SENTIMENT ANALYSIS MODEL FOR ONLINE PUBLIC PARTICIPATION FORUMS

OMELA MANASES MALACHI

A Thesis Submitted to the Institute of Postgraduate Studies of Kabarak University in Partial Fulfilment of the Requirements for the Award of Master of Science in Information Technology

KABARAK UNIVERSITY

NOVEMBER, 2023

DECLARATION

1. I do hereby declare that:
   i) This thesis is my own work and to the best of my knowledge, it has not been presented for the award of a degree in any university or college.
   ii) That the work has no incorporated material from other works or a paraphrase of such material without due and appropriate acknowledgement
   iii) That the work has been subjected to processes of anti-plagiarism and has met Kabarak University 15% similarity index threshold
2. I do understand that issues of academic integrity are paramount and therefore I may be suspended or expelled from the University or my degree may be recalled for academic dishonesty or any other related academic malpractices.

Signature: …………………………………….. Date: …………………………

Omela Manases Malachi

GMI/N/2517/05/18

RECOMMENDATION

To the Institute of Postgraduate Studies and Research:

The research thesis entitled "**Sentiment Analysis Model for Online Public Participation Forums**" written by **Omela Manases Malachi** is presented to the Institute of Postgraduate Studies and Research of Kabarak University. We have reviewed the thesis and recommend it to be accepted in partial fulfilment of the requirement for the Award of the degree of Master of Science in Information Technology.

Signature: …………………………………….. Date: …………………………

Dr. Moses M. Thiga

School of Science Engineering and Technology

Kabarak University

Signature: …………………………………….. Date: …………………………

Dr Nelson Masese

School of Science Engineering and Technology

Kabarak University

ACKNOWLEDGEMENT

Gratitude is the most equitable thing to those who do well to others; I am humbled, grateful and indeed thankful to the following persons for their kind contribution which helped accomplish this work. First and foremost, I give glory to God who has taken care of me this far. I also extend my sincere appreciation to my supervisors, Dr Moses Thiga and Dr. Nelson Masese for their kindness and well executed professional and intellectual guidance offered to me while writing this thesis. Special thanks to my lovely family for their patience while I was away looking for this second degree. Finally, to my course mates Antony, Mercy and Daisy, I feel honoured to have shared learning space with such great minds.

DEDICATION

I wish to dedicate this work to my beloved parents for their patience and support during times of great apprehension and work.

TABLE OF CONTENTS

DECLARATION .... **ii**
RECOMMENDATION .... **iii**
COPYRIGHT .... **iv**
ACKNOWLEDGEMENT .... **iv**
DEDICATION .... **vi**
TABLE OF CONTENTS .... **vii**
LIST OF FIGURES .... **xi**
LIST OF TABLES .... **xiii**
ACRONYMS AND ABBREVIATIONS .... **xiv**
DEFINITION OF TERMS .... **xv**
ABSTRACT .... **xvi**
CHAPTER ONE **1**
INTRODUCTION **1**
1.1 Background to the Study 1
1.1.1 Sentiment Analysis 4
1.2 Statement of the Problem 7
1.3 Objectives of the Study 8
1.4 Research Questions 9
1.5 Justification of the Study 9
1.6 Scope of the Study 10
1.7 Limitations of the Study 10
1.8 Assumptions of the Study 11
CHAPTER TWO **12**
LITERATURE REVIEW **12**
2.1 Introduction 12
2.2 Sentiment Analysis Overview 12
2.3 Challenges facing Acquisition of Sentiments in public participation forums 12
2.4 Sentiment analysis Models and Tools 14
2.4.1 Sentiment Analysis Models 14
2.3.2 Sentiment Analysis Tools and Techniques 17
2.3.3 Sentiment Analysis Algorithms 19
2.3.4 Language Translating Applications 29

2.4 Research Gaps 32
2.5 Conceptual Framework 34
2.5.1 Tier I Conceptual Framework 34
2.5.2 Tier II Conceptual Framework 35
CHAPTER THREE 37
RESEARCH DESIGN AND METHODOLOGY 37
3.1 Introduction 37
3.2 Research Paradigm 37
3.3 Research Design 37
3.4 Problem Identification 39
3.4.1 Target Population 39
3.4.2 Sampling techniques and sample size 40
3.5 Model design 42
3.6 Data Collection Procedures 43
3.7 Data Analysis and Presentation 44
3.8 Model Implementation 44
3.8.1 Subjectivity Detection and Polarity Detection 46
3.8.2 Model Workflow 47
3.9 Model Evaluation 48
3.10 Ethical Considerations 49
CHAPTER FOUR 50
DATA ANALYSIS, INTERPRETATION AND MODEL IMPLEMENTATION . 50
4.1 Introduction 50
4.2 Demographics 50
4.2.1 The County Category 50
4.2.2 Gender 51
4.2.3 Age 51
4.2.4 Level of Education 52
4.2.5 Occupation 52
4.2.6 Participation 53
4.3 Challenges faced in obtaining Sentiments in Public Participation Forums. 54
4.3.1 The extent of Public Participation 57
4.3.2 Factor Analysis 58
4.3.3 Correlation Analysis 60

4.3.4 Regression Analysis 61
4.3.5 ANOVA 61
4.4 Model Design 62
4.4.1 Design Process 62
4.4.2 Modular Design 63
4.5 System Implementation 65
4.5.1 Purpose of the Model 65
4.5.2 Computation of popularity and Subjectivity from the Comments 66
4.5.3 The functional Overview 66
4.5.4 Computation Procedure 66
4.5.5 Module Interfaces 67
4.6 Model Validation 81
4.6.1 Validation Metrics 82
4.6.2 Validation Variables 84
4.6.3 Paired Sample t-test Validation 90
CHAPTER FIVE 94
CONCLUSIONS AND RECOMMENDATIONS 94
5.1 Introduction 94
5.2 Conclusions 94
5.2.1 The challenges encountered when obtaining sentiments in public participation forums in county governments 94
5.2.2 Suitable design for sentiment analysis model for county government's public participation forums 95
5.2.3 Howsentiment analysis model for public participation forums in county governments be implemented 96
5.2.4 Howsentiment analysis model for public participation forums in county governments perform 96
5.3 Recommendations 97
REFERENCES 99
APPENDICES 103
Appendix I: Source Codes 103
Appendix II: Questionnaire for Respondents 109
Appendix III: Research Authorization Letter 112
Appendix IV: NACOSTI Research Permit 113

Appendix V: Evidence of Conferences 114
Appendix VI: List of Publication 115

LIST OF FIGURES

Figure 1: Experimental Path for Sentimental Analysis Models 15
Figure 2: Tier I Conceptual Framework 34
Figure 3: Tier II Conceptual Framework 35
Figure 4: Design Thinking Process Approach 38
Figure 5: Design thinking process in this study 39
Figure 6: Functional decomposition process 42
Figure 7: Incremental Prototyping Approach 45
Figure 8: Model Workflow 47
Figure 9: Number of times in participation. 54
Figure 10: System Architecture 63
Figure 11: User Registration Form 68
Figure 12: Registration Flowchart 69
Figure 13: User Registration Page 70
Figure 14: Users View and Management 70
Figure 15: Admin Login Form 71
Figure 16: Participants Login Form 72
Figure 17: County Management Form 73
Figure 18: Forum Creation Form 73
Figure 19: Forums View 74
Figure 20: Forums View 75
Figure 21: Forums View 75
Figure 22: New Roles Form 76
Figure 23: Permission Assignment 76
Figure 24: Sentiment Analysis 77
Figure 25: Scores and Polarity Indicator 78
Figure 26: Sentiment score values computation 79
Figure 27: Participants Sentiment Summary 80
Figure 28: Expert SurveyParticipants. 81
Figure 29: Validation metrics. 83
Figure 30: Validation Variable 1 Results. 84
Figure 31: Validation Variable 2 Results 85
Figure 32: Validation Variable 3 Results 86

Figure 33: Validation Variable 4 Results 87
Figure 34: Validation Variable 5 Results 88
Figure 35: General feeling Results from Expert Survey 89

LIST OF TABLES

Table 1: Tools for sentiment analysis and their techniques 18
Table 2: Research Gaps 33
Table 3: Target population across the project areas 40
Table 4: Sampled population per County 50
Table 5: Comparison of gender and County 51
Table 6: Respondents Age 51
Table 7: Education Level 52
Table 8: Challenges faced in obtaining Sentiments in Public Participation Forums 55
Table 9: The extent of Public Participation 57
Table 10: KMO and Bartlett's Test 58
Table 11: Rotated Component Matrix[a] 59
Table 12: Correlations matrix 60
Table 13: Model Summary 61
Table 14: ANOVA[a] 61
Table 15: Validation Metrics Questions and Responses 82
Table 16: Paired Samples T-test 91
Table 17: Paired Samples Statistics 92
Table 18: Paired Samples Correlations 92
Table 19: Paired Samples Test 93

ACRONYMS AND ABBREVIATIONS

AIDS	-	Acquired Immuno-Deficiency Syndrome
ANEW	-	Affective Norms for English Words
CDF	-	Constituency Development Fund
DFRD	-	District Focus for Rural Development
EWGA	-	Entropy Weighted Genetic Algorithm
FRN	-	Feature. Relation Network
HIV	-	Human Immuno-Deficiency Virus
ICT	-	Information and Communications Technology
IGRTC	-	Inter-Governmental Relations Technical Committee
ITSs	-	Intelligent Transportation Systems
KMO	-	Kaiser-Meyer-Olkin
LASDAP	-	Local Authority Service Delivery Action Plan
LATF	-	Local Authority Transfer Fund
LDA	-	Latent Dirichlet allocation
NACOSTI	-	National Commission for Science, Technology and Innovation
NLP	-	Natural Language Processing
PANAS	-	Positive and Negative Affect Schedule
PLSA	-	Probabilistic Latent Semantic Analysis
PP	-	Public Participation
SA	-	Sentiment analysis
SID	-	Society for International Development
SRDP	-	Special Rural Development Program
TSA	-	Traffic Sentiment Analysis
USSD	-	Unstructured Supplementary Service Data

DEFINITION OF TERMS

County Governments These are geographical units envisioned by the 2010 Constitution of Kenya as the units of devolved government. They are essentially second tier subnational governments. The powers are provided in Articles 191 and 192, and in the fourth schedule of the Constitution of Kenya and the County Governments Act of 2012 (Ministry of Devolution and Planning, 2016 *County Public Participation Guidelines*).

NRC Emotion Lexicon Is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments - negative and positive (Zad, Jimenez, &Finlayson, 2021)

Public Participation Can be any process that directly engages the public in decision-making and gives full consideration to public input in making that decision. Public participation is a process, not a single event (Quick, Kathryn &Bryson, 2022).

Public Participation Forums These are formal events organized by the county governments for discussions of the pertinent county affairs as required by law in Kenya (Uraia, 2022).

Sentiment Analysis Is the process of algorithmically identifying and categorizing opinions expressed in text to determine the user’s attitude toward the subject of the document or post (Shivanandhan, 2020)

ABSTRACT

Public participation (PP) is a key constitutional principle outlined in the Constitution of Kenya. It promotes democratic and accountable exercise of power. It gives the citizens opportunity to enhance self-development and service delivery while accounting for their leaders' actions. However, lack of/insufficient public participation in Kenyan county governments is impeding effective devolution process. Among the reasons advanced for this development are inadequate communications. Still even in cases where PP has been successfully carried out, capturing and analysing the sentiments of the participants still remain a serious challenge. Therefore, an online PP tool with embedded sentiment analysis algorithms specifically designed for the counties can be quite resourceful under the circumstances. The main objective of the study was to develop a sentiment analysis model for use in public participation forums in County Governments in Kenya. The specific objectives are to; evaluate the difficulty in obtaining sentiments; determine the challenges faced in the design of an effective sentiment analysis model for public participation forums; design a sentiment model for public participation forums in county governments, and; evaluate the performance of sentiment analysis model for public participation forums in county governments. The study was conducted through the design thinking process. The population of interest in this study comprised of county management and staff also area residents in Nakuru, Busia and Baringo counties who have participated in public participation forums before. A sample size of 106 respondents comprising 23 county administrators and 83 residents were purposively sampled for the project. The sentiment analysis model was developed by implementing cloud NLP package and Bidirectional Encoder Representations from Transformers (BERT) algorithm to get magnitudes of user sentiments. Cross validation was then used to evaluate the performance metrics at the design stage and users participated in the evaluation of the model. The overall conclusion of validation is that the model performed as expected and recorded instrumental results in increasing effective public participation in county governments in Kenya and strengthen the devolution process. This study recommends that the model can be cascaded to all the counties in Kenya to improve the efficiency of public participation.

Keywords: *County Governments, Public Participation, Sentiment Analysis, Sentiment Analysis Algorithms*

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

A key constitutional principle outlined in the Kenya Constitution of 2010 is public participation under Article 10(2) a (Transparency International Kenya, 2018). It prescribes public participation as a key aspect of Kenyan national values and principles of governance. One of the key means for achieving this is through devolution as is set out in article 174 (c) of the constitution. The article presents devolution as the means to give powers of self-governance to the people and enhance their participation in the decision-making processes of government (Ngigi & Busolo, 2019). The significance of public participation is to bridge the gap between the citizens, civil society, private sector and the government (Quick, Kathryn &Bryson, 2022). It promotes democratic and accountable exercise of power. Public participation gives the citizens opportunity to enhance self-development and service delivery while accounting for their leaders' actions. It also fosters national unity by recognising diversity among it's' citizens (Uraia, 2022).

Public participation takes many forms which include; vetting, electing and recalling leaders, vying for public positions, paying taxes, maintaining peace and order, being informed on public issues, signing a petition on government policy or action and participating in elections and citizen forums (Quick, Kathryn &Bryson, 2022). The main approach to public participation in Kenya is through public forum meetings or hearings commonly known as *barazas.* Barazas are social group gathering formed by residents in their communities. Public barazas are a means for raising awareness of community, forming relationships, sharing knowledge and ideas. Civic education is often done in the barazas. Other approaches are through membership into lobby groups and citizens' fora which is a memorandum of understanding between public service providers and representatives of the people (Kituyi & Moi, 2021).

There have been some clear successes in public participation in social programmes such as the family planning campaigns in the 1970s/80s and the anti-HIV/AIDS campaign in the 1990s/2000s.The programs were heavily supported by development partners, community initiatives and participation was encouraged with very good results (Inter-Governmental Relations Technical Committee., 2016).The efforts to decentralise and facilitate public participation have however not been with its' own challenges if we consider the initial attempts at decentralised governance through the Special Rural Development Program (SRDP) of 1967 , the District Focus for Rural Development (1983), the Local Authority Transfer Fund (LATF/ the Local Authority Service Delivery Action Plan (LASDAP) or the Constituency Development Fund (CDF). These decentralisation efforts fell short of effective public participation resulting in public apathy and low levels of engagement. This resulted in limited impact on citizen empowerment and development.

One of the main challenges of public participation is the inadequate standard measure of effective public participation (Transparency International Kenya, 2018). In both the national and the county governments, efforts have been put into public participation but there are challenges in the lack of clarity on what constitutes adequate participation, nature of participation that meets the constitutional threshold or the most effective mechanism for public participation (Mbithi, Ndambuki & Juma, 2019).

There exist challenges in both the public and the government ends (Ministry of Devolution, 2020). The public perception of the government is that it does not understand their plights and that the governments' policies impact them negatively.

There is also the issue of inadequate representation. The special groups may not be well represented in fora. Other authorities in the government may make a conclusion that the community may object to a certain policy without even a trial. Restrictions on policy timelines do not allow the public ample time to understand, prepare and research effectively in order to participate. Effective participation needs transparency (Mayienda, 2020). Transparency in public' actions and transparency in leadership and administration. Openness is affected through access to information. Inadequate access leads to difficulty in interpreting the policies, services and programs. Public apathy is the indifference, lack of concern in development. When there is apathy among the public means that there are disinterested leading to them withdrawing from participation (Kituyi & Moi, 2021).

Technology is viewed as a solution to social issues. Public participation is not an exception. Every county government in Kenya has an official website which is often used to access information. The requisite documents for public participation are posted on the website (Transparency International Kenya, 2018). Social media such as Facebook, Twitter, WhatsApp, and short messaging service have often been used to share and discuss various issues. In a study on public participation in Kisumu County 39.8% of information on county forums was passed through the media; local radio stations and newspapers (Transparency International Kenya, 2018; Wacera, 2016) cited in his study that 44% of his respondents cited media; print, television and radio as their source of information on public participation. Petitioning is a tool for public participation. Writing of a petition involves identification of a target, researching on the subject matter, clear communication and finding ways to promote the petition (Ministry of Devolution, 2020). There has been a rise in online petitions in the recent past. A petitioner would write a petition on a website and mobilize the public through social media in signing the petition. The major problem with e-participation is the feedback mechanism. One does not get feedback on whether their contribution was taken into account. Blogging is an online approach that is yet to be utilised for online public participation by any county government in Kenya.

1.1.1 Sentiment Analysis

As democratic space expands across the globe and more nations are granting their citizens unfettered freedom of expression, masses of people are taking to several platforms to express their sentiments about the developments in virtually every sphere of life. Several platforms such as; Hadoop, MapReduce, Amazon Web Services, Oracle Advanced Analytics, provide SA services of big data due to its proximity to social media analysis. Social media is increasingly being relied upon for product reviews (Sharef, Zin, & Nadali, 2016). Some of the use cases of online platforms used for public participation includes; GANA Pienso and Consul. GANA Pienso is an online platform created by the Nariño government in order to provide the citizens access and the ability to voice their views, opinions, concerns about any government's projects, policies and proposals (Scott, 2019). Consul is a free, customisable, secure and open software that is used in 33 countries and 130 institutions (Consul, 2019). It provides opportunity for debates, proposals, participatory budgeting, voting systems and collaborative legislation.

The Cambridge English Dictionary (Press, 2019) defines sentiment as "a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something." Sentiments in public participation are very important because they represent one's attitude towards a certain policy or a process. The users of the internet are across gender, nationalities, age, tribe and class to share, air out their impressions and experiences on different subjects (Rambocas & Gama, 2013).

Opinions and attitudes are expressed in several ways; the type of vocabulary used, slang and lingual variations, the context of writing, tone of the conversation, sarcasm and the number of details given (Rambocas & Gama, 2013). With this immense volume of data, manual sentiment mining can, therefore, be tedious for an average human reader. This means that opinion mining is effectively undertaken because enough data logs are available. Data logs are often disorganised, bulky and disintegrated across the different online platforms. Sentiment detection is done at different levels either at single words, phrases, or complete sentences. Technological advances such as sentiment analysis, natural language processing and innovative text analytics are employed to extract and classify data.

Sentiment analysis (SA) which is also referred to as emotion AI or opinion mining can be defined as the process of automating mining of opinions, views, attitudes, emotions and phrases through Natural Language Processing (Kharde & Sonaware, 2016). It is the application of text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. According to a definition supplied by Shivanandhan (2020), sentiment analysis is the process of algorithmically identifying and categorizing opinions expressed in a text to determine the user's attitude toward the subject of the document (or post).

Sentiment analysis is widely applied to the voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Sentiment Analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world. Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. In politics, SA has been widely used by political strategists to gauge public opinion to policy announcements and campaign messages ahead of elections. This enabled them to quickly appraise the sentiment behind everything from news articles to forum posts and be able to strategize and plan for the future.

Humans are fairly intuitive when it comes to interpreting the tone of a piece of writing because they are able to discern the context. However, when it comes to software analysis of sentiments, the context is always absent and this has gross implications on the accuracy of the analysis. Therefore, when it comes to using social and online data to understand consumer opinions, sentiment accuracy is incredibly important. Sentiment analysis classifies the opinions into “positive”, “negative” or “neutral”. Sentiment analysis can also be done through statistics or machine learning. There are several approaches to sentiment analysis; machine learning, lexicon analysis and a hybrid approach (Andrea, Ferri, Grifoni, & Guzzo, 2015). The machine learning approach is used to predict the polarity of sentiments based on training. Lexicon uses a predefined list of words, where each word is associated with a specific sentiment. The hybrid approach combines both approaches (Beigi, Hu, Maciejewski & Liu, 2016; Kharde & Sonaware, 2016; Pollyanna, Benevuto, Araújo & Chu, 2013).

Kenya is one of the countries in the world with a very high digital penetration rates and coverages; mobile phone coverage is over 90% while internet penetration is close to 80% and growing (Kibuacha, 2021). Kenyans are also among the most active citizens in online applications from search engines to social media; they are also rated as being prolific online content creators. Imperatively, the public sector has also taken to online applications to leverage its policy of the efficient government and as a way to reach the masses through instant messaging. Therefore, the government’s digital footprints are growing online and have, as such, acquired several online handles and downloadable applications.

Curiously, though Kenyan's participation in public affairs online is unquestionable, online applications dedicated to public participation which is a key component of governance as per the constitution- are lacking. As such, tracking citizens' sentiments on issues emanating from specific contexts such as counties which are of interest to the present study remains a difficult task with the tools available. Reports from various commissioned and academic studies already suggest that public participation in counties as stipulated by the constitution is poor owing to issues such as communications, logistics and security among other things. This means there is at best only minimal grassroots element in policy and development initiatives in the county governments. This does not augur well with the devolved model of governance which requires more grassroots inputs in the management of public affairs. Even in the event they turn up for such forums, their freedom to expressly give their sentiments is not guaranteed and also the tools to capture and analyse their sentiments are also questionable.

The present study, therefore, proposed to address this gap by developing a tool for SA for use in county governments for public participation. The application is expected to not only encourage public participation at large but also to increase the volumetric levels of PP and provide high accuracy SA to topical issues generated or related for public participation.

1.2 Statement of the Problem

The gravity of public participation in the country's governance system has been observed when successful court petitions were used to halt important policy implementations and government projects due to lack of or insufficient public participation in the process.

However, despite this, public participation in governance affairs is stipulated in the constitution is remarkably low. Lack of quorum in PP was cited specifically as a major hindrance in effective citizen input. Among the reasons advanced for this development are inadequate communications, lack of county legal provisions for PP, fear of victimization, venues and logistics.

Carrying out the discussions online could improve the quality of the debates and bring out other salient issues. It could also allow for rapid evaluation of the discussions using software to establish the prevailing themes. Therefore, an online PP tool with embedded sentiment analysis algorithms specifically designed for the counties can be quite resourceful under the circumstances. Already, there are several applications in the market such as Brand watch Analytics which use algorithms to capture and analyse users' sentiments though most are used commercially by marketers and not for public policy. Locally, such tools are not available for public participation and citizens' views on governance have had to be captured and analysed using traditional means like physical surveys which interestingly also fail in their accuracy of SA. The present study, therefore, endeavours to design, implement and evaluate the performance of a local PP sentiment analysis model for county governments in Kenya.

1.3 Objectives of the Study

The main objective of the study was to develop a sentiment analysis model for use in public participation forums in County Governments in Kenya. The specific objectives are:

i. To determine challenges faced in obtaining sentiments in public participation forums for county governments.

ii. To design a sentiment analysis model for public participation forums in county governments

iii. To implement the sentiment analysis model for public participation forums in county governments.

iv. To evaluate the performance of sentiment analysis model for public participation forums in county governments.

1.4 Research Questions

i. What are the challenges encountered when obtaining sentiments in public participation forums in county governments?

ii. What is a suitable design for sentiment analysis model for county government's public participation forums?

iii. How can the sentiment analysis model for public participation forums in county governments be implemented?

iv. How does the sentiment analysis model for public participation forums in county governments perform?

1.5 Justification of the Study

The main focus of public participation is to get the publics' ideas, opinions and views freely and in a manner that is inclusive and with reduced time constraints. Therefore, effective public participation requires an open, accountable and structured process where the public can interact and influence decisions. However, due to several challenges on both sides the leadership and the public, this has not always been the case. In a report by Mbithi, Ndambuki & Juma (2019), a challenge on the manipulation of public participation forums by diverse interests was mentioned. The public often expects to be compensated for their time and makes demands for allowances to attend the forums. For a county government to organise a public participation *baraza,* finances, human resources and time is required. An online forum for the public to present their view means that most of the costs of PP are reduced and the scope of participation increased.

This enables the residents to get more time to react to public participation issues raised in the platform. The project is also expected to enrich the future research and development in this area of ICT by providing insight into how domestic sentiment analysis models work and how they can be adapted and developed for various applications in analysing sentiments.

1.6 Scope of the Study

The model developed was not a social media application *per se* but an android based downloadable application that incorporates elements of Web 2.0; microblogging and social media applications, and sentiment analytics that could capture the mood of users. However, the model can be integrated into such platforms when and where possible. The model was designed, implemented and tested in three county governments in Kenya, namely; Nakuru, Busia and Baringo Counties because they were deemed to have different demographic patterns and internet usage. As such, the application allowed successful user registration, authentication, public participation, sentiment analysis and administration thereof.

1.7 Limitations of the Study

The main limitation of this study was the fact that the model implementation and testing was confined to public participation frameworks in the three county governments in Kenya. This meant that the unique and dynamic requirements and challenges of PP in the other 44 counties might not be necessarily addressed. However, further requirement analysis, application modification and testing may be necessary in order to successfully cascade the framework and its software to other county governments in Kenya and beyond. Apart from the scope of application, the other enviable limitation was the sourcing of software development tools for sentiment analysis application. However, with the assistance of software specialists, the appropriate software resources were sourced and applied to the framework.

Thirdly, the challenges associated Data collection challenges which were anticipated to rise from cooperation from the county officials due to the sensitivity of the date being sought were actually witnessed. However, this limitation was mitigated by explaining the nature of the project and also furnishing the respondents with the research permits from relevant authorities and assuring them of ethical considerations guiding the study.

1.8 Assumptions of the Study

From the onset, this study assumed that the necessary licenses and permits from relevant government authorities were going to be availed to enable collection of data to be possible. It also assumed that the presentation of the licenses and permits to the respondents and assuring them of ethical considerations applied in the study will make them cooperate and provide relevant data required for the study. The study further assumed that a sizable population in the three counties would participate in giving relevant data as well as testing the model. In the event of these assumptions proved false, then the study and the associated tasks were to be futile.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of current approaches to public participation, challenges to these approaches, sentiment tools and the current sentiment analysis tools.

2.2 Sentiment Analysis Overview

The Internet has enabled people to share knowledge, experiences and thoughts with the world by using social media platforms such as blogs, forum discussions, and social networks (Andrea, Ferri, Grifoni, & Guzzo, 2015). The web has been drastically changing to the extent that billions of people all around the globe are freely allowed to conduct activities such as interacting, sharing, posting and manipulating contents without geographical boundaries.

According to Beigi et al., (2016), sentiment analysis is one class of computational techniques that automatically extracts and summarizes the opinions of the immense volume of data which the average human reader is unable to process.

2.3 Challenges facing Acquisition of Sentiments in public participation forums

Despite the constitutional provisions, there have been outcries in some counties over certain decisions undertaken by the county government (Donald, Guyo & Moronge, 2020). A very low percentage of Kenyans partake in the governance of their counties. However, citizens in Mandera, Uasin Gishu and Baringo counties demonstrated the highest levels of commitment to public participation (Uraia, 2017). Further, the ministry of devolution, Kenya, enumerated the challenges to public participation to include; Negative attitudes and public apathy, lack of information or difficulty to access them, high cost of public participation, language and literacy structural barriers among others (Ministry of Devolution, 2020).

According to Uraia Trust (Uraia Trust, 2016) 60% of the respondents had not attended a public hall meeting in the past one year, while the 40% had attended. The reasons as to such a low percentage of attendance are as follows: lack of information regarding the forum (59%), lack of time for such forums (20%), insignificance of participation by the individual (10%) and proximity of the venue (9%) among others (Uraia, 2017). Inadequate information was also stated by an overwhelming majority of respondents in a study by Kituyi & Moi, 2021). They cited that they were not aware of the public hearing meeting dates or their responsibilities as the public in scrutinising the accounts of the county government. Civic education aids in the sensitization of the responsibilities of the public in governance however civic education is also not done fully in most counties.

Before a public participation forum, it is the obligation of the county government to provide all the information on the subject matter of discussion, mechanisms of engagement and inform the public on what is expected of them (Ministry of Devolution, 2020). Information can be availed to the public through websites, social media platforms, radio programmes, notice boards, religious meetings, television, and posters. The Kenya's ministry of devolution enumerated some strategies to counter the challenge of public participation. Amon the strategies highlighted include, but not limited to; to have county governments ensure that they make public participations accessible through varied channels, increase the quality of civic education, allocate adequate budget to public participation, and ensure diversity of methods applied in public participation (Ministry of Devolution, 2020).

Ngigi & Busolo (2019) highlight the challenges of devolution to include; Insufficient Public Participation/Gender inequality. As such, public participation is seen as a time-consuming process or not necessary. Scheduling of public forums has an effect on public participation (Donald, Guyo & Moronge, 2020). Most public forums are held on weekdays when most of the participants expected to attend are at work. Demand for money-in-exchange for participation by communities is a trend that is rampant in most counties (Mbithi, Ndambuki & Juma, 2019).A study by (Kaseya &Kihonge, 2016) indicated that 62.5% of Civil Society Organizations and Government officers found financial incentives to be effective in public participation, encouraging attendance of public participation forums. While the remaining 37.5% indicated that offering financial incentives boosts participants' morale. Some county administrators also demand financial incentives for their participation and boycotted the meetings where it is not available (Kituyi & Moi, 2021).

Absence of a structured feedback mechanism from the county officials who are generally perceived as none receptive to such efforts is a hindrance to effective public participation (Hakijamii, 2017). Lack of feedback mechanisms from earlier hearings held discourages participation (Moi, 2019). The public need to know whether or not their inputs were received, and whether and why they were or were not incorporated into the relevant plans or budgets (Hakijamii, 2017).

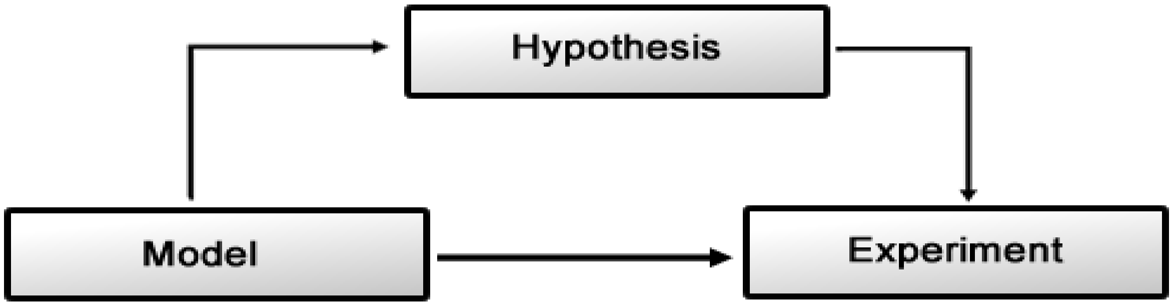
2.4 Sentiment analysis Models and Tools

2.4.1 Sentiment Analysis Models

A model as defined by Sedgewick (2007) is a theoretical tool of describing the physical or natural world, whose actual behaviour and relationships can then be tested and validated through hypothesizing the model constructs and/or carrying out an experiment to confirm model assumptions. A conceptual description of a model is given in Figure 1.

Figure 1

Experimental Path for Sentimental Analysis Models



The model in Figure 1 shows the two possible paths used in carrying out an experiment to evaluate the effectiveness of sentiment analysis models. Models can be tested directly to discover their efficacy in sentiment analysis. Also, specific aspects of the models can be hypothesized and tested through experiments. However, in the latter case, only the hypothesized variable(s) can be modified after the experiment. Some of the existing models of sentiment analysis are discussed as follows.

Unigram Language Model

A language model is a probability distribution over the words in that language. It assigns probability mass to the existing words of the considered vocabulary of that language. A unigram language model is a model that does not consider the order of words. Under this model, there is a single multinomial distribution that all the words of every document are drawn independently from it. In order to address the obvious shortcoming of the unigram model that ignores the order of terms in a document different other language models (bigram, trigram, n-gram) have been proposed. These models capture the dependency among constituting terms of a document by considering the previous term (bigram), two previous terms (trigram) and n−1 previous terms (n-gram).

Mixture of Unigrams

A mixture of unigrams model assumes that there exist K topics in the whole corpus and each document in the collection is about a single topic (Nigam et al. 2000). It introduces a discrete topic assignment variable for each document and once the topic of the document is known then all words of that document are drawn independently from the corresponding distribution. Each topic is modelled as a distribution over words and assigns a probability to words according to their likelihood of occurrence in that topic.

Probabilistic Latent Semantic Analysis

A mixture of unigrams model simply assumes that each document is generated from only a single topic. However, it is quite likely that a document could be about multiple topics. Probabilistic Latent Semantic Analysis (PLSA) is another method for modelling documents in a corpus that does not have the assumption of a mixture of unigrams (Hofmann, 1999). The notion of a topic (a probability distribution over words) also exists in PLSA and it models each document as a mixture of topics. Each document in the training set has its own distribution over topics and this distribution determines the proportion of each topic in the document. Knowing the topic assignment of each word of the document is a sample from it and then the word itself is drawn from the corresponding topic. PLSA is a mixture model for the probability distribution of words (a mixture of multinomial).

Each mixture component is a topic and each document is represented by a list of mixing coefficients of these components. In PLSA the topic distribution is only learned for those documents that are in the training set, so it cannot be seen as a generative model for generating a previously unseen document. Also, another drawback of PLSA is that since it learns a distribution over topics for every single document in the training set, the number of model parameters grows linearly with the size of the training set.

Latent Dirichlet Allocation

LDA is the simplest topic model that overcomes the limitations of PLSA. LDA treats the topic mixture weights as a K dimensional hidden random variable rather than a large set of individual parameters which are explicitly linked to the training set (Blei et al. 2003). LDA considers the topic distribution as a random vector and models it using a Dirichlet prior, however, PLSA learns a distribution over a topic for each document in the training set. Topic models have been developed in order to organize a collection of unstructured documents by extracting the main themes from a large collection of documents.

2.3.2 Sentiment Analysis Tools and Techniques

There have been many studies that provide tools and methods for sentiments analysis. The most used tool for detecting feelings polarity, negative and positive effect, of a message, is based on emoticons (Andrea et al., 2015). Emoticons are face-based and symbolize sad or happy feelings commonly known as *emojis*. Another method is the Linguistic Inquiry and Word Count (Dudău & Sava, 2021). It allows analysing of not only positive and negative but also emotional, cognitive, and structural components of a text-based on the use of a dictionary containing words and their classified categories. Happiness Index is a sentiment scale that uses the popular Affective Norms for English Words (ANEW) (Kapucu et al., 2021). It rates a text between 1and 9, indicating the amount or level of happiness. Another tool is the SentiStrength (SentiStrength, 2019) that is termed as "the most popular stand-alone sentiment analysis tool" (Nandwani&Verma, 2021). It uses a sentiment lexicon for assigning scores to negative and positive phrases in a text.

According to Sharma (2021), SentiWordNet is a tool widely used in opinion mining, based on an English lexical dictionary called WordNet. The lexical dictionary groups adjectives, nouns, verbs and other grammatical classes into synonym sets called synsets.

SentiWordNet associates three scores with synset from the WordNet dictionary to indicate the sentiment of the text: positive, negative, and objective (neutral) (Pollyanna, Benevuto, Araújo, & Chu, 2013). Other tools for sentiment analysis and their techniques are shown in Table 1.

Table 1

Tools for sentiment analysis and their techniques

Tools for sentiment analysis	Techniques used by tools	Weaknesses
SenticNet	Natural language processing approach for inferring the polarity at a semantic level	Has limitations in matching context compatibility to word similarity
PANAS-t	Eleven-sentiment psychometric scale	The PANAS does not encompass higher order mood states.
Sentiment140	API that allows classifying tweets to polarity classes positive, negative and neutral.	Words with similar contexts and opposite polarity can have similar word vectors.
NRC	Large set of human-provided words with their emotional tags.	Context-word compatibility
EWGA	Entropy-weighted genetic algorithm	Bias towards a specific polarity class.
FRN	Feature relation network considering syntactic n-gram relations	Lower positive and negative recall values.

SumView, is a Web-based system developed to summarize product reviews and customer opinions (Wang, Tsai, Liu, & Chang, 2013). It integrates review crawling from Amazon.com, automatic product feature extraction along with a text field where users can input their desired features, and sentence selection using a proposed feature-based weighted non-negative matrix factorization algorithm. The most representative sentences are selected to form the summary for each product feature. In a business domain, sentiment analysis is used for brand reputation, online advertising and e-commerce. Tweetfeel is an application that performs real-time analysis of tweets that contain a search term entered by a user (Crunchbase, 2019).

In the political context, the voting-advise applications represent an application of sentiment analysis. It enables campaign teams to track how voters feel about different issues. Sentiment analysis can be used to clarify politicians' positions, such as what public figures support or oppose. (Lu, et al., 2021) presents traffic sentiment analysis (TSA) applied on the area of traffic. The TSA allows analysing the traffic problem in a human way. Modern intelligent transportation systems (ITSs) represent a new emerging sentiment analysis domain. For the completeness of ITS space, it is necessary to collect and analyse the public opinions exchange.

2.3.3 Sentiment Analysis Algorithms

Several algorithms and methods for implementing sentiment analysis systems have been developed and are mostly classified as Rule-based systems that perform sentiment analysis based on a set of manually crafted rules; Automatic systems that rely on machine learning techniques to learn from data, and; Hybrid systems that combine both rule-based and automatic approaches. Both the rule-based and automatic approaches have lower levels of precisions which can be improved through hybridization. The concept of hybrid methods is very intuitive it just requires the combination of the best of both worlds, the rule-based and the automatic ones. Usually, by combining both approaches, the methods can improve accuracy and precision (Devlin and Chang, 2018).

Rule-based approaches normally define a set of rules in some kind of scripting language that identify polarity, subjectivity, or the subject of opinion. The rules may use a variety of inputs, such as the following: Classic NLP techniques like stemming, tokenization, part of speech tagging and parsing. Other resources, such as lexicons (that is, lists of words and expressions). This system is very naïve since it doesn't take into account how words are combined in a sequence. More advanced processing can be made, but these systems get very complex quickly. They can be very hard to maintain as new rules may be needed to add support for new expressions and vocabulary. Besides, adding new rules may have undesired outcomes as a result of the interaction with previous rules. As a result, these systems require important investments in manually tuning and maintaining the rules (Swartz, 2019).

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. The sentiment analysis task is usually modelled as a classification problem where a classifier is fed with a text and returns the corresponding category, e.g., positive, negative, or neutral (in case of polarity analysis is being performed). In the automatic approach, two classes of algorithms are mainly used that is the Machine Learning Algorithm where pairs of feature vectors and tags are fed into to generate a model and the Classifier Algorithms.

Ruleset for AAC (Adverb-Adjective Combination)

All NLP tasks commonly utilize parts of speech information to disambiguate sense and then use this to guide feature selection (Catelli et al., 2022). Adverbs and adjectives are mostly employed by researchers as features for identification of sentiment in a document or text. A description of algorithms used for identifying and analyzing adjectives, adverbs and AACs is discussed below.

Algorithm 2.1: Variable Scoring Algorithm

If adv $\in$ AFF $\cup$ STRONG, then:

fVS (adv, adj) = score (adj) + (5-score(adj)) * score (adv). … (1)

If score(adj) > 0. If score(adj) < 0,

fVS (adv, adj) = score (adj) - (5-score (adj)) * score (adv). … (2)

If adv $\in$ WEAK $\cup$ DOUBT, VS reverses the above and returns

fVS (adv, adj) = score (adj) - (5-score (adj)) * score(adv). … (3)

If score(adj) > 0. If score (adj) < 0, it returns

fVS (adv, adj) = score(adj) + (5-score (adj)) * score (adv). … (4)

Variable Scoring Algorithm (VSC)

The algorithm 2.1 changes the score of the AAC by using the function ***f*** defined as follows (Bonta, & Janardhan, 2019).

Description of the Algorithm: If an adverb is categorized as a strongly intensifying or affirmative adverb and the score assigned to the adjective is greater than 0, that is, the adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (1). For example, suppose score of positive adjective "good" is 3 and score of the adverb "really" belonging to affirmative or strongly intensifying adverbs is 0.4. Then, the final score of "really good" will be as given in (5).

Score (really good) = Score (good) + (5-score (good)) *score (really)

= 3 + (5-3) *0.4

= 3.8 ……………………………………………………………… (5)

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (4.6). But, if an adverb belongs to weakly intensifying adverb or adverb of doubt and score of an adjective is greater than 0, i.e., an adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (3).

For example, suppose score of adjectives “good” is 3 and score of the adverb “very” belonging to weak or doubt intensifying adverbs is 0.3. Then, the final score of “very good” will be as given in (6).

$$\begin{aligned}\text{Score (very good)} &= \text{Score (good)} + (5\text{-score (good)}) \text{ *score (very)} \\ &= 3 + (5-3) \text{ *}0.3 \\ &= 3.6 \quad \ldots\ldots (6)\end{aligned}$$

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (4). Thus, the score of “very good” is slightly lower than the score of “really good” because a score of the adverb “very” is less than the score of the adverb “really”.

Adjectives Priority Scoring Algorithms

Adjectives are most commonly used as features amongst all parts of speech. There is a strong correlation between adjectives and subjectivity of text. Even all the parts of speech play an important role, but only adjectives as features depict the sentiments with high accuracy. Accuracy of around 82.8% has been achieved in movie review domains by using adjectives only as features (Catelli et al., 2022). The algorithm 2.2 gives priority to adjectives over the adverbs and alters the score of the adjective by weight r. Then this weight r determines the extent to which an adverb affects the score of an adjective. It denotes the importance of adverb compared to an adjective that it modifies. The larger the value of r, the greater is the effect of the adverb (Bonta, & Janardhan, 2019).

Algorithm 2.2: Adjective Priority Scoring Algorithm

If adv $\in$ AFF $\cup$ STRONG, then:

$$\text{fAPSr (adv, adj)} = \min (5, \text{score (adj)} + r * \text{score (adv)}). \qquad \dots (7)$$

If score (adj) > 0. If score (adj) < 0,

$$\text{fAPSr (adv, adj)} = \min (5, \text{score(adj)} - r * \text{score(adv)}). \qquad \dots (8)$$

If adv $\in$ WEAK $\cup$ DOUBT, APSr reverses the above and returns

$$\text{fAPSr (adv, adj)} = \max (0, \text{score (adj)} - r * \text{score(adv)}). \qquad \dots (9)$$

If score (adj) > 0. If score (adj) < 0, it returns

$$\text{fAPSr (adv, adj)} = \max (0, \text{score (adj)} + r * \text{score (adv)}) \qquad \dots (10)$$

Description of Algorithm 2.2: If an adverb belongs to affirmative or strongly intensifying adverb and score of adjectives is greater than 0, i.e., adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (7). For example, consider the weight r between 0 and 1, i.e., 0.1. An adverb “really” belonging to affirmative or strongly intensifying adverb has a score of 0.4 and positive adjective “good” has a score of 3. Then, final score of “really good” according to adjective priority algorithm is as given in (11).

$$\text{Score of (really good)} = \min (5, \text{score (good)} + r * \text{score (really)}) = \min (5, 3+0.1*0.4)$$

$$= \min (5, 3.04)$$

$$3.04 \ldots\ldots\ldots\ldots (11)$$

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (8). If an adverb belongs to weakly intensifying adverb or adverb of doubt and score of an adjective is greater than 0, i.e., an adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (9). For example, consider the weight r between 0 and 1, i.e., 0.1. An adverb "very" belonging to weakly intensifying adverb or adverb of doubt has a score of 0.3 and positive adjective "good" has a score of 3. Then, the final score of "very good" according to adjective priority algorithm is as given in (12).

$$\begin{aligned} \text{Score of (very good)} &= \min(5, \text{score (good)} + r * \text{score (very)}) \\ &= \min(5, 3+0.1*0.3) \\ &= \min(5, 3.03) \\ &= 3.03 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12) \end{aligned}$$

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (10). Thus, the final score of "very good" is less than the score of "really good"

Algorithm 2.3: Adverb Priority Scoring Algorithm

If adv ∈ AFF ∪ STRONG, then:

fAdvPSr (adv, adj) = min (5, score (adv) + r * score (adj)). …….. (13)

If score (adj) > 0. If score (adj) < 0,

fAdvPSr (adv, adj) = max (0, score (adv) - r * score (adj)). …….. (14)

If adv ∈ WEAK ∪ DOUBT, then:

fAdvPSr (adv, adj) = max (0, score(adv) - r * score (adj)). …………..…(15)

If score (adj) > 0. If score (adj) < 0,

fAdvPSr (adv, adj) = min (5, score (adv) + r * score (adj)). …………….. (16)

Adverbs Scoring Algorithms

Adverbs have no prior polarity. But, adverbs can play a major role in the identification of sentiment of a sentence when used with sentiment bearing adjectives. The sentiment value of adjectives gets altered when adverbs are used (Surve et al., 2004). On the basis of the level to which adverbs can modify the sentiment value; these can be classified as shown in Table 4.2. The algorithm 2.3 is also similar to previous algorithm 2.2 except the parameter weight r is applied to adjective rather than adverb (Bonta, & Janardhan, 2019).

Description of Algorithm: If an adverb belongs to affirmative or strongly intensifying adverb and a score of an adjective is greater than 0, i.e., adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (13). For example, consider the weight r between 0 and 1, i.e., 0.1. An adverb “really” belonging to affirmative or strongly intensifying adverb has a score of 0.4 and positive adjective "good" has a score of 3. Then, the final score of "really good" according to adverb priority algorithm is as given in (17).

Score (really good) = min (5, score (really) + r * score (good))

= min (5, 0.4+0.1*3)

= min (5, 0.7)

= 0.7 ………………………………………………… (17)

So, final score of "really good" is 0.7.

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (14). If an adverb belongs to weakly intensifying adverb or adverb of doubt and score of an adjective is greater than 0, i.e., an adjective is positive then the final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (15). For example, consider the weight r between 0 and 1, i.e., 0.1. An adverb "very" belonging to weakly intensifying adverb or adverb of doubt has a score of 0.3 and positive adjective "good" has a score of 3. Then, the final score of "very good" according to adverb priority algorithm is as given in (18).

Score (very good) = max (0, score (very) + r * score good))

= max (0, 0.3+0.1*3)

= max (0, 0.6)

= 0.6 ………………………………………. (18)

Similarly, if the score of an adjective is less than 0, i.e., an adjective is negative then final score of the Adverb-Adjective combination is the score of the adjective that is accordingly modified with the effect of the adverb as given in (16). Thus, the final score of "very good" is less than the score of "really good".

Resulting Sentiment Computing Algorithm

The algorithm 2.4 assigns the sentiment to text on the basis of the final score calculated. The sentiment assigned to text can be positive, moderately positive, highly positive, negative, moderately negative, highly negative or neutral.

Description of the Algorithm: The final sentiment to text is assigned according to algorithm 4.4. According to this algorithm, if the final score is greater than 0 but less than or equal to 0.3 then text represents positive sentiment. But if the final score is greater than 0.3 but less than or equal to 0.5 then text represents moderately positive sentiment. Otherwise, if the final score is greater than 0.5 then text represents highly positive sentiment. Similarly, if the final score is less than 0 and also greater than or equal to -0.3 then text represents negative sentiment. But if the final score is less than -0.3 and greater than – 0.5 then text represents moderately negative sentiment. Otherwise, the text represents a highly positive sentiment. But if the final score is equal to 0 then text represents no sentiment or neutral sentiment.

Algorithm 2.4: Resulting Sentiment Computing Algorithm

If score > 0:

If score <=0.3: then text represents positive sentiment,

else if (score > 0.3) and (score <=0.5):

then text represents moderately positive sentiment,

else:

then text represents highly positive sentiment,

else if score < 0:   If score > = - 0.3:

then text represents negative sentiment,

else if (score < -0.3) and (score >=-0.5):

then text represents moderately negative sentiment,

else:

then text represents highly negative sentiment,

BERT Algorithm

According to Barry Schwartz (2019), the Bidirectional Encoder Representations from Transformers (BERT) algorithm is a technique for natural language processing (NLP) pre-training based on Google's neural network. BERT pre-trained model is based on other generative Pre-Training models such as ELMo and ULMFit and can be tuned to perform sentiment analysis tasks with improved levels of accuracy as opposed to training datasets from scratch. BERT contextualizes words using bidirectional representation; that is, considering preceding words (left-context) as well as succeeding words (right-context). This is an improvement of old models such as ELMo or ULMF which used either unidirectional representation (left or right contexts) or slowly bidirectional representation.

According to Devlin and Chang (2018), BERT considers words provided by the user as input, masks out 15% of the words, and then predict the masked words by running the whole sequence through a deep bidirectional Transformer encoder. The model can then be trained on simple tasks obtained from monolingual corpus to learn the relationship that may exist between sentences. Suppose, there are two sentences A and B, the algorithm based on contents of B whether B is just a random sentient in the corpus or is actually the next sentence that comes after A. As such, it labels B as "***IsNextSentence***" or "***NotNextSentence***". BERT therefore can be obtained as a large model (12-layer to 24-layer Transformer) trained on a large corpus (Wikipedia + BookCorpus) over a long period time (1M update steps)

The ability to adapt to many types of NLP tasks very easily makes pre-trained BERT one of the most significant models where the users can implement by just fine-tuning to perform their specific tasks.

2.3.4 Language Translating Applications

Since the project entails multilingual users, embedded language translation application is required both for the users and for the purpose of analytics. Social media is designed to maximize interaction by allowing users to communicate in text in their natural language and aided by features like emojis and multimedia such as pictures, videos and the Graphics Interchange Format (GIF). While emojis are fairly easy to understand in terms of the mood of the user, the mood in the textural communication can be quite difficult to understand due to the fact that users use words and phrases according to their culture. Therefore, social media applications recourse to different applications to help translate the language to its approximate form and this can then be matched with the emoji were used to estimate the moods of the user.

Two leaders of online global communication, Google and Facebook, are focused on improving their own machine translation systems in order to allow users to share their lives with ease and have conversations across the world. The ability to localize, rather than simply translate, is an important part of this process: Facebook and Google's users want their translated text to sound like the language they use every day. This means that machine translation systems need to understand idioms and the everyday evolutions of language. For example, French teenagers are creating new variations of the English word "wow," like "uau." Facebook's algorithms picked up on the trend and can now translate these phrases. Right now, two major types of machine translation are being tested by Google and Facebook. Each type has their own benefits, but Facebook's recent decision to use convolutional neural networks (CNNs) over the more common recurrent neural networks (RNNs) is showing promise in its ability to produce translations that more closely resemble localized text.

Typically, computers translate text by reading a sentence in one language and predicting a sequence of words in another language with the same meaning. RNNs operate on this principle, translating text one word at a time in a strict left-to-right or right-to-left order (Pascale, 2020). The most commonly used type of RNNs are long short-term memories (LSTMs): given a sequence of words they predict the probability of each word given the previous words. Google Translate and many other text applications use RNNs to search through a database of texts and uses statistical analysis to suggest the most likely translation to the user. Facebook recently tested a form of machine translation based on a CNN approach, which has typically been used for image recognition tasks. Unlike RNNs, which process information linearly and methodically, CNNs can process information in a hierarchy, which allows them to look for non-linear relationships in data. When it comes to translation, this means that a CNN can more easily grasp contextual meaning and translate accordingly. However, as these technologies are still under development and as such may not be readily available, the present study opted for the crowd translation application Linqapp.

Linqapp was developed by Sebastian Ang and David Vega in 2014 in Taiwan when they were studying Chinese. While adjusting to life in Taipei City with minimum language skills, the two constantly encountered problems that no app or Web site could help them solve, such as handwritten food menus that flummoxed even the most sophisticated optical-character recognition (OCR) program or confusing public transit maps. According to Ang (2014), the limits of computer-based translation are reached easily. “Bus schedules in Taipei are completely unreadable for foreigners. It’s a perfect example of how something like Google Translate can’t help you. It was a situation where I thought if I had just one native speaker to ask, that would be perfect. I could just take a picture of the schedule and ask ‘does this go there or there?"”

Linqapp is essentially a crowdsourced language assistance app where users can request real-time help from native speakers of a foreign language whenever they find themselves in linguistic trouble (Gino, 2015). This premise means that Linqapp can be used for virtually anything: from language learning to translation assistance to help with cultural nuances. This also makes the new Linqapp Live feature central to the selling proposition of the app. “Live” connects users to instant assistance — within a minute user can reach native speakers of the language with which they need assistance.

The study applied for license to use the Linqapp and embed it into the application to facilitate translation. This enabled the analyzer tool to interpret the words used by the participants in the public participation process even when they use Swahili language or a mixture of English and Swahili. The words are then correlated with the emojis to determine the mood of the participants.

2.4 Research Gaps

Several algorithms and methods for implementing sentiment analysis systems have been developed and are mostly classified as rule-based systems, automatic systems and hybrid systems. Technically, majority of these systems are not independent special purpose systems and their lack of customization may mean that they cannot be applied to more specific purposes like sentiment analysis of government policies.

Table 2

Research Gaps

S/No.	Paper Title Author(s) & Year	Specific Objectives	Findings	Research Gaps
1	Dodds, P., & Danforth, C. (2009). Measuring the happiness of large-scale written expression: songs, blogs, and presidents.	To examine the measurements of happiness of large-scale written expression: songs, blogs, and presidents.	Happiness Index is a sentiment scale that uses the popular Affective Norms for English Words (ANEW). It rates a text between 1and 9, indicating the amount or level of happiness.	The Happiness Index tool analyses sentiments on the basis of happiness score and uses only one language, the English language
2	Dudău, &Sava, (2021). The psychological meaning of words: Liwc and computerized text analysis methods.	To investigate the psychological meaning of words using Linguistic Inquiry and Word Count	Linguistic Inquiry and Word Count allows analyzing of not only positive and negative but also emotional, cognitive, and structural components of a text-based on the use of a dictionary containing words and their classified categories	The Linguistic Inquiry and Word Count tool lacks a weighting system for the sentiments
3	Nandwani&Verma (2021). Sentiment strength detection in short informal text.	To identify the causal effect of interest we use variation in foreign countries import tariffs that are plausibly exogenous to domestic firms	SentiStrength is termed as “the most popular stand-alone sentiment analysis tool” It uses a sentiment lexicon for assigning scores to negative and positive phrases in a text.	The SentiStrength tool analyses sentiments on the basis of polarity not through a weighting system. This may not reflect moods in a continuous scale
4	Esen, Simdi & Erguze Andrea, A. D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and applications for Sentiment Analysis Implementation.	To examine SA Approaches, Tools and applications for Sentiment Analysis Implementation	The most commonly used tool for detecting feelings polarity, negative and positive effect, of a message, is based on emoticons	The emoticons tool relies exclusively on emojis which biases the subscribers to the opinion of thread writer. It means they can only agree or disagree using preset graphics

In terms of empirical research, fewer studies exist in this area of computing and even where they are done, the focus is on commercial applications with public sector applications of SA being relatively unexplored. Therefore, the present study seeks to not only develop and application for the public sector but also provide empirical information that can be used for future research in this area.

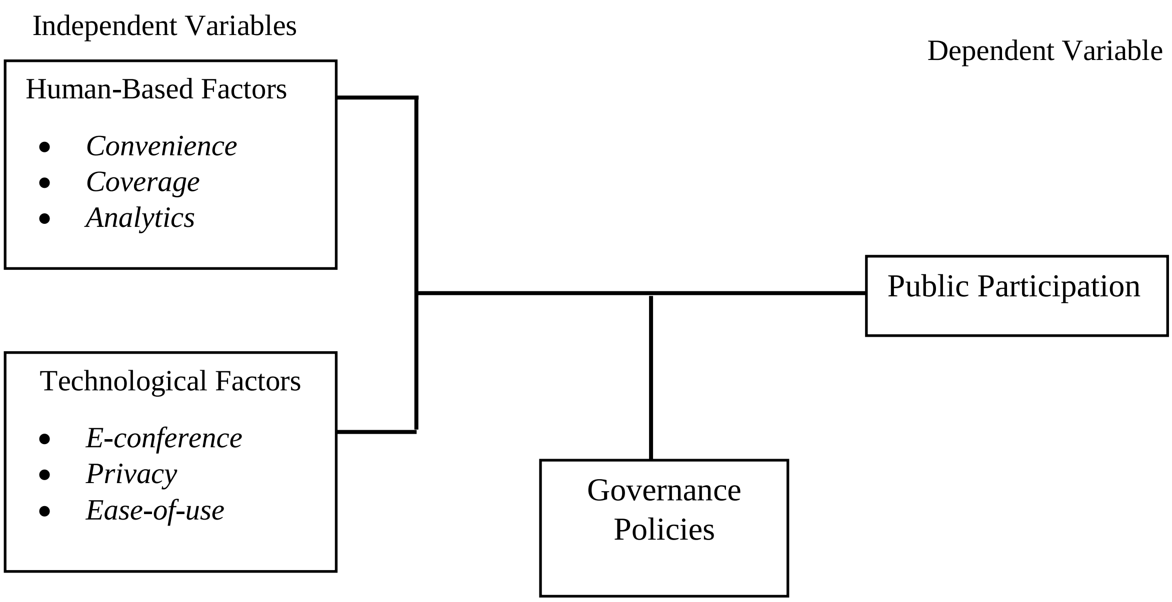
2.5 Conceptual Framework

The conceptual framework that directed the study was done in two stages, namely; Tier I conceptual framework and Tier II conceptual framework. Tier I conceptual framework presents the independent and moderating variables that influences public participation while Tier II conceptual framework presents the implementation of sentiment analysis model. Tier I and Tier II conceptual frameworks are presented in figures 2 and 3 respectively.

2.5.1 Tier I Conceptual Framework

Figure 2

Tier I Conceptual Framework:



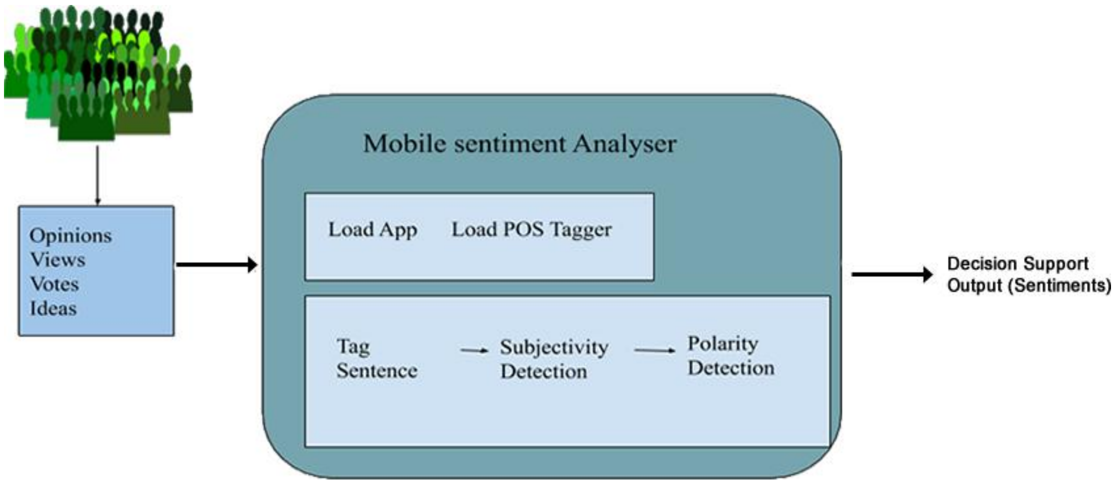
Source: Researcher (2023)

The Figure 2 above presents the conceptual framework that guided the study. The uptake of public participation is affected by technological factors as well as human based factors. In the case of this study, the human-based factors include, but not limited to; time constraints and availability herein referred to as convenience, the inability to express themselves fully and the scope of topics covered herein referred to as coverage, and sentiment analysis challenges herein referred to as analytics. Technological factors, on the other hand, include online platforms that allow participants to access information herein referred to as e-conference, the honesty of the participants to present their views herein referred to as privacy as well as the ease of use of the online public participation platform.

2.5.2 Tier II Conceptual Framework

Figure 3

Tier II Conceptual Framework:



Source: Researcher (2023)

Tier 2 conceptual framework presented in figure 3 above demonstrates the implementation of sentiment analysis model. The users include public participants who post their opinions, views and ideas to sentiment analysis model. Users also include system administrators, county government officials and other entities that access the model for reports.

The Mobile Sentiment Analyser incorporates a mobile application where the public can give their opinions on a subject matter provided by the county government and a sentiment analyser. The sentiment analyser includes the following components:

(a) **Text pre-processing -**A text might contain different paragraphs which have to be cut into sentences based on English symbols. Using the position of speech tagging to identify the types of words in the sentence.

(b) **Subjectivity Detection –**This involves using the POS tags to identify opinion lexicons in the sentence, whether the sentence is subjective or objective.

(c) **Polarity Detection –**This stage also utilizes the POS tags to indicate positive or negative expressions.

The output is of the sentiment analysis model is a cumulative general feeling that represents how participants consider the ideas being discussed. This is significant in guiding decision making in matters that affect the participants as citizens of county governments.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter comprises of the methodology that was used for the research study. Included are the research methods that were used to achieve each objective of the study. These include the research design, problem identification, model design, model implementation, subjectivity detection and polarity detection, evaluation, experimental testing on application, data collection procedures and data analysis and presentation.

3.2 Research Paradigm

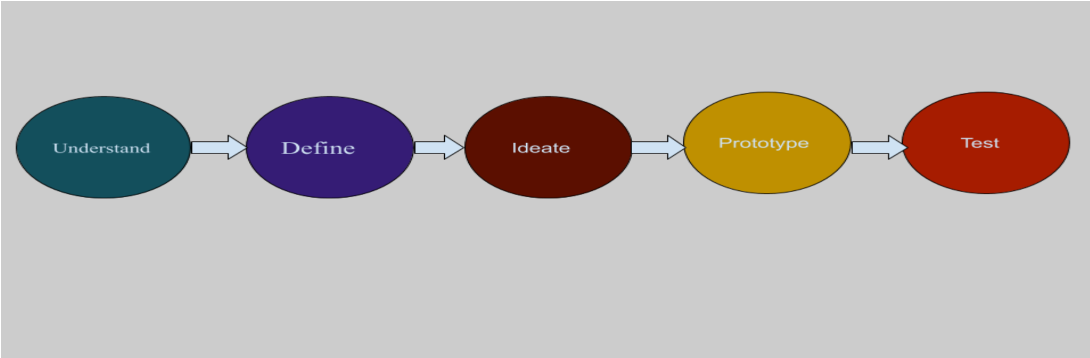
According to USC Libraries (2023), a research paradigm is technique, model, or pattern for performing research. It incorporates notions, convictions, or comprehensions that allow for the operation of theories and practices. This study adopted two paradigms, namely, scientific paradigm and design science paradigm. On the one hand, the scientific paradigm was significant for gathering data required about the challenges of current public participation models. On the other hand, the design science approach was relevant when the model was implemented.

3.3 Research Design

Implementation of the model was done through the design thinking process approach. It is the process by which the core principles of design are used to solve problems and identify innovative solutions that enhance user experience (Adams & Nash, 2016). The three elements to design thinking approach include; understanding the need and the user experience, brainstorming and coming up with a range of possibilities and ideas, and building and testing out the concepts to select a solution that fit the user's problem (Brown, 2009).

Figure 4

Design Thinking Process Approach



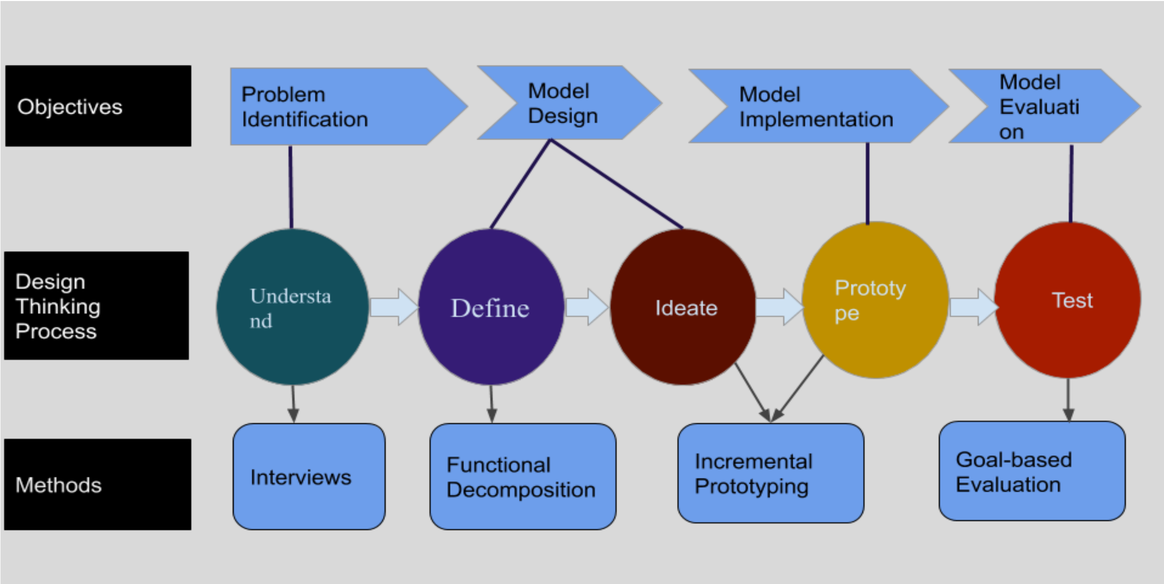
1) **Understand:** It involves using the empathic process to get the user's needs and experiences. The objective of empathy is to understand the other person. There are four phases to the empathic discovery process, which include; discovery, immersion, connection and detachment (Kouprie& Visser, 2009). Discovery is the decision to leave one's comfort zone to understand the client's world. Immersion is the action of stepping into the client's world. This may include acts such as observations, interviews or site visiting. It is a way to collect baseline data. Collection phase includes sharing of the feelings from the previous phase with the client. Detachment involves stepping out of the client's world and using professional expertise to enhance the client's life (Kouprie& Visser, 2009).
2) **Define:** This involves creating a point of view that is based on the insights and needs of the user. It includes redefining and focusing questions based on the insights gathered from the first stage. It analyses collected data and identifying which users' needs to be addressed with the design solution.
3) **Ideate:** This is the exploration phase. It is diverging on large quantities of possible ideas that could evolve into solutions.
4) **Prototype:** It is putting the ideas to the test. It includes developing some of the ideas into tangible objects. It is building the design.

5) **Test:** This involves the interaction of the prototype with the users, learning how they interact with it, allowing for refinement of the ideas.

The design thinking process was employed as follows:

Figure 5

Design thinking process in this study



3.4 Problem Identification

i. Understand –The study intended to understand more about public participation at the county governments. It sought to empathize with the use of questionnaires. It sought to determine the challenges in getting sentiments during public participation forums and the approaches the county undertakes when it comes to public participation.

3.4.1 Target Population

According to the Ministry of Devolution and Planning (2018), three counties were reported to have taken part in active public participation, namely, Nakuru, Busia and Baringo. The population of interest for this study therefore comprised of county management and area residents of these three counties. As such, data from the three counties indicated that cumulatively 491 local residents and 23 county administrators participated in the last public participation forums in 2018 as shown in Table 3.

Table 3

Target population across the project areas

County	Last Attendance in Public Participation Forums		
	County Administrators	Residents	Percentage (%)
Nakuru	7	211	42
Busia	9	165	34
Baringo	7	115	24
Total	23	491	100

Source: Nakuru County Governments (2018), Busia County Governments (2018) and Baringo County Governments (2018)

Therefore, the project targeted 218 persons in Nakuru, 174 persons in Busia and 122 persons in Baringo counties respectively. The three counties have been selected due to the fact that they have disparities in their demographic patterns and also public participation patterns and internet usage.

3.4.2 Sampling Techniques and Sample Size

According to Hodge (2020), sampling is the procedure of selecting members of a research sample from the accessible population which ensures that conclusions from the study can be generalized to the study population. Since the population of the local residents is large enough to warrant simple random sampling, the formula proposed by Nassiuma (2000) was be used to arrive at the desired sample size as;

$$n = \frac{Nc^2}{c^2 + (N-1)e^2}$$

Where n = sample size, N = population size, c = coefficient of variation (≤ 50%), and e = error margin (≤ 5%). This formula enables the researchers to minimize the error and enhance the stability of the estimates (Nassiuma, 2000). Substituting into the formula:
Thus, a sample size of 83 residents obtained from the above formula and to these was added 23 county administrators who were purposively sampled for the project, thus, bringing the total accessible population to 106. This figure closely agrees with Kathuri and Pals (1993) and Denscombe (2007) who recommend a minimum of 100 subjects as ideal for survey research in social sciences.

Two sampling methods were used in the study to select respondents, the purposive sampling method for the county administrators and the snowball sampling method for area residents who have participated in the previous public participation forums in their respective counties. According to Ames, Glenton and Lewin (2019), purposive sampling of primary studies for inclusion in the synthesis is one way of achieving a manageable amount of data. Purposive sampling and allows for key informant selection such as county administrators who participate in the public participation forums in their counties.

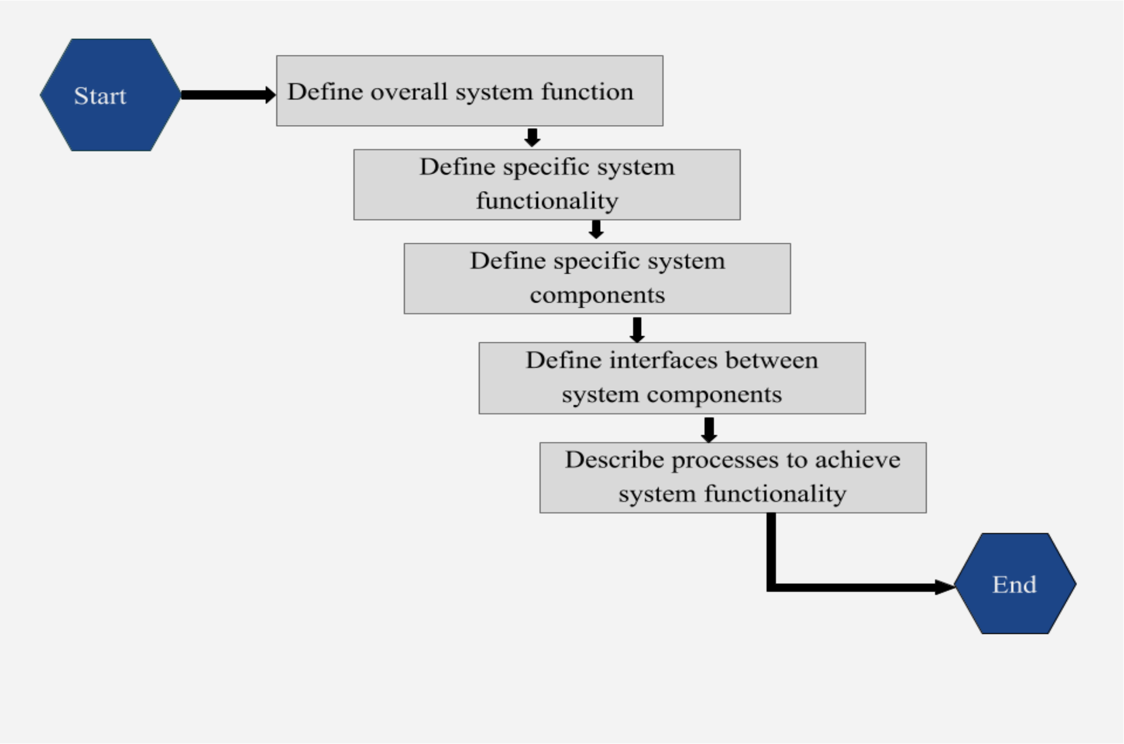
Snowball sampling may simply be defined as a technique for finding research subjects where one subject gives the researcher the name of another subject, who in turn provides the name of a third, and so on (Atkinson & Flint, 2001). The “snowball” sampling is a non-probabilistic research technique through survey and data registration which is usually recommended when: the population cannot be strictly delimited or detailed; the characteristics of the sample are rare; a good research method when the study is on behaviors, perceptions, customs, for the description of “typical” cases which cannot be generalized for entire populations (Etter & Perneger, 2000; Dragan & Isaic-Maniu, 2013).

3.5 Model design

i. **Define:** Based on the analysis of the results of the questionnaires, users could be factored into the design solution. Definition phase sought to design and explore the functional capability of a local sentiment analysis framework for public participation. Analysis and design of the sentiment analysis model for online participation forum utilised the functional decomposition approach. Functional decomposition begins with a broad abstract function of a system and levels down to the algorithmic functions that were translated into code. It focused on the functions, sub-functions and the interfaces between them as shown below.

Figure 6

Functional decomposition process:



Source: Ganney et.al (2020)

ii. **Ideate:** This involved brainstorming through the different sentiment analysis algorithms, sentiment classification techniques and selecting the best-suited algorithms and technique based on the overall function of the system as defined in the define stage.

For optimal precision, the proposed SA model used the hybrid approach where the pre-trainedmBERT NLP algorithms that would accommodate Swahili language was used to extract features (sentiments) as well as for classification. This approach that uses both rule-based and automatic approach algorithms is expected to increase the precision levels of the SA output. To evaluate the effectiveness of the sentiment analysis applications, first, the applications was subjected to pre-trained mBERT NLP analysers using the language commonly used by the users and this was then evaluated using the Resulting Sentiment Computing Algorithm. The researcher also compared the results with his own understanding of the language. Apart from the challenges arising from the cooperation of the public and infrastructure, the projects carefully analysed the limitation of the software in the design of the SA model. This was analysed for such issues as polarity, neutrality etc when it comes to the language of the intended user. The results of the first two steps were used in the architecture of the sentiment analysis model for public participation forums in county governments. Finally, the SA model was evaluated for performance through a computing system mainly comprising of tokenization and pattern matching.

3.6 Data Collection Procedures

Before the research instruments are distributed the researcher first obtained the necessary research authorizations from the University and the National Commission for Science, Technology and Innovation (NACOSTI) make prior visits to the selected counties headquarters to seek appointments with the respondents. The researcher identified the respondent through Public Participation office management. In the subsequent visit, the research instruments were administered to the respondents by the researcher and thereafter collected for analysis after they were duly filled.

3.7 Data Analysis and Presentation

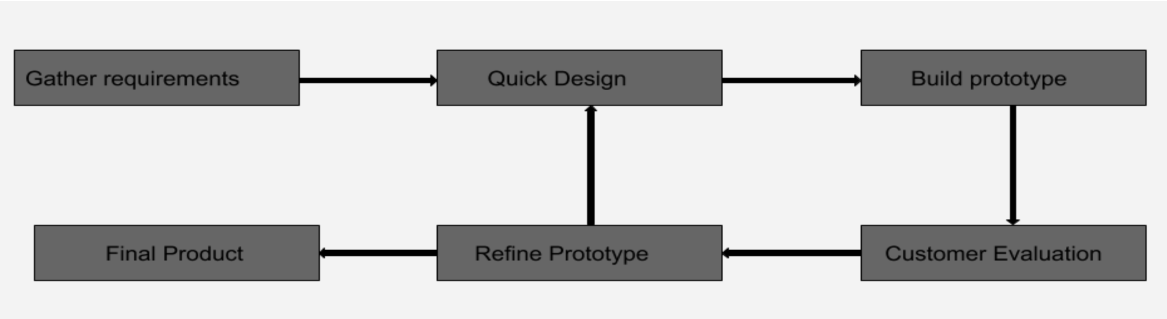
The data for the applicability testing was analysed using both descriptive and inferential statistical methods. Descriptive analysis was done using frequencies, percentages and means to describe the basic characteristics of the population. Inferential statistics involved spearman correlation to analyse whether the relationship between challenges faced in obtaining sentiments in public participation forums and level of public participation. Moreover, ANOVA was run to test whether there are significant differences in means between Baringo, Nakuru and Busia counties regarding public participation.

3.8 Model Implementation

The model employed incremental prototype approach comprising of six steps to implement the sentiment analysis model. The key features of the model design and implementation were derived from similar models of sentiment analysis and also from user and client perspectives to design and after the design. There are several approaches to the implementation of technology models which includes the incremental prototype approach. A prototype is an early approximation of a final system. It demonstrates what the system looks like, what it does and how it performs. Figure 7 shows the incremental prototyping approach that was applied in this study;

Figure 7

Incremental Prototyping Approach:



Source: Ganney et al., (2020)

1. Gather requirements: The requirements for the model were informed by the collected data obtained in the first objective.
2. Quick Design: This involves the following process:

   i. **Identifying the mobile operating system to be used**

   The android operating system was utilised. Android is a popular operating system with 82.26% of the Kenyan population using it (GlobalStats, 2019).

   ii. **Identifying the POS tagger tool**

   A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads the text in some language and tags parts of speech to each word, such as noun, verb, adjective, etc. Stanford POS tagger was employed.

   iii. **Identifying the subjectivity and polarity detection mechanisms**

   Subjectivity detection is a key component success of the polarity detection step. The study employed lexical resources, WordNet and SentiWordNet to infer the degree of subjectivity in each sentence.

3. Build Prototype: This is the development of the model. It used the following tools to build the prototype:

(a) Android Development Setup

i. **Android Studio v3.4.1**

Android Studio is the official integrated development environment tool that supports all Android SDKs for android application development. It contains libraries, debugger, an emulator, documentation, sample codes and tutorials for each API level of the released versions of Android.

ii. **Java SE Development Kit v11.0.3**

JDK is a package of tools for developing Java-based software. Allows developers to create Java programs that can be executed and run by the Java Virtual Machine and Java Runtime Environment.

(b) POS tagger Development

i. **Stanford Tagger v3.9.2**

The Stanford POS Tagger is an implementation of a log-linear part-of-speech tagger. It is effectively language-independent, usage on data of a particular language always depends on the availability of models trained on data for that language.

3.8.1 Subjectivity Detection and Polarity Detection

i. **SentiWordNet v3.0**

SentiWordNet is a lexical resource in which each WordNet synset is associated to three numerical scores Obj(s), Pos(s) and Neg(s), describing how objective, positive, and negative the terms contained in the synset are.

ii. **WordNet v3.1**

It is the most well developed and widely used lexical database for English. It organizes lexical information in terms of word meanings, rather than word forms.

4. Customer Evaluation: After the prototype has been built, it is presented to the client for test user evaluation. User test helps in realizing which aspects of the design is not working and what needs to be done to improve it.
5. Refine Prototype: Based on the user test evaluations, the prototype is refined to suit the client.
6. Final product: The final software product is delivered.

3.8.2 Model Workflow

Figure 8

Model Workflow

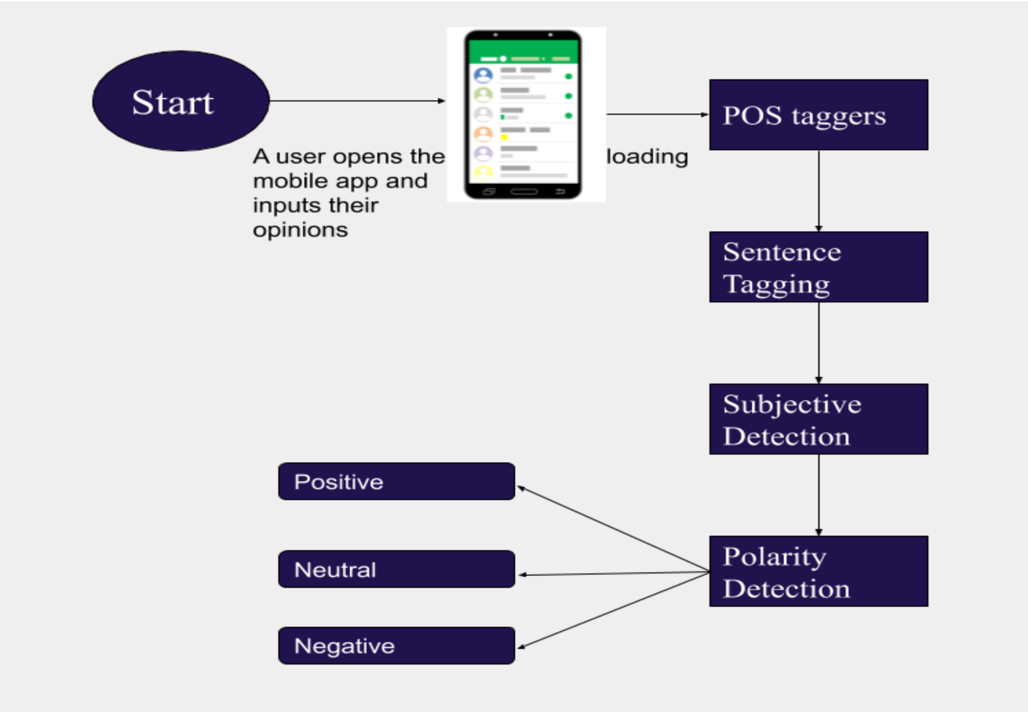


Figure 8 above presents the workflow of the sentiment analysis model. The Mobile Sentiment Analyser incorporates a mobile application where the public can give their opinions on a subject matter provided by the county government and a sentiment analyser. The sentiment analyser includes the following components:

(a) **Text pre-processing -**A text might contain different paragraphs which have to be cut into sentences based on English symbols. Using the position of speech tagging to identify the types of words in the sentence.

(b) **Subjectivity Detection –**This involves using the POS tags to identify opinion lexicons in the sentence, whether the sentence is subjective or objective.

(c) **Polarity Detection –**This stage also utilizes the POS tags to indicate positive or negative expressions.

3.9 Model Evaluation

After the implementation of the model, the performance of the model was evaluated to assist in understanding its capabilities and the user attitude towards it. The measures to be used in the evaluation are usefulness, ease of use, applicability, under stability, security, accuracy, stability, limit of detection, specificity and interact ability. Its capabilities was measured in terms of capturing and accurately analysing content against the context where a threshold of 60% correctly analysed content was deemed acceptable. Various studies (Devon Et Al., 2007; Oso & Onen, 2009; Wilson, Pan, & Schumsky, 2012) recommend that when 60% of constructs are rated by the assessors as essential, then the measurements are valid. Concomitantly, a voluntary user subscription of 60% and above of the target population was considered agreeable for the project and hence the model can only be improved and not redesigned (Wilson et al., 2012). Evaluation of the model was conducted where 33 persons comprising, the county officials and local residents participate in validating the model. This was done by encouraging the respondents to subscribe for free by downloading the Application on their mobile phones and then evaluating its usage and give feedback. Paired samples t-test was used to analyse feedbacks for validity of the model.

3.10 Ethical Considerations

Ethical issues to consider in carrying out research are privacy, confidentiality, sensitivity to cultural differences, gender and anonymity (Kitchin & Kate, 2000). The researcher did not require the participants in the project to reveal their true names or identity during the study. Issues that the participants are not comfortable with were addressed personally by the researcher. Adequate and clear explanations clarifying the intention of the research was given to all the respondents. The researcher disclosed the real purpose of the research and gave all the facts pertinent to the research so that the respondents may be able to make informed decisions on whether to participate in the study or to decline. The participants in the study were assured of anonymity and confidentiality throughout the research process. Further, Respondents were not expected to write their names or those of their schools anywhere in the questionnaire. They further assured that the information given were only be used for the purpose intended.

CHAPTER FOUR

DATA ANALYSIS, INTERPRETATION AND MODEL IMPLEMENTATION

4.1 Introduction

This chapter entails the data analysis relating to response rate, background information which entails county category, gender, age, education level, occupation, and number of participations in the county. Descriptive analysis entails proportion and percentages. Finally, inferential analyses include ANOVA, Factor analysis, correlation, and regression analysis.

4.2 Demographics

The variables analysed includes County category, gender, age, level of education, occupation, and number of times in participation in the county government.

4.2.1 The County Category

Respondents were sampled from three different counties, namely Baringo, Busia and Nakuru. The results of the analysis are presented in Table 4.

Table 4

Sampled population per County

County	Frequency	Percent
Baringo County	131	26.0
Busia County	193	38.4
Nakuru County	179	35.6
Total	503	100.0

The analyzed data showed that Busia County had 38.4% of the respondents followed by Nakuru County with 35.6%. Finally, Baringo County had 26% of the total respondent representation.

4.2.2 Gender

Respondents' gender was analysed by cross tabulation with County of residence. Table 5 shows the findings.

Table 5

Comparison of gender and County

Gender	Baringo County	Busia County	Nakuru County	Total
Male	17.9%	21.5%	17.7%	57.1%
Female	8.2%	16.9%	17.9%	42.9%
Total	26.0%	38.4%	35.6%	100.0%

The analysis of respondents' gender indicated that 57.1% were males while 42.9% were females. The data was further analyzed by county and was established that the proportion of males and female distribution in Nakuru county was approximately equal (0.2% difference). Furthermore, there was a slight difference in gender representation in Busia County with 4.6% more males than females. Finally, there was a significant difference in gender representation in Baringo County (9.2% difference).

4.2.3 Age

Age of the respondents was analysed using descriptive statistics as shown in Table 6.

Table 6:

Respondents Age

Age Category	Frequency	Percent
25 years and below	85	16.9
26-35 years	191	38.0
36-45 years	137	27.2
46-55 years	79	15.7
over 55 years	11	2.2
Total	503	100.0

According to table 6, 38% of the respondents were between 26-35 years and followed by 27.2% who were within36-45 years of age. Furthermore, 16.9% and 15.7% of respondents were below 25 years of age and between 46-55 years, respectively. It was noted that only 2.2% were advanced in years (over 55 years).

4.2.4 Level of Education

The educational level of the participants was analysed and presented as indicated in Table 7.

Table 7

Education Level

Level	Frequency	Percent
Certificate	149	29.6
Diploma	183	36.4
Degree	150	29.8
Masters	19	3.8
Doctorate	2	.4
Total	503	100.0

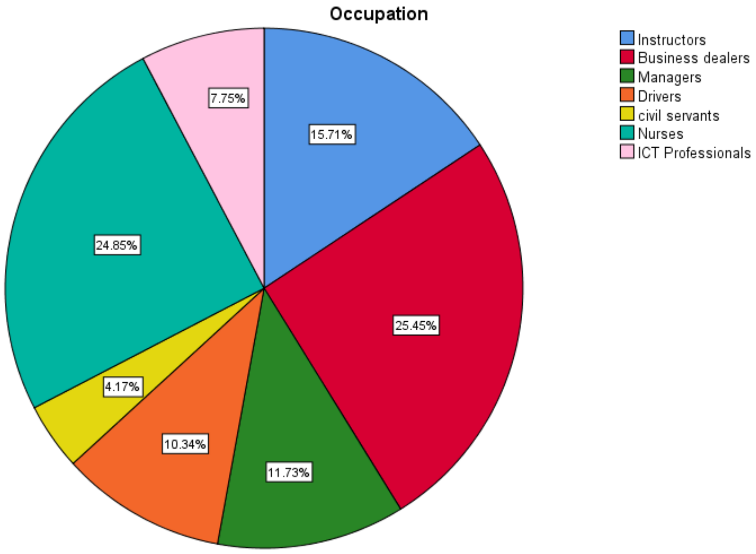
The level of education is important in determining the literacy level of the target population. From the analysed data, it was established that36.4% and 29.6% of the sampled respondents had diploma and certificate, respectively. Moreover, it was noted that 29.8% and 3.8% were qualified with degree and masters, respectively. Finally, 0.4% were qualified with doctorate degrees. This implied that all the respondents had basic literacy skills required for research to be effective.

4.2.5 Occupation

The occupational characteristics of the respondents was analyzed and presented in figure 8 as shown below.

Figure 8

Occupation



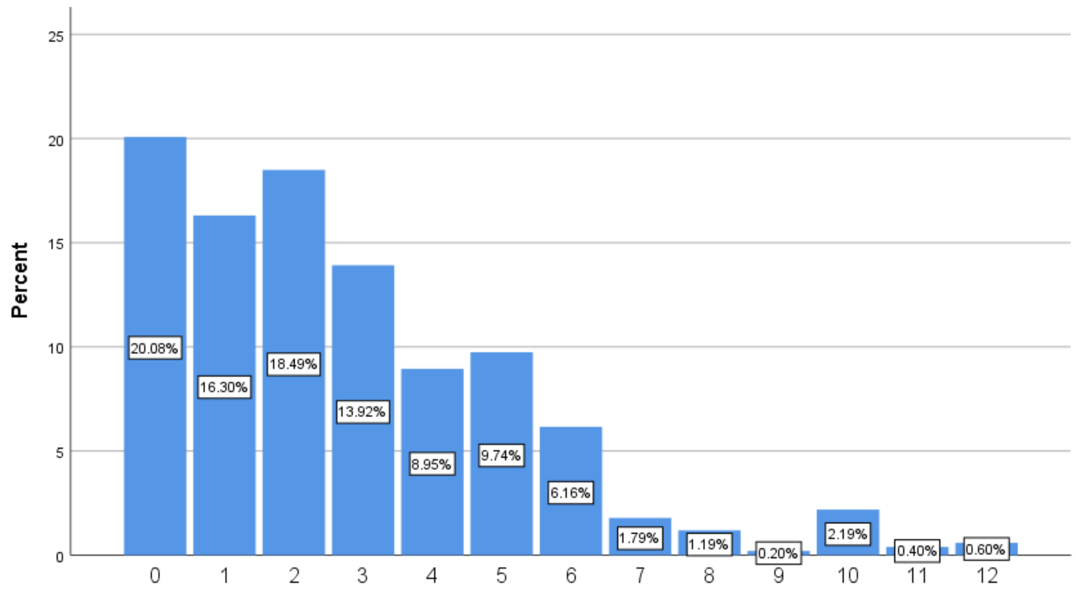
According to this analysis, it was recognized that majority of respondents were Business community (25.4%) and (9.9%) and Nurses (24.9%). Similarly, instructors and Managers represented 15.7% and 11.7% respectively. Furthermore, drivers and ICT professionals were represented by 10.3% and 7.8% respectively.

4.2.6 Participation

Respondents were asked to indicate the number of times they had participated in public participation forums in the county since devolution began. The findings are presented in bar-graph as shown in Figure 9.

Figure 9

Number of times in participation



The findings shows that a total of 79.9% of the respondents had participated in affairs at the county Government while 20.1% had not participated in public participation forums before. Those who had participated varied with the number of times they participated. For example, 18.5% had participated twice while 0.6% had done it up to 12 times.

4.3 Challenges faced in obtaining Sentiments in Public Participation Forums.

In an attempt to analyse the challenges faced in obtaining sentiments in public participation forums, respondents were asked to indicate the level of agreement with the identified factors. The findings are presented in Table 8.

Table 8

Challenges faced in obtaining Sentiments in Public Participation Forums

Statements	SD(%)	D(%)	N(%)	A (%)	SA (%)
We usually have limited time for everyone to fully contribute in the Public Participation discussions	16.1	6.4	11.3	22.9	43.3
Often few people get the chance to express their views in the PP	16.5	6.4	11.4	29.9	35.9
We are often unable to exhaust all the items in the PP forums	16.5	7.6	15.1	25.9	34.9
We are not able to capture each participants reactions adequately	17.0	7.0	16.2	28.5	31.3
We have challenges capturing the sentiments expressed by the participants in full	17.7	9.6	16.7	24.3	31.7
Often, we have difficulty in finding the right words to express our feelings towards a subject	18.7	13.7	15.7	27.8	24.1
We have challenges analyzing the sentiments of the participants in the PP	18.5	8.3	13.9	32.8	26.4
We would prefer the discussions on a subject begin online before the PP so that only the critical issues can be discussed in the PP sittings	16.3	5.4	14.9	33.8	29.6
We would prefer the discussions on a subject continue online after the PP so that we can exhaust subjects being discussed	17.1	4.6	15.9	34.8	27.6
Online discussions will enable everyone to have time to have time to adequately air their views on a subject and other members react to them	20.3	10.3	20.1	30.2	19.1
Online discussions will enable the participants to be very honest in their views	18.3	8.2	19.7	33.0	20.9
Through online discussions, we will be able to access adequate information of the discussion material	19.3	6.8	20.7	32.6	20.7

According to Table 8, 66.2% of respondents affirmed that they usually have limited time for everyone to fully contribute in the PP discussions. Furthermore, 65.8% agreed that few people get the chance to express their views in the PP. Furthermore, 60.8% agreed that they were often unable to exhaust all the items in the PP forums. These findings agrees with that of the Ministry of Devolution (2020) who reported that, reasons, such as absent leaders, time constraints were stipulated as the constraints to public participation.

The findings also established that 59.8% agreed that they were not able to capture each participant’s reactions adequately. This view was supported by 56% of the respondents who asserts that they have challenges capturing the sentiments expressed by the participants in full. The results similarly showed that 51.9% agreed that they frequently had difficulty in finding the right words to express their feelings towards a subject while 59.2% also agreed that they have challenges analyzing the sentiments of the participants in the PP. According to the Ministry of Devolution and Planning & Council of Governors (2016), before a public participation forum is initiated, it is the obligation of the county government to provide all the information on the subject matter of discussion, mechanisms of engagement and inform the public on what is expected of them.

It was also established that 63.4% agreed that they would prefer the discussions on a subject begin online before the PP so that only the critical issues can be discussed in the PP sittings. This view was maintained by 62.4% of respondents who agreed that they would prefer the discussions on a subject continue online after the PP so that we can exhaust subjects being discussed. Furthermore, 49.3% affirmed that online discussions will enable everyone to have time to have time to adequately air their views on a subject and other members react to them. Similarly, 53.9% of the respondents agreed that online discussions will enable the participants to be very honest in their views while 53.3% acknowledged that through online discussions, we will be able to access adequate information of the discussion material.

4.3.1 The extent of Public Participation

Public participation focuses on getting the thoughts, viewpoints, and views of the public openly and in a way that is inclusive and time limitations are minimized. Successful public participation therefore includes a transparent, accountable, and organized mechanism where decisions can be communicated with and informed by the public. An analysis was done to establish the extent of public participation. The findings are presented in Table 9.

Table 9

The extent of Public Participation

Statement	To a small extent	To some extent	To a moderate extent	To a great extent	To a very great extent
Incorporate a wide range of public values	16.3%	7.8%	11.5%	29.4%	35.0%
Be available to all public interests	18.3%	10.5%	15.9%	25.8%	29.4%
Allow for new participants over time	17.7%	9.3%	14.1%	32.2%	26.6%
Protect participants' identities when necessary	21.3%	12.5%	19.1%	29.8%	17.3%

The analysis revealed that 64.4% of respondents affirmed that public participation incorporates a wide range of public values to a great extent whereas 16.3% affirmed that public participation incorporates a wide range of public values to a small extent. Similarly, 55.2% were of the view that public participation was available to all public interests to a great extent though 16.3% stated that public participation incorporates a wide range of public values to a great extent while 18.3% and 10.5% of the respondents affirmed that public participation incorporates a wide range of public values to a small extent.

According to the table 9,58.8 % of the respondents correspondingly agreed that public participation allow for new participants over time to a great extent while 17.7% believed participation allowed for new participants over time to a small extent. In conclusion, it was noted that 47.1% of the respondents agreed that there was an element of protection of participants' identities when necessary to great extent while a whopping 21.3% believed there was an element of protection of participants' identities when necessary to a small extent.

4.3.2 Factor Analysis

Factor analysis is a statistical technique for identifying which underlying factors are measured by a (much larger) number of observed variables. During factor analysis, Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of Sphericity is employed. The findings are presented in Table 10.

Table 10

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.935
Bartlett's Test of Sphericity	Approx. Chi-Square	6183.600
	df	66
	Sig.	0.000

Kaiser-Meyer-Olkin (KMO) Test is a measure of how suited the data is for Factor Analysis. The test measures sampling adequacy for each variable in the model and for the complete model. The KMO value should be between 0.6 and 1 it indicates that the sampling is adequate (Stephanie, 2021) Otherwise KMO values less than 0.6 indicate the sampling is not adequate. Moreover, The Bartlett's Test of Sphericity should be significant ($p<0.05$). In this research, the KMO value was 0.935 while the Bartlett's Test of Sphericity is significant.

After factor analysis was done, rotated component matrix was generated and the findings presented in Table 11.

Table 11

Rotated Component Matrix[a]

	Variables	1	2
1.	We usually have limited time for everyone to fully contribute in the Public Participation discussions.	.836	
2.	Often few people get the chance to express their views in the PP	.872	
3.	We are often unable to exhaust all the items in the PP forums	.867	
4.	We are not able to capture each participants reactions adequately	.873	
5.	We have challenges capturing the sentiments expressed by the participants in full	.866	
6.	Often, we have difficulty in finding the right words to express our feelings towards a subject	.831	
7.	We have challenges analyzing the sentiments of the participants in the PP	.822	
8.	We would prefer the discussions on a subject begin online before the PP so that only the critical issues can be discussed in the PP sittings		.615
9.	We would prefer the discussions on a subject continue online after the PP so that we can exhaust subjects being discussed		.700
10.	Online discussions will enable everyone to have time to have time to adequately air their views on a subject and other members react to them		.861
11.	Online discussions will enable the participants to be very honest in their views		.893
12.	Through online discussions, we will be able to access adequate information of the discussion material		.895

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

The results in table 11 shows the rotated component matrix. This matrix indicates a group of variables that measures a given factor. Technically, a factor (or component) represents whatever its variables have in common. In this research, the component matrix (above) shows that the first component is measured by **seven** variables (variable 1-variable 7) while the second component is measured by **five** variables (8-12). The first component was then named as **human-based factors** while the second component was termed as **technological factors**. Generally, all the items in the two components had a factor loading above 0.3 leading to retention of all the variables for subsequent inferential analysis.

4.3.3 Correlation Analysis

An analysis was done to determine the relationships between the independent variables and the dependent variables. Spearman rho statistics was used, and the results presented in Table 12.

Table 12

Correlations matrix

			Public Participation
Spearman's rho	Human -Based Factors	Correlation Coefficient	.789**
		Sig. (2-tailed)	.000
		N	503
	Technological Factors	Correlation Coefficient	.697**
		Sig. (2-tailed)	.000
		N	503

**. Correlation is significant at the 0.01 level (2-tailed).

The results of the analysis revealed that there exist a statistically significant relationship between human-based factors and public participation ($r=0.789^{**}$;$p<0.05$).This implies that an improvement in human -based factors have a potential in enhancing public participation in County governments.

Moreover, the results indicated that there exist a statistically significant relationship between technological factors and public participation ($r=0.697^{**}$;$p<0.05$).This implies that advancement and uptake of technologies will improve public participation in County governments.

4.3.4 Regression Analysis

The goal of regression analysis is for prediction. Multiple Regression analysis was computed to predict public participation using human-based factors and technological factors. The model summary is presented in Table 13.

Table 13

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.943[a]	.889	.888	.39

a. Predictors: (Constant), Technological Factors, Human -Based Factors

The models R Square and the Adjusted R Square is 0.889 and 0. 888. This confirms that up to 88.8% variation in public participation is influenced by the variation of Human-based and technological factors with 11.2% as the unexplained variation which could be influenced by factors outside the model. The standard error of the estimate is 0.39.

4.3.5 ANOVA

The model significance was tested at 0.05 alpha as presented in Table 14.

Table 14

ANOVA[a]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	610.179	2	305.089	1993.479	.000[b]
	Residual	76.522	500	.153		
	Total	686.701	502			

a. Dependent Variable: Public Participation

b. Predictors: (Constant), Technological Factors, Human -Based Factors

The model is statistically significant at 0.05 alpha level, $F (2,500) = 1993.479; p < 0.05$. This implies that the predictors were significant in predicting the dependent variable.

4.4 Model Design

This section answers the research question on how the model was designed.

4.4.1 Design Process

The model design was done using software engineering and modern design processes. The following are particular design processes and activities that were applied;

(a) **Requirements Analysis**: To achieve the deliverables of the model, enumeration of all the system requirements was made and design and development of each expected deliverable was planned. PHP programming language was used to program system logic because of its ability to integrate with mBERT Google NLP. Other languages used included, bootstrap library, JavaScript and MySQL database.

(b) **Specification**: Module by module planning was done to establish the features that each module had, and how different modules were integrated.

(c) **Software architecture**: Relationship of different components was drawn. This provided an abstraction of the relationships that different components of the sentiment analysis system had.

(d) **Implementation**: different components of the model were developed primarily using PHP, JavaScript, MySQL and bootstrap library. Besides Google cloud NLP and Client translate packages (mBERT) were integrated. Rapid prototyping method was considered most ideal and therefore was applied in implementation of the sentiment analysis model.

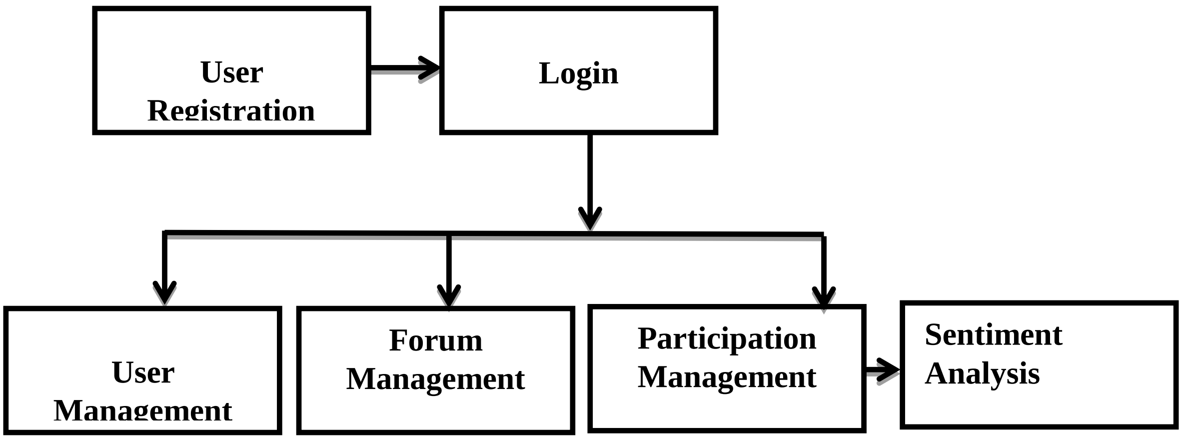
(e) **Evaluation**: System evaluation was done though objective-based evaluation method where various expected deliverables of the model were outlined and each deliverable tested against the specification. Where the deliverables did not meet the expected specifications, then particular module was redesigned and retested until the specifications were achieved. This ensured that the codes of different modules work together as one unit.

4.4.2 Modular Design

The sentiment analysis model contains several modules that are integrated, namely, user registration, login and roles management module, forum management module, participation management module, sentiment analysis module. Figure 10 below presents the integration of different modules that form Jumuika sentiment analysis model.

Figure 10

System Architecture



Source: Researcher (2023)

The following section presents a summary discussion of the different modules that form sentiment analysis model design;

(a) **Registration Module:** The section that allows the users of the system (admins and public users) register either through self-registration as is the case with public users or by admins through user management module. The registered users can therefore access other modules of the system based on their roles. The bio data that the user fills in during registration becomes vital in authenticating the users.

(b) **Login Module:** This is the module that authenticate users who are registered and authorizes them to explore other modules based on their roles by setting up sessions for them. This module therefore acts as an entry point to the sentiment analysis system for registered users.

(c) **User Management Module:** This module allows the admin users to add other uses to the system, drop users, update user details and assign roles to the system users.

(d) **Forum Management Module:** The forum management module allows the system admins to create new forums where other users can engage, edit existing forums or drop forums.

(e) **Participation Management Module:** The primary aim of this study was to develop a model that would assist in automating public participation on projects that county governments wish to implement. This module therefore is fundamental such that it allows the public users to post their comments about various topical issues that require participation and the admins to read general feeling of the public about the projects in discussion. To do so, the admin can read the comments (raw as well as translated).

(f) **System Navigation:** This module allows the users to traverse the system easily by based on their roles. To achieve this, system has a side menu that contains different menu items that guide the user in exploring different system features.

(g) **Sentiment Analysis module:** This module computes the average sentiment and magnitude scores then display the average feeling from participations ranging from positive to negative. This will guide the decisions by the county governments regarding implementation of projects for which public participation was sought.

(h) **System Dashboard:** This component displays quick view of system statistics regarding the number of comments made by participants, the number of participations that users can engage in, the number of forums created and the number of public user or citizens registered.

4.5 System Implementation

On the overall, this section offers a discussion of how sentiment analysis model was implemented. Section 4.5.1 presents the purpose for which the model was developed while section 4.5.2 presents the Computation of popularity and Subjectivity from the Comments. Further, section 4.5.3 presents the functional overview of the system whereas sections 4.5.4 and 4.5.5 provide discussions on computational procedures for sentiment analysis metrics and design processes respectively.

4.5.1 Purpose of the Model

The primary purpose of the study was to develop a sentiment analysis model for use in public participation forums in County Governments in Kenya. As such, this model aimed at easing the otherwise difficult process of obtaining sentiments from the public about county government projects as required by the law. The automation of public participation process through this makes it more effective for county governments to obtain and analyse sentiments from the public participations and make appropriate decisions thereof.

4.5.2 Computation of popularity and Subjectivity from the Comments

To obtain the average feeling from participations, the model gets the average sentiment scores and magnitude scores. The magnitude and score values were obtained using as set of open-source PHP package libraries for natural language processing (NLP), namely; cloud natural language processing to analyze comments sentiments and cloud translate client provide translation for Swahili/Sheng translation to English in order to analyze sentiment in English. Swahili and Sheng as local languages are supported by the packages during sentiment analysis.

4.5.3 The functional Overview

The overall expectation of the system was to allow user logins, admin logins, creation of projects and participation forums by admins, posting of comments by public users, and translation and analysis of sentiments by the model. As such, it was expected to take into consideration the inputs from respondents in English, Swahili or sheng languages then translate them to English for analysis of sentiments.

4.5.4 Computation Procedure

The model computes sentiments and magnitude scores by; first, setting up the configuration code which bares the path to the host of json code that accesses the cloud service with cloud NLP package. Secondly, detection of comment language is done to establish if the language is English or not. If the language is not English, then Translate Client package is invoked that will translate the comment to English. The translated comment is then used to analyse the sentiment using language client class under NLP package. Finally, pre-trained Bidirectional Encoder Representations from Transformers (mBERT) algorithm is used returns the magnitude and score values from cloud service containing basic language model.

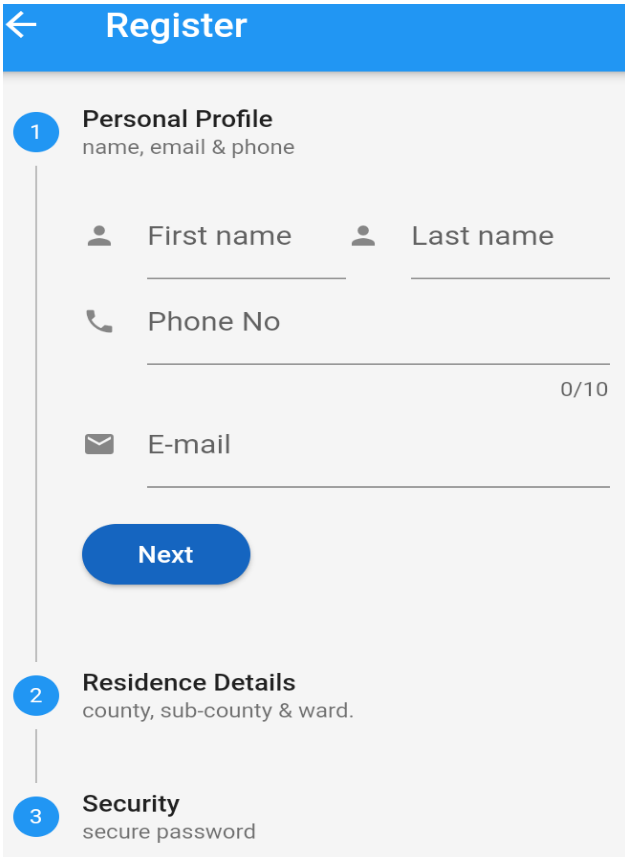
4.5.5 Module Interfaces

(a) Participant Registration Module

Participants do self-registration using Jumuika mobile application for them to get entry to the system and participate in projects posted in the platform that require public participation. Participant's information required during registration include; Participant's names, phone number, email address, details of residence (county, sub-county and ward), and their password. When the registration form is dully filled, the participant can then login and give their views about the projects on the platform that require them to participate. The system enforces password policy and stores user passwords as encrypted text on the database. Figure 11 presents the registration module view while Figure 12 presents the flowchart of the registration module.

Figure 11

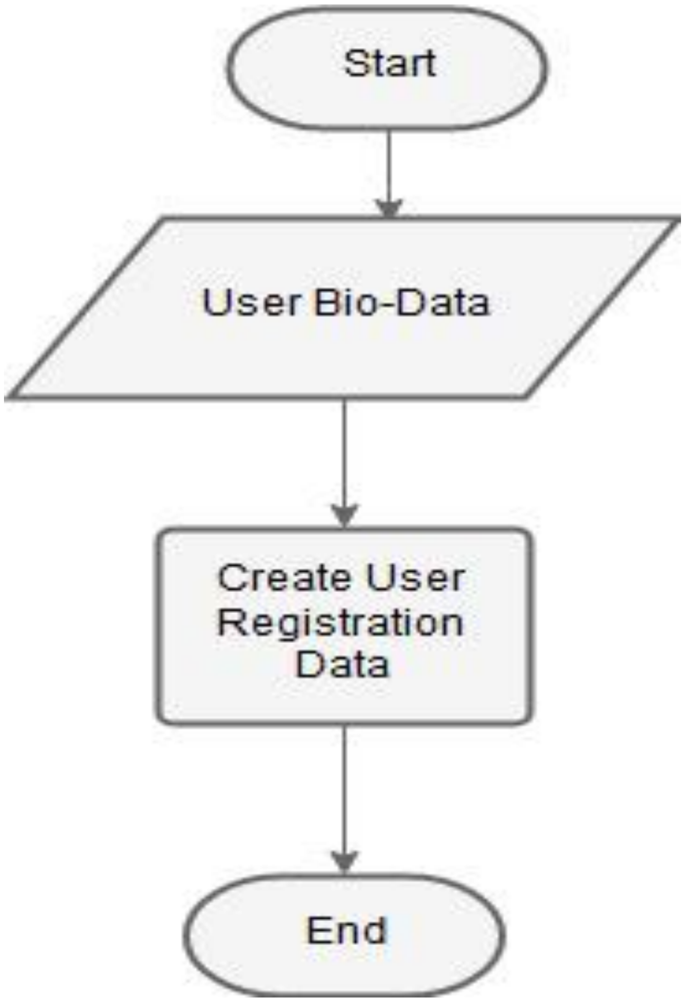
User Registration Form



Source: Researcher (2023)

Figure 12

Registration Flowchart



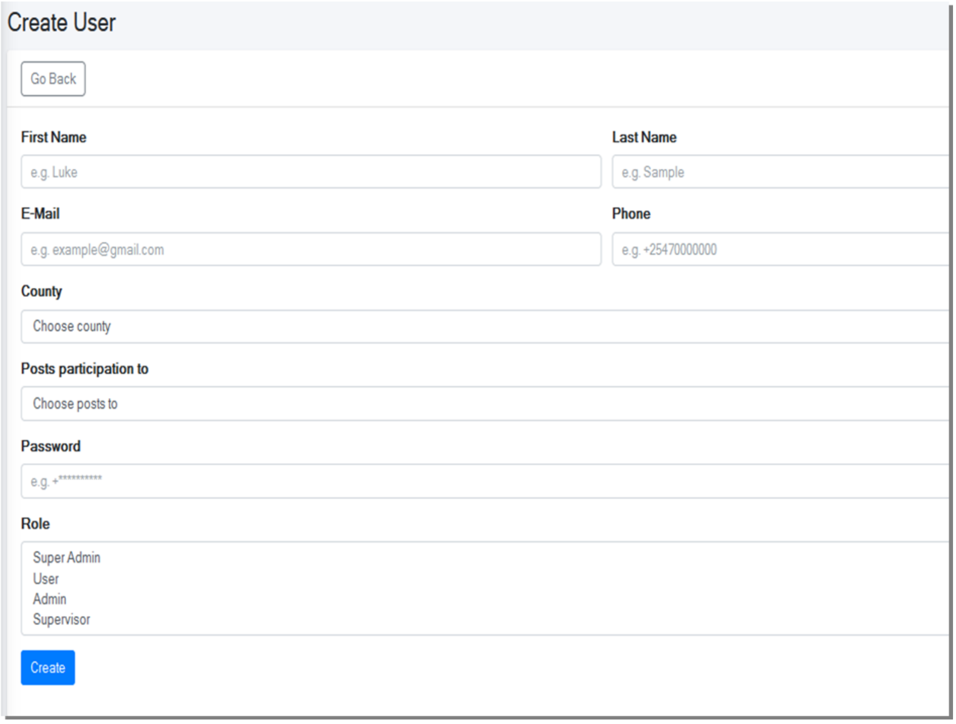
Source: Researcher (2023)

(b) User Management Module

The admins can create new users of the system, view all registered users and can further manage users by editing their user details, assigning roles to the users, and can remove active users from the platform. Figure 13 below shows the admin user registration module where admins can register new users. Figure 14 shows the users view page where the admin can edit the user details, assign roles or remove the users from the system.

Figure 13

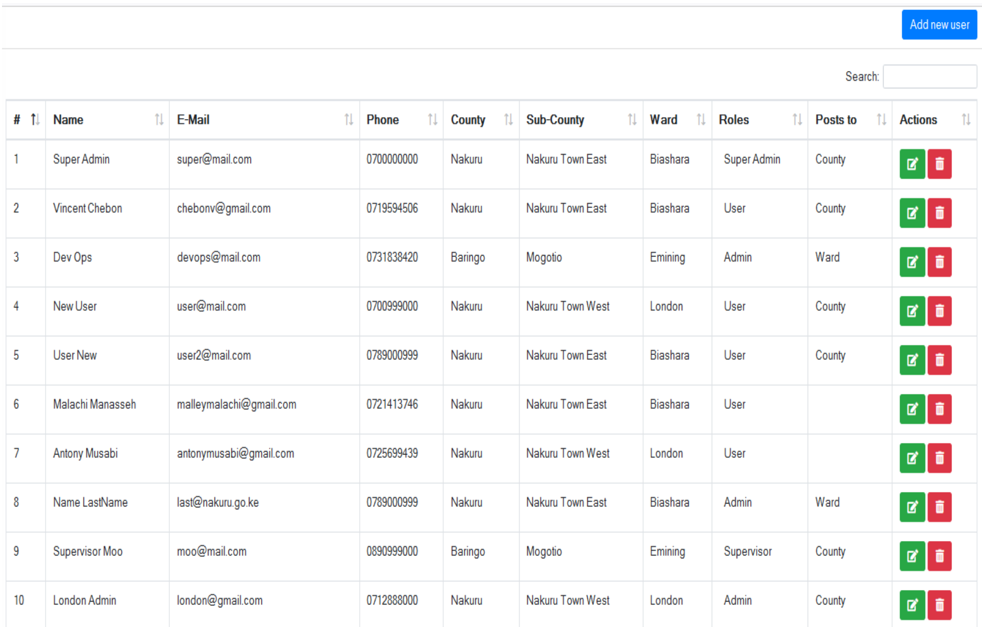
User Registration Page



Source: Researcher (2023)

Figure 14

Users View and Management


Add new user

Search:

#	Name	E-Mail	Phone	County	Sub-County	Ward	Roles	Posts to	Actions
1	Super Admin	super@mail.com	0700000000	Nakuru	Nakuru Town East	Biashara	Super Admin	County	
2	Vincent Chebon	chebonv@gmail.com	0719594506	Nakuru	Nakuru Town East	Biashara	User	County	
3	Dev Ops	devops@mail.com	0731838420	Baringo	Mogotio	Emining	Admin	Ward	
4	New User	user@mail.com	0700999000	Nakuru	Nakuru Town West	London	User	County	
5	User New	user2@mail.com	0789000999	Nakuru	Nakuru Town East	Biashara	User	County	
6	Malachi Manasseh	malleymalachi@gmail.com	0721413746	Nakuru	Nakuru Town East	Biashara	User		
7	Antony Musabi	antonymusabi@gmail.com	0725699439	Nakuru	Nakuru Town West	London	User		
8	Name LastName	last@nakuru.go.ke	0789000999	Nakuru	Nakuru Town East	Biashara	Admin	Ward	
9	Supervisor Moo	moo@mail.com	0890999000	Baringo	Mogotio	Emining	Supervisor	County	
10	London Admin	london@gmail.com	0712888000	Nakuru	Nakuru Town West	London	Admin	County	

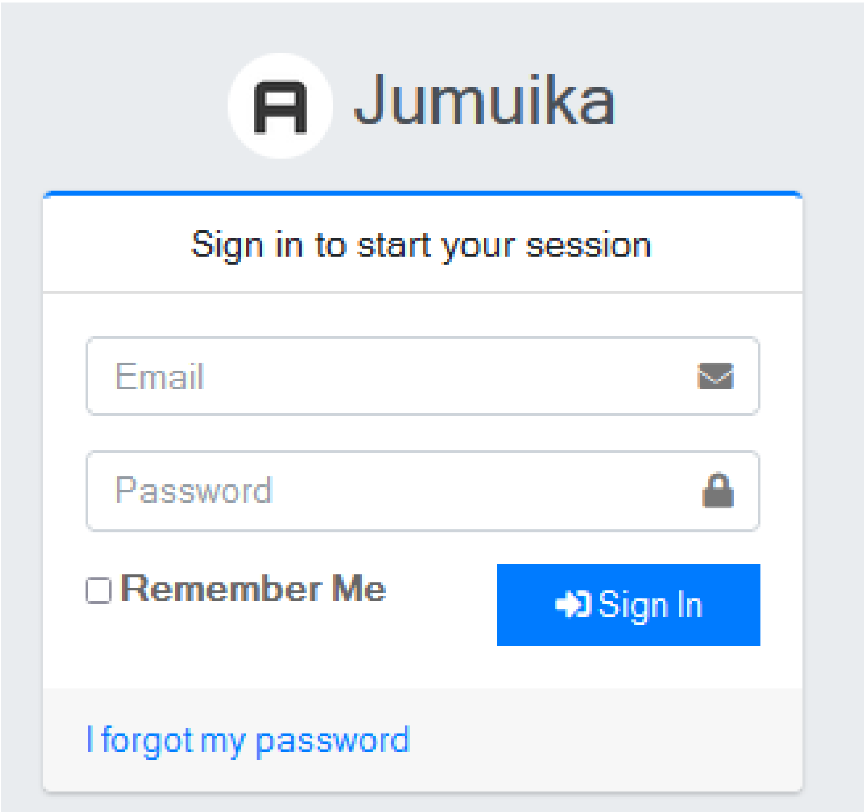
Source: Researcher (2023)

(c) Login Module

The public participation system has two login pages, one; the participant's login page from the mobile application where the registered users can gain entry to comments section upon authentication and authorization; two, the admin login where admins can access the platform based on their Figures 15 shows admin login form while figure 16 presents the public user login form.

Figure 15

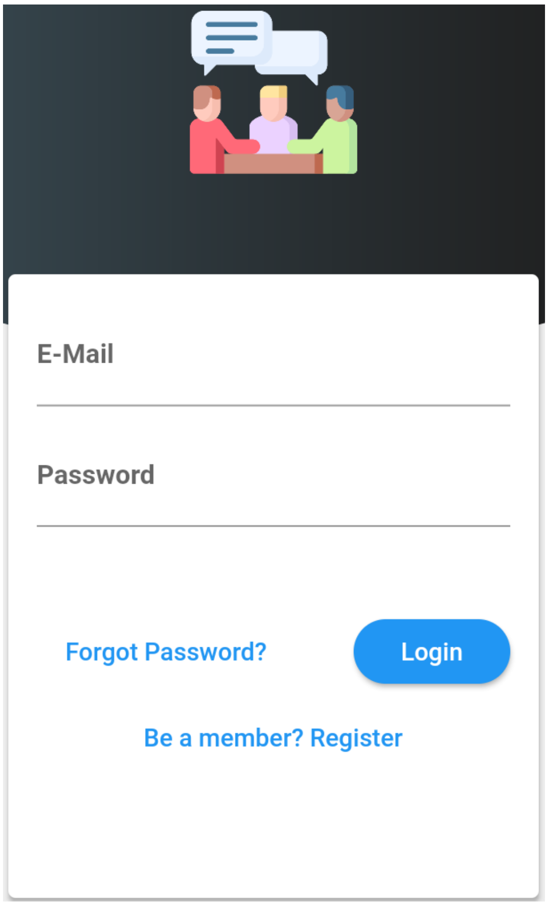
Admin Login Form



Source: Researcher (2023)

Figure 16

Participants Login Form



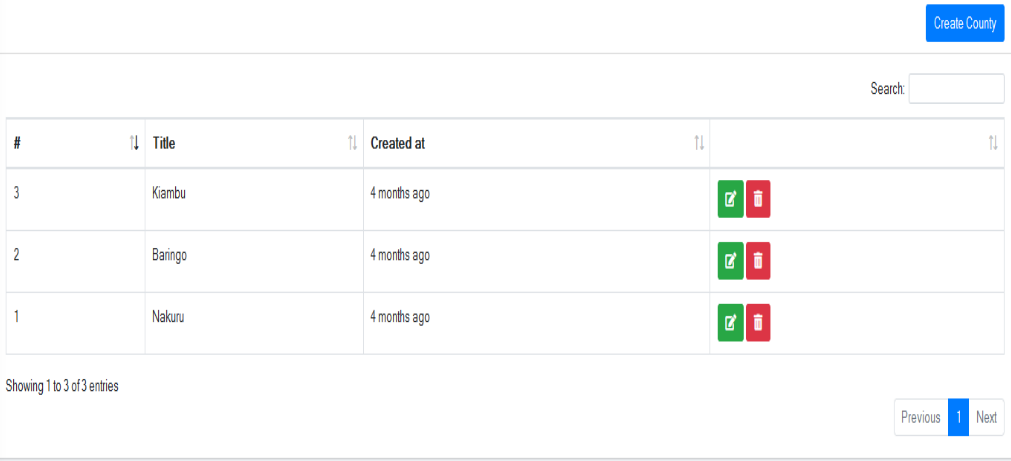
Source: Researcher (2023)

(d) Management of Counties

The model is scalable as to allow inclusion of additional counties, sub-counties and wards beyond the primary ones used in this study. Scalability is achieved through the system's ability to add new counties, sub-counties and wards. County management capability is only available to users with allocated roles to create, edit or delete the counties, sub-counties and wards. Figure 17 below shows county management page.

Figure 17

County Management Form



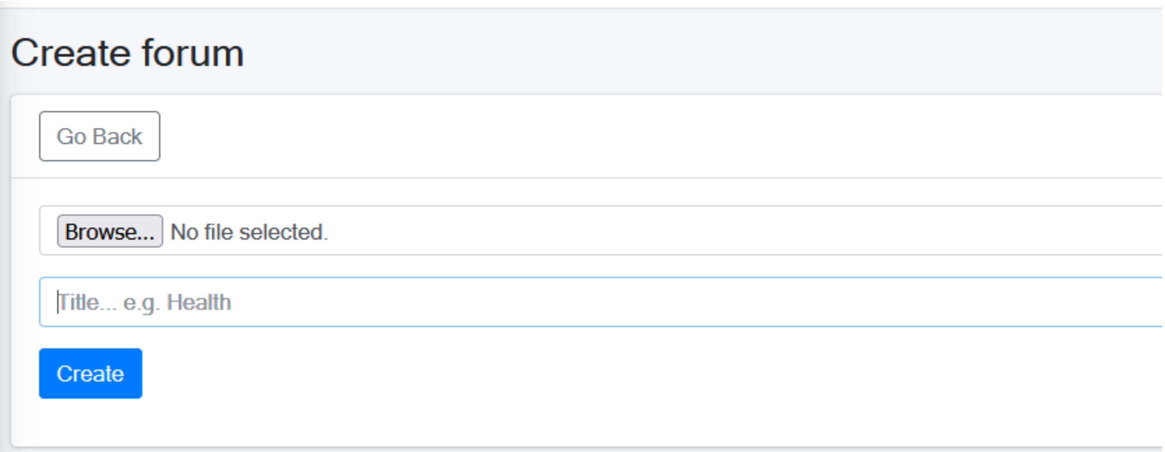
Source: Researcher (2023)

(e) Forums Management

The system allows the admins to create forums for which participation from public users can be collected and analysed. As shown in figure 18 below, the title, description and project photo of the forum are entered. Further admins can view all forums and can further manage the forums by editing or deleting created forums. Figure 19 below shows the admin forums' view.

Figure 18

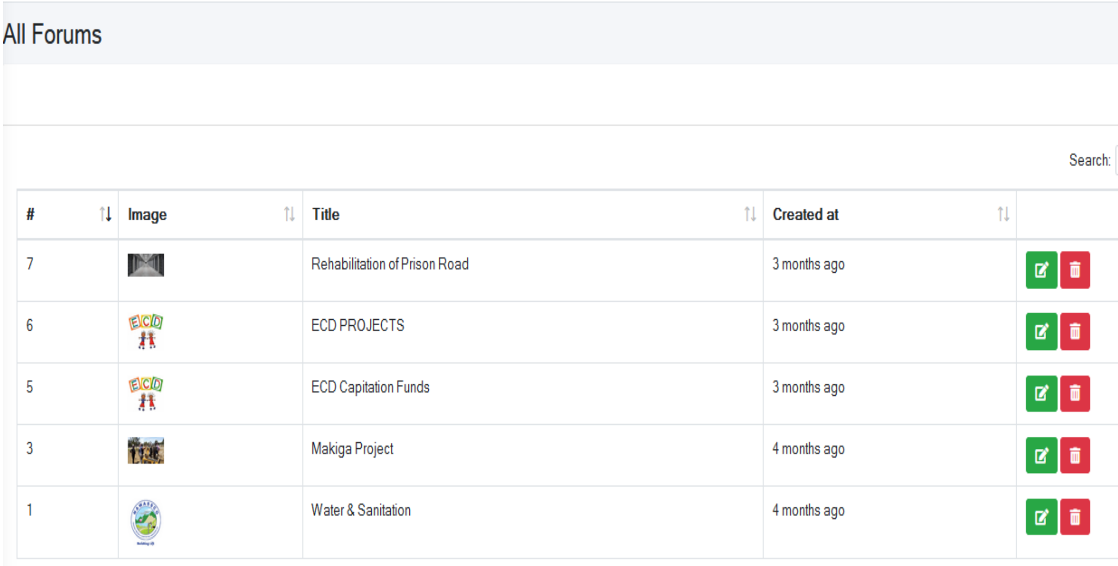
Forum Creation Form



Source: Researcher (2023)

Figure 19

Forums View



All Forums

Search:

#	Image	Title	Created at	
7		Rehabilitation of Prison Road	3 months ago	
6		ECD PROJECTS	3 months ago	
5		ECD Capitation Funds	3 months ago	
3		Makiga Project	4 months ago	
1		Water & Sanitation	4 months ago	

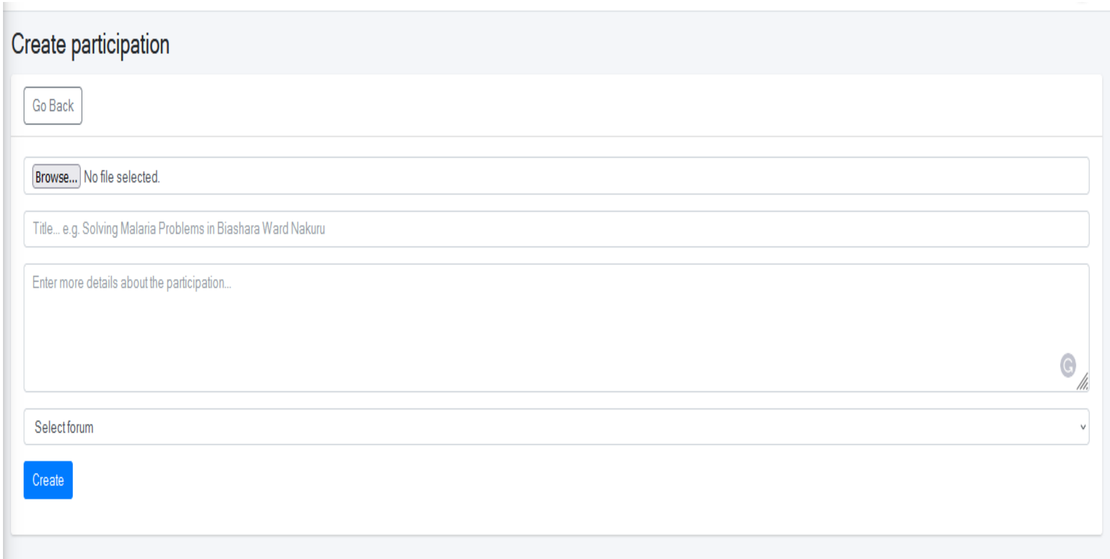
Source: Researcher (2023)

(f) Participations Management Module

This module enables the system admins to create and manage participation which contains vital information about the county projects or other issues that the county government would wish to get sentiments about. To achieve this, the admin registers the name of the project or issue, append relevant photos, give a detailed description about what the project entails and attach the participation to a relevant forum. Upon submission of the dully filled participation form shown in figure 20 below, the public users can access them and post the comments about how they feel about the projects or issues raised. The system admin can view all participations registered as presented in figure 21 and can further manage the participations by editing them or deleting them according to the roles they possess.

Figure 20

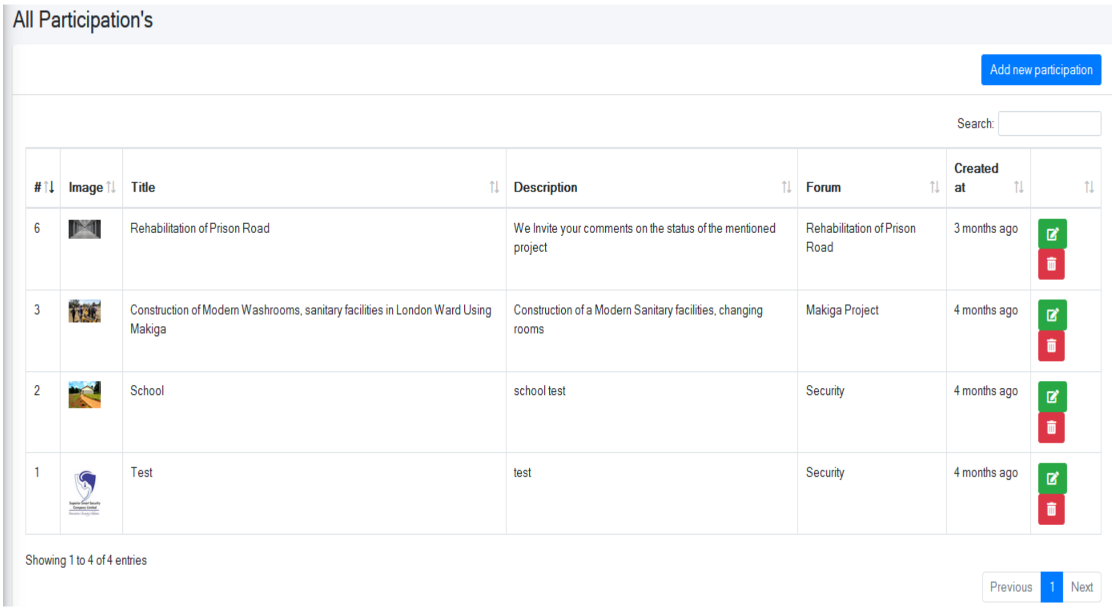
Forums View



Source: Researcher (2023)

Figure 21

Forums View



#	Image	Title	Description	Forum	Created at	
6		Rehabilitation of Prison Road	We Invite your comments on the status of the mentioned project	Rehabilitation of Prison Road	3 months ago	
3		Construction of Modern Washrooms, sanitary facilities in London Ward Using Makiga	Construction of a Modern Sanitary facilities, changing rooms	Makiga Project	4 months ago	
2		School	school test	Security	4 months ago	
1		Test	test	Security	4 months ago	

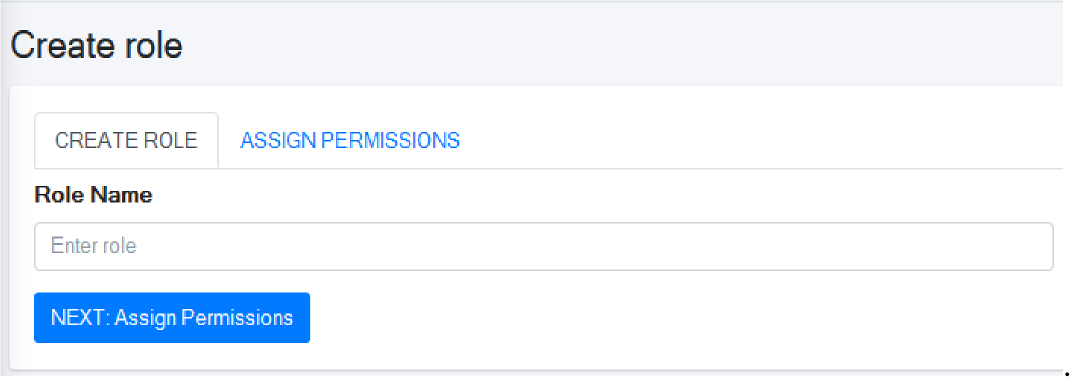
Source: Researcher (2023)

(g) User Roles Management Module

This module enables the system admin to manage user roles and by extension define what each user category can do in the system. To manage roles, the system admin can create new role types as shown in figure 22 below and assign permissions to those roles as shown in figure 23. The permissions grantable to the user roles include, and not limited to; read comments, create forums, update forum, edit forums, delete forums, create participations, read, and read participations.

Figure 22

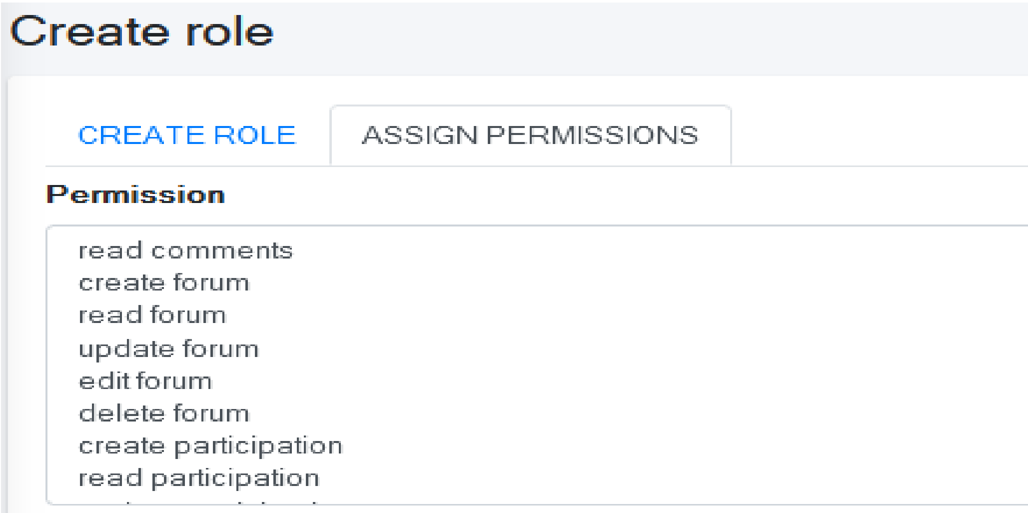
New Roles Form



Source: Researcher (2023)

Figure 23

Permission Assignment



Source: Researcher (2023)

(h) Sentiment Analysis Module

This section presents the epicenter of this study – to develop a sentiment analysis model. When the public user posts their comments on what their feelings are about the projects, first, their comments are translated if presented in Kiswahili or Sheng to English then sentiment as shown in figure 24 and magnitude and sentiment scores are computed using pre-trained mBERT algorithm incorporated into the system and the output is as shown in figure 25. As such, the general feeling predicted and the polarities of the comments are also predicted based on the scores. Besides, the cumulative feeling is computed which guides the county administration on the decision to make about the project based on the sentiments.

Figure 24

Sentiment Analysis

All Comments

Show 10 entries

#	Name	Comment	Polarity
8	Anonymous	hi samo imeweza sana " hi samo has been very successful "	Positive
7	Anonymous	tunashukuru sana " we are very grateful "	Positive
6	Anonymous	iko poa " iko poa "	Slightly Positive
5	Anonymous	ndio " yes "	Slightly Positive
4	Anonymous	Asante sana " Thank you very much "	Positive
3	Anonymous	good work	Positive

Source: Researcher (2023)

Figure 25

Scores and Polarity Indicator

General Comments Analysis

Avg. Sentiment Score	0.55
Avg. Magnitude Score	1.0416666666667
General feeling:	Slightly Positive 😊
Participation Verdict: Majority of the general population are slightly positive with this participation to proceed.	

Source: Researcher (2023)

Figure 25 above represents the output of the users' sentiments about a single issue that was asked to give comments about, namely; "Nakuru airport Project". The general feeling of the project by the participants indicated slightly positive.

The PHP code snippet presented below figure 26 demonstrates how magnitude and sentiment score values are computed as well as how the polarities are determined.

Figure 26

Sentiment score values computation

```
$config = [
   'keyFilePath' => config('services.keyFilePath'),
   'projectId' => config('services.projectId'),
];

$targetLanguage = 'en';  // Language to translate to

$translate = new TranslateClient($config);
$detect = $translate->detectLanguage($request->comment);
$comment  = new Comment();
$comment->user_id  = $request->user_id;
$comment->participation_id  = $request->participation_id;
$comment->comment  = $request->comment;
# Instantiates  a client
$language  = new LanguageClient($config);

# The text to analyze
$text = $request->comment;
if ($detect['languageCode'] != "en") {
   $result  = $translate->translate($request->comment,  [
      'target' => $targetLanguage,
   ]);
   $comment->translated_comment  = $result['text'];
   $text = $result['text'];
   Log::debug("Source language: " . $result['source']);
   Log::debug("Translation:  " . $result['text']);
}

# Detects the sentiment of the text
$annotation  = $language->analyzeSentiment($text);
$sentiment  = $annotation->sentiment();

$comment->subjectivity  = 'Sentiment Score: ' . $sentiment['sco
$sentiment['magnitude'];
$comment->score  = $sentiment['score'];
$comment->magnitude  = $sentiment['magnitude'];
```

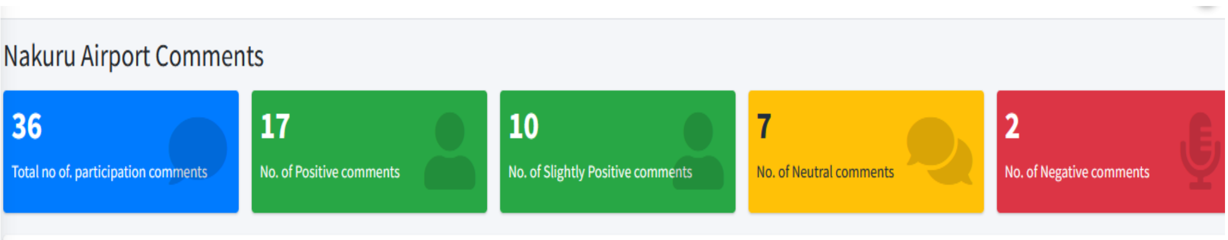
Source: Researcher (2023)

The average of all comments magnitude and score value based on the participation was done to obtain general sentiment about the participation. The chart below presents the bounds for sentiment and magnitude scores that were implemented in the code snippet of the model presented in figure 26 and corresponding output shown in Figure 25.

Figure 27

Participants Sentiment Summary

Sentiment	Compute Values
Clearly Positive*	“score” >= 0.8 &&“score” <= 1.0, “magnitude” >= 3.0
Positive	“score” >= 0.8 &&“score” <= 1.0, “magnitude” < 3.0 or “score” >= 0.3 &&“score” <= 0.8, “magnitude” >= 3.0
Slightly Positive	“score” >= 0.3 &&“score” <= 0.8, “magnitude” < 3.0
Clearly Negative*	“score” >= -1.0 &&“score” <= -0.6, “magnitude” >= 4.0
Negative	“score” >= -0.6 &&“score” <= -0.1, “magnitude” < 4.0 or “score” >= -0.6 &&“score” <= 0.1
Neutral	“score” >= 0.0 &&“score” <= 0.1, “magnitude” < 4.0 or “score” >= 0.1 &&“score” <= 0.3
Mixed	“score” >= 0.0 &&“score” <= 0.1, “magnitude” >= 4.0



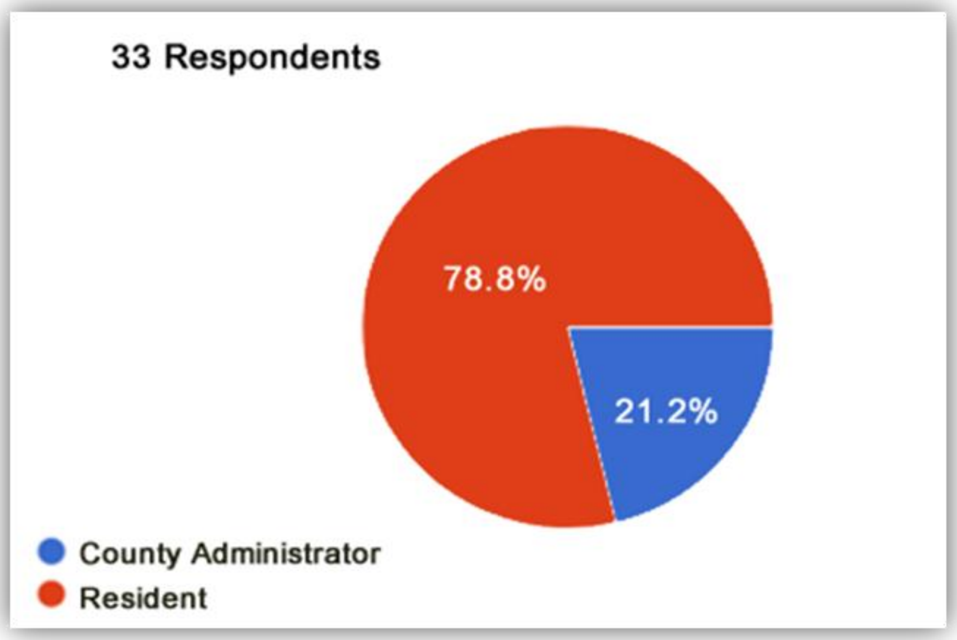
Source: Researcher (2023)

4.6 Model Validation

This section answers the last question of the study on how the sentiment analysis model was validated. To establish the validity of the model, expert survey was conducted which involved a total of thirty-three (33) participants, 26 (78.8%) of whom were county residents and 7(21.2%) county administrators as shown in figure 28 below. A set of variables were used derived from one issue in Nakuru County, namely; "Nakuru Airport". All the thirty-three respondents were expected to give their comments about the singular issue using the model and also give their feedback through a form provided during validation process. The responses were captured and analysed for validity of the model as far as sentiment analysis is concerned. The validation results are provided in the sections below.

Figure 28

Expert Survey Participants



Source: Researcher (2023)

4.6.1 Validation Metrics

The model was further evaluated for four software metrics, namely; usability, reliability, efficiency, and functionality. As such, Six (6) expert survey questions were structured to help obtain corresponding data (responses) for the four validations metrics. The data for validation of the four metrics were captured as summarised in Table 15 below;

Table 15

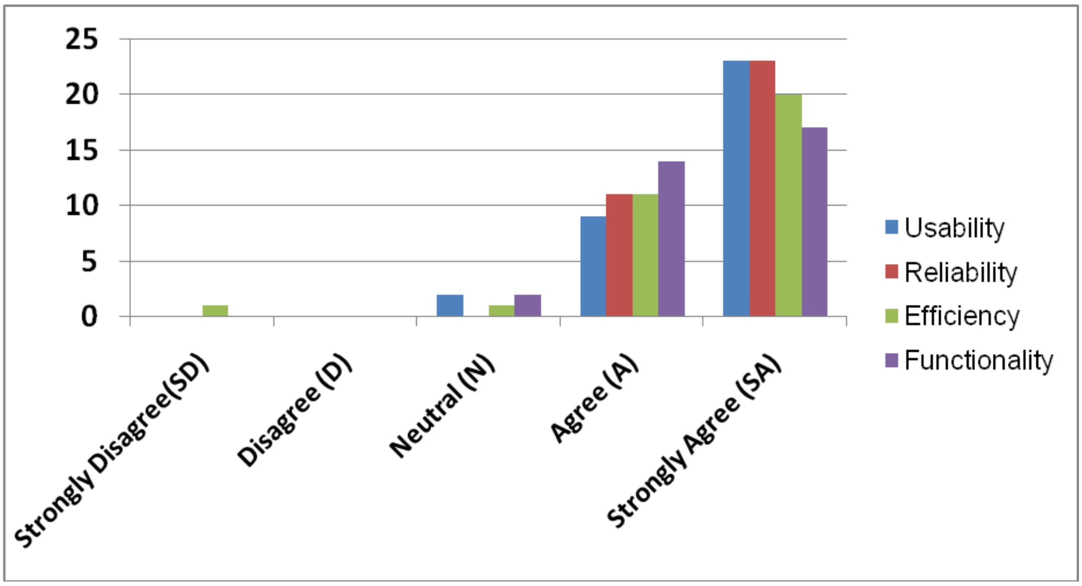
Validation Metrics Questions and Responses

Metrics		Question	SD	D	N	A	SA	TOTAL Responses
Usability	Q1	I like the overall experience I had with the public participation application	0	0	0	7	26	33
	Q2	I was able to navigate through the application without any challenge	0	0	3	11	19	33
		Usability Means	0	0	2	9	23	33
Efficiency	Q3	I could not use the application without registering and logging in.	2	0	2	13	16	33
	Q4	The system allowed me to register and take part in public participation for my county	0	0	0	9	24	33
		Efficiency Means	1	0	1	11	20	33
Reliability	Q5	I managed to complete the assessment without a challenge	0	0	0	10	23	33
Functionality	Q6	I think my opinions were captured well and will count on the overall.	0	0	2	14	17	33

Mean of responses for questions capturing the same metrics were obtained and analysis for specific metrics described. The results indicated that all the experts (100%; SA=26, A=7) considered that the sentiment analysis model was usable based on the overall experience they got using the application while a majority of them (90.9%; SA=19, A=11) considered it usable based on their ability to navigate through the application without challenge. An average of 94% (SA=20, A=11) further agreed that the application was efficient based on authentication and authorization. All the experts (100%; SA=23, A=10) agreed that the application was reliable while 94% (SA=17, A=14) considered the system to be functional. The analysis, as indicated in Figure 29 below, showed that all the responses for all the four metrics were skewed towards Agreed and Strongly Agreed. This means that the model was Usable, Functional, Reliable and its Efficiency was valid.

Figure 29

Validation Metrics



Source: Researcher (2023)

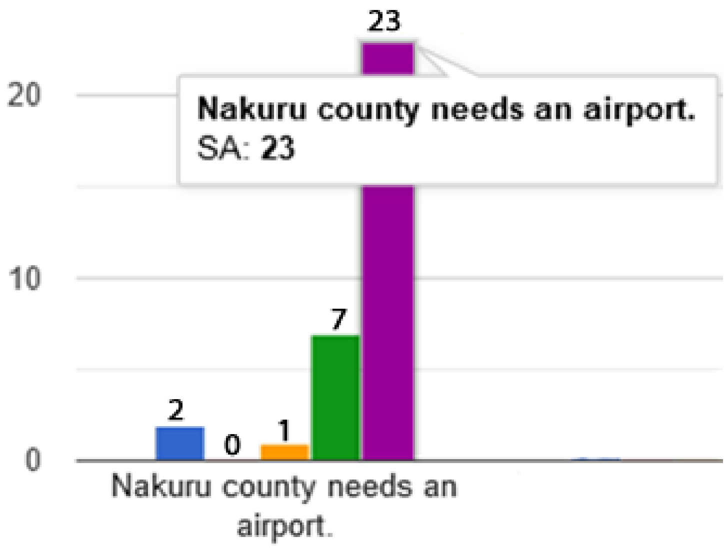
4.6.2 Validation Variables

(a) Validation Variable 1: Nakuru County needs an airport

From the responses, 90.9% (30 respondents, SA=23 and A=7) commented that having an airport in Nakuru county would be a good idea. Less than 10% of the respondents were either neutral or disagreed to having the airport in the county. As such, the model may be considered as valid basing on the rule of majority. Figure 30 below shows the distribution of responses for validation variable 1.

Figure 30

Validation Variable 1 Results

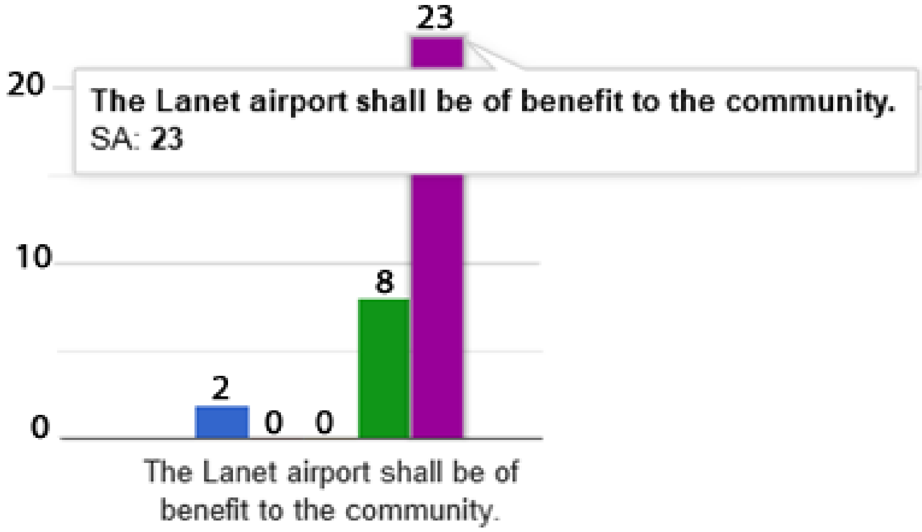


Source: Researcher (2023)

(b) Validation Variable 2: The Lanet airport shall be of benefit to the community. Participants were required to give their opinions on how they envisage the airport benefit the community. 93.9% (31 respondents, SA=23 and A=8) of the respondents indicated that they agree that the airport will benefit the community. Again, from the principle of majority this indicates that the model may be considered valid. Figure 31 below shows the distribution of the responses for variable 2.

Figure 31

Validation Variable 2 Results



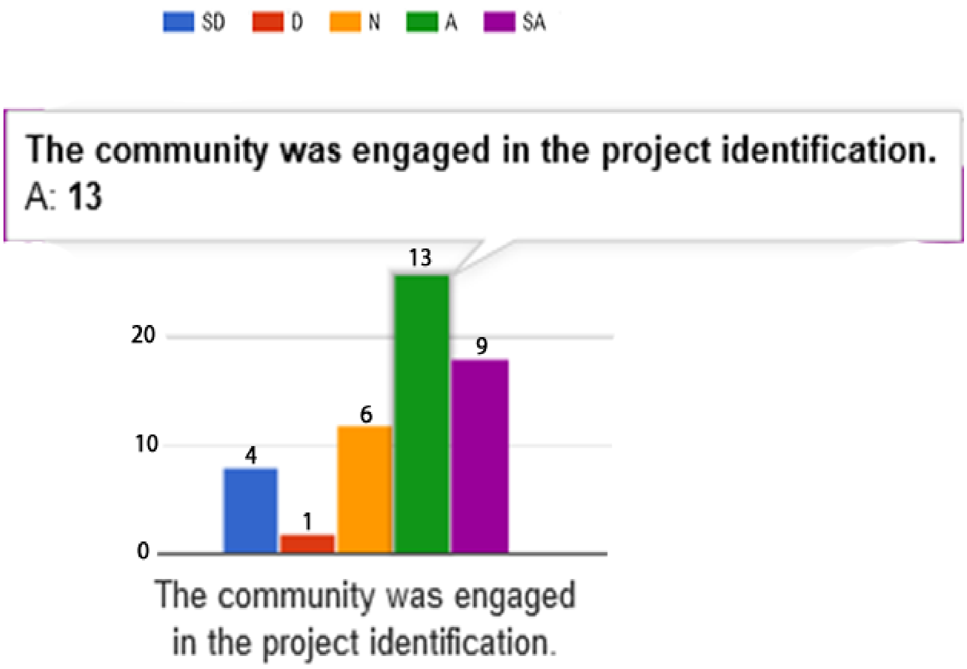
Source: Researcher (2023)

(c) Validation Variable 3: The community was engaged in the project identification.

Validation process required participants to comment on the engagement of the community in project identification. From the responses, 67% agreed or strongly agreed that the community was involved in the airport project identification. This means that the public participated in the airport project from the onset. Figure 32 below presents the results of variable 3 which, by principle of majority, validate the system for public participation.

Figure 32

Validation Variable 3 Results



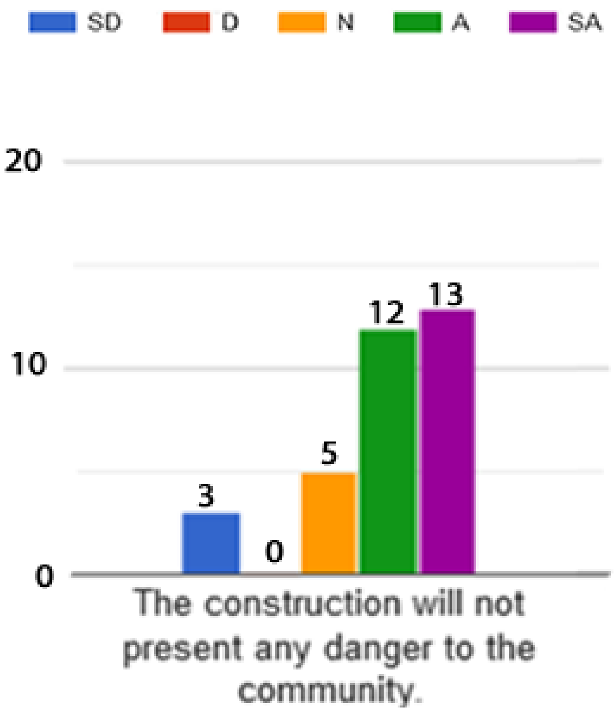
Source: Researcher (2023)

(d) Validation Variable 4: The construction will not present any danger to the community.

The survey also required comments from participants about how they think the airport will pose any challenge to the community. The responses indicated that a majority at 81.8% (27 respondents, SA=13 and A=12) agreed that the construction of the airport will not pose any danger to the community. The model can be considered again as valid based on this majority principle on variable 4. Figure 33 below indicates the distribution of responses for variable 4.

Figure 33

Validation Variable 4 Results



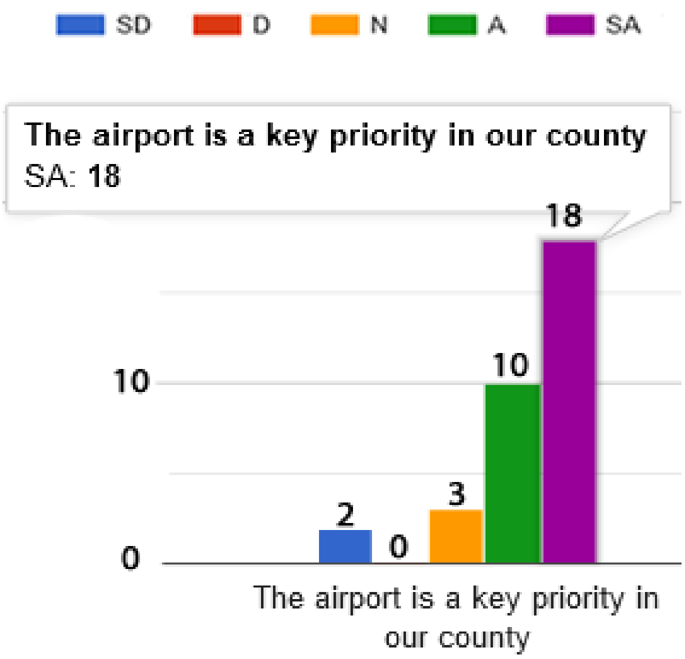
Source: Researcher (2023)

(e) Validation Variable 5: The airport is a key priority in our county

Finally, comments were required about whether the respondents consider the airport project as a priority or not. The results indicated that the majority of the participants (84.4%; SA=18, A=10) considered that the project was a priority. Figure 34 below shows the distribution of the responses on variable 5. Based on the results, the model can be considered as valid.

Figure 34

Validation Variable 5 Results



Source: Researcher (2023)

(f) Feedback Form Results versus Model Participation output

The feedback from the survey forms on the same airport project coincided with a general feeling of comments posted to the system about the same issue as positive. The system however goes further to translate the comments and compute the sentiment analysis ratings. This therefore further validates the system as useful in public participation for county governments. Figures 35 and 36 present the survey comments and system comments respectively.

Figure 35

General feeling Results from Expert Survey

On your own opinion kindly give your **general feeling towards the construction of the Lanet Airport with respect to its need, benefits, community engagement, safety and priority.**

33 responses

Good idea

Nakuru County needs an airport for effient movement of goods and services. It's safe and as a community we were fully involved

The project is timely because it will open Nakuru to the international Market. Improve service delivery because the community was fully involved and all safety measures were put into consideration

It is a good initiative

This project is timely because Nakuru will benefit greatly to boost the economy, enhance tourism and export of goods and services.The location is safe and the community is certisfied with it.

It will be a useful infrastructure towards the growth of the county.

A good initiative brought to the Nakuru county community

It will attract tourists in nakuru County

Source: Researcher (2023)

Figure 36

General feeling Results from the Model

All Comments

Show 10 entries

#	Name	Comment	Polarity	Subjectivity	Emoji
123	Anonymous	we are eagerly waiting for the airport because it will serve well as Nakuru as a whole and also benefit us youths by creating employment for us and business within Nakuru will bloom .	Positive	Sentiment Score: 0.9, Magnitude: 0.9	🙂
122	Anonymous	The project is good. It will be very convinient and will save us time.	Slightly Positive	Sentiment Score: 0.7, Magnitude: 1.5	😊
121	Anonymous	It is a fantastic project. If this is implemented with alot of care and attention then this will lift up the standards of Nakuru city together with it's residents. It will also attract the tourists from all areas in and outside our country.	Slightly Positive	Sentiment Score: 0.6, Magnitude: 1.8	😊
120	Anonymous	the project is very good.	Positive	Sentiment Score: 0.8, Magnitude: 0.8	🙂
119	Anonymous	such a great idea,can really bring change in Nakuru.	Positive	Sentiment Score: 0.9, Magnitude: 0.9	🙂
118	Anonymous	eagerly waiting for the airport..	Positive	Sentiment Score: 0.9, Magnitude: 0.9	🙂
117	Anonymous	very interesting	Positive	Sentiment Score: 0.9, Magnitude: 0.9	🙂
116	Anonymous	Inawezekana " It is possible "	Neutral	Sentiment Score: 0.1, Magnitude: 0.1	😐
115	Anonymous	Jambo la busara sana kwa uchukuzi mji wa nakuru.nina kubaliana. " A very wise thing for transportation in the city of Nakuru. I agree. "	Positive	Sentiment Score: 0.8, Magnitude: 1.6	🙂
114	Anonymous	such brilliant idea.I support	Positive	Sentiment Score:	🙂

Source: Researcher (2023)

4.6.3 Paired Sample t-test Validation

A paired sample t test was run to determine if there was a statistically significant differences between two issues that were analysed via Jumuika application and from the survey responses. In analysing the means of sentiments, data was collected from 33 respondents using Jumuika app.Similarly, the same respondents were given questionnaire with the same issues raised in the Jumuika application. Data was collected and coded in the SPSS software. The collected data was transformed to obtain a scale variable for the first and the second pair. Paired sample t-test was run to establish whether there were significant differences in the means in the two groups of data collection. The data that were used in computing the means are presented below:

Table 16

Paired Samples T-test

ID NO	Category	Issue: Nakuru County needs an airport for efficient movement of goods and services.	Four trials (Manual)=X1 1st case	2nd case	3rd case	4th case	Mean= X1	App=X2 Data from Jumuika App=X2
81	County Administrator	,,	5	4	5	5	4.75	5.00
83	County Administrator	,,	5	5	4	5	4.75	4.00
84	Resident	,,	4	4	4	3	3.75	5.00
85	Resident	,,	5	5	3	4	4.25	5.00
92	Resident	,,	5	5	4	3	4.25	3.00
93	County Administrator	,,	5	5	5	5	5.00	4.00
94	County Administrator	,,	4	4	4	4	4.00	4.00
95	Resident	,,	5	5	4	3	4.25	5.00
96	Resident	,,	5	5	3	5	4.50	5.00
97	Resident	,,	3	5	4	4	4.00	4.00
98	Resident	,,	1	1	1	1	1.00	3.00
99	Resident	,,	5	5	5	5	5.00	4.00
100	Resident	,,	5	5	3	5	4.50	3.00
101	Resident	,,	5	5	5	5	5.00	5.00
102	Resident	,,	4	5	4	5	4.50	3.00
103	Resident	,,	5	4	4	5	4.50	5.00
104	Resident	,,	4	4	4	4	4.00	1.00
105	Resident	,,	5	5	1	5	4.00	3.00
106	Resident	,,	5	5	5	5	5.00	3.00
107	Resident	,,	4	4	3	4	3.75	4.00
108	Resident	,,	5	5	3	5	4.50	5.00
109	Resident	,,	5	5	5	4	4.75	5.00
110	Resident	,,	4	4	2	4	3.50	4.00
111	Resident	,,	5	5	4	5	4.75	3.00
112	County Administrator	,,	5	5	4	5	4.75	4.00
113	County Administrator	,,	1	1	1	1	1.00	4.00
114	Resident	,,	5	4	4	4	4.25	5.00
115	Resident	,,	5	5	5	5	5.00	5.00
116	Resident	,,	5	5	5	5	5.00	3.00
117	Resident	,,	5	5	5	5	5.00	5.00
118	County Administrator	,,	5	5	1	5	4.00	5.00
119	Resident	,,	5	5	4	4	4.50	5.00
120	Resident	,,	4	5	3	4	4.00	5.00

By computing the means X1 and X2, it was then entered to the SPSS software which provided the analysis in table 16,17 and 18 respectively.

Table 16

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Nakuru Airport Sentiment Index	4.2348	33	.93737	.16318
	Data from Jumuika App On opinion concerning construction of the Nakuru Airport	4.1212	33	.99240	.17275

Table 17

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Nakuru Airport Sentiment Index & Data from Jumuika App On opinion concerning construction of the Nakuru Airport	33	.170	.344

The paired samples correlations indicates that the two variables are not significantly correlated to each other (r=0.170; p=0.344).

Table 18

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Dev	Std. Error Mean	Lower	Upper	t	Df	Sig. (2-tailed)
Pair 1	Nakuru Airport Sentiment Index - Data from Jumuika App On opinion on construction of the Nakuru Airport	.11364	1.24388	.21653	-.32742	.55470	.525	32	.603

There was no significant difference between the sentiment questionnaire and the results from Jumuika application (t_{32}=0.525, p>0.05). This implies that the solution developed captures the sentiments in the same way as the alternative approaches.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents conclusions and recommendations for the design and implementation of sentiment analysis model. As such summary of how each research question was answered is discussed as well as, areas for further study from this research.

5.2 Conclusions

The primary focus of this study was to develop a sentiment analysis model to aid county governments in the jurisdiction of Kenya to automate public participations on projects and issues they wish to undertake. As such the project was supposed to improve compliance with the law by county governments as well as engage the citizens who are the major stakeholders in developments projects and other goings on in the county government.

The study finally delivered web-based as well as mobile-based integrated applications to enable citizens to actively participate in goings on within the county and shape the progress of development within their areas of concern. With Jumuika application, the county government administration can read public views from the citizens and make informed decisions thereof based on population's general feeling. The Jumuika solution can manage users, roles, participations, and forums and compute magnitude scores, sentiment scores, polarity and general feelings. Sections 5.2.1 through 5.2.4 present conclusions on how each of the four research questions was answered.

5.2.1 The challenges Encountered when obtaining sentiments in Public Participation Forums in County Governments

This study established that the uptake of traditional public participation exercises organized by county governments faced great challenges because, more than sixty per cent of citizens who would like to participate indicated that they lacked time to do so and therefore few would-be participants do not get the chance to do so. Even when some citizens get the chance to participate, the study further established that attendees – on usual cases - do not sufficient time to express all their views about projects or issues requiring public participation, therefore views that are collected in the process are either insufficient or not representative enough. Other constraints to public participated highlighted by the study included; absent leaders, and time constraints that makes it near-impossible to capture views from each participant adequately.

To mitigate these challenges, the respondents agreed that online forum discussions would help to engage the leaders freely and give as-many-as-possible participants time to present their sentiments. This gives many people time to exhaust their views and express themselves in local languages (Swahili or Sheng). Online participation was therefore preferred to be the most convenient means to discuss freely and honestly.

5.2.2 Suitable Design for Sentiment Analysis Model for County Government's Public Participation Forums

Requirement analysis from the respondents guided the design of the sentiment analysis model. The overall challenges were noted then and considered as driving forces for the design. Regression analysis established that public participation was a function of human based factors as well as technological factors. Review of useful open-source NLP libraries were considered during design to facilitate language translation and sentiment analysis once implemented.

5.2.3 How Sentiment Analysis Model for Public Participation Forums in County Governments be Implemented

The sentiment analysis model was implemented as an integrated platform of mobile application and web application. The mobile application was developed using android and hosted on Google play store to enable citizens to easily access and install. The citizens use the mobile application to register, login, and participate in county development projects. The web application enables the admins to login manage participations, users, roles, forums and review scores and sentiments from participation forums. To obtain the average feeling from participations, the model gets the average sentiment scores and magnitude scores using as set of natural language processing (NLP), namely; mBERT algorithm to classify and analyze sentiments of comments made in English, Swahili and/or Sheng languages. As such, translation of comments to English is made in order to analyze sentiment in English. Swahili and Sheng as local languages are supported by the mBERT during sentiment analysis.

5.2.4 How Sentiment Analysis Model for Public Participation Forums in County Governments Perform

To assess whether the sentiment analysis model performed or not, the model was tested for functionality, usability and performance. First, the application was validated using goal-based evaluation by the developers to ascertain that the requirements were delivered. This testing actually ascertained the system performed as required. Besides, external online users largely from the selected county governments were given access to register and use the system for public participated on selected projects. The system recorded success as well because users, on one hand, were able to register, login, navigate to forums, and participate by posting their views and getting feedback. Admins on the other hand were able to login, create forums and participations, administer users, view sentiment scores and manage users and their roles.

In validating the model further, an issue from the county that needed opinions from the general public was identified. The issue was then posted on the system. Moreover, a sample of 33 study participants were identified and asked to give feedback on the system. After sometime, the same respondents were given the same issue in a 5-point Likert scale document where they were asked to rate from strongly disagree, disagree, neutral, agree and strongly disagree. A comparison was made of the results of sentiment analysis from the system and the rating scale using the paired sample t-test. The paired samples correlations indicates that the two groups were not significantly correlated to each other since one was a sentiment analyser application while the other method was the conventional manual-based approach ($r=0.170$; $p=0.344$). The paired sample t-test indicated that there was no significant difference between the sentiment questionnaire and the results from Jumuika application ($t_{32}=0.525$, $p>0.05$). This implies that the solution developed captures the sentiments in the same way as the alternative approaches.

5.3 Recommendations

Kenya, on one hand, has forty-seven (47) county governments which get funding from the national government to undertake development and other projects. Besides, the same devolved governments are required by law to account for the projects they undertake right from the approval by the public communities to judicial spending. As such, these county governments should implement automated mechanisms to gather the general feelings of the public on projects they plan to undertake before they do so. Uptake of ICT by the devolved units is paramount because it reaches more people faster. Counties should therefore deploy sentiment analysis applications for their public participations' fora during this information age. This study focused on three counties for development of the model which precisely succeeded. It is therefore recommended that the model can be cascaded to all the 47 counties of Kenya for efficient and effective public participation exercises. The study further recommends for public participation policies to be reviewed in order to enforce automated public participation. For the citizens that cannot access the internet and the model for participation, the study recommends that further enhancement can be made to incorporate unstructured supplementary service data (USSD) to accommodate them.

REFERENCES

Allen, M. (2017). Post Hoc Tests: Tukey Honestly Significant Difference Test. *sage encyclopedia of communication research methods*, doi: 10.4135/9781483381411.

Ames, H., Glenton, C., & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication. *BMC medical research methodology*, *19*(1), 26.

Andrea, A. D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, *125*(3), 26-33.

Atkinson, R., & Flint, J., (2001). Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies.: *Social Research Update*, No. 33, http://www.soc.surrey.ac.uk/sru/SRU33.html (19. 3. 2018).

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An Overview of Sentiment Analysis in social media and its' application in Disaster Relief. *Sentiment Analysis and Ontology Engineering. Studies in Computational Intelligence, 639*, 313-340. doi: 10.1007/978-3-319-30319-2_13

Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11*(3), 374.

Consul. (2019). *CONSUL*. Retrieved from Consul Project: http://consulproject.org/en/

Crunchbase. (2019). *TweetFeel*. Retrieved March 01, 2019, from Crunchbase Inc.: https://www.crunchbase.com/organization/tweetfeel

Devlin, J. and Chang, M., 2018. *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*. [Online] Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. Available at: <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html> [Accessed 22 July 2022].

Dodds, P., & Danforth, C. (2009). Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies*, *11(4)*, 441-456.

Donald, M. K., Guyo, W., & Moronge, M. (2020). Moderating Effect of Public Participation on the Relationship between Political and Economic Relations on Service Delivery of County Governments in Kenya.

Dragan, I-M., & Isaic-Maniu, A. (2013). Snowball Sampling Completion. *Journal of Studies in Social Sciences*. 5(2), 160-177

Dudău, D. P., & Sava, F. A. (2021). Performing multilingual analysis with Linguistic Inquiry and Word Count 2015 (LIWC2015). An equivalence study of four languages. *Frontiers in Psychology*, *12*, 570568.

Etter, J., F., & Perneger, T., V., (2000). Snowball sampling by mail: application to a survey of smokers in the general population. *International Journal of Epidemiology*, 29 (1), 43 - 45.

Ganney, P. S., Pisharody, S., & Claridge, E. (2020) Chapter 9-Software engineering. *Clinical Engineering*, 131-168

Hakijamii. (2017). *Barriers and facilitators of citizen participation in governance processes in Nairobi County, Kenya.* Nairobi: Economic and Social Rights Centre.

Hodge, S. R. (2020). Quantitative research. In *Routledge Handbook of Adapted Physical Education* (pp. 147-162). Routledge.

Initiative for Social and Economic Rights. (2018). *An assessment of the role and effectiveness of Barazas in decision making processes.* Kampala, Uganda.

Kapucu, A., Kılıç, A., Özkılıç, Y., &Sarıbaz, B. (2021). Turkish emotional word norms for arousal, valence, and discrete emotion categories. *Psychological reports*, *124*(1), 188-209.

Kibuacha, F. (2021) *Mobile penetration and growth in Kenya*, *GeoPoll*. Available at: https://www.geopoll.com/blog/mobile-penetration-kenya/ (Accessed: December 26, 2022).

KenyaLaw. (2017). *LAWS OF KENYA: County Governments Act.* Nairobi: National Council for Law Reporting with the Authority of the Attorney-General.

KEO. (2018). *The Legal Framework for Public Participation in Kenya.* Retrieved from http://www.keo.or.ke/cispkenya/

Kharde, V. A., & S.S. Sonaware. (2016). Sentiment Analysis of Twitter Data: A survey of Techniques. *International Journal of Computer Applications, 139*(11), 5-15.

Kituyi, O., & Moi, E. (2021). Determinants of Public Participation in Budget Making Process in Devolved Governments in Kenya: A Case of Trans Nzoia County. *Journal of International Business, Innovation and Strategic Management*, *5*(1), 37-51.

Lu, H., Zhu, Y., Yuan, Y., Gong, W., Li, J., Shi, K. ... & Wang, F. Y. (2021). Social signal-driven knowledge automation: A focus on social transportation. *IEEE Transactions on Computational Social Systems*, *8*(3), 737-753.

Mayienda, R. (2020). *An Assessment of factors influencing public participation in Kenya Urban Roads Authority projects* (Doctoral dissertation, Strathmore University).

Mbithi, A., Ndambuki, D., & Juma, F. O. (2019). Determinants of public participation in Kenya county governments. *Journal of Asian and African Studies*, *54*(1), 52-69.

Ministry of Devolution and Planning & Council of Governors. (2016). *County Public Participation Guidelines.* Nairobi, Kenya.

Ministry of Devolution (2020) *Challenges of public participation*, *AHADI toolkit*. Available at: https://countytoolkit.devolution.go.ke/challenges-of-public-participation (Accessed: January 26, 2023).

Ministry of Devolution and Planning. (2016). *Devolution and public participation in kenya: Civil education handouts for participants.* Nairobi, Kenya: USAID-AHADI.

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, *11*(1), 81.

Ngigi, S., & Busolo, D. N. (2019). Devolution in Kenya: the good, the bad and the ugly. *Public Policy and Administration Research*, *9*(6), 9-21.

Omolo, A. (2011). *Policy proposals on citizen participation in devolved governance in Kenya.*Nairobi, Kenya: The Institute for Social Accountability.

Pollyanna, G., Benevuto, F., Araújo, M., & Chu, M. (2013). *Comparing and Combining Sentiment Analysis Methods.* Boston, MA: ACM. doi:10.1145/2512938.2512951.

Press, C. U. (2019). *Dictionary*. Retrieved from Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/sentiment

Quick, Kathryn S., and John M. Bryson. "Public participation." In *Handbook on theories of governance*, pp. 158-168. Edward Elgar Publishing, 2022.

Rambocas, M., & Gama, J. (2013). *Marketing Research: The role of sentiment analysis.* University of Porto, Porto, Portugal.

Ronoh, G., Mulongo, L. S., &Kurgat, A. (2018). Challenges of integrating public participation in the devolved system of governance for sustainable development in Kenya. *International Journal of Economics, Commerce and Man, VI* (1), 476-491.

Scott, F. (2019). *GANA Pienso: An Online Bridge between Citizens & Government*. Retrieved from Participedia: https://participedia.net/case/5654

Sedgewick, R. (2007). *The role of the scientific method in software development.* New Jersey: Princeton University.

Sharef, N. M., Zin, H. M., & Nadali, S. (2016). Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data. *Journal of Computer Sciences, 12*(3), 153-168. doi:10.3844/jcssp.2016.153.168

Shivanandhan, M. (2020) *What is sentiment analysis? A Complete Guide for Beginners*, *freeCodeCamp.org*. FreeCodeCamp.org. Available at: https://www. freecodecamp.org/news/what-is-sentiment-analysis-a-complete-guide-to-for-beginners/ (Accessed: January 24, 2023).

Schwartz, B., 2022. *Welcome BERT: Google's latest search algorithm to better understand natural language*. [Online] Search Engine Land. Available at: <https://searchengineland.com/welcome-bert-google-artificial-intelligence-for-understanding-search-queries-323976> [Accessed 22 July 2022].

Sharma, S. (2021) *Sentiment analysis using the SentiWordNet Lexicon*, *Medium*. Medium. Available at: https://srish6.medium.com/sentiment-analysis-using-the-sentiwordnet-lexicon-1a3d8d856a10 (Accessed: January 24, 2023).

Seminar, M. T. Meaning Based Multilingual Search Engine.

Stephanie (2021) *Kaiser-Meyer-Olkin (KMO) test for sampling adequacy*, *Statistics How To*. Available at: https://www.statisticshowto.com/kaiser-meyer-olkin/ (Accessed: January 26, 2023).

Transparency International Kenya. (2018). *A case study of public participation frameworks and processes in Kisumu county.*

Transparency International Kenya. (2018). *Editor Public Participation Processes in Kisumu.*

Uraia. (2017). *PUBLIC PARTICIPATION IN KENYA; What are the milestones, challenges and prospects*. Uraia Trust.

Uraia (2022) *Kenya's National Civic Education Programme, Uraia*. Available at: https://uraia.or.ke/?r3d=public-participation (Accessed: January 22, 2023).

USC Libraries (2023) *Research guides: Organizing your Social Sciences Research Paper: Types of research designs*, *Types of Research Designs - Organizing Your Social Sciences Research Paper - Research Guides at University of Southern California*. Available at: https://libguides.usc.edu/writingguide/researchdesigns (Accessed: January 30, 2023).

Wacera, D. C. (2016). *Effect of Citizen Participation on budget implementation in Kenyan Counties: A case study of Nyandarua County.* Nairobi: University of Nairobi.

Wang, C., Tsai, M., Liu, T., & Chang, C. (2013). Financial Sentiment Analysis for Risk Prediction. *Sixth International Joint Conference on Natural Language Processing*, (pp. 802-808).

Zad, S., Jimenez, J., & Finlayson, M. (2021, August). Hell hath no fury? Correcting bias in the nrc emotion lexicon. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)* (pp. 102-113).

APPENDICES

Appendix I: Source Codes

Code 1: Log In Code

```
import 'dart:convert';
import 'package:http/http.dart' as http;
import 'package:jumuika_app/models/auth.dart';
import 'package:jumuika_app/models/error.dart';
import 'package:jumuika_app/network/APIs.dart';
class LoginController {
  // login code
  static authenticate(String email, String password) async {
    final http.Response response = await http.post(
Uri.parse(API.LOGIN),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
      },
      body: jsonEncode(<String, String>{
        "email": email,
        "password": password,
      }),
    );

    if (response.statusCode<= 500) {
      return Auth.fromJson(
json.decode(response.body),
      );
    } else {
      return Error(
        code: 600,
        status: "We are having troubles with login. Try again later.",
      );
    }
```

```
  }
}
```

Code II: Register Code

```
import 'dart:convert';
import 'package:flutter/foundation.dart';
import 'package:http/http.dart' as http;
import 'package:jumuika_app/models/auth.dart';
import 'package:jumuika_app/models/county.dart';
import 'package:jumuika_app/models/error.dart';
import 'package:jumuika_app/models/sub_county.dart';
import 'package:jumuika_app/models/ward.dart';
import 'package:jumuika_app/network/APIs.dart';
import 'package:shared_preferences/shared_preferences.dart';
class RegisterController {
  // login code
  static signUp(String firstName, String lastName, String phone, String email,
      String county, String subCounty, String ward, String password) async {
    final http.Response response = await http.post(
Uri.parse(API.REGISTER),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
      },
      body: jsonEncode(<String, String>{
        "first_name": firstName,
        "last_name": lastName,
        "phone": phone,
        "email": email,
        "county": county,
        "sub_county": subCounty,
        "ward": ward,
        "password": password,
      }),
    );
```

```
    if (response.statusCode<= 500) {
      return Auth.fromJson(
json.decode(response.body),
      );
    } else {
      return Error(
        code: 600,
        status: "We are having troubles with sing up. Try again later.",
      );
    }
  }
  static Future<List<County>>getCounties(http.Client client) async {
    final response = await client.get(
Uri.parse(API.COUNTY),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
      },
    );
    // Use the compute function to run parseSliders in a separate isolate.
    return compute(_parseCounties, response.body);
  }
  // A function that converts a response body into a List<County>.
  static List<County> _parseCounties(String responseBody) {
    final parsed = jsonDecode(responseBody).cast<Map<String, dynamic>>();
    return parsed.map<County>((json) =>County.fromJson(json)).toList();
  }
  static Future<List<SubCounty>>getSubCounties(http.Client client, String countyId)
async {
    final response = await client.get(
Uri.parse(API.SUB_COUNTY + '/' + countyId),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
      },
    );
    // Use the compute function to run parseSliders in a separate isolate.
```

```
    return compute(_parseSubCounties, response.body);
  }
  // A function that converts a response body into a List<SubCounty>.
  static List<SubCounty> _parseSubCounties(String responseBody) {
    final parsed = jsonDecode(responseBody).cast<Map<String, dynamic>>();
    return parsed.map<SubCounty>((json) =>SubCounty.fromJson(json)).toList();
  }
  static Future<List<Ward>>getWards(http.Client client, String subCountyId) async {
    final response = await client.get(
Uri.parse(API.WARD + '/' + subCountyId),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
      },
    );
    // Use the compute function to run parseSliders in a separate isolate.
    return compute(_parseWards, response.body);
  }
  // A function that converts a response body into a List<Ward>.
  static List<Ward> _parseWards(String responseBody) {
    final parsed = jsonDecode(responseBody).cast<Map<String, dynamic>>();
    return parsed.map<Ward>((json) =>Ward.fromJson(json)).toList();
  }
}
```

Code III: Comments Code

```
import 'dart:convert';
import 'package:http/http.dart' as http;
import 'package:jumuika_app/models/comment.dart';
import 'package:jumuika_app/models/error.dart';
import 'package:jumuika_app/network/APIs.dart';
import 'package:shared_preferences/shared_preferences.dart';
class CommentController {
  static post(String comment, int participationId) async {
SharedPreferencesprefs = await SharedPreferences.getInstance();
```

```
    int userId = prefs.getInt('id') ?? 0;
    final http.Response response = await http.post(
Uri.parse(API.COMMENT),
      headers: <String, String>{
        'Content-Type': 'application/json; charset=UTF-8',
        'Authorization': 'Bearer ' + prefs.getString('token'),
      },
      body: jsonEncode(<String, String>{
        "comment": comment,
        "participation_id": participationId.toString(),
        "user_id": userId.toString(),
      }),
    );
    print(response.body);
    if (response.statusCode<= 500) {
      return Comment.fromJson(json.decode(response.body));
    } else {
      return Error(
        code: 600,
        status: "We are having troubles with posting comments. Try again later.",
      );
    }
  }
}
```

Code IV: Participation Code

```
class ParticipationModel {
  final int id;
  final int forumId;
  final String image;
  final String title;
  final String description;
  final String date;
```

```
ParticipationModel({this.id,this.forumId ,this.image, this.title, this.description,
this.date});
  factory ParticipationModel.fromJson(Map<String, dynamic> json) {
    return ParticipationModel(
      id: json['id'] as int,
forumId: json['forum_id'] as int,
      image: json['image'] as String,
      title: json['title'] as String,
      description: json['description'] as String,
      date: json['created_at'] as String,
    );
  }
}
```

Appendix II: Questionnaire for Respondents

Section A: Background Information- tick (√) where appropriate

1. Gender Male [ ]
   Female [ ]
2. Age Bracket 25 years and below [ ]
   26 - 35 years [ ]
   36 - 45 years [ ]
   46 - 55 years [ ]
   Over 55 years [ ]
3. Highest Level of Education
   Certificate [ ]
   Diploma [ ]
   Degree [ ]
   Masters [ ]
   Doctorate [ ]
4. Occupation …………………………………………………………………………
5. How many times have you participated in public participation forums in the county since devolution began? ....

***Section B:* Challenges faced in obtaining sentiments in public participation forums**

Using the scale below, please indicate your level of agreement to the following propositions on cchallenges faced in obtaining sentiments in public participation forums.

5 – Strongly Agree; 4 - Agree; 3 - Neutral; 2 – Disagree; 1 – Strongly Disagree

		5	4	3	2	1
6.	We usually have limited time for everyone to fully contribute in the Public Participation discussions					
7.	Often few people get the chance to express their views in the PP					
8.	We are often unable to exhaust all the items in the PP forums					
9.	We are not able to capture each participants reactions adequately					
10.	We have challenges capturing the sentiments expressed by the participants in full					
11.	Often we have difficulty in finding the right words to express our feelings towards a subject					
12.	We have challenges analyzing the sentiments of the participants in the PP					
13.	We would prefer the discussions on a subject begin online before the PP so that only the critical issues can be discussed in the PP sittings					
14.	We would prefer the discussions on a subject continue online after the PP so that we can exhaust subjects being discussed					
15.	Online discussions will enable everyone to have time to have time to adequately air their views on a subject and other members react to them					
16.	Online discussions will enable the participants to be very honest in their views					
17.	Through online discussions, we will be able to access adequate information of the discussion material					

18. Of the following emojis (characters), which ones will make you very happy, happy, sad, very sad, agree, strongly agree, disagree, strongly disagree, laugh, wonder, angry (Please rate them by writing their numbers against the emojis).

Very Happy – 1 Happy - 2 Sad -3 Very Sad - 4 Agree - 5 Strongly Agree – 6 Disagree - 7 Strongly Disagree - 8 Laugh - 9 Wonder - 10 Angry - 11

Public Participation

Using the scale below, please indicate the extent of public participation of your engagement as provided by county government.

(5= To a very great extent; 4= To a great extent; 3= To a moderate extent; 2= To some extent;1=To a small extent).

No	Statement	5	4	3	2	1
19	Incorporate a wide range of public values					
20	Be available to all public interests					
21	Allow for new participants over time					
22	Protect participants’ identities when necessary					

Appendix III: Research Authorization Letter

KABARAK UNIVERSITY

BOARD OF POST GRADUATE STUDIES

Private Bag - 20157
KABARAK, KENYA
http://kabarak.ac.ke/institute-postgraduate-studies/

Tel: 0773265999
E-mail: directorpostgraduate@kabarak.ac.ke

15th July, 2020

The Director General
National Commission for Science, Technology & Innovation (NACOSTI)
P.O. Box 30623 – 00100
NAIROBI

Dear Sir/Madam,

RE: MALACHI OMELA MANASES – GMI/N/2517/05/18

The above named is a candidate at Kabarak University pursuing Master's degree in Information Technology. He is carrying out a research entitled "***Sentiment Analysis Model for Online Public Participation Forums***". He has defended his proposal and has been authorised to proceed with field research.

The information obtained in the course of this research will be used for academic purposes only and will be treated with utmost confidentiality.

Please provide the student with a research permit to enable him to undertake the research.

Thank you.

Yours faithfully,

Dr. Wilson O. Shitandi
DIRECTOR, INSTITUTE OF POST GRADUATE STUDIES

KABARAK UNIVERSITY
DIRECTOR
POST GRADUATE STUDIES
15 JUL 2020
PRIVATE BAG 20157, KABARAK

Kabarak University Moral Code
As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord. (1 Peter 3:15)

Kabarak University is ISO 9001:2015 Certified

Appendix IV: NACOSTI Research Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **743248**

Date of Issue: **21/July/2020**

RESEARCH LICENSE

This is to Certify that Mr.. Malachi Omela Manases of Kabarak University, has been licensed to conduct research in Baringo, Busia, Nakuru on the topic: SENTIMENT ANALYSIS MODEL FOR ONLINE PUBLIC PARTICIPATION FORUMS for the period ending : 21/July/2021.

License No: **NACOSTI/P/20/5895**

743248

Applicant Identification Number

Director General

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix V: Evidence of Conferences

KABARAK UNIVERSITY

Certificate of Participation

Awarded to

Malachi Omela Manases

for successfully participating in the Kabarak University International Research Conference on Computing and information systems 2021 on 4th and 5th October 2021 and presented a paper entitled ***"Sentiment Analysis Model for Online Public Participation Forums."***

Conference Theme

Development, Protection And Commercialization Of Intellectual Property

Dr. Peter Rugiri
Dean School of Science, Engineering and Technology

Dr. Miriam Muga
Ag. Director Research, Innovation and Outreach

Kabarak University Moral Code

As members of Kabarak University family, we purpose at all times and in all places, to set apart in one's heart, Jesus as Lord.

(1 Peter 3:15)

Kabarak University is ISO 9001:2015 Certified

Appendix VI: List of Publication

East African Scholars Journal of Engineering and Computer Sciences

OPEN ACCESS

Abbreviated Key Title: East African Scholars J Eng Comput Sci

ISSN: 2617-4480 (Print) & ISSN: 2663-0346 (Online)

Published By East African Scholars Publisher, Kenya

Volume-5 | Issue-5 | Sept-2022 |

DOI: 10.36349/easjecs.2022.v05i05.00X

Original Research Article

Sentiment Analysis Model for Public Participation Forums in County Governments

Malachi Omela Manases[1*], Moses Thiga[2], Nelson Masese[2]

[1]Master Student, Department of Information Technology, Kabarak University, Kenya

[2]Senior Lecturer, Department of Information Technology, Kabarak University, Kenya

Article History

Received: 19.08.2022

Accepted: 24.09.2022

Published: N/A

Journal homepage:

https://www.easpublisher.com

Quick Response Code

Abstract: Public participation is important because it helps to close the gap between the public, private sector and the government. However, a successful devolution process in Kenya is hampered by a lack of/inadequate public participation in county governments. Communications gaps are one of the arguments made for this development. The main objective of the study was to develop a sentiment analysis model for use in public participation forums in County Governments in Kenya. The study was conducted through the design thinking process. The population of interest of this study comprised of county management and staff also area residents in Nakuru, Busia and Baringo counties who have participated in public participation forums before. The Bidirectional Encoder Representations from Transformers (BERT) approach was used to create the cloud NLP package and obtain user sentiment magnitudes for the sentiment analysis model. Following that, cross validation was utilized to assess the performance indicators during the design stage, and users took part in the model's assessment. The overall conclusion of validation is that the model performed as expected and recorded instrumental results in increasing effective public participation in county governments in Kenya and strengthen the devolution process. This study recommends that the model can be cascaded to all the counties in Kenya to improve the efficiency of public participation.

Keywords: Public participation, (BERT), Sentiment Analysis Model.

INTRODUCTION

Technology is viewed as a solution to social issues. Public participation is not an exception. Every county government in Kenya has an official website which is often used to access information. The requisite documents for public participation are posted on the website (Transparency International Kenya, 2018). Effective participation needs transparency (Daudi, 2016). Transparency in public' actions and transparency in leadership and administration. Openness is affected through access to information. Inadequate access leads to difficulty in interpreting the policies, services and programs. Public apathy is the indifference, lack of concern in development. When there is apathy among the public means that there are disinterested leading to them withdrawing from participation (Obora, 2016).

Sentiment Analysis

Sentiment analysis (SA) which is also referred to as emotion AI or opinion mining can be defined as the process of automating mining of opinions, views, attitudes, emotions and phrases through Natural Language Processing (Beigi, Hu, Maciejewski & Liu, 2016). It is the application of text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

Sentiment analysis is widely applied to the voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Sentiment Analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

***Corresponding Author:** Malachi Omela Manases

Master Student, Department of Information Technology, Kabarak University, Kenya