ADDIS ABABA UNIVERSITY
ADDIS ABABA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Quality of Experience Modeling for Fixed Broadband Internet Using Machine Learning Algorithms

By

Abayneh Mekonnen

Advisor

Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Electrical and Computer Engineering, Addis Ababa University in Partial Fulfillment of the Requirement for the Degree of Master's of Science in Communication Engineering

March 07, 2024

Adis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering
Communication Engineering Graduate Program

Quality of Experience Modeling for Fixed Broadband Internet Using Machine Learning Algorithm

by:

Abayneh Mekonnen

Approval by Board of Examiners	Signature	Date
Chairman, School Graduate Committee:	________	________
Advisor's Name:	________	________
Internal Examiner's Name:	________	________
External Examiner's Name:	________	________

Declaration

I declare that this thesis is entirely my own work and does not include any content from other educational institutions without proper acknowledgement. To the best of my knowledge, it does not contain previously published material by another person without recognition.

Name of the Student	Signature	Date
________________	____________	________

This thesis has been submitted for examination with my approval as a university advisor

Name of the Advisor	Signature	Date
________________	____________	________

Addia Ababa, Ethiopia

Abstract

As the demand for dependable fixed broadband internet services continues to grow, ensuring an excellent Quality of Experience (QoE) for end-users is essential. This thesis centers on QoE modeling, employing advanced machine learning techniques, specifically Support Vector Machine (SVM) and Random Forest algorithms. The study utilizes subjective assessments and Quality of Service (QoS) metrics, including latency, upload speed, download speed, uptime, packet loss, and jitter, to comprehensively comprehend and model the factors influencing user satisfaction.

The research incorporates an exhaustive feature selector to extract pertinent features from the dataset, enhancing the precision of the models. Hyperparameter optimization is carried out through a Grid Search approach to fine-tune the models for optimal performance. To assess the models, a robust cross-validation methodology is implemented.

The results indicate that SVM surpasses Random Forest in QoE modeling for Virtual Internet Service Providers (vISPs) like Websprix and ZERGAW Cloud with average accuracy score of 92% and 70% respectively. Conversely, Random Forest proves to be the more suitable model for predicting QoE in the case of the national ISP, ethio telecom with average accuracy value of 88%. This comparative performance analysis offers valuable insights into the distinct strengths of each model for different service providers.

The research findings also indicate that employing both subjective and QoS metrics in combination to model the user QoE yields superior model performance and predictive outcomes compared to relying solely on subjective assessments and QoS metrics.

These findings contribute to the ongoing discussion on QoE enhancement in fixed broadband internet services, providing practical recommendations for service providers based on observed model performances. The application of machine learning, feature selection, and hyperparameter optimization techniques underscores the importance of these methodologies in customizing QoE models to specific service contexts, ultimately enhancing user satisfaction in diverse fixed broadband Internet environments.

Keywords: QoE, QoS, Support Vector Machine, Random Forest, vISP, Fixed Broadband Internet, Machine Learning Algorithm.

Acknowledgments

Foremost, I extend my heartfelt appreciation to the Almighty for providing me with this opportunity and endowing me with the capability to navigate through this journey successfully. Secondly, I wish to convey my profound gratitude to my advisor, Dr. -Ing. Dereje Hailemariam, whose invaluable contributions in the form of stimulating suggestions and unwavering encouragement played a pivotal role in the organization of my thesis. Without his steadfast support and guidance, the completion of this thesis would not have been conceivable.

I am also grateful to extend my thanks to my families and friends, whose inspiration, support, and encouragement served as a driving force in reaching the culmination of this academic endeavor. Their belief in my abilities has been a constant source of motivation.

Furthermore, I express my sincere appreciation to ethio telecom and the vISP for graciously providing the essential data integral to my study. Their cooperation and willingness to facilitate access to valuable information significantly contributed to the depth and quality of my research.

In summary, I acknowledge and thank all those who have played a role, both seen and unseen, in making this academic achievement possible.

Contents

List of Tables

List of Figures

.

List of Acronyms

2G Second Generation

3G Third Generation

4G Fourth Generation

5G Fifth Generation

ADSL Asymmetric Digital Subscriber Line

AI Artificial Intelligence

CPE Customer Premise Equipment

DSL Digital Subscriber Line

FBB Fixed Broadband

FBBI Fixed Broadband Internet

FBBN Fixed Broadband Network

FCC The Federal Communications Commission

FTTH Fiber to the Home

FTTN Fiber to the Node

GR Gender

ISS Internet Service Satisfaction

IFs Influencing Factors

LU Location Used

MOS Mean Opinion Score

ML Machine Learning

NAP's Network Access Points

OTN's Optical Transport Networks

PRTG Paessler Router Traffic Grapher

QoE Quality of Experience

QoS Quality of Service

RADSL Rate Adaptive Digital Subscriber Line

RF Radio Frequnecy

SDSL Single-line Digital Subscriber Line

SE Service Experience

SNR Service to Noise Ratio

SSM Speed of Service Maintenance

SUH Service Usage History

SVM Support Vector Machine

SVR Support Vector Regression

TCA Time Taken to Reach Call Center Agents

vISP Virtual Internet Service Provider

VDSL Very-high-rate Digital Subscriber Line

Chapter 1

Introduction

Fixed broadband Internet has revolutionized the way we access and interact with the digital world. With rapid advancements in technology and communication infrastructure, high-speed Internet connections are now widely available and affordable in many parts of the world. Fixed broadband Internet enables a wide range of applications, including video streaming, online gaming, remote work, telemedicine, and e-learning. It is also essential for accessing a variety of online services, such as education, healthcare, e-commerce, and entertainment [1].

Ethio telecom, the state-owned telecommunications provider in Ethiopia, is playing a key role in expanding access to fixed broadband Internet services across the country. Since June 2017 the operator is wholesaling bandwidth to vISP's to meet the surging demand for fixed broadband Internet in Ethiopia [2]. This has helped to bridge the digital divide and bring connectivity to both urban and rural areas. Fixed broadband services in Ethiopia encompass various technologies, including Digital Subscriber Line (DSL) and fiber optics. These technologies provide reliable and high-speed Internet connections, which are crucial for accessing a wide range of online services [2].

Quality of Service (QoS) and Quality of Experience (QoE) are two pivotal concepts in the realm of telecommunications and digital services, shaping the way we perceive and interact with technology [3]. QoS refers to the set of technical parameters and standards that ensure a consistent and reliable level of service delivery. It encompasses factors like network bandwidth, latency, packet loss, and jitter [3]. QoS mechanisms are integral to network management, allowing providers to prioritize and allocate resources efficiently, ensuring that real-time services like voice and video communication receive the required bandwidth and perform seamlessly. On the other hand, QoE is a user-centric concept, focusing on the end-user's perception of the service's quality. It takes into account not only the technical aspects but also subjective elements that influence the overall experience, such as ease of use, responsiveness, and content quality [3]. This interplay between QoS and QoE is a pivotal area of study and optimization for Internet service providers and researchers, as it directly impacts the user's digital experience [3].

As discussed by [4] achine Learning (ML) is a powerful tool that can be used to improve the QoS and QoE for fixed broadband Internet services. As digital technologies become more widespread, the performance of broadband services increasingly relies on the precise management of network resources, fault detection, and performance optimization. ML can revolutionize how QoS and QoE are monitored, assessed, and improved. ML algorithms can analyze large datasets to predict and prevent network disruptions, optimize bandwidth allocation, and enhance service reliability. This enables service providers to proactively address technical issues that could compromise QoS, such as latency or packet loss, before they impact the end user [4].

For this research, QoE modeling for fixed broadband users of ethio telecom and vISP's were implemented. This modeling was conducted utilizing SVMs and the Random Forest machine learning algorithms. The comprehensive model will take into account a range of parameters, including QoS metrics such as upload speed, download speed, jitter, latency, packet loss, and uptime duration. Additionally, subjective metrics, including user demographic information, location, service experience, end device type, and other influential factors are considered.

1.1 Statement of the Problem

The research initiative is driven by a commitment to tackle several critical challenges within the domain of Fixed Broadband Internet(FBBI), with a focused objective of providing effective resolutions to identified problems. One primary concern is the lack of an extensive comparative evaluation of machine learning algorithms tailored to enhance the QoE in FBBI. This research seeks to fill this void by conducting a thorough assessment of diverse machine learning methodologies, aiming to determine their efficacy in elevating the overall user experience. Furthermore, the inadequacy of current QoE metrics, which may not fully encapsulate the various factors influencing user satisfaction within fixed broadband services, is acknowledged. The study aims to refine and expand existing QoE metrics, ensuring a more comprehensive understanding of the intricate factors contributing to user satisfaction. Another critical challenge addressed is the high customer churn rate within FBBI, attributed to the absence of robust QoE modeling. By developing effective QoE models, the research seeks to contribute to customer retention strategies. Lastly, the absence of accurate machine learning models for predicting QoE in fixed broadband services hampers proactive user experience management. The research endeavors to develop refined machine learning models capable of accurate QoE prediction, facilitating providers in taking preemptive measures for enhanced user satisfaction and overall service quality.

1.2 Objective

1.2.1 General Objective

The main objective of the research is to develop and evaluate machine learning-based methods specifically SVM and Random Forest for modeling the quality of experience of fixed broadband service by incorporating QoS parameters.

1.2.2 Specific Objectives

To achieve the general objective of the thesis the below specific objectives have been done :

- Evaluate the QoE influential metrics for fixed broadband services to identify key areas for improvement.
- Conduct a comparative performance measurement for the proposed machine learning algorithms.
- Investigate the impact of different network conditions, such as latency, jitter, and packet loss, on the user's perception of quality.
- Optimization the parameters of the selected Machin learning algorithm to enhance the model performance.
- Develop recommendations for ISP/vISPs to enhance the FBB user experience using the research findings.

1.3 Literature Review

Fixed broadband Internet is becoming more and more popular, and it is important for ISPs to be able to accurately monitor, measure, and report on the QoE that their customers are having. ISPs can use this information to improve their services. Researchers are increasingly using machine learning and other techniques to help ISPs achieve this goal. Machine learning can be used to predict QoE metrics dynamically based on training data. This section provides an overview of existing QoE estimation models that use machine learning and other methods for a variety of services.

In one study [5], the QoE model was devised for social media video streaming services employing an ensemble machine learning algorithm. The researcher identifies there is not a study conducted on social media video streaming service that comprises the effect of user's device. The study used QoS measurements and user end device parameter to measure the user experience on the video streaming service using Mean Opinion Score (MOS) value using the ensemble machine learning algorithm. The results indicated a reasonable level of user satisfaction with the video streaming service. The study show that Ensemble machine learning proves robust in QoE modeling by amalgamating outcomes from diverse sub-models, enhancing prediction accuracy. However, this study fell short in encompassing all influential factors significantly impacting user service experiences, such as content type and location.

Another effort [6] introduced a method to model QoE in enterprise multimedia networks, utilizing both QoS and user subjective data (MOS). Machine learning algorithms, including regression and classification techniques, were employed to construct the models. QoS data was collected using Paessler Router Traffic Grapher (PRTG), while MOS data was gathered via Google Forms from network users. The study result shows among the employed machine learning algorithms scrutinized, the Random Forest Classifier exhibited superior accuracy in model performance. Nevertheless, the study had limitations as it did not consider all factors influencing QoE, including multimedia content types and user devices. Moreover, the dataset's size was relatively small, potentially yielding different results with a larger dataset.

In a separate study [7], the performance of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) in modeling user QoE for video streaming services was compared. The researcher motivated to conduct the user experience on the video streaming service by incorporating additional influential factors like video resolution and content type that are forget by previous studies. The researcher used Npref tool to measure the QoS metrics and used a survey questionary to collect the subjective data. The collected data are modeled using support vector regression and multivariant linear regression and the results demonstrated that SVR outperformed MLR in predicting user service perceptions. This emphasizes the importance of selecting the appropriate machine learning algorithm based on data characteristics. SVR suits datasets with non-linear data distributions, while MLR is better suited to linear data distributions. Nonetheless, this study omitted the consideration of Radio Frequency (RF) factors, such as signal-to-noise ratio (SNR), which can substantially affect video streaming quality. Thus, accounting for RF factors is crucial in QoE modeling.

In another investigation [8], the study measured QoE of users for the selected cloud service called "WeChat" The researcher motivated to emphasize the effect on mobile phone internal storage and RAM side effect on the user experience on the cloud service. Two mobile devices (Samsung and HTC) are used to conduct the study by providing saturated internal storage and ram size for Samsung phone and providing sufficient memory space for HTC device. The measure result shows, that users had a superior service experience using HTC devices compared to Samsung phones, primarily due to differences in internal storage and RAM. However, the study did not account for the effect of QoS on the "Wechat" cloud service experience. Incorporating technical metrics would provide a more realistic outcome than solely considering device parameters for QoE modeling.

Lastly, in a different study [9], both subjective and objective QoE modeling were employed on fixed broad band users. The in existence of earlier stud on QoE modeling on fixed broad band Internet service motivated the researcher. The researcher used Testmy.net for measuring QoS and Questionnaire survey to collect subjective data. The collected QoS were modeled using fuzzy logic machine learning algorithm Findings revealed that fixed broadband customers experienced superior QoE during off-peak hours than during on-peak hours. This shows that user satisfaction on fixed broadband Internet service vary in accordance with the time variation. However, the study could not ascertain whether this QoE disparity existed among customers using different technologies. This is noteworthy as the technology employed by customers can significantly influence their QoE. For instance, fiber connections tend to offer a superior QoE compared to copper connections. This limitation could introduce bias into the model's predictions for future data, potentially hindering accurate QoE predictions for customers using diverse technologies. In addition, the study not consider all the key quality indicator metrics for the Internet services, this will have an impact on the model predication performance.

1.4 Methodology

The method used in this thesis includes the below procedures:

- Subjective and Objective (QoS) are collected from both ethio telecom and vISP customers
- Label encoder applied to encoder the categorical measured data to numerical value before fitting to SVM and Random Forest ML algorithms.
- Standard scalar used to normalize the data with mean of 0 and unity variance, to reduce the effect of SVM and Random Forest model biasing. If the data is not scaled, the model may focus on the features with the largest scale and ignore the features with the smallest scale. In addition, standard scalar is a suitable method to identify the outlier-recorded data in each of the independent features
- To reduce the model complexity and prevent the model to learn from the nose data, machine learning based feature elimination technique is applied (recursive feature elimination) to sort the best-selected feature combination that can give the most model accuracy.
- To more enhance the model performance hyper parameter optimization is applied using a grid search technique to choose the optimal kernel type, C and Gamma parameter.
- The model performance evaluation conducted using k fold cross validation technique.
- The model will be trained using the selected features and hyper parameters.
- Finally, model prediction evaluation implemented using confusion matrix technique.

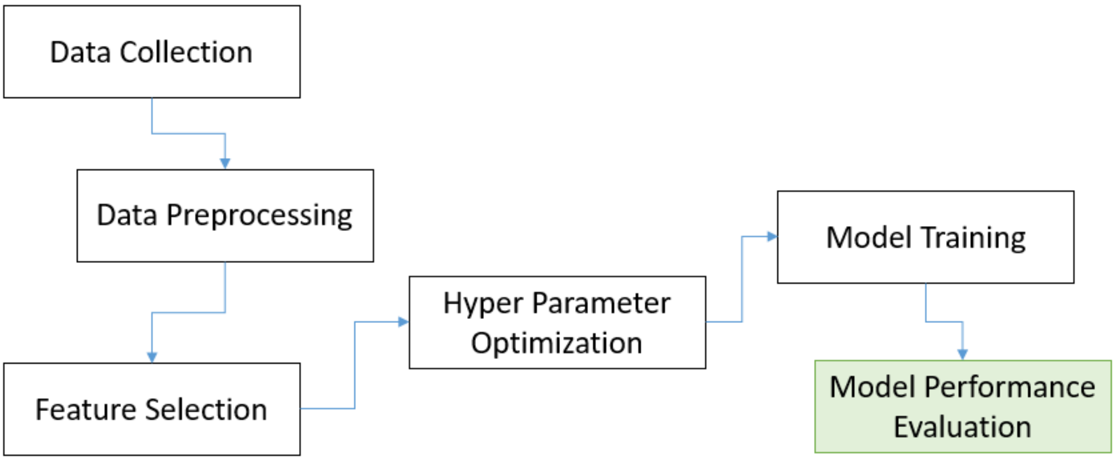

Figure 1.1: General System Model.

1.5 Scope and Limitations

The thesis scope focus on fixed broadband Internet users in Addis Ababa, Ethiopia both for the ISP (ethio telecom) and vISPs (Websprix and ZERGAW Cloud).

The below are the limitations of the thesis :

- Limitation of lack of all required QoS measurement from the vISP customers.
- Gathering the substantial amount of data from ISP and vISPs takes a significant amount of time.
- SVM and Random Forest models are complex and can be computationally expensive to train and deploy.
- Lack of similar studies on the subject matter.

1.6 Contributions

The primary contributions of this work can be outlined as follows: Firstly, it focuses on overseeing the Quality of Experience (QoE) within fixed broadband services and implementing corrective measures in response to influential factors. Specifically, it centers on accurately predicting customer perceptions regarding the service, a prediction that holds significant value for service providers. This aids in proactive network resource management to sustain high levels of QoE. Secondly, it involves assigning priorities to influential factors that have a substantial impact on the overall customer service experience. This prioritization allows for targeted efforts to address the most critical aspects affecting customer satisfaction. Lastly, it provides a comprehensive understanding of customer perceptions regarding the service by integrating both subjective and objective data sources. This approach offers a holistic view of the customer experience, enabling service providers to make informed decisions and enhancements to improve overall customer satisfaction and loyalty.

1.7 Thesis Organization

This thesis is structured into five main chapters. In Chapter 1, we delve into the introduction, problem statements, thesis objectives, a comprehensive literature review, the chosen methodology, the scope of the thesis, and its limitations. We also outline the contributions made by this study. Chapter 2 takes an in-depth look at background of fixed broadband services. This includes an examination of ethio telecom's fixed broadband service architecture and the proposed system model for vISPs. Moving on to Chapter 3, we explore ML approach for fixed broadband interent QoE modeling, with a focus on the utilization of the SVM and Random Forest machine learning algorithm. Chapter 4 delves into the methodology employed for designing the QoE model using SVM and Random Forest, offering a concise overview of each process and method employed. Chapter 5 is dedicated to presenting the results of our research and providing interpretations of these findings. Lastly, Chapter 6 serves as the conclusion of this thesis, encapsulating key insights and recommendations for future research endeavors.

Chapter 2

Overview of Fixed Broadband Network

2.1 Definition of Broadband Network

The definition of broadband varies across countries or telecommunication regulatory bodies; they have specific criteria for what qualifies as broadband, specifying minimum speed thresholds or other technical requirements. The Federal Communications Commission (FCC) defines broadband in terms of speed. The current standard specifies at least 25 megabits per second (Mbps) for downloading and 3 Mbps for uploading. In essence, broadband is a service that enables reliable, high-speed transfer of data, voice, and video over the Internet. In the 1990s, the only option to access the Internet was through a dial-up connection using your home phone line and a modem, but technologies have since changed and evolved. As interest and Internet applications grew, solutions were needed to meet the demands for greater data transmission [10]. However, in Ethiopia, ethio telecom ISP offers broadband service subscriptions beginning at 4 Mbps, as outlined in the company's fixed broadband circular guideline released in October 2023.

According to [11] the terms 'broadband network' and 'broadband internet' refer to different aspects of the communication infrastructure, emphasizing distinct elements of the connectivity landscape. A broadband network denotes a high-capacity communication infrastructure capable of transmitting multiple signals concurrently. It encompasses the physical infrastructure that enables communication within a specific area or across a broader region, facilitating the exchange of data, voice, and video signals. However, broadband internet refers to the service that provides users with access to the internet through the broadband network.

2.2 Types of Broadband Network

There are two groups of broadband access technologies. Fixed Broadband Technologies and Mobile Broadband Technologies [10].

2.2.1 Fixed Broadband Technologies

Based on [10] fixed broadband refers to those technologies where the end-user must remain at the same location to use the broadband service. The access network is associated with a specific physical location. Fixed broadband can be provided by wireline, wireless. The below diagram illustrates the three main layers of the fixed wired broadband network architecture: access layer, aggregation layer, and core layer.

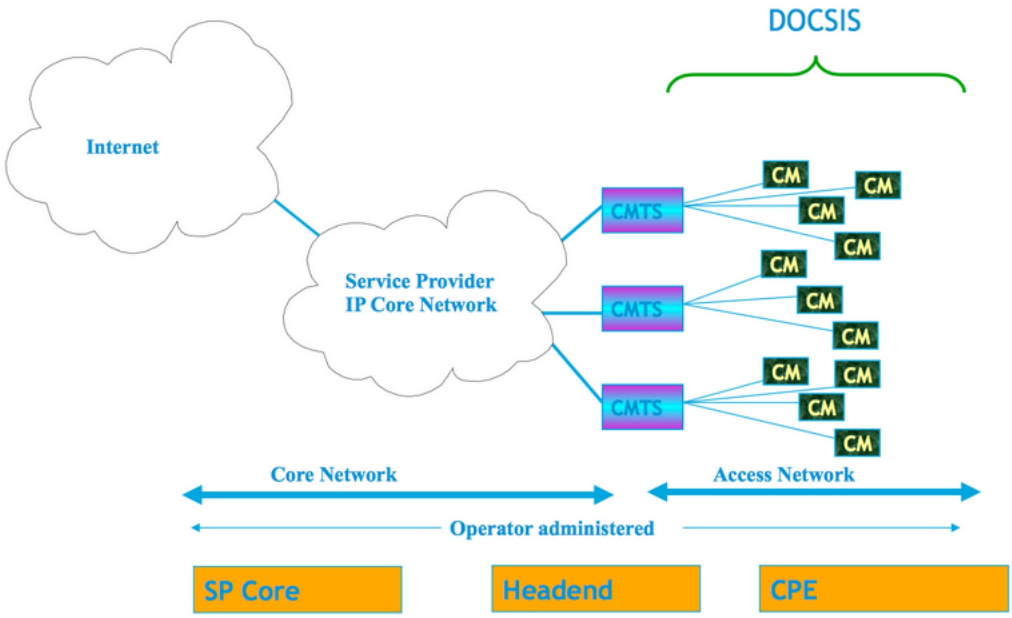

Figure 2.1: Fixed Broadband Access Network Architecture [10].

Access Layer connects the subscriber's home or office to the service provider's network. Various access technologies can be used, including fiber-to-the-home (FTTH), fiber-to-the-node (FTTN), DSL, and cable modem [10].

Aggregation layer aggregates traffic from multiple access networks and routes it to the core layer. It ensures that traffic is directed to the correct destination [10].

Core layer forms the backbone of the fixed broadband network. It transports high-bandwidth traffic between different parts of the network and to other networks, such as the Internet [10].

The CPE comprises essential components, including the modem, router, and additional devices strategically positioned at the subscriber's premises. These devices play a crucial role in establishing a seamless connection between the subscriber's home or office and the expansive network maintained by the service provider. Concurrently, NAPs serve as pivotal interconnection hubs, facilitating the seamless integration of the service provider's network with other networks, including the vast expanse of the Internet. NAPs play a critical role in enabling peering activities with other networks and facilitating the exchange of traffic [10]. In the realm of high-bandwidth traffic transportation over extended distances, OTNs take center stage. Leveraging technologies like dense Wavelength Division Multiplexing (DWDM) and Carrier Ethernet, OTNs achieve remarkable speeds, ensuring efficient and reliable transmission [10].

There are several types of fixed wired broadband connections, each utilizing different technologies to deliver high-speed internet access. Here are some common types [11]:

Digital Subscriber Line (DSL): It is a traditional type of internet connection that is established between computer and an ISP using a modem. An inexpensive internet connection type is slow in speed. For accessing internet connection through a dial-up connection, telephone connection is needed since phone number is required on the computer. Along with that, a modem is required since it acts as an interference between computer and telephone line. Dial up connection uses either Serial Line Internet Protocol (SLIP) or Point-to-Point Protocol (PPP). Such internet types of connection are analog in nature since they do not allow users to receive or make phone calls via home phone services while using Internet [11]. The below are the DSL families :

Asymmetric Digital Subscriber Line(ADSL): The most promising of the DSL technologies is ADSL or Asymmetric Digital Subscriber Line. ADSL looks to make the most impact in residential access and the SOHO (Small Office Home Office) market. ADSL is asymmetric, meaning that the downstream bandwidth is higher than the upstream bandwidth. Downstream refers to traffic in the direction towards the subscriber, and upstream refers to data sent from the subscriber back to the network. The asymmetric nature of ADSL lends itself well to applications like the web and client server applications [12].

High-data-rate Digital Subscriber Line (HDSL): The most common DSL deployed today is HDSL. Telephone companies mostly use HDSL to provision other services. HDSL symmetrically delivers 1.544 Mb/s over two sets of copper twisted pair. Which is the same rate as a T-1 type connection. This allows Telco's (short for telephone companies) to use HDSL to deliver T-1 services. HDSL's operating range is about 12,000 feet, and it is possible to extend that by using repeaters along the line to the customer. HDSL is mostly used to deploy PBX network connections, interexchange POP's (Point Of Presence), and directly connecting servers to the Internet [12].

Single-line Digital Subscriber Line (SDSL): Also known as, Symmetric Digital Subscriber Line Similar to HDSL, SDSL delivers the same 1.544 Mb/s, but it does it on a single set of twisted pair of copper. This limits SDSL's reach to 10,000 feet. SDSL could take hold in niche markets like residential video conferencing or connecting LAN's over short distances [12].

Very-high-rate Digital Subscriber Line (VDSL): VDSL technology operates on a single set of copper twisted pair, and delivers data in the range of 13 Mb/s to 52 Mb/s. The VDSL standard is still in the works but there are already applications for the technology. One use for it is in getting high data rate services from the telephone companies' central office to the subscriber via a FTTN (Fiber to the Neighborhood) network. VDSL would be used to connect premises distribution networks to the Optical Network Unit or ONU. The optical network unit is in turn connected via fiber optical line to the Telco's central office [12].

Rate Adaptive Digital Subscriber Line(RADSL): RADSL helps to provide a more consistent service for its subscribers by taking the uncertainties of line conditions out of the equation when setting up a DSL connection. RADSL can adjust line speed based on the gauge of the wire, the distance between subscriber and the central office, and the condition of the line. It also takes care of fluctuations that the weather can induce into the line [12].

Cable: Cable broadband uses coaxial cable that connects to your computer through a modem to transmit data. Coaxial cable was originally designed for transmitting television signals, and it has far less signal deterioration and much faster broadband speeds compared to telephone lines. Another advantage of cable is its capacity to handle a greater volume of audio and visual signals. This capability allows you to get your phone and digital TV as well as Internet connection services from your cable broadband provider. Cable is either buried underground or uses same utility poles already in place for your telephone and electricity [11].

Fiber Optic: Fiber optic technology uses glass polymer fibers for data transmission. This technology converts electrical signals to light, which is sent through transparent glass fibers. It enables significantly faster data transmission over greater distances compared to other broadband technologies. Due to the fact that the transmission material is glass, which cannot generate electricity, fiber optic broadband is immune to interference, and this minimizes data deterioration. Other wireline services such as cable broadband often incorporate fiber into some portion of their infrastructure [11].

Based on information provided by [13] fixed Wireless broadband connects a home or business to the Internet using a radio link between the customer's location and the service provider's facility. Wireless technologies using longer-range directional equipment provide broadband service in remote or sparsely populated areas where DSL or cable modem service would be costly to provide or fiber network installations may be too capital intensive. The below figure depics, fixed wireless broadband architecture encompasses access point, base station nodes and WiMAX base station.

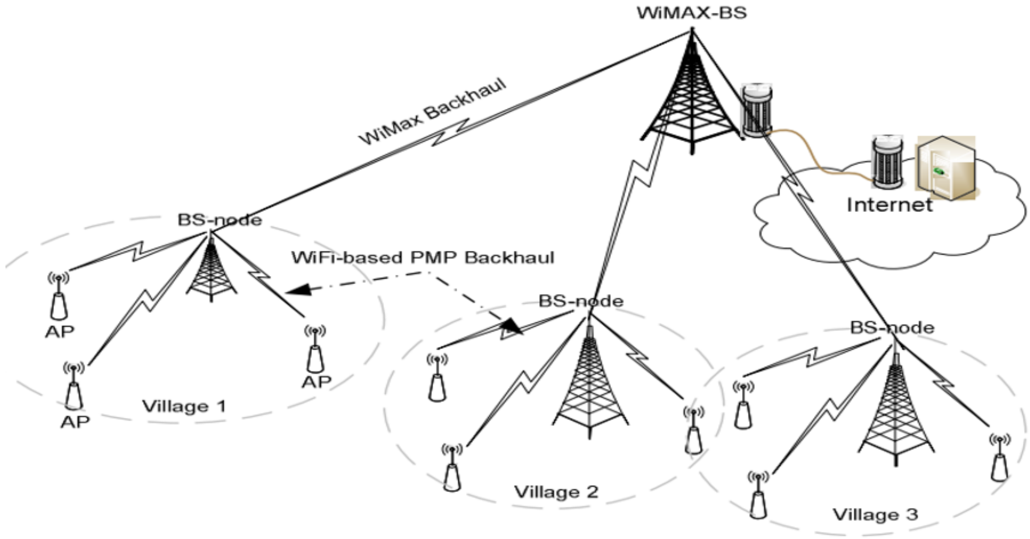

Figure 2.2: Fixed wireless broadband architecture [13].

Types of fixed wireless broadband connection type :

Satellite: Another way to connect to the Internet is through a satellite connection. Satellite Internet is good for people who live in areas where other Internet options aren't available. However, it is a more expensive option for Internet access. With satellite Internet, you are connecting a computer to the Internet with a modem and satellite dish. Satellite Internet provides a way for remote and rural areas to obtain Internet access. However, the weather will affect the signal path, so its expected to have poor Internet quality during windy or rainy weather. It is a costly Internet connection. The speed is relatively slower than other connections [14].

Air Fiber (Aironet): Air Fiber or Aironet is a technology that uses radio signals to establish high-speed wireless connections. It is often employed for point-to-point communication, linking two fixed locations. This can be useful for building-to-building connections or providing internet access in areas without wired infrastructure.

Worldwide Interoperability for Microwave Access (WiMAX): WiMAX is a wireless communication standard that provides high-speed Broadband Wireless Access (BWA). It was developed to offer wireless broadband connectivity over long distances and was intended to be an alternative to traditional wired broadband technologies [14].

2.2.2 Mobile Broadband Technologies

According to [11] mobile broadband refers to a high-speed wireless internet access service that allows users to connect to the internet using mobile devices, such as smartphones, tablets, or other wireless modems, without the need for physical cables. This type of broadband service is provided by mobile network operators and utilizes technologies such as 3G, 4G LTE, and 5G to deliver fast and reliable internet connectivity to users on the move . Mobile broadband enables users to access the internet, browse websites, stream videos, and engage in various online activities while being connected to the mobile network infrastructure. The portability and flexibility of mobile broadband make it a popular choice for individuals who require internet access in different locations or while traveling. Users typically subscribe to mobile broadband plans offered by service providers, which include data allowances for a specified period [11].

2.3 Application of Fixed Broadband Network

Fixed broadband networks provide high-speed internet access to homes and businesses through a wired connection. This type of connection is typically more reliable and consistent than wireless broadband options, such as satellite or cellular internet. [11]. Fixed broadband networks have a multitude of key applications. Firstly, they provide high-speed connectivity essential for various online activities such as web browsing, email communication, social media engagement, and online gaming [11]. Secondly, these networks serve as the backbone for online video streaming services like Netflix, Hulu, and YouTube, necessitating high-speed connections for seamless video playback [11]. Additionally, fixed broadband networks enable Voice over Internet Protocol (VoIP) technology, increasingly favored as a cheaper alternative to traditional phone services. Furthermore, they are crucial for online gaming, offering low latency and high bandwidth for real-time multiplayer experiences [11]. Moreover, fixed broadband facilitates teleworking by providing reliable, high-speed connectivity for remote work tasks like accessing files, participating in video conferences, and collaborating with colleagues. These networks also support distance learning, allowing individuals to pursue online courses and degrees without the need for physical attendance at traditional schools. Furthermore, they enable smart home applications, connecting appliances and devices for remote control and automation [11]. Additionally, fixed broadband networks are utilized for security and surveillance purposes, linking cameras and surveillance equipment for remote monitoring of properties. Moreover, they play a significant role in healthcare by connecting medical devices and facilitating the delivery of health services remotely [11]. Lastly, they are instrumental in government operations, connecting offices and providing online services to citizens [11].

2.4 Roles of vISPs in Fixed Broadband Internet

2.4.1 Definition of vISPs

vISPs is a company that offers internet services under its own brand name but does not own or operate its own infrastructure. Instead, it leases internet access from a wholesale ISP, such as ethio telecom, and then resells it to end users. Beyond the fundamental internet service, vISPs have the opportunity to offer supplementary Value Added Services such as Video on Demand, IPTV, and others [2]. The below figure shows, ethio telecom virtual Internet service provisioning market segment approach.

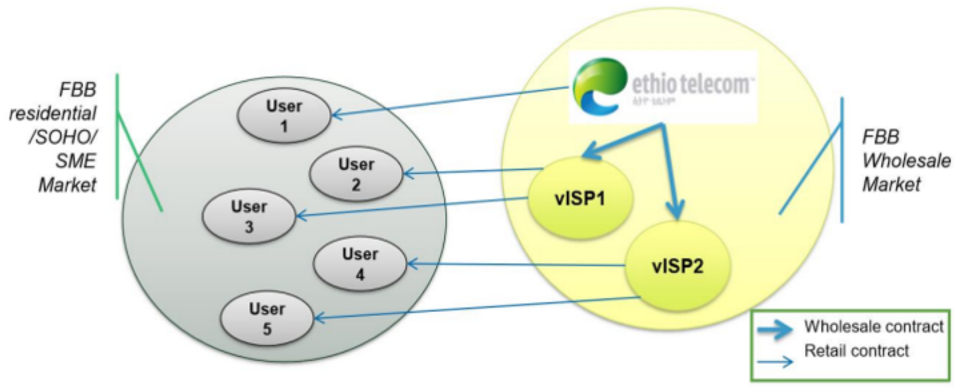

Figure 2.3: ethio telecom and vISPs market segment approach [2].

2.4.2 Technical Model of vISPs

According to the provided commercial circular of ethio telecom ISP, there are two service-provisioning methods available for potential partners to choose from based on their preferences [2]:

vISPs Technical Model 1 : Under this model, the vISPs will establish its own infrastructure or secondary network following its access devices [2].

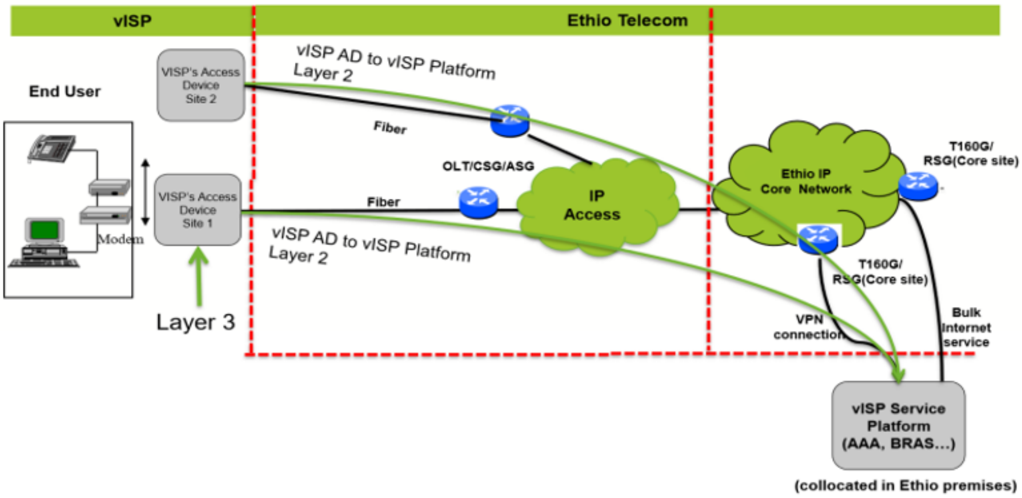

Figure 2.4: vISPs technical model one [2].

The above figure depicts, the vISP has subscribed to wholesale internet capacity to retail to its end users, connecting it to its service platform within ethio telecom's premises and linked to the ethio network. ethio telecom will provide the connection up to the vISP's Access Device via fiber, utilizing fiber cores from various ethio equipment. The vISP's access device should support multiple services across various technologies to manage current customer demands. vISPs will handle secondary network deployment, internal cabling, and establish their own service platform [2].

vISPs Technical Model 2 : In this model, the vISPs will utilize ethio telecom's secondary network after installing its access devices [2]. The below figures shows, ethio telecom will permit vISPs to utilize its secondary network infrastructure in areas where deployment is already completed. vISPs will be required to pay monthly rent per line for the use of this infrastructure. ethio telecom will assume responsibility for the maintenance of these secondary lines in the event of failure [2].

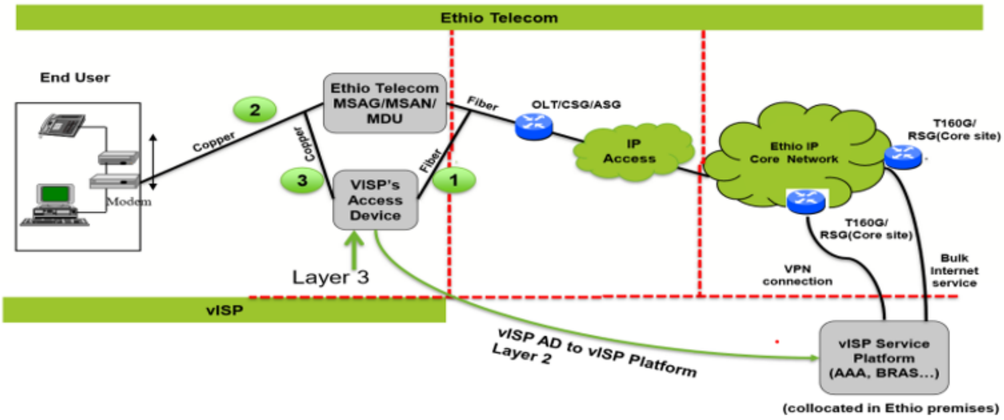

Figure 2.5: vISPs technical model two [2].

2.4.3 Benefit of vISPs

As per the information provided by [15] one of the key advantages of vISPs is their ability to offer flexible and customizable plans to their customers. Since they do not own the infrastructure, vISPs are not bound by the limitations of a specific network. This allows them to tailor their services to meet the unique needs of their customers. Whether it's a small business looking for a high-speed connection or a residential user seeking a reliable and affordable option, vISPs can provide a solution that fits their requirements.

Another benefit of vISPs is their ability to offer nationwide coverage. By leasing network capacity from multiple ISPs, vISPs can provide internet access to customers in remote or undeserved areas where traditional ISPs may not have a presence. This has been particularly beneficial for rural communities and developing regions, where access to reliable internet has been a challenge. With vISPs, these communities can now enjoy the benefits of being connected to the digital world [15].

In addition to their flexibility and wide coverage, vISPs also offer competitive pricing. Since they do not have the overhead costs associated with owning and maintaining infrastructure, vISPs can pass on the savings to their customers. This has made them an attractive option for budget-conscious consumers who are looking for affordable internet access without compromising on quality [15].

2.4.4 Challenges of vISPs

vISPs grapple with substantial operational and competitive hurdles. Firstly, the significant costs associated with constructing and maintaining essential infrastructure, encompassing access devices and service platforms, pose a formidable challenge [2]. Secondly, vISPs are heavily reliant on wholesale internet services from primary providers, rendering them susceptible to disruptions or issues with the wholesale service, such as outages, directly impacting service delivery and compromising reliability. The intense competition in the market serves as another major obstacle for vISPs, as direct rivalry with primary service providers makes securing market share and establishing a distinct identity challenging in the face of established brands and extensive resources [2]. Customer acquisition and retention present additional challenges, necessitating effective marketing strategies and adaptive retention programs in a fiercely competitive environment [2]. Lastly, expanding geographical coverage to reach undeserved areas poses logistical and financial challenges for vISPs, limiting their ability to broaden service availability. Addressing these multifaceted challenges is imperative for ensuring the sustainable growth and success of vISPs [2].

Chapter 3

Machine Learning for Fixed Broadband Internet QoE Modeling

3.1 Introduction to Machine Learning

According to [16] learning is the process of converting experience into expertise or knowledge. The input to a learning algorithm is training data, representing experience, and the output is some expertise, which usually takes the form of another computer program that can perform some task. Machine Learning (ML) is one subset of Artificial Intelligence (AI) that aims to build analytical models by learning from existing data. After decades of incremental development and technological innovation, ML has emerged as a powerful discipline for a wide range of scientific research and industrial applications, with a particular strength in discovering patterns in complex, high-dimensional data and examining non-linear relationships [17].

In the vast landscape of data, machine learning excels at uncovering hidden patterns and relationships, serving as an effective tool. It addresses challenges faced by human-designed machines in real-world environments, enabling on-the-job improvement and adaptation based on experiences. In tasks requiring extensive knowledge, machine learning allows machines to gradually acquire and internalize information. The adaptability of machine learning models to evolving environments reduces the need for constant redesign, enhancing the resilience of AI systems. In the ever-expanding realm of human knowledge, machine learning's continuous learning approach contributes to the robustness and relevance of AI systems in an ever-changing landscape [18].

The ML has been greatly expanded in various industries, including electricity, water, healthcare, etc. However, the telecommunication sector should be the leading industry hosting the other sectors. Explicitly, each machine should be able to learn the execution of a particular task, to keep the performance to a certain level based on the experience, where the system aims to keep reliably repeating the execution of the task to improve the performance. ML in telecom sector providing benefits in network optimization, fraud detection, customer experience, and resource management. ML algorithms enhance network performance by predicting and addressing issues proactively, optimize resource allocation, and contribute to dynamic pricing strategies [18].

Machine Learning algorithms are mainly divided into four categories: Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, and Reinforcement Learning [19]. Supervised learning is typically the task of machine learning to learn a function that maps an input to an output based on sample input-output pairs. It uses labeled training data and a collection of training examples to infer a function. Supervised learning is carried out when certain goals are identified to be accomplished from a certain set of inputs i.e., a task-driven approach. The most common supervised tasks are "classification" that separates the data, and "regression" that fits the data [19]. Unsupervised learning analyzes unlabeled datasets without the need for human interference, i.e., a data -driven process. This is widely used for extracting generative features, identifying meaningful trends and structures, groupings in results, and exploratory purposes. The most common unsupervised learning tasks are clustering, density estimation, feature learning, dimensional reduction, finding association rules, anomaly detection, etc [19]. Semisupervised learning can be defined as a hybridization of the above-mentioned supervised and unsupervised methods, as it operates on both labeled and unlabeled data. Thus, it falls between learning "without supervision" and learning "with supervision". In the real world, labeled data could be rare in several contexts, and unlabeled data are numerous, where semi-supervised learning is useful. The ultimate goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labeled data alone from the model. Some application areas where semi-supervised learning is used include machine translation, fraud detection, labeling data, and text classification [19]. Reinforcement learning is a type of machine learning algorithm that enables software agents and machines to automatically evaluate the optimal behavior in a particular context or environment to improve its efficiency,i.e., an environment-driven approach. This type of learning is based on reward or penalty, and its ultimate goal is to use insights obtained from environmental activists to take action to increase the reward or minimize the risk. It is a powerful tool for training AI models that can help increase automation or optimize the operational efficiency of sophisticated systems such as robotics, autonomous driving tasks, manufacturing and supply chain logistics, however, not preferable to use it for solving the basic or straightforward problems [19].

3.2 Support Vector Machine

In reference to [20] SVM were introduced by Vladimir Vapnik and colleague in the late seventies. SVM are a relatively new learning method used for binary classification [21]. The basic idea is to find a hyperplane which separates the d-dimensional data perfectly into its two classes. However, since example data is often not linearly separable, SVM introduce the notion of a “kernel induced feature space” which casts the data into a higher dimensional space where the data is separable. Typically, casting into such a space would cause problems computationally, and with overfitting. The key insight used in SVM is that the higher-dimensional space doesn’t need to be dealt with directly (as it turns out, only the formula for the dot-product in that space is needed), which eliminates the above concerns. Furthermore, the VC-dimension (a measure of a system’s likelihood to perform well on unseen data) of SVM can be explicitly calculated, unlike other learning methods like neural networks, for which there is no measure. Overall, SVM are intuitive, theoretically well- founded, and have shown to be practically successful. SVM have also been extended to solve regression tasks (where the system is trained to output a numerical value, rather than “yes/no” classification).

In pattern recognition, we are given training data of the form $(x_1, y_1), \dots, (x_l, y_l) \in \mathbb{R}^n \times \{+1, -1\}$, that is n-dimensional patterns (vectors) x_i and their labels y_i. A label with the value of $+1$ denotes that the vector is classified to class $+1$, and a label of -1 denotes that the vector is part of class -1. We thus try to find a function $f(x) = y : \mathbb{R}^n \to \{+1, -1\}$ that, apart from correctly classifying the patterns in the training data (a relatively simple task), correctly classifies unseen patterns too. This is called generalization [21].
The SVM are based on the class of hyperplanes:

$$(w \cdot x) + b = 0; \quad w \in \mathbb{R}^n,\ b \in \mathbb{R} \tag{3.1}$$

which divide the input space into two: one part containing vectors of class -1 and the other containing those that are part of class $+1$. If there exists such a hyperplane, the data is said to be linearly separable (non-separable otherwise). To find the class of a particular vector x, we use the following decision function [21]. (3.1)

$$f(x) = \mathrm{sign}((w \cdot x) + b) \tag{3.2}$$

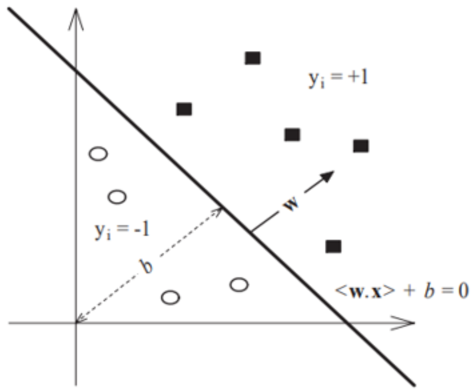

Figure 3.1: A separating hyperplane for a two-dimensional (2D) space [20].

3.2.1 Types of SVM

Linear SVM

It is possible to prove that the optimal hyperplane, defined as the one with the maximal margin of separation between the two classes, stems from the function class with the lowest capacity. This hyperplane can be constructed uniquely by solving a constrained quadratic optimization problem whose solution has an expansion in terms of a subset of training patterns that lie closest to the boundary. These training patterns, called support vectors, carry all relevant information about the classification problem [20]. There is just one crucial property of the algorithm that needs to be emphasized: Both the quadratic programming problem and the final decision function depend only on dot products between patterns, which is precisely what lets us generalize to the nonlinear case [20].

$$x_i \cdot w + b \geq 1 \quad \text{when } y_i = +1 \tag{3.3}$$
$$x_i \cdot w + b \leq -1 \quad \text{when } y_i = -1 \tag{3.4}$$

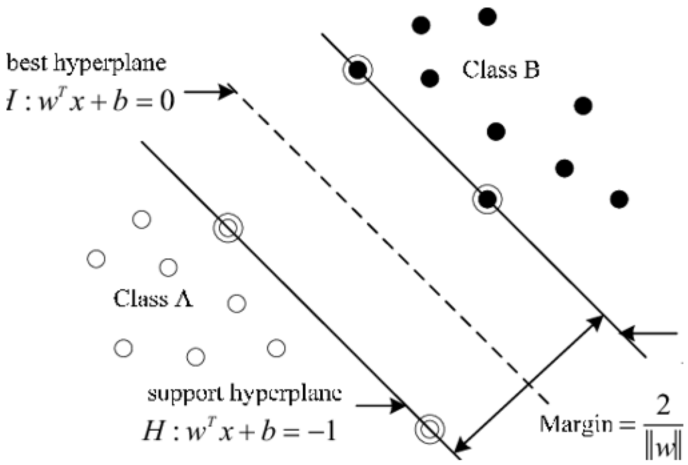

Figure 3.2: The hyperplane of SVM [21].

The key to set up the optimization problem of finding a maximal margin hyperplane is to observe that for two points x^+ and x^- that lie nearest to it, it is true that [20] :

$$(w, x^+) + b = +1 \tag{3.5}$$
$$(w, x^-) + b = -1 \tag{3.6}$$
$$(w, (x^+ - x^-)) + b = 2 \tag{3.7}$$
$$\left(\frac{w}{\|w\|}, (x^+ - x^-)\right) = \frac{2}{\|w\|} \tag{3.8}$$

Objective: $\min\|w\|$

$$s, t \quad y_i[(w, x_i) + b] \geq 1 \tag{3.9}$$

This equation can be solved by constructing a Lagrangian function:

$$\mathcal{L}(w, b, \boldsymbol{\alpha}) = \frac{1}{2}(w, w) - \sum_i \alpha_i[y_i((w, x_i) + b) - 1], \quad \alpha_i \geq 0 \tag{3.10}$$

Where the coefficient α is a Lagrange multiplier, and by transforming it into the corresponding dual Lagrangian by imposing the optimal conditions. Notice that the nearest points, called support vectors, determine its position [20].

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_i y_i \alpha_i x_i = 0 \tag{3.11}$$

$$\frac{\partial \mathcal{L}}{\partial b} = w - \sum_i y_i \alpha_i = 0 \tag{3.12}$$

The result is a quadratic programming problem with linear constraints m:

$$w(\boldsymbol{\alpha}) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i, x_j) \tag{3.13}$$

$$\alpha_i \geq 0 \tag{3.14}$$

$$\sum_i \alpha_i y_i = 0 \tag{3.15}$$

That presents just a global maximum and can always be exactly solved efficiently. The resulting solution has the property that:

$$w = \sum_i y_i \alpha_i x_i \tag{3.16}$$

In fact, often most of the coefficients α_i are equal to zero. The only positive coefficients correspond to the points that lie closest to the hyperplane, and for this reason, such points go under the name of support vectors. The final decision function can be written as [20]:

$$f(x) = (w, x) + b = \sum_i y_i \alpha_i (x_i, x) + b \tag{3.17}$$

Non-Linear SVM

If data is linear, a separating hyper plane may be used to divide the data. However, it is often the case that the data is far from linear and the datasets are inseparable. To allow for this kernels are used to non-linearly map the input data to a high-dimensional space. The new mapping is then linearly separable [22].

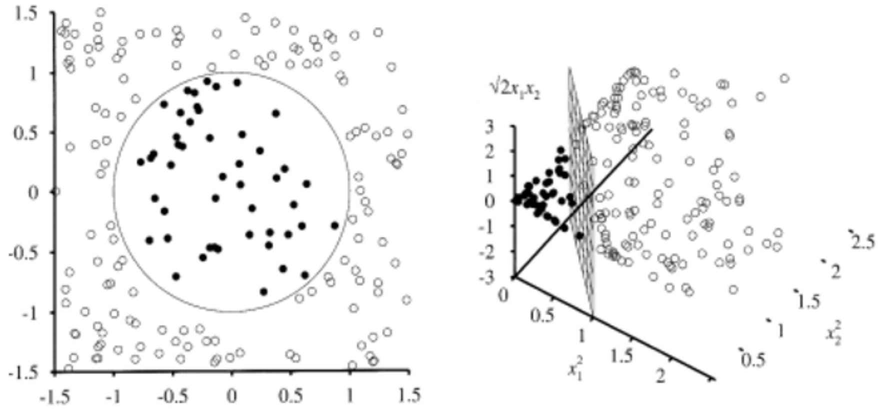

Figure 3.3: Feature Space Representation [22].

This mapping is defined by the Kernel:

$$K(x, y) = \phi(x) \cdot \phi(y) \tag{3.18}$$

Transforming the data into feature space makes it possible to define a similarity measure based on the dot product. If the feature space is chosen suitably, pattern recognition can be easy [22].

$$(x_1 \cdot x_2) \leftarrow K(x_1, x_2) = \phi(x_1) \cdot \phi(x_2) \tag{3.19}$$

Kernel Functions: The idea of the kernel function is to enable operations to be performed in the input space rather than the potentially high-dimensional feature space. Hence the inner product does not need to be evaluated in the feature space. We want the function to perform mapping of the attributes of the input space to the feature space. The kernel function plays a critical role in SVM and its performance. It is based upon reproducing Kernel Hilbert Spaces [22]. The different kernel functions are listed below :

$$\textbf{Linear kernel:} \quad K(x, x') = x_i \cdot x_j \tag{3.20}$$

$$\textbf{Polynomial function:} \quad K(x, x') = ((x, x') + c)^d \tag{3.21}$$

$$\textbf{Radial Basis Function (RBF):} \quad K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{3.22}$$

$$\textbf{Sigmoid kernel:} \quad K(x, x') = \tanh(\alpha x_i \cdot x_j + c) \tag{3.23}$$

3.2.2 SVM Multi Classification Approach

SVM are originally designed for binary classification. However, it can be effectively extended to multiclass classification by breaking down the multiclass classification problem into a series of binary classification problems [23]. The two most popular multiclass SVM methods: One Vs Rest and One Vs One [23].

One Versus Rest SVM

The one-vs-rest method construct k number of SVM models where k is the number of classes [23].

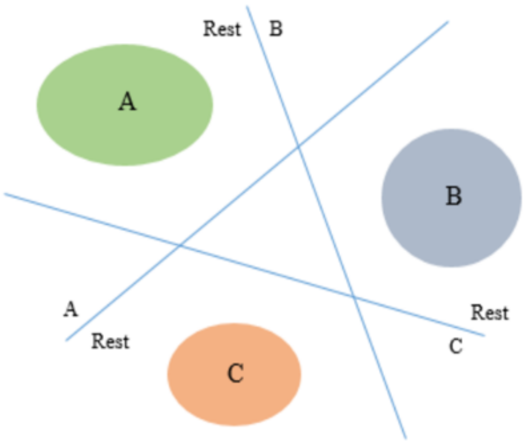

Figure 3.4: One versus rest method: classification between three classes [23].

One Vs One SVM

The one-vs-one method aims to get rid of the imbalance problem of the one-vs-rest method by training binary classifiers only with the data belonging to two original classes designated by each classifier. The one-vs-one method constructs $k(k-1)/2$ classifiers, given that k is the number of classes [23].

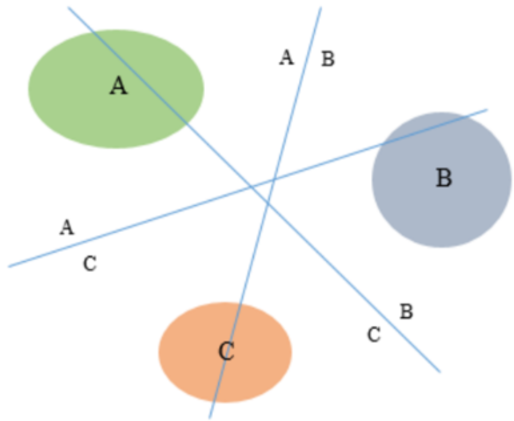

Figure 3.5: One versus one method: classification between three classes [23].

3.2.3 Support Vector Machine Hyperparamters

Choosing the right values for C and γ in a SVM is crucial for optimizing its performance. These hyperparameters significantly impact the decision boundary and, consequently, the accuracy and generalization of the model [24]. Here's a breakdown of their roles:

C: Regularization Parameter, controls the trade-off between model complexity and margin maximization.

Higher C:

- Encourages the model to fit the training data more closely, even at the expense of a smaller margin and increased risk of overfitting [24].
- May lead to increased sensitivity to noise in the data.

Lower C:

- Allows the model to prioritize a larger margin between classes, potentially sacrificing some training accuracy. Can lead to underfitting if the margin is too large [24].

Gamma: Determines the influence of each data point on the decision boundary.

Higher Gamma:

– Results in a more flexible decision boundary, able to capture complex relationships between data points. Can lead to overfitting and increased sensitivity to noise [24].

Lower Gamma:

– Creates a smoother decision boundary, less prone to overfitting. May not be able to capture intricate data patterns effectively [24].

3.3 Random Forest Machine Learning Algorithm

Random forests are known as ensemble learning methods used for classification and regression. Random forests are essentially a collection of decision trees that are each fit on a subsample of the data. While an individual tree is typically noisey and subject to high variance, random forests average many different trees, which in turn reduces the variability and leave us with a powerful classifier [25]. Random forests are also non-parametric and require little to no parameter tuning. They differ from many common machine learning models used today that are typically optimized using gradient descent. Models like linear regression, support vector machines, neural networks, etc. require a lot of matrix based operations, while tree based models like random forest are constructed with basic arithmetic. In other words, to build a tree all we're really doing is selecting a hand full of observations from out dataset, picking a few features to look through, and finding the value that makes the best split in our data [25].

3.3.1 Random Forest Learning Techniques

Bagging (Bootstrap Aggregating)

Bagging is a fundamental ensemble technique that aims to reduce variance and improve the stability of a model. It does so by training multiple instances of the same learning algorithm on different subsets of the data [25].
Steps:

Bootstrapped Sampling : Like Random Forests, Bagging involves sampling the training data with replacement to create multiple subsets [25].

Parallel Training : Each subset is used to train a separate model in parallel [25].

Aggregation: The final prediction is obtained by aggregating the predictions of all individual models (e.g., averaging for regression, and voting for classification) [25].

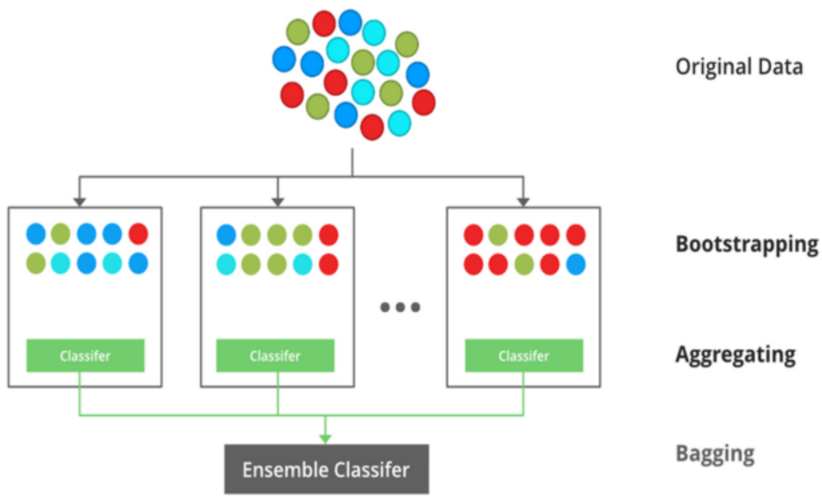

Figure 3.6: Bootstrap aggregating technique random forest learning [25].

Strengths and Applications:

Variance Reduction: Bagging reduces the variance of a model, which is especially beneficial for high-variance algorithms like decision trees [25].

Stability: It makes the model more robust to noise and outliers in the data [25].

Parallelization: Bagging allows for parallel training of models, making it computationally efficient [25].

Boosting

Boosting is an ensemble technique that focuses on improving the performance of weak learners (models that are slightly better than random guessing) by combining them sequentially [25].

Steps:

Sequential Training: Boosting builds models sequentially, with each subsequent model aiming to correct the errors of its predecessor [25].

Instance Weighting: Each data point is assigned a weight, and misclassified points receive higher weights to prioritize them in the next iteration [25].

Aggregation with Weighting: The final prediction is obtained by aggregating the weighted predictions of all individual models [25].

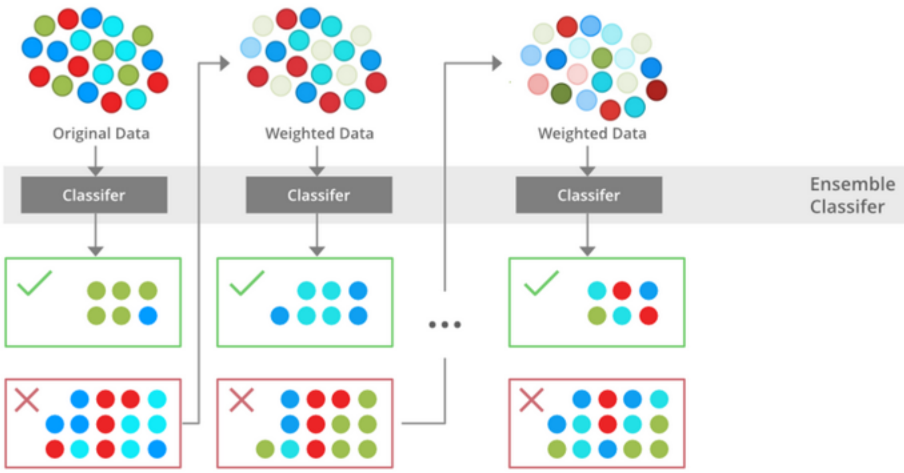

Figure 3.7: Boosting technique random forest learning [25].

Strengths and Applications:

High Accuracy: Boosting can achieve high accuracy even with simple base learners [25].

Handling Imbalanced Data: It is effective at handling imbalanced datasets by assigning higher weights to the minority class [25].

Adaptability: Boosting can be adapted to different types of weak learners, making it versatile [21].

Ensemble techniques like Random Forests, Bagging, Random Subspace, and Boosting provide powerful tools to improve the accuracy and stability of machine learning models. Understanding the principles behind each technique allows data scientists and machine learning practitioners to choose the most appropriate approach for their specific tasks. By leveraging the collective intelligence of multiple models, ensemble learning continues to be a cornerstone of modern machine learning practice [25].

3.3.2 Variable Selection Criterion

There are several ways to evaluate which features are the most important, but two of the most used criteria are Information Gain and Gini Impurity [26]. Information Gain is based on the principle of Entropy, which essentially measures the uncertainty in the data [27]. The entropy for a dataset D is defined as:

$$Entropy(D) = -\sum_{i=1}^{j} P_i \log_2(P_i) \tag{3.24}$$

where P_i represents the proportion of elements of class i in the dataset. The information gain is given by the reduction in entropy, or uncertainty, after splitting on a certain feature. If a split significantly decreases the uncertainty, it indicates that a considerable amount of information can be extracted from that feature and that it is therefore important for the classification [27]. The information gain of splitting on a feature A in the dataset D is given by:

$$Gain(D, A) = Entropy(D) - \sum_{v=1}^{V} \frac{|D_v|}{|D|} Entropy(D_v) \tag{3.25}$$

where D_v is the subset of D where all elements have the value v for attribute A. Thus, the best feature to split on becomes the one which minimizes the weighted sum of the entropy in all new sub-branches [27]. An alternative is to split features using the Gini impurity. It measures the probability of an element in the dataset being misclassified when labeled randomly according to the label distribution in the dataset [27]. The Gini impurity of a dataset D is given by:

$$GINI(D) = 1 - \sum_{i=1}^{j} P_i^2 \tag{3.26}$$

where P_i is the relative frequency of class i in D. If the dataset is split on attribute A into two subsets L_1 and L_2 with sizes N_1 and N_2 respectively [27]. GINI is calculated as :

$$GINI_A(L) = \frac{N_1}{N} GINI(L_1) + \frac{N_2}{N} GINI(L_2) \tag{3.27}$$

Reduction in impurity is calculated as:

$$\Delta GINI(A) = GINI(L) - GINI_A(L) \tag{3.28}$$

3.3.3 Random Forest Hyperparameters

Along with the two criteria for splitting, there are a number of other relevant hyperparameters that affect the learning phase [27]:

Max Features: The number of features to consider when looking for the best split.

Max Depth: The maximum depth of the tree.

Min Samples Split: The minimum number of samples required to split a non-leaf node.

Min Samples Leaf: The minimum number of samples required to form a leaf node.

Max Leaf Nodes: The maximum number of leaf nodes.

3.4 QoS and QoE in Fixed Broadband Network

3.4.1 Definition of QoS and QoE

QoS is the most widely used service quality metric. It examines network objectively and determines the network quality by measuring a series of criteria such as throughput, delay, jitter, error probability, packet loss, available bandwidth, etc. QoS does not consider user's subjective factors that directly reflect the user perceived degree of services. However, QoE, takes user' opinion into account and directly reflects the acceptance degree of services for users by integrating subjective and objective influence factors [3]. The international Telecommunication Union Telecommunication Standardization Sector (ITU-T) has defined QoE as the "overall acceptability of an application or service, as perceived subjectively by the end user". Different from the above definition, which only considers user's subjective perception as QoE influence factor, many researchers also consider objective human factors to make QoE as a blueprint. It includes both human subjective and objective quality requirements and experiences from the interaction of a human with technology and business entities in a particular situation [3]. Therefore, QoE is considered as a multidimensional construct, including both objective and subjective aspects. In this situation, QoE is related to both objective QoS, which belongs to an objective, technology-oriented concept and User Experience, which is primarily regarded as a subjective, user-oriented concept.

3.4.2 Role of QoS and QoE in Fixed Broadband Service

QoS and QoE are crucial aspects in the provision of fixed broadband services. A well-implemented QoS framework ensures that the network meets certain performance standards, minimizing disruptions and maintaining consistent service delivery. This, in turn, positively influences QoE by creating an environment where users can enjoy high-quality multimedia content, fast internet browsing, and reliable connectivity [28]. Service providers must strike a balance between technical optimizations through QoS and meeting user expectations to achieve a holistic approach that fosters customer satisfaction in the realm of fixed broadband services [28].

3.4.3 QoE Influential Factors

QoE is influenced by various factors typically, Human, System, and Context influential factors that collectively shape the overall satisfaction and perception of users interacting with a service or product [3].

Human IF: This demonstrates characteristics that can describe the socioeconomic and demographic background, the physical and mental organization, and the user's emotional state (e.g., gender, age, educational background) [3].

System IF: Applies to characteristics or properties that determine the produced quality of a service or application. They are related to technical aspect metrics (e.g., bandwidth, jitter, delay, loss, throughput, resolution) [3].

Context IF: Embraces any situational property to describe the user's environment in terms of physical, temporal, social, economic, task, and technical characteristics (e.g., location, duration, frequency of use, cost, brand of the service/system) [3].
The below are the identified influential factors for fixed broadband Internet service input for QoE modeling for the study :

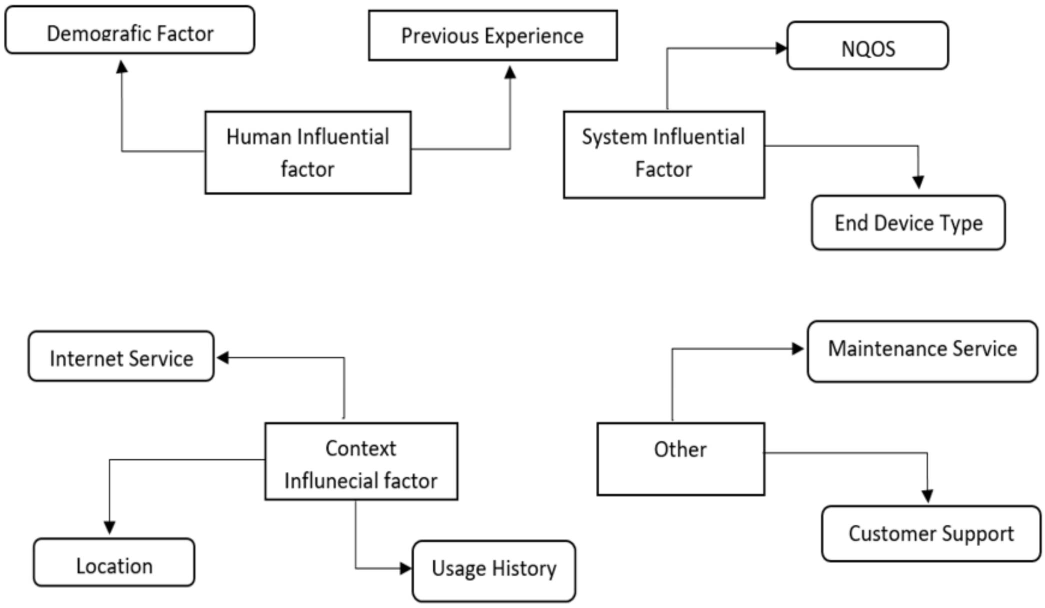

Figure 3.8: Identified QoE influential factor for fixed broadband internet.

3.4.4 QoE Modeling Techniques

Depending on the type of measured parameters, QoE evaluation approaches of service quality can be divided into three categories: subjective assessment, objective assessment, and hybrid assessment [29]. Subjective assessment approach is based on psychological experiments, subjective quality assessment is the most fundamental and reliable but also the most complex and expensive approach to evaluate users' QoE. The approach starts from the user's perception to assess operation. Most of the results from subjective evaluation experiments are quality ratings of users' opinions and ultimately averaged into MOS [29]. For the case of objective assessment in order to reduce large time-consuming and expensive perception tests when measuring the quality of applications or systems, the objective assessment approach is commonly considered as the computing model of QoE. To some extent, it can supply a similar result compared with the subjective assessment of actual users. Objective QoE assessment is a crucial method for evaluating the quality of experience (QoE) of a service or application by measuring quantifiable performance metrics. These metrics are primarily based on the inherent characteristics of the service or application itself, encompassing factors such as bandwidth, latency, jitter, and packet [13]. Hybrid QoE modeling combines objective metrics, subjective assessments, machine learning techniques, network performance monitoring, contextual information, and adaptive streaming algorithms to create a comprehensive and accurate representation of the user experience. This holistic approach is instrumental in understanding, improving, and optimizing services delivered [30].

QoE modeling using machine-learning algorithms is a dynamic and impactful approach to understand and predict user satisfaction across various domains, such as multimedia streaming, network services, and internet applications. In this study, fixed broadband internet is the focus point, the two powerful classification based machine-learning algorithm, support vector machine, and random forest is implemented. Support Vector Machines are well suited for scenarios where the relationship between input features and QoE may be non-linear. SVM work by finding an optimal hyperplane that separates different QoE classes, or predicts continuous satisfaction scores. This makes SVM particularly effective when dealing with complex and high-dimensional data, allowing them to capture intricate patterns and relationships in the dataset.

Random Forest, an ensemble learning algorithm, is another powerful tool for QoE modeling. By constructing multiple decision trees trained on different subsets of the data, Random Forest excels at capturing diverse relationships and mitigating overfitting. This versatility is especially valuable in QoE scenarios where various factors contribute to the overall user experience. Random Forests can handle both linear and non-linear relationships, making them well-suited for mapping the complex interactions affecting QoE.

Chapter 4

Method Deployed for Fixed Broadband Internet QoE Modeling

4.1 Detail System Model

The detailed system model depicted in the below figure illustrates the approach taken to model the quality of experience for fixed broadband internet service. This modeling framework is designed to address both ethio telecom and vISP's namely Websprix IT Solution PLC and ZERGAW Cloud.

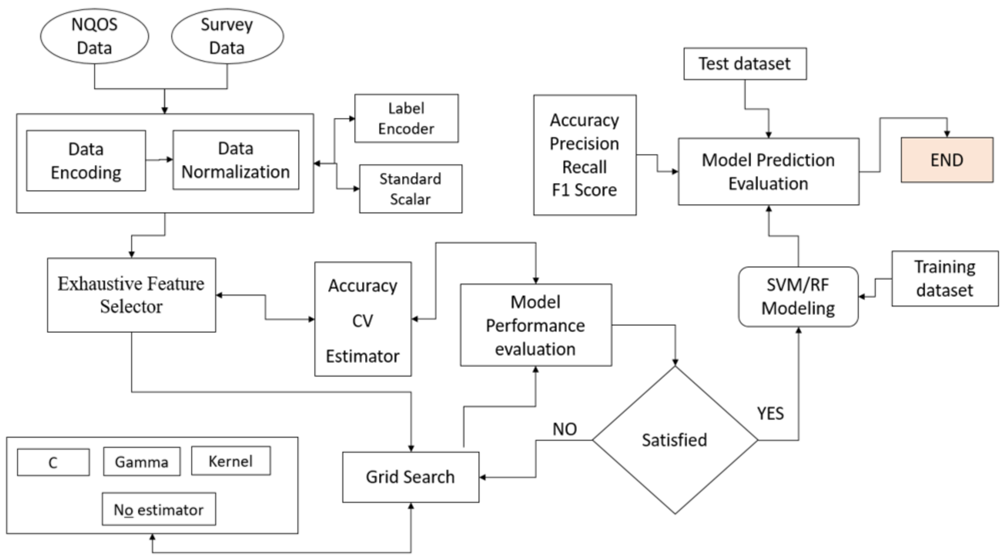

Figure 4.1: Detail System Model.

4.2 Data Collection

Data is collected based on identifying the influential factors that contribute to user experience on fixed broadband internet service.

Identified Influential Factors

The identified influential factors for fixed broadband internet services encompass a range of elements that play a crucial role in shaping the user experience. These factors are categorized into subjective and objective factors, along with other significant considerations.

Subjective Factors:

A. Human influential factors : Includes both user demographic factors and service experience. User demographic factors, such as age, gender, income, education, and occupation, play a pivotal role in shaping internet usage patterns, preferred service types, and overall engagement. For instance, younger users are often more inclined towards mobile-focused activities and favor streaming services, whereas older users may lean towards landline internet connections and news websites. Additionally, service experience holds significant weight in influencing user behavior, with past encounters significantly impacting current actions. Positive experiences foster trust and encourage continued engagement, while negative experiences can result in dissatisfaction and prompt users to seek alternative options.

B. Contextual Influential Factors : Encompass various elements that shape the user experience and satisfaction with internet services. The types of internet services utilized, such as video streaming, web browsing, or VoIP, significantly influence user expectations and contentment. The speed provided by fixed broadband internet services directly correlates with the level of user satisfaction during the utilization of specific internet services. Moreover, the location where users access the internet, whether at home, in the office, or in public hotspots, impacts their usage patterns. For instance, home users may prioritize entertainment and streaming, while office users might focus on communication and productivity tools. Additionally, users' service usage history, reflecting past patterns of internet usage, reveals their preferences and needs. For example, frequent online gaming requires reliable and high-speed connections, whereas casual browsing might be satisfied with basic plans. Understanding these contextual factors is crucial for optimizing internet service offerings and enhancing user experiences.

C. Other Factors Several other factors play significant roles in shaping user experiences with internet services. Service maintenance, whether scheduled or unexpected, can disrupt user activity and lead to frustration. Maintaining transparency and providing effective communication during such periods are crucial to minimizing negative impacts on users. Additionally, the quality and accessibility of customer support greatly influence user satisfaction and loyalty. Efficient problem-solving and responsive channels can build trust and encourage continued engagement with the service. Recognizing and addressing these factors are essential for providing satisfactory internet services and fostering positive user relationships.

Objective Factors: Quality of Service Factors

Upload Speed: The opposite of download speed, it is the speed at which information travels from your internet-connected device to the internet. Upload speed is beneficial for activities such as uploading files, sending emails with attachments, video conferencing, online gaming, and other tasks that require sending data from your device to the internet [31].

Download Speed: Signifies the pace at which data is transmitted from the internet to the device that is currently connected. It characterizes the rate of information transfer, delineating how swiftly content travels through the online network and reaches your specific device. Measured in bits per second (bps), kilobits per second (Kbps), megabits per second (Mbps), or gigabits per second (Gbps), depending on the scale of the transfer [31].

Jitter: The delay variation introduced by the variable transmission of delay of the packets over the network. This can occur because of routers' internal queues behavior in certain circumstances (e.g., flow congestion), routing changes, etc. This parameter can seriously affect the quality of streaming audio and/or video [31].

Delay: Intrinsic to communications, it represents the time information consumes to reach the other side. Also referred to as latency, delay time can be increased if packets face long queues in the network (congestion) or take a less direct route to avoid congestion. The delay can be measured either one-way (total time from the source that sends a packet to the destination that will receive it) or round-trip (one-way latency from source to destination plus the one-way latency from the destination back to the source) [31].

Service Up Time: Refers to the amount of time during which a particular service, system, or application is operational and available for use. Often expressed as a percentage, it represents the reliability and availability of a service over a specific period [31].

- **Packet Loss:** Occurs when one or more packets of data being transported across the internet or a computer network fail to reach their destination. Wireless and IP networks cannot guarantee that packets will be delivered, and some may be dropped if they arrive when their buffers are already full. Packet loss can be caused by factors like signal degradation, high loads on network links, corrupted packets being discarded, or defects in network elements [31].

4.3 Data Collection tool and Sample Size

To gather subjective influential factors from the customer perspective Google form used as a tool for data collection. Simultaneously, the PRTG system is utilized to gather QoS data from ethio telecom customers. Furthermore, for Websprix, the Observium system is employed, while the Ubiquity network management system is utilized for ZERGAW Cloud. The subjective data collection process incorporates the Mean Opinion Score (MOS) methodology, a widely recognized metric for gauging user perception concerning the chosen influential factors. The MOS serves as a valuable measure to capture and quantify the subjective opinions of users, providing essential insights into their perspectives on the factors under consideration. This comprehensive approach to data collection ensures a multifaceted understanding of both objective QoS metrics and subjective user perceptions, contributing to a more holistic evaluation of the overall system performance.

Table 4.1: MOS Scale and Ratings.

MOS Scale	Rating Meaning
5	Very Satisfied
4	Satisfied
3	Neutral
2	Dissatisfied
1	Very Dissatisfied

Determining the optimal sample size for machine learning algorithms like Random Forest and Support Vector Classification (SVC) is a complex task influenced by multiple factors. However, there are practical rules of thumb to provide a starting point. One such guideline suggests aiming for 10-20 data points per feature, a crucial consideration especially for SVC due to its vulnerability to the "curse of dimensionality" when confronted with numerous features and limited data [29]. Utilizing the rule of thumb as a preliminary framework, modeling the Quality of Experience (QoE) for fixed broadband service users entails gathering a reasonable number of samples from both the Internet Service Provider (ISP), Ethio Telecom, and the virtual Internet Service Providers (vISPs) like Websprix and Zergaw Cloud, while also factoring in their respective customer bases. This approach ensures a balanced representation of data from diverse sources, enabling a comprehensive analysis of QoE factors across different service providers and customer demographics.

Table 4.2: Service Provider and Sample Size

Service Provider Name	Sample Size
ethio telecom	313
Websprix IT Solution PLC	297
Zergaw Cloud	50

4.4 Data Preprocessing

The gathered data undergoes a preprocessing stage to prepare it for model training. Given that the subjective data collected entails categorical measurements, a label encoder is employed to convert these categories into numerical values, mapping each categorical variable to a unique integer value. Additionally, the collected data includes measurements that are not normalized, posing a risk of bias to the model. To address this, a standard scalar is applied to normalize the data, ensuring it has a mean of 0 and a unit variance.

The use of a standard scalar is particularly impactful as it not only normalizes the data but also serves as a robust technique for identifying outliers within the dataset, which can significantly influence the model's performance [32]. This comprehensive preprocessing approach aims to enhance the quality and reliability of the data for effective model training.

$$X_{\text{new}} = \frac{X - \mu}{\sigma} \tag{4.1}$$

Where:

X : input data set
μ : Mean value
σ : Standard deviation

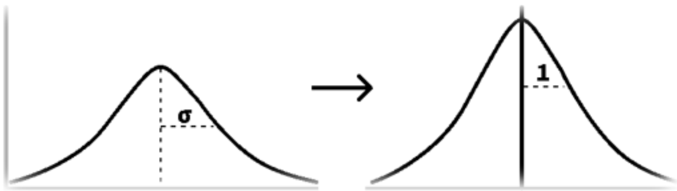

Figure 4.2: Standard Scalar [[32]

In order to selectively identify and isolate the most impactful factors while safeguarding the model from potential noise attributed to the intricacies of the data, an exhaustive feature selector is employed. This method represents a wrapped feature selection technique, characterized by its comprehensive approach. The exhaustive feature selector systematically explores all possible combinations of features and assesses their impact on the model's performance [33]. By evaluating each feature combination exhaustively, this technique aims to discern and select the optimal feature set that contributes to the highest model performance. In essence, the exhaustive feature selector acts as a robust mechanism to streamline the dataset, focusing on the most influential variables and mitigating the influence of extraneous factors that may introduce noise into the model's learning process.

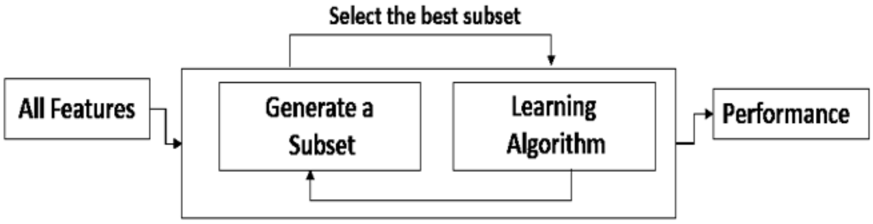

Figure 4.3: Exhaustive feature selection [33].

To enhance the overall performance of the model, a hyperparameter optimization process is implemented through a grid search. This meticulous approach involves systematically exploring a range of hyperparameter values to iteratively assess the model's performance. The grid search method meticulously evaluates the impact of different hyperparameter values and ultimately provides the optimal combination that yields the best model performance. Specifically, for this study, the hyperparameters under consideration include C, Gamma, and Kernel values for the Support Vector Machine, and the number of estimators for the Random Forest machine learning algorithm. This methodical optimization process is designed to fine-tune the model parameters, thereby ensuring an optimized configuration that contributes to superior model performance.

4.5 Model Training and Validation

From the entire dataset, 80% is utilized for training the model, while the remaining 20% is allocated for testing purposes. The training of the model for the Random Forest algorithm involves employing the Bootstrap Aggregating (bagging) method. Following the separate training of the Support Vector Machine and Random Forest models, their performances are evaluated using a cross-validation score, representing the mean score value of each model's performance. To assess the predictive capability of the models, the test dataset is employed. The accuracy, precision, recall, and F1 score are computed by generating a confusion matrix table for the ISP (ethio Telecom) and vISPs (Websprix IT Solution and Zergaw Cloud). A model comparison test is then conducted based on the results to validate the most suitable model.

Accuracy (AC): The proportion of the total number of predictions that were correct [34].

$$AC = \frac{TP + TN}{TP + FP + FN + TN} \tag{4.2}$$

Precision: A useful metric that shows, out of those predicted as positive, how accurate the prediction was [34].

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

Recall: The ratio of correctly predicted outcomes to all predictions. Also known as sensitivity or specificity [34].

$$Recall = \frac{TP}{TP + FN} \tag{4.4}$$

F1 Score: Combines precision and recall into a single score, providing a balance between these two metrics [34].

$$F1Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.5}$$

Where:
TP: True Positive - The actual value was positive and the model predicted a positive value [34]
FP: False Positive - Your prediction is positive, and it is false. (Also known as the Type 1 error) [34]
FN: False Negative - Your prediction is negative, and the result is also false. (Also known as the Type 2 error) [34]
TN: True Negative - The actual value was negative, and the model predicted a negative value [34].

Chapter 5

Results and Discussion

5.1 Collected Data Overview

From the gathered demographic information, the analysis of gender distribution reveals that the proportion of males is greater than that of females among respondents for both internet service providers (ethio telecom) and vISPs (Websprix IT Solution PLC and ZERGAW Cloud). The specific distribution rates are illustrated in the figure below.

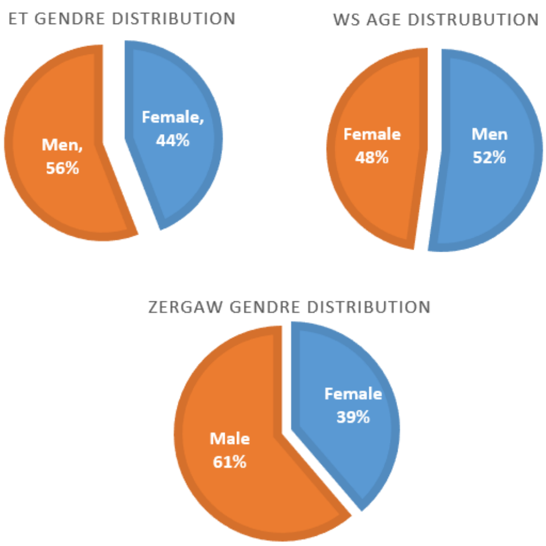

Figure 5.1: Gender distribution of respondents.

The figure below illustrates the breakdown of age and educational background according to the gender of the respondents. Based on the data collected from ethio telecom, we observe that among male respondents, the age group of 25-30 holds the highest proportion, with 50 individuals, predominantly possessing a degree-level educational background. Similarly, within the female respondent category, the age distribution of 25-30 also commands the largest share, comprising 31 individuals with a degree-level education. Notably, a minimal number of respondents are aged above 46 in both gender categories.

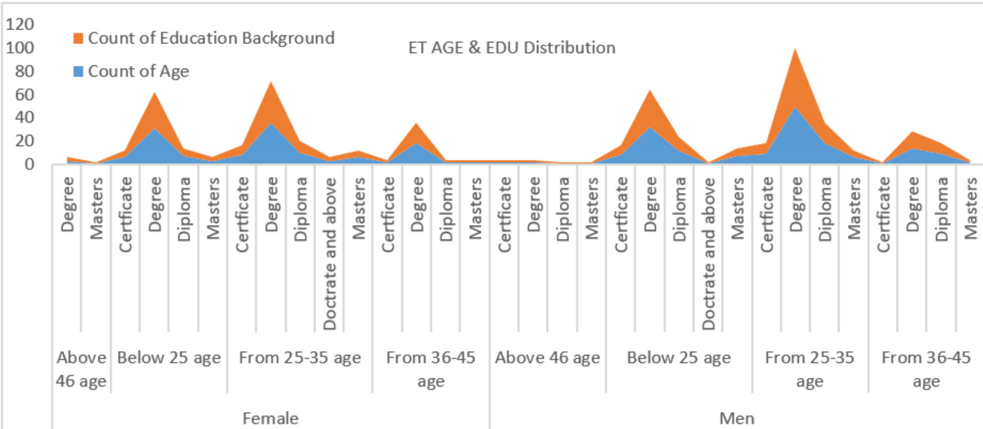

Figure 5.2: Age and Education distribution for ethio telecom fixed broadband internet users.

The data collected from Websprix IT Solution PLC and ZERGAW Cloud Fixed broadband internet users reveals that the 25-35 age group dominates among male respondents. Specifically, 39 individuals with a degree-level education constitute this category for Websprix customers, compared to 13 for ZERGAW Cloud customers. Similarly, among female respondents aged 25-35, Websprix customers have the highest proportion with 29 individuals holding a degree, while ZERGAW Cloud customers indicate a value of 6 for this group. This data highlights the prevalence of individuals with a degree in the specified age range among both male and female respondents for these two entities.

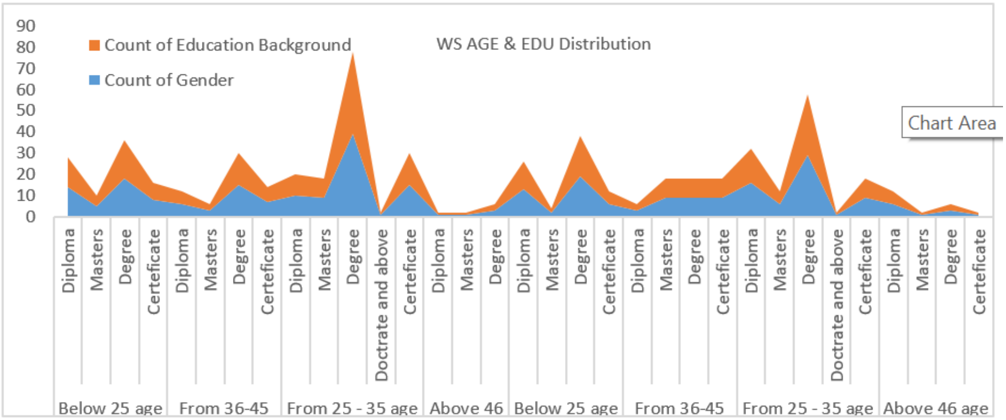

Figure 5.3: Age and Education distribution for Websprix fixed broadband internet users.

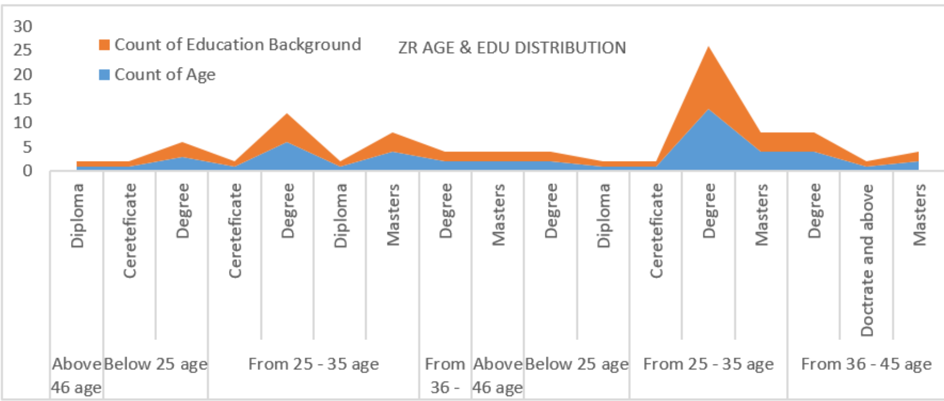

Figure 5.4: Age and Education distribution for ZERGAW Cloud fixed broadband internet users.

5.2 Data After Preprocessing

The data obtained from both internet service providers (ethio telecom) and virtual ISPs (Webprix IT Solution PLC and ZERGAW Cloud) for fixed broadband internet users undergoes preprocessing, starting with a label encoder. This encoder is employed to convert categorical values, including age, education background, service usage history, service experience, location used, and end device type, into numerical format for analysis. Subsequently, a standard scaler technique is applied to normalize the measured data, ensuring that it has a mean of 0 and unit variance. This normalization is crucial to prevent bias in the model due to variations in the scales of measured values. The below figures depict the data after label encoding and standard scaling for respondents from both ISP and virtual ISPs.

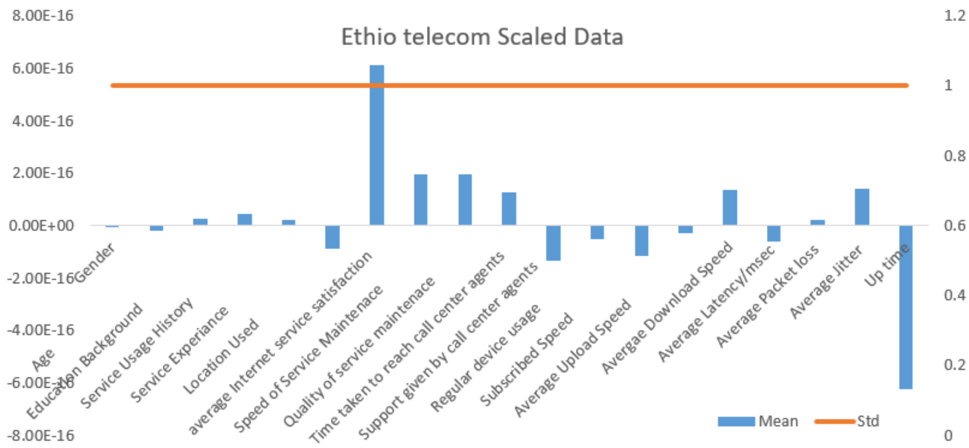

Figure 5.5: Scaled data from ethio telecom fixed broadband internet users.

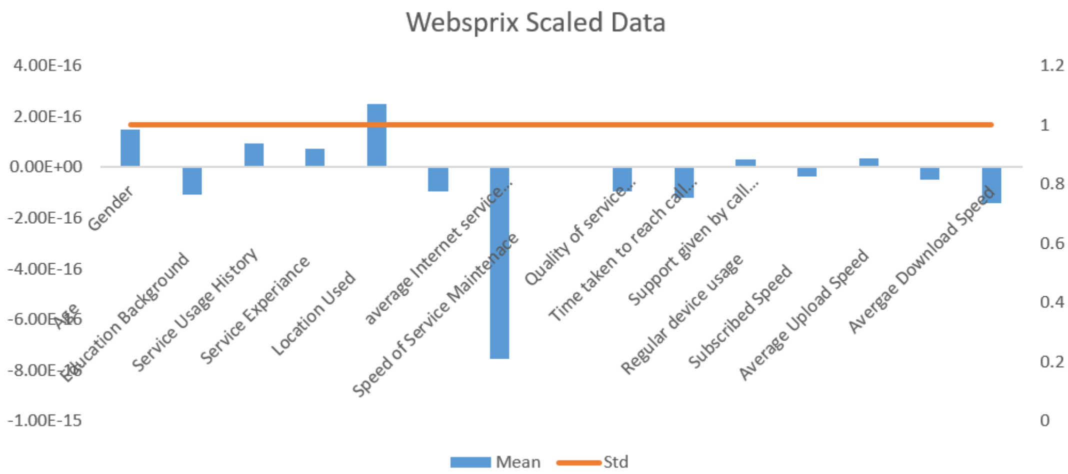

Figure 5.6: Scaled data from Websprix fixed broadband internet users.

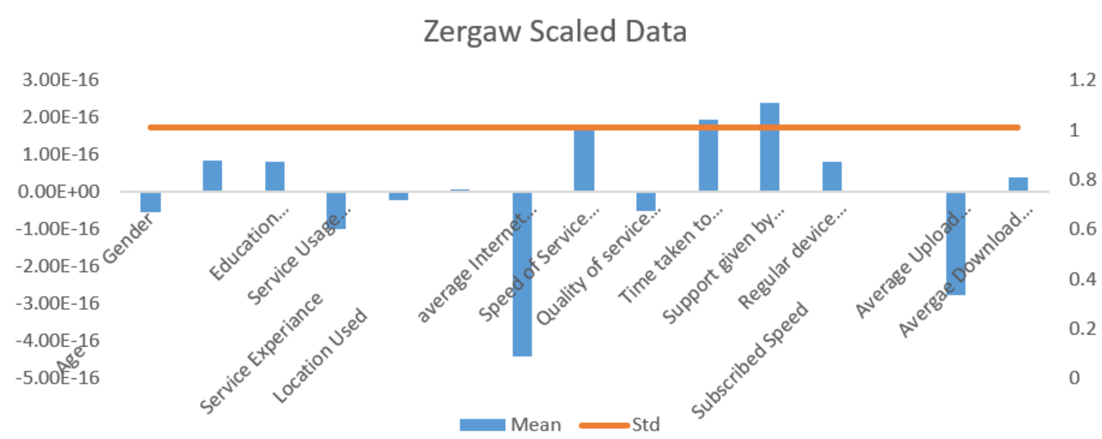

Figure 5.7: Scaled data from ZERGAW Cloud fixed broadband internet users.

5.3 Important Feature Selection

From the subjective data gathered from both ISP and vISPs, it is evident that numerous factors play a significant role. In order to streamline the model complexity and mitigate any potential bias, an Exhaustive Feature selection approach has been implemented. This aims to identify the primary influential factors for both machine learning algorithms, namely SVM and Random Forest. The ensuing graph illustrates the chosen influential factors derived from the twelve collected features for both Random Forest and Support Vector Machine. This method of feature selection contributes to a more refined model, ensuring a more accurate representation of the essential elements impacting the user experience from the subject matter.

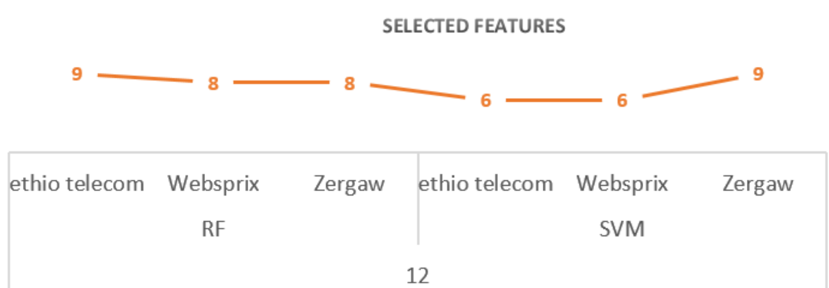

Figure 5.8: Selected influential factors from subjective data.

The below figure shows, the selected features coefficient, that reflects their impact on the user service satisfaction rate:

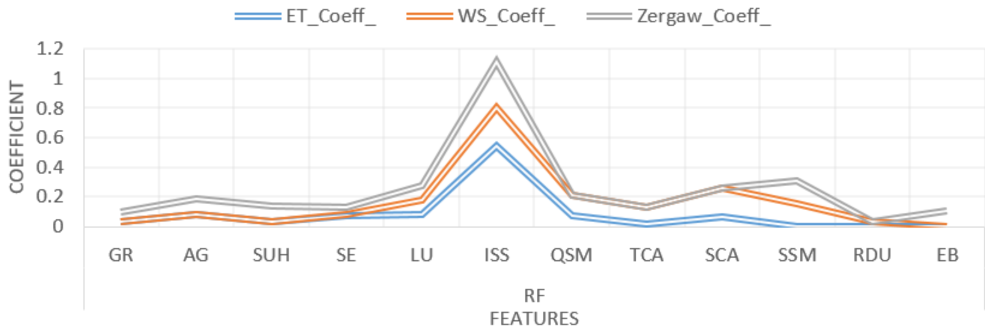

Figure 5.9: Selected features coefficient for Random Forest.

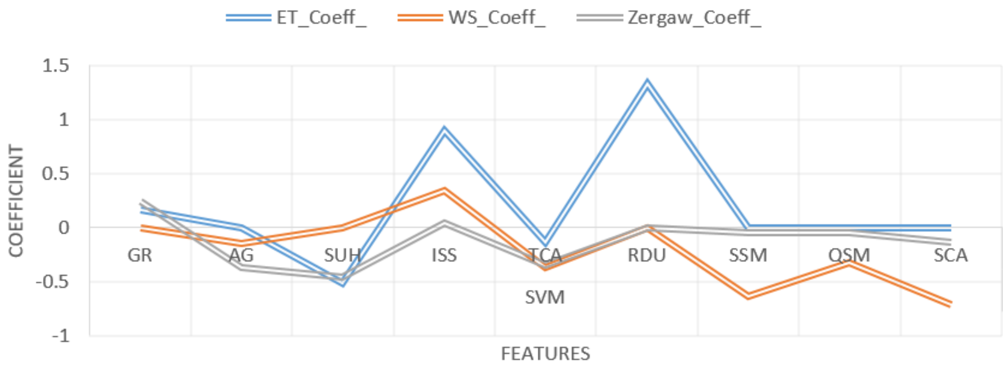

Figure 5.10: Selected features coefficient for SVM.

5.4 Hyper Parameter Optimization

The grid search method is implemented to determine the most effective hyperparameter values for the chosen machine algorithms, specifically the random forest and support vector machine. In weighing the balance between computational expense and model accuracy, only grid search is utilized for the designated hyperparameters and their corresponding range values.

Support Vector Machine Initialized Hyperparameters:

- Kernel type: ['linear', 'poly', 'rbf', 'sigmoid']
- C: [0.1, 1, 10, 100, 1000]
- Gamma: [1, 0.1, 0.01, 0.001, 0.0001, 'scale']

Random Forest Hyperparameter Initialization:

- n_estimators: [20, 50, 100], and use the default value for other hyperparameters (max_depth, min_samples_split, and others) considering the high computational cost.

Grid search result for Support Vector Machine using subjective data :

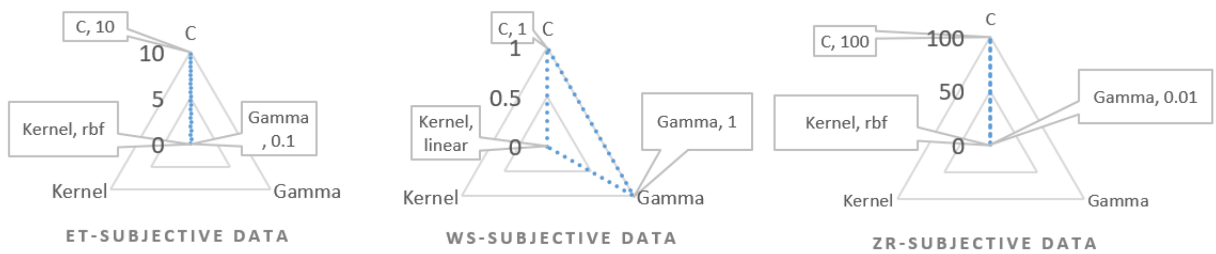

Figure 5.11: Selected hyperparameters for SVM using subjective data.

Grid search result for Support Vector Machine using QoS data :

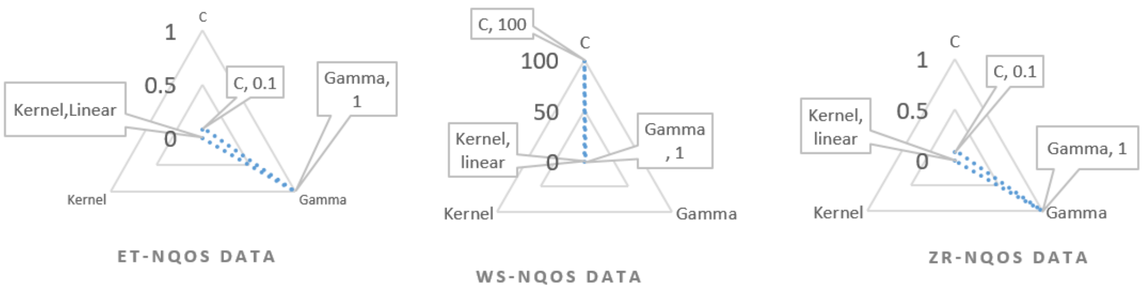

Figure 5.12: Selected hyperparameters for SVM using QoS data.

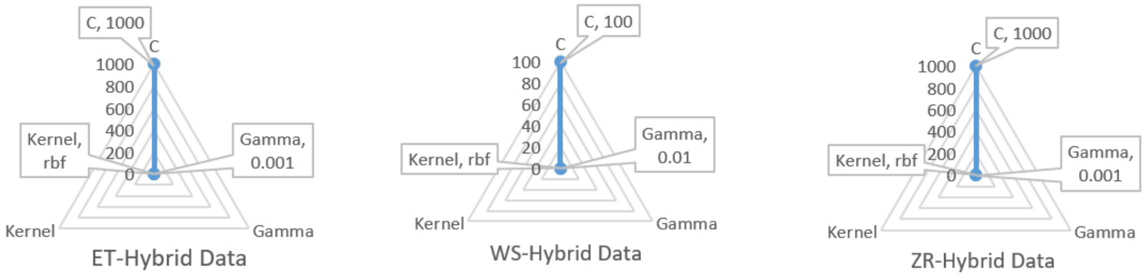

Figure 5.13: Selected hyperparameters for SVM using Hybrid data

For Random forest, the grid search selected, 100 number of estimator for both subjective and NQoS data sets.

5.5 Model Training

The support vector machine and random forest get learned with the selected influential factors and optimal hyper parameters values using k fold cross validation techniques. K-fold cross-validation is a technique used in machine learning to assess the performance and generalizability of a model. The basic principle involves dividing the dataset into K subsets or "folds," training the model on K-1 of these folds, and then evaluating the model on the remaining fold [35]. This process is repeated K times, with each of the K folds used exactly once as the validation data. The results from each iteration are then averaged to obtain a more robust performance metric [35].

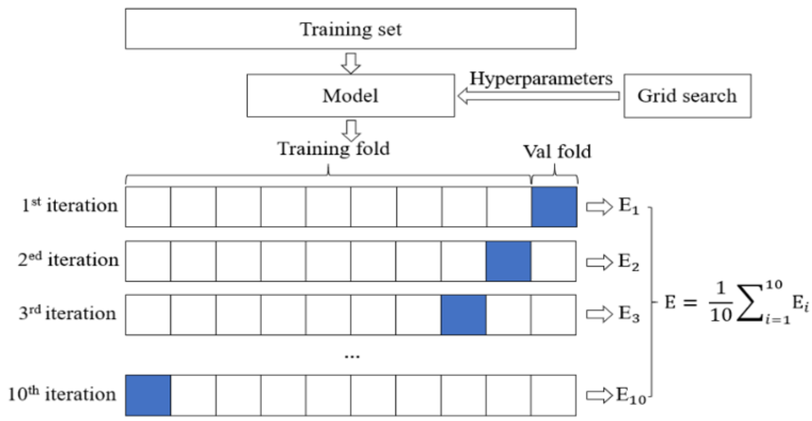

Figure 5.14: K fold cross validation model training [35]

For this study three k fold are used to train the model's considering the computational cost. The below figure shows how support vector machine try to learn from the training dataset for hybrid QoE modeling.

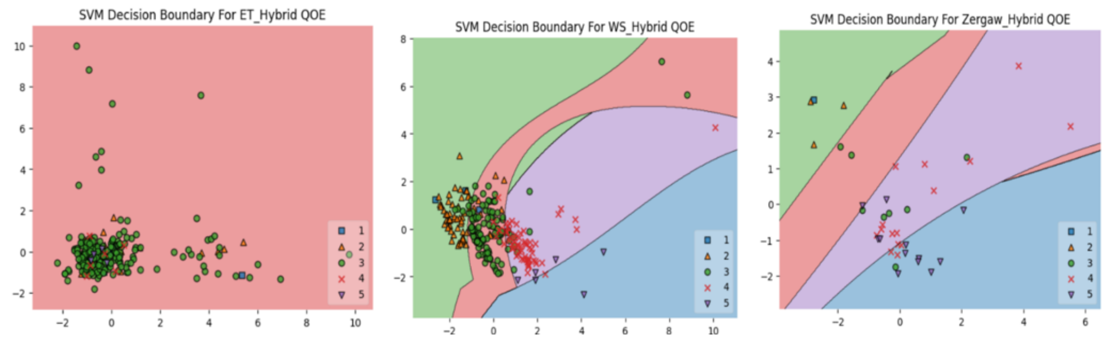

Figure 5.15: Support Vector Machine decision boundary

The depicted illustration illustrates the process by which a support vector machine seeks the optimal hyperplane for classifying a given problem. In the case of the ethio telecom dataset, the data exhibits an imbalance, posing challenges in visualizing the decision boundary in 2D plot. Contrastingly, for the vISP datasets (Websprix and ZERGAW Cloud), the support vector machine adeptly visualizes and plots the decision boundary, in reference to the Mean Opinion Score (MOS) values.

5.6 Model Performance Comparison and Selection

The assessment of model performance has been conducted for both Support Vector Machine and Random Forest, encompassing evaluations for both the ISP(ethio telecom) and vISPs(Websprix and ZERGAW Cloud). To enhance comprehension, the model performance analyses were carried out individually using subjective data, QoS data, and a composite of the two (hybrid QoE). The evaluations were performed employing the K-fold cross-validation technique. The following figure illustrates comparable model performance for ethio telecom data with model accuracy of 88% for both Support Vector Machine and Random Forest in the realm of subjective QoE modeling. In contrast, when considering objective QoE, the Support Vector Machine outperforms with a superior model accuracy of 80%. Remarkably, for Hybrid QoE modeling, the Random Forest exhibits improved model performance accuracy, achieving a value of 88%.

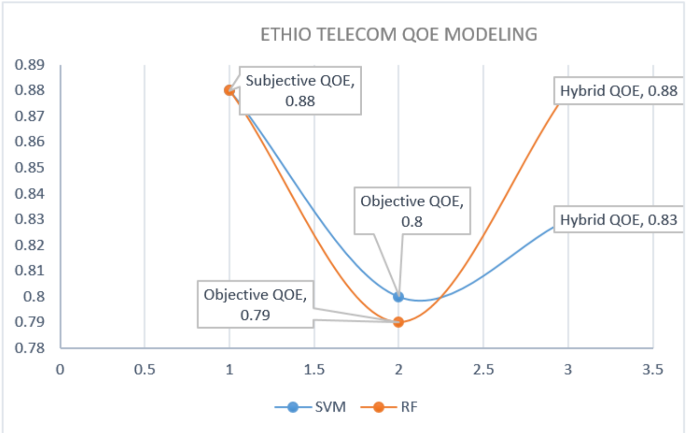

Figure 5.16: Model performance evaluation for ethio telecom data.

In the context of Websprix QoE modeling, the Support Vector Machine outperforms the Random Forest model in subjective, objective, and hybrid QoE scenarios, achieving accuracy scores of 79%, 82%, and 92%, respectively. Conversely, in the case of ZERGAW Cloud QoE modeling, the Support Vector Machine demonstrates superior performance for both subjective and objective QoE, with accuracy scores of 69% and 43%, respectively. Notably, even in the hybrid QoE scenario for ZERGAW Cloud, SVM continues to exhibit better performance, with an average accuracy score of 70%.

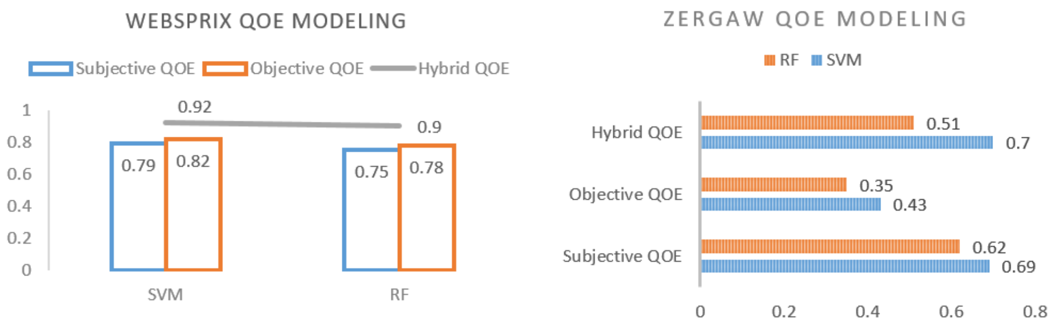

Figure 5.17: Model performance evaluation for vISP's

When considering QoE modeling collectively for both the ISP's(ethio telecom) and vISPs (Websprix and ZERGAW Cloud), hybrid QoE modeling consistently yields superior model performance for both Support Vector Machine and Random Forest, compared to modeling QoE solely based on subjective and QoS metrics. Moreover, opting for hybrid QoE modeling provides a comprehensive overview of customer experience, serving as valuable input for future service optimization efforts to manage and enhance user experiences effectively.

The evaluation of model performance through k-fold cross-validation also indicates that the collected data is adequate for training the chosen model, resulting in a commendable accuracy score for both the ISP (ethio telecom) and vISPs (Websprix). However, in the case of Zergaw ISP, the model's accuracy appears relatively lower compared to ethio telecom and Websprix. This discrepancy can be attributed to the smaller dataset available for training, stemming from a smaller active customer base utilizing the fixed broadband internet service offered by Zergaw ISP.

5.7 Model Prediction Performance Evaluation

To evaluate the model performance of Support Vector Machine and Random Forest, a confusion matrix is implemented to get the Accuracy,Precison,Recall and F1 score value. A confusion matrix is a table that is commonly used to evaluate the performance of a classification algorithm. It provides a comprehensive summary of the model's predictions by comparing them to the actual outcomes. The matrix is particularly useful for tasks with multiple classes (categories).

Confusion matrix for ethio telecom subjective and objective QoE:

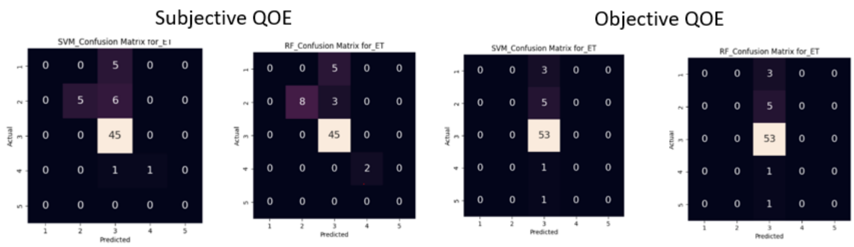

Figure 5.18: Confusion matrix for ethio telecom

Confusion matrix for Websprix subjective and objective QoE :

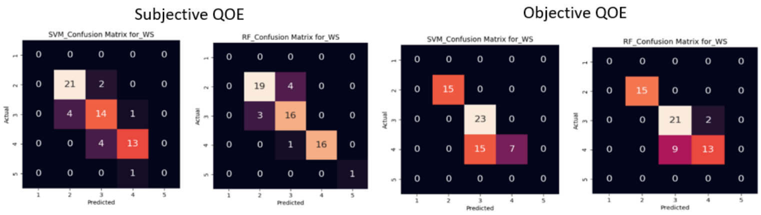

Figure 5.19: Confusion matrix for Websprix QoE

Confusion matrix for ZERGAW Cloud subjective and objective QoE

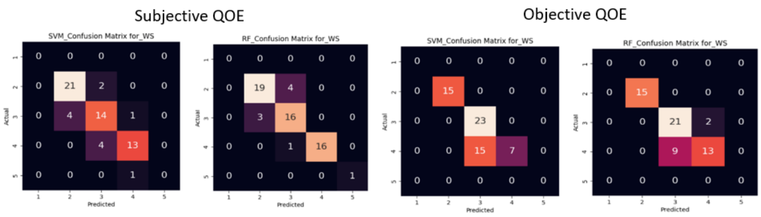

Figure 5.20: Confusion matrix for ZERGAW Cloud QoE

Confusion matrix for hybrid QoE both for ethio telecom and vISP's

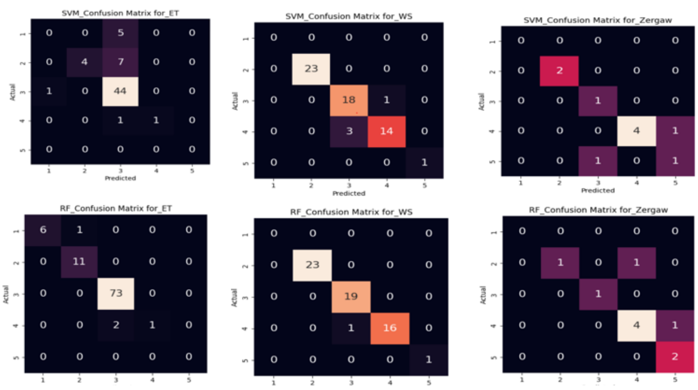

Figure 5.21: Confusion matrix for Hybrid QoE both for ISP and vISP's

The data provided in the below charts depicts the computed accuracy, precision, recall, and F1 score. These metrics were derived from the confusion matrix, and the evaluation was conducted separately for each of the subjective, objective, and hybrid models used in QoE modeling. Both the Support Vector Machine and Random Forest models underwent assessment based on these performance metrics to provide a comprehensive understanding of their effectiveness in each modeling scenario.

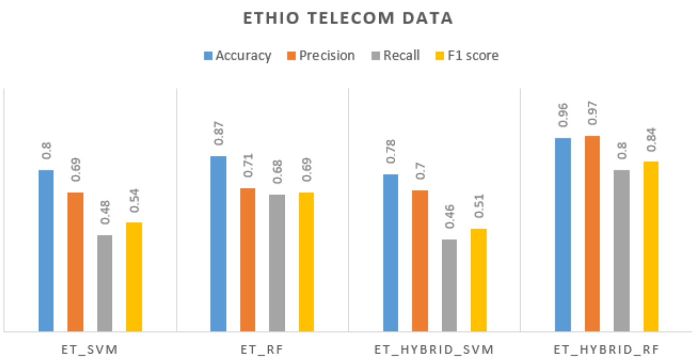

Figure 5.22: Model prediction performance of subjective Vs hybrid QoE for ethio telecom.

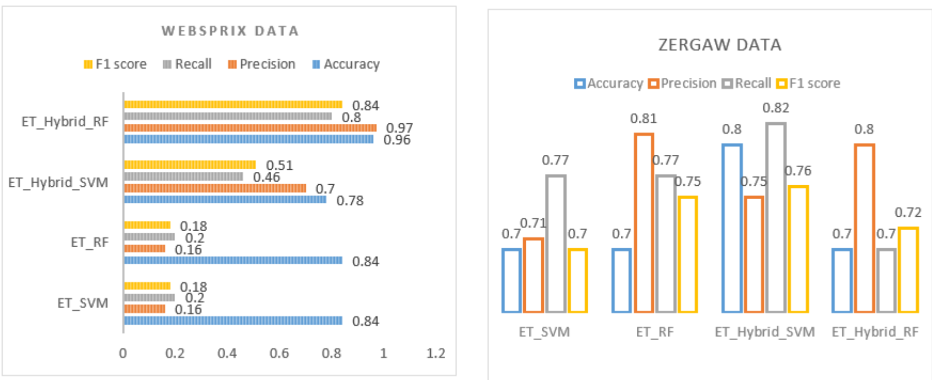

Figure 5.23: Model prediction performance of subjective Vs hybrid QoE for vISP's

The result presented above reveal that employing a random forest machine learning algorithm for Hybrid (QoE modeling among ethio telecom fixed broadband internet users yields improved model prediction results subjective QoE modeling. The accuracy, precision, recall, and F1 score stand at 96%, 97%, 80%, and 84%, respectively. Similarly, in the case of Websprix fixed broadband internet users, the random forest model for Hybrid QoE demonstrates superior predictive capabilities, achieving accuracy (97%), precision (97%), recall (97%), and an impressive F1 score of 96%. Contrastingly, for ZERGAW Cloud fixed broadband internet users, the Support Vector Machine stands out in Hybrid QoE modeling, shows superior predictive performance compared to other modeling techniques. The accuracy, precision, recall, and F1 score values 80%, 75%, 82%, and 76% respectively.

The figures below illustrate a comparative analysis of model prediction performance between Hybrid QoE and Objective QoE, employing both random forest and support vector machine algorithms. Once again, the results affirm that for ethio telecom and Websprix fixed broadband internet users, the random forest model for Hybrid QoE surpasses the performance of Objective QoE. Conversely, for ZERGAW Cloud fixed broadband internet users, the support vector machine's Hybrid QoE modeling yields superior predictive outcomes.

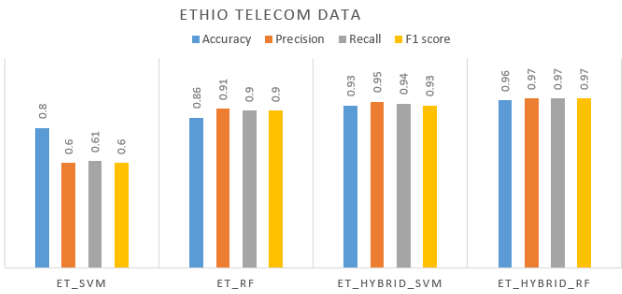

Figure 5.24: Model prediction performance of objective Vs hybrid QoE for ethio telecom

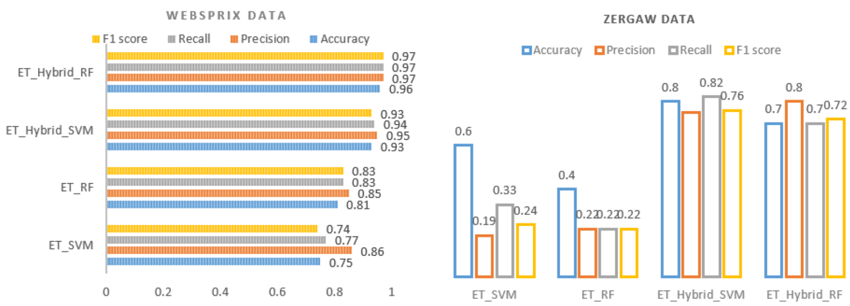

Figure 5.25: Model prediction performance of objective Vs hybrid QoE for vISP's.

The utilization of a hybrid QoE model, combining both subjective and objective data, offers distinct advantages in enhancing the performance and prediction accuracy of SVM and Random Forest models. Relying solely on subjective or objective data has inherent limitations, and the hybrid approach aims to capitalize on the strengths of both types of information. By considering user opinions and feedback alongside technical parameters and network statistics, the hybrid model provides a more comprehensive insight into the user experience. This holistic approach addresses the difficulties of subjectivity and objectivity, mitigating the limitations of each and yielding a more balanced evaluation. The inclusion of both types of data enhances the model's robustness, adaptability to diverse scenarios, and overall predictive accuracy. The synergy between subjective and objective data leads to a higher level of predictive accuracy, with subjective insights capturing user preferences and experiences, and objective metrics providing technical details. Moreover, the hybrid model has the potential to reduce bias by considering multiple perspectives, avoiding the pitfalls of over-reliance on a single type of information. In summary, the adoption of a hybrid QoE model for SVM and Random Forest models represents a nuanced and comprehensive approach, contributing to a more informed, adaptable, and accurate prediction of user experience.

Chapter 6

Conclusion and Recommendations

6.1 Conclusion

The study employed data from ethio telecom and virtual ISPs (Websprix IT Solution PLC and ZERGAW Cloud) for fixed broadband Internet users, The preprocessing of data involved the utilization of a label encoder to convert categorical values into numerical format, followed by standard scaling to normalize the measured data. This normalization was crucial to prevent bias in the subsequent machine learning model due to variations in the scales of measured values. The subjective data gathered from both ISPs and vISPs revealed the significance of various factors in influencing user experience. To streamline model complexity and mitigate potential bias, an Exhaustive Feature Selection approach was employed. This approach aimed to identify the primary influential factors for two machine learning algorithms, namely SVM and Random Forest. Grid search determined optimal hyperparameter values, and K-fold cross-validation ensured robust learning with selected factors. Model performance assessments, using subjective, objective, and hybrid QoE metrics, indicated superior accuracy (88%) for both SVM and Random Forest in subjective QoE. Random Forest outperformed in hybrid QoE (88%), with SVM excelling in objective QoE (80%).

The evaluation of model performance for Support Vector Machine and Random Forest was conducted on ISP(ethio telecom) and vISPs(Websprix and ZERGAW Cloud). Ethio telecom showed comparable model accuracy of 88% for both models in Subjective QoE, while in Objective QoE, Support Vector Machine outperformed with 80% accuracy. In Hybrid QoE modeling, Random Forest excelled with an accuracy of 88%. For Websprix, Support Vector Machine demonstrated superior performance across subjective, objective, and hybrid QoE scenarios with accuracy values of 79%, 82%, and 92%, respectively. In contrast, for ZERGAW Cloud, Support Vector Machine exhibited better performance for subjective and objective QoE (69% and 43%, respectively), but Random Forest outperformed in hybrid QoE with an average accuracy of 70%. Adopting a hybrid QoE model, combining subjective and objective data, enhanced SVM and Random Forest prediction accuracy. The approach mitigated limitations of individual data types, providing a comprehensive understanding of user experience. The synergistic use of subjective insights and technical metrics improved model robustness and adaptability. The hybrid model's potential to reduce bias and offer a balanced evaluation underscores its value in predicting user experience accurately. In summary, adopting a hybrid QoE model for SVM and Random Forest represents a nuanced and comprehensive approach, contributing to a more informed, adaptable, and accurate prediction of user experience.

6.2 Recommendations

Derived from the analysis conducted in this thesis, the subsequent recommendations are proposed for future research endeavors :

- Delve into the specific characteristics of fixed broadband technologies, such as Fiber and Copper, during the data analysis. Understanding how different technologies impact user experience can provide valuable insights for both service optimization and infrastructure planning.

- Enhance the robustness of the machine learning models by leveraging a larger dataset. Increasing the volume of data used for training and evaluation can lead to more accurate predictions and a better representation of user behaviors and preferences.

- Extend the model comparison by incorporating additional machine learning algorithms. A diverse set of models, beyond SVM and Random Forest, can offer a more nuanced evaluation, providing insights into the strengths and weaknesses of various approaches for predicting user experience.

- Enhance the model's performance by conducting an extensive exploration of hyperparameter optimization. Broaden the search space to include a wide range of hyperparameter values for the chosen machine learning algorithms. This exhaustive approach aims to identify the most effective combination of hyperparameters, further improving the accuracy and generalizability of the models

- Broaden the scope of the study by including a diverse set of vISPs. This inclusion can capture a wider range of user experiences, ensuring that the developed models are adaptable and effective across different service providers and user demographics.

Appendix A

Mean K fold cross validation result

The below result shows average performance of the models through three-fold cross-validation for both Support Vector Machine and Random Forest, employing the Subjective, Objective, and Hybrid QoE modeling approaches.

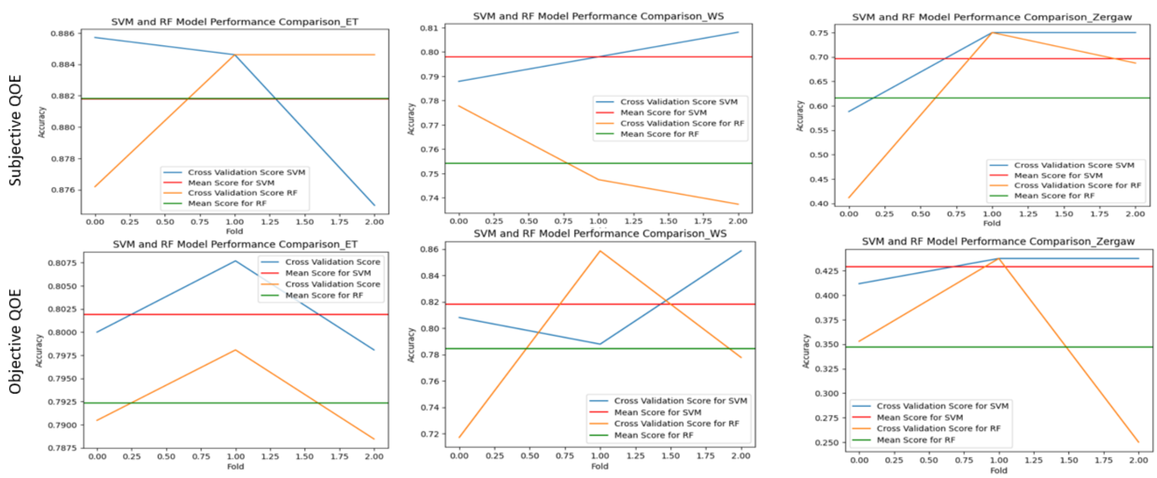

Figure A.1: Average model performance comparison for subjective and objective QoE

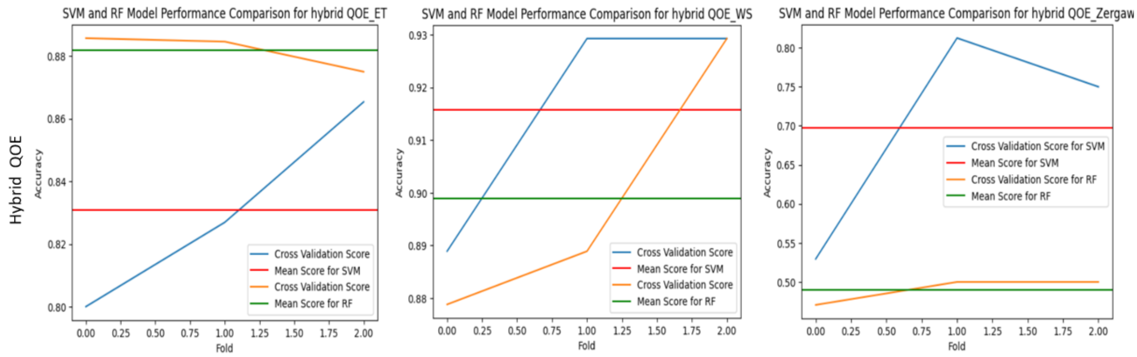

Figure A.2: Average model performance comparison for subjective and objective QoE

Appendix B

Survey Questionnaire

The survey is done to investigate the customer's perception and satisfaction level regarding the fixed broadband internet provided to its customers. The survey was designed to identify the influencing factors that affect the level of customer satisfaction with the service. Thus, your honest and prompt response is a valuable resource for the superior and effective success of the study.

We would like to assure you that all the information you provide will be kept confidential and will only be used for research purposes.

Thank you for taking your time to fill out this questionnaire.

Personal Information

1. Gendre :

   O Male O Female

2. Age :

   O Below 25 O 25-35 O 36-45 O Above 45

3. Education level :

   O Certificate O Diploma O Degree O Masters

   O Doctorate and Above

Service Related Information

4. Location used the fixed broadband internet ? : ——————

5. Service number or User ID ? : ——————

6. Subscribed Bandwidth (Mbps) ? : ————

7. For how long used the fixed broadband service ?

O Below 1 Year O From 1 to 2 years O From 2 to 3 years
O Above 3 years

8.Did you have previous service experience from different service operator ?

O Yes O No

Internet service related questions

9. Did you use video service ?

O Yes O No

If yes,

9.1 How much satisfied with the quality of the video ?

O Very Dissatisfied O Dissatisfied O Neutral O Satisfied O Very Satisfied

9.2 How much satisfied with the time taken to open the video ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

9.3 How satisfied are you with the availability of video content ?

O Very Dissatisfied O Dissatisfied O Neutral

O Satisfied O Very Satisfied

9.4 How satisfied are you with the frequency of cuts/reloads while watching video ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

10. Did you use web browsing service ?

O Yes O No

If yes,

10.1 How much satisfied with the time taken to open the web link

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

10.2 How much satisfied with the time taken to open web content ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

10.3 How satisfied are you with the interaction with the web page ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

11. Did you use voice over internet service ?

O Yes O No

If yes,

11.1 How much satisfied with the quality of voice ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

11.2 How much satisfied with the sound delay ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

11.3 How satisfied are you with access to calls? ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

12. Did you use chatting service ?

O Yes O No

If yes,

12.1 How satisfied are you with the time it takes to start the chat service ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

12.2 How satisfied are you with the time it takes to open multimedia content ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

12.3 How satisfied are you with the time it takes to deliver multimedia content ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

12.4 How satisfied are you with your interaction on the chat service ?

Service Maintenance Related Questions

13. How satisfied are you with the speed of maintenance service ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

14. How satisfied are you with the quality of maintenance service ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

Customer Support Related Questions

15. How satisfied are you with the timetakes to reach call center call center agents ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

16. How satisfied are you with the support received from the call center agents ?

O Very Dissatisfied O Dissatisfied O Neutral
O Satisfied O Very Satisfied

Bibliography

[1] H. S. Na, J. Hwang, and H. Kim, "Digital content as a fast Internet diffusion factor: Focusing on the fixed broadband Internet," *Information Development*, vol. 36, no. 1, pp. 97–111, 2020.

[2] A. Ababa, "St. mary's university school of post graduate studies," no. May, 2021.

[3] M. Yang, S. Wang, R. N. Calheiros, and F. Yang, "Survey on QoE assessment approach for network service," *IEEE Access*, vol. 6, no. c, pp. 48374–48390, 2018.

[4] G. Kougioumtzidis, V. Poulkov, Z. Zaharis, and P. Lazaridis, "Machine Learning for QoE Management in Future Wireless Networks," *2021 34th General Assembly and Scientific Symposium of the International Union of Radio Science, URSI GASS 2021*, no. August, 2021.

[5] W. Tadesse and A. Ababa, "QoE Model for Social Media Video Streaming Service Using Ensemble Method - The Case of Addis Ababa," 2022.

[6] H. O. Hamidou, J. P. Kouraogo, O. Sie, and D. Tapsoba, "Machine learning based Quality of Experience (QoE) Prediction Approach in Enterprise Multimedia Networks," 2023.

[7] A. S. Fite, "QoE Assessment Model for Addis Ababa LTE Video Streaming Service Using Machine Learning Techniques," no. February, 2020.

[8] A. A. Laghari, H. He, M. Shafiq, and A. Khan, "Impact of storage of mobile on quality of experience (QoE) at user level accessing cloud," *2017 9th IEEE International Conference on Communication Software and Networks, ICCSN 2017*, vol. 2017-Janua, no. October 2018, pp. 1402–1409, 2017.

[9] T. Submitted, "A Thesis Submitted to the School of Electrical and Computer Engineering Addis Ababa Institute of Technology," 2020.

[10] Technology & Information Policy Institute (TIPI), "Broadband-What is it?," pp. 1–8, 2022.

[11] S. Aarnikoivu and J. Winter, "Mobile broadband wireless access," *Innovation in Telecommunications*, no. January 2006, p. 65, 2006.

[12] R. A. Saeed, Z. M. Yusof, and A. B. Musa, "Cross-layer design and performance analysis of TDMA-based backhaul network," *BROADBANDCOM 2009 - Selected Papers on Broadband Communication, Information Technology and Biomedical Applications*, no. July 2009, pp. 1–7, 2009.

[13] C. B. Ameur, "Streaming Chiheb Ben Ameur TCP Protocol Optimization for HTTP Adaptive Streaming," no. April, 2016.

[14] M. Bausha and V. Riabov, "Intro to Networking: Broadband Connection, DSL vs. Cable Modem," p. 22, 2001.

[15] A. HECHT, "What Is a Virtual GPU?," 2018.

[16] N. J. Nilsson, "INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK Department of Computer Science," *Machine Learning*, vol. 56, no. 2, pp. 387–99, 2005.

[17] J. Peng, E. C. Jury, P. Dönnes, and C. Ciurtin, "Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges," *Frontiers in Pharmacology*, vol. 12, no. September, pp. 1–18, 2021.

[18] H. H. H. Mahmoud and T. Ismail, "A Review of Machine learning Use-Cases in Telecommunication Industry in the 5G Era," *16th International Computer Engineering Conference, ICENCO 2020*, no. December 2020, pp. 159–163, 2020.

[19] O. Omankwu and C. Chidimma, "Machine Learning Types and Techniques AND," vol. 11, no. 12, pp. 56–58, 2021.

[20] B. Schölkopf, "An Introduction to Support Vector Machines," *Recent Advances and Trends in Nonparametric Statistics*, pp. 3–17, 2003.

[21] S. Busuttil, "Support Vector Machines," no. 1.

[22] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," *School of EECS, Washington State University*, pp. 1–13, 2011.

[23] D. Prasanna and W. R. Don, "Digital Commons @ Georgia Southern Multiclass Classification Using Support Vector Machines," 2018.

[24] S. Liu, C. Joe-Wong, J. Chen, C. G. Brinton, C. W. Tan, and L. Zheng, "Economic Viability of a Virtual ISP," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 902–916, 2020.

[25] A. Cutler, "Random Forest for Regression and Classification," *Books*, vol. 1, no. 1, p. 21, 2010.

[26] S. I. Dimitriadis and D. Liparas, "How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: From Alzheimer's disease neuroimaging initiative (ADNI) database," *Neural Regeneration Research*, vol. 13, no. 6, pp. 962–970, 2018.

[27] R. Nygren and A. Petkov, "Evaluation of Hyperparameter Optimization Methods for Random Forest Classifiers," *Kth Royal Institute of Technology School of Electrical Engineering and Computer Science*, pp. 1–21, 2021.

[28] E. Kokkinos and G. Liodakis, "" Quality of Experience (QoE) for Real Time Broadband Services "," no. September, 2018.

[29] T. Hossfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming," *2014 6th International Workshop on Quality of Multimedia Experience, QoMEX 2014*, no. December 2015, pp. 111–116, 2014.

[30] L. Zhang, "Network Capacity, Coverage Estimation and Frequency Planning of 3GPP Long Term Evolution," *Diva-Portal.Org*, 2010.

[31] F. C. D. Gouveia and T. Magedanz, "QUALITY OF SERVICE IN TELECOMMUNICATION M ES PL C E O – M ES PL C E O –," vol. II.

[32] L. B. de Amorim, G. D. Cavalcanti, and R. M. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, pp. 1–37, 2023.

[33] K. Bouzoubaa, Y. Taher, and B. Nsiri, "Dos attack forecasting: A comparative study on wrapper feature selection," *2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020*, no. June, 2020.

[34] K. M. Ting, "Confusion Matrix," *Encyclopedia of Machine Learning and Data Mining*, no. October, pp. 260–260, 2017.

[35] P. Guo, W. Meng, M. Xu, V. C. Li, and Y. Bao, "Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning," *Materials*, vol. 14, no. 12, 2021.