Decision support system for bandwidth expansion according to the increment of users

U J Hashan Dinuka

Index No-169311G

Faculty of Information Technology

University of Moratuwa

2019

Decision support system for bandwidth expansion according to the increment of users

U J Hashan Dinuka

Index No-169311G

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the Requirements of the Honors Degree of Bachelor of Science in Information Technology.

February 2019

Declaration

We declare that is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student	Signature of Student
U J Hashan Dinuka	……
	Date: ……..

Supervised by

Name of Supervisor	Signature of Supervisor
Mr. Saminda Premaratne	…….
	Date: ……...

Acknowledgements

First and foremost, it would be my pleasure to manifest heartfelt gratitude to the supervisor Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance without whom, this thesis would not have been possible. I should emphasize the thorough instructions from Prof. Asoka S. Karunananda helped me a lot. Also I would like to convey my sincere gratitude with all the professors who taught us throughout the M.Sc program for molding me to make this thesis a success.

I would also share my gratitude with the Lanka Bell (pvt.) Ltd. for the constant assistant throughout this process in terms of obtaining data. I would also like to thank all the batch mates of the M.Sc. in IT degree program who supported me with guidance and with their genuine and valuable opinions which led me to make the necessary improvements and finalized the results of the research and my family for the support they provided me.

Abstract

Considering the local context, in Sri Lanka internet speediness has become a strain issue domestically and in organization status making the promising promises of internet providers of delivering the fastest internet speed, an abortive attempt. One of the major reasons to that is giving a fixed bandwidth to the area, or fixed number of EnodeB s (EUTRAN Node B / Evolved Node B in 4g LTE). But with the increase of usage that given bandwidth becomes insufficient causing a lot of network traffic, mostly in peak times it almost becomes unusable as same as with EnodeBs becomes insufficient because too many clients connecting to the same node. The thorough analysis about this particular circumstance leads to understand that this issue has been occurred in different countries and most of the expertise has looked upon the matter. With the understanding that bandwidth expansion according to the user increment is a quite common concern which has been undergone with several suggested solutions; this research is going to explore the matter in terms of the local context considering the sample of consumer usage data of Lankabell (pvt) Ltd. The collected data will be processed through data mining algorithms. The objectives of this paper are to identify the internet speed decrements, high-risk area and low-risk area customers by using the data mining techniques like clustering.

KEYWORDS - Data mining, Clustering, Internet speed, Bandwidth, Broadband speed

Table of Contents

List of Tables

List of Figures

Chapter 01

Introduction

1.1 Prolegomenon

In the first chapter, the initial objective is providing an introduction to the complete project as a whole. In other words, through this chapter it will provide clarity on what is the problem this thesis would investigate on and what are the expected objectives from conducting this research along with the scope of this research.

Internet usage is in this genre is getting increased vigorously compared to the previous years. In the mid of 2017, it is reported that exceeding 50% of population in the world is using internet. This was caused by the tremendous technological advancement occurred in this very decade. Considering this rapid global internet usage, an increment in the volume of information per unit of time that a transmission medium can handle has to adjust accordingly. As it was stated in the Nielsen's Law of Internet Bandwidth, “Users' bandwidth grows by 50% per year” confirming the need of decision support system for bandwidth expansion according to the increment of users

In current IT sphere the highway is the web with fast, more information and considerably lower cost. On the off chance that we don't have over three focuses in our nation network access it would be another ordinary street, we as a whole know typical streets takes hours, True that the ride is smooth per say but when it comes to the peak hours, fastness is questionable.

Local Internet benefit has been poor mainly because of these reasons.

- Low information given by the specialist co-op – "Information resembles fuel in web, with restricted fuel the sum you can go in web is constrained, this puts Sri Lanka way being different nations in intensity"

- Web Cost is high - "Both versatile and landline"

- Slower speed – "It resembles when different nations utilize present day fuel effective autos we are as yet utilizing old, obsolete, wasteful gas chugging Morris Minors and Humber autos"

- Substantial tax collection on the telephone charge – "This intensely debilitates landline Broadband and Fiber"

In the Sri Lankan context locals are intensely dependent on versatile Internet utilizing a device named as dongles. As affording a “Broadband/ Fiber” is not a practically capable thing for the mass community in out of Colombo; They used to fulfill their internet needs using portable based web utilizing dongles or mobile phones for a more convenient and economical aspect.

Generally, opposed to the developed countries, Sri Lankan fixed line internet usage is very low. Thus if anyone is conversing with actualize extending web should offer need to the versatile Internet hence that is the most straightforward and also the speediest approach to spread Internet crosswise over the country additionally from at any rate, not recommending to disregard the landline based web "Broadband/Fiber" Such considerably higher capital should be invested in internet (landline based) across the country as well.

1.2 Research Aim & Objective

Aim

The researcher is starting this dissertation with the interest of implementing a decision support system for bandwidth expansion for provide an uninterrupted, better internet facility to the people in Sri Lanka..

Objectives

1. Critically appraise the data mining technologies apropos of the Bandwidth expansion
2. Keen investigation of methods of data mining with the interest of understanding the best method to use in this particular project
3. Data pre-preparation, processing and the implementation of user interface
4. Implementing and executing the suggested system
5. Investigate the limitations of the process of conducting the project for the future perusal
6. Inscribe the project as a thesis (dissertation)

It is expected to obtain the following mentioned benefits for the consumer as a result of implementation of this particular system. Such as,

- Increase of consumption can be easily handled.
- Customer Satisfaction goes up.
- Quality of the service goes up.

1.3 Problem Statement

Considering the local context, in Sri Lanka internet speediness has become a strain issue domestically and in organization status making the promising promises of internet providers of delivering the fastest internet speed, an abortive attempt. One of the major reasons to that is giving a fixed bandwidth to the area, or fixed number of EnodeB s (EUTRAN Node B / Evolved Node B in 4g LTE). But with the increase of usage that

given bandwidth becomes insufficient causing a lot of network traffic, mostly in peak times it almost becomes unusable as same as with EnodeB s becomes insufficient because too many clients connecting to the same node. So to keep up the promised bandwidth ISPs have to increase the given bandwidth to particular area or increase the nodes at the given area. Currently they do not have any method of foreseeing the bandwidth increments. As a result customer complaints and dissatisfied customer count grows

1.4 Scope of the research

In conducting the project, it was visible that there are some limitations and challenges to be sorted and settled. At the very beginning, it was observed that a very few number of researches are there in this particular phenomenon. Other than lack of theoretical support which is the fundamental backbone of any successful research; there are few other limitations to be identified in a through perception such as,

- Doubtfulness of accuracy of considering data
- Difficulty getting accessibility to the commercial- grade software for an affordable cost
- Data that are supposed use as the input, can be available in variety of formations (images, geo data, text, social, numeric, etc.) as they gained from numerous sources which is also known as “Data Variety”
- Online machine learning utilization demands constant update of new incoming data which is also known as “Data Velocity”.
- Distribution approaches are required when dealing with massive data sets which are also known as "Big Data"

1.5 Proposed Solution

As per my knowledge, the solution is to find the bandwidth expansion needs using data mining techniques like rapidminer and I propose a web based system as above mentioned; a decision support system for bandwidth expansion according to the increment of users to overcome this matter which will be a valuable tool for the telecommunication industry.

1.6 Summery

1.6.1 Structure of the Thesis

This thesis will cascade as follows. The initial first chapter is the introduction to the thesis, problem background, need of analyzing this problem and expected outcomes. The second chapter will be a critical review of the literature of the focused topic which is creating a decision support system for bandwidth expansion according to the increment of the users using data mining techniques while the third chapter will be about the technologies that uses to gather data and the tools that will be used in data mining in terms of conducting this particular research. The forth chapter will be about approach with users, inputs, outputs, process and features. The fifth chapter would be about the procedure of designing and implementing a system as suggested. Chapter six will be a discussion on the implementation process while chapter seven would provide a conclusion for the whole research with recommendations for further reference.

1.6.2 Chapter summery

As mentioned in the beginning, purpose of this chapter is to provide clarity on what this thesis will be about and what will be the project under the topic of “Decision support system for bandwidth expansion according to the increment of users” intend to obtain in the end. This chapter also brought forth the main objectives and the need of this project in a practical perspective. The researcher has mentioned a nutshell of the problem statement is there to understand that why a decision support system is critical for the subjected topic; Bandwidth expansion with respect to the increment of users.

Chapter 2

Review of data mining techniques and bandwidth expansion in substantive findings

2.1 Introduction

Reviewing the academic literature and the theoretical contribution from the already conducted publications can provide an idea about where this scrutinizing of decision making system for bandwidth expansion according to increment of users will lead to. In this chapter, first knowledge would be about the concept of data mining. Then I will be exploring about how data mining helped certain successful researches and afterword I will be discussing about the researches which was conducted about the bandwidth expansion and about the conclusions the respective experts achieved. Another reason behind reviewing the literature is to find about the limitation of conducting my research. Through executing a literature review, it provides a clear path to conduct this research as in where to overlook and a summary of problem definition along with the challenges in the practical context which will eventually leads to refine my project.

2.2 Related findings with reference to Data mining

Data mining is the way towards dealing with huge data collections to recognize designs and set up connections to take care of issues through data analysis. Data mining devices enable companies to forecast future trends. In other words, data mining is extracting knowledge hidden in large volumes of data and identifying potentially useful and understandable data. When it comes to the telecommunication industry aspect in data mining, it is to be well understood that Telecommunication organizations create a tremendous measure of data. These data incorporate call detail data depicting calls crossing media transmission systems, organizing data depicting the condition of the system's equipment and programming segments, and customer data describing the telecommunications customers. This thesis depicts how Data Mining can be utilized to reveal valuable data covered inside these data sets. A few Data Mining applications are

depicted and together they show that Data Mining can be utilized to distinguish media transmission misrepresentation, enhance showcasing viability, and recognize organize shortcomings. In simple terms, it can be identified that data mining in telecommunication has been used to detect frauds, to define customers‟, to retain customers‟, to recognize the products and services yield highest amount of profit and to emphasize the factors that influence customers to call more at certain times.

In the current context, Decision support system for bandwidth expansion according to the increment of users has taken the developers attention scrupulously. However several perspectives of this particular phenomenon have been assessed by numerous enthusiastic specialists over various research articles. It is reasonable to presume that it has encountered the prioritized observation of developers in the path to prominence of web development in software engineering enhancement in product development. With regard to scrutiny of Decision support system for bandwidth expansion, an incisive sense can be gathered by overlooking at explorations by experts in the sphere. As in this particular thesis, discussion is supposed to be about Decision support system for bandwidth expansion according to the increment of users through the data collected by data mining technique in telecommunication, the past perusals will be considered in both regards in the following.

2.3 Related findings with reference to Bandwidth Expansion requirement

In the thesis “A Bandwidth Allocation Policy for Helpers in Cloud-assisted P2P Video-on-demand Systems” by Guowei Huang, Lingjing Kong, Keke Wu and Zhi Chen it is thoroughly analyzed the bandwidth allocation issue of cloud helpers in the cloud- obliged Peer to Peer based Video on Demand streaming system. It shows the issue as a direct advancement issue numerically and proposes an ideal distribution arrangement in light of the logical model, which is useful to enhance the fulfillment degrees of consumers. The researches have conducted several experiments in the matter to confirm the factualness of the matter. (Guowei Huang, 2017)

When adjudging the Bandwidth Allocation Algorithms in the Short Video Sharing System article by Dan Huang, Changjia Chen, it can be seen that they have concentrated on the keynote in the short video sharing system; users upload bandwidth. Thus it has presented the task of uploading bandwidth of peers in the video accelerator system based on P2P. In a nutshell authors have described programming and heuristic strategy as a proposal to control how much portion of upload bandwidth should be allocated to the request. In conclusion they suggest having "centralized algorithm is near optimal while the one of distributed algorithm outperforms the one using equal allocation and is eminently suitable for practical deployment. The simulation results also demonstrate that both algorithms can lead to significant performance improvement" (Dan Huang)

In the article "End-to-end Internet Speed Analysis of Mobile Networks with Map Reduce" by Mete UZUN and Osman ABUL, it is been discussed that mobile data activity is developing by soaring step by step. More mobile application administrations and keen gadgets are developing each day. As a result of expanding versatile data utilization, data activity investigation is an exceptionally major issue in portable system administration and improvement for portable specialist organizations. In their research paper, they have suggested that test frameworks run intermittently speed tests in a few test regions with genuine use situations and follow TCP parcels and process downloaded and transferred information sum every second and RTTs as per each speed test. Test comes about, as extensive informational indexes, originate from different test frameworks, handled and dissected utilizing Map Reduce programming model on distributed computing stage based Hadoop Environment. Additionally, local speed attributes delineate rendered by territory based examined huge test outcomes. They have built up the Internet Speed. Analysis System which can scale end to end organize examination and subsequently portrays area based system speed quality, client experience and issues without client grumblings with huge information investigation. Thusly, the framework follows information activity by end to end and accurately catches nature of experience as indicated by huge information examination. (Mete Uzun, 2016)

“Why Is My Internet Slow?: Making Network Speeds Visible” is an interesting journal section of this very phenomenon proceeded by Marshini Chetty, David Haslem, Andrew Baird, Ugochi Ofoha, Bethany Sumner and Rebecca E. Grinter. They have depicts that, with far reaching broadband selection, more family units report troublesome speeds. Not only are moderate paces disconcerting, they can show buyers don't get the administrations they pay for from their network access providers. However, it is uncomfortable to decide the speed and source of moderate downs given that there are few instruments for broadband administration. After investigating aftereffects of a field trial with 10 families utilizing a visual system test intended to address these issues, they have come across the following suggestions. On account of worries around least broadband paces to American homes, they have planned and conveyed a test to indicate family units their broadband velocities and transfer speed use. The test was generally welcomed, and recommends the requirement for more instruments that give customers propelled control over their broadband associations. All the more imperatively, such apparatuses can fill in as a mechanical purchaser guard dog and can
enable family units to have a voice in discusses around broadband velocities. (Marshini Chetty, 2011)

“Pushing the Wireless Data Rate to the Internet Speed” was another very thorough research article that has been published by Shaofang Gang and Magnus Karlsson which has answered the inquiry how to accomplish remote information rates that can make up for lost time with current Internet speed, from a fundamental material science perspective. It is demonstrated that the customary electric circuit hypothesis and plan technique that have been utilized for ages are shockingly not sufficient for remote correspondences later on. Rather, troublesome methodologies, for example, six-port modulators for handling of electromagnetic waves and optical heartbeats, ought to be utilized to push up the remote information rate over 100 Gb/s. The key variables to consider for high-speed digital communications are bandwidth, modulation order, and signal-to-noise ratio. In principle, it should be possible to achieve a wireless data rate at 100 Gb/s within the frequency spectrum below 20 GHz. In summery they have concluded their analysis by suggesting; the key factors to consider for fast advanced interchanges

are bandwidth speed, modulation order and signal-to-noise ratio. On a basic level, it ought to be conceivable to accomplish a remote information rate at 100 Gbps inside the recurrence range underneath 20 GHz. To understand that, one should utilize Maxwell's conditions rather than Kirchhoff's electrical current and voltage laws for outlining correspondence frameworks. That is, we should utilize approaches treating electromagnetic waves, rather than electrical current and voltage signals. In this regard, the six-port modulator can be a decent contender for problematic advancement, particularly when baseband information in optical fibers is utilized straightforwardly to do tweak. This is a straightforward engineering, yet it fulfills the necessities for a huge transmission capacity, high modulator request and high flag to-clamor proportion. Therefore, it can possibly achieve the greatest information rate that physical standards allow. (KARLSSON, 2017)

This occurrence was a key note in 2014 International Conference on Intelligent Computing Applications and in their report under “INCREASING INTERNET SPEED AND BANDWIDTH BY USING LAWS OF PHYSICS” S.Sharad has discussed that the internet speeds assumes an essential part in all web based utility/benefit people use in their everyday life. The speed of the web is subject to the bandwidth which is given by the Internet Service Provider (ISP).In any discourse quick web availability is an essential necessity, for everybody who utilizes a PC. Consequently, this thought advanced, which is a technique for enhancement of flag quality, which will bring about a quicker internet. This is finished by contorting cables over the current transmission link. The twisting of the cable around the existing transmission cable is thus found to increase the internet speed and bandwidth by 20-50 % depending on the data plan and other factors. around the current transmission link is along these lines found to build the web speed and bandwidth by 20-50 % relying upon the information design and different variables. The improvements performed and proposed in this paper offers a most extreme pick up with least expenses. This undertaking serves well for a straightforward LAN and the future works incorporates improving the paces significantly more by some product sidekick for help. The future work additionally incorporates into upgrading the scope range of the execution locale to other system topologies and furthermore different sorts of systems

like MAN, WAN, intranet and furthermore web .In future an investigation on how the OFC (optical Fiber Cable) has a transmission misfortune and how this thought can be adjusted and connected to the OFC to enhance the transmission rate in them is likewise to be finished. As, India heads on the race to end up noticeably a super-control it needs a speedier web network to back it up in this advanced age and this is the initial phase toward that path. (S.Sharad, 2014)

In the “Optimization of 4G Wireless Access Network Features by Using Reverberation Chambers: Application to High-Speed Train LTE Users” as a Proceedings of the 46th European Microwave Conference by Massimo Barazzetta, Davide Micheli and Riccardo Diamanti; They says that a real 4G base station associated with live system of Telecom Italia and transmitting into a resonation chamber is utilized to examine exhibitions of quick moving terminals on leading group of rapid trains. The outcomes concern a testing effort by duplicating genuine rapid prepare spread situations, similar to the prepare running in open space or in burrows, with the guide of a resonation chamber. This is incompletely filled by retaining materials to approach the multipath engendering that is ordinary of a prepare mentor, and outfitted with two turning paddles keeping in mind the end goal to duplicate the flag fluctuations that are run of the mill for the specified situations. The impacts of Doppler move and its quick varieties are introduced, together with an examination of the efficiency of various transmission answers for burrows. As a summery the have gathered that the effect of high speed on LTE FDD access technology cannot be encompassed in the Doppler Effect concept alone. Doppler shift may have some negative impacts on throughput, but this is not the most important issue. Experience from this campaign and also feedbacks from the operator led us to conclude that access to the network could be problematic. After 250 km/h, the activation of the restricted set of cyclic shifts in PRACH should be considered; otherwise the orthogonality between signatures could be compromised by the high relative speed between the transmitting and the receiving points. Another important aspect to consider is the outage after a radio link failure. The UE may persist in the out-of sync condition if the radio link is not re-established through a new access to PRACH. In this case, the effect on the perception of quality is bad since the signal strength is rather good but the UE is not able to get access

to the network. Possible coverage solutions for tunnels were presented. The usage of a SISO solution that replicate the signal from the external cell is cheap and rather simple, but it presents strong throughput degradation when increasing speed. This problem is mitigated again using the restricted set of cyclic shifts in PRACH, with additional enhancements like the decorrelation of the two signal contributions. (Massimo Barazzetta, 2016)

The article “A Bandwidth Allocation Scheme to Meet Flow Requirements in Mobile Edge Computing” by Yusuke Ito and Hiroyuki Koga mentioned that distributed computing has empowered clients to appreciate different Web administrations gave by server farms through a wide assortment of access systems and correspondence terminals at whenever and any place. In distributed computing, the quality of latency-sensitive services relies upon the separation amongst clients and server farms. In this manner, versatile edge registering, which finds edge servers on get to systems, has been proposed it enables lower latency and higher-speed communication services than can be realized through computing solely through data server farms. Different administrations in edge distributed computing, nonetheless, have diverse round-trip times (RTTs) and necessities. In this circumstance, a RTT-shamefulness issue emerges on the grounds that such administrations normally utilize TCP. In this examination, we propose a data transfer capacity distribution plot in light of collectable data to meet the prerequisites of each stream in versatile edge processing; this is an expansion of our beforehand revealed conspire. We have affirmed the adequacy of this approach through recreation assessments. Simulations evaluations have indicated that the proposed scheme can allocate transmission rates among flows according to their requirements independently of each flow’s RTT as well as achieve high link utilization. In our future work, we will consider an effective algorithm to allocate unutilized bandwidth to other competing flows in a more realistic environment as mentioned. (Y. Ito, H. Koga and K. Iida, 2017)

In the journal article “Architecture Design of an Area Efficient High Speed Crypto Processor for 4G LTE” by Anastasios N. Bikos, Nicolas Sklavos; they argues that the entire security design of LTE/SAE (Long Term Evolution/System Architecture

Evolution) is being comprised of four primary equipment situated cryptographic calculations: KASUMI piece figures, SNOW-3G stream figure, the MILENAGE calculation set, and the 4G advancement of ZUC calculation. This paper introduces a FPGA organization of an all-inclusive security engineering crypto processor for 4G LTE, comprising of both four figures empowering everyone on request, which depends on two novel outline standards. One is a more smart usage of the four calculation's substitution boxes (S-boxes) in light of a typical crossing point presumption of their substance. The second incorporates the utilization of a typical information way equipment square sent along the four figure's building plan. This all inclusive security engineering crypto processor demonstrates to decrease the region space no less than 1.5 times, and furthermore gives twofold throughput contrasted and the cutting edge acknowledge of the individual figures, something which is a high need while creating parts for the requesting post-4G cell advertise. As a conclusion system administrators and chip producers have awesome advertising viewpoints towards the sending and use of 4G arranges in the up and coming decades. The tremendous increment of portable supporters makes awesome difficulties regarding guaranteeing privacy and trustworthiness of the two information and flagging transmissions. A productive and smaller Full Rolling Architecture crypto processor was depicted in this archive, to merge the aggregate arrangement of 4G LTE/SAE security board, alongside the aftereffects of its usage in FPGA innovation. The plan procedures and novel reexaminations forced in building this security processor have ended up being extremely valuable. Not exclusively does this proposition accomplish a higher execution, throughput, and region/defer change; however is one of the Advanced most novel outlines as far as bringing together the S-boxes of 4G Security into just a single. What's more, it figures out how to copy the operation of a convoluted procedure, the left move of bit stream into a prepared circuit. We trust that the usage addresses the issues for a vigorous figuring design which is minimized, coherent and sufficiently built to spare silicon space and has bring down power utilization. (Anastasios N. Bikos, 2016)

A Bandwidth-Conscious Caching Scheme for Mobile Devices is an article by Badari N. Thyamagondlu, Victor W. Chu and Raymond K. Wongwhich depicts while a generous measure of huge information are expended through mobile phones, getting to content by

means of remote information associations on cell phones has its own arrangement of difficulties. Among these difficulties, speed of information exchange is generally our first need. In spite of the fact that there are many quick information associations accessible for Web surfing (3G, LTE and so on.), the genuine association speed could shift significantly among various districts where quick associations may not be constantly accessible. Thus, the client experience of viewing in formation varies with different type of data connections in various areas. This paper proposes using the Type of Data Connection (Bandwidth) to decide if a dataset should be stored or pre-gotten to decrease the reaction time and along these lines giving a superior client encounter. The part of the cell phone proprietor will frame the premise of dataset development criteria by utilizing the procedure of part mining. As the cell phones are not confined to a specific space, a push to follow the proprietor's development figures out where the proprietor with the gadget is heading towards. This distinguishes the distinctive association speed designs along the proprietor's way; with the goal that diverse reserving or pre-getting system can be sent already to go for reliable nature of administrations. After a thorough discussion they discovered that these days, a significant measure of enormous information is expended by means of cell phones because of the notoriety of cell phones. While the innovation of portable system is constantly enhancing, association speed could in any case shift contingent upon what sort of association is accessible in an area and the association capacity of the gadget. In this paper, we have proposed utilizing information association sort to control dataset pre-getting subject to the settings of client activities, and consequently this control spares information space on cell phones where information stockpiling is constrained. Aside from space sparing, we likewise feature that this dynamic data can prompt much better client encounter (or to keep up a homogeneous client encounter) by pre-bringing dataset that is foreseen to be utilized as a part of an area where just slower information association is accessible. The idea of area relationship mining was utilized to anticipate the development of gadget proprietors. Facilitate headway to this model can be made by considering extra parameters, e.g., client like and abhorrence, favored areas, parts that the client may have a place with in a more extensive group, and enhancing likelihood estimation and fusing different factors that may influence the pre-getting component. Also, a thin portable customer application can be

worked to track information utilization examples to take into account the fine tuning of the dataset pre-getting control display. (Badari N. Thyamagondlu, 2013)

For the discourse of Decision support system for bandwidth expansion according to the increment of users, the research “Bandwidth Consumption and Broadband Reliability” is also contributing a considerable logic. Thus, it is an examination of the fast development of Internet-associated gadgets in the house is changing the way broadband administrations are devoured. In the previous year and a half, the level of family units with at least four gadgets getting to the web expanded from 32 to 43 percent. In the meantime, the measure of bandwidth utilization per home is developing and is required to develop to four times by 2015. The development in portable cell phones and tablets speaks to the quickest developing area of associated gadgets, and offloading information activity to Wi-Fi systems is further driving the development of home systems administration. This expanded use over different gadgets makes another arrangement of difficulties for broadband suppliers. To begin with, it expands the cost to set up and bolster these more mind boggling home systems, contrasted with the customary "one PC" home. Second, it makes conflict inside the home system for a common asset - the broadband pipe. At long last, it expands general interest for bandwidth, which thus adds center limit cost to the specialist organization. The Cisco Prime™ Home application and gadget administration platform was conveyed to the market to deliver issues identified with the principal issue, particularly the experience of setting up, keeping up, and adding esteemed applications to the oversaw home system, and is in far reaching arrangement today. This white paper will address the second issue of saw execution and bandwidth dispute, and will address the specialist organizations' reaction to the third issue - use based charging and bandwidth tops.

After an absolute scavenge the researches have concluded their research with the followings. The Bandwidth Consumption and Broadband Reliability contemplate uncovers some fascinating patterns and new difficulties developing in the home system. Some broad conclusions can be drawn from this information such as the home system keeps on developing in intricacy, with more gadgets expending wealthier applications,

this development makes asset dispute with in the home, which drives bolster calls and professional truck rolls, consumers for the most part don't have approaches to comprehend and resolve issues alone, nor would they be able to track and deal with their broadband utilization

The changing scene of customer gadgets, applications, and substance conveyance will no uncertainty keep on putting weight on specialist co-ops to convey solid administration. Consolidating this with the want to advance toward a usage based charging model would appear to demonstrate the requirement for better customer training and bolster apparatuses past what is right now accessible. Cisco trusts that specialist organizations should grasp innovations that advance a superior comprehension of utilization, dispute, and execution for the end client. The outcome will be a superior general ordeal for the broadband client and a diminished operational cost for the suppliers. (The Cisco Prime Home solution, 2012)

Finally this background scout can be concluded by exploring “How the speed of the internet will develop between now and 2020” by Ir. Tommy van der Vorst, Ir. ing. Reg. Brennenraedts, Ir. David van Kerkhof, Dr. ir. ing. Rudi Bekker. The key note of the research is "How will transfer and download bandwidth request have created by 2020?" to give the appropriate response, the study initially need to address the accompanying arrangement of sub-questions: Such as whatever degree do as of now accessible applications add to upstream traffic?, What exactly degree has the requirement for upstream movement of at present accessible applications changed as of late?, Which business applications require high upstream movement?, What exactly degree will shoppers utilize more business applications by 2020? What exactly degree will shoppers utilize different applications with a popularity for upstream activity by 2020?, Which upstream and downstream speeds will be adequate for future request? This examination concentrates primarily on customers in the Netherlands and other West European nations with exceedingly created broadband markets. Inquiries 1, 2 and 3 concentrate on the present circumstance, while the last three inquiries manage the future improvement of shopper interest for internet based activity. It is fundamental to settle on a period skyline

with a specific end goal to plan conclusions pertinent for characterizing approach and in addition methodology. As the broadband market is exceedingly unique, we have picked a period skyline of seven years (until 2020). After the exploration they researches came to a conclusion that the future bandwidth interest for existing administrations can be assessed by joining information on current utilization of administrations with projections of the appropriation and development rates of administration use power. Future progressive administrations are required to assume a noteworthy part in the development of bandwidth request. Such administrations are however hard to anticipate. We have created a next best‟ estimation model for these administrations, where we displayed a likelihood appropriation of the effect of unrests and their normal event recurrence.

The strategy created in this examination was connected to information on private memberships in Western Europe. The researches have anticipated the compound yearly development rate (CAGR) of upstream and downstream movement request to be 44% and 40% individually. While request in 2013 is by and large 15.3 Mbit/s downstream and 1.6 Mbit/s upstream, in 2020 request is relied upon to increment to 165.4 Mbit/s downstream and 20.1 Mbit/s upstream. Huge contrasts can be found between the sorts of administrations and the client gatherings. Power clients, constituting 2% of the aggregate clients, will require 1,155 Mbit/s downstream and 315 Mbit/s upstream by 2020, though the slouches will just need 6.6 Mbit/s downstream and 0.8 Mbit/s upstream at that point. (Ir. Tommy van der Vorst, 2014)

2.4 Post Script

In considering the literature related to the project of implementing a Bandwidth expansion support system through data mining to bear the accordance to increment of usage, it should be duly noted that there is no research that has conduct which can absolutely related to the above mentioned topic. But however the reviewing the literature of data mining techniques and bandwidth expansion needs and approbation of bandwidth and its related notes can give an understanding about how important this research is and how useful it will be in the future and also some directions to properly conduct the expected project.

2.5 Summary

Evidently this chapter provided a thorough description of the current measure of the particular matter in the focus and related literature was reviewed to get an idea to execute this project. It presented a comprehensive analytical and critical appraise of data mining and bandwidth related keynotes. Even though it is not entirely connected to the selected topic or utterly relate to telecommunication industry, it delivered a clear idea to what technologies to use and possible ways to address this selected topic to enact a successful project.

Chapter 03

Methodology

3.1 Introduction to the chapter

The chapter one was about the span of which this research “Decision support system for bandwidth expansion according to the increment of users: Critical Review” supposed to discover and the chapter two was the analysis of how data mining and bandwidth expansion has been researched by various expertise in the past out of which to make a prediction of where this research can lead to. As per this chapter, it is to discuss the assistance of technologies in the process of developing a support system.

3.2 Technologies used for the Tool

3.2.1 What is Data mining

Every person works in the IT sector has to deal with complex data in large volumes which may generates via computers or networks or humans. Maintaining these data sets is one of the top most priorities in companies and organizations; thus they have contributed considerable amount of resources to conserve and preserve these large volume data. But it should be also noted that in terms of decision making, experts uses a sample of these large scale data since it is not practical to manage such large sets of data or rather complicated and time/ cost consuming to do operations with such large data bindles. The largeness has made it comprehensive to analyze. Therefore in the current context, the experts are keen on extracting the useful knowledge concealed in these large data sets which does add value to an effective analysis. The “Data Mining” is the computerized stratagem that is used to filter and discover the knowledgeable data.

Data mining is a process for analyzing and summarizing large scale data bundles in to useful processed data which means information via different techniques with respect to the various needs and perspectives. The process is known as a non-trivial process for identifying valid, genuinely useful and literally comprehensible data patterns. Finding useful knowledge within the raw data is usually done by a sequel of steps. First and foremost, application domain has to be defined and fixing a target data set referring the Intelligent method of data selection by focusing on a subset of variable or data samples,

data cleaning and preprocessing, data reduction and projection, data mining task selection, data mining algorithm selection, interpretation of clever patterns and consolidation of knowledge. Ergo the data mining process can be considered as a total solution comprising the understanding data and preparation of business phases, modeling, evaluation and deployment in an interactive manner as shown in the figure 1

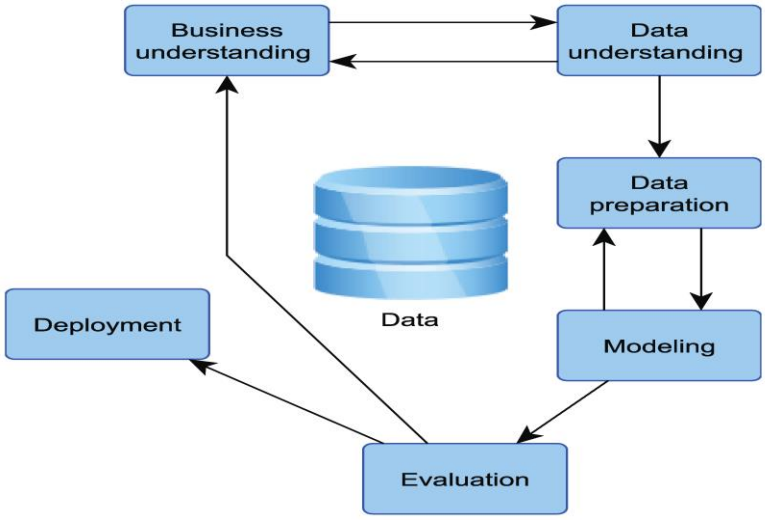


Figure 1 Data Mining Process

Being predictive or descriptive are the two preliminary goals of having a data mining. Depending on the composition of the project, different techniques can and will be used. If the target is predictive; classification, regression and deviation detection techniques can be used and if the project needs a descriptive interpretation; association rules, cluster analysis techniques are used. Predictive algorithms tends to deliver predictions of the values of variables when given input data whereas in the descriptive algorithm tends to deliver a summarization of given input data. The convoluted task would be to find the ideal approach depending on the requirement. (Figure 2)

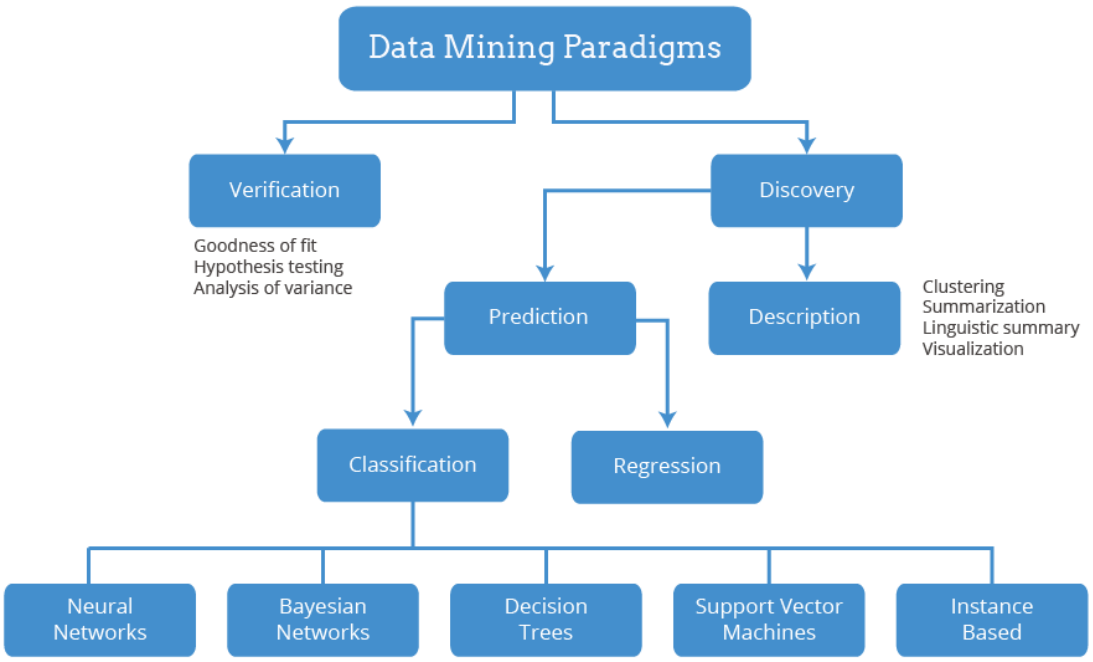


Figure 2 Data Mining Paradigm

This beneficial and useful tool is currently used by experts because of the reliability of the results. Especially analysts use this technique in the telecommunication industry purposes. Even the astronomy, pattern discovery in biology, marketing, fraud detection in finance, defect findings in manufacturing and governance uses this interesting tool and its various applications like web content mining, web usage mining and stream data mining for the daily professional processes.

3.2. 2 Cross-industry standard process

Cross-industry standard process or CRISP as generally termed in its acronym is used for data mining and it is data mining processes demonstrate that portrays generally utilized methodologies that data mining specialists use to handle issues which remain the most mainstream approach for investigation and data mining. CRISP-DM breaks the procedure of data mining into six criterions. The grouping of stages is not strict and it is constantly necessary to swing back and forth between different stages. The bolts in the process outline demonstrate the most critical and successive conditions between stages. he external hover in the outline symbolizes the cyclic idea of data mining itself. In a

nutshell, through this thesis data analyzing process will be followed by starting Business understanding where the concentration will be focused on understanding the task Destinations and necessities from a business point of view and, subsequently, the transformation of this information into a definition of data mining issues and a preparatory arrangement to achieve the objectives and followed by data understanding Which begins with the collection of the underlying data and continues with exercises taking into account the end goal of familiarizing with the data, distinguishing data quality issues, finding first bits of knowledge in the data or identifying intriguing subsets to shape theories for concealed data preparation- where all exercises to develop the last dataset from the underlying raw data are concealed. (Undertakings incorporate table, record, and quality determination and change and cleaning of data for demonstrating apparatuses) will cover to Modeling – in which different demonstrating procedures are chosen and connected to Evaluation- in the phase in which it will be possible to constructed a model (that seems to have high caliber, from a data investigation viewpoint. Before continuing to conclusive arrangement of the model, it is vital to all the more altogether assess the model, and audit the means executed to build the model, to be sure it legitimately accomplishes the business goals. A key objective is to determine whether there are some important business issues that have not been adequately addressed. At the end of this stage, a choice on the utilization of the results of data mining should be reached and finally concluded with the deployment of the model. Whatever the possibility that the motivation behind the model is to build data information, the learning gathered should be sorted out and introduced in a way that is valuable to the customer. The organizational phase can be as straightforward as reporting or mind - boggling as performing a repeatable data scoring or data mining process.

3.2.3 Data mining tool- RapidMiner

Data mining can be utilized for the procedure of extraction of fascinating, nontrivial, verifiable, already obscure and conceivably helpful examples or learning from immense measures of renting customers‟ information, and for the forecast of future inclinations of client installments. It is the arrangement of exercises used to discover new, covered up or sudden examples in data or strange examples in data. Utilizing data contained inside information distribution center.

In this research, well acknowledge RapidMiner Tool was used in data mining. It was the free trial version offered to the university student for a year period. RapidMiner utilizes a customer/server demonstrate with the server offered as either on-introduce, or openly or private cloud infrastructures. RapidMiner uses can be stretched out with extra modules which are made accessible by means of RapidMiner Marketplace. The RapidMiner Marketplace gives a stage to designers to make information examination calculations and distribute them to the network.

Thus, the carried out research thesis has used the RapidMiner version 7.0 which provides updates to its getting started materials, an updated user interface, and improvements to its data preparation capabilities.

3.2.4 Java/ JSF/HTML

Implementation of the frontend interface of the suggested system will be created using Java/ JSF/ HTML languages. Java is acknowledged for its adaptability from different computers. It has the ability to run one specific program on many different systems as preferred is a duly significant advantage to World Wide Web software. At both the source and binary levels Java is being platform-independent hence it can succeed in the above mentioned advantage.

3.2.5 QoS Attribute

The Quality of Service (QoS) alludes to the capacity offering a service with a specific quality system. It is possible to identify the nature of a service with various parameters. Much utilized parameters are:

- Availability of a link
- Number of bit errors
- Latency (delay in the network)
- Jitter

Which parameter is important depends on the service Voice and video service, for example, require low dormancy but still endure a certain error rate. Non - specific information applications cannot be mistaken by differentiating, but dormancy is not fundamental.

In terms of the output result representation, it get delivers as in high, low or medium in the actual record.

In reference to this research, the researcher has used QoS as one of the indirect component to measure the quality of the service in terms of bandwidth expansion need that the clients face. Even though it will not provide a spick and span idea of the matter in the hand, the researcher has identified that it will still provide an idea along with a hazed direction for the blueprint of the intended decision support system implementation.

3.2.6 Average cause for Record closing (AVG_CAUSE_FOR_REC_CLOSING)

Via this field, the user can obtain the indications as in "The reason behind the session drop once a user connected to the session"

In terms of the output result representation, it delivers as in Time Limit (when the given time to surf ceased), Volume Limit (when the give volume to surf the internet ceased), Max Changed Condition (when the user gets terminated), Normal Release (where some abnormality arises in terms of network or hardware failure) and Management Intervention (when the service provider cease the session due to a reason)

3.2.7 MS SQL

MS SQL or the Microsoft SQL Server is a system which in technically a relationship management system offered to the market my Microsoft. This particular database server can be recognize as a abet product for software applications which primarily provide the users the ability to store and retrieve data depending on the requests done by other software application. It is a software product which can be used in the same or a different computer in a network implying the internet as well for that matter.

The development of the backend of the decision support system of bandwidth expansion according to the increment of users has benefited from this tool.

3.2.8 MS Excel

Microsoft Excel will be used in this project as spreadsheet software since the project is associated with large set of data bundles which needs to sort accordingly. It was selected due to MS Excels compatibility with other application which has being used in the project for the purpose of organizing and formatting data.

3.3 Summery

As it was expressed in this chapter, data mining is a tool which can be used to analyze large set of data with regards to the purpose. In this chapter, I was keen on explaining the data mining tool that I was expecting to use in the development process of this decision support system which are as mentioned, CRISP, Rapidminer, ava/ JSF/HTML, MS SQL, QoS Attribute, MS Excel and Average cause for Record closing.

Chapter 04

Novel approach to predict the Bandwidth where the infrastructure upgrade is need the most according to the user/ usage

4.1 Introduction

Chapter three presented the technology to be used to solve the research problem. This chapter thoroughly examines and analyzes the data gathered on the sample focusing the independent variables which is the increment of data usage and the dependent variable which is the bandwidth expansion. The findings of this research study and the subsequent evaluation carried out on the responses reflected the key areas of increase of data usage and the link amongst the factors on the need of bandwidth expansion. This chapter described our approach to address the problem Bandwidth decreasing due to the insufficient amount of e node b problem. Analyzing real time data to predict where new e node bs are required is the solution, which addressed by the Decision support system for bandwidth expansion according to the increment of users. We present our approach by highlighting hypothesis, input, output, process, users and features of the CRIC-Win Predictor solution.

4.2 Hypothesis

We hypotheses that the issue of uses getting frustrated about the lower bandwidth issue due to insufficient number of e node b's can be solved by increasing e node bs and introducing a system to analyses user data to predict where to focus service providers attention will make the process easier. This hypothesis was influenced by the fact that freely available cricket data can be liable to make competitive advantages by simulating game in given situation. The hypothesis of this research is that the focus areas of the Sri Lanka according to the increment of the users can be predicted by using classifier analysis. Researcher is going to use various classify technologies such as Naïve Bayes, Decision Tree, K-NN and then finally picked up most accuracy classifying techniques based on data model.

4.3 Process Model

The development of the proposed system is in 4 steps as follows

- Data gathering,
- Data preprocessing,
- Data mining,
- Data visualization.

One of the data mining techniques is “Data Preprocessing” that involves transforming raw data into a readable or an understandable format. Real-world data is often inadequate, conflicting, and additionally ailing in specific practices or inclines, and is probably going to contain numerous mistakes. Data preprocessing is a demonstrated method of settling such incompleteness’s. Data preprocessing initially prepares raw data for processing further. Inpre processing data runs under the following steps,

I. Data Cleaning: Via a method like filling in missing values, smoothing the noisy data or resolving the inconsistencies in the data, data gets cleansed.
II. Data Integration: To sort the conflicts within the data, different representations are put together
III. Data Transformation: Data is undergoing in normalizing, aggregation and generalization.
IV. Data Reduction: After considering the data in a data warehouse, by undergoing this step; it is expected to get a reduced representation of them.
V. Data Discretization: Involves the reduction of a number of values of a continuous attribute by dividing the range of attribute intervals.

This step is really important in the final decision making, because unnecessary data can lead to misguided decisions in the end. Different data types have to be handled in this phase. Data has to be transformed into other representation in order to improve the use of different algorithms in the analysis process. Because some analysis algorithms require nominal data binomial or Boolean

Scenes the collected data set have large number of unnecessary attributes as well next important thing is to extract the relevant subset.

After finishing data preprocessing the researcher has the ability to select a proper mining technique to go forward data mining there are few Good techniques which can used for the cause, such as clustering, classification, association, regression analysis.

Classification is a function of data mining which assigns items to target categories or classes in a collection. The classification objective is to predict the target class accurately in the data for each case. A classification model, for example, could be used to identify loan applicants as low, medium or high credit risks.

A classification task starts with a set of data that knows the class assignments. For example, a classification model that predicts credit risk could be developed over a period of time on the basis of data observed for many loan applicants. Besides the historical credit rating, the data could track job history, home ownership or rental, years of residence, number and type of investment, etc. Credit rating would be the target, predictors would be the other attributes, and a case would be the data for each customer.

Classifications are unobtrusive and do not imply order. Continuous, floating - point values would indicate a target that is numerical and not categorical. A numerical target predictive model uses a regression algorithm, not an algorithm for classification.

Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes; called clusters that help users understand the natural grouping or structure in a data set.

Clustering is an unattended classification technique that is either used as a stand - alone tool to gain insight into the distribution of data or as a pre - processing step for other algorithms. K - means acceptance is mainly because it is simple. The algorithm is also suitable for clustering large data sets C because it has much less computational complexity by increasing the data points linearly. Decide on the simplicity of this technique it suffers from certain disadvantages such as determining the number of clusters by the user effectively from the outer Data of high dimensional data and sensitivity to industrial cluster centers and this possibility of being trapped in a local minimum can reduce the efficiency of the k - means algorithm.

4.4 System Over view

The process will start from collecting the data and then analyzing the collected data to figure out where users will be frustrated the most due to bandwidth depletion compared to the user need and infrastructural issues in order to fix it before it get worse or out of hand.

Figure 3 System Overview

4.5 Raw Data Presentation & Inputs

These collected data was taken as a sample of the total data users in Sri Lanka in an average day. Particularly this was a random sample which was distinguished from the client tail of Lanka Bell (pvt) ltd on an arbitrary day within the twenty four hours.

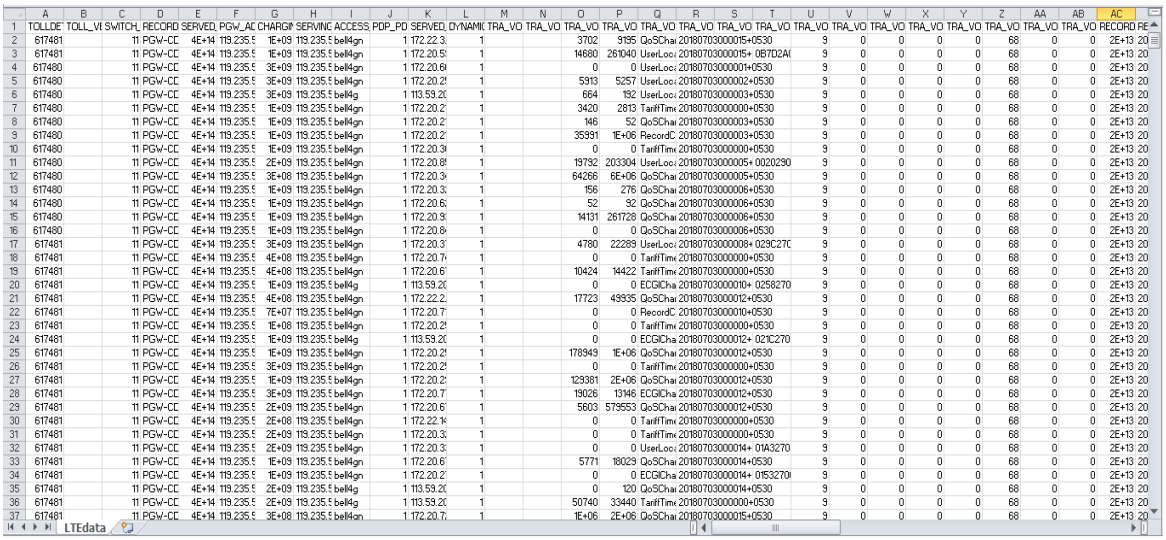


Figure 4 Raw Data Sample

4.5.1 Inputs

Input for conducting research, I collected 4GLTE data from Lanka Bell (Pvt.) Ltd for duration of 5 months. Attributes Selected for the classifier analysis were MSISDN, DOWNLINK SPEED, UPLINK SPEED,BTS NAME (for the regions where, already have more than one node),QOS, record closing reason. And researcher assumed if they do not meet the specs the company said the will provide user will not be satisfied and if those meet what they said the will provide users will be satisfied.

Table 1 Attributes (Inputs)

SERVED_MSISDN	VOL_VOLUME_UP_LINK	VOL_VOLUME_DOWN_LINK	REGION	AVG_CAUSE_FOR_REC_CLOSING	NETWORK_QOS	USER_STATUS
94110210545	0	0	Aluthgama	TimeLimit	High	Satisfied
94210100646	2471961	13584523	Aluthgama	MaxChangeCond	High	Satisfied

There are lots of data mining techniques as mentioned earlier, but researcher try to focus on two techniques: clustering and classification. Selected algorithms from each technique, trained using rapidminor.

4.6 Output

As main output of this process will be predicting, which area will need a infrastructure upgrade in near future. Web based software tool that accept the rapidminor generated results (the predictions and the use clusters) as an excel file to give user the predicted output for a better understanding. This tool is called decision support system for bandwidth expansion according to the increment of usage and it analyses given data to make strategic predictions about hardware upgrade on the ground in order to provide a better service to the consumer. The solution includes following components.

- Data Analytic Engine
- Excel Data files
- Web UI

4.7 Stakeholders of the solution

Local telecommunication companies will be the users and this will be a huge advantage for them. Local telecommunication companies do not invest in these sorts of subordinate tools if they can successfully accomplish the primitive need which is providing internet facilities as per this scenario. They are in a rush to cover their geographic arena entirely from signaling e node b s but with the user increment/ usage increment and other infrastructure troubles customers get disappointed. This solution will be tailor made for such users.

4.8 Features of the solution

The features of the solution include the following in a broader sense. Among other features, the best thing in the solution is module based development; this contains three layers/ modules UI Module, prediction Engine and database Module. The prediction engine can be included to any other program if needed which implements low coupling. So prediction engine is separated rapid minor module as well as it can be act as individual component in future development. There are several other advantages of the solution.

- User Friendliness
- Learnability
- Security
- Operator independency

4.9 Summery

This chapter brought forth the data mining approach for the particular research. The researcher has mentioned the hypothesis, input, output, process along with the features to provide the expected understanding. Accurately conducting this would lead to more reliable decision making support system for bandwidth expansion according to the increment of users in the end.

Chapter 05

Implementation

5.1 Introduction

In the chapter four, the research conductor has explained the process along with the overview of the system or in simple term it was presented the approach to develop the solution by using data mining technology. This chapter will depict the overall picture of the proposed solution and the design of the solution with its modules. This solution mainly has three modules namely, User Interface, Data mining Engine and Database module. So this chapter will describe overall architecture, the system design and the individual modules of the solution.

5.2 Top Level Architecture Design (System Design) of the solution

Under this criterion, it will be discussed how the overall architecture will be designed and as it was explained previously this solution will be a modular solution out which the rapidminer analytic engine module will process the given inputs and deliver the predictions as the output whereas the Web- UI will display the results in the front end in a more of a user friendly convenient manner for the end user.

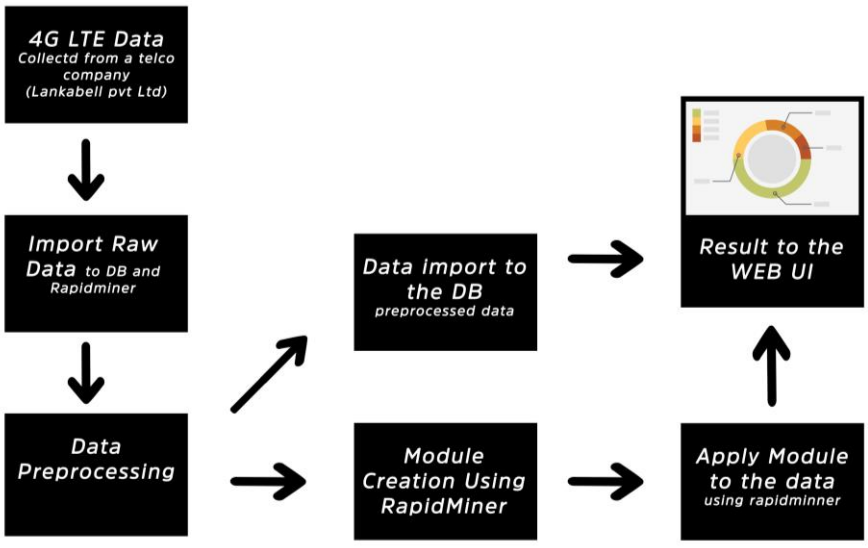


Figure 5 System Design

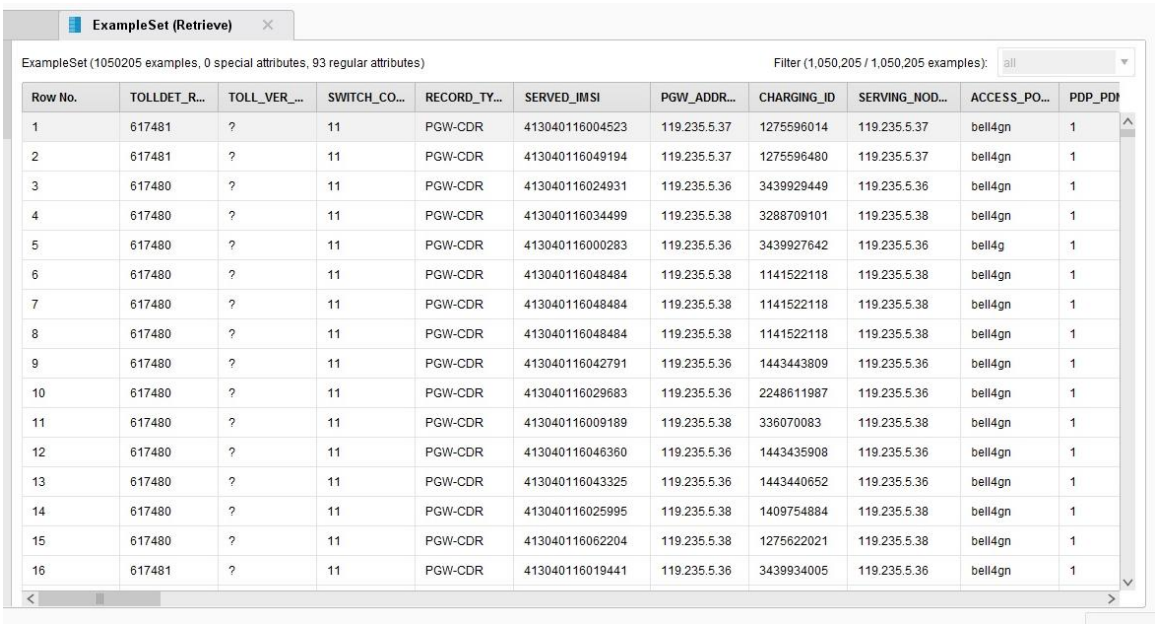
As it is thoroughly visible at a glance in the above figure, consumer usage 4G LTE data will be collected from a telecommunication company (Lanka Bell (Pvt.) Ltd apropos to this thesis) which will be subjected to rapidminer preprocessing process hence there could be certain outliers, missing values, anomalies and etc. which the researcher is willing to get rid of; after which that above mentioned cleaned data will be export in to the main DB and to the analytic engine.

It should be noted that while conducting this research, the researcher has doubts with regards to the absolute data mining method would provide the best results. Therefore the researcher is going to build an analytic engine in both supervised and unsupervised manner using the clustering algorithms and classification algorithms in order to select the ideal prediction method.

5.3 Data Collection and Data Preprocessing

5.3.1 Data Collection

The Raw data collection



ExampleSet (Retrieve)

ExampleSet (1050205 examples, 0 special attributes, 93 regular attributes) Filter (1,050,205 / 1,050,205 examples): all

Row No.	TOLLDET_R...	TOLL_VER_...	SWITCH_CO...	RECORD_TY...	SERVED_IMSI	PGW_ADDR...	CHARGING_ID	SERVING_NOD...	ACCESS_PO...	PDP_PDI
1	617481	?	11	PGW-CDR	413040116004523	119.235.5.37	1275596014	119.235.5.37	bell4gn	1
2	617481	?	11	PGW-CDR	413040116049194	119.235.5.37	1275596480	119.235.5.37	bell4gn	1
3	617480	?	11	PGW-CDR	413040116024931	119.235.5.36	3439929449	119.235.5.36	bell4gn	1
4	617480	?	11	PGW-CDR	413040116034499	119.235.5.38	3288709101	119.235.5.38	bell4gn	1
5	617480	?	11	PGW-CDR	413040116000283	119.235.5.36	3439927642	119.235.5.36	bell4g	1
6	617480	?	11	PGW-CDR	413040116048484	119.235.5.38	1141522118	119.235.5.38	bell4gn	1
7	617480	?	11	PGW-CDR	413040116048484	119.235.5.38	1141522118	119.235.5.38	bell4gn	1
8	617480	?	11	PGW-CDR	413040116048484	119.235.5.38	1141522118	119.235.5.38	bell4gn	1
9	617480	?	11	PGW-CDR	413040116042791	119.235.5.36	1443443809	119.235.5.36	bell4gn	1
10	617480	?	11	PGW-CDR	413040116029683	119.235.5.36	2248611987	119.235.5.36	bell4gn	1
11	617480	?	11	PGW-CDR	413040116009189	119.235.5.38	336070083	119.235.5.38	bell4gn	1
12	617480	?	11	PGW-CDR	413040116046360	119.235.5.36	1443435908	119.235.5.36	bell4gn	1
13	617480	?	11	PGW-CDR	413040116043325	119.235.5.36	1443440652	119.235.5.36	bell4gn	1
14	617480	?	11	PGW-CDR	413040116025995	119.235.5.38	1409754884	119.235.5.38	bell4gn	1
15	617480	?	11	PGW-CDR	413040116062204	119.235.5.38	1275622021	119.235.5.38	bell4gn	1
16	617481	?	11	PGW-CDR	413040116019441	119.235.5.36	3439934005	119.235.5.36	bell4gn	1

Figure 6 Raw Data Collection

The status of the above Raw Data

	A	B	C	D
1		Type	Missing	Statistic
2	TOLLDET_REF_NUM	Integer	0	617583.291
3	TOLL_VER_NUM	Polynominal	1050205	
4	SWITCH_CODE	Integer	0	11
5	RECORD_TYPE	Polynominal	0	PGW-CDR (1050205)
6	SERVED_IMSI	Real	0	4.13E+14
7	PGW_ADDRESS	Polynominal	0	119.235.5.36 (428312), 119.235.5.38 (326771), ... [1 more]
8	CHARGING_ID	Real	0	1.78E+09
9	SERVING_NODE_ADDRESS	Polynominal	0	119.235.5.36 (428312), 119.235.5.38 (326771), ... [1 more]
10	ACCESS_POINT_NAME_NI	Polynominal	0	bell4gn (947327), bell4g (99787), ... [7 more]
11	PDP_PDN_TYPE	Integer	0	1
12	SERVED_PDP_PDN_ADDRESS	Polynominal	0	172.22.5.220 (20911), 172.20.85.174 (15082), ... [22846 more]
13	DYNAMIC_ADDRESS_FLAG	Integer	0	0.994441085
14	TRA_VOL_QOS_REQUESTED	Polynominal	1050205	
15	TRA_VOL_QOS_NEGOTIATED	Polynominal	1050205	
16	TRA_VOL_VOLUME_UP_LINK	Integer	0	1601209.211
17	TRA_VOL_VOLUME_DOWN_LINK	Integer	0	1.58E+07
18	TRA_VOL_CHANGE_CONDITION	Polynominal	0	ECGIChange (420188), UserLocationChange (329587), ... [3 more]
19	TRA_VOL_CHANGE_TIME	Polynominal	0	20180703010000+0530 (25092), 20180703080000+0530 (20493), ... [46876 more]
20	TRA_VOL_FAILURE_HANDLING_CC	Polynominal	1050205	
21	TRA_VOL_USER_LOCATION_INFO	Polynominal	302467	02C12A0040F314EC0340F31418 (23411), 0020290040F314ED0340F31418 (23152), ... [1372 more]
22	TRA_VOL_E_QCI	Integer	0	8.980221004
23	TRA_VOL_E_MAX_REQ_BW_UL	Integer	0	118.3210307
24	TRA_VOL_E_MAX_REQ_BW_DL	Integer	0	118.3210307
25	TRA_VOL_E_GUARANT_BIT_RATE	Integer	0	80.61105975
26	TRA_VOL_E_GUARANT_BIT_RATE	Integer	0	80.61105975
27	TRA_VOL_E_ARP	Integer	0	67.90749396
28	TRA_VOL_E_APN_AGG_MAX_BITF	Integer	0	0
29	TRA_VOL_E_APN_AGG_MAX_BITF	Integer	0	0
30	RECORD_OPENING_TIME	Real	0	2.02E+13
31	RECORD_OPENING_TIME_FULL	Polynominal	0	20180703010000+0530 (25091), 20180703000000+0530 (21890), ... [46874 more]
32	UNIX_TIME_STAMP	Integer	0	1.53E+09
33	DURATION	Integer	0	873.6314767
34	CAUSE_FOR_REC_CLOSING	Polynominal	0	MaxChangeCond (776642), TimeLimit (211290), ... [4 more]
35	DIAGNOSTICS	Polynominal	1017604	ManufacturerSpecificCause:1 (30248), ManufacturerSpecificCause:96 (1009), ... [5 more]
36	RECORD_SEQUENCE_NUMBER	Integer	0	4936.404173

Figure 7 Raw Data Status

5.3.1.1 Pre- Preparation

After gathering data and reviewing the problem, it was duly noted the need and the vital characteristics of the system that is supposed to be implemented. For the streamlining, a requirement, process and outline were plotted and the skeleton came out as the following.

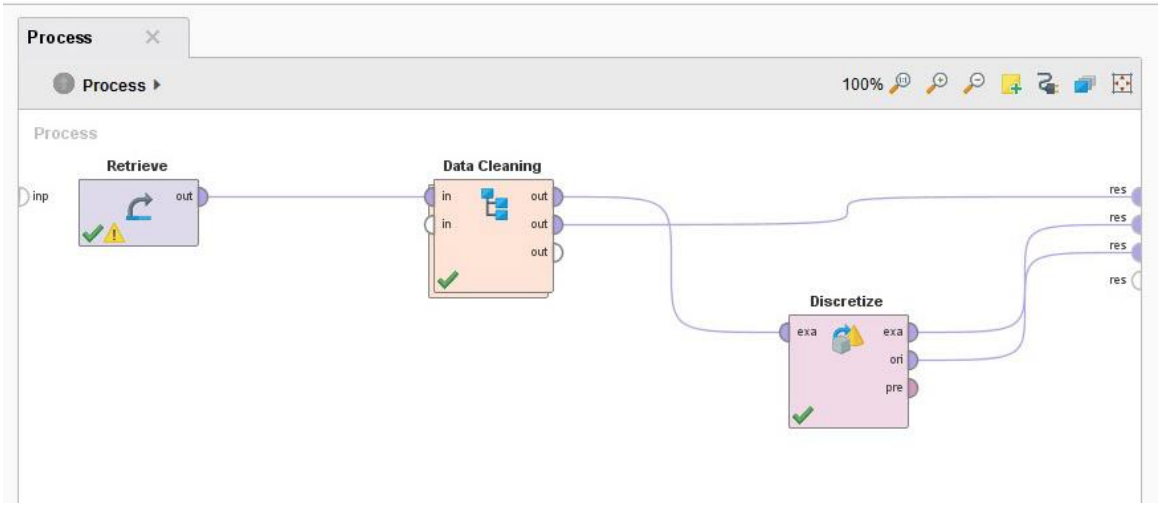


Figure 8 Data Cleaning (Pre-preparation)

Using the RapidMiner tool, the data cleaning part was sorted in the inceptive juncture.

5.3.2 Data Preprocessing

Significantly the elementary steps demanded in data preprocessing are Data Cleaning, data integration, data reduction, and data transformation.

For the phase of data cleaning, it has been done through filling in missing values, noisy data smoothing, to spot or take away outliers, and to resolve inconsistencies are conducted. Therefore as to filling missing values, ignore the tuple (when the category label is missing), fill within the missing values manually, using a scope of global to fill within the values that are missing, use a live of central tendency for the attribute to fill within the missing value, use the attribute mean or median for all samples affiliated to identical category because the given tuple and use the foremost probable value to fill within the missing worth.

Data integration conjointly handle major role of data mining. The target goal was to merge information from multiple data stores. Hence to assist scale back and avoid redundancies and inconsistencies within the ensuing data set, keen data integration was attentively conducted. Thus the accuracy and speed of the next data {processing} process is going to be improved. Data reduction are often accustomed get a shrunken illustration of the selected data set that's abundant lesser in volume, nonetheless ultimately maintains the integrity of the initial raw data. This stated data reduction methods embody spatiality reduction, number reduction, and information compression.

The reason behind the conducted dimensionality reduction was to diminish the quantity of irregular factors or traits under the scope. Dimensionality diminishment strategies are wavelet transforms, central segments investigation, and trait subset choice. Numerosity reduction strategies supplant was used in the initial data volume into understandable microscope forms of data presentations.

In this thesis, it has to be noted that without considering the technique if it is a parametric or nonparametric it has been used for the need of storing the data parameters rather than the real data (regression and log-direct models). The transformations are connected in data compression to acquire a diminished or "compressed" representation of the initial original data. It should be also observed that the data is subjected transformations or consolidation into several forms accordingly for mining in data transformation

5.3.3 Data Cleaning

The statistic after the process of data cleaning and the well sort of assortment is as follows.

	A	B	C	D
1		Type	Missing	Statistics
2	PGW_ADDRESS	Polynominal	0	119.235.5.36 (280954), 119.235.5.38 (203443), ... [1 more]
3	SERVING_NODE_ADDRESS	Polynominal	0	119.235.5.36 (280954), 119.235.5.38 (203443), ... [1 more]
4	ACCESS_POINT_NAME_NI	Polynominal	0	bell4gn (622234), bell4g (64479), ... [7 more]
5	CAUSE_FOR_REC_CLOSING	Polynominal	0	MaxChangeCond (480611), TimeLimit (161379), ... [4 more]
6	APN_SELECTION_MODE	Polynominal	0	MSorNetworkProvidedSubscriptionVerified (687603)
7	CH_CH_SELECTION_MODE	Polynominal	0	ServingNodeSupplied (687603)
8	USER_LOCATION_INFORMATION	Polynominal	0	1814F34003ED14F34000292000 (12365), 1814F34003EC14F340002AC102 (12065), ...
9	SER_D_CHARGING_RULE_BASE_NAME	Polynominal	0	cpsvlt_abb04_50g (505951), cpsvlt_abb04_70g (99543), ... [24 more]
10	SER_D_SERVICE_CONDITION_CHANGE	Polynominal	0	ECGIChange (251655), UserLocationChange (208514), ... [6 more]
11	SERVING_NODE_TYPE	Polynominal	0	gTPSGW (687603)
12	CELL_ID	Polynominal	0	2750 (17692), 2A75 (13775), ... [348 more]
13	BTS_NAME	Polynominal	0	Ketawalamulla_LB_10064 (17692), Boralesgamuwa Egodawatta Rd_SE_10869 (13775), ...
14	SERVED_IMSI	Real	0	4.13E+14
15	DYNAMIC_ADDRESS_FLAG	Integer	0	0.995171632
16	TRA_VOL_VOLUME_UP_LINK	Integer	0	2436232.092
17	TRA_VOL_VOLUME_DOWN_LINK	Integer	0	2.41E+07
18	DURATION	Integer	0	1031.310118
19	SERVED_MSISDN	Real	0	9.42E+10
20	SERVED_IMEISV	Real	0	7.93E+15
21	SER_D_DATA_VOL_FBC_UP_LINK	Integer	0	2427297.089
22	SER_D_DATA_VOL_FBC_DOWN_LINK	Integer	0	2.41E+07
23				

Figure 9 Assortment of Data

These above assorted data will be thoroughly scrutinized as mentioned in the below plot.

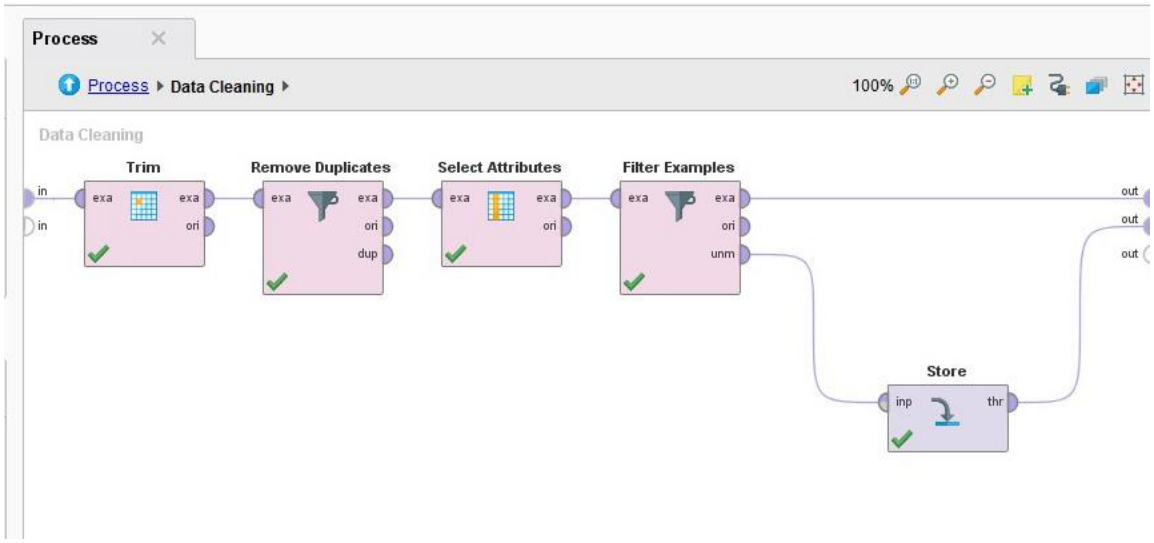


Figure 10 Data assorting Process

5.4 User Interface Model

Through the User Interface Model, it allows the use to continue the interactions with the system for the output requirement. Specifically since the module provides direct access to the main back-end database and analytic engine, it makes this system more useful. The limitation goes beyond handling the output. In terms of the telecommunication industry, they typically get to work with larger data sets hence they would need much predictable output at a glance and this module can facilitate that convenience for them.

5.5 Rapidminer Analytic Process

This process primitively contributes as data mining engine and delivers the results of data preprocessing, filtering the relevant attribute and data mining algorithm. Also in this particular scenario of this thesis classification and clustering algorithm will be contributed by the Rapidminer.

But the outcome of the clustering or the unsupervised algorithm is complicated and not practically useful to make any expected decisions. Hence moving forward, the researcher will use the rapidminer for a classification algorithm to get the expected output which the researcher can actually make predictions and come to decisions based on.

As per the next step, the researcher will get a subset of the preprocessed data set as a training data set. It should be noted that the subset data set will be included the final result whether the user is satisfied or not with the given parameters to get basis to make predictions in the actual context.

Afterwards, the system is fully fledged to provide predictions of the user status when the actual data sets provided.

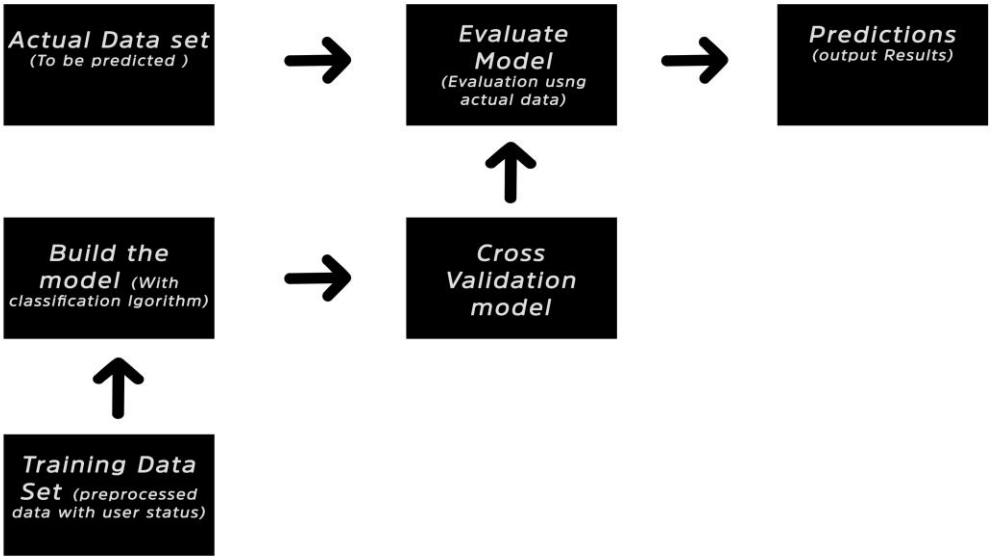


Figure 11 Rapidminer analytic process

5.6 Back end Database

Even though the back-end database does not play the keynote role in handling ongoing data mining in this project, it will act as a warehouse to keep the preprocessed data until it get sent to analytic engine and to hold the results provided by the analytic engine.

5.7 Summery

This chapter brought forth the design of the decision support system and how exactly the researcher is going to construct the system in steps. User interface module, Rapidminer Analytical process and the Back end database was described further in this chapter to provide clarity on how the collaboration of these tools can lead to the accomplishment of the decision support system for bandwidth expansion.

Chapter 6

Discussion

6.1 Introduction

Apropos of preceding chapter, it was discussed the blueprints of the proposed solution for the focused quandary. In this chapter it will bring forth how the implementation process will executed in a detailed manner with regards to each of the individual module mentioned before and its collaboration.

6.2 Overall Implementation

The suggested decision support system for bandwidth expansion according to the increment of users is basically for the system users to get notified when the bandwidth is shared with more than the expected amount of users and is subjected to crowding.

The built system will require the Broadband Data as an input. In terms of broadband data, the researcher meant usage data, uplink/ downlink speed, connected node, QOS, node distribution, connected data times.

The purpose of this demonstrated decision support system is to identify the bandwidth low points, potential overloads and predictions of that regard and of where to increase nodes hopefully to provide a satisfactory service throughout all the time to the users.

To accomplish the above mentioned solution, the proposed system is a web based system which is accompanied by Java Sever Faces (JSF).

Apropos of the back end implementation, the language used is Java and the frame work is spring. The data base was connected using hibernate. This connected data base is a mysql which runs in a hiediSQL platform. Through these front end will display the output result obtained by the Rapidminer analytical process as per which is visible below.

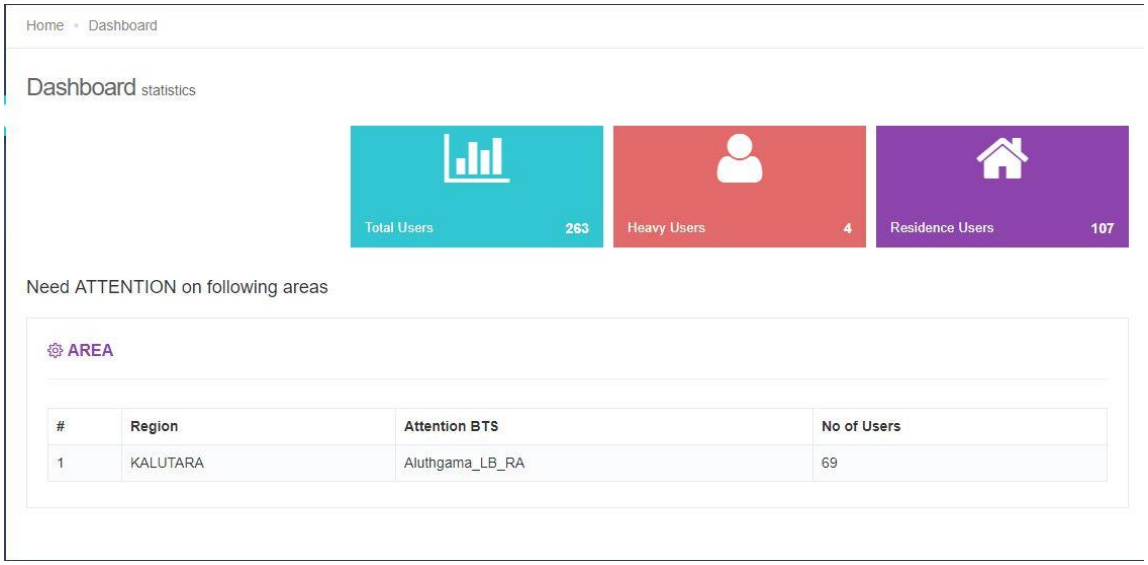


Figure 12 Dashboard Statistics

6.3 Implementation of Rapidminer Analytical Engine

In this decision support system implementation, the core is to deliver the predictions which will perform through the Rapidminer analytical process. The researcher is using Rapidminer 8.2 educational version to be exact to attain this. After completing the preprocessing, the data which will deliver as the output will send to the prediction process. In this engine, the data that has to be predicted get assigned to a role as a label (whereas it will not be considered as an algorithm). These labeled data will sent through the cross validation attribute where data goes through the algorithm like decision tree, naive bayes or K-NN. After the training data set goes through this algorithms, module get trained according to the behavior of the training data whereas when actual date set goes through the same, it will be able to predict the satisfaction status of the users.

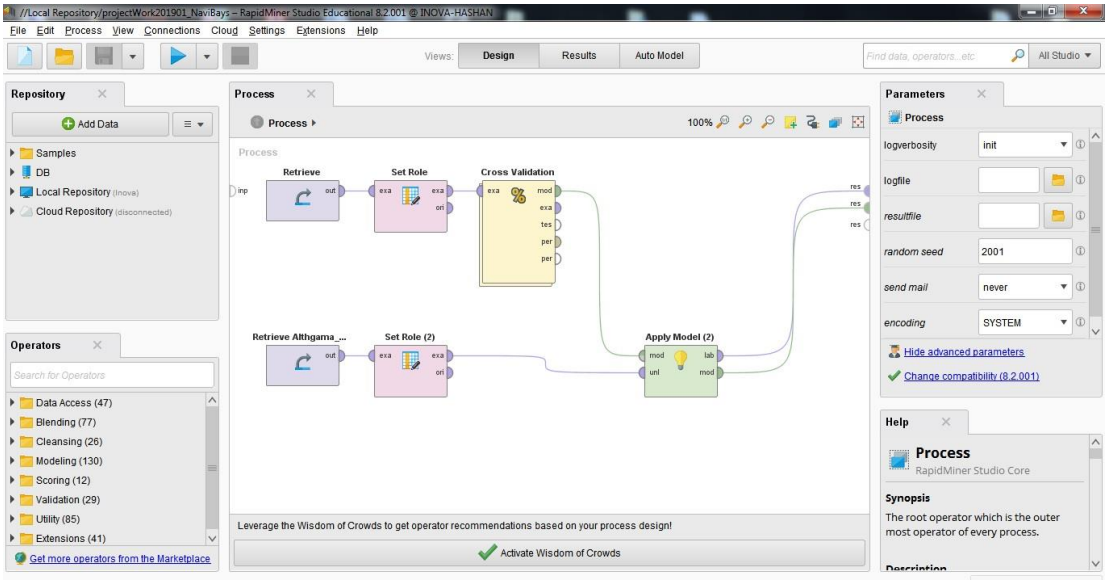


Figure 13 Implementation of analytic process

The above figure will depict the implementation of the analytic process.

When it comes to the cross validation, there are numerous algorithms which can be used accomplish the task such as,

- Linear Classifiers
- Decision Trees
- Boosted Trees
- Random Forest and etc.

In the process of conducting this research, with the interest of getting more accurate results, the same data was entered and inspects the outcome through three different algorithms namely; Decision Tree, Naive Bayes and K-NN. Depending on the accuracy percentage of the predictions, which algorithm to be used when proceeding further would be decided?

Following snaps are the confidence percentages of the predictions of all three algorithms considered.

Decision Tree structure

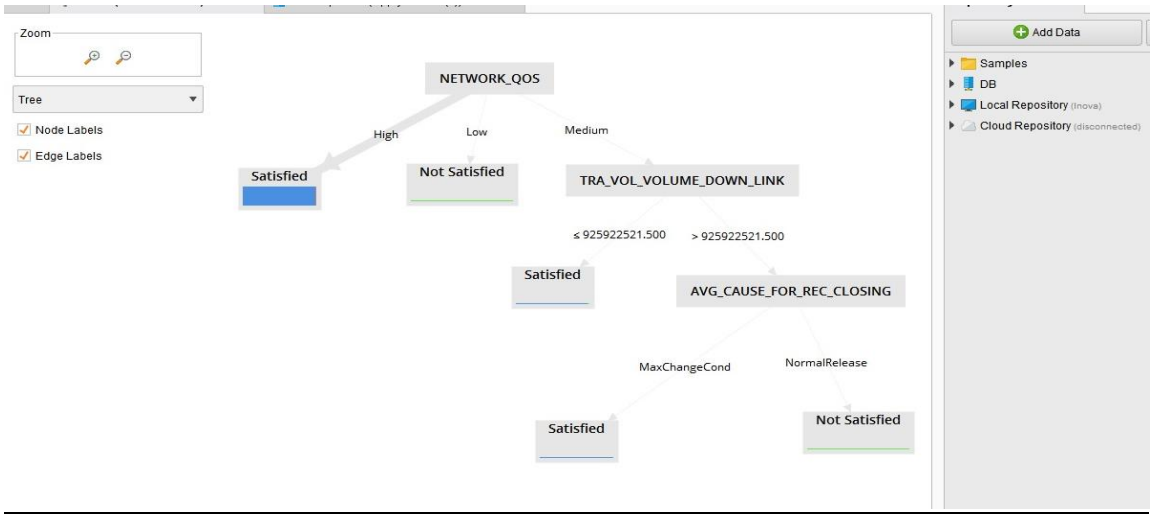
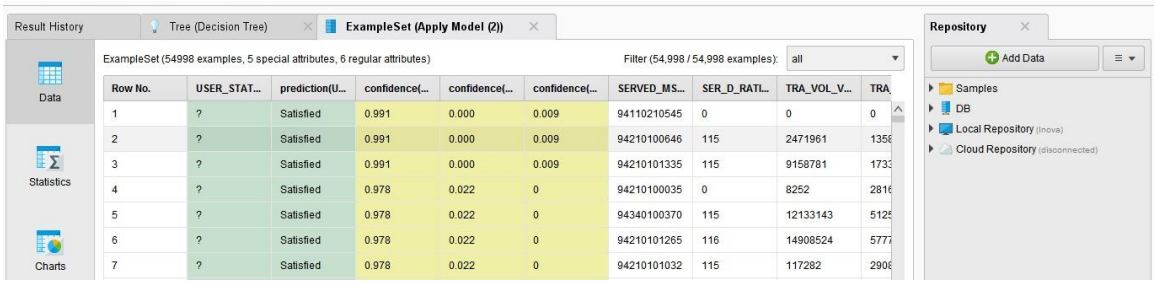


Figure 14 Decision Tree structure

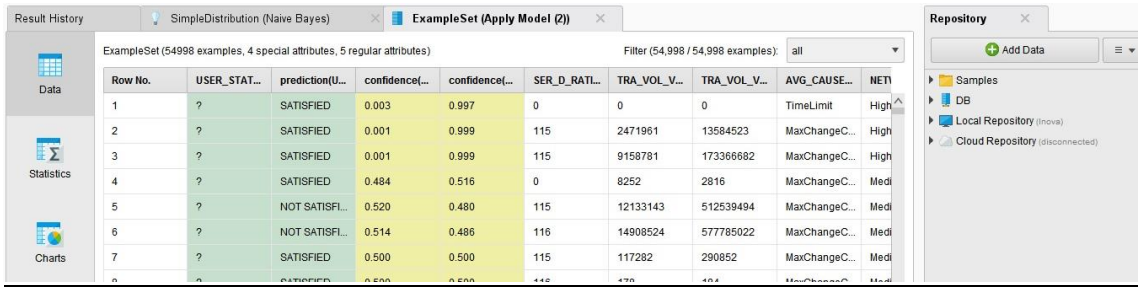
Decision tree predictions



Row No.	USER_STAT...	prediction(U...	confidence(...	confidence(...	confidence(...	SERVED_MS...	SER_D_RATI...	TRA_VOL_V...	TRA
1	?	Satisfied	0.991	0.000	0.009	94110210545	0	0	0
2	?	Satisfied	0.991	0.000	0.009	94210100646	115	2471961	1358
3	?	Satisfied	0.991	0.000	0.009	94210101335	115	9158781	1733
4	?	Satisfied	0.978	0.022	0	94210100035	0	8252	2816
5	?	Satisfied	0.978	0.022	0	94340100370	115	12133143	5125
6	?	Satisfied	0.978	0.022	0	94210101265	116	14908524	5777
7	?	Satisfied	0.978	0.022	0	94210101032	115	117282	2908

Figure 15 Decision Tree Predictions

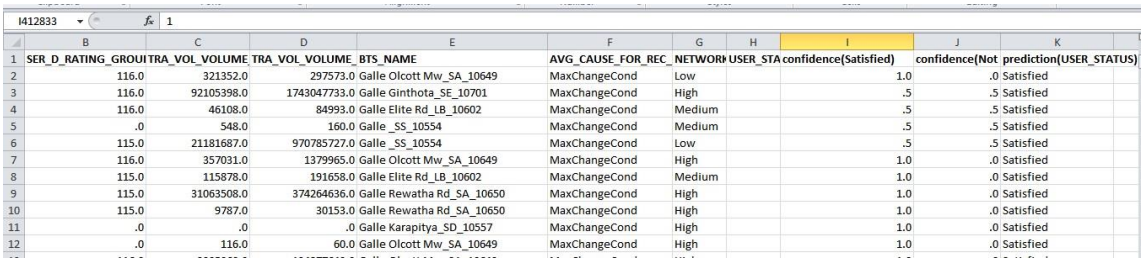
Naive Bayes Prediction



Row No.	USER_STAT...	prediction(U...	confidence(...	confidence(...	SER_D_RATI...	TRA_VOL_V...	TRA_VOL_V...	AVG_CAUSE...	NET
1	?	SATISFIED	0.003	0.997	0	0	0	TimeLimit	High
2	?	SATISFIED	0.001	0.999	115	2471961	13584523	MaxChangeC...	High
3	?	SATISFIED	0.001	0.999	115	9158781	173366682	MaxChangeC...	High
4	?	SATISFIED	0.484	0.516	0	8252	2816	MaxChangeC...	Medi
5	?	NOT SATISFI...	0.520	0.480	115	12133143	512539494	MaxChangeC...	Medi
6	?	NOT SATISFI...	0.514	0.486	116	14908524	577785022	MaxChangeC...	Medi
7	?	SATISFIED	0.500	0.500	115	117282	290852	MaxChangeC...	Medi

Figure 16 Naive Bayes Predictions

K-NN Prediction



I412833 | 1

	B	C	D	E	F	G	H	I	J	K
1	SER_D_RATING_GROUI	TRA_VOL_VOLUME	TRA_VOL_VOLUME_	BTS_NAME	AVG_CAUSE_FOR_REC_	NETWORK	USER_STA	confidence(Satisfied)	confidence(Not	prediction(USER_STATUS)
2	116.0	321352.0	297573.0	Galle Olcott Mw_SA_10649	MaxChangeCond	Low		1.0	.0	Satisfied
3	116.0	92105398.0	1743047733.0	Galle Ginthota_SE_10701	MaxChangeCond	High		.5	.5	Satisfied
4	116.0	46108.0	84993.0	Galle Elite Rd_LB_10602	MaxChangeCond	Medium		.5	.5	Satisfied
5	.0	548.0	160.0	Galle _SS_10554	MaxChangeCond	Medium		.5	.5	Satisfied
6	115.0	21181687.0	970785727.0	Galle _SS_10554	MaxChangeCond	Low		.5	.5	Satisfied
7	116.0	357031.0	1379965.0	Galle Olcott Mw_SA_10649	MaxChangeCond	High		1.0	.0	Satisfied
8	115.0	115878.0	191658.0	Galle Elite Rd_LB_10602	MaxChangeCond	Medium		1.0	.0	Satisfied
9	115.0	31063508.0	374264636.0	Galle Rewatha Rd_SA_10650	MaxChangeCond	High		1.0	.0	Satisfied
10	115.0	9787.0	30153.0	Galle Rewatha Rd_SA_10650	MaxChangeCond	High		1.0	.0	Satisfied
11	.0	.0	.0	Galle Karapitya_SD_10557	MaxChangeCond	High		1.0	.0	Satisfied
12	.0	116.0	60.0	Galle Olcott Mw_SA_10649	MaxChangeCond	High		1.0	.0	Satisfied

Figure 17 K-NN Predictions

After a logical consideration, it was decided that both Naïve Bayes algorithms and K-NN algorithms provides results in a higher accuracy percentage. But due to the lagging of the processing the result in the K-NN algorithm and as the prediction confidence along with the reliability is low, K_NN algorithm was eliminated moving forward. Thus in conclusion, here after this system will present results using the Naïve Bayes algorithms.

6. 4 Summery

In the chapter six, it brought forth the overall implementation process of the Decision support system for bandwidth expansion according to the increment of user. How the each module will contribute in the implementation process along with how training data set will establish the parameter and the process will continue working with actual data sets.

Chapter 07

Evaluation

7.1 Introduction

This chapter is to evaluate and justify the suggested system as a solution for the focused problem of bandwidth expansion based on the increment of users. The main purpose of this chapter is to test the above used strategies within the know parameters to evaluate the accuracy of the system.

7.2 Evaluation of Classification techniques

In the used classification techniques, the trained collected data set went through different classifications such as Naïve Bayes, Decision tree and K-NN by the help of Rapidminer Analytic tool. In terms of evaluating the classifier quality, there are few matrixes that can be used for various measurements such as Accuracy and Classification Error. The below mentioned are few of such matrixes for the reference.

Table 2 Classification Evaluation Measurements

Measurement	Formula	Description
Accuracy	(TP+TN)/(P+N)	The percentage of predictions those are correct, therefore the higher the accuracy better it is.
Classification Error	(FP+FN)/(TP+TN+FP+FN)	Relative number of miss-classified data or in other words percentage of incorrect predictions.

With referring the above mentioned formulas, the actual data will be used to run the models using Decision Tree and Naïve Bayes to get an idea of the accuracy level and the classification error measurement. When it comes to a perfect test, the score has an area of

1.00. Practically best models should have higher accuracy and lower classification rate close to 1.00

With reference to the used classification techniques in this thesis, the accuracy rate and the classification error rates are mentioned in the below table.

Table 3 Comparison of classification techniques to determine overall outcome

Algorithm	Accuracy	Classification Error
Naïve Base	99.90	0.10
Decision Tree	99.10	0.90

According to the above table, it is clear that accuracy rates and almost similar in both Naïve Bayes and Decision Tree the classification Error seems to be lower in the Naïve Bayes. Evidently Naïve Bayes is a better fit in all aspects for this thesis.

7.3 Evaluation of the solution

In the process of conducting this research, Data was collected out from Galle, Aluthgama and Kollupitiya regions from 1st of August 2018 to 14th of January 2019. Out of the data which gathered, from 1st of August to 31st of December 2018 data was used as the training data bundle and from then to 14th of January data was predicted accordingly in the system. The following is the result that the researcher obtained.

The following is the predicted result obtained by the Galle set of Data using the Naïve Bayes algorithm.

	A	B	C	D	E	F	G	I	J	K	L
1	SERVED_MSISDN	SER_D_RA	TRA_VOL_V	TRA_VOL_VOLL	BTS_NAM	AVG_CAU	NETWORI	confide	confid	prediction(USER_STATUS)	
2	2147483647.0	500.0	4928.0	10606.0	Galle Elite	MaxChang	Medium	.3	.7	Not Satisfied	
3	2147483647.0	115.0	15424265.0	226072270.0	Galle Elite	MaxChang	Medium	.6	.4	Satisfied	
4	2147483647.0	115.0	7933516.0	41121566.0	Galle Gint	MaxChang	High	1.0	.0	Satisfied	
5	2147483647.0	115.0	9659867.0	194870326.0	Galle Elite	MaxChang	Medium	.6	.4	Satisfied	
6	2147483647.0	116.0	38922181.0	1858187505.0	Galle Kara	MaxChang	Medium	1.0	.0	Satisfied	
7	2147483647.0	115.0	9907952.0	428566450.0	Galle Olcc	MaxChang	Medium	1.0	.0	Satisfied	
8	2147483647.0	116.0	16428729.0	256544855.0	Galle Olcc	MaxChang	Medium	1.0	.0	Satisfied	
9	2147483647.0	115.0	7909426.0	178484198.0	Galle Kara	MaxChang	Medium	1.0	.0	Satisfied	
10	2147483647.0	116.0	25752.0	24112.0	Galle Olcc	MaxChang	Medium	1.0	.0	Satisfied	
11	2147483647.0	115.0	5854537.0	214380502.0	Galle Elite	MaxChang	Medium	.6	.4	Satisfied	
12	2147483647.0	.0	.0	.0	Galle Gint	MaxChang	High	1.0	.0	Satisfied	
13	2147483647.0	116.0	62.0	.0	Galle Gint	MaxChang	High	1.0	.0	Satisfied	
14	2147483647.0	116.0	25672950.0	1648896551.0	Galle Elite	MaxChang	Medium	.4	.6	Not Satisfied	
15	2147483647.0	115.0	82752556.0	2764663631.0	Galle Olcc	MaxChang	Medium	.9	.1	Satisfied	
16	2147483647.0	116.0	186881.0	526056.0	Galle Kara	MaxChang	Medium	1.0	.0	Satisfied	
17	2147483647.0	116.0	8420371.0	54204442.0	Galle Olcc	MaxChang	Medium	1.0	.0	Satisfied	
18	2147483647.0	.0	.0	.0	Galle Kana	MaxChang	High	1.0	.0	Satisfied	
19	2147483647.0	115.0	1404.0	4379.0	Galle Elite	MaxChang	Medium	.6	.4	Satisfied	
20	2147483647.0	116.0	6380631.0	182552413.0	Galle Kara	MaxChang	Medium	1.0	.0	Satisfied	
21	2147483647.0	115.0	13156204.0	738123459.0	Galle Kara	MaxChang	Medium	1.0	.0	Satisfied	
22	2147483647.0	115.0	1151367.0	14267562.0	Galle Kale	MaxChang	High	1.0	.0	Satisfied	
23	2147483647.0	.0	.0	.0	Galle _SS_	MaxChang	Medium	.6	.4	Satisfied	
24	2147483647.0	116.0	42966225.0	3570333091.0	Galle Kara	MaxChang	Medium	.7	.3	Satisfied	
25	2147483647.0	116.0	13567.0	53598.0	Galle Olcc	MaxChang	Medium	1.0	.0	Satisfied	

Figure 18 Predicted feedback of Users

The original feedback collected from users is in the screenshot below.

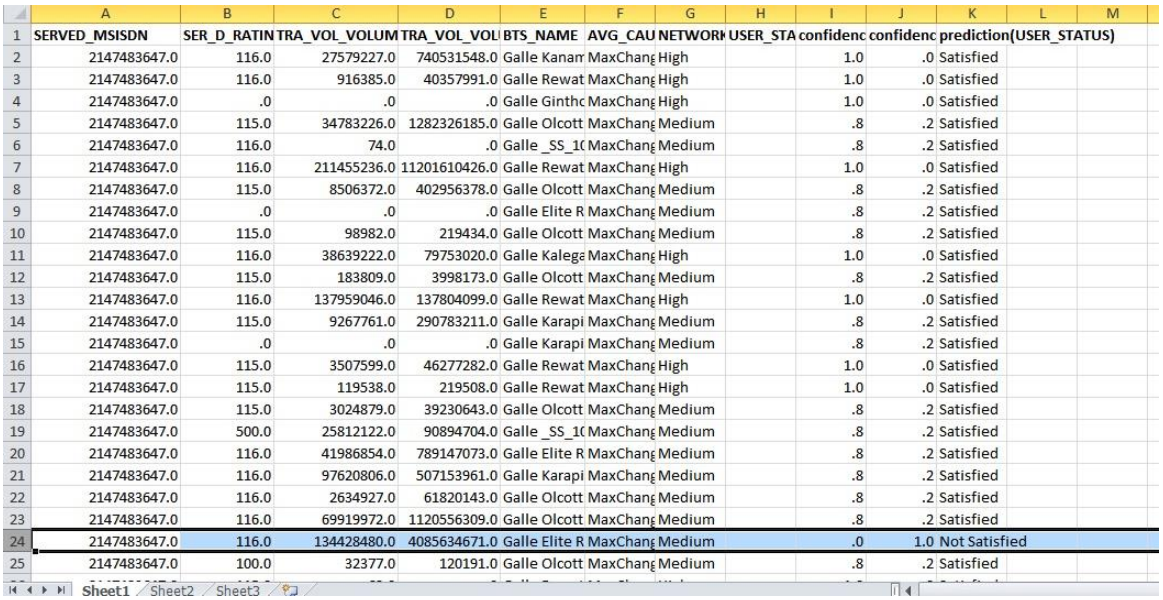
	F	G	H	I	J	L	M	N	O
381481	500	4928	10606	3600	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Low	Satisfied
381482	115	15424265	226072270	28800	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Havy	Satisfied
381483	115	7933516	41121566	32400	Galle Ginthota_SE_10701	MaxChangeCond	High	Low	Satisfied
381484	115	9659867	194870326	32400	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Havy	Satisfied
381485	116	38922181	1858187505	39600	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Havy	Satisfied
381486	115	9907952	428566450	6938	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Havy	Satisfied
381487	116	16428729	256544855	10716	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Havy	Satisfied
381488	115	7909426	178484198	32400	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Havy	Satisfied
381489	116	25752	24112	3600	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Low	Satisfied
381490	115	5854537	214380502	32400	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Havy	Satisfied
381491	0	0	0	25200	Galle Ginthota_SE_10701	MaxChangeCond	High	Low	Satisfied
381492	116	62	0	29	Galle Ginthota_SE_10701	MaxChangeCond	High	Low	Satisfied
381493	116	25672950	1648896551	41872	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Havy	Not Satisfied
381494	115	82752556	2764663631	27458	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Havy	Satisfied
381495	116	186881	526056	899	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Low	Satisfied
381496	116	8420371	54204442	21600	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Low	Satisfied
381497	0	0	0	2	Galle Kanampitiya_SA_10651	MaxChangeCond	High	Low	Satisfied
381498	115	1404	4379	14	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Low	Satisfied
381499	116	6380631	182552413	7200	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Havy	Satisfied
381500	115	13156204	738123459	32400	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Havy	Satisfied
381501	115	1151367	14267562	18300	Galle Kalegana_SM_10657	MaxChangeCond	High	Low	Satisfied
381502	0	0	0	327	Galle _SS_10554	MaxChangeCond	Medium	Low	Satisfied
381503	116	42966225	3570333091	16519	Galle Karapitya_SD_10557	MaxChangeCond	Medium	Havy	Satisfied
381504	116	13567	53598	161	Galle Olcott Mw_SA_10649	MaxChangeCond	Medium	Low	Satisfied
381505	116	62	0	32	Galle Elite Rd_LB_10602	MaxChangeCond	Medium	Low	Satisfied

Formula Bar

Sheet1 Sheet2 Sheet3

Figure 19 Original feedback from users

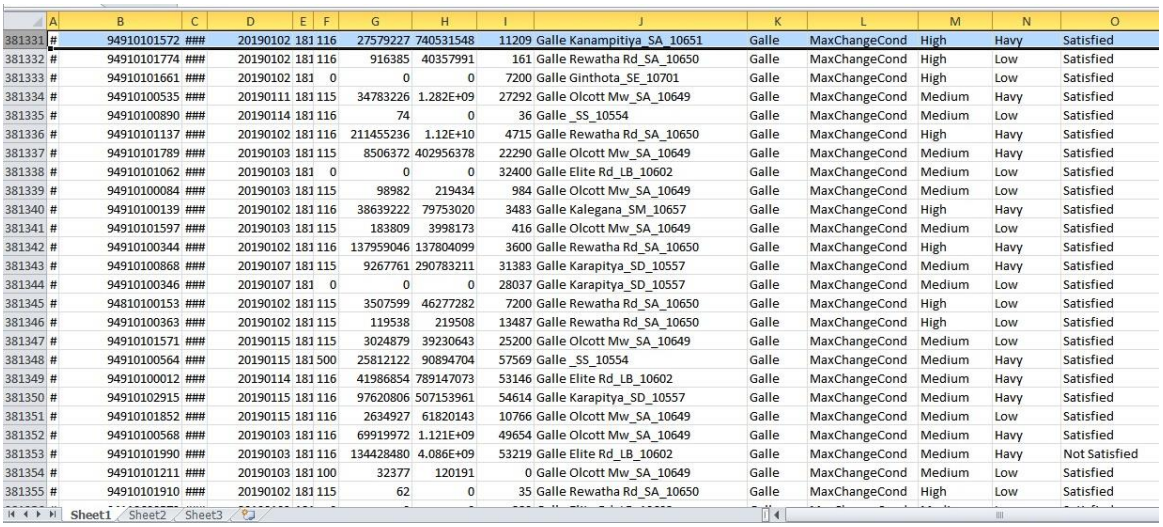
The following mentioned is the predicted result gained from the same Galle set of data using the decision tree algorithm



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	SERVED_MSISDN	SER_D_RATIN	TRA_VOL_VOLUM	TRA_VOL_VOL	BTS_NAME	AVG_CAU	NETWORK	USER_STA	confidenc	confidenc	prediction(USER_STATUS)		
2	2147483647.0	116.0	27579227.0	740531548.0	Galle Kanam	MaxChang	High		1.0	.0	Satisfied		
3	2147483647.0	116.0	916385.0	40357991.0	Galle Rewat	MaxChang	High		1.0	.0	Satisfied		
4	2147483647.0	.0	.0	.0	Galle Ginth	MaxChang	High		1.0	.0	Satisfied		
5	2147483647.0	115.0	34783226.0	1282326185.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
6	2147483647.0	116.0	74.0	.0	Galle _SS_1(	MaxChang	Medium		.8	.2	Satisfied		
7	2147483647.0	116.0	211455236.0	11201610426.0	Galle Rewat	MaxChang	High		1.0	.0	Satisfied		
8	2147483647.0	115.0	8506372.0	402956378.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
9	2147483647.0	.0	.0	.0	Galle Elite R	MaxChang	Medium		.8	.2	Satisfied		
10	2147483647.0	115.0	98982.0	219434.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
11	2147483647.0	116.0	38639222.0	79753020.0	Galle Kalega	MaxChang	High		1.0	.0	Satisfied		
12	2147483647.0	115.0	183809.0	3998173.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
13	2147483647.0	116.0	137959046.0	137804099.0	Galle Rewat	MaxChang	High		1.0	.0	Satisfied		
14	2147483647.0	115.0	9267761.0	290783211.0	Galle Karapi	MaxChang	Medium		.8	.2	Satisfied		
15	2147483647.0	.0	.0	.0	Galle Karapi	MaxChang	Medium		.8	.2	Satisfied		
16	2147483647.0	115.0	3507599.0	46277282.0	Galle Rewat	MaxChang	High		1.0	.0	Satisfied		
17	2147483647.0	115.0	119538.0	219508.0	Galle Rewat	MaxChang	High		1.0	.0	Satisfied		
18	2147483647.0	115.0	3024879.0	39230643.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
19	2147483647.0	500.0	25812122.0	90894704.0	Galle _SS_1(	MaxChang	Medium		.8	.2	Satisfied		
20	2147483647.0	116.0	41986854.0	789147073.0	Galle Elite R	MaxChang	Medium		.8	.2	Satisfied		
21	2147483647.0	116.0	97620806.0	507153961.0	Galle Karapi	MaxChang	Medium		.8	.2	Satisfied		
22	2147483647.0	116.0	2634927.0	61820143.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
23	2147483647.0	116.0	69919972.0	1120556309.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		
24	2147483647.0	116.0	134428480.0	4085634671.0	Galle Elite R	MaxChang	Medium		.0	1.0	Not Satisfied		
25	2147483647.0	100.0	32377.0	120191.0	Galle Olcott	MaxChang	Medium		.8	.2	Satisfied		

Figure 20 Predicted results of Galle Data set

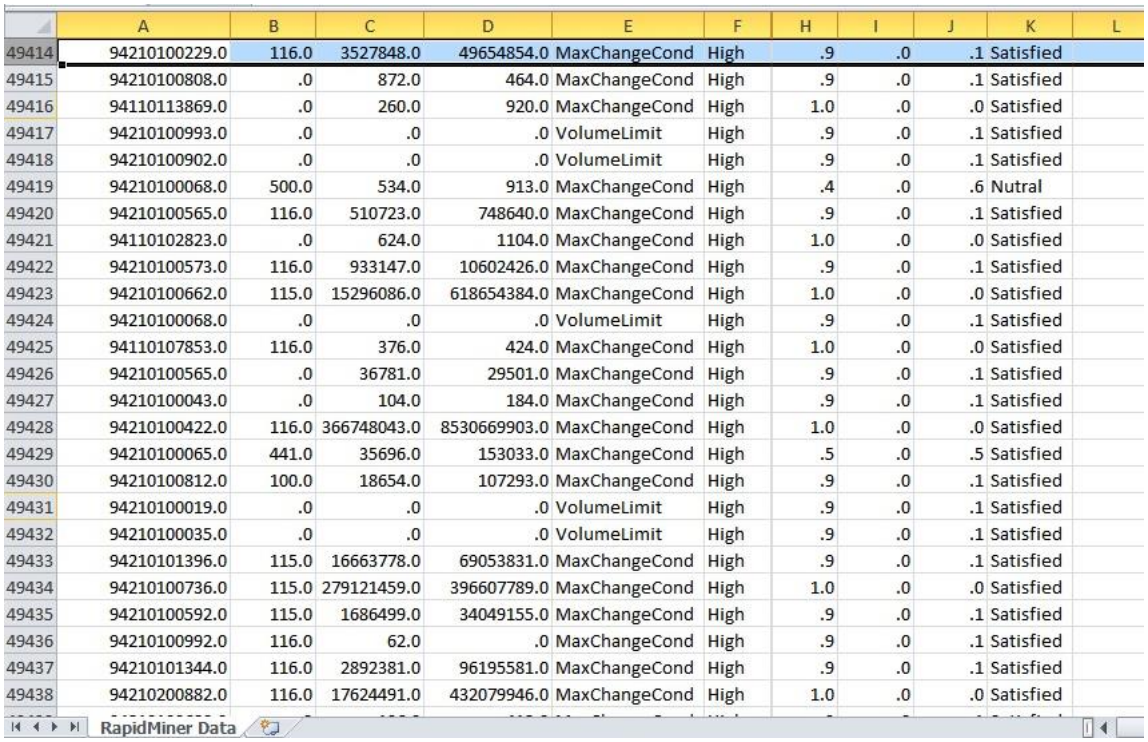
The following mentioned is the actual result of the Galle data set.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
381331	#	94910101572	###	20190102	181	116	27579227	740531548	11209	Galle Kanampitiya_SA_10651	Galle	MaxChangeCond	High	Havy	Satisfied
381332	#	94910101774	###	20190102	181	116	916385	40357991	161	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381333	#	94910101661	###	20190102	181	0	0	0	7200	Galle Ginthota_SE_10701	Galle	MaxChangeCond	High	Low	Satisfied
381334	#	94910100535	###	20190111	181	115	34783226	1.282E+09	27292	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381335	#	94910100890	###	20190114	181	116	74	0	36	Galle _SS_10554	Galle	MaxChangeCond	Medium	Low	Satisfied
381336	#	94910101137	###	20190102	181	116	211455236	1.12E+10	4715	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Havy	Satisfied
381337	#	94910101789	###	20190103	181	115	8506372	402956378	22290	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381338	#	94910101062	###	20190103	181	0	0	0	32400	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Low	Satisfied
381339	#	94910100084	###	20190103	181	115	98982	219434	984	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381340	#	94910100139	###	20190102	181	116	38639222	79753020	3483	Galle Kalegana_SM_10657	Galle	MaxChangeCond	High	Havy	Satisfied
381341	#	94910101597	###	20190103	181	115	183809	3998173	416	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381342	#	94910100344	###	20190102	181	116	137959046	137804099	3600	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Havy	Satisfied
381343	#	94910100868	###	20190107	181	115	9267761	290783211	31383	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Havy	Satisfied
381344	#	94910100346	###	20190107	181	0	0	0	28037	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Low	Satisfied
381345	#	94810100153	###	20190102	181	115	3507599	46277282	7200	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381346	#	94910100363	###	20190102	181	115	119538	219508	13487	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381347	#	94910101571	###	20190115	181	115	3024879	39230643	25200	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381348	#	94910100564	###	20190115	181	500	25812122	90894704	57569	Galle _SS_10554	Galle	MaxChangeCond	Medium	Havy	Satisfied
381349	#	94910100012	###	20190114	181	116	41986854	789147073	53146	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Havy	Satisfied
381350	#	94910102915	###	20190115	181	116	97620806	507153961	54614	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Havy	Satisfied
381351	#	94910101852	###	20190115	181	116	2634927	61820143	10766	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381352	#	94910100568	###	20190103	181	116	69919972	1.121E+09	49654	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381353	#	94910101990	###	20190103	181	116	134428480	4.086E+09	53219	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Havy	Not Satisfied
381354	#	94910101211	###	20190103	181	100	32377	120191	0	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381355	#	94910101910	###	20190102	181	115	62	0	35	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied

Figure 21 Actual results of Galle Data set

In here Researcher State the result comparison, when the data set from Aluthgama region went through the solution. First is the predicted data set.

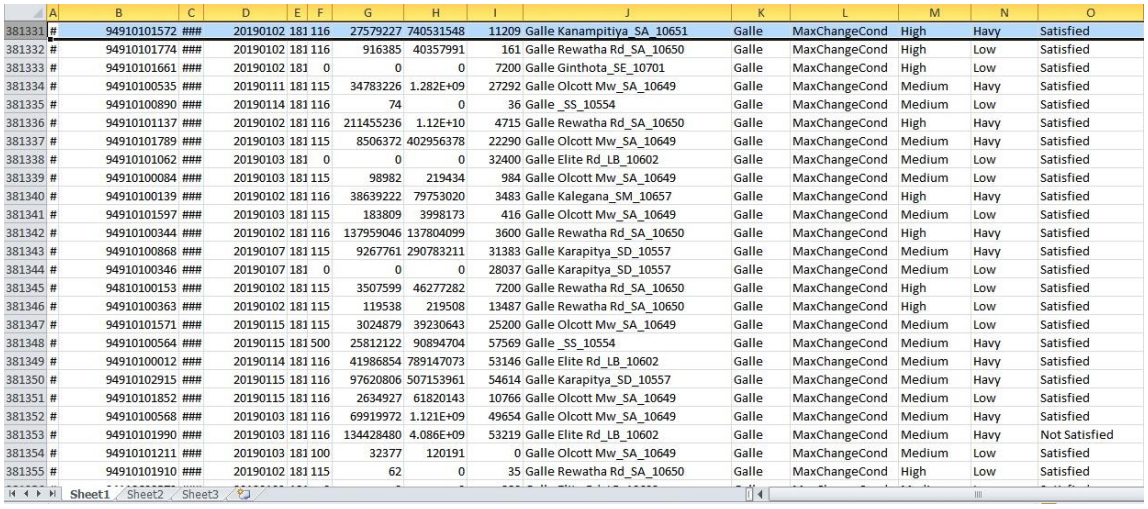


	A	B	C	D	E	F	H	I	J	K	L
49414	94210100229.0	116.0	3527848.0	49654854.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49415	94210100808.0	.0	872.0	464.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49416	94110113869.0	.0	260.0	920.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49417	94210100993.0	.0	.0	.0	VolumeLimit	High	.9	.0	.1	Satisfied	
49418	94210100902.0	.0	.0	.0	VolumeLimit	High	.9	.0	.1	Satisfied	
49419	94210100068.0	500.0	534.0	913.0	MaxChangeCond	High	.4	.0	.6	Nutral	
49420	94210100565.0	116.0	510723.0	748640.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49421	94110102823.0	.0	624.0	1104.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49422	94210100573.0	116.0	933147.0	10602426.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49423	94210100662.0	115.0	15296086.0	618654384.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49424	94210100068.0	.0	.0	.0	VolumeLimit	High	.9	.0	.1	Satisfied	
49425	94110107853.0	116.0	376.0	424.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49426	94210100565.0	.0	36781.0	29501.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49427	94210100043.0	.0	104.0	184.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49428	94210100422.0	116.0	366748043.0	8530669903.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49429	94210100065.0	441.0	35696.0	153033.0	MaxChangeCond	High	.5	.0	.5	Satisfied	
49430	94210100812.0	100.0	18654.0	107293.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49431	94210100019.0	.0	.0	.0	VolumeLimit	High	.9	.0	.1	Satisfied	
49432	94210100035.0	.0	.0	.0	VolumeLimit	High	.9	.0	.1	Satisfied	
49433	94210101396.0	115.0	16663778.0	69053831.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49434	94210100736.0	115.0	279121459.0	396607789.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	
49435	94210100592.0	115.0	1686499.0	34049155.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49436	94210100992.0	116.0	62.0	.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49437	94210101344.0	116.0	2892381.0	96195581.0	MaxChangeCond	High	.9	.0	.1	Satisfied	
49438	94210200882.0	116.0	17624491.0	432079946.0	MaxChangeCond	High	1.0	.0	.0	Satisfied	

RapidMiner Data

Figure 22 Predicted data set - Alutgama

Here is the Actual user feedback for the January month.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
381331	#	94910101572	###	20190102	181	116	27579227	740531548	11209	Galle Kanampitiya_SA_10651	Galle	MaxChangeCond	High	Havy	Satisfied
381332	#	94910101774	###	20190102	181	116	916385	40357991	161	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381333	#	94910101661	###	20190102	181	0	0	0	7200	Galle Ginthota_SE_10701	Galle	MaxChangeCond	High	Low	Satisfied
381334	#	94910100535	###	20190111	181	115	34783226	1.282E+09	27292	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381335	#	94910100890	###	20190114	181	116	74	0	36	Galle _SS_10554	Galle	MaxChangeCond	Medium	Low	Satisfied
381336	#	94910101137	###	20190102	181	116	211455236	1.12E+10	4715	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Havy	Satisfied
381337	#	94910101789	###	20190103	181	115	8506372	402956378	22290	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381338	#	94910101062	###	20190103	181	0	0	0	32400	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Low	Satisfied
381339	#	94910100084	###	20190103	181	115	98982	219434	984	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381340	#	94910100139	###	20190102	181	116	38639222	79753020	3483	Galle Kalegana_SM_10657	Galle	MaxChangeCond	High	Havy	Satisfied
381341	#	94910101597	###	20190103	181	115	183809	3998173	416	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381342	#	94910100344	###	20190102	181	116	137959046	137804099	3600	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Havy	Satisfied
381343	#	94910100868	###	20190107	181	115	9267761	290783211	31383	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Havy	Satisfied
381344	#	94910100346	###	20190107	181	0	0	0	28037	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Low	Satisfied
381345	#	94810100153	###	20190102	181	115	3507599	46277282	7200	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381346	#	94910100363	###	20190102	181	115	119538	219508	13487	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied
381347	#	94910101571	###	20190115	181	115	3024879	39230643	25200	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381348	#	94910100564	###	20190115	181	500	25812122	90894704	57569	Galle _SS_10554	Galle	MaxChangeCond	Medium	Havy	Satisfied
381349	#	94910100012	###	20190114	181	116	41986854	789147073	53146	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Havy	Satisfied
381350	#	94910102915	###	20190115	181	116	97620806	507153961	54614	Galle Karapitya_SD_10557	Galle	MaxChangeCond	Medium	Havy	Satisfied
381351	#	94910101852	###	20190115	181	116	2634927	61820143	10766	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381352	#	94910100568	###	20190103	181	116	69919972	1.121E+09	49654	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Havy	Satisfied
381353	#	94910101990	###	20190103	181	116	134428480	4.086E+09	53219	Galle Elite Rd_LB_10602	Galle	MaxChangeCond	Medium	Havy	Not Satisfied
381354	#	94910101211	###	20190103	181	100	32377	120191	0	Galle Olcott Mw_SA_10649	Galle	MaxChangeCond	Medium	Low	Satisfied
381355	#	94910101910	###	20190102	181	115	62	0	35	Galle Rewatha Rd_SA_10650	Galle	MaxChangeCond	High	Low	Satisfied

Sheet1 Sheet2 Sheet3

Figure 23 Actual feedback for January (Result)

7.4 Summery

Within the scope of this chapter, the researcher evaluated the methodology used in conducting this research along with the results, he obtained. The implementation process was assessed to ensure the reliability of the system.

Chapter 08

Conclusion and further work

8.1 Introduction

According to the literature, reviewed in this paper there are no any direct researches conduct on using customer internet usage data to analyze the speed decrements on internet usage though there are many researches conduct to find the solutions for having good steady internet speed. while the developed countries which have archived greater speeds in the context of internet, countries like us not even focused to maintain the speed which ISP s promised to give in the first-place.

While there are many reasons to not having a proper internet speed in Sri Lanka, limited bandwidth given to large capacity crowed and limited nodes installation for large geographical areas takes a higher rank. This resulting large number of people using internet by connecting to the same node while that cost in decreased internet or buffered internet lines.

The best way to address this situation is finding a way to maintain a steady internet speed by any means like increasing the bandwidth which the isp s are providing, increasing data transfer ratios or increasing the EnodB antennas which locates in a given are because in a time the whole world is talking about data sharing, globalizing, virtual reality, IaaS future is depending on the internet Sri Lanka does not want to miss out.

In the current context, it was exposed via this research that user satisfaction positivity depends on following infrastructure factors.

- Network QOS
- Record Closing Factor
- Bandwidth Allocation

The following was discovered from this research,

- Out of the machine learning techniques such as classification, clustering and association; classification is the best fit for this particular scenario.

- Out of the classifier algorithms, Naïve Bayes is the ideal algorithm for this data bundle.
- Telecommunication companies can make sure the customer satisfaction out of which they could be guarantee the sustainability as a benefit of this system.

8.2 Limitation

As mentioned this is a beneficial tool for the telecommunication companies. But however there could be scenarios where it can be deviated from what is suspected from this given system and where as it will not be able to capture the absolute cause for the matter.

For an instant, Network QOS and Record closing factors can be subjected to unavoidable affects like when the copper cables get stricken externally.

Also, due to the changes in the atmosphere between the network towers and client's receiving end, signal strength can be weakening.

Even the weather can have a considerable impact on the predicted outcome since this research's data depend on the wireless data communications.

8.3 Further Development

It should be noted that within the scope of this research was tighten among the direct factors that can influence the decrease in promised bandwidth. But there are several other factors which contribute this problem in a hidden manner such as usage patterns, seasonal changes and etc. In terms of further development, these factors can be considered, when making predictions in the variations of the promised bandwidth.

For a keener and deeper attempt to smoothen up the expected bandwidth, telecommunication companies can consult the complaints with reference to internet speed that customer care unit gets and inspect for common ground for those complaints and cluster them and include them to the data mining and make the system more effective as well.

8.4 Summery

This chapter is to conclude the whole thesis for a bird eye view of Decision support system for bandwidth expansion according to the increment of users. Whereas it brought forth the vital reason behind conducting this thesis along with the limitations the researcher observed and the recommendations in terms of further development for the deployment process offered by the researcher.

Bibliography

Anastasios N. Bikos, N. S. (2016). *Architecture Design of an Area Efficient High Speed Crypto Processor for 4G LTE* .

Badari N. Thyamagondlu, V. W. (2013). *A Bandwidth-Conscious Caching Scheme for Mobile Devices.*

Dan Huang, C. C. (n.d.). *Bandwidth Allocation Algorithms in the Short Video Sharing System.*

Guowei Huang, L. K. (2017). A Bandwidth Allocation Policy for Helpers in Cloud-assisted P2P Video-on-demand Systems. *Fifth International Conference on Advanced Cloud and Big Dat*, 12.

Ir. Tommy van der Vorst, I. i. (2014). *How the speed of the internet will develop between now and 2020.*

KARLSSON, S. G. (2017). *Pushing the Wireless Data Rate to the Internet Speed.*

Marshini Chetty, D. H. (2011). *Why Is My Internet Slow?: Making Network Speeds Visible.*

Massimo Barazzetta, D. M. (2016). *Optimization of 4G Wireless Access Network Features by Using Reverberation Chambers: Application to High-Speed Train LTE Users.* London, UK.

Mete Uzun, O. A. (2016). *End-to-end Internet Speed Analysis of Mobile Networks with Map Reduce.*

S.Sharad. (2014). *INCREASING INTERNET SPEED AND BANDWIDTH BY USING LAWS OF PHYSICS.*

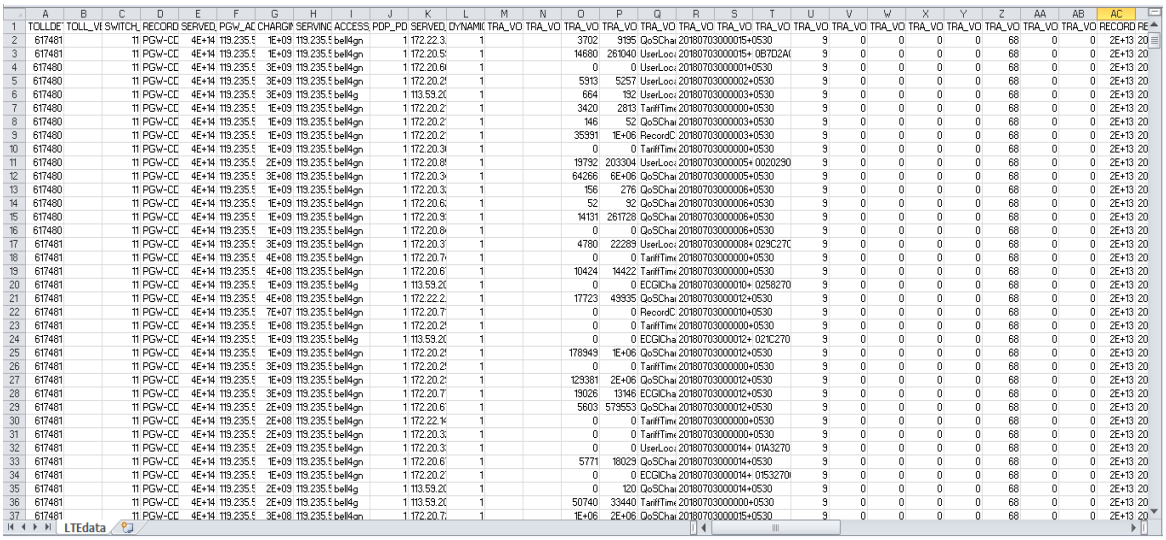
Sharad, S. (2014). Increasing Internet Speed and Bandwidth by Using Laws of Physics. *2014 International Conference on Intelligent Computing Applications*, Coimbatore.

The Cisco Prime Home solution. (2012). *Bandwidth Consumption and Broadband Reliability.*

Y. Ito, H. Koga and K. Iida. (2017). *A bandwidth allocation scheme to meet flow requirements in mobile edge computing.* Prague.

Appendixes - A

Raw data sample



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	TOLLDE	TOLL_VE	SWITCH_	RECORD	SERVED_	PGW_AC	CHARGI	SERVING	ACCESS_	PDP_PD	SERVED_	DYNAMI	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	TRA_VO	RECORD_RE
2	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.22.3	1			3702	9135	QoSChai	20180703000015+0530		9	0	0	0	0	68	0	0	2E+13	20
3	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.5	1			14680	261040	UserLoci	20180703000015+	0B7D2A	9	0	0	0	0	68	0	0	2E+13	20
4	617480		11	PGW-CD	4E+14	119.235.5	3E+09	119.235.5	bell4gn	1	172.20.6	1			0	0	UserLoci	20180703000001+0530		9	0	0	0	0	68	0	0	2E+13	20
5	617480		11	PGW-CD	4E+14	119.235.5	3E+09	119.235.5	bell4gn	1	172.20.2	1			5313	5257	UserLoci	20180703000002+0530		9	0	0	0	0	68	0	0	2E+13	20
6	617480		11	PGW-CD	4E+14	119.235.5	3E+09	119.235.5	bell4g	1	113.59.20	1			664	132	UserLoci	20180703000003+0530		9	0	0	0	0	68	0	0	2E+13	20
7	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			3420	2813	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
8	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			146	52	QoSChai	20180703000003+0530		9	0	0	0	0	68	0	0	2E+13	20
9	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			35991	1E+06	RecordC	20180703000003+0530		9	0	0	0	0	68	0	0	2E+13	20
10	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.3	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
11	617480		11	PGW-CD	2E+14	119.235.5	2E+09	119.235.5	bell4gn	1	172.20.8	1			19792	203304	UserLoci	20180703000005+	0020290	9	0	0	0	0	68	0	0	2E+13	20
12	617480		11	PGW-CD	4E+14	119.235.5	3E+08	119.235.5	bell4gn	1	172.20.3	1			64266	6E+06	QoSChai	20180703000005+0530		9	0	0	0	0	68	0	0	2E+13	20
13	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.3	1			156	276	QoSChai	20180703000006+0530		9	0	0	0	0	68	0	0	2E+13	20
14	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.6	1			52	92	QoSChai	20180703000006+0530		9	0	0	0	0	68	0	0	2E+13	20
15	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.3	1			14131	261728	QoSChai	20180703000006+0530		9	0	0	0	0	68	0	0	2E+13	20
16	617480		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.8	1			0	0	QoSChai	20180703000006+0530		9	0	0	0	0	68	0	0	2E+13	20
17	617481		11	PGW-CD	3E+14	119.235.5	3E+09	119.235.5	bell4gn	1	172.20.3	1			4780	22289	UserLoci	20180703000008+	029C270	9	0	0	0	0	68	0	0	2E+13	20
18	617481		11	PGW-CD	4E+14	119.235.5	4E+08	119.235.5	bell4gn	1	172.20.7	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
19	617481		11	PGW-CD	4E+14	119.235.5	4E+08	119.235.5	bell4gn	1	172.20.6	1			10424	14422	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
20	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4g	1	113.59.20	1			0	0	ECGIChai	20180703000010+	0258270	9	0	0	0	0	68	0	0	2E+13	20
21	617481		11	PGW-CD	4E+14	119.235.5	4E+08	119.235.5	bell4gn	1	172.22.2	1			17723	49335	QoSChai	20180703000012+0530		9	0	0	0	0	68	0	0	2E+13	20
22	617481		11	PGW-CD	4E+14	119.235.5	7E+07	119.235.5	bell4gn	1	172.20.7	1			0	0	RecordC	20180703000010+0530		9	0	0	0	0	68	0	0	2E+13	20
23	617481		11	PGW-CD	4E+14	119.235.5	1E+08	119.235.5	bell4gn	1	172.20.2	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
24	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4g	1	113.59.20	1			0	0	ECGIChai	20180703000012+	021C270	9	0	0	0	0	68	0	0	2E+13	20
25	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			178949	1E+06	QoSChai	20180703000012+0530		9	0	0	0	0	68	0	0	2E+13	20
26	617481		11	PGW-CD	4E+14	119.235.5	3E+09	119.235.5	bell4gn	1	172.20.2	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
27	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			129381	2E+06	QoSChai	20180703000012+0530		9	0	0	0	0	68	0	0	2E+13	20
28	617481		11	PGW-CD	4E+14	119.235.5	3E+09	119.235.5	bell4gn	1	172.20.7	1			19026	13146	ECGIChai	20180703000012+0530		9	0	0	0	0	68	0	0	2E+13	20
29	617481		11	PGW-CD	4E+14	119.235.5	2E+09	119.235.5	bell4gn	1	172.20.6	1			5603	579553	QoSChai	20180703000012+0530		9	0	0	0	0	68	0	0	2E+13	20
30	617481		11	PGW-CD	4E+14	119.235.5	2E+08	119.235.5	bell4gn	1	172.22.14	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
31	617481		11	PGW-CD	4E+14	119.235.5	2E+09	119.235.5	bell4gn	1	172.20.3	1			0	0	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
32	617481		11	PGW-CD	4E+14	119.235.5	2E+09	119.235.5	bell4gn	1	172.20.3	1			0	0	UserLoci	20180703000014+	01A3270	9	0	0	0	0	68	0	0	2E+13	20
33	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.6	1			5771	18029	QoSChai	20180703000014+0530		9	0	0	0	0	68	0	0	2E+13	20
34	617481		11	PGW-CD	4E+14	119.235.5	1E+09	119.235.5	bell4gn	1	172.20.2	1			0	0	ECGIChai	20180703000014+	0153270	9	0	0	0	0	68	0	0	2E+13	20
35	617481		11	PGW-CD	4E+14	119.235.5	2E+09	119.235.5	bell4g	1	113.59.20	1			0	120	QoSChai	20180703000014+0530		9	0	0	0	0	68	0	0	2E+13	20
36	617481		11	PGW-CD	4E+14	119.235.5	2E+09	119.235.5	bell4g	1	113.59.20	1			50740	33440	TariffTime	20180703000000+0530		9	0	0	0	0	68	0	0	2E+13	20
37	617481		11	PGW-CD	4E+14	119.235.5	3E+08	119.235.5	bell4gn	1	172.20.7	1			1E+06	2E+06	QoSChai	20180703000015+0530		9	0	0	0	0	68	0	0	2E+13	20

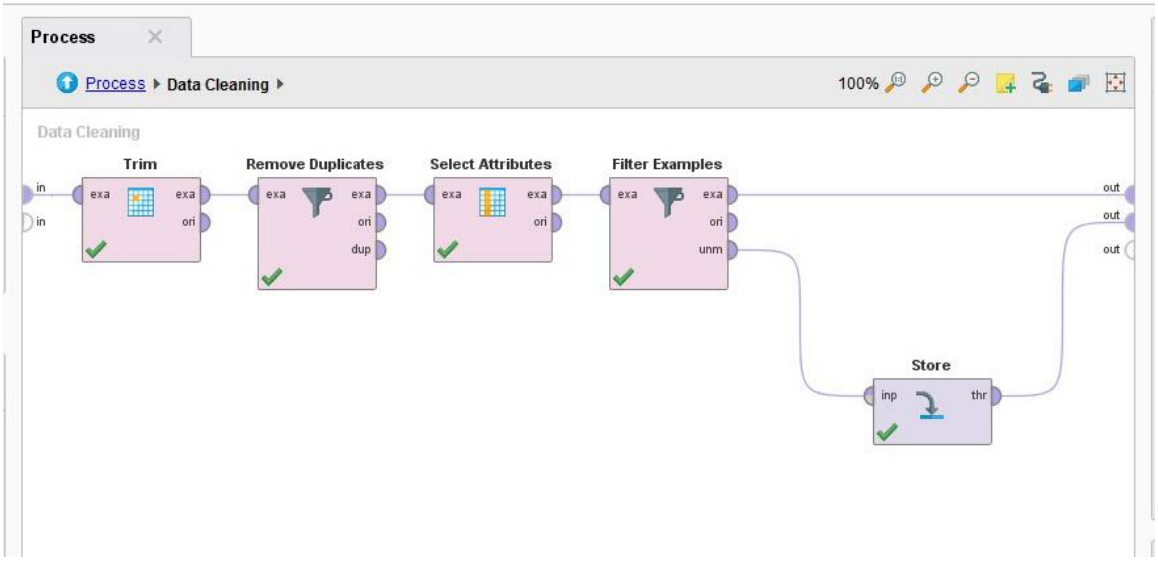
LTEdata

Raw data status

	A	B	C	D
1		Type	Missing	Statistic
2	TOLLDET_REF_NUM	Integer	0	617589.291
3	TOLL_VER_NUM	Polynominal	1050205	
4	SWITCH_CODE	Integer	0	11
5	RECORD_TYPE	Polynominal	0	PGW-CDR (1050205)
6	SERVED_IMSI	Real	0	4.13E+14
7	PGW_ADDRESS	Polynominal	0	119.235.5.36 (428312), 119.235.5.38 (326771), ... [1 more]
8	CHARGING_ID	Real	0	1.78E+09
9	SERVING_NODE_ADDRESS	Polynominal	0	119.235.5.36 (428312), 119.235.5.38 (326771), ... [1 more]
10	ACCESS_POINT_NAME_NI	Polynominal	0	bell4gn (947327), bell4g (99787), ... [7 more]
11	PDP_PDN_TYPE	Integer	0	1
12	SERVED_PDP_PDN_ADDRESS	Polynominal	0	172.22.5.220 (20911), 172.20.85.174 (15082), ... [22846 more]
13	DYNAMIC_ADDRESS_FLAG	Integer	0	0.994441085
14	TRA_VOL_QOS_REQUESTED	Polynominal	1050205	
15	TRA_VOL_QOS_NEGOTIATED	Polynominal	1050205	
16	TRA_VOL_VOLUME_UP_LINK	Integer	0	1601209.211
17	TRA_VOL_VOLUME_DOWN_LINK	Integer	0	1.58E+07
18	TRA_VOL_CHANGE_CONDITION	Polynominal	0	ECGIChange (420188), UserLocationChange (329587), ... [3 more]
19	TRA_VOL_CHANGE_TIME	Polynominal	0	20180703010000+0530 (25092), 20180703080000+0530 (20493), ... [46876 more]
20	TRA_VOL_FAILURE_HANDLING_CO	Polynominal	1050205	
21	TRA_VOL_USER_LOCATION_INFO	Polynominal	302467	02C12A0040F314EC0340F31418 (23411), 0020290040F314ED0340F31418 (23152), ... [1372 more]
22	TRA_VOL_E_QCI	Integer	0	8.980221004
23	TRA_VOL_E_MAX_REQ_BW_UL	Integer	0	118.3210307
24	TRA_VOL_E_MAX_REQ_BW_DL	Integer	0	118.3210307
25	TRA_VOL_E_GUARANT_BIT_RATE	Integer	0	80.61105975
26	TRA_VOL_E_GUARANT_BIT_RATE_	Integer	0	80.61105975
27	TRA_VOL_E_ARP	Integer	0	67.90749996
28	TRA_VOL_E_APN_AGG_MAX_BITF	Integer	0	0
29	TRA_VOL_E_APN_AGG_MAX_BITF	Integer	0	0
30	RECORD_OPENING_TIME	Real	0	2.02E+13
31	RECORD_OPENING_TIME_FULL	Polynominal	0	20180703010000+0530 (25091), 20180703000000+0530 (21890), ... [46874 more]
32	UNIX_TIME_STAMP	Integer	0	1.53E+09
33	DURATION	Integer	0	873.6314767
34	CAUSE_FOR_REC_CLOSING	Polynominal	0	MaxChangeCond (776642), TimeLimit (211290), ... [4 more]
35	DIAGNOSTICS	Polynominal	1017604	ManufacturerSpecificCause:1 (30248), ManufacturerSpecificCause:96 (1009), ... [5 more]
36	RECORD_SEQUENCE_NUMBER	Integer	0	4936.404173
37	NODE_ID	Polynominal	0	isp.com (1050205)

Sheet1 Sheet2 Sheet3

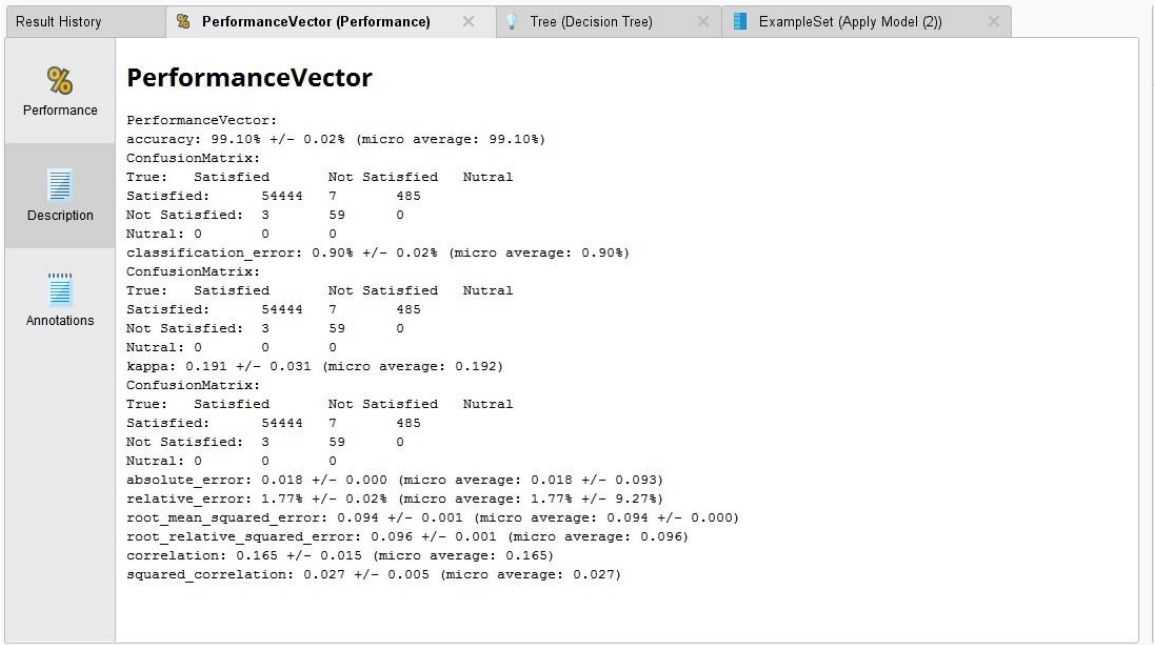
Data Preprocessing Process



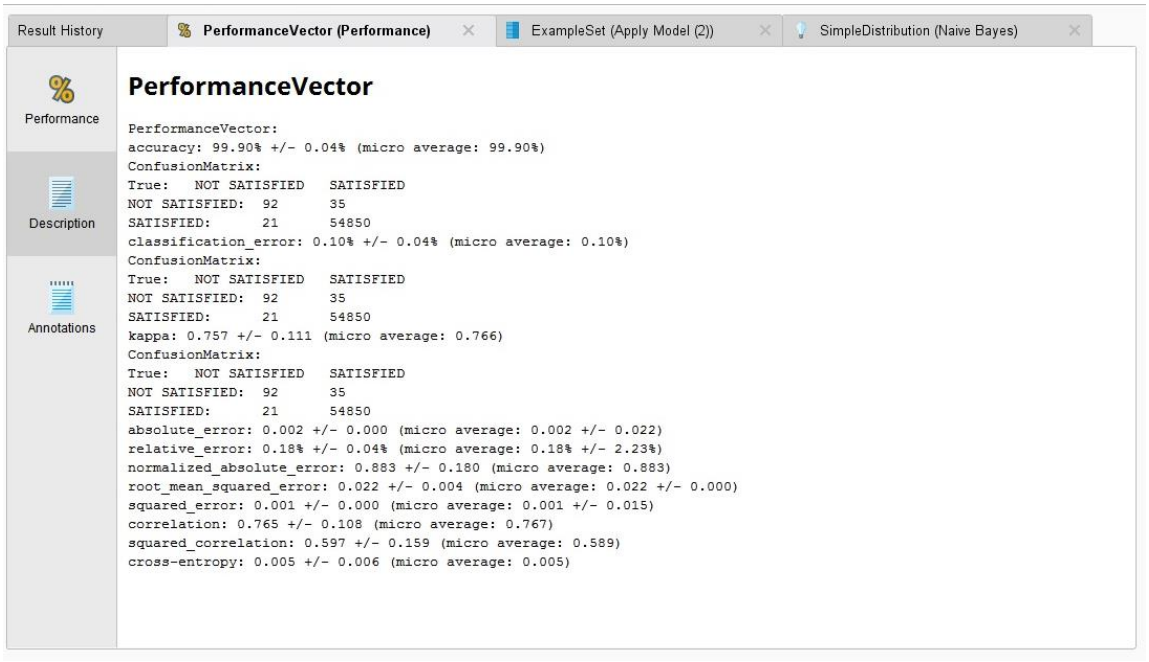
Appendixes - B

Model Evaluation summery.

Decision tree Summery

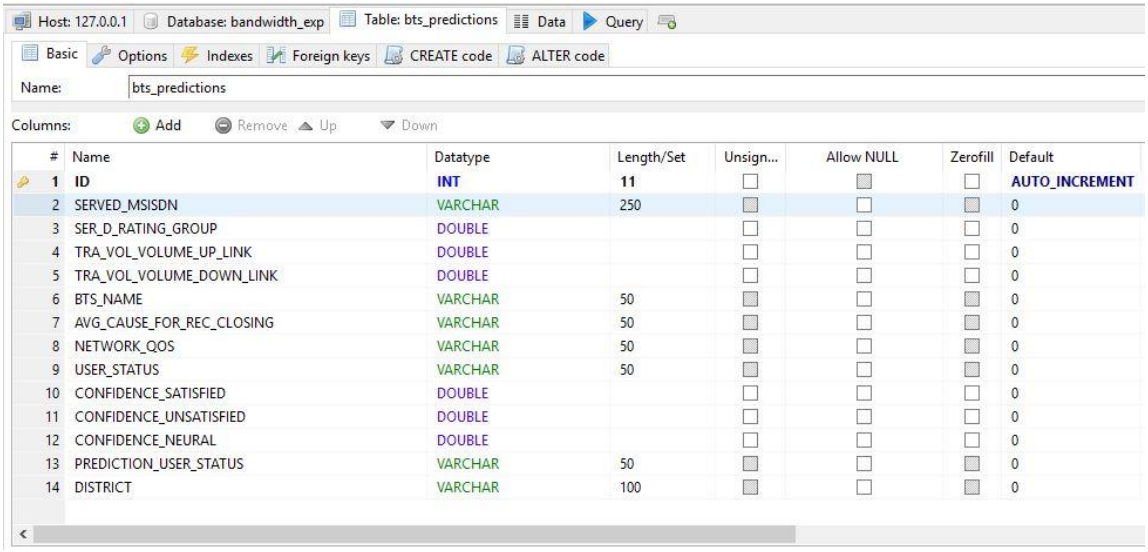


Naïve Base summery



Appendixes - C

Data Table 01



#	Name	Datatype	Length/Set	Unsign...	Allow NULL	Zerofill	Default
1	ID	INT	11	☐	☐	☐	AUTO_INCREMENT
2	SERVED_MSISDN	VARCHAR	250	☐	☐	☐	0
3	SER_D_RATING_GROUP	DOUBLE		☐	☐	☐	0
4	TRA_VOL_VOLUME_UP_LINK	DOUBLE		☐	☐	☐	0
5	TRA_VOL_VOLUME_DOWN_LINK	DOUBLE		☐	☐	☐	0
6	BTS_NAME	VARCHAR	50	☐	☐	☐	0
7	AVG_CAUSE_FOR_REC_CLOSING	VARCHAR	50	☐	☐	☐	0
8	NETWORK_QOS	VARCHAR	50	☐	☐	☐	0
9	USER_STATUS	VARCHAR	50	☐	☐	☐	0
10	CONFIDENCE_SATISFIED	DOUBLE		☐	☐	☐	0
11	CONFIDENCE_UNSATISFIED	DOUBLE		☐	☐	☐	0
12	CONFIDENCE_NEURAL	DOUBLE		☐	☐	☐	0
13	PREDICTION_USER_STATUS	VARCHAR	50	☐	☐	☐	0
14	DISTRICT	VARCHAR	100	☐	☐	☐	0

Data Table 02

Basic | Options | Indexes | Foreign keys | CREATE code | ALTER code

Name: region_lte

Columns: Add | Remove | Up | Down

#	Name	Datatype	Length/Set	Unsign...	Allow NULL	Zerofill	Default
1	ID	INT	11	☐	☐	☐	AUTO_INCREMENT
2	BTS_NAME	VARCHAR	255	☐	☑	☐	NULL
3	DURATION	INT	11	☐	☑	☐	NULL
4	NUMX	INT	11	☐	☑	☐	NULL
5	REC_DATE	INT	11	☐	☑	☐	NULL
6	DISTRICT	VARCHAR	255	☐	☑	☐	NULL
7	SER_D_RATING_GROUP	VARCHAR	255	☐	☑	☐	NULL
8	SERVED_IMSI	DECIMAL	19,2	☐	☑	☐	NULL
9	SERVED_MSISDN	DECIMAL	19,2	☐	☑	☐	NULL
10	TRA_VOL_VOLUME_DOWN_LINK	DECIMAL	19,2	☐	☑	☐	NULL
11	TRA_VOL_VOLUME_UP_LINK	DECIMAL	19,2	☐	☑	☐	NULL
12	USER_LOCATION_INFORMATION	VARCHAR	255	☐	☑	☐	NULL

Appendixes - D

Code Snippet

```
package lk.hashan.bandwidthexpdss.dao;

import java.math.BigInteger;

@Repository
public class RegionDao{

    @Autowired
    SessionFactory sessionFactory;

    public Integer getAllUserCount(Integer lowerDate, Integer upperDate){
        Session  session = sessionFactory.openSession();
        session.beginTransaction();
        List result = session.
                createQuery("select distinct r.servedMsisdn from RegionLte r where r.recDate >= :lowerDate and r.recDate <= :upperDate")
                .setInteger("lowerDate", lowerDate)
                .setInteger("upperDate", upperDate)
                .list();
        if(result != null && result.size() > 0){
            return result.size();
        }else{
            return 0;
        }

    }
```

```
public Integer getResidenceUserCount(Integer lowerDate, Integer upperDate){
    Session  session = sessionFactory.openSession();
    session.beginTransaction();
    Integer i = 0;
    @SuppressWarnings("unchecked")
    List<Long> result = session.
            createQuery("select count(l.id) from RegionLte l where l.recDate >= :lowerDate and l.recDate <= :upperDate group by l.servedMsisdn ")
            .setInteger("lowerDate", lowerDate)
            .setInteger("upperDate", upperDate)
            .list();

    for (Long count : result) {
        System.out.println(count);
        if(count > 20){
            i++;
        }
    }

    return i;

}
```

```
public Integer getHeavyUserCount(Integer lowerDate, Integer upperDate){
    Session  session = sessionFactory.openSession();
    session.beginTransaction();
    @SuppressWarnings("unchecked")
    List<BigInteger> result = session.
            createQuery("select SUM(r.volume) from RegionLte r where r.recDate >= :lowerDate and r.recDate <= :upperDate"
                    + "  group by r.servedMsisdn "
                    + " ")
            .setInteger("lowerDate", lowerDate)
            .setInteger("upperDate", upperDate)
            .list();
    Integer i = 0;
    for (BigInteger object : result) {
        System.out.println(object);

        if(object.compareTo(BigInteger.valueOf(Long.valueOf("64424509440"))) == 1){
            i++;
        }

    }
    return i;
}
```