

# Survey on Parallel Patient Treatment Time Management System

Radha R\*, Dr. Ravi Kumar V

Department of CSE Computer Science and Engineering, Vidyavardhaka College of Engineering,  
Mysuru, Karnataka

DOI: <https://doi.org/10.21467/proceedings.1.67>

\* Corresponding author email: radhar261@gmail.com

## Abstract

Successful patient queue administration to limit quiet hold up deferrals and patient congestion is one of the significant difficulties looked by healing centers. Pointless and irritating sits tight for elongated stretches give rise to generous person asset and too much time consumption and increment dissatisfaction persisted by sick persons. Every patient in the line, the aggregate healing time of the considerable number of patients before him is the time that he should wait. It is helpful and best if the patients could get the most well-organized treatment and know the anticipated holding up time in the course of a versatile application that updates progressively. Hence, we recommend a Patient Treatment Time Prediction (PTTP) calculation to foresee the sitting tight period for every treatment undertaking for a patient. We utilize reasonable patient information from different healing facilities to acquire a patient treatment time display for each task. In light of this expansive scale, reasonable dataset, the treatment time for every patient in the present line of each task is estimated. In view of the anticipated holding up time, a Hospital Queuing-Recommendation (HQR) framework is produced. HQR computes and determine an efficient and advantageous treatment plan suggested for the patient. As a result of the huge size, reasonable dataset and the necessity for constant reaction, the PTTP calculation and HQR framework command effectiveness and little-dormancy reaction. We utilize an Apache Start based cloud execution at the National Supercomputing Center in Changsha to accomplish the previously mentioned objectives. Broad experimentation and reenactment comes about to show the viability and materialness of our proposed model to suggest a powerful treatment anticipate patients to limit their waiting times in hospitals.

Index Terms- Apache spark, big data, cloud computing, hospital queuing recommendation, patient treatment time prediction,RF algorithm.

## 1 INTRODUCTION

Presently, health organizations are overfull and need viable persistent queue administration. Persistent request line organization and hold up time desire make a testing and puzzled work in light of the fact that each patient may require particular stages/undertakings, for instance, a



© 2018 Copyright held by the author(s). Published by AIJR Publisher in Proceedings of the 3<sup>rd</sup> National Conference on Image Processing, Computing, Communication, Networking and Data Analytics (NCICCNDA 2018), April 28, 2018. This is an open access article under [Creative Commons Attribution-NonCommercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0) license, which permits any non-commercial use, distribution, adaptation, and reproduction in any medium, as long as the original work is properly cited. ISBN: 978-81-936820-0-5

checkup, diverse tests, e.g., a sugar level or blood test, X-rays or a CT check, minor surgeries, in the midst of treatment. We call every one of these stages/activities as treatment tasks or errand in this paper. Every treatment errand have shifting time prerequisites for every patient, which sets aside a few minutes expectation and proposal exceedingly convoluted. A patient is typically required to experience examinations, investigations or tests (refereed as errands) as indicated by his condition. In such a case, more than one are not dependent, while others may need to wait for the fulfillment of dependent tasks. Most patients sit tight for unpredictable yet long time in queue, waiting for their swing to achieve every treatment undertaking. Here, we center on supporting patients finish their treatment assignments in an anticipated period and making health organization to accomplish efficient treatment time plan, avoid congestion and insufficient queues. We utilize enormous reasonable information from different healing centers to build up a patient healing period utilization demonstrates. Reasonable patient information is broke down precisely and thoroughly in light of imperative parameters, like patient treatment begin time, end time, age, and detail treatment record for each unique task. We perceive and gauge sitting tight time for various patients in light of their conditions and exercises performed in the midst of treatment. The work process of the patient treatment and hold up structure is depicted in Fig.1.

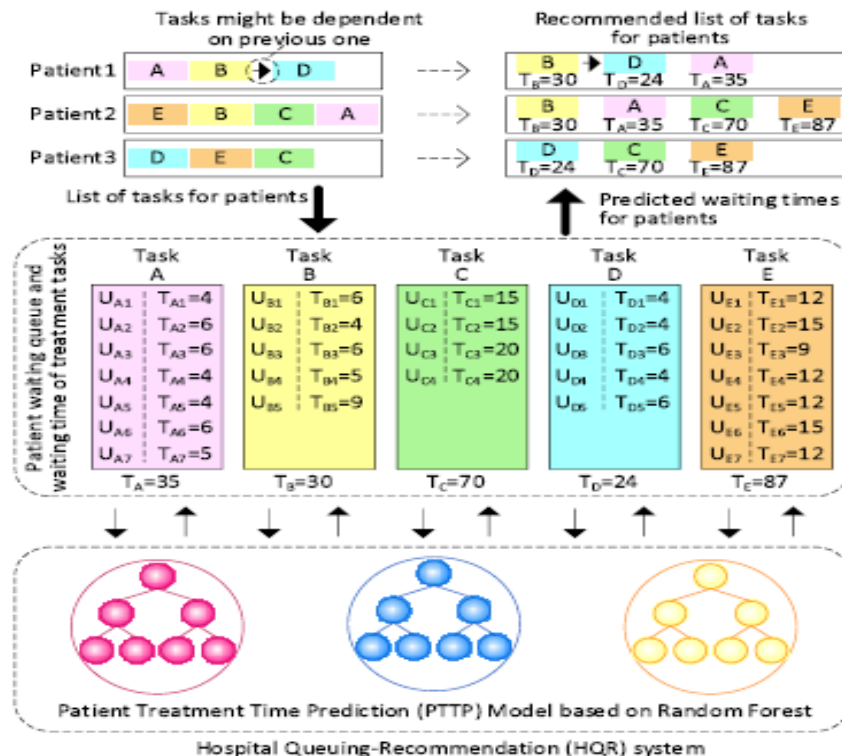


Fig.1. Patient Treatment And Wait Model.

Fig. 1 represents 3 patients (Patient1, 2, and 3) and an arrangement of treatment undertakings necessary for every tolerant. A few errands can be reliant on a earlier one, e.g., surgery is impossible prior to X-rays. Tasks {A, B, D} are essential for Patient1, though undertaking D must sit tight for the finishing of B. Tasks {E, B, C, A} are necessary for Patient 2, and Tasks {D, E, C} are required for Patient 3. Additionally, there are diverse quantities of patients holding up in line of each errand, for instance, 7 patients in the queue of Task A and 5 patients in the line of assignment B.

In this paper, a Patient Treatment Time Prediction (PTTP) model is prepared in light of healing facilities' chronicled information. The holding up period of every treatment undertaking is anticipated by PTTP, which is the entirety of overall patients' holding up times in the current line. At that point, as indicated by every patient's asked for treatment errands a Hospital Queuing-Recommendation (HQR) framework suggests an well-organized and helpful cure design with the least deferrals for patients.

The patient treatment time utilization of every sick person in the holding up line is evaluated through the prepared PTTP demonstrate. The entire holding up period of every undertaking at the present time can be anticipated, for example, {TA = 35(min); TB =30(min); TC = 70(min); TD = 24(min); TE = 87(min)}. At long last, the undertakings of every patient are arranged in a ascending order as indicated by the holding up period, excluding from the reliant tasks. A lining suggestion takes place for every sick person, for example, the suggested lining {B,D,A} for Patient1, {B, A, C, E} for Patient2, and {D,C, E} for Patient3.

Finally the greater part of the required treatment tasks in the most limited holding up time, the holding up time of each tasks is anticipated continuously. Since the waiting time of line for each errand refreshes, the lining suggestion is recomputed in continuous. In this manner, every sick person is encouraged to finish his healing activities in the most helpful path and with the limited waiting time.

Health organization's information is stored in the Apache HBase, parallel arrangement is utilized with the MapReduce and Resilient Appropriated Datasets (RDD) programming model. Segment 2 reviews the survey related work such as PTTP algorithm and a HQR framework. At last, concludes the paper.

## 2 LITERATURE SURVEY

To enhance the exactness of the information examination with consistent highlights, different streamlining strategies for classification, regression algorithms are projected. A self-versatile induction algorithm for cumulative structure of binary regression trees was obtainable [1]. Tyree et al. [2] presented a parallel regression tree calculation for web seeks positioning. In [3], a multi-branch decision tree calculation was proposed in light of a relationship part basis. Other enhanced classification and regression tree strategies were proposed in [4][6].

The random forest algorithm [7] is an outfit classification algorithm in view of a decision tree, which is a reasonable information digging calculation for huge information. The irregular

woodland calculation is broadly utilized as a part of numerous fields, for example, quick activity location by means of relevance random forest selection [8], vigorous ,exact form show coordinating utilizing random forest regression voting [9], and a huge information scientific system for shared botnet discovery utilizing random forest[10]. The test brings about the papers exhibit viability and appropriateness of the random forest algorithm. Bernard [11] proposed a dynamic preparing procedure to develop the exactness of the random forest algorithm. In [12], random forest algorithm system in view of weighted trees was proposed to group tremendous structural up various record. Nonetheless, the first arbitrary backwoods calculation utilizes a customary straight selection technique in the voting procedure. In this situation, the random forest containing boisterous decision trees prompt an erroneous anticipate an incentive for the validation dataset [13].

Different proposal calculations introduced furthermore, connected in associated field. Meng et al. [14] proposed a sign mindful administration suggestion method on MapReduce for huge data application. A movement recommendation calculation that extract individuals' properties and travel aggregate composes was projected in [15]. Yang et. al. [16] presented a Bayesian-inference based suggestion framework to connected informal communities, wherein a client proliferates a substance mark question alongside the informal community to his immediate and aberrant companions. Adomavicius and Kwon [17] presented latest proposal methods for multi-criteria rating frameworks. Adomavicius and Tuzhilin [18] presented an outline of the current generation of recommendation strategy, for example, content-based, collaborative, and hybrid-recommendation suggestion approaches. Regardless, there is no prescient count for understanding treatment time usage in the present examinations.

The swiftness of information mining and investigation intended for huge information is an essential factor [19]. Distributed compute, disseminated figuring, and supercomputers offer rapid registering control. Both the Apache Hadoop [20] and Spark [21] are acclaimed cloud stages that are generally utilized as a part of parallel processing and information investigation. Various parallel information mining calculations have been executed in view of the MapReduce [22] and RDD [23] models. In [24]-[27], different information pulling out calculations was proposed in view of the MapReduce programming model. Apache Spark is an proficient cloud stage that is reasonable for information mining and machine learning. In the Spark, information stored in memory, what's more, cycles for similar information come specifically from memory. Zaharia et al. [28] exhibited a quick and intuitive investigation over Hadoop information with Spark.

Anticipate the deferral time for every healing errand, we utilize the random forest calculation to prepare the patient Healing period consumption in light of both sick person and period attributes and after that fabricates the PTTP model. Since patient Healing period utilization is a constant variable, a Classification and Regression Tree (CART) demonstrate is utilized as a meta-classifier in the RF algorithm. Since inadequacies of the first RF algorithm and the

attributes of the sick person information, in this paper, the RF calculation is enhanced in 4 angles to get a successful outcome from huge range, high dimensional, consistent, and large patient information. Contrasted and the first RF calculation, our PTTP calculation in light of an enhanced RF calculation has significant preferences regarding accuracy and performance. Besides, there is no current research on healing facility lining administration and proposals. Consequently, we propose a HQR framework in view of the PTTP model. To the best of our insight, this paper is the first endeavor to illuminate the issue of patient deferral time for healing facility lining administration figuring. A treatment queue proposal with a proficient and helpful treatment design and the minimum deferral time is prescribed for every patient.

## **2.1 Apache Spark**

Apache Spark is an exceptionally quick bunch processing innovation, intended for quick calculation. It depends on Hadoop MapReduce and it stretches out the MapReduce representation to proficiently consume it for more calculations, which integrate intuitive questions and stream handling. The principle highlight of Spark is its in-memory group registering which builds the preparing rate of an application.

Spark is planned to cover a broad assortment of workloads, for instance, bunch applications, iterative estimations, instinctive request and spilling. Besides supporting all these workload in a specific system, it reduces the organization weight of keeping up detached isolated apparatuses.

Apache Spark has following highlights.

- Speed – Spark execute an application in Hadoop group, it is up to 100 times quicker in memory, and 10 times speedier when running on circle. This is possible by diminishing number of read/form undertakings to circle. It stores the widely appealing planning data in memory.
- Backings various dialects – Spark provide built in APIs in Java, Scala, or Python. like this, we can create applications in a variety of vernacular. Start imagine of 80 abnormal state administrators for intelligent questioning.
- Progressed Analytics – Start not simply sponsorships 'Map' and 'Reduce'. It in like manner supports SQL request, Streaming data, Machine learning (ML), and Graph calculation.

## **2.2 Big data**

Each business, huge or little, is dealing with a lot of information created through its different information focuses and business forms. Now and again, organizations can deal with this information utilizing exceed expectations sheets; get to databases or other comparable instruments. Notwithstanding, when information can't fit into such apparatuses, and human blunder cases increment above adequate constrains because of serious manual handling, the time has come to consider Big Data and examination. Big Data can be characterized utilizing the 3 Vs - Volume, Velocity and Variety.

- **Volume:** Inside the Social Media space for example, Volume suggests the measure of data created through locales, doors and online applications. Especially for B2C associations, Volume joins the available data that are out there and ought to be overviewed for relevance. Consider the going with - Facebook has 2 billion customers, Youtube 1 billion customers, Twitter 350 million customers and Instagram 700 million customers. Reliably, these customers add to billions of pictures, posts, accounts, tweets et cetera. You would now have the capacity to imagine the frantically broad total - or Volume-of data that is created every minute and reliably.
- **Velocity:** With Velocity we imply the speed with which data are being created. Staying with our web based systems administration case, each and every day 900 million photos are exchanged on Facebook, 500 million tweets are posted on Twitter, 0.4 million hours of video are exchanged on Youtube and 3.5 billion requests are performed in Google. This takes after a nuclear data impact. Tremendous Data makes the association hold this impact, recognize the moving toward stream of data and meanwhile strategy it brisk with the objective that it doesn't make bottlenecks.
- **Variety:** Collection in Big Data insinuates all the sorted out and unstructured data that has the probability of getting made either by individuals or by machines. The most typically included data are sorted out - compositions, tweets, pictures and chronicles. In any case, unstructured data like messages, voice messages, physically composed substance, ECG scrutinizing, sound narratives et cetera, are in like manner basic segments under Variety. Combination is about the ability to arrange the moving toward data into various characterizations.

### 2.3 Cloud computing

Distributed computing demonstration of using an arrangement of remote servers encouraged on the Internet to store, supervise, and process data, rather than an area server or a PC.

Top points of interest of conveyed processing:

Here are 6 customary reasons affiliations are swinging to disseminated figuring organizations:

- **Cost:** Appropriated registering takes out the capital cost of buying gear and programming and setting up and running adjacent datacenters—the racks of servers, the round-the-clock control for power and cooling, the IT masters for managing the system.
- **Speed :**Most dispersed figuring organizations are given self organization and on ask for, so even colossal measures of enlisting resources can be provisioned in minutes, normally with just a few mouse clicks, giving associations a lot of flexibility and taking the weight off degree evaluation.
- **Worldwide scale:** The upsides of appropriated registering organizations consolidate the ability to scale adaptable. In cloud talk, that infers passing on its ideal measure

resources—for example, practically figuring power, storing, exchange speed—right when its required and from the benefit geographic region.

- Profitability: On area datacenters customarily require an impressive measure of "racking and stacking"— gear set up, programming settling and different monotonous IT organization errands. Circulated figuring removes the necessity for enormous quantities of these endeavors, so IT gatherings can put vitality in achieving more essential business targets.
- Execution: The best appropriated figuring organizations continue running on a general arrangement of secure datacenters, which are routinely climbed to the latest period of brisk and compelling preparing hardware. This offers a couple of focal points over a single corporate datacenter, including reduced framework dormancy for applications and more significant economies of scale.
- Unwavering quality: Circulated registering makes data fortification, disaster recovery and business cognizance less difficult and more moderate, in light of the fact that data can be reflected at different tedious districts on the cloud provider's framework.

**Cloud administrations:** IaaS, PaaS, SaaS

Most disseminated registering organizations fall into three general characterizations: Infrastructure-as-a-service (IaaS) ,Platform as a service (PaaS) and Software as a service (SaaS). These are before called as conveyed registering stack, since they grow over each other. Perceiving what they are and how they are unmistakable makes it less requesting to accomplish your business goals.

- Infrastructure-as-a-service (IaaS) : Infrastructure-as-a-service (IaaS) The most basic order of conveyed processing organizations. With IaaS, you rent IT establishment— servers and virtual machines (VMs), storing, frameworks, working systems—from a cloud provider on a remuneration as-you-go start.
- Platform as a service (PaaS) : (PaaS) suggests appropriated processing organizations that supply an on-ask for condition for making, testing, passing on and directing programming applications. PaaS is planned to make it less requesting for originators to quickly make web or versatile applications, without struggling with setting up or managing the fundamental establishment of servers, amassing, framework and databases required for development.
- Software as a service (SaaS) : Software as a service (SaaS) is a procedure for passing on programming applications over the Internet, on ask for and customarily on an enrollment commence. With SaaS, cloud providers have and manage the item application and fundamental structure and handle any help, like programming updates and security settling. Customers connect with the application over the Internet, commonly with a web program on their phone, tablet or PC.



## 2.4 PTTP Model based on the improved RF Algorithm

To anticipate the deferral period for every patient treatment undertaking, the patient treatment time utilization in light of various quiet qualities and time attributes must be figured first. The time consumption of every treatment errand won't not be in same range, where it differs as per the substance of undertakings and special conditions, diverse periods, what's more, and unique states of patients.

In this way, we utilize the RF calculation to prepare understanding treatment time utilization in view of both sick person and time attributes and after that assemble the PTTP display.

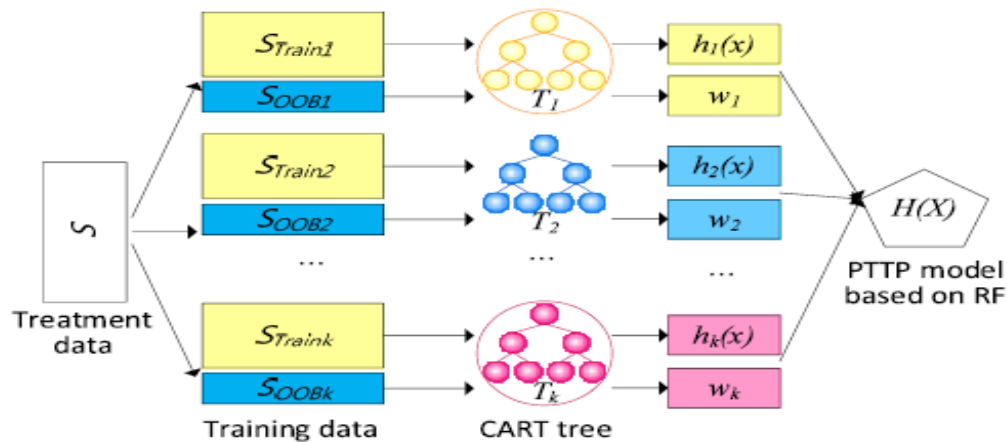


Fig.2. PTTP model for queuing system.

For some training data  $S$ , construct the training subsets  $S_{train}$ , at the same time non selected records in each sampling are made as an  $S_{OOB}$ . Split the data into forks that should denote right data subsets and left data subsets. Based on this subsets construct the multi branch Classification and Regression Tree (CART). Then evaluate mean value of leaf nodes after pruning of noisy data, calculate accuracy of every tree that is the weighted regression outcome  $H(X)$  of RF model for the data  $X$  is the average value of  $k$  trees, where  $w_i$ , the weight of tree  $h_i$ .

### Pseudo Code for PTTP model based on RF algorithm

**Input:** Set of data.

**K:** Large number of CART trees presented in RF structure.

**Output:**

Anticipated time for every treatment errand in light of Gender, age, time.

**Steps to Be Followed:**

**Step 1:** Create subset of training data.

**Step 2:** Create subset of OOB.



**Step 3:** For each set of Training data, do following steps.

**Step 4:** Calculate best split point for Feature variable.

**Step 5:** Split the data into two parts.

**Step 6:** Construct multi-branch for the CART Model.

**Step 7:** Calculate Mean estimation of Leaf hub after evacuation of uproarious information.

**Step 8:** Predicted time for every treatment errand will be put away in database.

Beneath specified Table .I indicates cases of treatment records, Table. II demonstrates configurations of the information for various treatment assignments. Table. III is case of highlights of treatment information for the PTTP calculation.

Table.I Sample of different treatment tasks.

| Patient No. | Gen  | Age | Task name     | Dept. name   | Doctor name | Start time             | End time               |
|-------------|------|-----|---------------|--------------|-------------|------------------------|------------------------|
| 0001        | Male | 15  | Checkup       | Surgery      | Dr. Chen    | 2015-10-10<br>08:30:00 | 2015-10-10<br>08:42:25 |
| 0001        | Male | 15  | Payment       | Cashier-6    | Null        | 2015-10-10<br>08:50:05 | Null                   |
| 0001        | Male | 15  | CT scan       | CT-5         | Dr. Li      | 2015-10-10<br>09:20:00 | 2015-10-10<br>09:27:00 |
| 0001        | Male | 15  | MR scan       | MR-8         | Dr. Pan     | 2015-10-10<br>10:05:06 | 2015-10-10<br>10:15:35 |
| 0001        | Male | 15  | Take medicine | TCM Pharmacy | Null        | 2015-10-10<br>10:42:03 | 2015-10-10<br>10:45:29 |
| ...         | ...  | ... | ...           | ...          | ...         | ...                    | ...                    |

Table.II Features of treatment data for the PTTP algorithm.

| Treatment task | Format of the data (Feature name)  |
|----------------|--|
| 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, start time, end time, 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, body region of scans, start time, end time, remark}    |
| Injection      | {Patient card number, patient name, gender, age, task name, department, doctor, start time, end time, drug name, drug number, remark } |
| Blood Tests    | {Patient card number, patient name, gender, age, task name, department, doctor, time of blood tests, time of report}                   |

Table.III Features of treatment data for the PTPP algorithm.

| No.      | Feature Name     | Value range of each feature subspace   |
|----------|------------------|--|
| $y_1$    | Patient Gender   | "Male", "Female".  |
| $y_2$    | Patient Age      | The age of the patient.  |
| $y_3$    | Department       | All departments in the hospital.   |
| $y_4$    | Doctor Name      | All doctors in the hospital.   |
| $y_5$    | Task Name        | Each treatment task in all treatment processes in the hospital.  |
| $y_6$    | Start Time       | The start time of the treatment task.  |
| $y_7$    | End Time         | The end time of the treatment task.  |
| $y_8$    | Week             | The day of week for the treatment time. The value is from Monday to Sunday.  |
| $y_9$    | Time Range       | The time range of treatment time in a day. The value is from 0 to 23.  |
| $y_{10}$ | Time Consumption | (1) End time - Start time, such as a CT scan, an MR scan. (2) Time interval between one patient and the next in the same treatment, such as payment. |

### 2.5 Hospital Queuing Recommendation System

After training the PTPP model for every healing errand utilizing chronicled healing center treatment information, a PTPP-based hospital queue recommendation framework is created. An proficient and helpful treatment design is made and prescribed to every patient to accomplish brilliant triage. Expect that there are different treatment assignments for every patient as indicated by the patient's situation, for example, tests and reviews. Let  $Tasks = \{Task_1, Task_2, \dots, Task_n\}$  be an arrangement of treatment assignments that the present patient must finish, and let  $U_i = \{U_{i1}, U_{i2}, \dots, U_{im}\}$  be an arrangement of patients in sitting tight the line for  $Task_i$ . The procedure of the HQR framework in light of the PTPP display is appeared in Fig. 3.

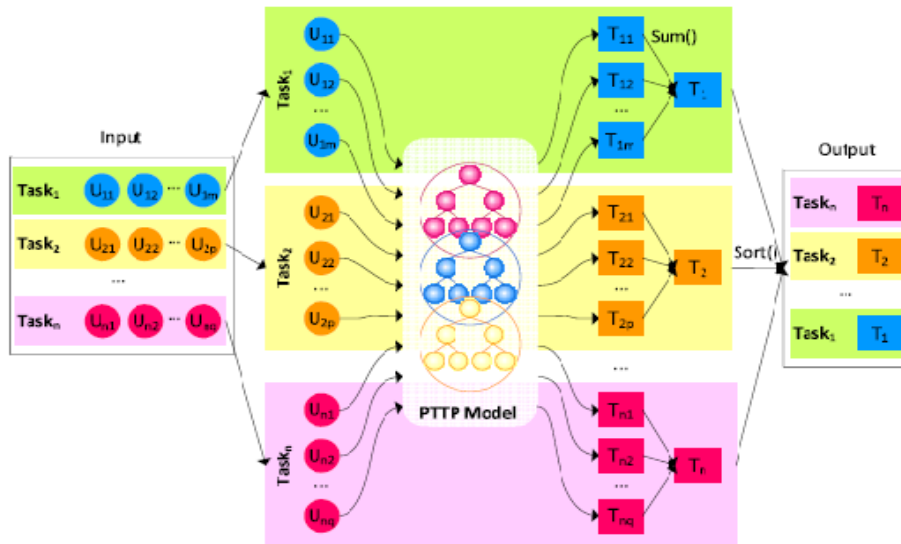


Fig.3. HQR system based on PTPP model.

This Predict the holding up time of the greater part of the treatment errands for the present patient, sort the majority of the treatment undertakings of the present patient in climbing request by holding up time at that point give a healing facility lining proposal to the present patient.

### **Pseudo code for HQR System Algorithm**

#### **Input:**

X Current patient Treatment data .  
PTTP model based on RF algorithm.

#### **Output:**

The suggested tasks with predicted deferral time.

#### **Steps to be followed in the above algorithm:**

**Step 1:** For every tasks present in X

**Step 2:** Create a list to store patients in waiting for each task

**Step 3:** calculate the task time for every task of user from PTTP algorithm

**Step 4:** Append the data in List

**Step 5:** Sort the data in ascending order.

**Step 6:** keeping records of the dependent task before that task.

Contribution for this framework is healing information for the current patient and PTTP calculation. Yield will be sort rundown of errands for every sick people. For each undertaking in the current client needs to get holding up time from PTTP calculation so that sick person get his appointment without any delay.

### **3 CONCLUSION**

Finally, a PTTP figuring in perspective of enormous data and the Apache Spark cloud condition is proposed. A self-assertive Random Forest is performed for the PTTP show. The line holding up time of each treatment task is foreseen in perspective of the arranged PTTP illustrate. A parallel HQR system is created, and a productive and supportive treatment configuration is recommended for each patient. Wide examinations and application comes to fruition exhibit that our PTTP count and HQR structure achieve high precision and execution. Centers' data volumes are extending every day. The workload of setting up the real data in every plan of mending focus manage recommendations is depended upon to be extraordinarily high, yet it require not be. Accordingly, an incremental PTTP estimation in light of spilling data and a more invaluable recommendation with constrained way care are prescribed for future work.

### **REFERENCES**

- [1] R. Fidalgo-Merino and M. Nunez, "Self-adaptive induction of regressiontrees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1659\_1672, Aug. 2011.
- [2] S. Tyree, K. Q. Weinberger, K. Agrawal, and J. Paykin, "Parallel boosted regression trees for Web search ranking," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2012, pp. 387\_396.

- [3] N. Salehi-Moghaddami, H. S. Yazdi, and H. Poostchi, "Correlation based splitting criterion in multi branch decision tree," *Central Eur. J. Comput.Sci.*, vol. 1, no. 2, pp. 205\_220, Jun. 2011.
- [4] G. Chrysos, P. Dagrizikos, I. Papaefstathiou, and A. Dollas, "HC-CART: A parallel system implementation of data mining classification and regression tree (CART) algorithm on a multi-FPGA system," *ACM Trans.Archit. Code Optim.*, vol. 9, no. 4, pp. 47:1\_47:25, Jan. 2013.
- [5] N. T. Van Uyen and T. C. Chung, "A new framework for distributed boosting algorithm," in *Proc. Future Generat. Commun. Netw. (FGCN)*, Dec. 2007, pp. 420\_423.
- [6] Y. Ben-Haim and E. Tom-Tov, "A streaming parallel decision tree algorithm," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 849\_872, Oct. 2010.
- [7] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5\_32, Oct. 2001.
- [8] G. Yu, N. A. Goussies, J. Yuan, and Z. Liu, "Fast action detection via discriminative random forest voting and top-K subvolume search," *IEEE Trans. Multimedia*, vol. 13, no. 3, pp. 507\_517, Jun. 2011.
- [9] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, "Robust and accurate shape model matching using random forest regression-voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1862\_1874, Sep. 2015.[10] K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, "Big data analytics framework for peer-to-peer botnet detection using random forests," *Inf. Sci.*, vol. 278, pp. 488\_497, Sep. 2014.
- [11] S. Bernard, S. Adam, and L. Heutte, "Dynamic random forests," *Pattern Recognit. Lett.*, vol. 33, no. 12, pp. 1580\_1586, Sep. 2012.
- [12] H. B. Li, W. Wang, H. W. Ding, and J. Dong, "Trees weighting random forest method for classifying high-dimensional noisy data," in *Proc. IEEE 7th Int. Conf. e-Business Eng. (ICEBE)*, Nov. 2010, pp. 160\_163.
- [13] G. Biau, "Analysis of a random forests model," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 1063\_1095, Apr. 2012.
- [14] S. Meng, W. Dou, X. Zhang, and J. Chen, "KASR: A keyword-aware service recommendation method on MapReduce for big data applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3221\_3231, Dec. 2014.
- [15] Y.-Y. Chen, A.-J. Cheng, and W. H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1283\_1295, Oct. 2013.
- [16] X. Yang, Y. Guo, and Y. Liu, "Bayesian-inference-based recommendation in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 642\_651, Apr. 2013.
- [17] G. Adomavicius and Y. Kwon, "New recommendation techniques for multicriteria rating systems," *IEEE Intell. Syst.*, vol. 22, no. 3, pp. 48\_55, May/June. 2007.
- [18] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734\_749, Jun. 2005.
- [19] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97\_107, Jan. 2014.
- [20] Apache. (Jan. 2015). *Hadoop*. [Online]. Available: <http://hadoop.apache.org>
- [21] Apache. (Jan. 2015). *Spark*. [Online]. Available: <http://spark-project.org>
- [22] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107\_113, Jan. 2008.
- [23] M. Zaharia et al., "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proc. USENIX NSDI*, 2012, pp. 1\_14.