TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

Classifying Philippine-Korean Speech Analysis based on various Machine Learning Heuristics.

Rolando B. Barrameda[1], Jobelle B. Baccay[2]
rbbarrameda@dlsud.edu.ph[1], jbbaccay@dlsud.edu.ph

Article Info
Volume 81
Page Number: 5027 - 5035
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *24 December 2019*

Abstract:
Every individual can be identified in their native language, and use them as communication expressing an opinion, sharing emotions and feelings through voices, speech and body language.Language is basic and fundamental to us.Numerous languages in terms of variants in the world today speak, a diverse manner of speech technique and method are common practices and a natural way of speaking.This paper presents a classification of Filipino and Korean speech assessment based on various machine learning.Data are learned using several algorithms including NaivesBayes, K-Means, Support Vector Machine and Multilayer Perceptron. The data collected from recorded voices tests and compares with those classifiers. There are 100 respondents and collect recorded audio. A dataset is split into half 50 came from Filipino and 50 came from Korean. The assessment and evaluation are measured based on its accuracy and correctness.Based on the results shown in the Multilayer Perceptron (MLP) figures, the highest and best performance was 99.49 percent, followed by the Support Vector Machine (SVM) with 98.6 percent, and the Naïve Bayes with 97 percent. With 94.78 percent, 74% and K-means fall to the lowest position.

Introduction

Everylanguage served as a mirror of each and every race, different country, may identify the diverse of voices and can be distinguished by listening and talking through, chat, dialogue, discussion and conversation.Technological changes happen very fast hence, linearly proportion are observed to extreme raising of innovation technology to developed and improve the emerging urges of society. The growth improvement in the field speech recognition are relevance for the researchers. Various method and technique are being done in industry and academic to worked on speech translation and Machine translation.Interaction between human and machine communication translate with different invented modules, speech pronunciation difference is ubiquitous in every race.

These phenomena of related disparityare absorbed by a certain phoneme like tonal and non-tonal languages for instance, Tagalog and Korean language both are classified as a non-tonal language [11]. According to AsiaSociety.org, Seoul languages in South Korea and Phyong'yang languages in North Korea are the two standard varieties in practice,These two dialects are prominent by each language policy. The modern korean writing system is called han'gul it consists of twenty-four (24) letters, fourteen (14) consonants and Ten (10) vowels. The pattern and groups of these letters characterize by 5 double consonants and 11 dipthongs. Letters are grouped in every cluster of 2, 3, and 4 syllables and words [6]. The form and meaning of root words is not affected with respect of the tone speech. However, Korean language has some minimal changes in accent pitch which evenly stress phrases and sentences. With regards to conversational and reading nuances is upward at the end of the sentences. It took time and effort

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

to learn the Korean language by heart. In fact, linguistic claim han'gul ranks among the world's three hardest language to master.

Similarly, Filipino dialect is non-tonal language in the sense that it does not have pitch or tone level that changes the meaning of a word if the chunk of its syllable is in the low, middle and high tone. Tagalog language has different stresses that has same spelling, but it varies the meaning like (baba vs. babâ / chin vs going down) Nevertheless, regardless of how high or low the voice pitch as long as substance and the stress on syllables is acceptable, it will stay the same in meaning. Tagalog has a verbal complexity system demonstrating morphological phenomena like stress changing, interchange of consonant, and repetition for formative parts of speech and voice which involves the customin grammar and word. Tagalog transform into a more complex language whichconstructsfewerprocedure of pointers and morphemes for regulating the rules of parts of speech and give emphasis as it does syntactic correct in manner [8]. Likewise, added study claim that Tagalog persuaded by other languages like English, Spanish and Indonesian with regards to phonological and lexical features. Although, morphosyntatic feature preserve completely Tagalog [9].

The study uses the power of WEKA software to classify datasets using recorded voice. The data set is evaluated with five classification algorithms, including Naïve-Bayes, K-Means, Support Vector Machine (SVM), KNN and Neural Network (MLP).

Review of Related Literature

In the field of Artificial Intelligence (AI), machine learning became a dominant knowledge expert to understand human-computer interaction specifically Natural Language Process (NLP). Researchers have more interest to explore and discover the deep knowledge about these machine learning algorithms where they can able classify and construct knowledge to accomplish tasks such as automatic speech recognition. Utilizing the NLP give more understanding to analyze language use for converting speech and automatically translating between languages. In the blog presented by google, released an open source neural network framework implementing a TensorFlow that provide a foundation of Natural Language Understanding (NLU).

Recently,the team of researchers from Microsoft has created the first machine translation that can translate news articles from Chinese to English with the same quality and accuracy as a person[5]. Google use ParseyMcParseFace algorithm which learn the linguistic structure of language in a manner of accurate model for automatic extraction of information, translation and other application [2].

Language is a vital and most essential way to communicate people. Identity and race can easily classify through communication and the tone of its languages. In the paper presented by Marimuthu et.al. (2014), entitled "A Study on Speech Recognition Today", discussed about system recognition to word spoken by extracting, characterizing and recognizing signal. The prime strength of these study is that speech produce signal is natural and obstructive. Many study and technique are used for speech recognition [3].Other study used three feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC) 2) Formant Estimation Coefficients (FEC) 3) Linear Predictive Codes (LPC)[4]. The development of efficient speech recognition using HMM, MFCC and Distance minimum algorithm. This technique worked to performed efficient and less complexity in term of running time and memory [7].

Innumerable researches evaluate, assess and scrutinize speech recognition appliedvarious algorithm inmultiple language features.Further tried to solve and utilize a novel DNN architecture which replace standard DSR pipeline modules to trained variation standard of back-propagation algorithm being used speech recognition and speech enhancement gradients through the network of DNN(Ravanelliet. al, 2017).

Additional study discussed toprobeseveralstructures for finding of various language using deep learning neural networkintegrates tonal and non-tonal language. The dataset is trained for Cantonese, Tagalog and Vietnamese and tested several hours using IARPA Babel. The table1 present the tested setup of 2 non-tonal language and 2 tonal language, calculating tone feature resulting a

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

small gain language. The integration of Tagalog using DBNF lessen the percentage of WER by 1.8% fit to the boundary DBNF. Even for the non-Babel English system, a minorincrease of 0.5% from 16.0% to 15.5% could be gained with this approach. In parallel to the tonal languages, remodify the system has no effect and improvementintonal features. However, the system urges further improvement of WER in terms of performance [11].

The paper conducted by hemlata (2018) investigates and examines the available algorithms of Waikato Environment for Knowledge Analysis (WEKA) in terms of correctness and accurateness. The paper gives a complete assessment of different classifiers which can be used to determine paramount and suited classifier while using this software. These open source tool used specifically for data mining and knowledge extraction huge amount of data [13].

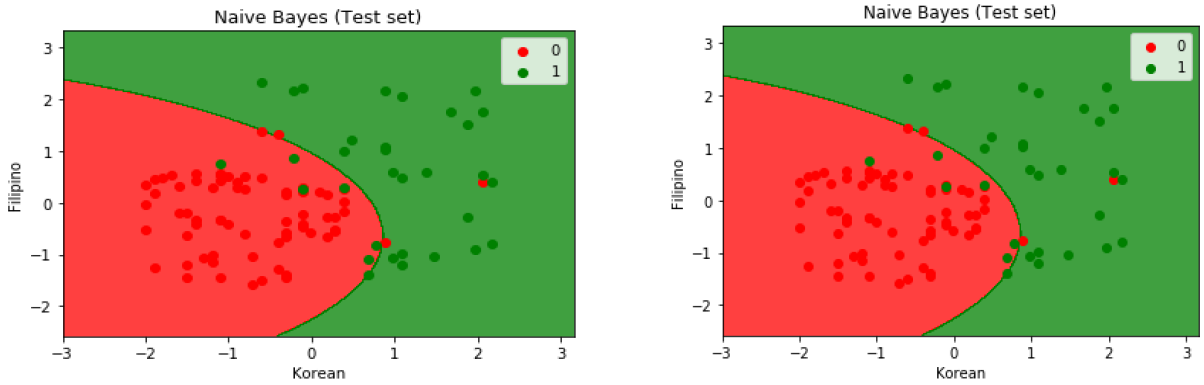
Methodology

This paper assesses and classify various algorithm using machine learning classification specifically Naïve Bayes, K-Means, SVM and MLP

Naïve-Bayes

Naïve Bayes classifier is based from base theorem. These class is for probabilistic type of class Given the plot observation for the dataset. What is the probability that the persons talk is Korea or Filipino. New data point is like supervised learning with certain features of observation from tagalog and korean. Bayes theorem formula:

- P(A|B) = P (B|A) * P(A)
  P (B)



(a) Plot of datasets (b) Plot of dataset using Bayes theorem

Figure 5. Naïve Bayes Plot

Figures show the feature class of a data and the applied classifier, based on the probability of speaker it turns out resembling to the Filipino speakers, after the processes the classifier attest the possibility.

Naives Bayes applied rules:

Step 1: P (Filipino | X) = P(X | Filipino) * P (F)
P (X)

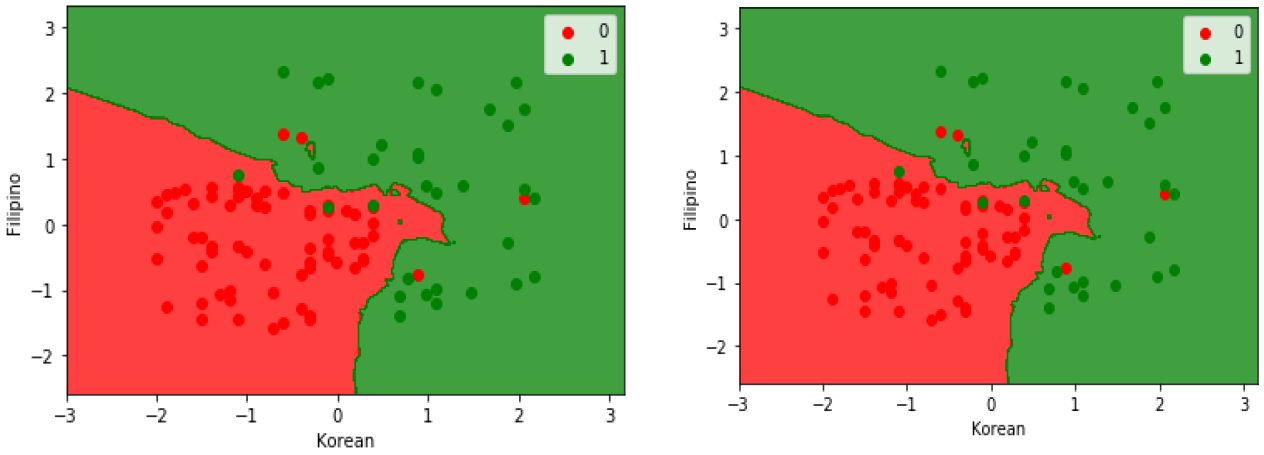
1. Calculate the prior probability using the given the feature.
   P(Filipino) = Number of Filipino Speech
   Total Observation
2. Calculate Marginal likelihood
   P(X) = Number of Similar Observation
   Total Observations
3. Calculate likelihood.
   P (X | Filipino) = No. of similar Observation among those who talk Filipino
   Total number of samples
4. Get the Posterior Probability

Step 2. P(Korean | X) = P(X | Korean) * P (Korean)
P (Korean)
Step 3. Compare the feature of X and Y
P(Filipino | X) vs P (Korean| X)

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

K-Means

This clustering algorithm is a convenient tool for discovering categories of groups in the dataset. The plot has variables for X and Y axis, these how the observation is configured according these two variables. Identifying groups in this plot can create different numbers groups. What actually k-means does is takes out the complexity from this making process allows us to easily identify those clusters are actually called of data points in your dataset. These is very simplified exam to group the classes in a two-dimension X and Y but in can be work with complex calculation multiple object to it. So, algorithm designed to find these cluster in the dataset.



(a) Plot of datasets

(b) Plot of dataset using Bayes theorem

Figure 6.K-Means Scatter plot

The scatter plot in this figure are generated upon processing using python language. The dataset are plots the number of samples (a) and cleans the data sample (b) utilizing the clustering data.

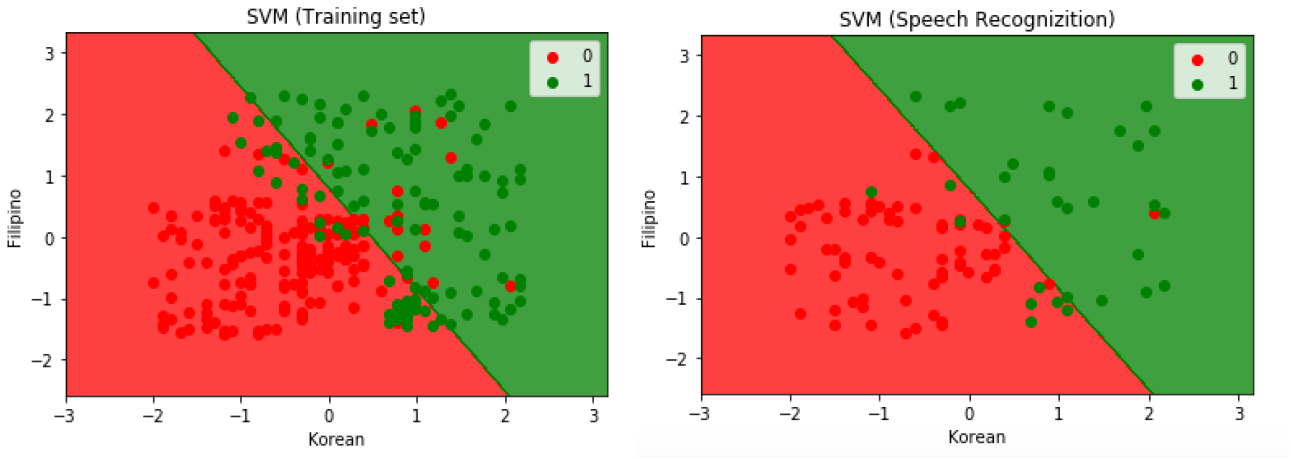
K-Mean applied rules:

1. By selecting the number of K-Clusters and randomly selecting the K pointsthat are the centroid, select any random points in the scatter plot and then select a certain number of centroids that are going to equate the number of clusters you decided upon.
2. Assign each data points to the closest centroid that forms k clusters. Euclidean can be used to measure distance for instance.
3. Recalculate the centroid and compute the new position of centroid of each cluster.
4. Reassign each data points to the new closest centroid. If a reassignment took place proceed to step 4 otherwise Model is ready

Support Vector Machine (SVM)

SVM got some observation that was already classified, the challenge here is how to separate points them in a line. The decision boundary is going to be important, so create a boundary between these two find optimal one. Separator line can help us to separate to classes. Line is search through maximum number. These are vectors.

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035



(a) Plot of the dataset (b) Plot of dataset using SVM

Figure 7. SVM Scatter Plot

The scatter plot in this figure are generatedusing python language for data visualization. Plot (a) are trained using recorded voices, plot (b) visualizing the test set result using SVM.

MultiLayer Perceptron (MLP)
This neural network is a class of feedforward, backpropagation techniques classify as supervised learning. Figure show the single layer perceptron.

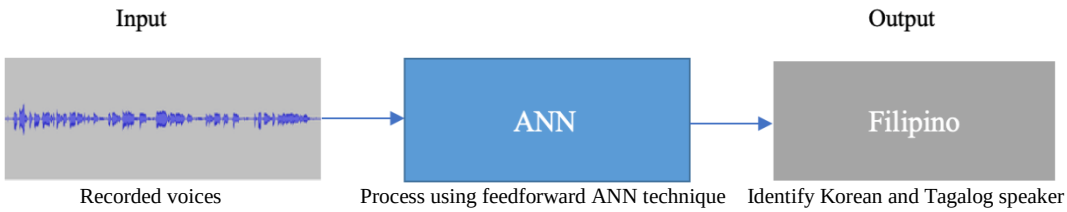


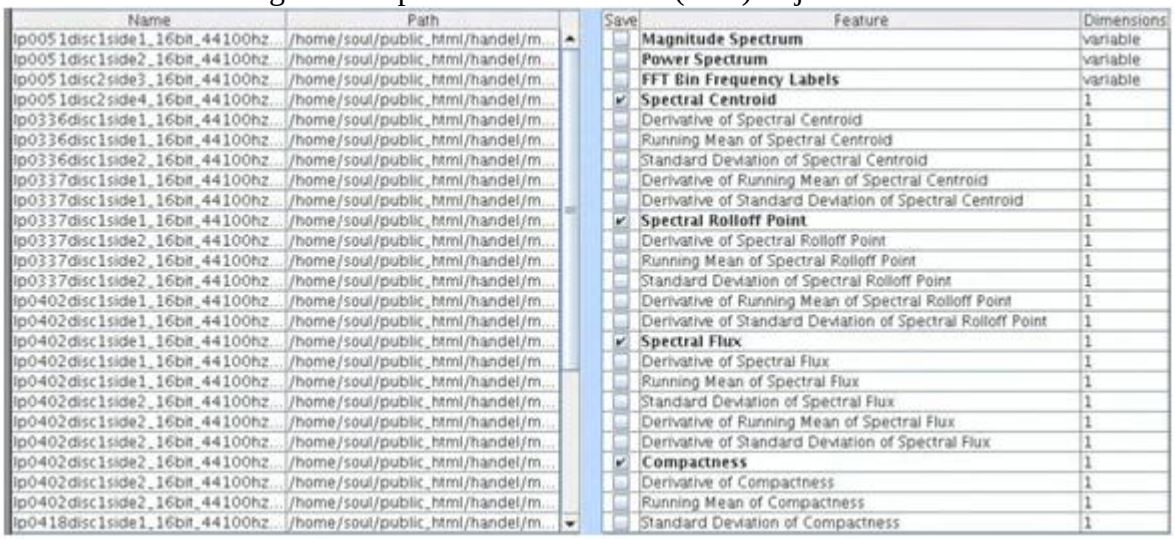
Figure 8. MLP Plot

The dataset is trained using MLP, the raw data is process through rectifier activation with some number of epoch iteration. By predicting the test result given the condition of threshold where the above threshold became 1 and below the threshold became 0, with the help of parameter tuning using techniques like k-fold cross validation where we can get better accuracy result.

Result and Discussion
This study utilizes some software tools to clean, extract and classify the dataset features with the support of Audacity, jAudio, Format Factory and WEKA. First is Audacity,anopen source software being used to eliminate unnecessary noise of an audio files.Second, is open source format factory for converting wide range of format. Third, with the help of jAudio which applied some algorithm feature extraction and Lastly, WEKA used for data mining task in straight forward manner for classification, clustering, preparation of data and visualization. A collective recorded voice of Filipino-Koreanisbuilt and classify using weka.

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

Figure1.Graphical User Interface (GUI) of jAudio



The tool is a framework for feature extraction intended to eradicate the repetition of work in analyzing features from an audio signal. This mechanism occurs the requirements of MIR researchers by delivering a collection of evaluating algorithms that are meet for a wide-rangingcollection of MIR tasks. The tooloffers features with a minimal learning curve, which process of selecting needed feature. In addition, this tool provides anexceptional method of managing multidimensional features to avoididenticalresults. In this figure,the user selects the prefer features to be extracted to solve dependency problems. Batch processing can be saved either ACE XML or ARFF file result [12].

The empiricalstudy managed collect one hundred (100) respondent. Equal number of selected individuals, fifty (50) Korean's and fifty (50) Filipino's. The respondents are requested to read and record the paragraph that was written in the paper.

The importance of reading in English classrooms in Korea has remained strong for a number of reasons. Writing in English, however, is also emerging as an important skill in English in many Korean students studying in the Philippines. As writing takes on more weight in the testing system, the emphasis on instruction and assessment of foreign languages has changed.

This study reports on the perceived needs of Ko rean English learners regarding their learning of reading and writing, and how integrated readin g and writing instruction is provided.

The recorded data are collected. It varies the according to size, time, length ranging from 16-24 seconds it consists of 100 recorded data. Recorded files was collected and mark/tag T – Filipino and K for Korean. Afterwards, raw data was pre-processing using Audacity and extract the feature using jAudio then classified using Weka software.

Figure 3. Recorded data

Recorded raw data was using mobile phone to record to voices. Table 2 shows, records files that was extracted using jAudio, data produced with 36 features.

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

Table 2: Extracted features.

@relation jAudio
@ATTRIBUTE "Spectral Centroid0" NUMERIC
@ATTRIBUTE "Spectral Rolloff Point0" NUMERIC
@ATTRIBUTE "Spectral Flux0" NUMERIC
@ATTRIBUTE "Compactness0" NUMERIC
@ATTRIBUTE "Spectral Variability0" NUMERIC
@ATTRIBUTE "Root Mean Square0" NUMERIC
@ATTRIBUTE "Fraction Of Low Energy Windows0" NUMERIC
@ATTRIBUTE "Zero Crossings0" NUMERIC
@ATTRIBUTE "Strongest Beat0" NUMERIC
@ATTRIBUTE "Beat Sum0" NUMERIC
@ATTRIBUTE "Strength Of Strongest Beat0" NUMERIC
@ATTRIBUTE "LPC0" NUMERIC
@ATTRIBUTE "LPC1" NUMERIC
@ATTRIBUTE "LPC2" NUMERIC
@ATTRIBUTE "LPC3" NUMERIC
@ATTRIBUTE "LPC4" NUMERIC
@ATTRIBUTE "LPC5" NUMERIC
@ATTRIBUTE "LPC6" NUMERIC
@ATTRIBUTE "LPC7" NUMERIC
@ATTRIBUTE "LPC8" NUMERIC
@ATTRIBUTE "LPC9" NUMERIC
@ATTRIBUTE "Method of Moments0" NUMERIC
@ATTRIBUTE "Method of Moments1" NUMERIC
@ATTRIBUTE "Method of Moments2" NUMERIC
@ATTRIBUTE "Method of Moments3" NUMERIC
@ATTRIBUTE "Method of Moments4" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs0" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs1" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs2" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs3" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs4" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs5" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs6" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs7" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs8" NUMERIC
@ATTRIBUTE "Area Method of Moments of MFCCs9" NUMERIC

Recorded datawas process and clean using Audacity to eliminate the unwanted sound. Noticeably, noises and unwanted sound are still present. murmuring and

pulsatingechoes heard in the speech. These are the factors that can affect result of speech waveform leading to low classification of correctness and accuracyoutcome. Noise reduction algorithm uses to lessen and or completely remove the noise. Research article presented as comparative study between discrete-time and discrete-frequency Kalman filtering algorithms to optimize the fraction of noise reduction.The quality of the estimated speech signal in the output of each filter was evaluated using segmental signal to noise ratio. It turns out that Kalman filtering is more suited to lessen noise of speech signals [14].

Figure 4. Reduce noise.

Conclusion and Recommendation

Classification using four various algorithms were used in this study namely Naïve Bayes, K-Means, SVM and MLP. MultiLayer Perceptron (MLP) and Support Vector Machine (SVM) were used as a classifier performed and implement with WEKA. All figures below show each model performances using 10-fold cross-validation. Based on the results shown in the Multilayer Perceptron (MLP) figures, the highest and best performance was 99.49 percent, followed by the Support Vector Machine (SVM) with 98.6 percent, and the Naïve Bayes with 97 percent. With 94.78 percent, 74% and K-means fall to the lowest position.

Table 3. Result of accuracy and efficiency of the four Algorithms.

Classifier	Correct Classified Instances	Incorrect Classified Instances	Kappa Static	Mean absolute error	Root Mean squared error	Relative Absolute error	Root relative squared error
Naïve Bayes	97.7452%	2.2548 %	0.95	0.0226	0.1481	4.5188 %	29.6287%
K-Means	94.7482%	5.2518%	0.89	0.526	0.2292	10.523%	45.8435%
SVM	98.6125%	1.3875%	0.97	0.0119	0.1178	2.7767%	29.5656%
MLP	99.4935	0.5065%	0.98	0.0054	0.0629	1.0843%	12.5776%

The result of the classification is shown, it was surprisingly good in general, because majority provide a highest correctness and accuracy result. Though, in terms of running time and memory MLP and SVM are good one. All the Methods and technique are mark and perform very well. However, this study simple and straight forward because of its limitation. Future research may explore more and experiment it with a large number of data that could be a Big Data for more accurate and acceptable results. It would be possible to compare four or more race speeches. Test and performwidely effective algorithm such as deep learning, Natural language processing or may be emphatic computing. This study can also be a good foundation to a more and highly relevant research. One of which is to improve the speech and emotion of child with Autism.

References

[1]. F. Metze. 2013. MODELS OF TONE FOR TONAL AND NON-TONAL LANGUAGES. IEEE Workshop on Automatic Speech Recognition and Understanding. https://ieeexplore.ieee.org/document/6707740/references#references.

[2]. M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio. 2017. A Network of Deep Neural Network for distant speech recognition.

[3]. https://asiasociety.org/education/korean-language.

[4]. K.M. Adlaon. 2017, MAG-Tagalog: A Rule-Based Tagalog Morphological Analyzer and Generator. Proceedings of the 17th Philippine Computing Science Congress (PCSC 2017)

November-December 2019
ISSN: 0193-4120 Page No. 5027 - 5035

[5]. N. Hans. 2004. A Two-level Engine for Tagalog Morphology and a Structured XML Output for PC-Kimmo. Dissertations. 133. http://scholarsarchive.byu.edu/etd/133
[6]. https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity
[7]. S. Shankaranand, M.M. Sharma, N.A.S, Roopa K.S, K.V. Ramakrishnan2014. An Enhanced Speech Recognition System. International Journal of Recent Development in Engineering and Technology. Volume 2, Issue 3, March 2014. www.ijrdet.com (ISSN 2347-6435).
[8]. Arul.V.H,R. Marimuthu. 2014, A Study on Speech Recognition Technology. Journal of Computing Technologies Volume 3 Issue 7 pp. 2278 pp. 3814.
[9]. Manar M. Salah El.Dina ,Micheal N. Mikhaelb , Hala A. Mansour. 2018, The Improvement of Real-Time ASR by using Gender recognition. International Journal of Scientific & Engineering Research Volume 9, Issue 7, July-2018.
[10]. M. Ravanelli, P. Brakel, M. Omologo,Y.Bengio. 2017. A Network ofDeep Neural Network for distant speech recognition.
[11]. H. Hemlata.2018. Comprehensive Analysis of Data Mining Classifiers using WEKA
[12]. L. Silva, G. Carrijo, E. Vasconcelos, R. Campos, C. Goulart , R. Lemos. 2018. Comparative Study between the Discrete-Frequency Kalman Filtering and the Discrete-Time Kalman Filtering with Application in Noise Reduction in Speech Signals. Journal of Electrical and Computer Engineering Volume 2018. https://doi.org/10.1155/2018/1895190
[13]. F. Metze. 2013. MODELS OF TONE FOR TONAL AND NON-TONAL LANGUAGES. IEEE Workshop on Automatic Speech Recognition and Understanding. https://ieeexplore.ieee.org/document/6707740/references#references.1
[14]. D. McEnnis, C.M, I. Fujinaga, P. Depalle. 2005. JAUDIO: AFEATURE EXTRACTION LIBRARY. http://jmir.sourceforge.net/publications/ISMIR_2005_jAudio.pdf
[15]. https://blog.algorithmia.com/introduction-natural-language-processing-nlp
[16]. https://asiasociety.org/education/korean-language.