamics is accustomed to the Induced model. The unknown dynamics mentioned previously include abrupt and gradual function drift, noise and virtual drift and when certain regions of one's function changes. Both symbolic and numeric attributes can be handled. In order for obtaining new patterns, their recommended algorithms have the capability of automatically altering its modelled structure and internal parameters. This is highly dependent on the data stream's current dynamics. SAIRT helps in monitoring the useful attributes of the nodes and may leave out examples that come out of certain regions that are selected; thereby storing the ones that are remaining in local windows that are associated with the leaves of the tree. Depending on our problems dynamics current methods showing regression desire a careful configuration on particular conditions. Then certain algorithms that are suggested procures better outputs than current algorithms while handling input data or streams implicating changes regarding levels of noise, differences of speed, example of Sampling distribution and changes in partial or unconditional underlying functions, we call it experimentation.

3. System Analysis and Design

1. Proposed System

The focus of the of the aforementioned paper is on helping hospitals to schedule every treatment procedure in a convenient and predictable predictable time period and help patients complete their treatment procedures without complete their treatment procedures without overcrowding hospital lobbies and formation of inefficient queues. The model of a patient treatment time consumption is developed and predicted by making use of massive realistic data derived from various hospitals. This realistic input data of a patient's are then analysed carefully, rigorously and moderately by the use of important parameters taken into consideration like the start time and end time of a patient's of a patient's and end time of a patient's of a patient's time of a patient's time and end time of a patient's of a patient's and end time of a patient's of a patient's time of a patient's treatment, the age of the patient admitted and the detailed contents of a patient's treatment treatment for every single different task. We estimate the condition of a patient and take into consideration the operations performed during the treatment and this estimation leads us to

identify various waiting times for particular patients. Historical data provided by hospitals of the various patients treatment is then is then used to train a patients treatment time prediction (PTTP) model. Taking a current queue patient treatment time prediction model then estimates that the waiting time of each patient standing in the queue is the total of all the patients waiting times before him. Hospital queuing recommendation recommendation system is then used to recommend an efficient and convenient plans of treatment of treatment according to the requested treatment task of each patient with the least waiting time for a patient.

2. New Methodology

a. Gathering input data that is derived from different procedures, tasks and strategies: medium sized Hospital Hospital has the capacity of 8000 to 12000 number of patients per day depending on certain statistics and figures 120000 to 2 lacs are the number of of number of of treatment records that is input in hospital hospital in hospital database. These inputs are collected from many different procedures conducted in a medical institution like medical examination of a patient, registration of admission, Drug Delivery, inspection, payment of bills and many other procedures. Different formats are used to store the input data for different treatment procedures.

b. Choosing the data dimensions that are identical: the input data generated inside a hospital for treatment procedures from various tasks are created in different contents and different formats in addition to the different dimensions. We have to decide on similar features of the input data in order to train the model for the time utilisation of a patient for each procedure conducted. The subject or patient information can be one such feature. The information can be a patient's card number, gender, age of the patient, etc. The information related to the treatment procedure and the time taken can be another feature. There are more innovative attributes of the treatment data which will not be taken for the reason that they are useless to the PTTP algorithm. These include data such as name of the patient, telephone number or address.

c. Calculation of feature variables that are new to the input data: various features that are importance to the input data has to be calculated in order to train the PTTP model. These attributes or features include the consumption of time taken by a patient for every treatment procedure, the time of the treatment like the day and the time range or treatment that is conducted.

3. Algorithms

a. SVM Algorithm: Challenges of classification and regression are pre owned by certain machine learning algorithms that are supervised called support vector machine. However these algorithms are more commonly useful in problems where classification is involved. The support vector algorithm states that every single input data item must be plotted as a point in a n dimensional space. Here n is taken as the number of features or attributes. The plotting is conducted with each feature value being a particular coordinate's value. Classification process is then carried out by the finding of a hyperplane that can differentiate the classes into two very well.

b. KNN Algorithm: All of the available and visible cases and the new cases that have been classified are stored by an algorithm in accordance with a measure of similarity called The k-nearest Neighbours algorithm. One example of the KNN algorithm is the distance function. An approach which was already nonparametric and apply in the probability valuation and the pattern recognition in this paper is the KNN algorithm.

4. Modules

Module 1: Patient Treatment Time Prediction
The data which is received in hospitals in real time is what is used as input to train the patient treatment time prediction model which

delivers the total of the waiting time of all the patients in the current queue.

Module 2: Patient laboratory test time prediction

Input data received from the clinical laboratory tests have been extracted which form the input data for testing done by patients and this data can be used as a proof of the concept. Q in this paper we have derived the test results from other test by predicting the time taken for a patients' test procedure using a variety of machine learning algorithms.

Module 3: Suggesting a Doctor to the Patient based on the Feedback

After the test time has been predicted based on reviews derived of all hospitals nearby the patient can get an idea of which hospital is better to upload and which doctor is better for diagnosis. This most done by the patient treatment time algorithm will be extremely helpful for every particular patient.

Table 1: Example of Treatment Records

Patient No	Gender	Age	Task Name	Dept. Name	Doctor Name	Time Elapsed
001	Male	15	Checkup	Surgery	Dr.Shyam	5 min
001	Male	15	Payment	Cashier	Null	Null
001	Male	15	CT Scan	CT-5	Dr.Raj	10 min
001	Male	15	M R Scan	MR-8	Dr.Li	7 min
001	Male	15	Take Medicine	Pharmacy	Null	3 min
...	...	...	...	...	...	...

Table 2: Format of the data for different Treatment Tasks

| Treatment Task	Format of the Data
Registration	{Patient Card number,Patient Name,Gender,Age, Telephone number, Address,Task name, Operation time}
Checkup	{Patient Card number,Patient Name,Gender,Age,Task name, Department,Doctor name,Doctor position,Time elapsed,Context}
Payment	{Patient Card number,Patient Name,Task name, Amount, Operation time}
Take medicine	{Patient Card number,Patient Name,Task name, Dispensary, Time of Compounding, Time of issue}
CT Scan	{Patient Card number,Patient Name,Gender,Age,Task name, Department,Doctor name,Body region of scans,Time elapsed,remarks}
Injection	{Patient Card number,Patient Name,Gender,Age,Task name, Department,Doctor name,Time elapsed,Drug name, Remark}
Blood test	{Patient Card number,Patient Name,Gender,Age,Task name,Department, Doctor name,Doctor position,Time of the blood test,Time of report}
...	...
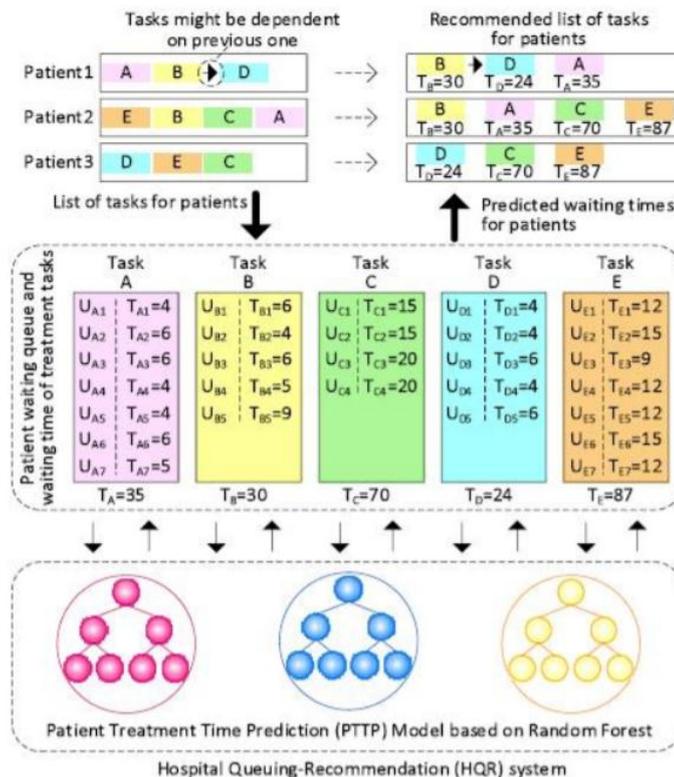
Fig. 3: Architecture Diagram
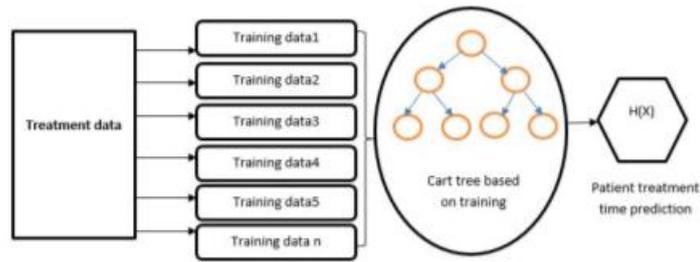
Fig. 4: Data Flow Diagram
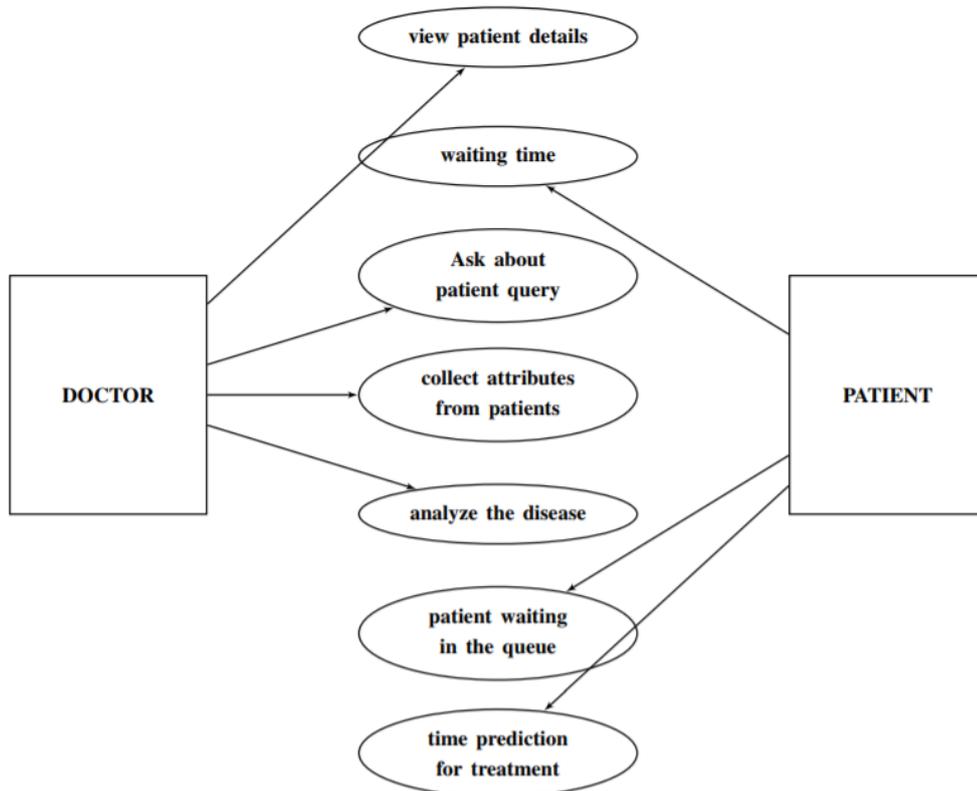
Fig. 5: Use Case Diagram
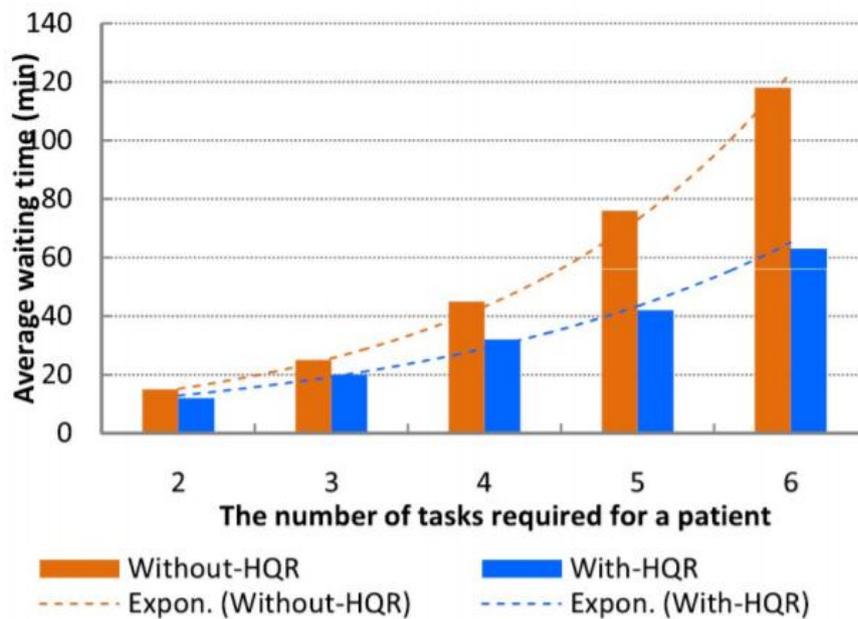
Fig. 6: Average waiting time for patients

4. Result

Using the technology of big data and Cloud Computing in this study we propose a PTTP algorithm. The algorithms that where further applied on the PTTP algorithm where the naive Bayes and the K nearest neighbours and support vector machine techniques. Prediction of the waiting time in a queue of every single treatment task can be done with the trained PTTP model. This follows the development of a hospital glowing recommendation system which has the ability to expect what the disease is in response to the symptoms diagnosed and the patient gets an efficient and a convenience treatment plan. Once the patient has completed the treatment in a particular doctor as a particular Hospital review or a feedback is taken and then used for the prediction of the best doctor with respect to certain aspects like the cost of treatment, the procedures done, the effectiveness of the treatment suggested, the waiting time taken, etc. This will in future anywhere any human who is suffering from any particular disease to search for the best

doctor in a hospital nearby based on the symptoms using an Internet connection from his home. After conducting extensive bouts of experimentation and verifying the results of the application so that the patient treatment time algorithm and the Hospital queuing recommendation system have achieved great physician with respect to its applicability and its performance has been deemed efficient.

The training data set was made by the collection and screening of large data from more than a few hospitals and laboratories and the algorithms were applied to the large data to delete the outliers. The big data techniques were applied to predict the disease based on the symptoms and the slots were booked to respective doctor with respective to the predicted time.

After the doctor's checkup, based on the prescription the laboratory tests slots were booked with respective to the predicted time. At the end a feedback form was given to the patient to get the feedback, based on that feedback the best doctor was suggested to the new patient.
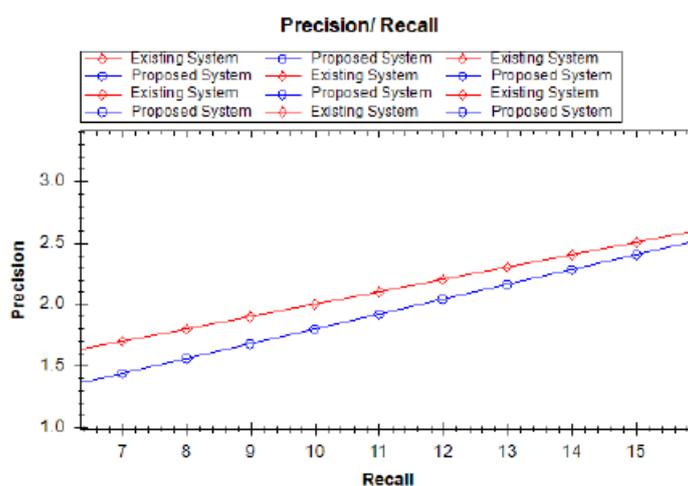

Fig. 7: Comparision of Existing System and Proposed System

5. Conclusion

The data volumes inside medical Institutions like hospitals and clinics are increasing on a day-to-day level. It is not necessary but the training of the data that is archived in the hospital database is a very heavy workload which requires a set of Hospital guide recommendations. For the future work of this paper it is suggested to take an incremental patient treatment time algorithm which is based on data streaming and Recommendation which is much more convenient which follows the minimised path awareness.

References

[1] Adomavicius, G. and Tuzhilin, A. (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE transactions on knowledge and data engineering, 17(6), 734–749.

[2] Ben-Haim, Y. and Tom-Tov, E. (2010). "A streaming parallel decision tree algorithm." Journal of Machine Learning Research, 11(Feb), 849–872.

[3] Chrysos, G., Dagritzikos, P., Papaefstathiou, I., and Dollas, A. (2013). "Hc-cart: A parallel system implementation of data mining classification and regression tree (cart) algorithm on a multi-fpga system." ACM Transactions on Architecture and Code Optimization (TACO), 9(4), 47.

[4] Dean, J. and Ghemawat, S. (2008). "Mapreduce: simplified data processing on large clusters." Communications of the ACM, 51(1), 107–113.

[5] Fidalgo-Merino, R. and Nunez, M. (2011). "Self-adaptive induction of regression trees." IEEE transactions on pattern analysis and machine intelligence, 33(8), 1659– 1672.

[6] Li, K., Tang, X., Veeravalli, B., and Li, K. (2015). "Scheduling precedence constrained stochastic tasks on heterogeneous cluster systems." IEEE Transactions on computers, 64(1), 191–204.

[7] Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014). "Data mining with big data." IEEE transactions on knowledge and data engineering, 26(1), 97–107.

[8] Yang, X., Guo, Y., and Liu, Y. (2013). Bayesian-inference-based recommendation in online social networks, Vol. 24. IEEE.

[9] Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M. J., Shenker, S., and Stoica, I. (2012). "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, USENIX Association. 2–2.