INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

A Comparative Study of Machine Learning Models for Gender Recognition from Voice Samples

Manasi Manoj Sukale*, Pradip Ravindra Jagdale

Department of Statistics, Dr. D. Y. Patil Arts, Commerce, & Science College Pimpri, Pune-411018, Maharashtra, India

***Corresponding Author**

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP038

Received: 26 June 2025; Accepted: 30 June 2025; Published: 25 October 2025

Abstract: Voice recognition for gender has come a prominent area of study in machine literacy and speech processing. Dimorphism, or the clear physiological and aural distinctions between man and woman voices, is a point of mortal voices that allows automated systems to determine gender grounded on oral traits like pitch, frequency, accentuation, and speech rate. This study investigates how aural features taken from recorded speech can be used to classify gender using machine literacy algorithms. The delicacy and effectiveness of several bracket algorithms are compared through perpetration and evaluation. According to the analysis, woman voices have slightly advanced frequentness than man voices. Mean frequency of man and woman voice is thick between 0.15- 0.20.

Keywords: Data mining Classifiers: logistic regression, Decision tree, SVM, ANN, Naive Bayesian classifier, Python.

I. Introduction

Gender identity is one of the biggest problems today. Detecting gender from acoustic data, ie medium, frequency, etc. Machine learning provides promising results for classification problems in all research domains. There are several performance matrix to evaluate the algorithm of a region. Our comparative model algorithm to evaluate different machine learning algorithms from acoustic data. Dimorphism is the characteristic of the voice that is strongly observed in humans. Intonation, speech speed and duration are some properties that separate human voices, mainly male and female voices. Voting datasets are converted to various parameters such as vocal power, pitch, frequency, Q21, Q25, etc., they are then trained and tested with different algorithms to predict gender based on algorithms. In the real world, it is possible for a person to confirm a person through voice. The voice is filled with many linguistic properties. These voting features are considered a voice print to recognize gender on the speaker. The registered voice is considered the entrance to the system, which is then a system process to achieve voice functions. Check the input and compare it with a trained model, calculated based on the algorithm used and provides the latest matching output. Gender recognition can be used with various other applications. There are some: male to feel sad, female anger, etc., separate sound and video using tags, helping individual assistants answer questions with gender -specific results, etc. The applications include effective advertising and marketing strategies in Customer Relations Management (CRM) system

II. Literature Review

Harb and Chen (2005) showed one of the earliest research studies in the area of voice-based gender recognition, particularly in multimedia systems. They sought to design a system for the identification of a speaker's gender based on acoustic signals alone, which has direct applications in personalized content delivery, surveillance, and human-computer interaction.23

Max Kuhn and Kjell Johnson (2013) Applied Predictive Modeling is a core book that gives a complete picture of the end-to-end process of developing predictive models. Although the book doesn't concentrate on gender identification by voice, it gives crucial methodologies and techniques directly implementable in such tasks, especially in feature selection, model evaluation, and classification algorithms.

Ahmad and Tariq (2020) present an elaborate overview of voice-based gender classification approaches and applications. They compare both traditional and contemporary approaches, highlighted featureextraction methods such as MFCC, pitch, formants, spectral and prosodic features, and classifiers such as SVM, KNN, Naïve Bayes, decision trees, and neural networks. They point out an important trend: hybrid feature sets with ensemble or deep models result in best accuracy (~95%+)

III. Methodology

This database was designed to identify a voice as a man or woman based on the acoustic properties of voice and speech. The dataset consists of 3,168 registered voice samples, collected from male and female speakers. The voice samples are pre-processed by acoustic analysis in R using the seewave and tuneR packages, with an analyzed frequency range of 0hz-280hz (human vocal range). Missing values has been checked and removed if any available by using descriptive analysis. Graphical presentation used for comparison between male and female voice frequency, comparison between male and female mean dominant voice frequency, Distribution of Variable, Density plot for mean frequency**,** outlier detection. Heat map was used to investigate the dependence between multiple variables at the same time. The resultant Graph contains the correlation coefficients between each variable. For variable selection, the Gini Index is calculated by subtracting the sum of the squared probabilities of each class from one. After data

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

pre-processing the logistic regression, support vector machine, K Nearest Neighbour and naïve Bayesian classifier algorithms, Artificial Neural Network, Decision tree was used.

Descriptive Analysis:

Table 1: Descriptive Analysis of data

Sr. No	Variable	Type	Min	Mean	median	max	Missing value
1	Meanfreq	Numeric	0.0394	0.1809	0.1848	0.2511	0
2	Sd	Numeric	0.0184	0.0571	0.0592	0.1153	0
3	Median	Numeric	0.0110	0.1856	0.1900	0.2612	0
4	Q25	Numeric	0.0002	0.1405	0.1403	0.24730.	0
5	Q75	Numeric	0.0429	0.2248	0.2257	0.2735	0
6	IQR	Numeric	0.0146	0.0843	0.0946	0.2522	0
7	Skew	Numeric	0.1417	3.1402	2.1971	34.7255	0
8	Kurt	Numeric	2.0685	36.5685	8.3185	1309.612	0
9	sp.ent	Numeric	0.7387	0.8951	0.9018	0.9820	0
10	Sfm	Numeric	0.0369	0.4082	0.3963	0.8429	0
11	Mode	Numeric	0.0000	0.1653	0.1866	0.2800	0
12	Centroid	Numeric	0.0394	0.1809	0.1848	0.2511	0
13	Meanfun	Numeric	0.0556	0.1428	0.1405	0.2376	0
14	Minfun	Numeric	0.0098	0.0368	0.0461	0.2041	0
15	Maxfun	Numeric	0.1031	0.2588	0.2712	0.2791	0
16	Meandom	Numeric	0.0078	0.8292	0.7658	2.9577	0
17	Mindom	Numeric	0.0049	0.0526	0.0234	0.4590	0
18	Maxdom	Numeric	0.0078	5.0473	4.9922	21.8672	0
19	Dfrange	Numeric	0.0000	4.9946	4.9453	21.8438	0
20	Modindx	Numeric	0.0000	0.1738	0.1394	0.9324	0
21	Lable	Factor	0.0000	0.0000	0.0000	0.0000	0

No missing value is present in the data

Graphical Presentation: Comparison Between Male and Female Voice Frequency

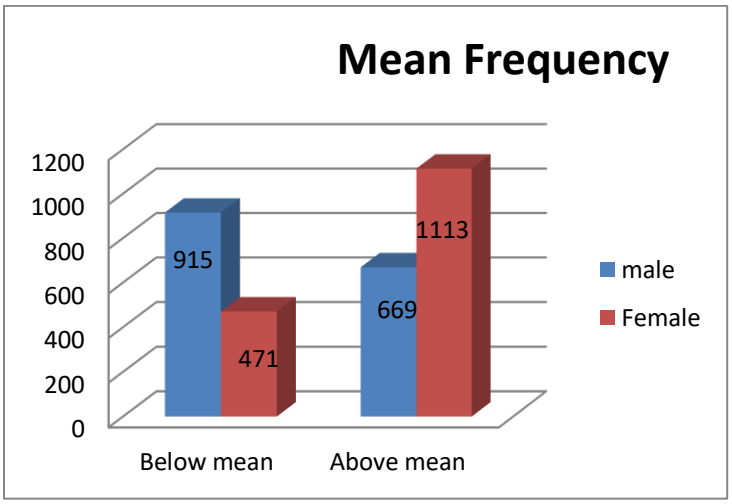


Fig. 1: Comparison between male and female voice frequency

The voice frequency of female is slightly greater than male.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Comparison Between Male and Female Mean Dominant Voice Frequency

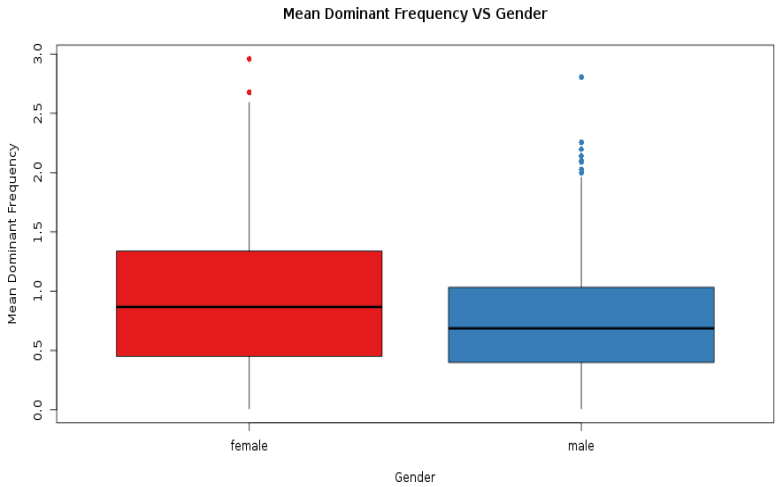


Fig. 2: Comparison between male and female mean dominant voice frequency

The average mean dominant frequency of female is greater than male frequency.

Distribution of Variable:

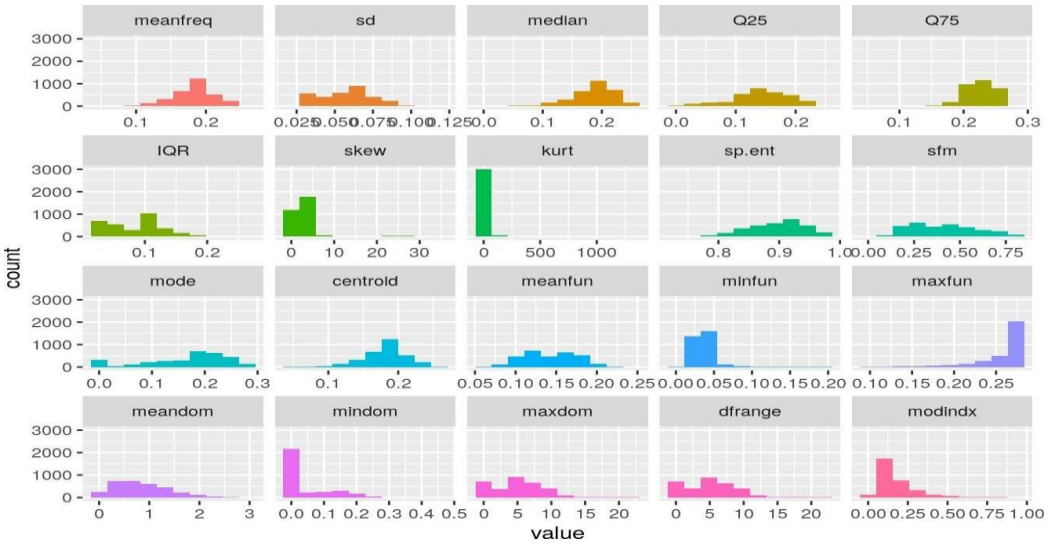


Fig. 3 Distribution of variable

From above, no variable has normal distribution.

Density plot for mean frequency:

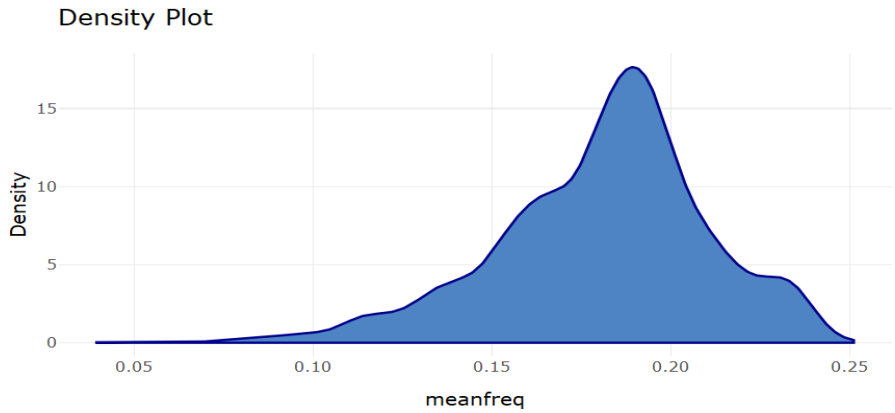


Fig. 4: Density plot

Mean frequency of male and female voice is dense between 0.15-0.20.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Outlier Detection:

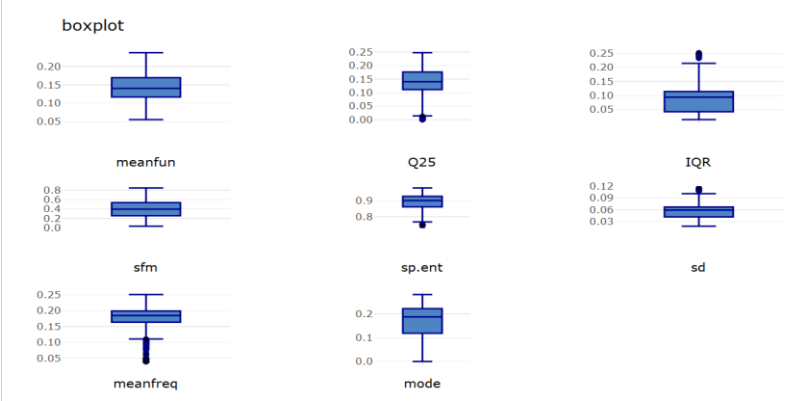


Fig. 5: Outlier Detection

From above boxplots, only variable which have outliers is the mean frequency.

Heatmap: It is used to investigate the dependence between multiple variables at the same time. The resultant Graph contains the correlation coefficients between each variable.

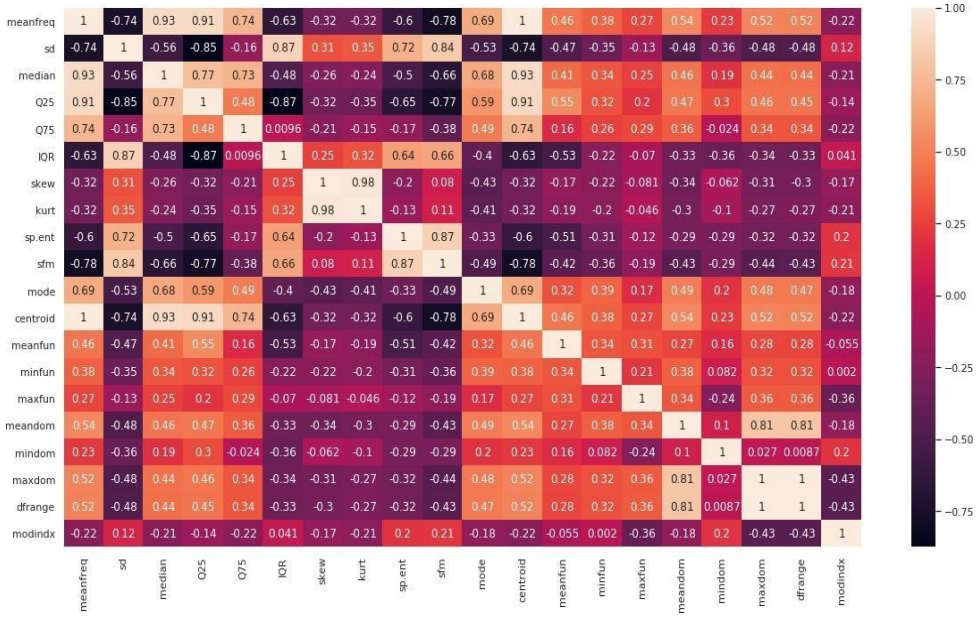
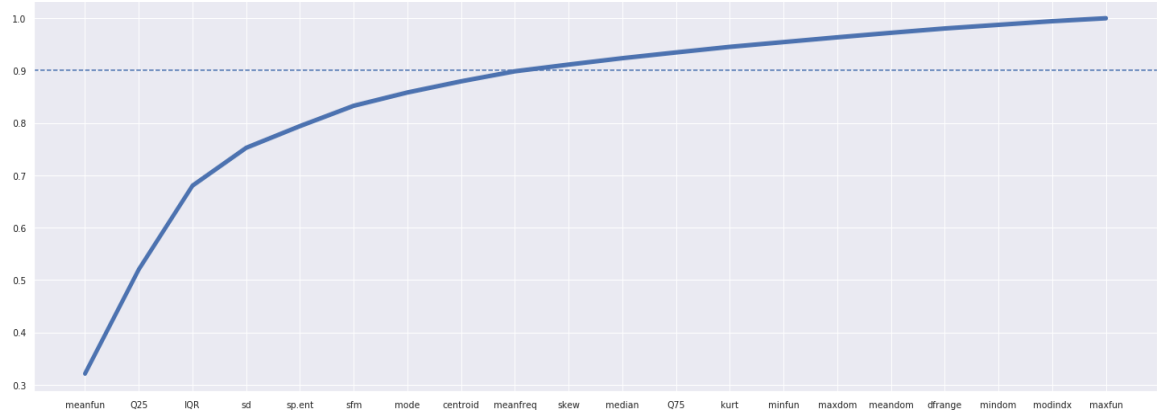


Fig. 6: Heatmap

The values of correlation coefficients are grater for some variables. There is dependency in variables. Hence select those variables which gives more accuracy.

Variable Selection: Gini index has used as variable selection method. The Gini Index is calculated by subtracting the sum of the squared probabilities of each class from one.



ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

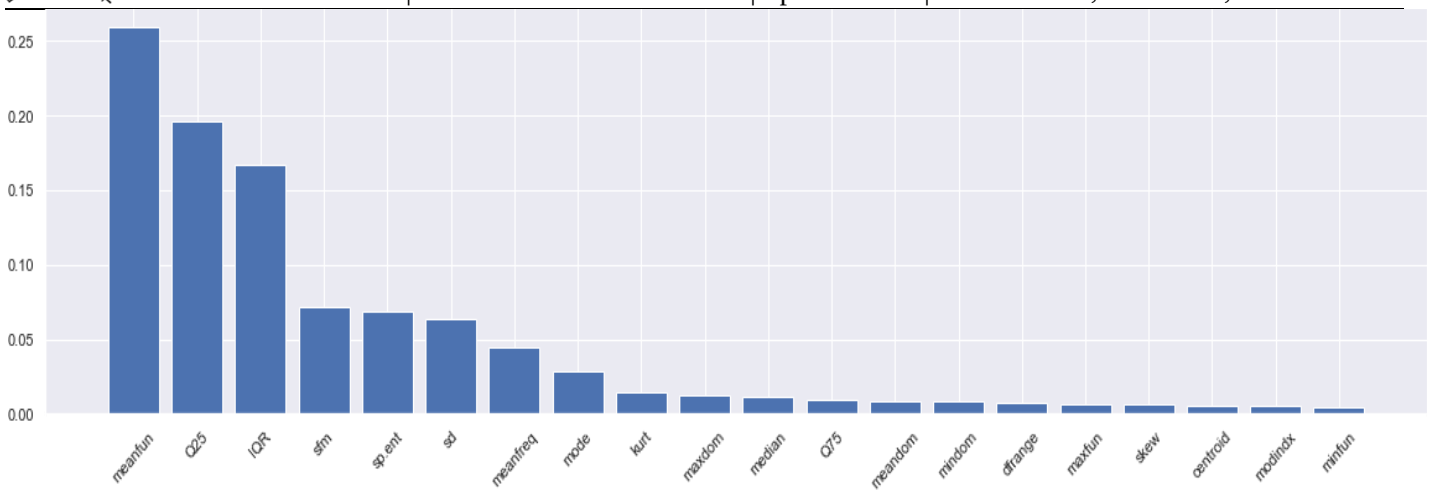


Fig. 7: Variable Selection

Here 12 features out of total 21 features having considerable importance. Cumulative feature importance shows that 9 out of total 21 features cover 90% of feature importance.

Model building: The methodology and performance of every model is as follows:

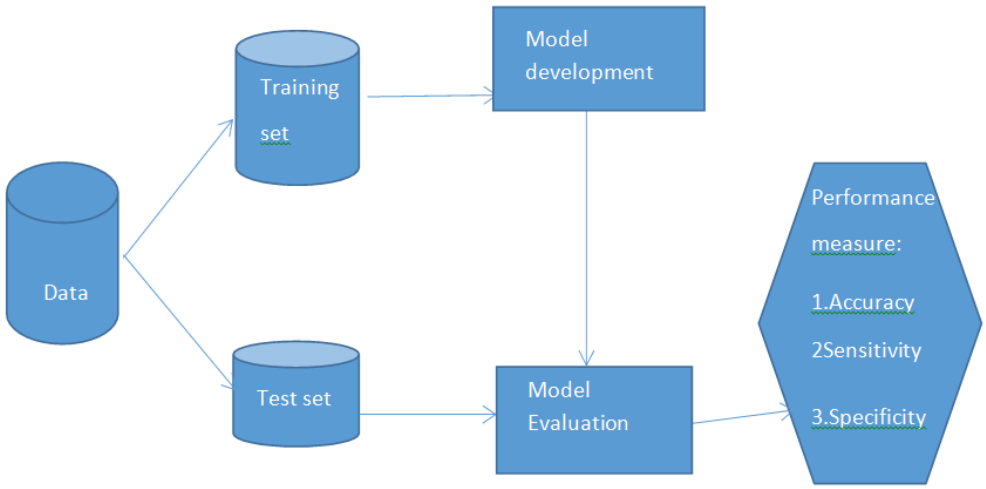


Fig. 8: Model building

1) **Logistic Regression:** Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is binary. It computes the probability of an event occurrence. We selected this model for prediction purpose i.e. identify gender.

Using R software results are as follows:

Now after variable selection the data is devoided into training and testing data. The training data is used to build the model and testing data for prediction.

Confusion Matrix:

	Predicted	
Actual	Female	Male
Female	302	14
Male	6	310

2) **Decision Tree:** A Decision Tree is a simple representation of classifying examples. It is a Supervised Machine Learning method. It provides a graphical representation of how different factors influence on target. It breaks down a dataset into smaller and smaller subsets. In the literature there are so many popular decision algorithms like ID3, C4.5, C5 etc. As the name implies these techniques recursively separates observations in branches who construct a tree for the purpose of improving the prediction accuracy.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

	Predicted	
Actual	Female	Male
Female	304	10
Male	12	306

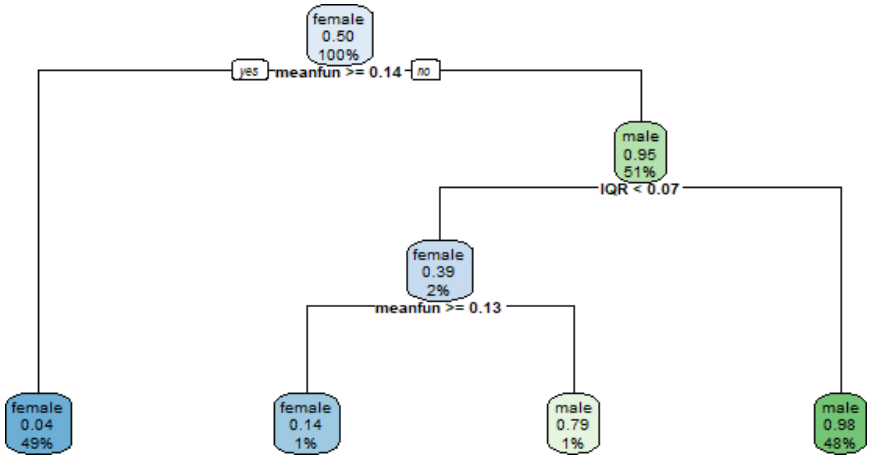


Fig. 9: Decision Tree

3) **Artificial Neural Network:** Artificial Neural Network is commonly known as biological inspired, sophisticated analytical technique, capable of modelling extremely complex non-linear function. We used a popular ANN architecture called multilayer perception (MLP) with black proposition. It is most commonly used and well-studied ANN architecture.

By using R software, model and corresponding result are as follows,

	Predicted	
Actual	Female	Male
Female	297	8
Male	19	308

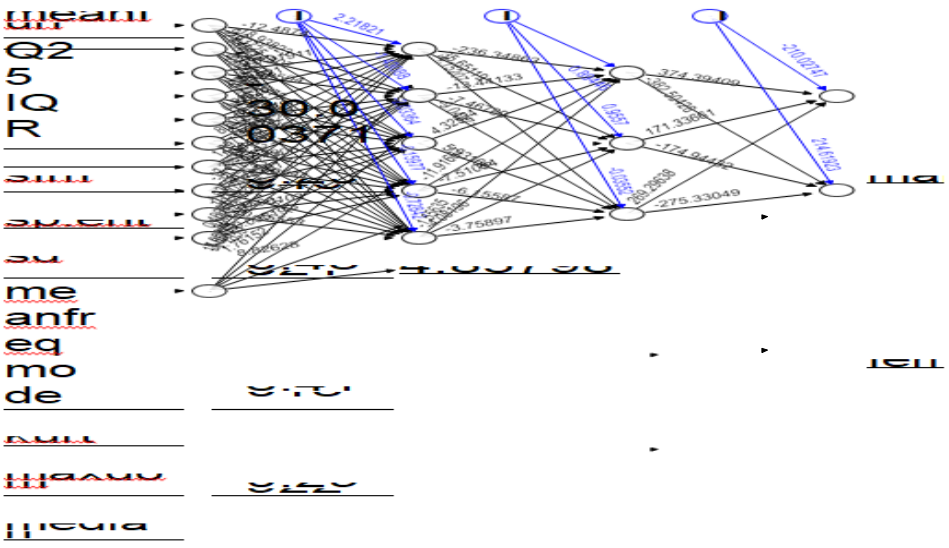


Fig. 10: ANN

4) **Support Vector Machine:** Support Vector Machine (SVM) is a supervised machine learning algorithm. It is used for classification and regression tasks. It tries to find the best boundary known as hyperplane that separates different classes in the data. It is useful when you want to do binary classification.

By using Python software, the model and corresponding result are as follows,

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

	Predicted	
Actual	Female	Male
Female	333	4
Male	11	286

5) **Naïve Bayesian Classifier:** The Naïve Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given dataset. The algorithm uses Bayes theorem and assumes all attributes to be independent given the value of class variable. Naïve Bayes classifier is based on Bayes theorem and the theorem of total probability. In this classifier we compute the conditional probability P(Cj/X) and assign X to those class Ci having large probability

i.e. $X \epsilon Cj$ if $P(Cj/X)>P(Ci/X)$ for all $i \neq j=1, 2, \ldots, m$

By using Python software, the model and corresponding result are as follows,

	Predicted	
Actual	Female	Male
Female	298	39
Male	38	259

Comparison of Models:

Sr. No.	Model	Accuracy	sensitivity	Specificity
1	Logistic Regression	0.9683	0.9810	0.9556
2	Decision Tree	0.9651	0.9622	0.9681
3	ANN	0.9572	0.9418	0.9737
4	SVM	0.9763	0.9629	0.9881
5	Naïve Bayes	0.8785	0.8720	0.8842

IV. Conclusion

The voice frequency of female is slightly greater than male. The distribution of male and female voice mean dominant frequency is near about same. No one variable have normal distribution. Mean frequency of male and female voice is dense between 0.15-0.20. The Support Vector Machine i.e. SVM model has the highest accuracy. Logistic regression has highest sensitivity.

V. Future Scope

This type of study can be done for the problems related with medical studies like predicting cancer survivability.

References

1. Harb, H., & Chen, L. (2005). Voice-based gender identification in multimedia applications. *Journal of Intelligent Information Systems*, *24*(2–3), 179–198.
2. Kuhn, M., & Johnson, K. (2013). *Applied predictive modelling*. Springer.
3. Nair, R. R., & Vijayan, B. (2019). Voice-based gender recognition. *International Research Journal of Engineering and Technology (IRJET)*, *6*(5), 2109–2112.
4. Tiwari, V., & Pandey, V. (2021). Gender classification using voice features and machine learning techniques. *International Journal of Computer Applications*, *174*(3), 28–34.
5. Venkatesh, S., & Patra, M. (2018). Gender identification from voice using spectral and prosodic features. *International Journal of Speech Technology*, *21*(2), 231–239.
6. Bocklet, J., & Cordier, J. (2019). Deep learning-based gender classification using speech signals. *Journal of Computer Science and Technology*, *34*(6), 1183–1192.
7. Khan, S., & Khan, S. (2018). Gender detection using voice: A hybrid approach based on MFCC and LPC. *International Journal of Artificial Intelligence & Applications*, *9*(5), 12–19.
8. Ahmad, S., & Tariq, S. (2020). Gender classification using voice signals: A review of methods and applications. *International Journal of Speech Processing*, *35*(1), 11–23.
9. Kumar, S., & Rani, R. (2020). Gender classification using voice features and deep learning techniques. *IEEE Transactions on Audio, Speech, and Language Processing*, *28*, 1550–1558.