Voice Based Gender Classification Using Machine Learning

[1]K.Geethika Anasuya, [2] D. Soujanya, [3] Dr. K. Venkata Ramana
[1]*Department of Computer Science and Systems Engineering, Andhra University College of Engineering (A), Andhra University, Visakhapatnam - 530003, India*
[2]*Andhra University College of Engineering (A), Andhra University, Visakhapatnam - 530003, India*

Abstract **- Gender identification is one of the major problem speech analysis today. Discovering the gender from acoustic data i.e., pitch, median. Frequency etc. Machine learning gives ominous results for classification problem in all the research domains. There are several performance metrics to assess algorithms of an area. Aim is to identify gender, with five different algorithms: Linear Discriminant Analysis, K-Nearest Neighbor, Characterization and Regression Trees, Random Forest, and Support Vector Machine on premise of eight unique techniques. The main parameter in assessing any algorithms is its performance. Misclassification rate must be less than in classification problems, which says that the accuracy rate must be high. Location and gender of the person have become crucial in economic markets in the form of AdSense. Here with this comparative model algorithm, we are using the different ML algorithms and find the best one for gender classification of acoustic data.**

Index Terms **- Voice Recognition, Machine Learning, Random Forest classifier, Decision Tree Classification, K-Nearest Neighbor, Gaussian Naïve Bayes, Support Vector Machine, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Logistic Regression, ADA Boost, Guardian Boosting.**

INTRODUCTION

Intonation, speech rate, and duration are certain characteristics that distinguish human voices, mainly male and female voices. The recognize dimorphism accounts for 98.8% which consists of the gender of the speaker and the respective frequencies. Variation in gender, however, cannot be estimated by vocal speech. Some voice pitch may vary between male and female so it is difficult to recognize male and female accurately. Gender recognition can be used along with various applications

Machine learning: Machine Learning is a sunset of artificial intelligence focusing on a specific goal. Machine Learning algorithm uses data to get better at a specific task.

Computers are fed structured data and learn to become better at evaluating and acting on that data over time. Think of structured data as data inputs you can put in columns and rows. You might create a category column in Excel called food and have row entries such as fruit or meat. This form of structured data is very easy for computers to work with and the benefits are obvious. Once programing is completed a computer can take in new data indefinitely, sorting and acting on it without the need for further human intervention. Over time, the computer may be able to recognize that fruit is a type of food even if you stop labeling your data. This self-reliance is so fundamental to machine learning that the field breaks down into subsets based on how much ongoing human help is involved.

Machine learning is related to many fields, industries, and has the capability to grow over time in fields like image recognition, speech recognition, medical diagnosis, and predictive analytics.

WORKING OF MACHINE LEARNING

Machine Learning algorithm is trained using a training data set to create a model. When new input data is introduced to the ML algorithm, it makes a forecast on the basis of the model.

The forecast is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning algorithm is deployed. In the event that the precision isn't satisfactory, the Machine Learning calculation is prepared over and over with an expanded preparing informational collection. This is only an extremely

undeniable level model as there are many variables and different advances included.

RELATED WORK

Discourse arrangement and handling and gender based acknowledgment and classification have been for long period of time. We used some concepts developed over time to implement gender recognition. Late investigations dependent on gender location shows that voice is changed over first into various boundaries. Main parameters contain pitch and frequency. Classification is done to different male, female and children. For this structure need to arranged first with the planning data and test information are presented and assessed for the presentation of the framework for these information. The outcomes acquired are diverse for various calculations and appears to deliver changed outcomes at various occasions. Gender based classification using F0 frequency and pitch to classify male, female and a child. Additional tuning this based on a binning technique to improve the efficiency of the results produced.

Voice based word extraction lab view shows that algorithm works best for this classification and it results better to extract the vowels in male samples. When the samples are trained and tested efficiently it produces a result It is likewise seen that by expanding the unvoiced aspect in discourse like the sound of ‘s' worth of pitch increments hampering the gender orientation identification in the event of tests of the male. Similarly, by increasing the voice part of speech like ‘a’, it decreases the value of the pitch. When the speaker speaks two different tones it fails to identify.

Speech recognition in adult shows that they are accomplished of spontaneous and vocal length adjustments and they can sound like masculine and feminine. So, it is difficult to classify the male and female.

Some Female voices are hard to analyze based on the pitch such as examination of one aspect of female voices does not fulfill our requirements. This paper pitch between male and female shows that female voice has to be identified with a different method of parameters where female clients and it differs by a female to female henceforth, informational index should be handled dependent on this before grouping of male and female.

According to the perception of pitch. Fundamental Frequency [f0] has a combination of linguistic, paralinguistic and nonlinguistic information of the speaker and three relates to male and female and it likewise relies upon high pitch and tone of the speaker. This figured out how to put a recurrence [f0] with no involvement in range and no syllable-outer data. This shows that the voice of the speaker varies between high and low pitches between speakers.

Identification of gender by SVM shows that speech of gender is analyzed by various diverse discourse components like packed discourse, discussing phone and contrast in dialects on. It passes on that male voice from pitch, period and Mel-Frequency is in the scope of around 100-146Hz and in the female of around 188-221Hz. Here, the voice is isolated dependent on the recurrence and it is removed and it is broke down.

GMM based classification of gender propose that speech is analyzed by means of age, words, etc., a Mixture model is utilized and it distinguishes up to 98% exactness and it is one of a productive strategy for discourse location and gender investigation. Order depends on consolidated boundaries of pitch and relative otherworldly perceptual straight prescient coefficient to show male and female discourse.

METHODOLGY

The speech emotion detection system is performed as a Machine Learning (ML) model. The steps of operation are similar to any other ML project, with supplementary fine-tuning systems to make the model function adequately. The fundamental action is data collection which is of prime importance. The model being generated will acquire from the data contributed to it and all the conclusions and decisions that a progressed model will produce is supervised data. The auxiliary activity, called as component designing, is a mix of different AI tasks that are performed over the assembled information. These systems approach the various data description and data quality problems. The third step is frequently investigated the embodiment of a ML project where an algorithmic based model is created. This model uses an ML algorithm to determine about the data and instruct itself to react to any new data it is exhibited to the ultimate step is to estimate the functioning of the built model. Habitually, engineers recreate the means of creating a model and assessing it to break down the

presentation of different calculations. Measuring outcomes help to choose the suitable ML algorithm most appropriate to the predicament

A. Dataset:

To recognize the voice we have taken the several persons audio clips to classify male or female. The data set available in kaggle.com. The second step is data cleaning and pre-handling under which initially copy esteems and missing esteems are evacuated. We used Microsoft Excel to clean adjust and standardize our dataset. Third step is correlation method is applied on dataset.

The data set of audio clips is taken in the form of values of frequencies of audio. The data is classified by the values and pitch of the voice. The data set has several data some of the data as shown below:

meanfreq	sd	median	Q25	Q75	IQR	skew	kurt	sp.ent	sfm	mode	centroid	meanfun	minfun	maxfun
0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.86346	274.4029	0.893369	0.491918	0	0.059781	0.084279	0.015702	0.275862
0.066009	0.06731	0.040229	0.019414	0.092666	0.073252	22.42329	634.6139	0.892193	0.513724	0	0.066009	0.107937	0.015826	0.25
0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.75715	1024.928	0.846389	0.478905	0	0.077316	0.098706	0.015656	0.271186
0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	4.177296	0.963322	0.727232	0.083878	0.151228	0.088965	0.017798	0.25
0.13512	0.079146	0.124656	0.07872	0.206045	0.127325	1.101174	4.333713	0.971955	0.783568	0.104261	0.13512	0.106398	0.016931	0.266667
0.132786	0.079557	0.11909	0.067958	0.209592	0.141634	1.932562	8.308895	0.963181	0.738307	0.112555	0.132786	0.110132	0.017112	0.253968
0.150762	0.074463	0.160106	0.092899	0.205718	0.112819	1.530643	5.987498	0.967573	0.762638	0.086197	0.150762	0.105945	0.02623	0.266667
0.160514	0.076767	0.144337	0.110532	0.231962	0.12143	1.397156	4.766611	0.959255	0.719858	0.128324	0.160514	0.093052	0.017758	0.144144
0.142239	0.078018	0.138587	0.088206	0.208587	0.120381	1.099746	4.070284	0.970723	0.770992	0.219103	0.142239	0.096729	0.017957	0.25
0.134329	0.08035	0.121451	0.07558	0.201957	0.126377	1.190368	4.78731	0.975246	0.804505	0.011699	0.134329	0.105881	0.0193	0.262295
0.157021	0.071943	0.16816	0.10143	0.21674	0.11531	0.979442	3.974223	0.965249	0.739693	0.096358	0.157021	0.088894	0.022069	0.117647
0.138551	0.077054	0.127527	0.087314	0.202739	0.115426	1.62677	6.291365	0.966004	0.752042	0.012101	0.138551	0.104199	0.019139	0.262295
0.137343	0.080877	0.124263	0.083145	0.209227	0.126082	1.378728	5.008952	0.963514	0.73615	0.108434	0.137343	0.092644	0.016789	0.213333
0.181225	0.060042	0.190953	0.128839	0.229532	0.100693	1.36943	5.4756	0.937446	0.53708	0.219827	0.181225	0.131504	0.025	0.275862
0.183115	0.066982	0.191233	0.129149	0.240152	0.111004	3.568104	35.38475	0.940333	0.571394	0.049987	0.183115	0.102799	0.020833	0.275862
0.174272	0.069411	0.190874	0.115602	0.228279	0.112677	4.485038	61.76491	0.950972	0.635199	0.050027	0.174272	0.102046	0.018328	0.246154
0.190846	0.06579	0.207951	0.13228	0.244357	0.112076	1.562304	7.83435	0.938546	0.53881	0.050129	0.190846	0.113323	0.017544	0.275862

Fig: Sample Dataset

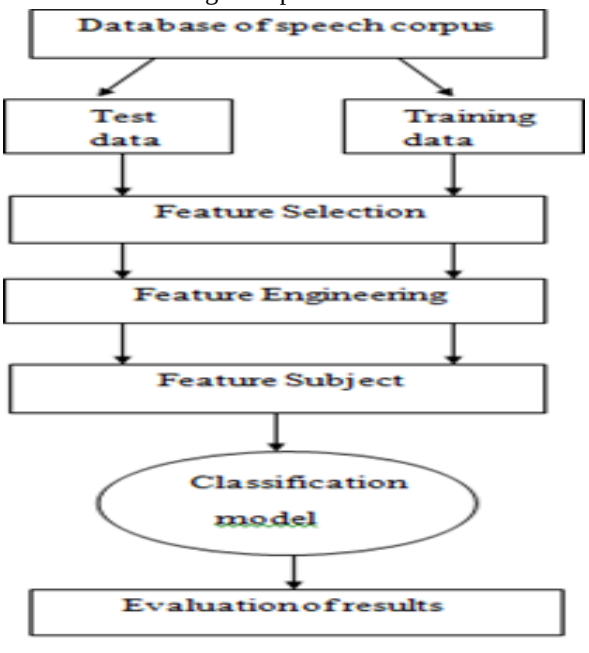

Fig: Data Flow Diagram

B. Implementation:

Data Collection: One of the first steps we perform during implementation is an analysis of the data. This was done by us in an attempt to find the presence of any relationships between the various attributes present in the dataset. The accuracy of any machine learning algorithm depends on the number of parameters and the correctness of the training dataset. In this project analysed multiple datasets collected from the Kaggle website and carefully selected the parameters that would give the best results. Many works done in this field have considered where as in some works only economic factors are taken into consideration. We have tried to combine both environmental parameters like rainfall, temperature, ph and soil parameters like soil nutrients to provide accurate and reliable recommendation to the farmer on which crop while most suitable for his land.

Data Pre-Processing: A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Information pre-handling is required errands for cleaning the information and making it reasonable for an machine learning model which likewise builds the exactness and productivity of an machine learning.

Data pre-processing is an important step as it helps in cleaning the data and making it suitable for use in machine learning algorithms. Most of the focus in pre-processing is to remove any outliers or erroneous data, as well as handling any missing values. Missing data can be delete with in two ways. The first method is to simply remove the entire row which contains the missing or erroneous value. While this easy to execute method, it is better to use only on large datasets. Using this method on small datasets can reduce the dataset size too much, especially if there are a lot of missing values. This can severely affect the accuracy of the result. Since ours is a relatively small dataset, we will not be using this method.

Training and Testing Data: The proposed model needs to be trained and tested under various conditions by altering here we used the different ML and DL algorithms to obtain reliable output.

Results and Analysis: The prediction results have been evaluated using following parameters:

Accuracy: It is the proximity of a computation to the true value which is calculated by taking true positive and true negative with a fraction of true positive, true negative and false positive with false negative.

Algorithms applied:

a) Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It tends to be utilized for both characterization and relapse issues in ML. It depends on the idea of group realizing which is a course of joining various classifiers to tackle a perplexing issue and to work on the presentation of the model.

b) Decision Tree is a supervised learning technique that can be used for both order and relapse issues however generally it is liked for tackling characterization issues. It is a tree-organized classifier, where inward hubs address the elements of a dataset, branches address the choice principles and each leaf hub addresses the outcome.

c) K-Nearest Neighbor is one of the simplest Machine Learning algorithm based on Supervised Learning technique. It expects the closeness between the new case and accessible cases and put the new case into the classification that is generally comparable

d) Gaussian Naïve Bayes is a variant of Naïve Bayes that follows Gaussian normal distribution and support continuous data. When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal distribution.

e) Support Vector Machines are a binary classification algorithm. Support vectors are the data points adjacent to the hyper-planes if the dataset is removed it will change the position of the dividing hyper-plane

f) Linear Discriminant Analysis is a dimensionality reduction technique that is commonly used for supervised classification problems. It is utilized for displaying contrasts in bunches i.e. isolating at least two classes. It is used to project the features in higher dimension space

g) Quadratic Discriminant Analysis is not really that much different from LDA except that you assume that the covariance matrix can be different for each class and so we will estimate the covariance matrix

h) Logistic Regression is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead. This can be stretched out to demonstrate a few classes of occasions, for example, deciding if a picture contains a feline, canine, flank, and so forth. Each object being detected in the image would be assigned a probability between 0 and 1 with a sum of one.

i) Ada-Boost calculation, short for Adaptive Boosting is a Boosting procedure utilized as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the loads are re-appointed to each occasion with higher loads alloted to erroneously arranged examples

j) Gradient boosting algorithm is one of the most powerful algorithm in the field of machine learning. As we realize that the mistakes in machine learning calculations are comprehensively characterized into two classifications I.e. Predisposition Error and Variance Error. As gradient boosting is one of the boosting algorithm it is used to minimize bias error of the model.

RESULTS

Accuracy for Random Forest classifier:

```
Accuracy with all variables: 96.63 %
Accuracy with 5 less variables: 96.81 %
```

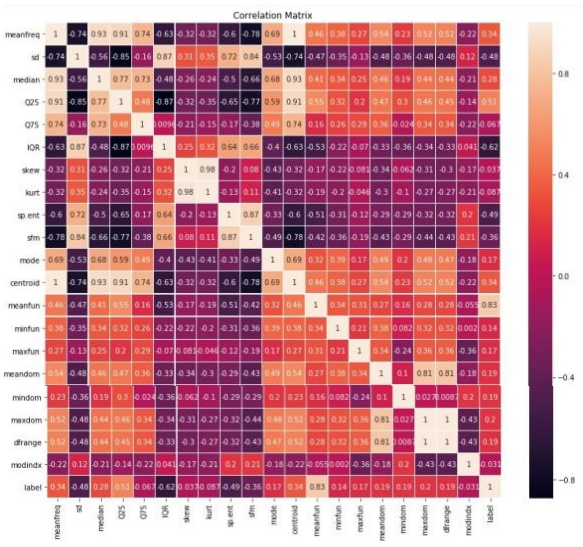

Fig: Correlation Matrix

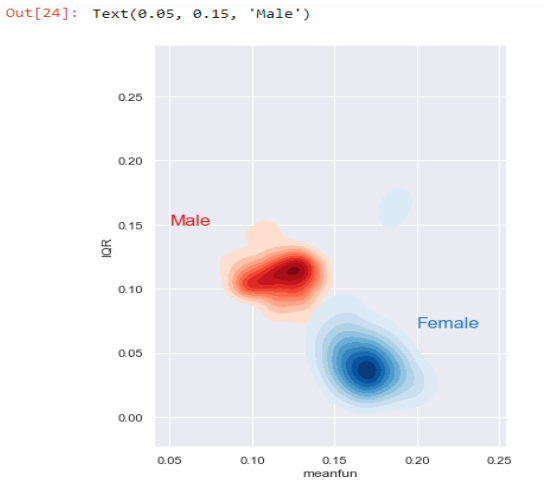

Fig: KDE plot

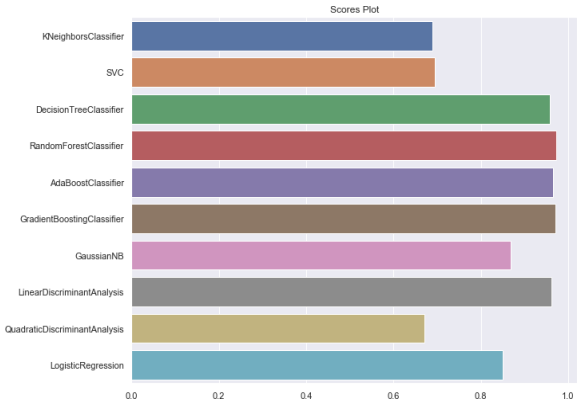

Fig: Bar plot for other classifiers

Fig: Bar plot for SVC

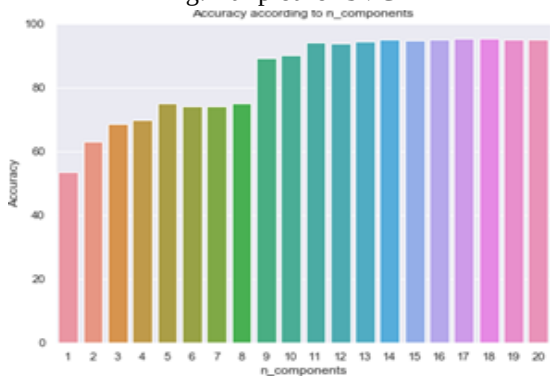

Fig: Bar plot for PCA

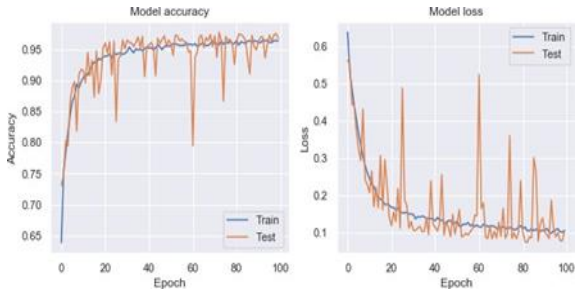

Fig: Accuracy and loss values

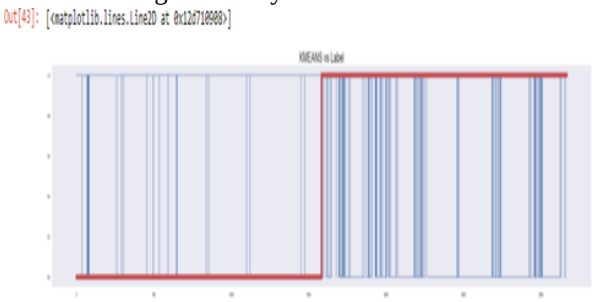

Fig: K-means versus label

CONCLUSION

By Comparative algorithm of comparing the above-mentioned different algorithms. The results obtained shows that Random Forest algorithm performs better in classification and with reduced error rate. These outcomes acquired utilizing this Comparative calculation are just for this voice gender dataset and it might change for another dataset. Random Forest tends to have more accuracy over another algorithm in classifying gender in spite of variations in pitch and frequency. Future work to add more algorithms to this Comparative model and to compare the performance with this work and to identify which algorithm in linear and Non-Linear performs better in the classification of gender in voice gender dataset.

REFERENCES

[1] Ericsdotter C and Ericsson A M 2001 Gender differences in vowel duration in read Swedish: Preliminary results Working Papers - Lund University Department of Linguistics 34 – 37 IOP Conf. Series: Materials Science and Engineering 263 (2017) 042083 doi:10.1088/1757-899X/263/4/042083

[2] Whiteside S P 1996 Temporal-Based Acoustic-Phonetic Patterns in Read Speech: Some

Evidence for Speaker Sex Differences, J International Phonetic Association 26 23–40

[3] Byrd D 1992 Preliminary results on speaker-dependent variation in the TIMIT database J AcoustSoc Am 92 593–596

[4] Henton C G 1989 Fact and fiction in the description of male and female pitch Language and Communication 9 299-311

[5] Bishop J and Keating P 2012 Perception of pitch location within a speaker's range: Fundamental Frequency, voice quality and speaker sex The Journal of the Acoustical Society of America 32-2 1100-1112

[6] Muhammad G, AlSulaiman M, Mahmood A and Ali Z 2011 Automatic voice disorder classification using vowel formants, 2011 Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '11) 1–6

[7] Gaikwad S, Gawali B, and Mehrotra S C 2012 Gender identification using SVM with combination of MFCC, Advances in Computational Research 4 69–73

[8] Machine Learning Random Forest Algorithm https://www.javatpoint.com/machine-learning-random-forest-algorithm(dateaccessed: August 11, 2021)

[9] Machine Learning Decision Tree Classification Algorithm https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm (date accessed: August 11, 2021)

[10] K-Nearest Neighbor(KNN) algorithm for machine learning https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning(date accessed: August 11, 2021)

[11] Linear Discriminant Analysis Algorithm for Machine Learning https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/(data accessed: October 6,2021)