INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

GENDER CLASSIFICATION THROUGH VOICE AND PERFORMANCE ANALYSIS BY USING MACHINE LEARNING ALGORITHMS

B. Satya Prasad
Vardhaman College of Engineering
Shamshabad, India
prasadsatya952@gmail.com

Abstract- There are various machine learning algorithms for classification of gender but to choose the best one is more difficult task. To select the best machine learning algorithm I conducted an experimental study for gender classification. In this experimental study, we analyzed the accuracy (performance) of various models for classification of gender through voice. So we conclude that ANN and SVM are giving the best results.

Keywords- Machine Learning; ANN; SVM; accuracy; classification;

1. INTRODUCTION

Prediction of gender is useful in many applications like interactive systems, targeted advertisements, health care systems through mobile, recognition systems, crime analysis (identification of gender taken part in crime through voice) and so on. In general, a voice and speech is used for gender classification. A natural voice recognition system is human ear. A human ear has a mechanism to distinguish the gender by voice and speech based on various factors. In the same way, a machine is to be trained to classify the gender through voice using machine learning algorithms. Classifying the gender through voice is a challenging problem in machine learning.

The data we collected from different repositories is of unstructured form. Hence, Deep Learning models can be used on unstructured data like images, video and audio. Deep learning model gives the better result when the data is too large. The dataset consists of 3168 male and female voice acoustic features to be trained and tested by using advanced machine learning algorithms.

Data Science is the used to extract insights from data by algorithms and scientific methods.

Through data science one can work on huge data, analysis the data and visualize the data. Machine learning is the field that makes machines to learn without being programmed. It is mainly used to combine the statistical tools with the data to predict the class label.

Some of its applications are financial services, healthcare, virtual personal assistants, videos surveillance, search engine and so on.

The problems can be solved by supervised learning, unsupervised learning, reinforcement learning and semi-supervised learning. In this paper we are also comparing the accuracy of various machine learning algorithms.

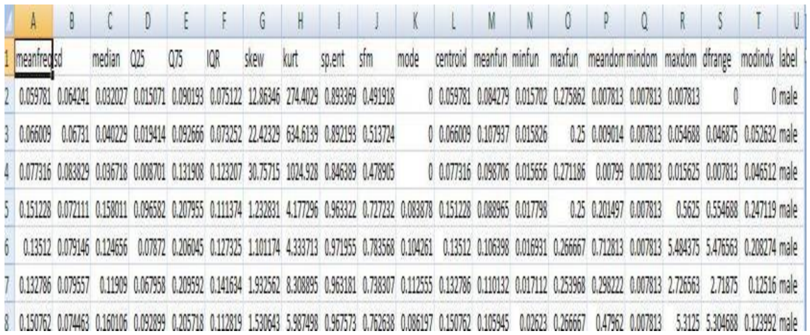
Vol.7 Issue 11, Pg.: 1-11
November 2019

2. LITERATURE REVIEW

There are various deep learning models and machine learning algorithms to classify the gender of a person based on voice. In [1], pitch used the Multi-Layer Perception Neural Networks for the classification of gender achieved with accuracy of 96%. In [2], Lee and Lang used the Support Vector Machine (SVM) for the gender classification. In [3], Silvosky and Nouza used the Gaussian Mixture Model to classify the gender. In [4] and [5] the machine learning algorithms like Support Vector Machine, Regression and classification can also be used. In [6] Support Vector Machine is used to classification purpose and achieved an accuracy of 95%.

3. VOICE DATASET

The dataset consists of total 3168 voice samples of male and female with voice acoustic properties. The acoustic properties are meanfreq, mode, sd, median, Q25, Q75, skew, IQR, kurt, sp.ent,
meanfun, minfun, centroid, maxfu, mindom, sfm, meandom, maxdom, dfrange, modindex, label.
The fields of voice dataset are described in TABLE 1.
The whole voice data set is stored in CSV format which is viewed in below Fig1.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	meanfreq	sd	median	Q25	Q75	IQR	skew	kurt	sp.ent	sfm	mode	centroid	meanfun	minfun	maxfun	meandom	mindom	maxdom	dfrange	modindx	label
2	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	0.007813	0.007813	0.007813	0	0	male
3	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	0.009014	0.007813	0.054688	0.046875	0.052632	male
4	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	0.00799	0.007813	0.015625	0.007813	0.046512	male
5	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	0.201497	0.007813	0.5625	0.554688	0.247119	male
6	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	0.712813	0.007813	5.484375	5.476563	0.208274	male
7	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	0.298222	0.007813	2.726563	2.71875	0.12516	male
8	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	0.47962	0.007813	5.3125	5.304688	0.123992	male

Fig. 1. Voice Dataset

TABLE 1. Acoustic properties

Properties	Description
Meanfreq	mean frequency
Mode	mode frequency
Sd	standard deviation of frequency
Median	median frequency
Q25	first quantile
Q75	third quantile
Skew	Skewness
IQR	interquantile range
Kurt	Kurtosis
Centroid	frequency centroid
sfm	spectral flatness
sp.ent	spectral entropy
Meanfun	average of fundamental frequency measured across acoustic signal
Minfun	minimum fundamental frequency measured across acoustic signal
Maxfu	maximum fundamental frequency measured across acoustic signal
Mindom	minimum of dominant frequency measured across acoustic signal
Meandom	average of dominant frequency measured across acoustic signal
Maxdom	maximum of dominant frequency measured across acoustic signal
Dfrange	maximum of dominant frequency measured across acoustic signal
Modindex	modulation index
Label	male or female

Vol.7 Issue 11, Pg.: 1-11
November 2019

4. ALGORITHMS

4.1 Logistic Regression

Logistic regression is applied when the target (dependent) value is categorical. For example, to predict whether the person is happy (1) or sad (0), to predict the gender male (1) or female (0) and so on. In Logistic regression, the target value is coded as 1(yes, success, good, living …) and 0(no, failure, bad, dead …). It can also be expressed in terms of prediction as p(Y=1) as a function of X.
Logistic Regression is used when the target variable is dichotomous. Hence, this model fits to the research paper because our target variable predicts ‘male’ and female’. The example how can we predict target variable ‘Happy’ or ‘Sad’ is shown in Fig.2.

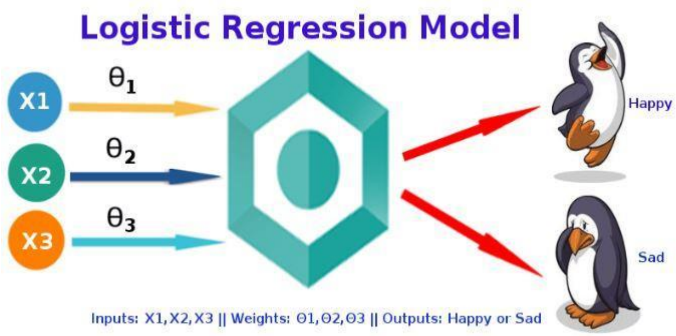


Fig. 2 Logistic Regression

Since the probability lies between 0% and 100% (or o and 1), when we plot the probability of target value it will give a ‘S’ shape curve as Fig3. That means all the probability of occurring ‘yes’ to one side and all the probability of occurring ‘no’ to other side.

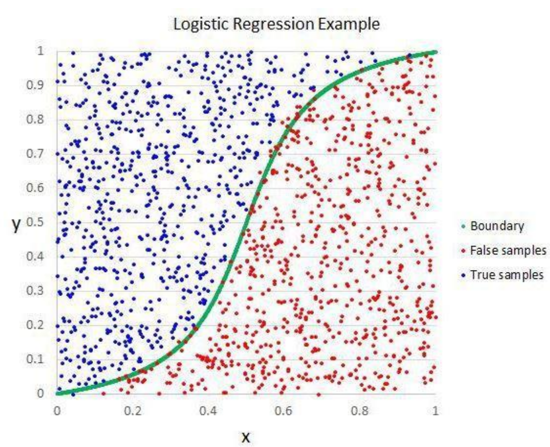
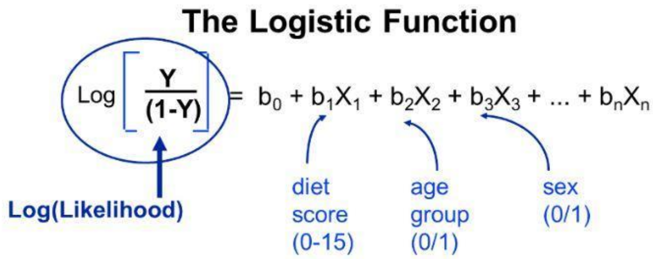


Fig.3 Logistic regression curve

Logit transformation= Log (p/1-p)= Log(probability of event occurring/probability of event not occurring)

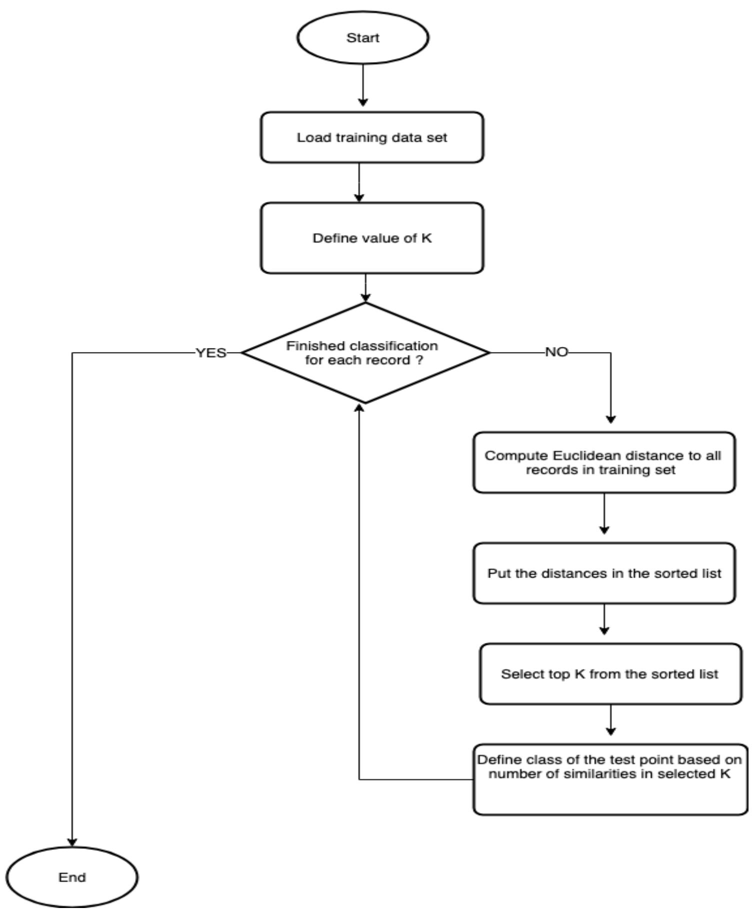
$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

Logistic regression can handle any number of categorical and numerical values.



4.2 KNN

K-Nearest Neighbor is generally used for both regression and classification predictive analysis.
KNN is used to predict on training set directly. Predictions are done for a instance (X) by searching the training set for the K most instances (the neighbors) which are similar and predicting the target variable for those K instances. To determine the similarity in the K most instances in the training set a distance measured is used. For real-valued variables the most common distance measure used is Euclidean distance.
Euclidean Distance(x, xi) = sqrt (sum (xj-xij)^2)).
The other distance measures are Hamming distance, Manhattan distance, Minkowski distance, Jaccard distance, Tanimoto distance, Cosine distance and Mahalanobis distance.



INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
www.ijrcar.com
Vol.7 Issue 11, Pg.: 1-11
November 2019

Fig.4 KNN

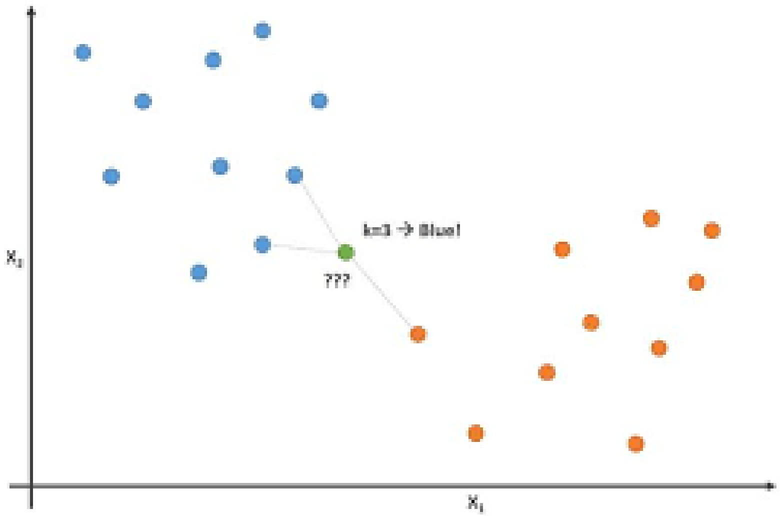


Fig.5 KNN plot

4.3 Naïve Bayes

A Naïve Bayes Classifier is a probalistic model used for classification. Using Bayes theorem we find the probability of ‘A’ happening that ‘B’ has occurred. Here, ‘A’ is the hypothesis and ‘B’ is the evidence. Here the assumption made here is that the features/predictors are independent.
Naïve Bayes model is mainly used for spam filtering, recommendation systems, sentiment analysis and so on.

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

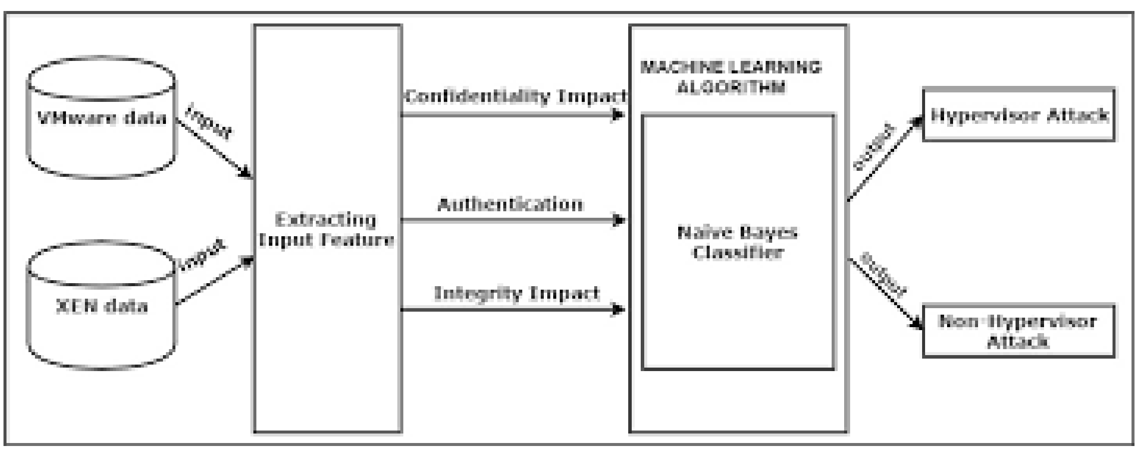


Fig.6 Naïve Bayes

4.4 Decision Tree

A decision tree is a classification algorithm and it is a supervised learning. A Decision tree consists of nodes, leaf nodes and edges. A decision tree is a tree structure in which each internal node represents attribute or a feature , the branch represents a decision rule and each leaf node represents the output or target variable. The below fig shows the decision tree algorithm works. The process for classification of dataset is showed in Fig.7 .

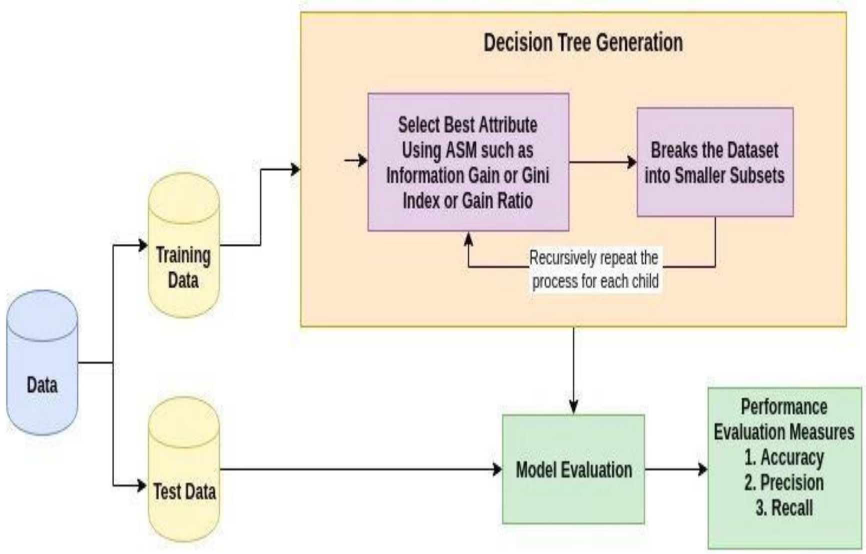


Fig.7 Decision Tree

The decision tree is constructed by information gain, gain ratio and gini index and how the decision tree is constructed by using the information gain, gain ratio and gini index is showed in the figure Fig.8.

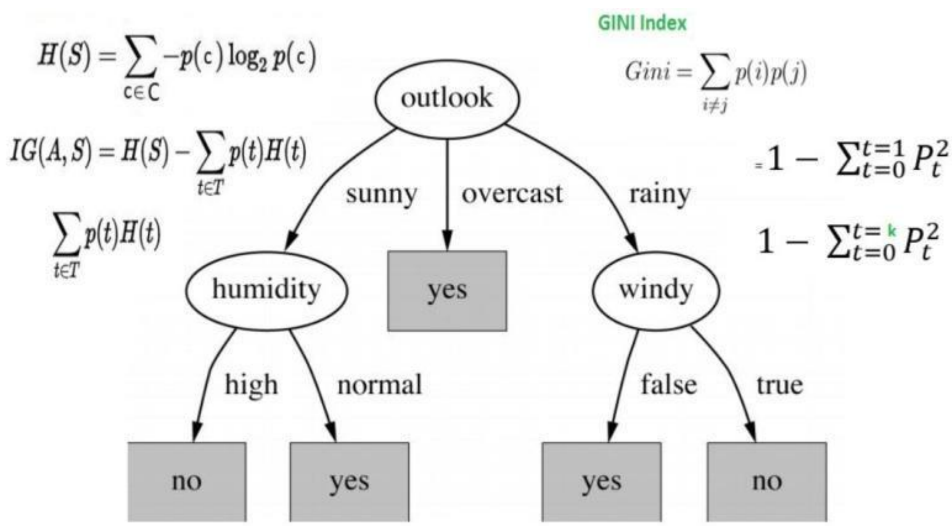


Fig.8 Construction of decision tree

4.5 Random Forest

Random forest model is an ensemble algorithm. Ensemble algorithm combines algorithms of similar type or different type for classifying variables or objects. This classifier creates decision trees from training dataset. It then check or calculate the votes from different decision trees to decide the target variable or final class of test object. The features of random forest are more efficiency working on large data sets, handles on more variables, estimates which variables are important during classification, estimates missing data, and so on. As shown in th e fig 9 we can construct many decision trees at a time and through voting provide the target variable.

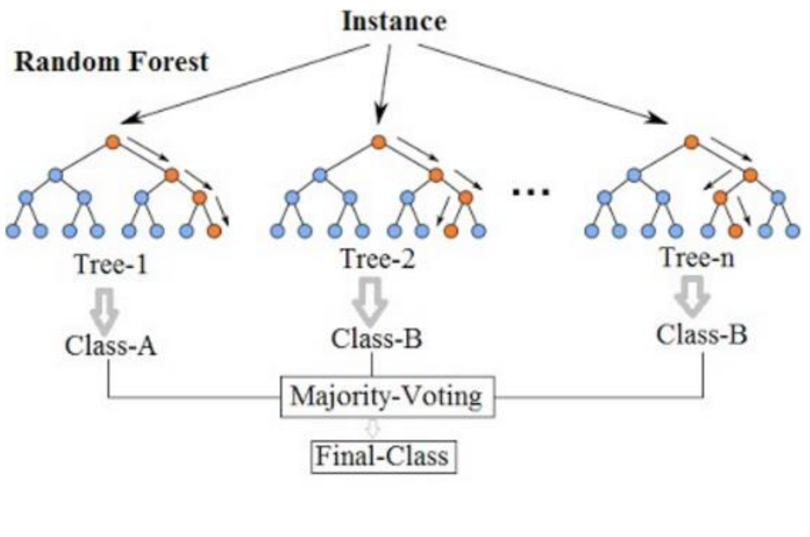
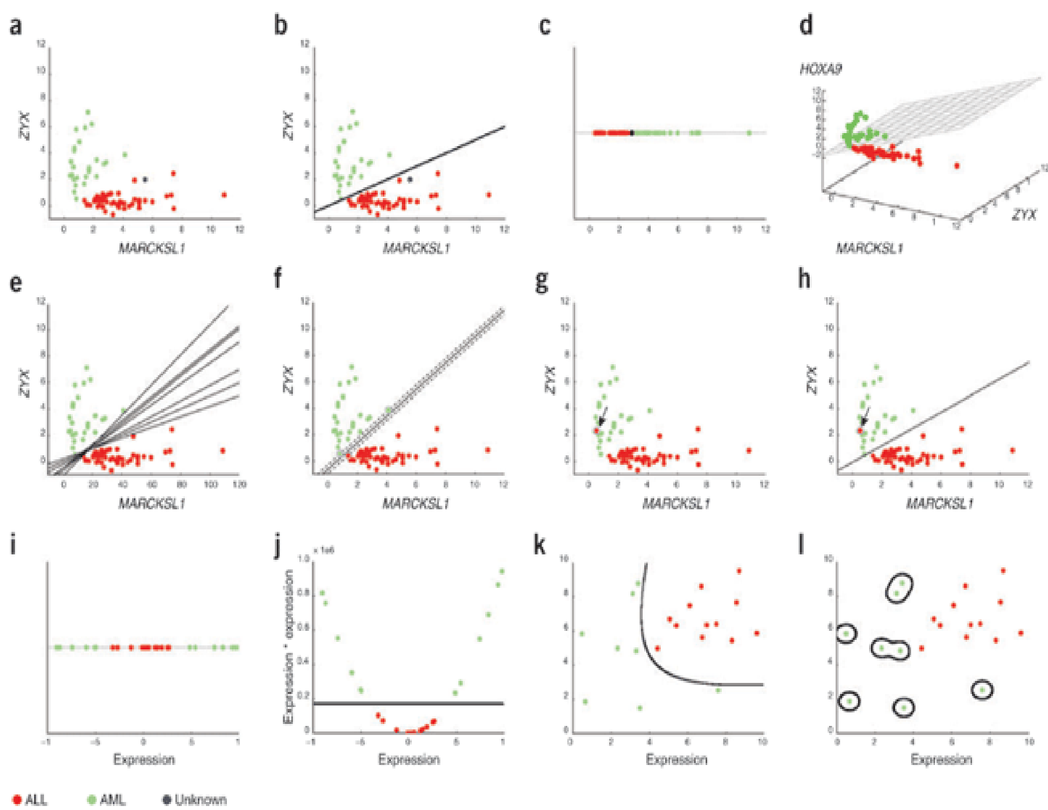


Fig.9 Random forest

4.6 SVM

A Support Vector Machine is a classifier is defined by separating with a hyperplane. It comes under supervised learning and the hyperplane classifies the target variables or data points. The hyperplane is used for categorizing data points. Data points which are on either side of hyperplane create their own classes.
As shown in the fig 10 the various representations of hyperplane which classifies the outputs and create the classes.



Fig, 10 SVM

Vol.7 Issue 11, Pg.: 1-11
November 2019

4.7 ANN

Artificial Neural Network is a computational algorithm. A neural network is a oriented graph consists of nodes connected by arcs. A neural network contains three layers they are input layer, hidden layer and output layer as shown in fig 12. Neural network is organized by layers. Layers are made by interconnecting the nodes which has a activation function. The structure of neural network also referred as architecture or topology.

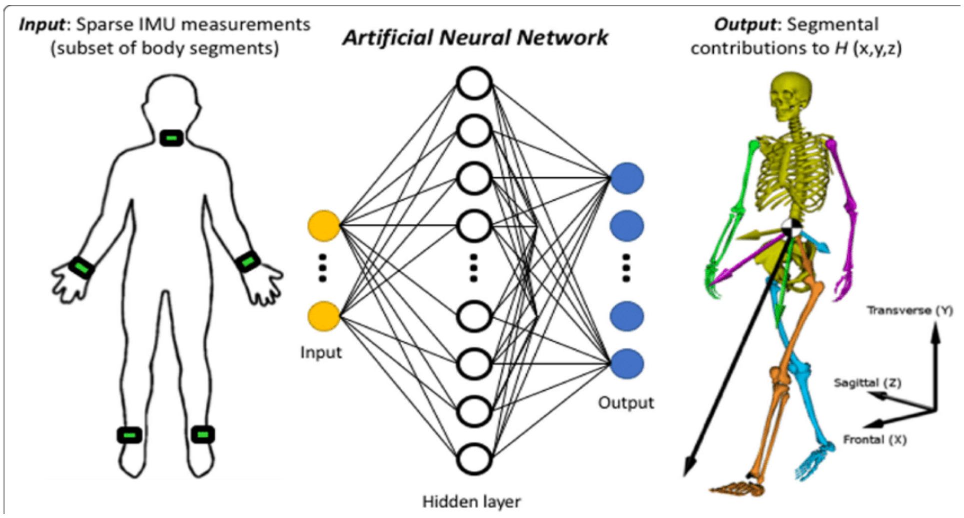


Fig11 ANN

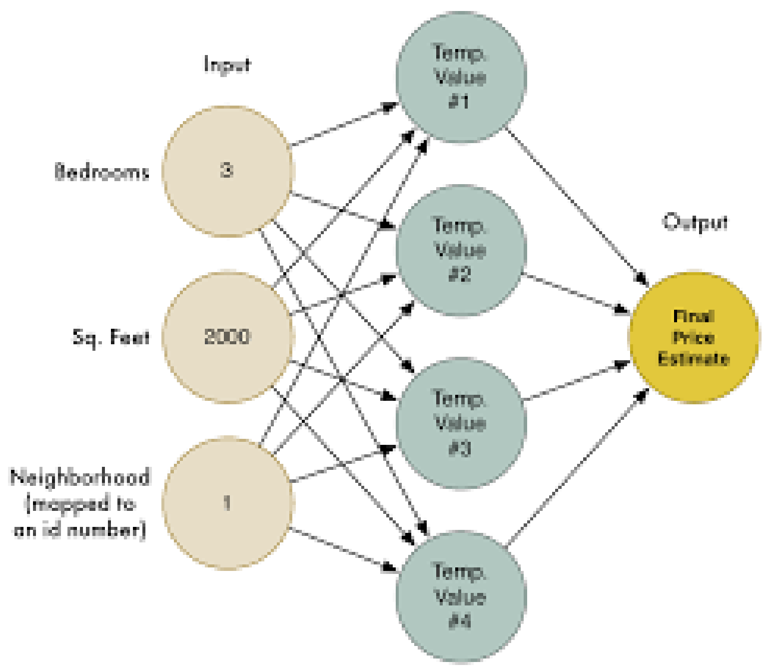


Fig.12 Neural network with layers

Vol.7 Issue 11, Pg.: 1-11
November 2019

5. PERFORMANCE ANALYSIS

Classification algorithms are generally used for identify the gender, spam detection, intruder detection, etc. In this paper we have compared various classification models using voice dataset.
Also we conducted experiments using machine learning algorithms on voice dataset. The train and test accuracies are observed and analyzed for seven classification algorithms. The seven classification algorithms we used are Logistic regression, SVM, ANN, Naïve bayes, KNN, Decision tree and Random Forest.
We used sklearn for data preprocessing, there are no missing values in the voice dataset and label encoder. We also used pandas library and numpy library to load dataset, perform calculations and sklearn for modeling the machine learning algorithm. Tensorflow and Keras are used in ANN. We used cross validation to train the models. The accuracies are shown in table 2 below

TABLE 2. Accuracies

Model	Training dataset	Testing dataset
KNN	98.5	97.72
ANN	99.9	98.35
Logistic Regression	97.26	97.7
Naïve Bayes	89.3	89.39
SVM	97.4	97.97
Random Forest	99.832	97.60
Decision Tree	99.99	96.71

If we observed the above table both the ANN and SVM are giving more accuracy than other algorithms. We can further improve these algorithms by parameter tuning.

6 RESULTS

The improved accuracies of both ANN and SVM are shown below table hence these two algorithms can be used to recognize gender whether male or female.

TABLE 3. After Tuning

Model	Training dataset	Testing dataset
ANN	98.27	98.611
SVM	99.7	99.8

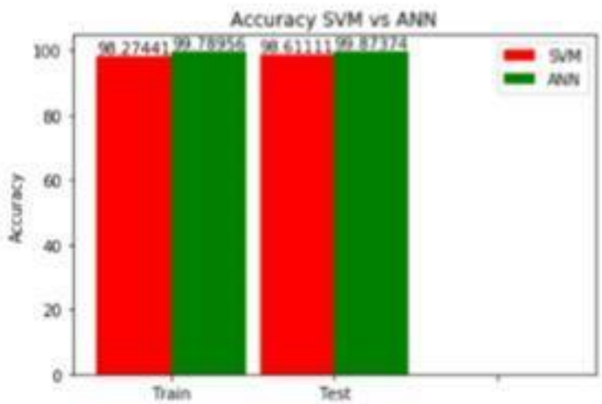


Fig. 13 ANN and SVM

7. FUTURE SCOPE

Through this research we can use this project in detecting the criminal gender when voice is recorded but image cannot be seen, used for recognize the emotion of a person , used to classify the gender for audio samples and so on.

8. CONCLUSION

Support-vector machines and neural networks are performing better on voice dataset. Parameter Tuning is giving the 98.6% accuracy with SVM and 99.87% with ANN. Hence we can conclude that ANN is the best way of recognizing gender through voice based and using acoustic properties.

9. REFERENCES

[1] S. M. R. Azghadi, M. R. Bonyadi, and H. Sliahhosseini, “Gender classification based on feed forward back propagation neural network,” in Artificial Intelligence and Innovations 2007: From Theory to Applications: Proceedings of the 4th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2007), C. Boukis, L.Pnevmatikakis, and L. Polymenakos, Eds., vol. 247 of IFIP The International Federation for Information Processing, pp. 299–304, Springer, Berlin, Germany, 2007.

[2] K.-H. Lee, S.-I. Kang, D.-H. Kim et al., “A Support Vector Machine- Based Gender Identification Using Speech Signal,” IEICE Transactions on Communications, pp. 3326-3329, 2008.

[3] J. Silovsky, and J. Nouza, “Speech, Speaker and Speaker's Gender Identification in Automatically Processed Broadcast Stream,”

[4] Cortes, Corinna & Vapnik, VN. (1995). Support Vector Networks. Machine Learning. 20. 273-297. 10.1023/A: 1022627411411.

[5] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, CRC Press, 1984.

[6] S. Gaikwad, B. Gawali, and S. C. Mehrotra, “Gender identification using SVM with combination of MFCC,” Advances in Computational Research, vol. 4, no. 1, pp. 69–73, 2012.

Vol.7 Issue 11, Pg.: 1-11
November 2019

AUTHOR PROFILE:

Buddavarapu Satya Prasad, currently pursuing B.Tech in the field of Information Technology at Vardhaman College of Engineering, Hyderabad, has published papers till now in various journals.

Email: prasadsatya952@gmail.com