International Journal of Electrical Electronics & Computer Science Engineering
Special Issue - NCSCT-2018 | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
March, 2018 | Available Online at www.ijeecse.com

Gender Determination of Humans Using Their Voice Samples

Garvit Burad[1], Ankita Sharma[2]
[1-2]UG Scholar, Computer Science and Engineering Department, PIET, Jaipur, Rajasthan, India
[1]garvit.burad@gmail.com, [2]ankitasharma5911@gmail.com

Abstract: ***The aim of this paper is to decide a person's gender as male or female, by using the sample of their voice. Generally, the human ear can undoubtedly distinguish the contrast between a male or female voice by listening to just few spoken words. Nonetheless, planning a PC program to do this ends up being somewhat trickier.In this paper, different classification algorithms are used to classify the voice samples into two classes of gender namely male and female with high accuracy.***

Keywords: Machine Learning, Classification, Voice Analysis, Natural Language Processing, Gender Classification, Support Vector Machine, NLP.

I. INTRODUCTION

This paper depicts the plan of a PC program to display acoustic investigation of voices and discourse for deciding the gender of the human being. The model is developed utilizing 3,168 recorded examples of male and female voices, speech, and utterances. The examples are prepared utilizing acoustic investigation and afterward connected to an artificial intelligence/machine learning algorithm to learn gender-specific traits. The learning from these machine learning models after training on the training dataset and testing it on the testing dataset can be applied in various practical scenarios like it can be used in voice assistants now available on our mobile devices and in various electrical equipment's to identify the gender of the user based on the voice sample and give the results tailored according to the gender of the user. The algorithm we are going to use are based on Support Vector Machine like default linear kernel, RBF kernel, polynomial kernel, k-cross validation on these and also perform grid search to find the best parameters to optimize the algorithm.

II. FEATURES

Clustering discourse acknowledgment innovation alludes to the capacity of a PC regarding listening input words and giving the rectify implications of words, as it were; it changes the input sound documents to content. This is called Automatic Speech Acknowledgment (ASR). The innovation can apply to PC also; enable the PC to recognize words from people which input discourse through a gadget, for example, an amplifier or phone gadget. The principle highlight of discourse acknowledgment application ought to need to see every one of the words accurately regardless of whether it is the expressions of anybody. This isn't restricted to the size of the vocabulary, the tumult, the nature of the speaker, and elocution of conceivable channels which are Text-to-Speech what's more, Speech-to-Text. The current research has demonstrated that the most famous working frameworks are Android and Microsoft.

The extent of Android is at 38.85% and Windows is at 36.96% as indicated by kStat Counter Global Stats. Along these lines, the mainstream discourse acknowledgment API is Microsoft and Google.

The dataset we used for this paper had following attributes

- Total number of samples: 3168
- Number of male: 1584
- Number of female: 1584
- 22 features
- 1 label- Male/Female

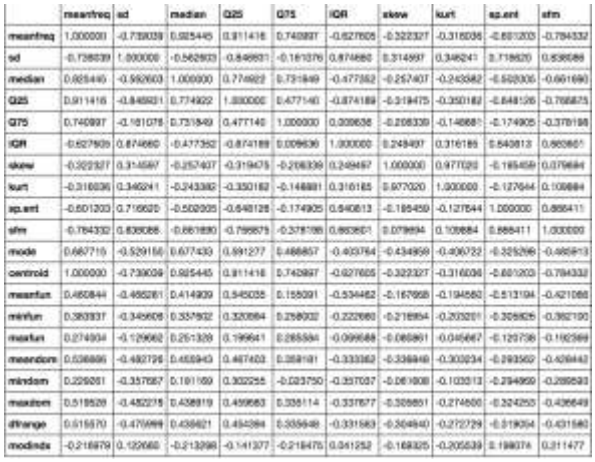

Fig. Correlation Factor

A. Voice Frequency:

A voice recurrence (VF) or voice band is one of the frequencies, inside piece of the sound range that is being utilized for the transmission of discourse.

In communication, the usable voice recurrence band ranges from around 300 Hz to 3400 Hz. It is thus that the ultra low recurrence band of the electromagnetic range in the vicinity of 300 and 3000 Hz is likewise alluded to as voice recurrence, being the electromagnetic vitality that speaks to acoustic vitality at baseband. The transfer speed assigned for a solitary voice-recurrence transmission channel is typically 4 kHz, including monitor groups, permitting a testing rate of 8 kHz to be utilized as the premise of the beat code balance framework utilized for the computerized PSTN. Per the Nyquist– Shannon examining hypothesis, the testing recurrence (8 kHz) must be no less than double the most elevated segment of the voice recurrence by

means of suitable separating preceding inspecting at discrete circumstances (4 kHz) for successful recreation of the voice flag.

B. Mean Frequency:

The mean frequency of a spectrum is calculated as the sum of the product of the spectrogram intensity (in dB) and the frequency, divided by the total sum of spectrogram intensity.

C. Quartiles:

A quartile is a type of quantile. The first quartile (Q1) is defined as the middle number between the smallest number and the median of the data set. The second quartile (Q2) is the median of the data. The third quartile (Q3) is the middle value between the median and the highest value of the data set.

The features present in the dataset are

- meanfreq: mean frequency (in kHz)
- sd: standard deviation of frequency
- median: median frequency (in kHz)
- Q25: first quantile (in kHz)
- Q75: third quantile (in kHz)
- IQR: interquantile range (in kHz)
- skew: skewness (see note in specprop description)
- kurt: kurtosis (see note in specprop description)
- sp.ent: spectral entropy
- mode: mode frequency
- centroid: frequency centroid (see specprop)
- peakf: peak frequency (frequency with highest energy)
- meanfun: average of fundamental frequency measured across acoustic signal
- minfun: minimum fundamental frequency measured across acoustic signal
- maxfun: maximum fundamental frequency measured across acoustic signal
- meandom: average of dominant frequency measured across acoustic signal
- mindom: minimum of dominant frequency measured across acoustic signal
- maxdom: maximum of dominant frequency measured across acoustic signal

Fig. Visualizing the Data Based on Various Features

III. ALGORITHMS

Below are the algorithms used:

A. SVM

Various leveled Support vector machines (SVMs) are a game plan of controlled learning methodologies used for gathering, backslide and special cases disclosure. The upsides of help vector machines are: Powerful in high dimensional spaces. Still intense in circumstances where number of measurements is more vital than the amount of tests. Usages a subset of planning centers in the decision limit (called bolster vectors), so it is similarly memory beneficial.

Versatile: unmistakable Kernel limits can be resolved for the decision limit. Customary pieces are given, notwithstanding it is also possible to demonstrate custom parts.

The damages of help vector machines include:

If the amount of features is altogether more imperative than the amount of tests, avoid over-fitting in picking Kernel limits and regularization term is basic.

SVMs don't direct give probability checks, these are processed using an expensive five-cover cross-endorsement (see Scores and probabilities, underneath).

Running SVM with default hyperparameter.

Accuracy Score: 0.9763406940063092

B. Linear Support Vector Classification:

Similar to SVC with parameter kernel='linear', however actualized as far as liblinear instead of libsvm, so it has greater adaptability in the selection of punishments and misfortune works and should scale better to vast quantities of tests.

a. This class supports both thick and scanty info and the multiclass support is taken care of as per a one-versus the-rest conspire.

b. Core points: These are the points which are available inside the bunch. A point is considered to be inside the bunch if the measure of data inside the area surpasses an exact edge cost.

c. Border points: These are the points which are available in the area of the center indicates and are considered be the part that group.

d. Exceptions: These are the points which are not the piece of the group and henceforth are thought to be the commotion.

Accuracy Score: 0.9779179810725552

C. Radial Basis Function (RBF) kernel SVM:

In machine learning, the (Gaussian) radial basis work kernel, or RBF kernel, is a famous kernel work utilized as a part of different kernelized learning algorithms. Specifically, it is usually utilized as a part of support vector machine classification.

The RBF kernel on two examples x and x', spoke to as feature vectors in some information space. the impact of the parameters gamma and C of the Radial Basis Function (RBF) kernel SVM.

Naturally, the gamma parameter portrays how far the effect of a lone getting ready case comes to, with low regards implying 'far' and high regards meaning 'close'. The gamma parameters can be seen as the regressive of the traverse of effect of tests picked by the model as help vectors. The C parameter trades off misclassification of planning bodies of evidence against straightforwardness of the decision surface. A low C settles on the decision surface smooth, while a high C goes for describing all readiness outlines precisely by giving the model adaptability to pick more cases as help vectors. The lead of the model is outstandingly fragile to the gamma parameter. In case gamma is excessively tremendous, the scope of the domain of effect of the help vectors just consolidates the help vector itself and no measure of regularization with C will have the ability to envision overfitting. Right when gamma is pretty much nothing, the model is unnecessarily constrained and can't get the diserse quality or "shape" of the data. The region of effect of any picked help vector would consolidate the whole getting ready set. The ensuing model will bear on relatively to a direct model with a course of action of hyperplanes that diverse the focal points of high thickness of any match of two classes.

Accuracy Score: 0.9763406940063092

D. K- Fold Cross Validation:

Cross-validation, once in a while called pivot estimation, is a model approval framework for reviewing how the delayed consequences of a quantifiable examination will entirety up to a free enlightening list. It is overwhelmingly used as a piece of settings where the goal is desire, and one needs to evaluate how definitely an insightful model will perform eventually. In a gauge issue, a model is ordinarily given a dataset of known data on which getting ready is run (planning dataset), and a dataset of cloud data (or first watched data) against which the model is attempted (called the approval dataset or testing set).The objective of cross validation is to portray a dataset to "test" the model in the arrangement organize (i.e., the approval set), in order to keep issues like overfitting, give a learning on how the model will total up to a self-governing dataset (i.e., a darken dataset, for instance from a bona fide issue) and so on.

One round of cross-validation includes apportioning an example of information into integral subsets, playing out the examination on one subset (called the preparation set), and approving the investigation on the other subset (called the validation set or testing set). To diminish inconstancy, in many strategies numerous rounds of cross-validation are performed utilizing diverse parcels, and the validation comes about are joined (e.g. found the middle value of) over the rounds to assess a last prescient model.

One of the principle purposes behind utilizing cross-validation as opposed to utilizing the regular validation (e.g. dividing the informational collection into two arrangements of 70% for preparing and 30% for test) is that there isn't sufficient information accessible to parcel it into isolated preparing and test sets without losing noteworthy demonstrating or testing ability. In these cases, a reasonable method to legitimately appraise display expectation execution is to utilize cross-validation as an effective general technique.

CV on Linear kernel

Mean Accuracy:0.9694132632752168

CV on RBF kernel

Mean Accuracy: 0.9659382214791815

Taking every one of the estimations of C and looking at the precision score with kernel as linear. The C parameter tells the SVM improvement the amount you need to abstain from misclassifying each preparation case. For huge estimations of C, the advancement will pick a littler edge hyperplane if that hyperplane completes a superior occupation of getting all the preparation focuses characterized effectively. On the other hand, a little estimation of C will cause the streamlining agent to search for a bigger edge isolating hyperplane, regardless of whether that hyperplane misclassifies more focuses. Therefore, for an extensive esteem we can cause overfitting of the model and for a little estimation of C we can cause underfitting. Thus

the estimation of C must be picked in such a way, to the point that it summed up the concealed information well.

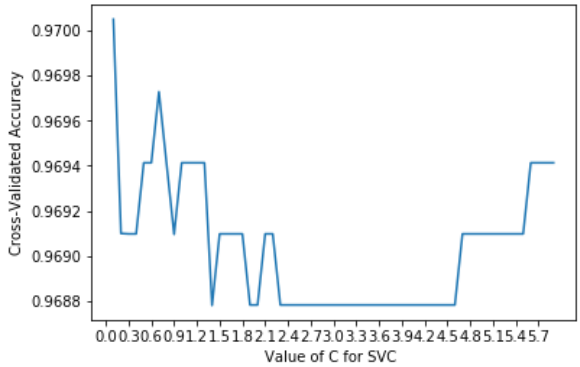

Fig Cross-validation Accuracy for C Parameters

From the above plot we can see that accuracy has been close to 97% for C=1 and C=6 and then it drops around 96.8% and remains constant.

E. Taking Kernel as RBF and Taking Diverse Values Gamma:

The gamma parameter is the reverse of the standard deviation of the RBF kernel (Gaussian capacity), which is utilized as similitude measure between two focuses. Naturally, a little gamma values a characterize a Gaussian capacity with an expansive change. For this situation, two focuses can be viewed as comparable regardless of whether are a long way from each other. In the other hand, an expansive gamma esteem implies characterize a Gaussian capacity with a little fluctuation and for this situation, two focuses are viewed as comparable just in the event that they are near each other.

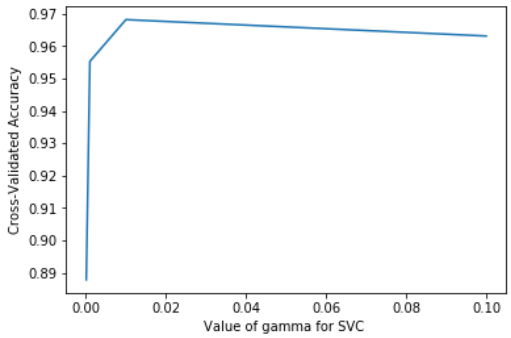

Fig Relation between CV and Gamma

Performing SVM by taking hyperparameter C=0.1 and kernel as linear

Accuracy: 0.9747634069400631

F. Performing Grid Search:

Grid search enables us to search all parameter combination of give hyper parameter space and evaluate models. After grid search you can obtain the best hyper parameter set. It's performed for SVC Linear kernel and SVC RBF Kernel

Accuracy : 0.9794790844514601

G. Learning Curve:

Learning curve enables us decide a model is over fitting to given training data and training under appropriate bias and variance balance.

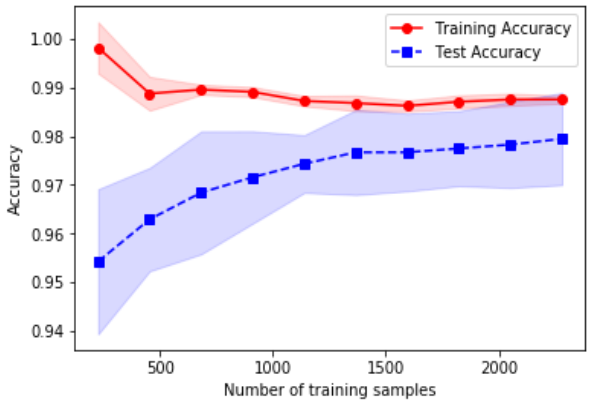

Fig. Learning Curve

H. Validation Curve:

So we can estimate the best value of C can be 1 if C is over 1 test accuracy is decreasing. It can cause overfitting of this model.

So how can we decide the best hyper parameter of this model without checking learning curve and validation curve one by one.

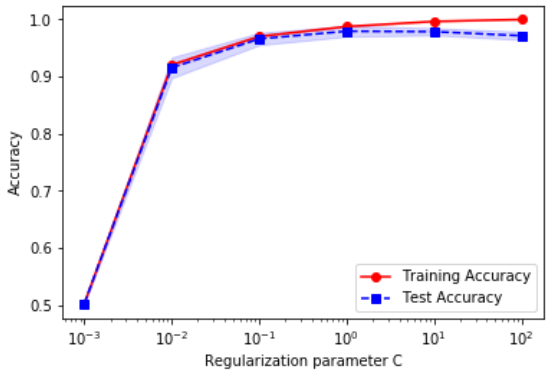

Fig. Validation Curve

IV. RESULT

The highest accuracy score was achieved by using the Grid search method to find the best hyperparameter and using that on the combination of SVC Linear kernel and SVC RBF Kernel.

Accuracy: 0.9794790844514601

V. CONCLUSION

We have achieve some very high accuracy results on the dataset with our approach of using Support Vector Machine for the classification of the voices into two gender categories namely male and female. The future

scope of this study is in implementation of this model in various voice assistants to give the personalized results.

ACKNOWLEDGMENT

This research paper was supported by Poornima Institute of Engineering and Technology as they have shared their pearls of wisdom with us during the course in this research. I would also like to thank my guide Dr. Rekha Jain who provided his support and expert insight towards the research, her comments has greatly improved the manuscript.

VI. REFERENCES

[1] The Harvard-Haskins Database of Regularly-Timed Speech.

[2] Telecommunications & Signal Processing Laboratory (TSP) Speech

[3] Database at McGill University

[4] VoxForge Speech Corpus

[5] Festvox CMU_ARCTIC Speech Database at Carnegie Mellon University FLEXChip Signal Processor (MC68175/D), Motorola, 1996.