QoE Model for Social Media Video Streaming Service Using Ensemble Method - The Case of Addis Ababa

BY: WUBALEM TADESSE

ADVISOR: EPHREM TESHALE (PhD)

A Thesis submitted to
School of Electrical and Computer Engineering
Addis Ababa Institute of Technology

in Partial Fulfillment of the Requirements for the Degree of Master of Science
(Telecommunication Engineering)

Addis Ababa University
Addis Ababa, Ethiopia
January 31, 2022

Declaration

I, the undersigned, declare that the thesis comprises my own work in compliance with internationally accepted practices; I have fully acknowledged and referred all materials used in this thesis work.

Wubalem Tadesse

Name

Signature

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

This is to certify that the thesis prepared by **Wubalem Tadesse**, entitled *QoE Model for Social Media Video Streaming Service Using Ensemble Method - The Case of Addis Ababa* and submitted in partial fulfillment of the requirements for the degree of Master of Science (Telecommunication Engineering) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner 1	____________	Signature	________	Date	______
Examiner 2	____________	Signature	________	Date	______
Advisor	Ephrem Teshale (PhD)	Signature	________	Date	______

____

Dean, School of Electrical and Computer Engineering

ABSTRACT

Nowadays, the number of Social Media (SM) users is increasing tremendously worldwide. An increase in the number of smartphone users and an increase in Internet coverage helps people expand their networks. The availability of SM helps users to find their friends, make a connections with new people with different skills, and improve their careers. The diversity of services on SM also attracts many new users to use SM in their day-to-day activities.

To deliver SM services, many SM application developers continuously work to improve their services and also to add new services to attract new customers and also to keep their users.

There are many stakeholders that are involved in end-to-end service delivery of SM. These include Telecom network providers, SM application owners, and end-user devices performance. Having a good network Quality of Service (QoS) may not guarantee good service quality on the customers' side. The quality of a given service perceived by an end-user, which is Quality of Experience (QoE) is a broad term and influenced by many Influencing Factors (IF).

There are many research papers done on the QoE model of different SMs. Most of the papers focus on the impact of network and application related QoS on the overall customer satisfaction level. Even though these papers incorporate different parameters as input features for the QoE model, to the best of the author's investigation, researches which are done in the context of Ethiopia didn't consider the users device parameters influence on the customers' satisfaction level.

The over all QoE of a service is influenced by many factors, the main focus of this thesis is to provide a QoE model for SM video streaming services by taking different IF as input parameters. From the network QoS parameter download speed, upload speed, latency, and jitter are used as inputs. From users' device parameters phone Random

Access Memory (RAM) size, phone's free internal storage size, and phone's screen resolution are taken as device IF. And from application parameters, video resolution is used as an input for the model.

The developed model is based on an ensemble technique which is a Machine Learning (ML) based approach. The model has good accuracy, which is 94.1% accuracy. In addition to the accuracy, based on the importance of each input feature to the final model, download speed takes the main influencing share by 52.357% from the total input parameters and from the users' mobile device parameter free internal storage space has 10.784% and mobile RAM size 9.4% on the final QoE model.

Generally, this work meets the initial objective by developing QoE model with a good accuracy, and shows the influence of other parameter other than the usual network QoS parameters and gives insight to the gaps that cause the customers' dissatisfaction of the SM video services by considering different IF.

Keywords: QoE, QoS, Social Media, Ensemble Method, Machine Learning.

ACKNOWLEDGMENTS

First, I want to express my deepest gratitude to the ALMIGHTY GOD who has given me the strength, power, wisdom and time to do this thesis.

I am grateful to all my instructors who had given us all the necessary courses in the MSc. Program, which had laid the foundations to develop the mindset, the knowledge and wisdom necessary to do this thesis.

My sincere gratitude goes out to my advisor Ephrem Teshale (PhD)who has aided me in the development of this thesis from scratch to finish. Moreover, I am grateful for his time, support and constructive comments up to the submission of my thesis.

My sincere gratitude goes out to my Examiners Dr.-Ing. Dereje Hailemariam and Dr. Yalemzewd Negash who have sharpened my critical thinking, given me precious ideas, hints, corrected me when necessary and have given me deeper insights. My deepest gratitude goes out to the Ethio telecom, who had faith in me and sponsored my study.

Finally, yet importantly, I would like to express my heartfelt gratitude to all my family members specially Woinitu Hailu and Lijalem Fetene for their unconditional love, encouragement, and assistance from start to finish. At last, my special acknowledgment goes to my lovely daughter Hemen Lijalem for being my strength to keep going when I feel like giving up!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

3G	Third Generation
4G	Fourth Generation
ACQUA	Application for PrediCting QUality of experience at Interne Access
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artifcial Neural Network
AUC	Area Under the Curve
CART	Classification And Regression Trees
CN	Core Network
DBW	Downlink Bandwidth
E2E	End to End
FN	False Negative
FP	False Posotive
FPR	False Positive Rate
FR	Full Reference
IF	Influencing Factors
IP	Internet Protocol
IPTV	Internet Protocol TV
ISP	Internet Service Provider
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KNN	K-Nearest Neighbor
KPI	Key Performance Indicator
LTE	Long Term Evolution
ML	Machine Learning
MOS	Mean Opinion Score
NR	No Reference
OTT	Over The Top
PCC	Pearson Correlation Coefficient
PPI	Pixel Per Inch
QoE	Quality of Experiance
QoS	Quality of Service
RAM	Random Access Memory
RAN	Radio Access Network
RF	Random Forest

ROC Reciever Operator Caracterstic
RR Reduced Reference
SM Social Media
SVM Support Vector Machine
TN True Negative
TP True Positive
TPR True Positive Rate
UE User Equipment
UMTS Universal Mobile Telecommunications System
VoD Video on Demand
Wi-Fi Wireless Fidelity

1 INTRODUCTION

1.1 BACKGROUND

Many papers gives different definition for SM, the book written by Varinder and Kanwar [1] defines SM as "All web based applications which allow for creation or exchange of user-generated content and enable interaction among the users can be classified as Social Media". It can be seen that SMs are designed to be Internet-based to provide different services for users. Usually, SM users from different geographical locations with different social, religious, and educational backgrounds are connected to pursue mutual interests via web-based applications on their desktops, laptops, smartphones, or tablets SM users. Delivery of all web-based systems to the customers depends on the telecom service provided, which means SM services directly rely on the available telecom network.

The services types delivered by SM are increasing from time to time. Some of the services include photo sharing, video sharing, blogs, products/services sharing, social networks, social gaming, and news sharing. In addition to the type of services, the number of SM users show a significant increase in the past few years. Based on Global Digital Overview statistical data[2] there are 4.48 billion SM users around the world as of July 2021 and there is an annual increment of 13.1%. In the case of Ethiopia, there are 6.7 million SM users as of January 2021, with 8.1% annual increments [3].

Telecom service providers provide network service to deliver Internet services to their customers. In addition, to provide the infrastructures the telecom operators examine their services whether they are delivered with the expected quality or not. There are different Key Performance Indicators (KPI) to measure the QoS for different services.

QoS is technical concept that is expressed and understood in terms of networks and network elements. The overall QoS from the users' perspective cannot be determined only by considering QoS, there are many IFs that determine the customers' satisfaction. Based on ITU-T Rec. G.100 / P.10 definition, the overall acceptability of an application or service, as perceived subjectively by the end-user is described by QoE. QoE is measured in terms of Mean Opinion Score MOS [4], which represents the user's opinion about a service using a scale from 1 to 5, (1 for bad, 5 for excellent).

User expectations about the service they use is influenced by many technical and non-technical factors. The technical system comprises a chain of components (user device, Internet network, and SM application server) where as non-technical factors include expectations, personal background, demography, urgency, task, gender, age, etc. This shows QoE assessment is not an easy task.

Out of many services provided by telecom operators SM services rely on the Internet provided by the telecom operator. As mentioned above the number of SM users increase from time to time, which indicates there will be more new customers that need the Internet connection. Ethio telecom is performing expansion works like deploying LTE and advanced LTE in the main cities of the country. Beyond infrastructure expansion, QoS measurement and assessment, Ethio telecom needs a QoE model for assessing, measuring, and predicting QoE for SM services.There are different researches which are done in this area. This research extends the previous works by incorporating different IFs which were not considered in previous works and develop a model for SM video streaming service. Thus this will give crucial feedback that would be used to improve the quality of SM service to improve customer satisfaction.

1.2 MOTIVATION

The total population of Ethiopia by world population review as of 2021 is around 115 million [3], from this population 44.86 million peoples have mobile connections and 23.96 million have Internet access [5]. When considering the current SM users' statistical data, the total active SM users are 6.7 million people, out of this 96.2% of the total users use their mobile phones to access SM in [6].

Out of the total population of the country 56.55% of the population has an age range from 15 to 64 and 39.92% from 0 to 14 [7]. From this data, the age group from $15 - 65$ is considered as an active working age and most SM users are considered to be under this age group. When looking at the age group under 15, which also constitute the large share of the population are expected to join the SM users in near future.

From this statistical data, there are many SM users by now, and there will be a large growth in SM users in near future. Ethio telecom as the only current network provider is working on many expansion and optimization works to provide the best network quality for this growing SM demand. But having a good network may not guarantee good service quality. There are many factors to be considered for the overall service quality. As an example when looking at the SM users' device choice, mobile smartphones are placed in the first place. This shows when thinking to work on SM service quality, one input to be considered should be the effect of the mobile device. There are also many other factors, and the effect of these factors on the SM service quality should be studied in detail. Ethio telecom, as a business company should work on SM customer satisfaction to maintain the existing customers and also to attract new potential customers. As far as my knowledge is concerned there is no QoE model for SM video streaming services that comprise the effect of users' devices in addition to other network-related metrics. This motivates me to model SM video streaming services QoE that is capable of assessing and predicting SM QoE.

1.3 STATEMENT OF THE PROBLEM

When looking the SM delivery chain, there are many parties that influence the overall QoS provided. Figure 1.1 demonstrates an End-to-End (E2E) chain of the SM delivery system. The end user's background like social status, demography, educational level, SM usage history and soon have a great impact on the customers' satisfaction level. The other factor that can influence the SM satisfaction level is different end-users

Equipment (UE) parameters. The third one that influences the satisfaction level of the SM users', which is under full control of the telecom operator is the network part. This part consists of the Radio Access Network(RAN) and Core Network (CN) part including the transmission network, which has a direct impact on the QoS delivered. The other factor is the SM application server parameters which also affect the overall service satisfaction level.

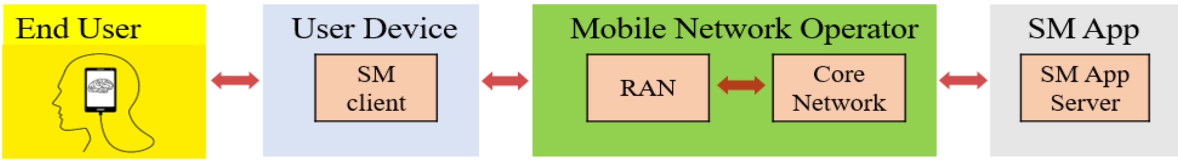


Figure 1.1: Social Media chain

Since there are many parties on the (E2E) chain of SM, delivering a good network QoS will not assure a service with the best satisfaction level. Therefore, other factors besides the network QoS part that have an impact on the overall service quality need to be assessed. This help to pinpoint the possible factors that influence the QoE and to work on possible solution beside guaranteeing QoS for an E2E delivery of quality service.

In the current Ethiopian context, in near future there will be a competitive environment between the current telecom operator Ethio telecom and the new telecom company Safaricom Ethiopia. This indicates Ethio telecom has to do a lot of work to keep its current customers and also attract new potential customers to keep its profitability. To do so Ethio telecom should assess its customer satisfaction level on each service the company provides. Pinpointing the satisfaction level will help to fill the possible gaps that cause complaints by the customers.

1.4 OBJECTIVE

1.4.1 *General objective*

The general objective of this thesis work is to develop QoE model for SM video streaming service using an ensemble method.

1.4.2 *Specific objectives*

The specific objectives of this thesis work are:

- Review different literature that focus on QoE modeling.
- Study the impact of network QoS parameters on QoE of SM video streaming services.
- Study the impact of device parameters on QoE of SM video streaming services.
- Measure the SM customers' satisfaction by conducting a survey.
- Model QoE by taking network, device, and application parameters as an IF using ML base ensemble method.

1.5 METHODOLOGY

The method used in this thesis aimed to provide a good QoE model with reasonable accuracy. To do so the first step was the data collection step by surveying selected SM users. The collected data consist of different metrics that influence the overall QoE of SM video streaming services. From network QoS download speed, upload speed, latency, and jitter are collected using nPerf which is a free mobile application for network QoS test. The other metrics are from users' mobile phone specifications which include mobile RAM size, screen resolution (number of pixels per inch), and the available internal storage size. From the application, the video resolution is taken as one metric.

After data gathering, the next step was data pre-processing which include data cleaning, mismatch handling, and data preparation for modeling. Then the selected ML base technique (ensemble method) is applied to train the model with the selected data set. After model training the last step will be to check the model validity and propose the final model. Figure 1.2 shows the method used in this thesis work.

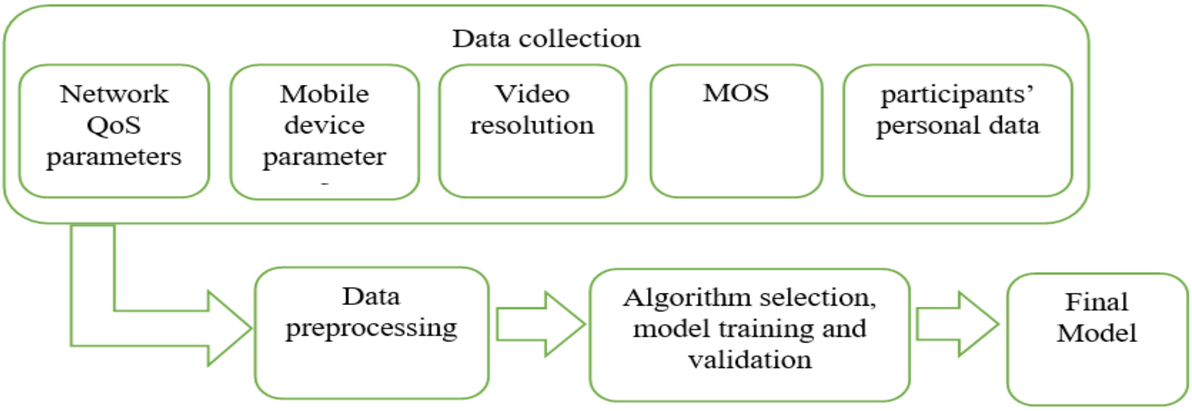


Figure 1.2: Research Method Used

1.6 SELECTED LITERATURE REVIEW

There are many papers about QoS and QoE of different service types on different network architecture. The approaches followed by these papers also vary based on the method deployed for analyzing QoE. The paper by W.Digis [8] examines the QoE of video streaming services specifically YouTube download streaming QoE. ML algorithms (Artificial Neural Network, k-Nearest Neighbors, and Random Forest) are used for developing QoE estimation models using QoS features. For QoS data gathering, the author use ACQUA which is a free Android tool that measures the user-level network traffic like signal strength, download and upload bandwidth, round-trip time, jitter and etc. A total of 230 participants are included in the data gathering process and their satisfaction level is collected by MOS values. The collected data is applied to different ML algorithms to model QoE. Out of which Random Forest perfectly fit the QoE prediction. This work takes only network QoS as the only IF of the model.

The other paper by A.Menbere [9] models the QoE of UMTS voice service using the Adaptive Neuro-Fuzzy Inference Approach. The data which is 3G voice network KPI is gathered from the Network Management System. For the subjective information, the author provide questionnaires' to evaluate the subjective perception of quality. Using these two data as an input to Adaptive Neuro-Fuzzy Inference system which is ML technique a new QoE model is developed for the 3G and 2G voice calls services.

The users service quality perception is determined by many factors. Beside the QoE models that incorporate network parameters as the main input, the paper by Reichl et al. [10] tries to fill the gap for by integration of user behavior characteristics and user context factors, as well as the consideration of appropriate temporal scales. The authors discuss a set of use cases, like traditional QoE, user behavior, charging and pricing models, the impact of user characteristics and problems related to energy consumption. They separate the technical perspective of the system from the user perspectives and the context related behavioral perspective relevant for the service provider.

The other approach for QoE is assess the users feedback by providing network emulator instead of the actual network. Casas et al. in [11] perform a survey to collect feedback from a different participant which uses YouTube and Facebook. Based on MOS values at different download speeds. Instead of the actual network, they use a network emulator to control and vary the bandwidth between $64kps$ and $4,096kps$ and capture all the packets before and after the emulator for analysis of YouTube videos. They use the YoMo free mobile tool for monitoring application layer parameters like the number and duration of video stall and they gather feedback from the participants on different download speeds. From their analysis, they show YouTube QoE is highly sensitive to down-link bottlenecks and video bit-rate and DBW has a strong correlation with the end-users' satisfaction and acceptability. Network emulation is also applied to measure QoE of different applications like Facebook, YouTube, Google maps, web browsing, and Whatsapp . The other paper [12] also repeated the QoE measurements in terms of MOS on the field and shows how customers satisfaction is affected by different network parameter values. This paper also includes the acceptance rate for different MOS values.

Beside modeling QoE by taking different IFs as an input parameters Gomez et al. In [13] perform a survey by making participants watch videos on Wi-Fi and UMTS network, in the meantime the network QoS parameters are measured. Then they perform a mapping from the network QoS to Application QoS (initial buffer time, rebuffing frequency, and mean buffer time) then from Application QoS to in terms of MOS.

The other papers focus on different QoE IF besides network parameters. Laghari et al in [14] shows how QoE is affected by the internal storage free space when users access cloud services from their mobile phones. When end users are accessing cloud services enormous amount of temporary/cache data is generated by apps, so internal storage of mobile devices is filled quickly. The mobile device without any space in internal storage has a huge impact on the performance when accessing the cloud services, which degrade the QoE of end-users for particular cloud apps and services.

The other paper by Buberwa and Mbise [15] proposed a QoE model by mapping combined effects from both network and device parameters on video streaming services. From the network QoS, they took packet loss, packet reordering, and delay were emulated using network emulator. Through analysis of variance, they found that packet loss had the highest impact, followed by device features, reordering, and delay on the video QoE , and from the combined effects, two-way interactions and three-way interactions had significant effects on the video QoE .

Different ML algorithms are used in many papers to model QoE [16] by Casas et al. performs QoE prediction for smartphone applications using different ML algorithms. They incorporate single models (Decision tree, Naïve Bayes, Neural network, Support vector Nearest neighbor) and ensemble-based (random forest, bagging, boosting, and stacking) methods. Based on their results ensemble methods shows the highest accuracy in prediction and on the overall classification performance. The other paper by Wassermann et al in [17] conduct survey analysis on large scale worldwide ,they suggest that QoE -based monitoring of YouTube mobile can be realized through ML models with high accuracy, relying only on network related features and without accessing any higher-layer metric to perform the estimations. They compare different ML-based models and they suggest random forest predicts QoE with a better performance. They also show the advantages of ensemble-learning techniques with respect to simpler models in terms of prediction accuracy.

From the literature review, it is seen that QoE modeling is influenced by many factors including the network QoS, device performance, and users' emotional behaviors. Most of the models incorporate the influence of the network QoS parameters and only a few papers try to study the impact of different users' device performance parameters on the users' satisfaction. So, in this research paper, the effect of user device parameters in addition to the network parameters are considered in the QoE model.

1.7 SCOPE AND LIMITATION

The scope of this thesis is to provide a QoE model for SM video streaming service, specifically Facebook, YouTube, and Tiktok users. Out of many QoE IF, this paper's concern is limited to Network QoS parameters (download speed, upload speed, jitter and latency), mobile phone Parameters (Phone RAM size, phone screen resolution and mobile phone internal storage free space) and from application parameter video resolution. The study only considers Ethio telecom customers which use their smartphones for accessing SM services in Addis Ababa Ethiopia. It does not include SM users on other devices like tablets, desktops, and laptops.

1.8 CONTRIBUTIONS

This thesis work is intended to contribute the following point.

- To assess SM video streaming service satisfaction level.
- Understand different QoE influencing parameters besides the commonly known network QoS parameters.

- To evaluate how aware, the customers are on the impact of their smartphone performance on their SM they use by themselves.
- Finally, this thesis aims to propose a QoE model with good accuracy for SM video streaming service using an ensemble method for Ethio telecom SM video streaming service.

1.9 THESIS ORGANIZATION

This thesis is organized in the following structure. Chapter one covers an introductory part on what SM are, what is their trend in the type of service and number of users through time and what factors influence the QoE of SM services. Then what is the motivation that forces me to work on this idea, what are the general and specific objectives and the methodologies followed to develop the final QoE model. Then other works which are done by other authors on the related topics are reviewed, finally, the scope, limitation, and papers contribution is included. In Chapter two the basic concepts of QoE, QoS, and other SM video streaming service QoE IF are discussed briefly. Chapter three covers the basic ideas behind an ensemble method which is used in this paper for modeling QoE like the techniques used ,different hyper parameters and model evaluation methods. Chapter four deals with the deployed system model and chapter five discuses the obtained results and interpretation that has been done. Finally the conclusion, recommendation and future work based on the result obtained is covered in Chapter six.

2

SOCIAL MEDIA SERVICES QOS AND QOE OVERVIEW

2.1 INTRODUCTION

SM has become an integral part of people's daily routines. There are many activities going on the SM including written posts, blogs, articles, Books, posting images videos, reviewing news, advertising services, chatting with friends, and soon. Getting all such services by SMs make them to be the first thing most people use after waking up. In addition to the type of services, the number of users is increasing from time to time. According to Digital 2021 report [2], there are 4.2 billion SM users by the start of 2021. On average, more than 1.3 million new users joined SM every day during 2020. Figure 2.1 shows the trend in the growth of SM users. Globally, more than 5.19 billion people now use mobile phones out of these more than 4.5 billion people use the internet. From this global trend, it is seen that mobile phones and internet access are crucial for SM usage.

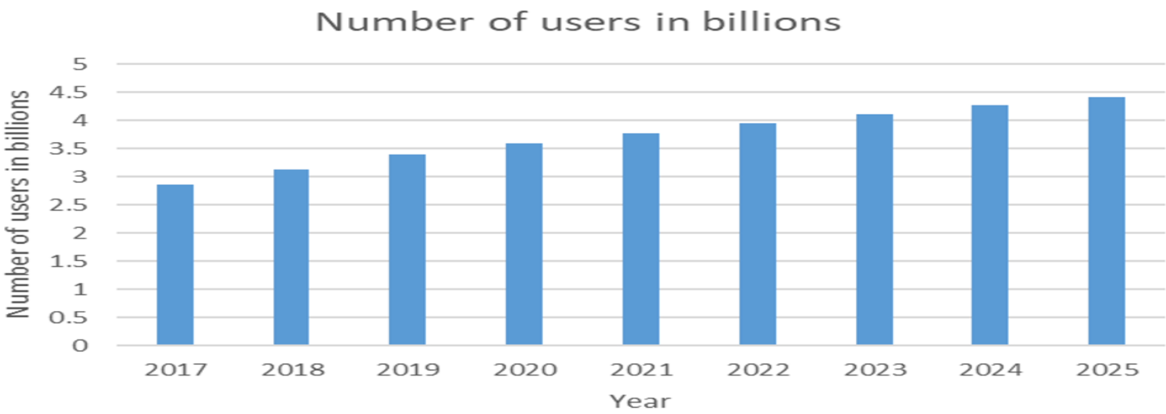


Figure 2.1: Social Media Users' Growth Trend Adopted from [2]

The service which is delivered by SM should satisfy the users on different context. When we see an E2E SM delivery chain,there are many parties that contribute to the overall service quality. For a long time, the QoS is defined based on network centric QoS parameters defined by ITU (ITU-T Rec. X.200) like download speed, upload speed, latency, and packet loss. Even though there is a visible relationship between network based QoS and the service quality observed by a user, meeting these and other sets of QoS criteria might not guarantee end-user satisfaction. The user's perspective about the service should be considered to achieve an E2E service quality.

To incorporate the users' centric perspective, a concept of QoE was introduced. ITU defines both QoS and QoEas follow, QoS is defined as: "Collective effect of service performance which determines the degree of satisfaction of a user of the service" Based on ITU-T Rec.E.800 [18] and Quality of Experience by ITU-T SG 12 in 2007 [19] is defined as: "The overall acceptability of an application or service, as perceived subjectively by the end-user. This includes the complete E2E system effects (client, terminal, network, services infrastructure, etc.) and by user expectations and context." From the above QoE definition user's perception about the service they use, like SM can be influenced by characteristic of a user, system, service, application and context.

2.2 QOE INFLUENCING FACTORS

The overall acceptability of a SM application perceived by the end users can be influenced by many factors. These factors can be categorized in to human-related IF, system-related IF, context-related IF and content related IF [20].

- Human-related IF: These are related to the characteristic of a human that influence the perceived quality. Which includes demographic and economic background, physical, mental and emotional status of user's emotional state.
- System-related IF: These are technological variables that affect the QoS, such as transmission, coding, storage, rendering, and reproduction/display, as well as information exchange from content creation to the user.

- Context-related IF: These are factors that embrace any situation property to describe the user's environment, in terms of, temporal, economic, task, and technical characteristics.
- Content-related IF: These are the information regarding the offered content by the service or application under study.

2.3 QOE ASSESSMENT METHODS

In the context of video streaming applications which is the focus of this research, QoS is not enough to assess the service quality. Therefore, different strategies are followed to study an E2E QoE assessment. These are subjective, objective and the hybrid assessments. "Subjective methods are conducted to obtain information on the quality of multimedia services using opinion scores, while objective methods are used to estimate the network performance using models that approximate the results of subjective quality evaluation" [21]

2.3.1 *Subjective Assessment*

Subjective evaluation is thought to be the most accurate method of determining the end user's perceived QoS. This method involves gathering of human opinion and assigning grades based on their point of view and perception to analyze service quality. The values obtained for each test sequence is based on five-point discrete MOS [1: bad, 2: poor, 3: fair, 4: good and 5: excellent] [4]. This assessment method is very expensive in terms of human resource, cost and time consumption.

2.3.2 *Objective Assessment*

A large variety of objective quality measurements have been established based on QoS metrics (parameters collected from the network) by employing different mathematical formulas or algorithms to estimate the QoE. One drawback of using this technique is it does not incorporate users' perception as an input for the QoE assessment. For the video services, the objective quality assessment models are categorized as full reference (FR), reduced reference (RR) and no reference (NR) models, depending on the availability of a reference (unprocessed) video signal for the assessment [20].

- Full Reference: The FR method accesses the original and received content in the evaluation process.
- Reduced Reference: Only some features of the reference signal are extracted and employed to evaluate the quality of the distorted signal.
- No Reference: Approach assesses the content quality level without any knowledge of the original material.

2.3.3 *Hybrid Assessment*

Another approach to assess QoE is to use combination of subjective and objective assessment, referred to as the hybrid method. It could be employed in real time, and it is categorized as the most accurate approach since it decreases the weaknesses of previous two approaches [20].

2.4 QOE MODELING APPROACHES

Currently using only QoS parameters is regarded as insufficient for accurately describing a service quality. Because QoS only considers technical components of a service, it excludes any human-related quality-affecting factors and does not reflect actual service quality.

The main focus of this paper which is QoE, is what the end-user experiences while using the service. However, when talking about measuring or predicting the QoE there are many different approaches that have been proposed. The main reason for this differentiation is the fact that QoE is a subjective measure and subjective quality assessment is based on subjective experiment which is expensive.

One way of QoE analysis method is by using statistical analysis of the subjective tests. The subjective test data was analyzed with the discriminant analysis method [22]. This method builds linear functions for each class or label that the data point can be associated with. The input parameter may differ based on the service and the type of technology under consideration.

Mapping function are also used on the QoE modeling method. Mapping function can be linear or non-linear [23]. The common mapping functions in the different literature are Logistic, Cubic, Exponential, Logarithmic, and Power functions. These different forms of mapping function correspond to differentQoS and QoE parameter measurements.

Another method used for QoE Modeling is based on ML methods. There are many QoE prediction models which are built by using ML methods like RF, KNN and ANN , SVM [24] and [8], Adaptive Neuro Fuzzy Inference System (ANFIS) [9], Ensemble method [16].

2.5 QOS MEASUREMENT POINTS AND TOOLS

When developing a QoS or QoE monitoring architecture, it is important to define data collection points. In general QoS measurement points are categorized in to five different points [25]. Each points have their own advantages and limitations. These measurement points, which are shown in Figure 2.2 were initially developed for Internet

Protocol Television (IPTV) and Video on Demand (VoD) applications but they can also be generally applied to any other service. The points are:

- *Monitoring Point 1 (PT1)*: Located at the border between Internet service provider (ISP) and Over The Top (OTT) content provider. At this point, it would be wise for ISPs to begin monitoring or attempting to foresee quality issues. This would aid ISPs in preventing problems farther down the transmission chain, which would necessitate more complex troubleshooting or monitoring closer to the user-end, which would be significantly more time consuming and costly in practice.
- *Monitoring Point 2 (PT2)*: This point marks the boundary between what is identified in the Figure below as "service provider" and "network provider". In practice, these are usually the same entities. At this point ISPs can gather information on quality and performance parameters of streaming media and check service-related policies.
- *Monitoring Point 3 (PT3)*: Defines the boundary between IP core and access networks. The purpose of monitoring at this point is to measure IP network related parameters such as: mean one-way delay, packet delay variation, transmission codecs and packet loss ratio.
- *Monitoring Point 4 (PT4)*: This point is located between the IP access network and end-user domain. Monitoring at this stage provides information on packet loss ratios, audiovisual content synchronization issues and IP network parameters. Unless ISPs gain access to end-user equipment, this is the last point at which service providers ISPs can monitor data that is as near to what end users receive as possible.
- *Monitoring Point 5 (PT5)*: This point is located at the end-user equipment. The information gathered here can be used to directly assess the quality of the user's experience. This is where user-end equipment or application-specific data might be collected.

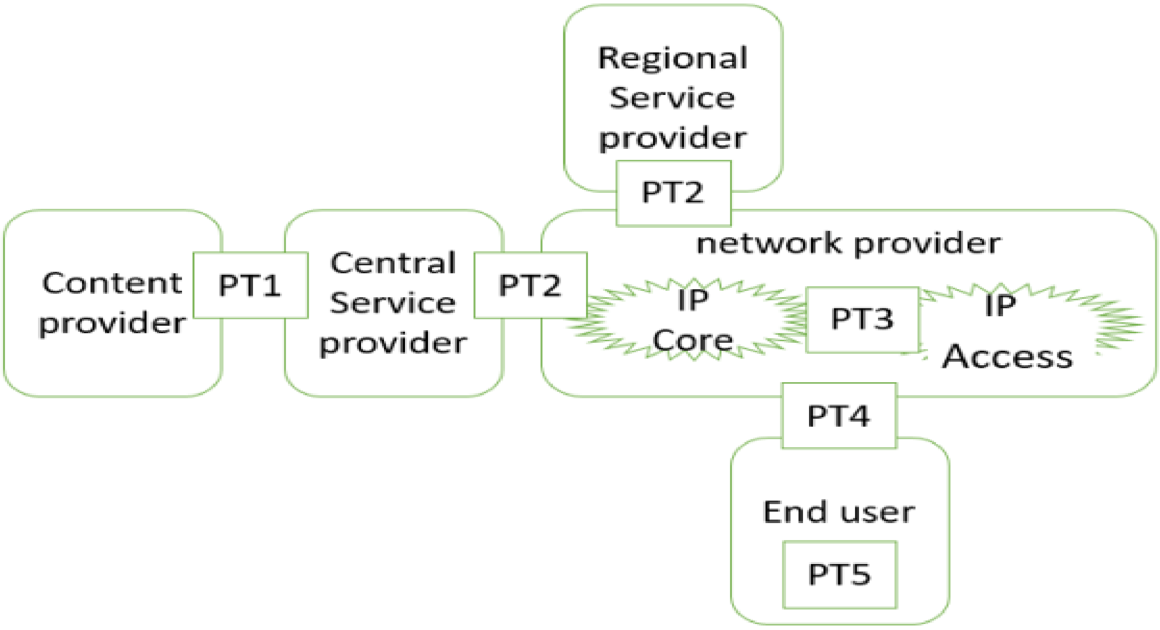


Figure 2.2: Network Measurement Points (Adopted from [25])

At each points there are different measuring approaches are followed. The approach deployed in this research paper is to measure the network parameters at measurement point 5 by using end user's mobile device. For end users which use smart phone, there are different freely available mobile applications for measuring different QoS parameters (packet loss, jitter, latency , up link and down link speed). On different papers [8], [26], [24], [27] different free mobile apps like Acqua, RTR, nPerf, YoMo App are used for measuring QoS Parameters.

3

ENSEMBLE METHOD ALGORITHMS

3.1 INTRODUCTION TO MACHINE LEARNING TECHNIQUES

Now a days ML become one of the most powerful tools which uses different algorithms to find patterns in data that generate insight and help you make better decisions and predictions. When there is a complex task or problem that involving a large amount of data-set with more variables there are conditions in which the existing formula or equation can't make sense of the data. In this case implementing ML techniques will be very helpful.

ML systems employ different algorithms or models to discover patterns in different type of data. The data-set may include structured data, unstructured data, numeric data and also multi media data.

ML algorithms can be supervised learning in which the prediction model is made based on known input and output data set , or it can be unsupervised learning when the data set is only input data and the learner tries to identify hidden patterns that connect different variables to form clusters of items which are similar to each other. Figure 3.1 shows machine learning techniques.

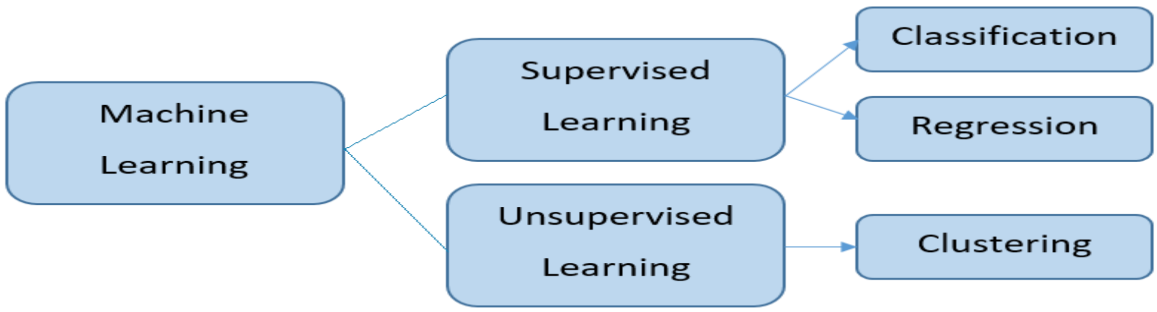


Figure 3.1: Machine Learning Techniques

Supervised learning is based on prior knowledge of the output values for the samples data. Therefore, the goal of supervised learning is to learn a function that, given a sample of data and desired outputs, best approximates the relationship between input and output observable in the data. where as in unsupervised learning, there is not labeled outputs, so its goal is to infer the natural structure present within a set of data points. In this research paper the data used have both known input and output values, therefore supervised learning based algorithm is used. Further supervised learning can be grouped in to two based on the technique used to develop the ML model.

- *Classification Techniques*: This technique is used when the input data can be tagged, categorized, or separated into specific groups or classes to develop a classification model. Common algorithms that use this techniques include support vector machine (SVM), boosted and bagged decision trees, k-nearest neighbor, Naïve Bayes, discriminant analysis, logistic regression, and neural networks.

- *Regression techniques*: This technique is used to develop a model if the type of response data is a real number. Common regression algorithms include linear model, nonlinear model, regularization, boosted and bagged decision trees, neural networks and adaptive neuro-fuzzy learning.

Both classification and regression ML algorithms learn a mapping function from the inputs to outputs data provided. On the process of learning there are errors made by learning algorithms which is described by two terms bias and variance.

The bias describes how close the model can capture the mapping function between inputs hand outputs. Where as the variance of the model measure how the performance of the model changes when it is fit on different training data. Both bias and the variance of a model's performance are related. A model with a low bias and low variance is preferable, even though this is very challenging modeling problem to achieve in practice. The bias can often be minimized easily by increasing the variance and the variance can also be minimized by increasing the bias.

3.2 ENSEMBLE METHOD

3.2.1 *Overview*

For a particular dataset, a single algorithm might not be able to generate the ideal accurate predictive model. Different ML algorithms have their own constraints, and creating a high-accuracy model is the main difficulty in modeling process. There are different techniques to increase the total accuracy of the model, one way to achieve high model accuracy is by building different sub models and combining the output of each models on the final model.

Ensemble techniques are a ML approach to combine multiple other models called base estimators in the modeling process. Ensemble learner overcome the challenges on ML methods, that use a single algorithm for the model. i.e. increase predictive performance by decreasing error on regression and increasing accuracy on classification.

3.2.2 *Ensemble Techniques*

There are different techniques used in ensemble techniques, but the most commonly used techniques in ensemble method are:

- *Bootstrap Aggregating (Bagging)*: In this method base estimators or base models are generated using the same algorithm.
- *Adaptive Boosting (AdaBoost)*: In this method the base estimators are weak learner algorithms to create the last strong model.

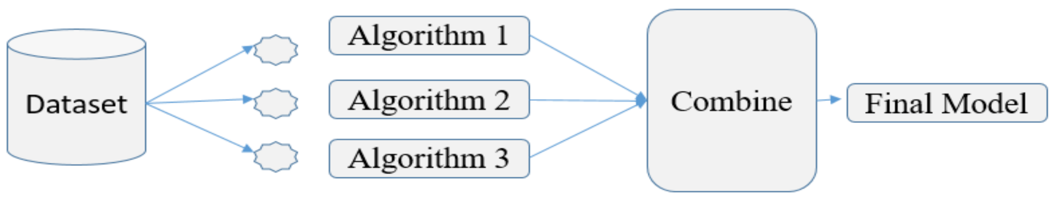


Figure 3.2: Ensemble Techniques Using Multiple Algorithm (Adopted from [28])

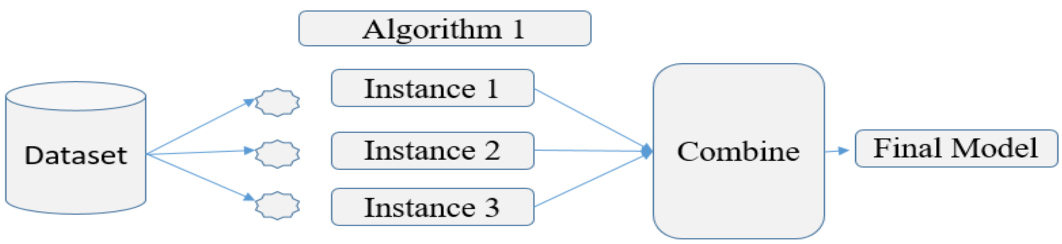


Figure 3.3: Ensemble Techniques With One Weak Learner Algorithm (Adopted from [28])

3.2.3 *Aggregating Predictions*

There are different ways of combining the base estimator to get the final model depending on the type of problem. Like, it a Classification Problems, Which focus on simply a categorization process or regression Problems which try to find the best fitting membership function. Based on the type of problem The following three main techniques can be used:

- *Majority Voting*: The final prediction is made based on majority voting for classification problems.
- *Averaging*: Typically used for regression problems where predictions are averaged.
- *Weighted Average*: when base models/algorithms have different weight than the others.

3.2.4 *Hyper Parameters*

There are parameters in ML algorithms which specify how to transform the input data into the desired output, we call these parameters model parameters. There are other parameters which cannot be directly trained from the data but used to define the architecture of the model, which are referred to as hyper parameters. On tree based ML algorithms the model architecture addresses the following ideas.

- The maximum depth allowed for the decision tree.
- The minimum number of samples required at a leaf node for the decision tree.
- The number of trees for the decision tree.

In order to know the optimal model architecture, the range of these hyper parameter possibilities should be explored. The process of searching for the ideal model architecture is referred to as hyper parameter tuning. There are three ways to search the optimal hyper parameter values

- *Grid search*: Construct a model for each possible combination of all hyper parameter values,evaluate each model, and choose the architecture that provides the best results.
- *Random Search*: Construct a model for randomly selected sample hyper parameters based on statistical distribution of each hyper parameter.

- *Bayesian optimization*: The above two methods performed method model optimization by various hyper parameter values independently and record the model performance for each case. In Bayesian optimization the results of our previous iteration is used to improve the sampling method of the next experiment. The iteration process continues until the model converges to an optimal value.

3.3 BOOTSTRAP AGGREGATING

Bootstrap Aggregating method is based on the idea of "bootstrap "which is a powerful statistical method for estimating a quantity from a data sample. In this method if the number of samples are small in number, the mean will have error in it. To improve the mean, we use the bootstrap procedure:

- Create many random sub-samples of data-set with replacement i.e. the same value can be selected multiple times.
- Calculate the mean for individual sub-sample.
- Calculate the average of all sub- samples' mean, and take the value as the mean of the original data.

This process can be used to estimate other quantities like the standard deviation and even quantities used in different ML algorithms. The idea behind Bagging is, it applies Bootstrap procedure to a high-variance ML algorithm, typically Classification and Regression Trees (CART) to reduce variance. In this technique multiple subset of data-set are created from the original data-set with replacement. Since the original training data-set is re-sampled with replacement, certain instances may be selected many times while others are not selected at all. Then each subset of data, which have an equal data size are trained parallel by the classification decision tree algorithm. finally, each models are combined by majority vote rule to produce the final model. This is shown in figure 3.4. The meta-estimator reduces the variance of the base estimator through the introduction of randomization into the construction method and then generating an ensemble from it [29].There are two hyper parameters in bagging technique that define the model architecture. These are

- *Number of learner*: Number of decision trees.
- *Maximum number of split*: The depth of decision.

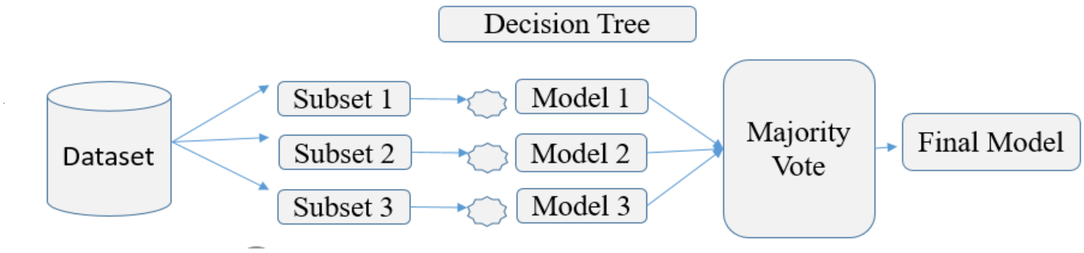


Figure 3.4: Bagging Technique (Adopted from [28])

Many popular ensemble algorithms are based on this approach, includes

- Bagged Decision Trees (canonical bagging),
- Random Forest,
- Extra Tree.

3.4 MODEL EVALUATION METRICS

The initial data-set is trained by a chosen algorithm to produce the final model in the ML modeling process. Before assessing the final model with multiple evaluation approaches, the k fold cross-validation procedure is used to ensure a stable model. The data set is shuffled randomly, then split into k subgroups, as is the standard approach in k-fold validation. One of the K's subgroups will be utilized to test the model, while the remaining k-1 will be used to train the model.

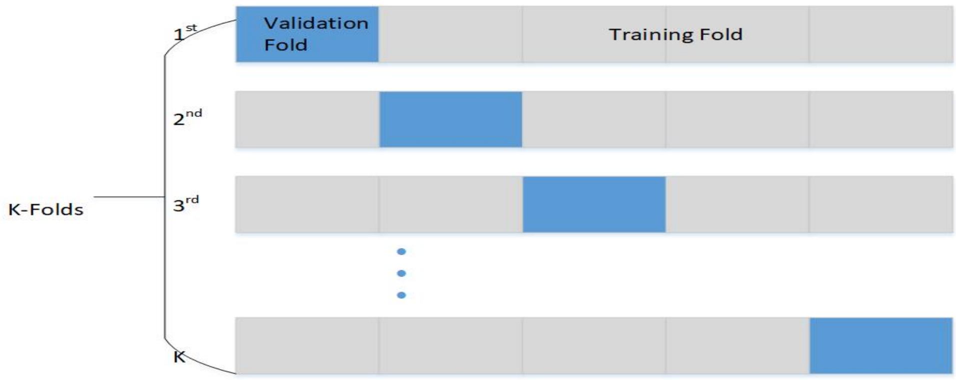


Figure 3.5: k-fold cross validation

After modeling, the model's performance is assessed using the loss function. Loss functions are metrics that compare expected values to actual values; the error or residual is the output of a loss function. The accumulated mistakes throughout the entire data set will be the overall model validation error.When evaluating the performance of a predictive model, there are several loss functions to choose from, each providing a distinct insight of the predicted accuracy and differing across regression and classification models. The following are some of the most commonly utilized functions:

- *MSE*: Mean squared error is the average of the squared error.it is used for regression based model evaluation. MSE is the most common error metric use the objective is to have minimum MSE value.
- *RMSE*: Root mean squared error, it simply takes the square root of the MSE metric. It is also for regression based models
- *Misclassification*: This is the overall error for classification model which calculate how much of the data are classified wrongly from the total data set. the objective is to have minimum value.

3.4.1 *Confusion Matrix*

A confusion matrix is a technique for evaluating the performance of ML classification models. It's just a matrix comparing actual categorical values to expected categorical values. By comparing the actual and predicted classifications, the confusion matrix visualizes a classifier's accuracy. When we accurately anticipate actual values, we call them a true positive, and when we predict incorrect values as incorrect, we call them a true negative. A false positive occurs when we anticipate a value that does not occur, and a false negative occurs when we do not predict a value that does occur.

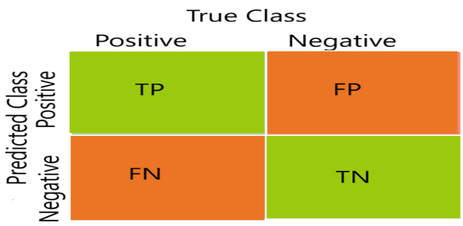


Figure 3.6: Binary confusion matrix (Adopted from[**CM**])

Where:

- *TP*- True Positive, which predict actual positive as positive
- *TN*- True Negative, which predict an actual negative as negative
- *FP*- False Positive, which predict Negative values as positive
- *FN*- False Negative, which Predict Positive values as negative

Note that: TP and TN predicts values correctly whereas FN and FP predict values incorrectly. Based on the confusion matrix values, the performance of the model is evaluated by the following metrics.

1. **Accuracy** The accuracy is used to find the portion of correctly classified values. It tells us how many of the correctly classified are predicted correctly.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

2. **Precision** Precision is used to calculate the model's ability to classify positive values correctly. It measures how likely the prediction of the positive class is correct.

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

3. **Recall** It is used to calculate the model's ability to predict positive values. It explains from all the positive classes, how many we predicted correctly.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3.3}$$

4. **F-Measure (F1-Score)** F1-Score: is a weighted average score of the true positive (recall) and precision.

$$\mathrm{F} - \mathrm{Measure} = \frac{2 * \mathrm{Recall} * \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{3.4}$$

3.4.2 *Receiver Operator Characteristic (ROC) curve*

Another tool for evaluating binary classification problems is the ROC curve. It's a probability curve that compares the TPR to the FPR at various thresholds. The Area Under the Curve (AUC) is a summary of the ROC curve that measures a classifier's ability to distinguish between classes. The AUC represents the model's ability to distinguish between positive and negative classes. The greater the AUC, the better the model's performance. The ROC curve for various AUC values is shown in Figure 3.7

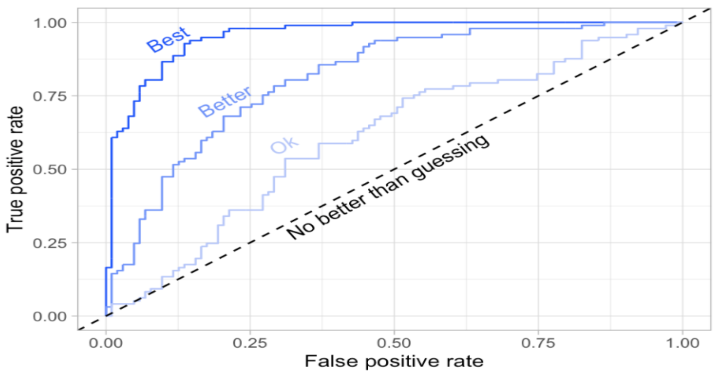


Figure 3.7: ROC Curve with Different AUC Values

- If AUC = 1, then the classifier is able to perfectly distinguish between all the Positive and the Negative class points correctly.
- IfAUC = 0, then the classifier would be predicting all Negatives as Positives, and all Positives as Negatives.

- If $0.5 < \text{AUC} < 1$, there is a high chance that the classifier will be able to distinguish the positive class values from the negative class values.
- If AUC=0.5, then the classifier is not able to distinguish between Positive and Negative class points. Meaning either the classifier is predicting random class or constant class for all the data points.

3.5 FEATURE IMPORTANCE

Feature importance refers to a class of techniques for assigning scores to input features to a predictive model that indicates the relative importance of each feature when making a prediction. Feature importance scores can be calculated for problems that involve predicting a numerical value, called regression, and those problems that involve predicting a class label, called classification. The scores are useful and can be used in a range of situations in a predictive modeling problem, such as:

- *Feature importance scores can provide insight into the data-set*: The relative scores can highlight which features may be most relevant to the target, and the converse, which features are the least relevant.
- *Feature importance scores can provide insight into the model*: Most importance scores are calculated by a predictive model that has been fit on the data-set. Inspecting the importance score provides insight into that specific model and which features are the most important and least important to the model when making a prediction.
- *Feature importance can be used to improve a predictive model*: This can be achieved by using the importance scores to select those features to delete (lowest scores) or those features to keep (highest scores).

4

METHOD DEPLOYED FOR MODELING

4.1 DEPLOYED SYSTEM MODEL

The system model used in this paper to model the QoE of SM video streaming services using an ensemble method is shown in Figure 4.1. The selected IF are network QoS parameter, users' mobile devices parameters and application parameter. These parameters are selected by providing questioner to the selected SM users of different ages, educational backgrounds, and gender in a different place in Addis Ababa.

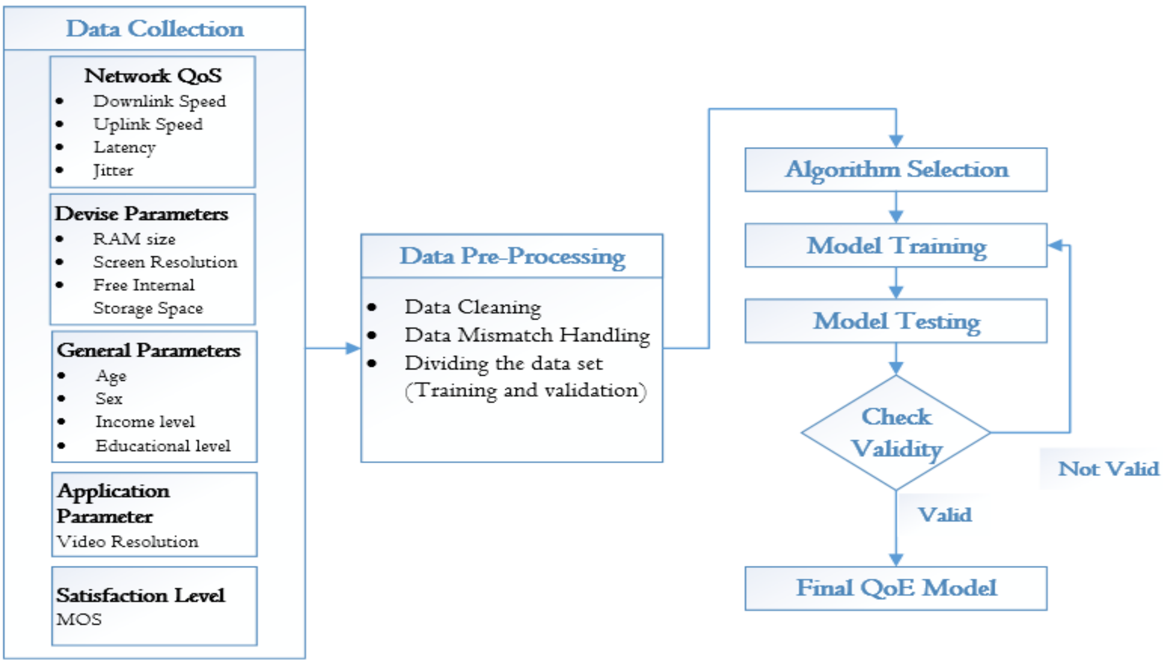


Figure 4.1: The Overall System Model Deployed

4.2 DATA COLLECTION

Data is collected by distributing questioners to selected SM users. The questioners are organized into four main parts.

- The first part is to find the SM applications that are mostly used by the participants.
- The second part aims to gather participant's personal data plus their device parameter.
- The third one is to collect the network QoS parameter.
- The last part is the participants' feedback according to MOS value.

The first part of the questioner focus to gather the top used SM applications that are mostly used by the survey participants and their preferred network type used for each application i.e. whether they use cellular data (3G or 4G) or Wi-Fi. In addition, their social media usage history is recorded from their cell phone device for the past two months. This is done to differentiate participants that use SM frequently from the one that uses rarely. Participants who have no or little SM usage history based on their data usage history on their mobile phone but have usage history on other devices like laptops, tablets, or desktop computers are excluded from the survey. The general data collection setup is shown in Figure 4.2. On the second part of the questioner, the participant's personal data plus their device parameters are collected. The personal data include the participants' education, income, age, and gender data from different location of Addis Ababa. The device parameters include the mobile phone RAM size, mobile phone screen resolution in a number of pixels per inch (PPI) which is calculated by dividing diagonal pixel resolution divided by diagonal size, internal storage free space in megabytes and the phone model.

The other issue to consider is to decide the number of participants on the survey. The sample size used for research purposes should be carefully fixed so that it will be adequate to draw valid and generalized conclusions.

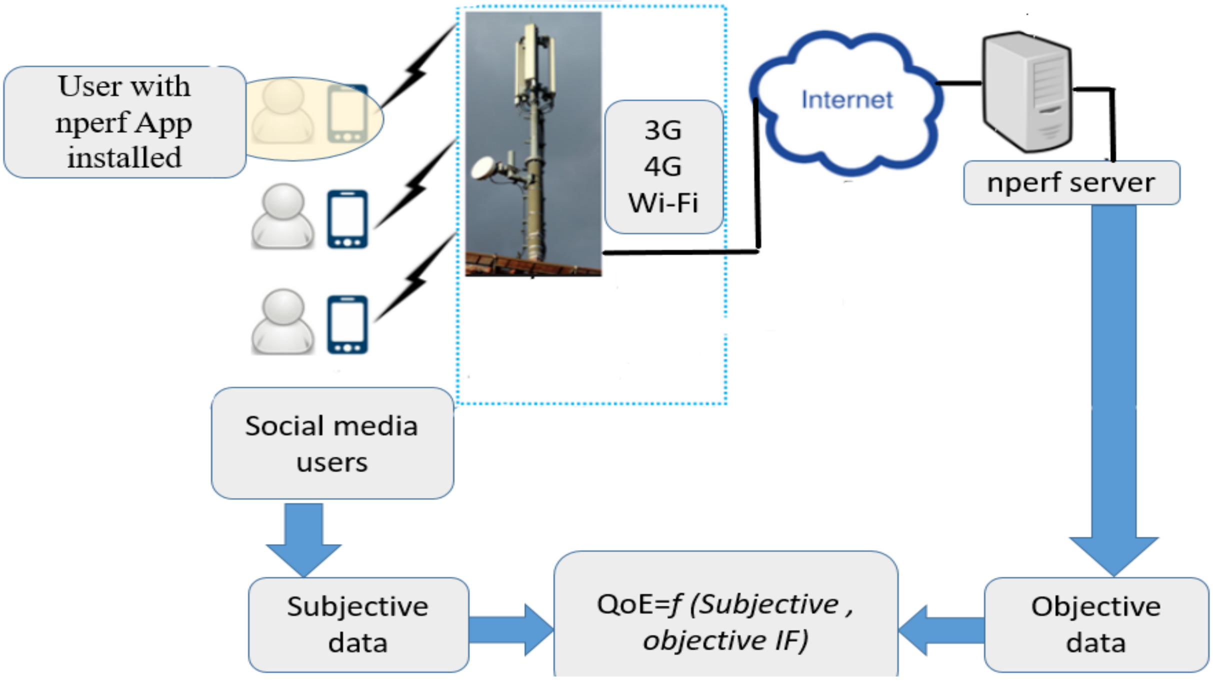


Figure 4.2: Data Collection setup

There are different ways to determine the sample size [30] some of the methods are sampling

- Using a Census for Small Populations by taking the entire population as the sample.
- Using a Sample Size of a Similar Study.
- Using Published Tables which provide the sample size for a given set of criteria.
- Using Formulas to Calculate a Sample Size.

There are many formulas used on different research papers to find the sample size. in this paper, a Simplified Formula for Proportions noted on (4.1) is used to calculate the number of samples.

$$n = \frac{N}{1 + (N * e^2)} \tag{4.1}$$

Where N= Population size, e= precision, and n= sample size

Since the focus area of this work is only in Addis Ababa, Ethiopia. The population size is determined by the population of Addis Ababa. Based on [31] the current population of Addis Ababa in 2021 is expected to be 5 Million, and this will be the value of N.

The other parameter is the level of precision e, sometimes called sampling error. It is the range in which the true value of the population is estimated to be. This range is often expressed in percentage points, by taking e = ±5% and N = 5 million, n, which is the minimum sample size become 400.

After deciding the number of participants the survey is done by randomly selecting participants which are believed to best represent the society. The participants are selected from different sub-cities of Addis Ababa (Bole, Nifas Silk, Akaki Kality, Yeka, and Arada sub-cities) and also with different age groups, educational backgrounds, gender, and income level. Out of the total collected 473 data 80 of the patricipants filled the survey more than once within different time of the day and different place.

The other parameter collected is network QoS parameters, which include download speed, upload speed, jitter, and latency. To collect these data free mobile application nPerf is used. This tool measure uploads and download bit rate by downloading or uploading a binary file with a simultaneous connection for a few seconds. The detailed working principle for measuring different QoS is described in [32]. To record

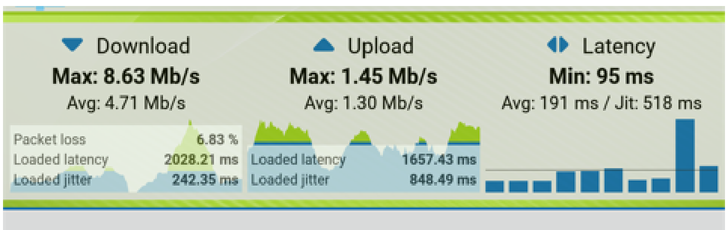


Figure 4.3: Sample nPerf Reading

these network QoS parameters the selected participants are subjected to watch a video of their interest on their preferred SM (YouTube, Facebook, or Tiktok) for a minute. In the meantime, the nPerf application runs to measure QoS parameters at the time of the video play.

After watching the video, on the last part of the questioner, the participants are asked to give their feedback based on MOS rating from 1 to 5.

Table 4.1: MoS Rating values

MOS scale	Rating Meaning
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

4.3 DATA PRE-PROCESSING

The data pre-processing step focuses on the data preparation for training the data set with the selected algorithm. A total of 10 features are collected from the survey as shown in Table 4.2. On the data collection process of the survey, some of the participants left the device parameters RAM, and screen resolution null. These missing data are completed by searching the parameters from the internet by using the mobile phone model, which is one provided by the participants. This makes the missing data complete. The other issue is data type variation, the collected data contains both numeric and non-numeric categorical data types. Non-numeric categorical features age group and sex are used as input parameters for the modeling. These parameters are converted to dummy variables for the sake of analysis. The other input parameter, device resolution is two-dimensional data which is pixel per inch for the length and width of a mobile device. For the analysis purpose, the multiple of the two dimensions is used as an input feature.

4.4 MODEL TRAINING AND VALIDATION

The total data set is divided in to training data set which is 80% and the rest for model validation.The training data set is trained by Bootstrap Aggregating (Bagging) method, which is a ML based ensemble technique. The final QoE model performance is evaluated by Accuracy, Precision, Recall and F-score based on confusion matrix, and using ROC curve. 20% of the total data set is used to measure the errors between the estimated and the actual collected MOS values. In this paper, a Regression analysis is used to check how fit the model is by Using the validation data set.

Table 4.2: The selected features used in the model

Feature	Unit	Description
Download Speed	Kbps	The amount of data received in one second from the video server
Upload Speed	Kbps	The amount of data send in one second to the video server
Latency	Millisecond (Ms)	It is the amount of instant time it takes for a packet to be captured, transmitted then received at its destination
Jitter	Millisecond (Ms)	It is a delay variation due to the inherent variability in arrival times of individual packets
RAM	Megabits	Mobile phones RAM size
Phone Screen Resolution	Pixel per inch	The number of pixel per inch of the mobile width * The number of pixel per inch of the mobile height
Phone Internal Storage Free Space	Megabits (Mb)	The available mobile phone free space
Video Quality	Pixel(P)	The number of pixel contained in each frame.
Age	-	The different age group of peoples that participate on the survey
Gender		Male or female
Satisfaction Level	MoS	collected from the user to determine how the customers are happy with the SM video streaming services

5

RESULTS ANALYSIS AND INTERPRETATIONS

5.1 THE COLLECTED DATA OVERVIEW

From the survey result YouTube, Tiktok, and Facebook have high precedence in terms of usage. This is shown in Figure 5.1 out of the total 473 samples YouTube is ranked first by 46%, Tiktok in second place by 39%, and Facebook in third place by 15% in terms of participants' choice for watching video services. In this thesis work, QoE is modeled based on data gathered from these three SM applications users. When looking at the network used by the participants 77% of them use 3G network while the rest 23% are LTE users. For video related activities 85% of the participants use Wi-Fi network instead of their cellular network.

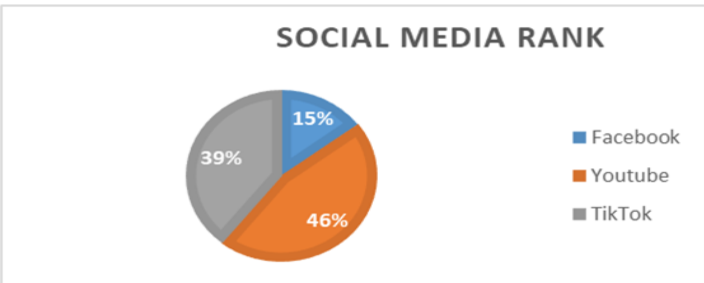
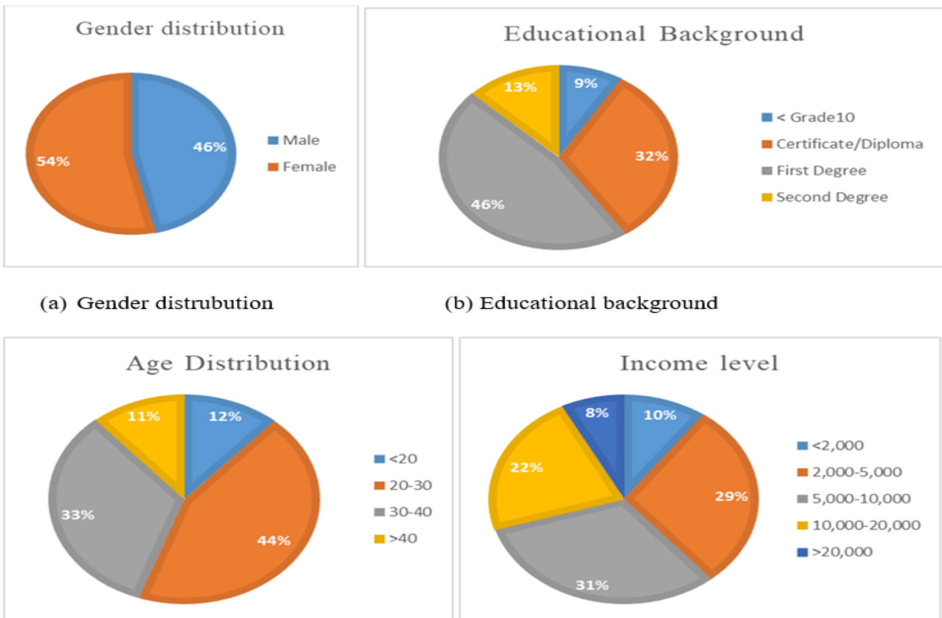


Figure 5.1: Social Media Application Preference Rank

Table 5.1: The participants phone data overview

Phone Model	Number of Subscribers
Samsung	132
Zte	10
Itel	92
Huawei	2
Infinix	89
Lenovo	29
Tecno	115
iPhone	4
TOTAL	473

Figure 5.2shows the selected participants data over view.



(a) Gender distrubution

(b) Educational background

(c) Age contribution

(d) Income level

Figure 5.2: survey Data Overview (a) Gender , (b) Education, (c) Age and (d) Income level

Based on the selected data the participants' feedback on the SM video streaming is summarized in Table 5.2. The average MOS value is 2.8 which shows the average satisfaction level of video streaming services on SM applications is fair based on the MOS scale.

Table 5.2: Collected MoS values

MOS	Count	Percentage Contribution
1	137	29%
2	33	7%
3	123	26%
4	147	31%
5	33	7%

The general statistical result of the collected data,the minimum, maximum, mean, and standard deviation of each features used in the model are described on the Table 5.3.

	Phone RAM size(Mb)	Phone free space (Mb)	Screen Resolution (L*W PPI)	Video Resolution (pixel)	Download speed(Kbps)	Upload speed(Kbps)	Jitter(Ms)	Latency (Ms)	MoS
Mean	2.31	625	1287216.8	377.4	2622.43	531.01	65.85	798.28	2.80
SDV	1.4180996	947.8	801237.5	127.15	4297.61	442.36	53.96	530.15	0.06
MIN	1	152	409920	144	12	50	1	250	1
MAX	8	7317	3686400	720	26374	2696	200	2275	5
Count	473	473	473	473	473	473	473	473	473

Figure 5.3: The statistical analysis of the input features

5.2 CORRELATION ANALYSIS OF INPUT FEATURES

The correlation analysis has been done for feature selection to reduce if there are redundant parameters. Furthermore, the correlation result was used to identify the impact of each parameter on the user perception. It is done by using Pearson Correlation

Coefficient (PCC) or R values by assuming the parameters are sampled from a population that exhibits an approximate normal (Gaussian) distribution. From the correlation analysis from heat map plot in Figure 5.4 the user perception MOS is highly correlated with the network QoS parameter download speed with $R = 0.528$. This shows the user satisfaction level is highly affected by the download speed. The other parameters jitter and latency also affect the perceived quality in inverse relation this indicates as the value of jitter and latency increase the MOS values decrease. From the users' device parameter, RAM size has a correlation value of $R = 0.35$. This also shows it has a significant influence on the users' perceived quality.

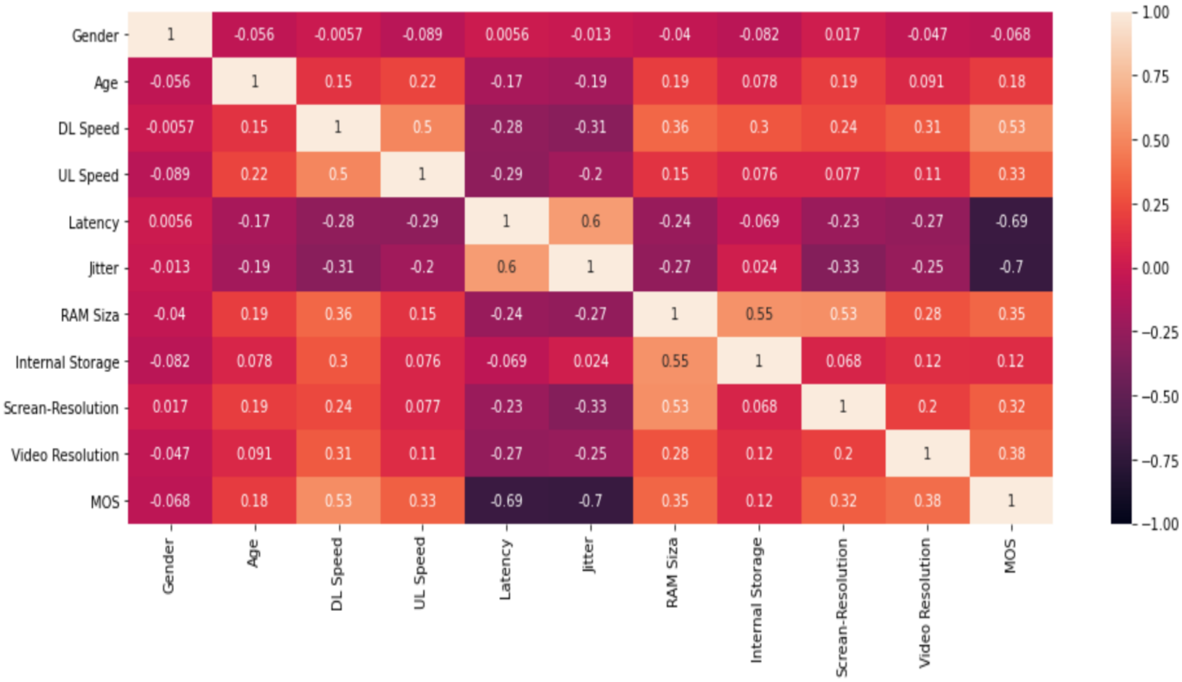


Figure 5.4: Heat map plot of the correlation matrix

5.3 TRAINING MODEL

From the supervised ML category, Classification predictive modeling is used to map input variables to discrete output variables. In our classification problems, the learning algorithm learns a function to map inputs to outputs where the output value is a discrete class MOS values. For developing the final model the total data set is divided into training and validation set,which is 80% are the training data and the rest 20% for validation. A total of 467 valid data set are collected from the survey, out of these 80%(374) data set are trained by ensemble technique using MATLAB 9.4 Release R2018a Statistics and Machine Learning Toolbox. For the ensemble method, the two hyper parameters used in this model are the number of decision trees and the depth of each tree to train the model as described in Section 3.2.4. From the two commonly used ensemble techniques, the bagging method which uses a Decision Tree algorithm is used for training the data-set. The two hyper parameter of this algorithm are:

- The depth of decision tree = the number of training data set =374.
- Number of decision trees = 500, which is the maximum possible value.

By taking these values, there are many possible combinations to train the model. That is

$500 * 374 = 187,000$, there are $187,000$ possible combinations of hyper parameters. This is a large number of iteration to check each combination to train the model. Instead of manually selecting these options, Matlab hyper parameter optimization within the Classification Learner app is used to automate the selection of hyper parameter values. Different combinations of hyper parameter values are tried by using an optimization scheme to minimize the model classification error and return a model with the optimized hyper parameters.

Table 5.3: Possible Hyper Parameter Values

Hyper parameter	Possible Value
Number of decision trees (Number of learner)	10 – 500
The depth of decision tree (Maximum number of split)	[1 - max (2, n–1) Where n is the number of observation (dataset) which is 374

The output of the hyper parameter optimization is a minimum classification error plot that shows the following information.

- *Estimated minimum classification error*: Computed by the optimization process when considering all the sets of hyper parameter values tried so far, including the current iteration.
- *Observed minimum classification error*: Which corresponds to the observed minimum classification error computed so far by the optimization process.
- *Best point hyper parameters*: Corresponds to the optimized hyper parameters.
- *Minimum error hyper parameters*: The iteration that corresponds to the hyper parameters that yield the observed minimum classification error.

Figure 5.5 shows a sample misclassification error plot for 1000 iteration. Model train-

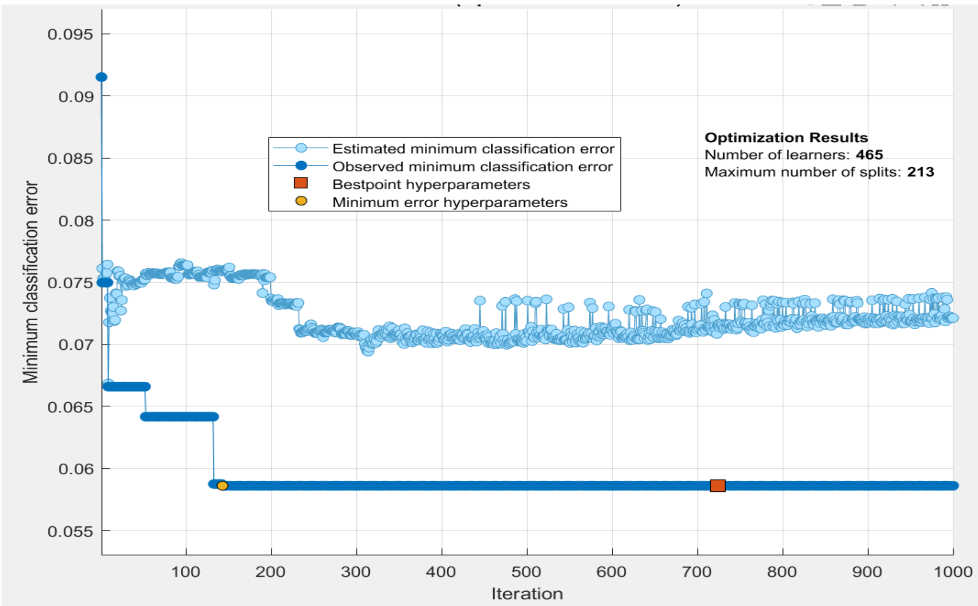


Figure 5.5: Minimum Classification Error Plot for 1000 iteration.

ing is done on a number of times to get a model with better accuracy. Each iteration lasts from a few seconds up to hours to finish the training process and suggest optimal hyper parameter values. Table 5.4 shows the accuracy and misclassification cost for sample iteration

Table 5.4: Model Training Iteration and The Corresponding Hyper Parameter Values

Iteration	Number of split	Number of Learner	Misclassification cost	Accuracy
200	30	25	27	92.8%
500	274	103	27	92.8%
500	50	473	28	92.5%
1000	120	489	24	93.6%
1000	213	463	29	92.2%
1500	60	470	24	93.6%
1	50	454	22	94.1%

5.4 TRAINED MODEL RESULT

From the above sample training and manually testing different values of hyper parameters, a model with the highest accuracy is selected as a final model. Different hyper parameter values and other model training results are listed in Table 5.5. The final model has an accuracy of 94.1% with a miss classification cost of 22 and hyper parameter values, 454 decision trees each having a depth of 50. Figure 5.6 shows correctly classified observations of each class on the diagonal cells and miss classified observations on the other cells. When considering class 1 that is MOS value of 1, 98.1% are correctly classified and the rest 1.9% are classified as MOS 2. For class 2, 3.3% are classified as MOS 1 and the other 3.3% are classified as MOS value of 3 and the rest 93.3% are correctly classified as MOS value of 2. For class 3, 1.1% are classified as MOS value 1 and 5.4% are classified as MOS value of 4, and the rest 93.5% are correctly classified. For class 4, 92.6% are correctly classified whereas the rest are misclassified. finally, class 5 has 11.1% misclassification.

Table 5.5: Result of the trained model

Parameter	Value
Accuracy	94.1
Misclassification cost	22
Training time	54.78 sec
Ensemble method	Bag
Learner type	decrision tree
Maximum number of split	50
Number of learner	454

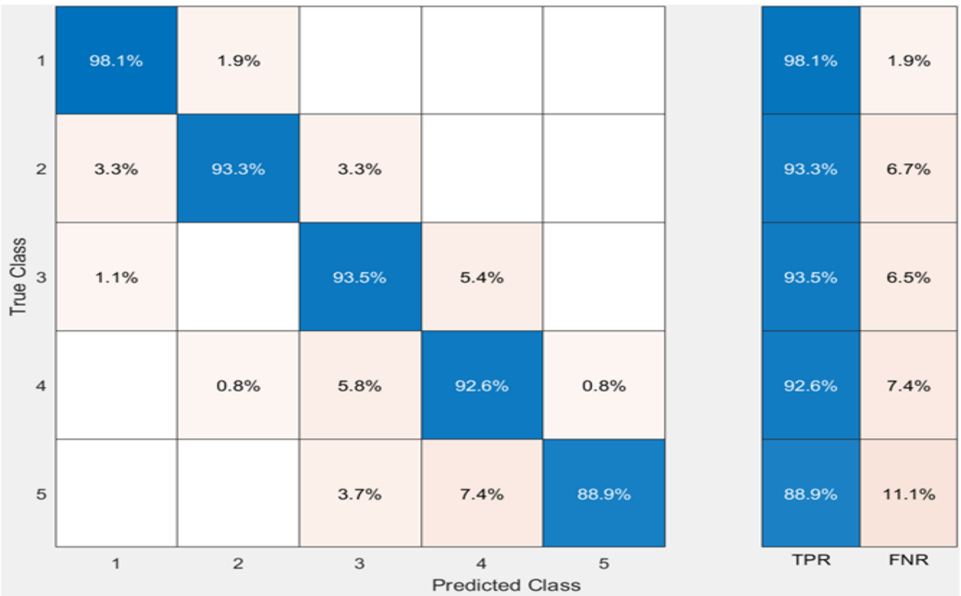


Figure 5.6: Number of correctly classified and miss classified class

5.5 MODEL PERFORMANCE EVALUATION

A confusion matrix is one of the performance measurement technique for ML classification models. Table 5.7 shows the trained model confusion matrix values.

class		Predicted	
	Actual	T	F
1	T	102	2
	F	2	268
2	T	28	3
	F	2	341
3	T	86	9
	F	6	276
4	T	112	7
	F	9	246
5	T	24	1
	F	3	346

Figure 5.7: Confusion Matrix Values

From the confusion metric the model accuracy for each class is computed with formula 3.3. Each class corresponds to the corresponding MOS values. MOS values 1 and 5 have the highest accuracy value 98.93% each and MOS value 4 have relatively less accuracy than the others, which is 95.72%.Figure 5.8 shows the the accuracy of each class. The other evaluation metric is precision, the precision of each class describes

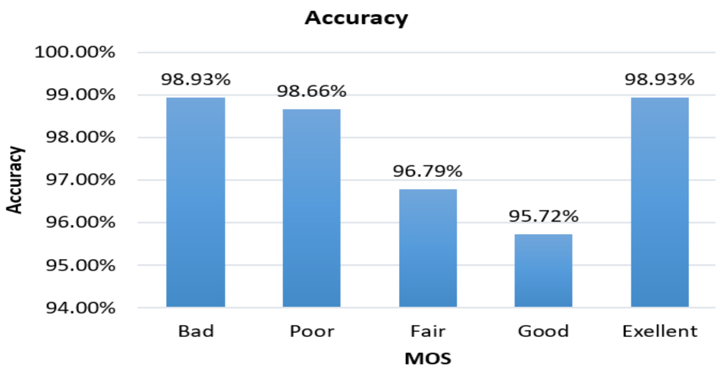


Figure 5.8: Model Accuracy

how much the model classifies positive values correctly. From Figure 5.9) MOS value 1 has the highest precision value 98.08% and MOS value 5 has less precision than the others' which shows the model classifies MOS value 1 more likely, Whereas the model classifies MOS values 5 as less likely. That means some values of them are classified as if they belong to other MOS values. The other evaluation metric, recall shows how much the model can predict different classes from all the positive classes. Which is

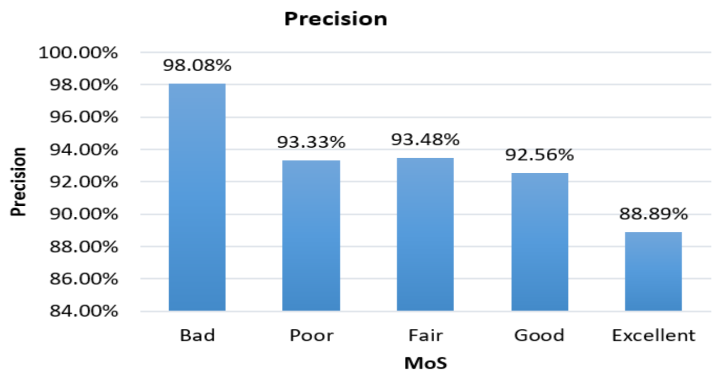


Figure 5.9: Model Precision

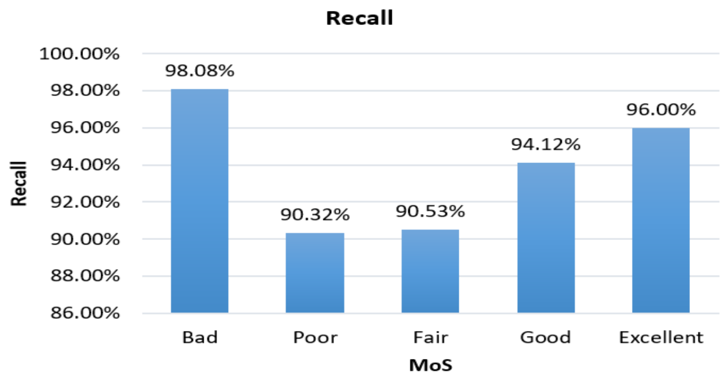


Figure 5.10: Model Recall

shown in Figure 5.10 again the model predicts MOS Value 1by98.08%. which is the highest and MOS value 2 has the lowest value.

The other metric F-score shows the average of the precision and recall value which shows how can the model classify and predict positive class. F-score values range from 0 to 1, if the F-score value approaches to 1 indicate model is a good model. In this model the F-score ranges from 0.92 to 0.98 as shown in figure 5.11.

The ROC curve is another evaluation metric that plot a probability curve of TPR against FPR at various threshold values. AUC is used as a measure of the ability of a model to distinguish between classes and is used as a summary of the ROC curve. AUC value ranges from 0to1,the maximum AUC=1 means that the model is perfect in the differentiation between the specific class and the rest of the classes while AUC=0 means the model incorrectly classify all classes. ROC curve for class 3 is shown in Figure 5.12 with AUC of 0.98 this value shows the model differentiate Fair satisfactory level from

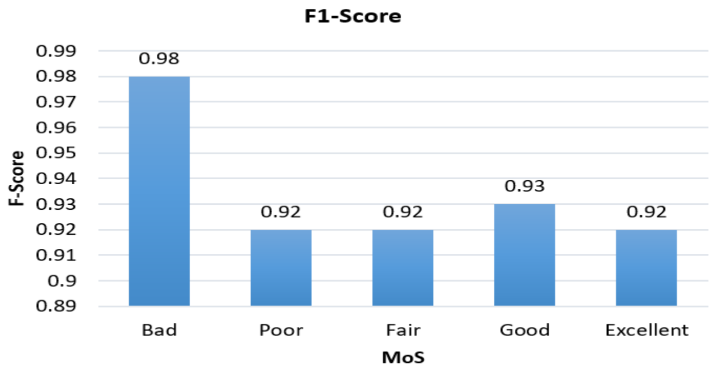


Figure 5.11: F-score Values for each class

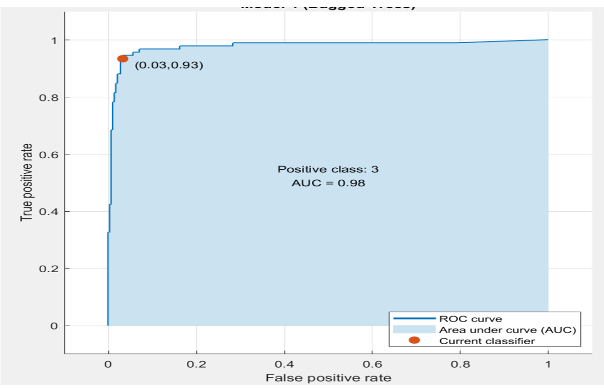


Figure 5.12: ROC Curve for Class 3

the other MOS values with high accuracy. The other ROC plot for the rest of the class is found in Appendix A1. As shown in Table 5.6 AUC values of each class ranges from 0.97to0.99 which shows the model classify the classes with high accuracy.

Table 5.6: AUC of Each Class

MoS	1	2	3	4	5
AUC	0.99	0.97	0.98	0.98	0.99

5.6 MODEL VALIDATION

After developing our QoE model, it is necessary to validate the model using another test data set. The test data which is 20% of the total data set is used to measure the errors between the estimated and the actual collected MOS values. In this paper, a

Regression analysis is used to check how fit the model is by using the validation data. Table 5.7 shows the model has good accuracy with R, R square, adjusted R square, and RMSE of $0.957, 0.916, 0.915, and 0.340$ respectively for 93 test data set. In addition, Figure 5.13 shows how fit are the actual MOS values and their corresponding predicted value by the developed model. From the above validation tests the model is good for our data set.

Table 5.7: Regression Between Validation and the Corresponding Predicted Data

R	R-Squared	Adjusted R-Squared	RMS	Observation
0.957	0.916	0.915	0.34	93

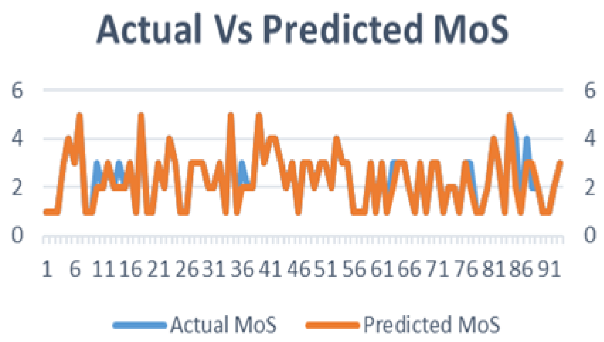


Figure 5.13: Actual and Predicted line graph

5.7 FEATURE IMPORTANCE ON THE MODEL

Different features have a different level of contribution to the model. The feature importance plot in Figure 5.14 shows the input features' rank based on their contribution to the model. The feature with a high feature importance score has high predictive power.

Download speed is the main IF of the model followed by jitter, phone internal free space and phone RAM size, latency, and video resolution. The rest parameters, gender, screen resolution, age, and upload speed have very small importance in the model. The feature importance score for each feature is listed in figure 5.14

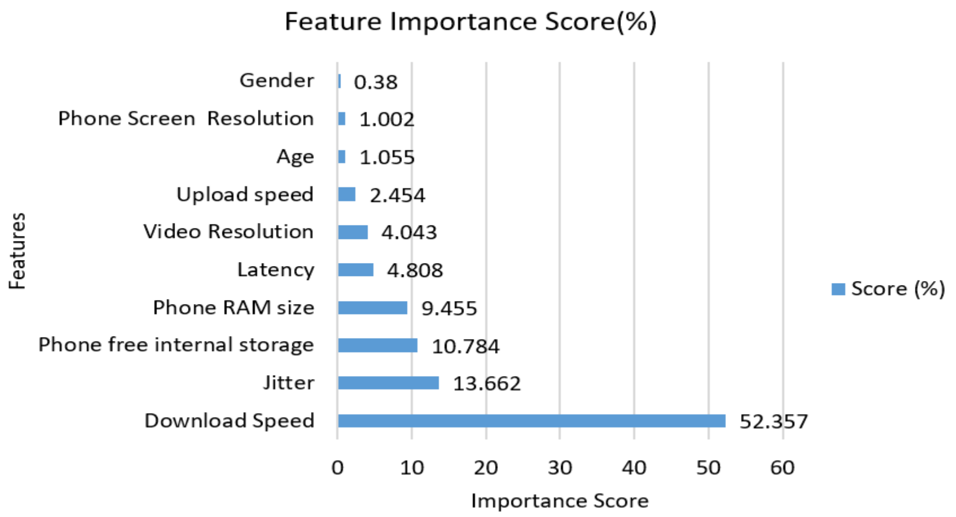


Figure 5.14: Feature Importance Plot

From the results obtained in this paper, the selected ensemble method technique, which is Bagging method, has good accuracy. The impact of each feature on the final model is described on feature importance score values. Which shows how much the QoE values are influenced by the IF. Apart from this, from the survey data, the overall satisfaction level of SM users is below fair. Generally, this thesis meets the initial objective, i.e to model QoE using different IF, demonstrates each features impact on the total satisfaction level, and assess the overall SM video streaming users' satisfaction level.

In doing so, there were different challenges like conducting the survey, on model training to find optimum hyper parameter values; it needs to train the data set on different iterations, which takes hours to finish one large iteration. To overcome this challenge, the maximum number of iterations has to be 1000, which improves the hours to be 8 hours.

6

CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The main purpose of this thesis is to propose a model that is capable of predicting the user perceived quality of SM video streaming services. Users perceived quality can be influenced by many factors, out of these, network QoS parameters, users' mobile phone parameters, and application-related parameters are included in this paper as an IF. From network QoS download speed, upload speed, latency, and jitter are selected as an input feature for the model. From mobile device parameters, phone RAM size, phone internal free space, and users' phone screen resolution are selected as input features. From the application side since the only concern of this thesis is SMvideo services video resolution is selected as an application parameter. SM users' satisfaction level is collected by providing questioners for selected users that are believed to be a good representative of the actual population. Each participant of the survey is subjected to watch their preferred video on either YouTube, Facebook, or Tiktok which are the SM that are selected for this thesis. On the survey, while the selected SM users are watching the videos, the network QoS parameters are collected by a network performance measuring tool called nPerf. In addition to these parameters, the participants' cell phone parameters are collected from each participant. An ensemble method which is a ML-based technique is used to develop the model. The proposed SM video streaming QoE model has an accuracy of 94.1%. further, the model is evaluated by accuracy precision and recall which are performance measurement techniques for ML classification models. The proposed model has an accuracy of $98.93\%, 98.63\%, 96.73\%, 95.7\% and 98.93\%$ for MOS values $1 to 5$ respectively. Further, the model is tested by test data set, which shows the actual and the corresponding predicted MOS values are correlated with correlation coefficient R of 95.7%. From these values we can say the developed model has good accuracy. When we see the importance of the features the download speed is the main IF followed by jitter, phone internal free space, phone RAM size, latency, and video resolution. The rest, gender, screen resolution, age, and upload speed has very small importance on the model. From the survey, the average satisfaction level for SM video services is 2.8, which shows the customers have a below fair satisfaction level. From the users' network preference, 85% of the participants use Wi-Fi networks for their SM video-related activities. From the above data, the download speed for most of the available Wi-Fi networks is not satisfactory, based on this finding Ethio telecom should work on this area to fill the gap between these large customers' demands and the actually available Wi-Fi service and study different ways to deliver Wi-Fi services at different places to its customer at a fair price. The other IF, users device parameters RAM size and the phone internal free storage space have a significant influence on the QoE of SM video services. From the survey, 68% of the participants clean their phones regularly whereas the rest do not have such a habit. Cleaning the mobile storage increases mobile performance this in turn increase the performance of the applications that is installed on the phone, creating such awareness is also one of the tasks to deliver the SM video services with the desired quality. The rest parameters, phone screen resolution, upload speed, age, and gender have less significance on the SM video streaming users' satisfaction level.

To the best of the author's investigation, there is no existing work in the literature that model the SMvideo streaming services QoE by incorporating users mobile phone parameters like RAM size, screen resolution, and free internal storage space, in addition to the commonly used QoS parameters. So, this work advances the developed model for Ethio telecom to give a new insight to improve its SM video streaming customers.

6.2 RECOMMENDATIONS

Based on this thesis analysis result, the following points are recommended as future works:

- Consider other SM applications other than the one used in this thesis and study their overall QoE.
- Consider SM services other than video streaming and study the users' feedback on different SM applications.
- Include other QoE IF in addition to the one used in this thesis and their impact on the overall users' satisfaction.
- This thesis uses an ensemble method to model the QoE, the model can be done by other ML methods.
- This thesis is limited only to Addis Ababa SM users. Users from different geographical places may have different satisfaction levels. So one study area in the future can be to incorporate SM users from other parts of the country.

BIBLIOGRAPHY

[1] Varinder Taprial and Priya Kanwar. *Understanding social media*. Bookboon, 2012.

[2] DataReportal-Global Digital Insights. *https://datareportal.com/social-media-users*. Online; accessed 21-October-2021.

[3] worldpopulationreview Ethiopea. *https://worldpopulationreview.com/countries/ ethiopia-population*. Online; accessed 21-October-2021.

[4] Standardization Sector and of ITU. *Series p: Terminals and subjective and objective assessment methods methods for objective and subjective assessment of speech quality*. 2013.

[5] datareportal. *https://datareportal.com/reports/digital-2021-ethiopia*. Online; accessed 21-October-2021.

[6] Paula Gilbert. *Ethiopians are online*. Online; accessed 16-December-2021.

[7] Aaron O'Neill. *Ethiopia: Age structure from 2010 to 2020*. Online; accessed 26-December-2021.

[8] Digis Weldu. "Machine Learning Based QoE Estimation Model for Video Streaming over UMTS Network." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2020.

[9] Menbere Asfaw. "Quality of Experience Model for Addis Ababa Voice Service Using Adaptive Neuro Fuzzy Inference Approach." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2020.

[10] Peter Reichl et al. "Towards a comprehensive framework for QoE and user behavior modelling." In: *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE. 2015, pp. 1–6.

[11] Pedro Casas et al. "YouTube & Facebook Quality of Experience in mobile broadband networks." In: *2012 IEEE Globecom Workshops*. IEEE. 2012, pp. 1269–1274.

[12] Pedro Casas et al. "Next to you: Monitoring quality of experience in cellular networks from the end-devices." In: *IEEE Transactions on Network and Service Management* 13.2 (2016), pp. 181–196.

[13] Gerardo Gómez et al. "YouTube QoE evaluation tool for Android wireless terminals." In: *EURASIP Journal on Wireless Communications and Networking* 2014.1 (2014), pp. 1–14.

[14] Asif Ali Laghari et al. "Impact of storage of mobile on quality of experience (QoE) at user level accessing cloud." In: *2017 IEEE 9th international conference on communication software and networks (ICCSN)*. IEEE. 2017, pp. 1402–1409.

[15] Beatrice Buberwa and Mercy Mbise. "A Model for Mapping Combined Effects of Quality of Service Parameters and Device Features on Video Streaming Quality of Experience." In: *Tanzania Journal of Science* 47.1 (2021), pp. 268–281.

[16] Pedro Casas et al. "Enhancing machine learning based QoE prediction by ensemble models." In: *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE. 2018, pp. 1642–1647.

[17] Sarah Wassermann, Nikolas Wehner, and Pedro Casas. "Machine learning models for YouTube QoE and user engagement prediction in smartphones." In: *ACM SIGMETRICS Performance Evaluation Review* 46.3 (2019), pp. 155–158.

[18] Sigit Haryadi. "The Concept of Telecommunication Network Performance and Quality of Service." In: (2018).

[19] Paul Coverdale et al. "Multimedia quality assessment standards in ITU-T SG12." In: *IEEE Signal Processing Magazine* 28.6 (2011), pp. 91–97.

[20] Khadija Bouraqia et al. "Quality of experience for streaming services: measurements, challenges and insights." In: *IEEE Access* 8 (2020), pp. 13341–13361.

[21] Abdelhamid Mellouk, Hai Anh Tran, and Said Hoceini. *Quality of Experience for Multimedia: Application to Content Delivery Network Architecture*. John Wiley & Sons, 2013.

[22] William R Klecka, Gudmund R Iversen, and William R Klecka. *Discriminant analysis*. Vol. 19. Sage, 1980.

[23] Mohammed Alreshoodi and John Woods. "Survey on QoE-QoS correlation models for multimedia services." In: *arXiv preprint arXiv:1306.0221* (2013).

[24] Addisu Shiferaw. "QoE Assessment Model for Addis Ababa LTE Video Streaming Service Using Machine Learning Techniques." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2020.

[25] ITU-T Study SERIES J(2005-2008). "Transmission of television, sound programme and other multimedia signals measurement of the Quality of Service ." In: (2000).

[26] Abera Reesom. "Quality of Experience Evaluation for Addis Ababa UMTS Enterprise Data Customers." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2018.

[27] Zelalem Fitiwe. *Quality of Experience Evaluation for Smartphone Video Streaming over 3G Network: In the case of Addis Ababa, Ethiopia*. MSC Thesis. 2019.

[28] Mohammed Alhamid. *Ensemble Models*. Tech. rep. Online; access 10-December-2021.

[29] Jason Brownlee. "Bagging and random forest ensemble algorithms for machine learning." In: *Machine Learning Algorithms* (2016), pp. 4–22.

[30] Ajay S Singh and Micah B Masuku. "Sampling techniques & determination of sample size in applied statistics research: An overview." In: *International Journal of economics, commerce and management* 2.11 (2014), pp. 1–22.

[31] worldpopulationreview Addis Ababa. *https://worldpopulationreview.com/countries/addisababa-population*. Online; accessed 21-October-2021.

[32] nPerf. *nPerf Internet speed test*. Tech. rep. Online; accessed 12-November-2021.

A

APPENDIX

A.1 ROC CURVE FOR THE FINAL MODEL

AUC value ranges from 0 to 1,the maximum AUC=1 means that the model is perfect in the differentiation between the specific class and the rest of the classes while AUC=0 means the model incorrectly classify all classes. As shown in the figure below AUC values of each class ranges from 0.97 to 0.99 which shows the model classify the classes with high accuracy.

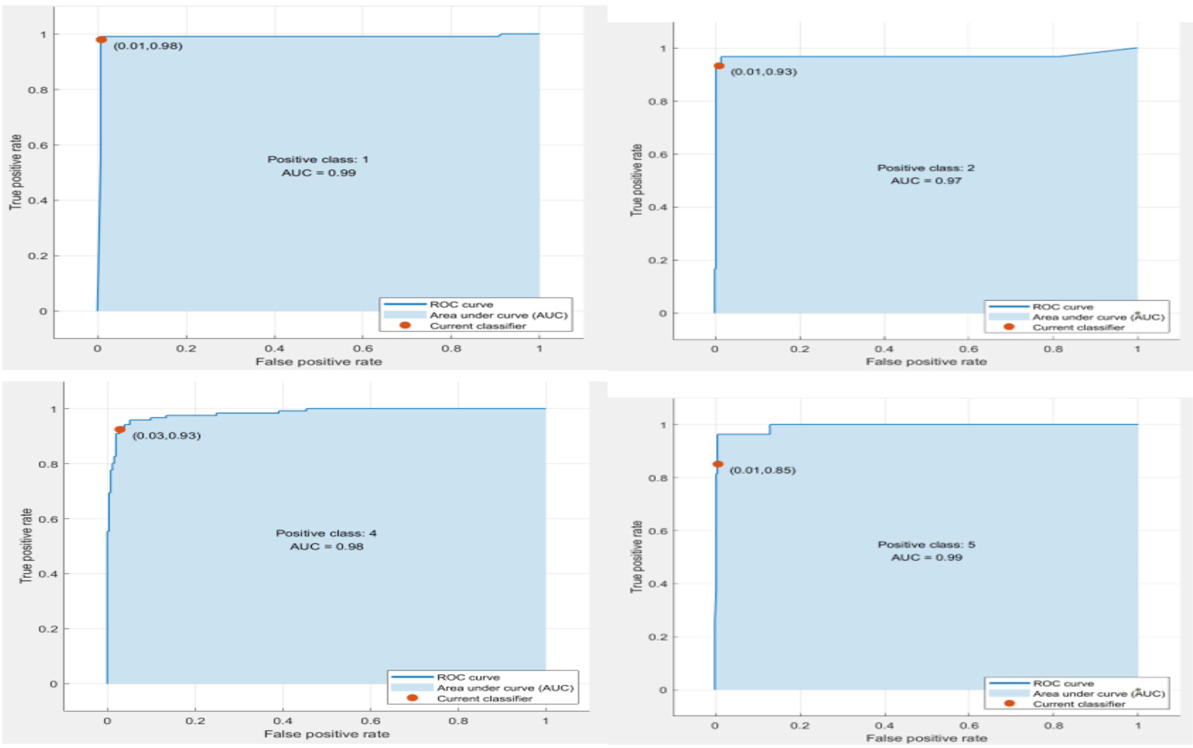


Figure A.1: ROC Curve for each Class

A.2 SURVEY QUESTIONER

Survey for Evaluating Social Media(SM) Quality of Experience

This survey is done to assess the Social media video services uses' satisfaction level by the existing network quality. During the survey, writing your name or phone number is not necessary. You are asked to fill the questioners after watching a video on selected social media video streaming applications (YouTube, Facebook and Tiktok). Question are organized into different categories, question number 1 to 4 is about your personal data, from 5 to 7 is to examine your Social media usage habits, from number 8 to 11 asks your mobile phone device parameters, Question number 12 asks your overall satisfaction level about the video you have watched, the next part of the questioner 13 is technical data filled from nPerf reading. Question 14 asks you to fill the video quality you have watched and finally the last question asks your habit on cleaning your phone storage.

1. Sex Male ☐ Female ☐

2. Age <20 ☐ 20-25 ☐ 26-30 ☐ 31-40 ☐ >41 ☐

3. Higher level of education

<Grade 10 ☐ Diploma ☐ 1st Degree ☐ 2nd degree ☐

4. Monthly incomes

<2,000 ☐ 2,000-5,000 ☐ 5,000-10,000 ☐ 10,000-20,000 ☐ >20,000 ☐

5. Rank the Social media you use for watching videos on your phone.

Facebook ___ YouTube___ Tik tok___

6. which network do you use most of the time for	Wi-Fi	Mobile data
YouTube		
Facebook		
Tiktok		

7.What is your average data usage per month in Mb (from your mobile setting->data usage) for 2 month	Mobile data	WI-FI	Mobile data	WI-FI
YouTube				
Facebook				
Tik tok				

The following data (9-11) are from your phone device setting

8. What is your phone RAM size ___________

9. what is your Phone screen resolution (width*height) in Pixel per inch ___________

10. What is your phone free internal storage space ___________

Please watch your preferred video on YouTube, Facebook or tik tok and fill your satisfaction

(please use the following rating values 1- 5)

11. What is your phone's model ___________

(5 - Extremely satisfied, 4 - Moderately satisfied, 3 - Slightly Satisfied, 2 - Neither satisfied nor dissatisfied, 1- Dissatisfied)

12. How much are you satisfied by the video you watch ___________

13. Write the network QoS parameters from nPerf reading

Download speed ___________Jitter ___________Latency ___________Upload Speed___________

14. What is the video quality from video setting (240p, 360,720p,1040p) ___________

15. Do you clean your phone's internal storage space regularly?

Yes ☐ No ☐

Thank you for your response

A.3 PUBLISHABLE MANUSCRIPT

QoE Model for Social Media Video Streaming Service Using Ensemble Method The Case of Addis Ababa, Ethiopia

1st Wubalem Tadesse Muse
School of Electrical and Computer Engineering
Addis Ababa University
Addis Ababa, Ethiopia
wubalemtadesse9@gmail.com

2nd Ephrem Teshale Bekele (PhD): Assistant Professor
School of Electrical and Computer Engineering
Addis Ababa University
Addis Ababa, Ethiopia
ephremteshale@gmail.com

***Abstract*—Nowadays, the number of Social Media (SM) users are increasing tremendously worldwide. An increase in the number of smartphone users and Internet coverage helps people to expand their netwcfxcorks on SM. There are many stakeholders that involved in an end-to-end service delivery of SM. These include the Telecom network providers, the SM application owners, and end-user devices performance. Having a good network Quality of Service (QoS) may not guarantee good service quality on the customers' side. The quality of a given service perceived by the end-user, which is Quality of Experience (QoE) is a broad term and influenced by many Influencing Factors .**

The main focus of this thesis is to provide a QoE model for social media video streaming services by taking download speed, upload speed, latency, and jitter from network QoS parameter, users' phone Random Access Memory (RAM) size, free internal storage size and screen resolution from users' mobile device parameters and video resolution from application parameter.

The developed model is based on an ensemble technique which is a Machine Learning (ML) based approach. The model has good accuracy, which is 94.1% accuracy. In addition to the accuracy, based on the importance of each input feature to the final model, download speed takes the main influencing share by 52.357% from the total input parameters and from the users' mobile device parameter free internal storage space has 10.784% and mobile RAM size 9.4% on the final QoE model.

***Index Terms*—QoE, QoS, Machine Learning, Ensemble method.**

I. Introduction

Many papers gives different definition for SM, the book written by Varinder and Kanwar defines Social medias as "All web based applications which allow for creation or exchange of user-generated content and enable interaction between the users can be classified as Social Media" [1]. It can be seen that Social medias are designed to be Internet-based to provide different services for the users. Usually, SM users from different geographical locations with different social, religious, educational backgrounds are connected to pursue mutual interests via web-based applications on their desktops, laptop, smartphones, or tablets SM users.Delivery of all web-based systems to the customers depends on the Telecom network provided, which means social media services directly rely on the available Telecom network.

The services types delivered by social media are increasing from time to time. Some of the services include photo sharing, video sharing, blogs, products/services sharing, social networks, social gaming, and news sharing. In addition to the type of services, the number of social media users show a significant increase in the past few years. Based on Global Digital Overview statistical data [2] there are 4.48 billion social media users around the world as of July 2021 and there is a 13.1% annual increment. In the case of Ethiopia, there are 6.7 million social media users as of January 2021, with 8.1% annual increments.

Telecom network providers provide network service to deliver Internet services to their customers. In addition, to provide the infrastructures the Telecom operators examine their services whether they are delivered with the expected quality or not. There are different Key Performance Indicators (KPI) to measure the QoS for different services. QoS is a technical concept that is expressed and understood in terms of networks and network elements. The overall quality of service from the users' perspective cannot be determined only by considering only QoS, there are many other technical and non-technical factors that influence the customers' satisfaction. The technical system comprises a chain of components (user device performance, Telecom network, and social media application server) and non-technical factors include expectations, personal background, demography, urgency, task, gender, age, etc. This shows QoE assessment is not an easy task.

Out of many services provided by Telecom operators SM services rely on the Internet provided by the Telecom operator.When looking the SM delivery chain, there are many parties that influence the overall QoS provided. Figure 1 demonstrates an end-to-end chain of the SM delivery system. The end user's background like social status, demography, educational level, SM usage history and soon have a great impact on the customers' satisfaction level.The other factor that can influence the SM satisfaction level is different end-Users Equipment (UE) parameters. The third one that influences the satisfaction level of the SM users', which is under full control of the Telecom operator is the network part. This part consists of the Radio Access Network(RAN) and Core Network (CN) part including the transmission network, which has a direct impact on the QoS delivered. The other factor is the SM application server parameters which also affect the overall service satisfaction level. Since there are many parties

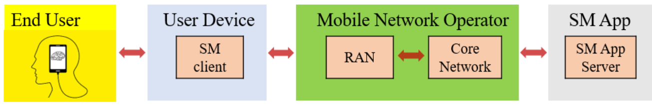


Fig. 1. Social Media chain

on the end-to chain of SM, delivering a good network QoS will not assure a service with the best satisfaction level. Therefore, other factors besides the network QoS part that have an impact on the overall service quality need to be assessed. This help to pinpoint the possible factors that influence the QoE and to work on possible solution beside guaranteeing QoS for an E2E delivery of quality service.

The rest of the paper is organized as follows. Section II presents an overview on some of the aforementioned papers relative to this study. Section II explains the methodology followed in developing the model. Section III discuss detail about the results achieved and their interpretations. Finally, the conclusion and future works points are discussed in section IV.

II. Literature Review

There are many research papers about QoS and QoE of different service types on different network architecture. The approaches followed by these papers also vary based on the method deployed for analyzing QoE. The paper by W.Digis [3] examines the QoE of video streaming services specifically YouTube download streaming QoE. ML algorithms (Artificial Neural Network, k-Nearest Neighbors, and Random Forest) are used for developing QoE estimation models using QoS features. For QoS data gathering, the author use ACQUA which is a free Android tool that measures the user-level network traffic like signal strength, download and upload bandwidth, round-trip time, jitter and etc. A total of 230 participants are included in the data gathering process and their satisfaction level is collected by MOS values. The collected data is applied to dif ferent ML algorithms to model QoE. Out of which Random Forest perfectly fit the QoE prediction. This work takes only network QoS as the only IF of the model.

The other paper by A.Menbere [4] models the QoE of UMTS voice service using the Adaptive Neuro-Fuzzy Inference Approach. The data which is 3G voice network KPI is gathered from the Network Management System. For the subjective information, the author provide questionnaires' to evaluate the subjective perception of quality. Using these two data as an input to Adaptive Neuro-Fuzzy Inference system which is ML technique a new QoE model is developed for the 3G and 2G voice calls services. The users service quality perception is determined by many factors. Beside the QoE models that incorporate network parameters as the main input, the paper by Reichl et al. [5] tries to fill the gap for by integration of user behavior characteristics and user context factors, as well as the consideration of appropriate temporal scales. The authors discuss a set of use cases, like traditional QoE, user behavior, charging and pricing models, the impact of user characteristics and problems related to energy consumption. They separate the technical perspective of the system from the user perspectives and the context related behavioral perspective relevant for the service provider.

The other approach for QoE is assess the users feedback by providing network emulator instead of the actual network. Casas et al. in [6] perform a survey to collect feedback from a different participant which uses YouTube and Facebook. Based on MOS values at different download speeds. Instead of the actual network, they use a network emulator to control and vary the bandwidth between 64kps and 4, 096kps and capture all the packets before and after the emulator for analysis of YouTube videos. They use the YoMo free mobile tool for monitoring application layer parameters like the number and duration of video stall and they gather feedback from the participants on different download speeds. From their analysis, they show YouTube QoE is highly sensitive to downlink bottlenecks and video bit-rate and DBW has a strong correlation with the end-users' satisfaction and acceptability. Network emulation is also applied to measure QoE of different applications like Facebook, YouTube, Google maps, web browsing, and Whatsapp . The other paper [7] also repeated the QoE measurements in terms of MOS on the field and shows how customers satisfac tion is affected by different network parameter values. This paper also includes the acceptance rate for different MOS values.

Beside modeling QoE by taking different IFs as an input parameters Gomez et al. In [8] perform a survey by making participants watch videos on Wi-Fi and UMTS network, in the meantime the network QoS parameters are measured. Then they perform a mapping from the network QoS to Application QoS (initial buffer time, rebuffing frequency, and mean buffer time) then from Application QoS to in terms of MOS.

The other papers focus on different QoE IF besides network parameters. Laghari et al in [9] shows how QoE is affected by the internal storage free space when users access cloud services from their mobile phones. When end users are accessing cloud services enormous amount of temporary/cache data is generated by apps, so internal storage of mobile devices is filled quickly. The mobile device without any space in internal storage has a huge impact on the performance when accessing the cloud services, which degrade the QoE of end-users for particular cloud apps and services.

The other paper by Buberwa and Mbise [10] proposed a QoE model by mapping combined effects from both network and device parameters on video streaming services. From the network QoS, they took packet loss, packet reordering, and delay were emulated using network emulator. Through analysis of variance, they found that packet loss had the highest impact, followed by device features, reordering, and delay on the video QoE , and from the combined effects, two-way interactions and three-way interactions had significant effects on the video QoE .

Different ML algorithms are used in many papers to model QoE [11] by Casas et al. performs QoE prediction for smartphone applications using different ML algorithms. They incorporate single models (Decision tree, Naïve Bayes, Neural network, Support vector Nearest neighbor) and ensemble-based (random forest, bagging, boosting, and stacking) methods. Based on their results ensemble methods shows the highest accuracy in prediction and on the overall classification performance. The other paper by Wassermann et al. in [12] conduct survey analysis on large scale worldwide ,they suggest that QoE -based monitoring of YouTube mobile can be realized through ML models with high accuracy, relying only on network related features and without ac cessing any higher-layer metric to perform the eshtimations. They compare different ML-based models and they suggest random forest predicts QoE with a better performance. They also show the advantages of ensemble-learning techniques with respect to simpler models in terms of prediction accuracy.

From the literature review, it is seen that QoE modeling is influenced by many factors including the network QoS, device performance, and users' emotional behaviors. Most of the models incorporate the influence of the network QoS parameters and only a few papers try to study the impact of different users' device performance parameters on the users' satisfaction. So, in this research paper, the effect of user device parameters in addition to the network parameters are considered in the QoE model.

III. Methodology

The method used in this thesis aimed to provide a good QoE model with reasonable accuracy. To do so the first step was the data collection step by surveying selected SM users. The collected data consist of different metrics that influence the overall QoE of SM video streaming services. From network QoS download speed, upload speed, latency, and jitter are collected using nPerf which is a free mobile application for network QoS test. The other metrics are from users' mobile phone specifications which include mobile RAM size, screen resolution (number of pixels per inch), and the avail able internal storage size. From the application, the video resolution is taken as one metric.

After data gathering, the next step was data pre-processing which include data cleaning, mismatch handling, and data preparation for modeling. Then the selected ML base technique (ensemble method) is applied to train the model with the selected data set. After model training the last step will be to check the model validity and propose the final model. Figure 2 shows the method used in this thesis work.

DATA COLLECTION : Data is collected by distributing questioners to selected SM users. The questioners are organized into four main parts.

• The first part is to find the SM applications that are mostly used by the participants.

• The second part aims to gather participant's personal data plus their device parameter.

• The third one is to collect the network QoS parameter.

• The last part is the participants' feedback according to MOS value.

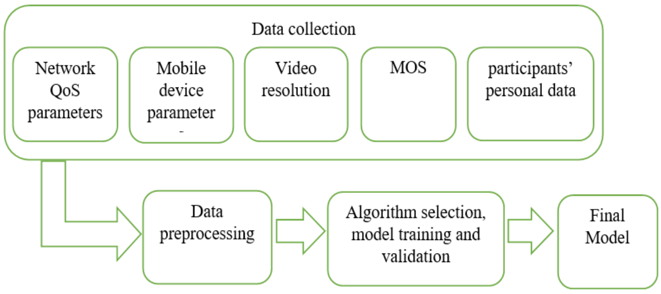


Fig. 2. Research Method Used

The first part of the questioner focus to gather the top used SM applications that are mostly used by the survey participants and their preferred network type used for each application i.e. whether they use cellular data (3G or 4G) or Wi-Fi. In addition, their social media usage history is recorded from their cell phone device for the past two months. This is done to differentiate participants that use SM frequently from the one that uses rarely. Participants who have no or little SM usage history based on their data usage history on their mobile phone but have usage history on other devices like laptops, tablets, or desktop computers are excluded from the survey. The general data collection setup is shown in Figure 3. On the second part of the questioner, the participant's personal data plus their device parameters are collected. The personal data include the participants' education, income, age, and gender data from different location of Addis Ababa. The device parameters include the mobile phone RAM size, mobile phone screen resolution in a number of pixels per inch (PPI) which is calculated by dividing diagonal pixel resolution divided by diagonal size, internal storage free space in megabytes and the phone model. The other issue to consider is to decide the

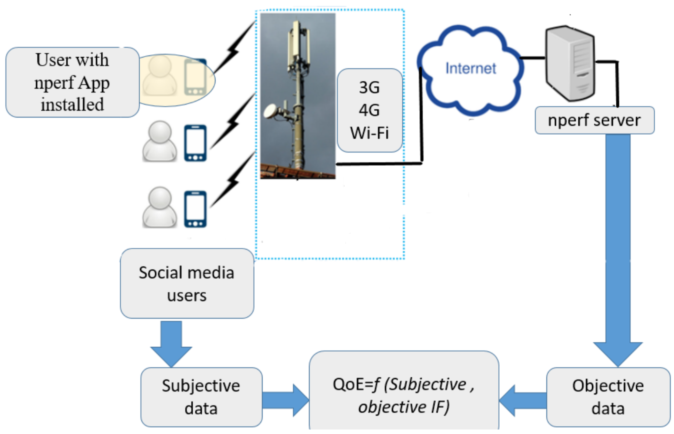


Fig. 3. Data Collection Setup

number of participants on the survey. The sample size used for research purposes should be carefully fixed so that it will be adequate to draw valid and generalized conclusions. There are different ways to determine the sample size [13] some of the methods are sampling

• Using a Census for Small Populations by taking the entire population as the sample.
• Using a Sample Size of a Similar Study.
• Using Published Tables which provide the sample size for a given set of criteria.
• Using Formulas to Calculate a Sample Size.

There are many formulas used on different research papers to find the sample size. in this paper, a Simplified Formula for Proportions noted on 1 is used to calculate the number of samples.

$$n = \frac{N}{1 + (N * e^2)} \tag{1}$$

Where N= Population size, e= precision, and n= sample size Since the focus area of this work is only in Addis Ababa, Ethiopia. The population size is determined by the population of Addis Ababa. Based on [15] the current population of Addis Ababa in 2021 is expected to be 5 Million, and this will be the value of N. The other parameter is the level of precision e, sometimes called sampling error. It is the range in which the true value of the population is estimated to be. This range is often expressed in percentage points, by taking e = ±5%and N = 5 million, n, which is the minimum sample size become 400. After deciding the number of participants the survey is done by randomly selecting participants which are believed to best represent the society. The participants are selected from different sub-cities of Addis Ababa (Bole, Nifas Silk, Akaki Kality, Yeka, and Arada sub-cities) and also with different age groups, educational backgrounds, gender, and income level. Out of the total collected 473 data 80 of the participants filled the survey more than once within different time of the day and different place. The other parameter collected is network QoS parameters, which include download speed, upload speed, jitter, and latency. To collect these data free mobile application nPerf is used. This tool measure uploads and download bit rate by downloading or uploading a binary file with a simultaneous connection for a few seconds. The detailed working principle for measuring different QoS is described in [14]. To record these network QoS parameters the selected participants are subjected to watch a video of their interest on their preferred SM (YouTube, Facebook, or Tiktok) for a minute. In the meantime, the nPerf application runs to measure QoS parameters at the time of the video play. After watching the video, on the last part of the questioner, the participants are asked to give their feedback based on the MOS rating from 1 to 5.

DATA PRE-PROCESSING : The data pre-processing step focuses on the data preparation for training the data set with the selected algorithm. A total of 10 features are collected from the survey done. On the data collection process of the survey, some of the participants left the device parameters RAM, and screen resolution null. These missing data are completed by searching the parameters from the Internet by using the mobile phone model, which is one provided by the participants. This makes the missing data complete. The other issue is data type variation, the collected data contains both numeric and non-numeric categorical data types. Non-numeric categorical features age group and sex are used as input parameters for the modeling. These parameters are converted to dummy variables for the sake of analysis. The other input parameter, device resolution is two-dimensional data which is pixel per inch for the length and width of a mobile device. For the analysis purpose, the multiple of the two dimensions is used as an input feature.

MODEL TRAINING AND VALIDATION :
From the supervised ML category, Classification predictive modeling is used to map input variables to discrete output variables. In our classification problems, the learning algorithm learns a function to map inputs to outputs where the output value is a discrete class MOS values.The training data set is trained by Bootstrap Aggregating (Bagging) method, which is a ML based ensemble technique using MATLAB 9.4 Release R2018a Statistics and Machine Learning Toolbox. For developing the final model the total data set is divided into training and validation set,which is 80% are the training data and the rest 20% for validation. For the ensemble method, the two hyper parameters used in this model are the number of decision trees and the depth of each decision tree. The maximum values of the two hyper parameter of this algorithm are:

• The depth of decision tree = the number of training data set =374.
• Number of decision trees = 500, which is the maximum possible value.

By taking these values, there are many possible combinations to train the model. That is $500 * 374 = 187,000$, there are $187,000$ possible combinations of hyper parameters. Matlab hyper parameter optimization within the Classification Learner app is used to automate the selection of hyper parameter values. Different combinations of hyper parameter values are tried by using an optimization scheme to minimize the model classification error and return a model with the optimized hyper parameters.

The final QoE model performance is evaluated by Accuracy, Precision, Recall and F-score based on confusion matrix, and using ROC curve. 20% of the total data set is used to measure the errors between the estimated and the actual collected MOS values. In this paper, a Regression analysis is used to check how fit the model is by Using the validation data set.

IV. Results Analysis and Interpretations

From hyper parameter optimization and manually testing different combinations of hyper parameters, a model with the highest accuracy is selected as a final model. The final model has an accuracy of 94.1% with a miss classification cost of 22 and hyper parameter values, 454 decision trees each having a depth of 50. Table I demonstrate the final model parameters with their corresponding values. Figure 4shows

TABLE I
RESULT OF THE TRAINED MODEL

Parameter	Value
Accuracy	94.1
Misclassification cost	22
Training time	54.78 sec
Ensemble method	Bag
Learner type	decrision tree
Maximum number of split	50
Number of learner	454

correctly classified observations of each class on the diagonal cells and miss classified observations on the other cells. When considering class 1 that is MOS value of 1, 98.1%are correctly classified and the rest 1.9% are classified as MOS 2. For class 2, 3.3% are classified as MOS 1 and the other 3.3% are classified as MOS value of 3 and the rest 93.3% are correctly classified as MOS value of 2. For class 3, 1.1% are classified as MOS value 1 and 5.4% are classified as MOS value of 4, and the rest 93.5% are correctly classified. For class 4, 92.6% are correctly classified whereas the rest are misclassified. finally, class 5 has 11% misclassification.

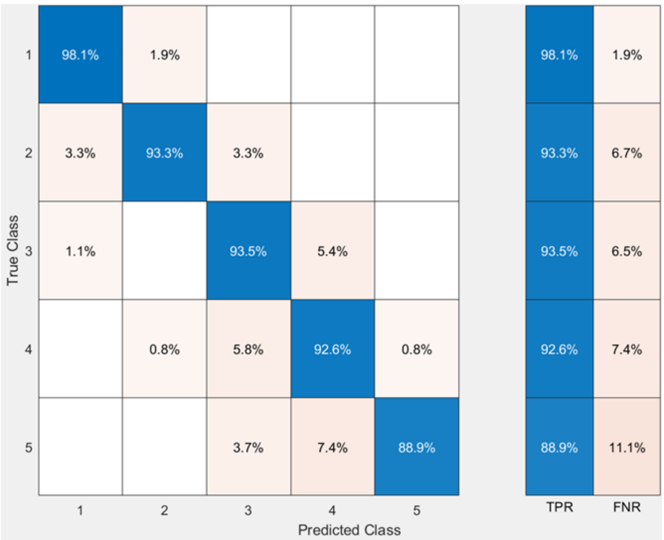


Fig. 4. Number of correctly classified and miss classified class

MODEL PERFORMANCE EVALUATION

A confusion matrix is one of the performance measurement technique for ML classification models. Table 5 shows the trained model confusion matrix values. From the confusion metric the model accuracy for each class is computed, each class corresponds to the corresponding MOS values. MOS values 1 and 5 have the highest accuracy value 98.93% each and MOS value 4 have relatively less accuracy than the others, which is 95.72%.Figure 6 shows the the accuracy of each class. The other evaluation metric is precision, the precision

class		Predicted	
	Actual	T	F
1	T	102	2
	F	2	268
2	T	28	3
	F	2	341
3	T	86	9
	F	6	276
4	T	112	7
	F	9	246
5	T	24	1
	F	3	346

Fig. 5. Confusion Matrix Values

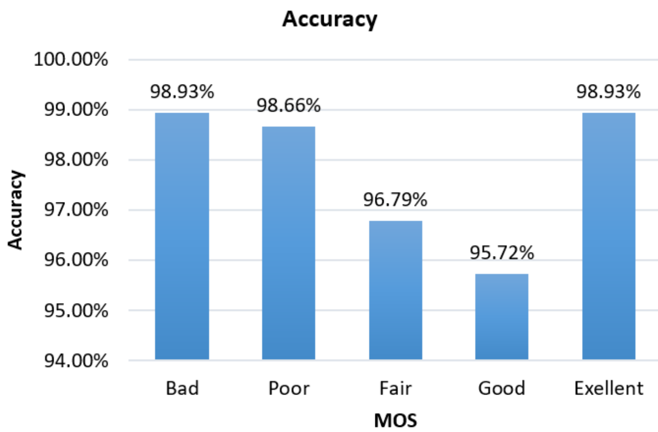


Fig. 6. Model Accuracy

of each class describes how much the model classifies positive values correctly. From Figure 7) MOS value 1 has the highest precision value 98.08% and MOS value 5 has less precision than the others' which shows the model classifies MOS value 1 more likely, Whereas the model classifies MOS values 5 as less likely. That means some values of them are classified as if they belong to other MOS values. The other evaluation metric, recall shows how much the model can predict different classes from all the positive classes. Which is shown in Figure 8 again the model predicts MOS Value $1 by 98.08\%$. which is the highest and MOS value 2 has the lowest value.

The other metric F-score shows the average of the precision and recall value which shows how can the model classify and predict positive class. F-score values range from 0 to 1, if the F-score value approaches to 1 indicate model is a good model. In this model the F-score ranges from 0.92 to 0.98 as shown in figure 9.

The ROC curve is another evaluation metric that plot a probability curve of True Positive Rate (TPR) against False Positive Rate (FPR) at various threshold values. AUC is used as a measure of the ability of a model to distinguish between classes and is used as a summary of the ROC curve. AUC value ranges from $0 to 1$,the maximum AUC=1 means that the

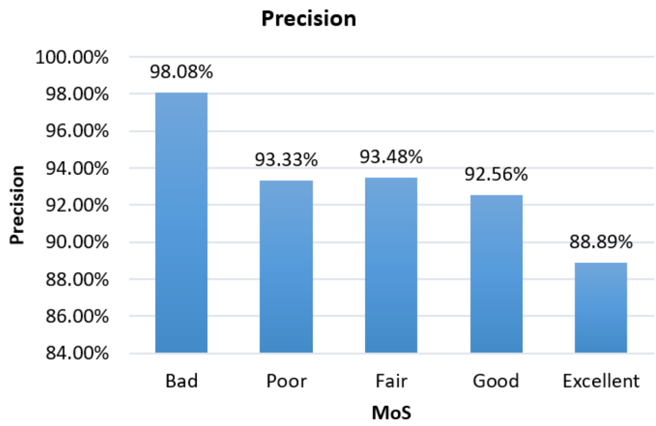


Fig. 7. Model Precision

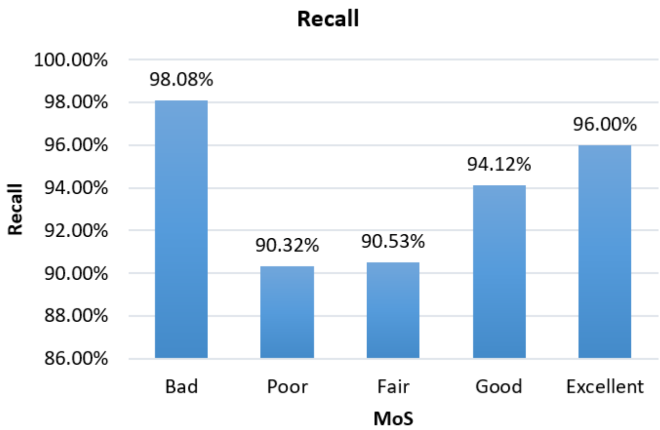


Fig. 8. Model Recall

model is perfect in the differentiation between the specific class and the rest of the classes while AUC=0 means the model incorrectly classify all classes. ROC curve for class 3 is shown in Figure 10 with AUC of 0.98 this value shows the model differentiate Fair satisfactory level from the other MOS values with high accuracy. As shown in Table II AUC values of each class ranges from $0.97to0.99$ which shows the model classify the classes with high accuracy.

TABLE II
AUC of Each Class

MoS	1	2	3	4	5
AUC	0.99	0.97	0.98	0.98	0.99

MODEL VALIDATION

After developing our QoE model, it is necessary to validate the model using another test data set. The test data which is 20% of the total data set is used to measure the errors between the estimated and the actual collected MOS values. In this paper, a Regression analysis is used to check how fit the model is by using the validation data. Table III shows the model has good accuracy with R, R square, adjusted R square,

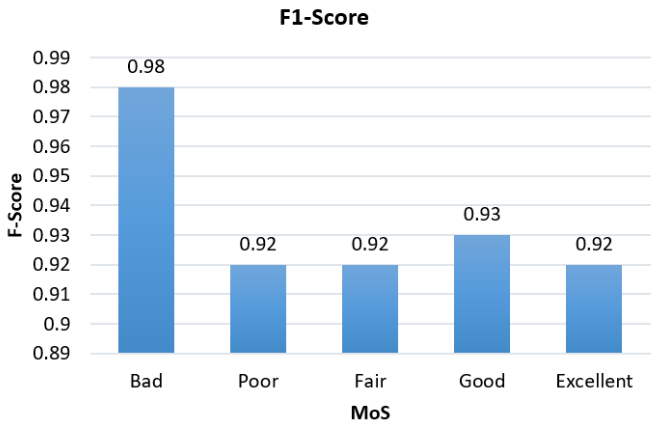


Fig. 9. F-score Values for each class

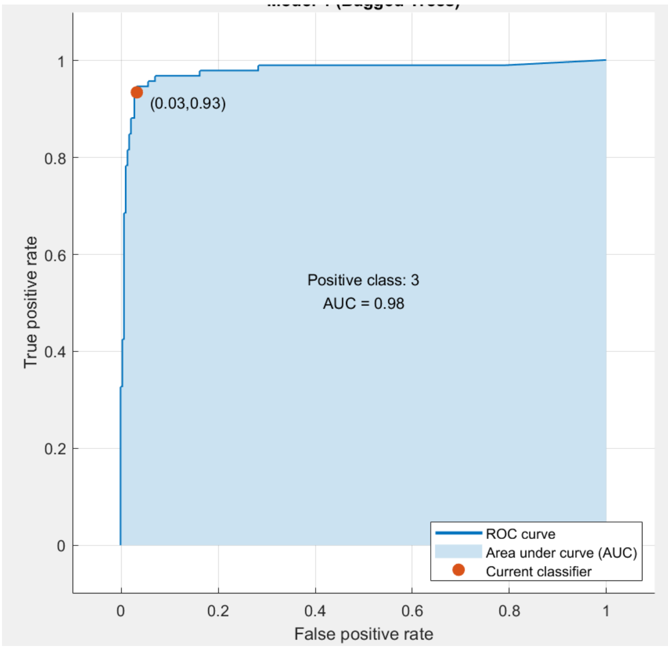


Fig. 10. ROC Curve for Class 3

and RMSE of $0.957, 0.916, 0.915, and 0.340$ respectively for 93 test data set. In addition, Figure 11 shows how fit are the actual MOS values and their corresponding predicted value by the developed model. From the above validation tests the model is good for our data set.

TABLE III
Regression Between Validation and the Corresponding Predicted Data

R	R-Squared	Adjusted R-Squared	RMS	Observation
0.957	0.916	0.915	0.34	93

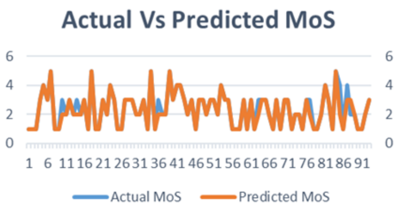


Fig. 11. Actual and Predicted line graph

FEATURE IMPORTANCE ON THE MODEL

Different features have a different level of contribution to the model. The feature importance plot in Figure 12 shows the input features' rank based on their contribution to the model. The feature with a high feature importance score has high predictive power.

Download speed is the main IF of the model followed by jitter, phone internal free space and phone RAM size, latency, and video resolution. The rest parameters, gender, screen resolution, age, and upload speed have very small importance in the model. The feature importance score for each feature is listed in figure 12

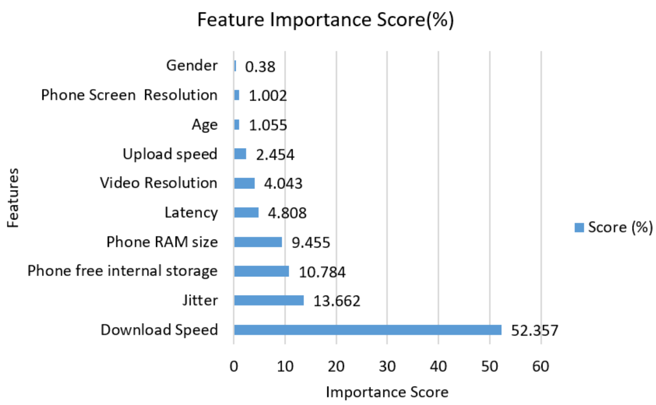


Fig. 12. Feature Importance Plot

From the results obtained in this paper, the selected ensemble method technique, which is Bagging method, has good accuracy. The impact of each feature on the final model is described on feature importance score values. Which shows how much the QoE values are influenced by the IF. Apart from this, from the survey data, the overall satisfaction level of SM users is below fair. Generally, this thesis meets the initial objective, i.e to model QoE using different , demonstrates each features impact on the total satisfaction level, and assess the overall SM video streaming users' satisfaction level.

In doing so, there were different challenges like conducting the survey, on model training to find optimum hyper parameter values; it needs to train the data set on different iterations, which takes hours to finish one large iteration. To overcome this challenge, the maximum number of iterations has to be 1000, which improves the hours to be 8 hours.

V. Conclusions and Recommendations

CONCLUSIONS

The main purpose of this thesis is to propose a model that is capable of predicting the user perceived quality of SM video streaming services. Users perceived quality can be influenced by many factors, out of these, network QoS parameters, users' mobile phone parameters, and application-related parameters are included in this paper as an IF. From network QoS download speed, upload speed, latency, and jitter are selected as an input feature for the model. From mobile device parameters, phone RAM size, phone internal free space, and users' phone screen resolution are selected as input features. From the application side since the only concern of this thesis is SMvideo services video resolution is selected as an application parameter. SM users' satisfaction level is collected by providing questioners for selected users that are believed to be a good representative of the actual population. Each participant of the survey is subjected to watch their preferred video on either YouTube, Facebook, or Tiktok which are the SM that are selected for this thesis. On the survey, while the selected SM users are watching the videos, the network QoS parameters are collected by a network performance measuring tool called nPerf. In addition to these parameters, the participants' cell phone parameters are collected from each participant. An ensemble method which is a ML-based technique is used to develop the model. The proposed SM video streaming QoE model has an accuracy of 94.1%. further, the model is evaluated by accuracy precision and recall which are performance measurement techniques for ML classification models. The proposed model has an accuracy of $98.93\%, 98.63\%, 96.73\%, 95.7\% and 98.93\%$ for MOS values $1 to 5$ respectively. Further, the model is tested by test data set, which shows the actual and the corresponding predicted MOS values are correlated with correlation coefficient R of 95.7%. From these values we can say the developed model has good accuracy. When we see the importance of the features the download speed is the main IF followed by jitter, phone internal free space, phone RAM size, latency, and video resolution. The rest, gender, screen resolution, age, and upload speed has very small importance on the model. From the survey, the average satisfaction level for SM video services is 2.8, which shows the customers have a below fair satisfaction level. From the users' network preference, 85% of the participants use Wi-Fi networks for their SM video-related activities. From the above data, the download speed for most of the available Wi-Fi networks is not satisfactory, based on this finding Ethio Telecom should work on this area to fill the gap between these large customers' demands and the actually available Wi-Fi service and study different ways to deliver Wi-Fi services at different places to its customer at a fair price. The other IF, users device parameters RAM size and the phone internal free storage space have a significant influence on the QoE of SM video services. From the survey, 68% of the participants clean their phones regularly whereas the rest do not have such a habit. Cleaning the mobile storage increases mobile performance this in turn increase the performance of the applications that is installed on the phone, creating such awareness is also one of the tasks to deliver the SM video services with the desired quality. The rest parameters, phone screen resolution, upload speed, age, and gender have less significance on the SM video streaming users' satisfaction level.

To the best of the author's investigation, there is no existing work in the literature that model the SM video streaming services QoE by incorporating users mobile phone parameters like RAM size, screen resolution, and free internal storage space, in addition to the commonly used QoS parameters. So, this work advances the developed model for Ethiotelecom to give a new insight to improve its SM video streaming customers.

RECOMMENDATIONS

Based on this thesis analysis result, the following points are recommended as future works:

- Consider other SM applications other than the one used in this thesis and study their overall QoE.
- Consider SM services other than video streaming and study the users' feedback on different SM applications.
- Include other QoE IF in addition to the one used in this thesis and their impact on the overall users' satisfaction.
- This thesis uses an ensemble method to model the QoE, the model can be done by other ML methods.
- This thesis is limited only to Addis Ababa SM users. Users from different geographical places may have different satisfaction levels. So one study area in the future can be to incorporate SM users from other parts of the country.

REFERENCES

[1] Varinder Taprial and Priya Kanwar. Understanding social media. Bookboon, 2012.

[2] DataReportal-Global Digital Insights. https://datareportal.com/social-media-users. Online; accessed 21-October-2021.

[3] Digis Weldu. "Machine Learning Based QoE Estimation Model for Video Streaming over UMTS Network." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2020.

[4] Menbere Asfaw. "Quality of Experience Model for Addis Ababa Voice Service Using Adaptive Neuro Fuzzy Inference Approach." MSC Thesis. Addis Ababa Institute of Technology School of Electrical and Computer Engineering, 2020.

[5] eter Reichl et al. "Towards a comprehensive framework for QoE and user behavior modelling." In: 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX). IEEE. 2015, pp. 1–6.

[6] Pedro Casas et al. "YouTube & Facebook Quality of Experience in mobile broadband networks." In: 2012 IEEE Globecom Workshops. IEEE. 2012, pp. 1269–1274.

[7] Pedro Casas et al. "Next to you: Monitoring quality of experience in cellular networks from the end-devices." In: IEEE Transactions on Network and Service Management 13.2 (2016), pp. 181–196.

[8] erardo Gómez et al. "YouTube QoE evaluation tool for Android wireless terminals." In: EURASIP Journal on Wireless Communications and Networking 2014.1(2014), pp. 1–14.

[9] Asif Ali Laghari et al. "Impact of storage of mobile on quality of experience (QoE) at user level accessing cloud." In: 2017 IEEE 9th international conference on communication software and networks (ICCSN). IEEE. 2017, pp. 1402–1409.

[10] Beatrice Buberwa and Mercy Mbise. "A Model for Mapping Combined Effects of Quality of Service Parameters and Device Features on Video Streaming Quality of Experience." In: Tanzania Journal of Science 47.1 (2021), pp. 268–281.

[11] Pedro Casas et al. "Enhancing machine learning based QoE prediction by ensemble models." In: 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS). IEEE. 2018, pp. 1642–1647.

[12] Arah Wassermann, Nikolas Wehner, and Pedro Casas. "Machine learning models for YouTube QoE and user engagement prediction in smartphones." In: ACM SIGMETRICS Performance Evaluation Review 46.3 (2019), pp. 155–158.

[13] Ajay S Singh and Micah B Masuku. "Sampling techniques & determination of sample size in applied statistics research: An overview." In: International Journal of economics, commerce and management 2.11 (2014), pp. 1–22.

[14] nPerf. nPerf Internet speed test. Tech. rep. Online; accessed 12-November-2021.

[15] worldpopulationreview AddisAbaba. https://worldpopulationreview.com/countries/addisababa-population. Online; accessed 21-October-2021.