Proceedings of the Kabarak University International Conference on Computing and Information Systems, October 2021 Nakuru, Kenya.

Sentiment Analysis Model for Online Public Participation Forums

Malachi MANASES[1], Moses THIGA[2], Nelson MASESE[3]
[1]*Kabarak University, P.O. Box Private Bag, Kabarak, 20157, Kenya*
Tel: +254 0721 413 746, Email: mmanasseh@kabarak.ac.ke
[2]*Kabarak University, P.O. Box Private Bag, Kabarak, 20157, Kenya*
Tel: +254 0722 555 999, Email: mthiga@kabarak.ac.ke
[3]*Kabarak University, P.O. Box Private Bag, Kabarak, 20157, Kenya*
Tel: +254 0737070029, Email: nmasese@kabarak.ac.ke

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 an 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 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 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. A sample size of 106 respondents comprising 23 county administrators and 83 residents were purposively sampled for the project. The findings indicate that there exists a statistically significant difference in public participation amongst the three counties (Baringo, Busia and Nakuru) at the 0.05 alpha level, $F(2, 500) = 100.296$, $p < 0.05$. The results of regression analysis revealed that human-based factors significantly influence public participation ($\beta=0.520$; $p<0.05$) while technological factors affect public participation significantly ($\beta=0.449$; $p<0.05$). These findings were incorporated in the model design.

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

1. Introduction

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. 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 (LawQuery, 2017).

The efforts to decentralize and facilitate public participation have however not been with its' own challenges if we consider the initial attempts at decentralized 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 decentralization 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 (InterGovernmental Relations Technical Committee., 2016).

1.1 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 (Kharde & Sonaware, 2016).

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).

According to Bannister (2018), 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.

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.

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 numerous commissioned and academic studies suggest that public participation in counties as specified by the constitution is poor owing to issues such as communications, logistics and security (Ngugi, & Oduor, 2015).

The present study, therefore, proposes 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 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 as 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

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

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.

2. Literature Review

2.1 Challenges faced in obtaining sentiments in public participation forums.

Despite the constitutional provisions, there have been outcries in some counties over certain decisions undertaken by the county government (Kaseya &Kihonge, 2016). An exceptionally low percentage of Kenyans partake in the governance of their counties. In a report by 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 (Daudi, 2016). 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. But civic education is also not done fully in most counties. In a study by (Kaseya &Kihonge, 2016), 68.5% of the respondents had not been sensitised. Civil society organisations were the biggest sensitizer voted at 29.4%, the county government following closely at 28.5% and the national government at 9.4%.

According to (Ronoh, Mulongo, &Kurgat, 2018), challenges involving public participation include public participation is seen as a time-consuming process. Scheduling of public forums influences public participation (Kaseya & Kihonge, 2016). 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 (InterGovernmental Relations Technical Committee., 2016).

Absence of a structured feedback mechanism from the county officials who are generally perceived as non-receptive to such efforts is a hindrance to effective public participation (Hakijamii, 2017). Lack of feedback mechanisms from earlier hearings held discourages participation (TISA, 2015). The public need to know whether their inputs were received, and whether and why they were or were not incorporated into the relevant plans or budgets (Hakijamii, 2017).

2.2 Sentiment analysis models for public participation forums

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 (Beigi et al., 2016). 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).

SentiWordNet (Esuli & Sebastiani, 2006) is a tool widely used in opinion mining, based on an English lexical dictionary called WordNet.

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.

2.3 Conceptual Framework

The conceptual framework for the study is presented in the following figure.

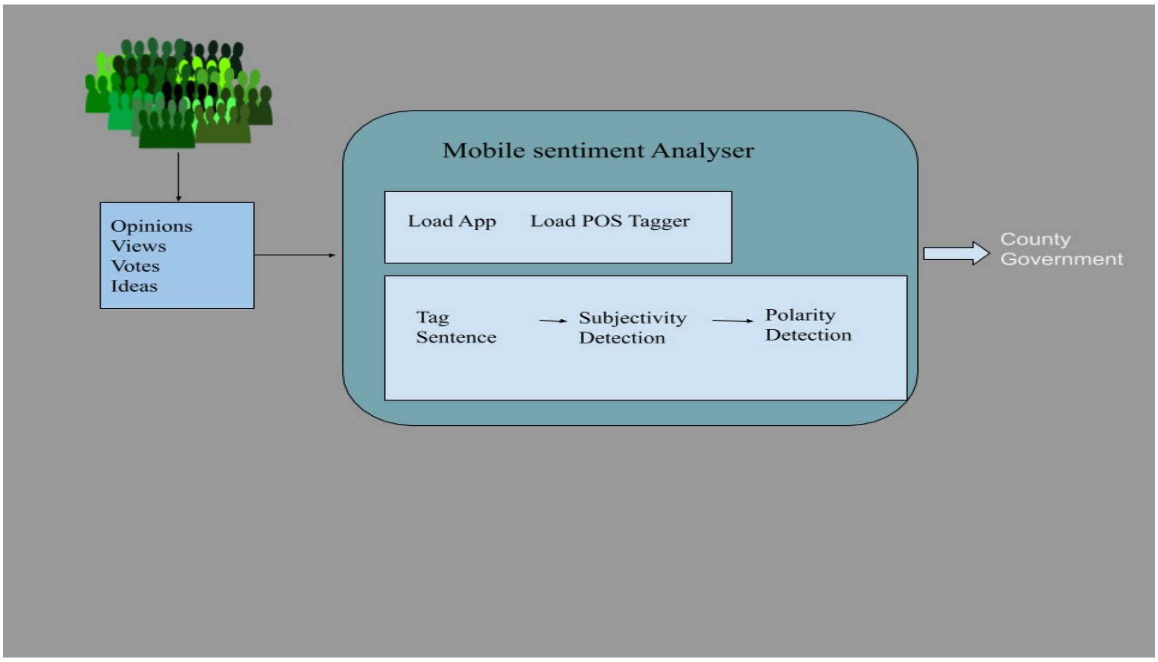

Figure 1: Conceptual Framework

The conceptual framework has the following concepts:

1. **Users:** These are the parties who are going to utilize the online participation forums. The participants of the system include:

i. **Public-** These are the individuals giving their opinions, views or ideas about a particular subject.
ii. **System Administrator –** This is the entity that is in charge of the system, updating information and ensuring the system is running efficiently.
iii. **County Government** – It is an entity that will utilise the system by checking on the analysis results to gauge the citizen's response.

2. **Mobile Sentiment Analyser –** This will incorporate 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 will include the following components:
   i. **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.
   ii. **Subjectivity Detection –**This involves using the POS tags to identify opinion lexicons in the sentence, whether the sentence is subjective or objective.
   iii. **Polarity Detection –**This stage also utilizes the POS tags to indicate positive or negative expressions.

2.4 Model Workflow

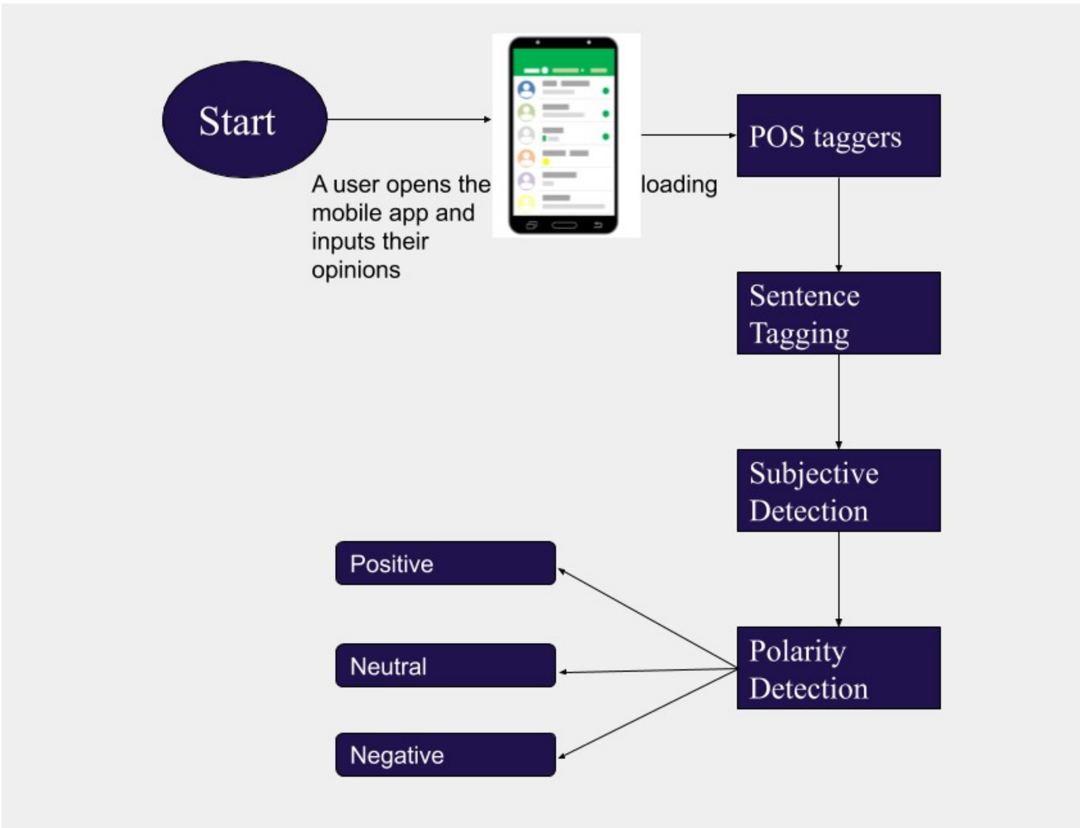

Figure 2: Model Workflow

3. Methodology

3.1 Research Design

The study was conducted 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.

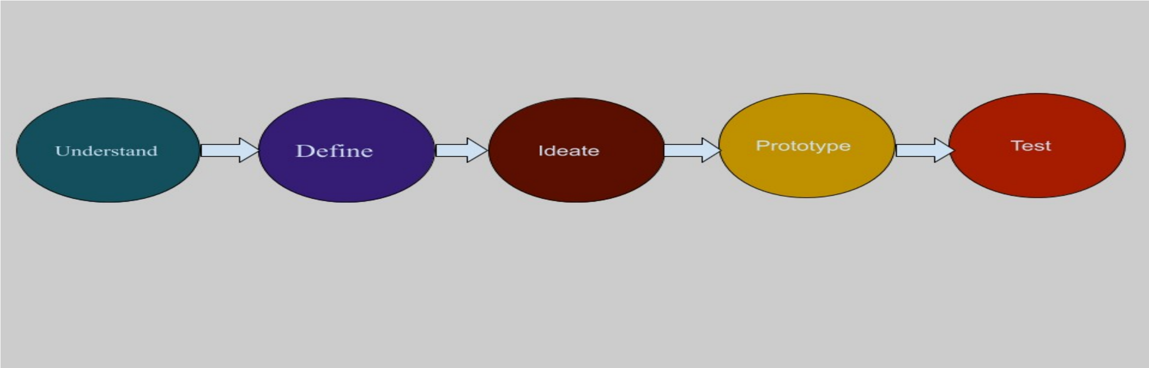

Figure 3: Design Thinking Process Approach

3.2 Target Population

The population of interest of this study comprised of county management and area residents in the three counties who have participated in public participation forums before. Data from the three counties Nakuru, Busia and Baringo indicate that cumulatively 491 residents and 23 county administrators participated in the last public participation forums in 2018 as shown in Table 2.

Table 2: 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.

3.3 Sampling techniques and sample size

Taherdoost (2016) define sampling as a 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) will 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%).

Thus, a sample size of 83 residents obtains 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.

4. Results

4.1 The County Category

Respondents were sampled from three different counties, namely Baringo, Busia and Nakuru. The results of the analysis are presented in Table 3.

Table 3: 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 analysed 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 Descriptive Analysis

Descriptive analysis was done using percentages to describe the characteristics of the population regarding the variables being investigated. The key variables under investigation was challenges faced in obtaining Sentiments indicators as well as the extent of public participation indicators.

Table 4: Challenges faced in obtaining Sentiments in Public Participation Forums

Statements	SD (%)	D (%)	N (%)	A (%)	SA (%)
We usually have limited time for everyone to in the PP to participate fully in the 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 4, 66.2% of respondents affirmed that they usually have limited time for everyone to in the PP to participate fully in the 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 Society for International Development (2016) 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 participants 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 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 Inferential Analysis

In this section, ANOVA, Post -hoc test, Factor analysis, correlation and regression analysis were utilized. The results of the analysis were presented in subsequent tables.

4.3.1 Testing difference in Public Participation between Counties.

In testing the difference in population means on the extent of public participation, one-way ANOVA was employed. The findings were presented in table 5

Table 5: ANOVA

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Between Groups	196.615	2	98.308	100.296	.000
Within Groups	490.085	500	.980		
Total	686.701	502			

Dependent variable: public participation

The finding showed that there exists a statistically significant difference in public participation amongst the three counties (Baringo, Busia and Nakuru) at the 0.05 alpha level, $F(2, 500) = 100.296$, $p < 0.05$.

4.3.2 Post- hoc test

Post hoc ("after this" in Latin) tests are used to uncover specific differences between three or more group means when an analysis of variance (ANOVA) F test is significant. A post hoc test is used only after it is found a statistically significant result and need to determine where our differences truly came from. The Tukey post hoc test is generally the preferred test for conducting post hoc tests on a one-way ANOVA (Allen, 2017).The findings are presented in table 6.

Table 6: Multiple Comparisons

Dependent Variable: Public Participation

Tukey HSD

(I) County	(J) County	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Baringo	Busia	-.25189	.11208	.064	-.5153	.0116
	Nakuru	1.13926*	.11383	.000	.8717	1.4068
Busia	Baringo	.25189	.11208	.064	-.0116	.5153
	Nakuru	1.39115*	.10273	.000	1.1496	1.6326
Nakuru	Baringo	-1.13926*	.11383	.000	-1.4068	-.8717
	Busia	-1.39115*	.10273	.000	-1.6326	-1.1496

*. *The mean difference is significant at the 0.05 level.*

The findings indicate evidence of a statistically significant difference in public participation between Baringo and Nakuru counties ($p<0.05$), as well as between Busia and Nakuru counties ($p<0.05$). However, there were no differences between the groups of respondents in Baringo and Busia Counties regarding public participation ($p=0.064$).

4.3.3 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 7.

Table 7: 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 (Dodge,2008). 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 8.

Table 8: Rotated Component Matrix[a]

	Rotated Component Matrix[a]	**Component**	
	Variables	1	2
1.	We usually have limited time for everyone to in the PP to participate fully in the 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		.89

	honest in their views		3
12.	Through online discussions, we will be able to access adequate information of the discussion material		.89 5
	Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.		
	a. Rotation converged in 3 iterations.		

The results in table 8 show 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.

4.3.4 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 9.

Table 9: 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 exists 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.

4.3.5 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 10.

Table 10: Model Summary

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.

4.3.6 ANOVA

The model significance was tested at 0.05 alpha as presented in table 11.

Table 11: ANOVA[a]

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.

Table 12: Coefficients[a]

Coefficients[a]						
Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
1	(Constant)	.039	.056		.699	.485
	Human -Based Factors	.520	.018	.574	29.266	.000
	Technological Factors	.449	.019	.462	23.553	.000
a. Dependent Variable: Public Participation						

The results shows that human-based factors affect 52% of the public participation while technological factors affects up to 44.9% of public participation. In other words, an increase in one unit in human factors increases public participation by 0.52 units. Similarly, a unit increase in technological factors increases public participation by 0.449 units.
The equation for the model that will be adopted is given below:

$$Y=C+\beta_1X_1+\beta_2X_2+\varepsilon$$

Where:

Y= Public participation
C-constant
β_1,β_2-Regression coefficients
X_1= human-based factors
X_2= technological factors
ε-Standard error of the estimate.
$Y=0.039+0.52X1+0.449X2+ 0.39$

5. System Implementation

The overall purpose of the study was to develop a sentiment analysis model as an online system that would allow participants to make comments about policies and projects to be undertaken by county governments. Access to the online system is based on roles of the users. System admins, on one hand, can set forums and participations, view participation comments, manages users and roles, and view sentiment analysis scores as shown in figure 5.1 below.

The system computes sentiment and magnitude scores by first taking into consideration the inputs from respondents in English, Swahili or sheng languages then translate them to English for analysis of sentiments. This is procedurally done, first, by 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, by detecting the language of the comment whether it is in 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 analyse the sentiment using language client class under NLP package. Finally, Bidirectional Encoder Representations from Transformers (BERT) algorithm returns the magnitude and score values from cloud service containing basic language model. Ultimately, the system computes the general feeling of the public about county projects.

The PHP code snippet presented below demonstrates how magnitude and sentiment score values are computed.

```php
$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'];
```

General Comments Analysis

Avg. Sentiment Score	0.15
Avg. Magnitude Score	1.65
General feeling:	Neutral 😐
Participation Verdict: Majority of the general population are neutral with this participation to proceed.	

Figure 4: Sentiment Analysis System

6. Conclusions

The study established that a large portion of the population would wish to give views on their feelings about county projects but lack time to do so, or when they get chance to do so they do not exhaust all the items in the public participation forums. Besides, it is a tall order for county administrators to interpret the general lingo used by the citizens and therefore it is equally difficult to obtain the sentiments and general feeling of the population about the projects. Jumuika platform therefore provides a more efficient way of involving more participants, interpreting common lingo (Swahili and Sheng) to English and computing magnitude and sentiment scores that indicates the general feeling of the population.

7. Recommendations and Areas for further study

This study therefore recommends application of online public participation platforms like Jumuika app to all counties in Kenya to enable the local and county governments to improve the levels of public participations. County governments should roll out such platforms to tap on the large population of smart mobile users as well as users of personal computers. Much sensitization therefore needs to be done to further improve personal participation and comply with the law and regulatory requirements.

References

Adams, C., & Nash, J. (2016). Exploring design thinking practices in evaluation. *Journal of MultiDisciplinary Evaluation*, *12*(26), 12-17. 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.

Bannister. K. (2018). *Understanding Sentiment Analysis: What It Is & Why It's Used*. Retrieved on 15th June 2020 from: https://www.brandwatch.com/blog/understanding-sentiment-analysis/.

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

Daudi, K. A. (2016). *Determinants of citizen participation in devolved governance in Kenya: A case study of Machakos County.* Nairobi: University of Nairobi.

Esuli, A., &Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *5th Conference on Language Resources and Evaluation*, *6*, pp. 417-422. Italy. doi:10.1.1.61.7217

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

InterGovernmental Relations Technical Committee. (2016). *Status of public participation in national and county governments.* Nairobi, Kenya.

Kaseya, C. N., &Kihonge, D. E. (2016). Factors Affecting the Effectiveness of Public Participation in County Governance in Kenya: A case of Nairobi County. *International Journal of Scientific Research and Publications, 6*(10), 476-487.

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.

LawQuery. (2017). Retrieved from LawQuery Kenya: https://www.lawquery.co.ke/areas-we-cover/elections-2017-public-participation/what-is-public-citizen-participation

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

Nassiuma D.K (2000). *Survey Sampling: Theory and Methods*. University of Nairobi Press, Nairobi.

Ngugi, R. W., & Oduor, C. (2015). Review of status of public participation, and county information dissemination frameworks: a case study of Isiolo Kisumu Makueni and Turkana Counties.

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.

Society for International Development (SID) (2016). *Voices from the Counties: Lessons from the 2015 citizen report card.* Nairobi.

TISA. (2015). *JIHUSISHE: Lesson in Participation in County Budget.* Nairobi: The Institute for Social Accountability.

Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. *How to Choose a Sampling Technique for Research (April 10, 2016)*.

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

Uraia Trust. (2016). *Citizen Participation Booklet.* Nairobi, Kenya.