Gender Recognition by Voice using Machine Learning Techniques

Sweta Jain
Neha Pandey
Vaidehi Choudhari
Pratik Yawalkar
Amey Admane
Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Gender Recognition using voice is of enormous prominence in the near future technology as its uses could range from smart assistance robots to customer service sector and many more. Machine learning (ML) models play a vital role in achieving this task. Using the acoustic properties of voice, different ML models classify the gender as male and female. In this research we have used the ML models- Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), Gradient Boosting, K-Nearest Neighbor (KNN), and ensemble method (KNN, logistic regression, SVM). To propose which algorithm is best for recognizing gender, we have evaluated the models based on results achieved from analysis of accuracy, recall, F1 score, and precision.

Keywords: Gender recognition, Gradient Boosting, Ensemble, Machine Learning.

1. INTRODUCTION

In this research, we aim to identify the gender using voice by training the Kaggle dataset using different machine learning algorithms and suggested the preeminent machine learning algorithms to realize the work. The Machine learning algorithms plays a prodigious role in gender acknowledgement.ML models needs huge data corpus to develop and infer the accurate results. The results of the ML models are analyze using standard evaluation parameters. The model with the highest accuracy score will be the paramount model to accomplish the task of gender recognition using voice.

The future purview of our research ranges from being useful in the customer service sector to high technological advances. There are many applications where gender recognition can be useful. It is used in personal assistants, automatic salutations, to detect the gender of criminals using voice, and many other. In countries like Japan, China, and Korea there are fully automated stores with no human employees and many other countries are planning to build these stores in the future. So, in these automated shops predicting gender by voice will help to recommend products based on gender.

2. RELATED WORK

In the past subsequent years, researchers had been extensively using machine learning models in the field of gender identification by voice. These machine learning models make use of different acoustic properties of voice like pitch, frequency, kurtosis, peak frequency, and many other properties to predict gender. For gender recognition by voice Ioannis E. Livieris, Emmanuel Pintelas and Panagiotis Pintelas proposed an ensemble-based self-labeled algorithm built using the three most efficient self-labeled methods such as tri-training, co-training, and self-training by using the ensemble as a base learner, which is termed as iCST-Voting. For evaluating the performance, authors used the Kaggle dataset and the Deterding dataset. The highest accuracy obtained is 98.42%. (Livieris, Pintelas, and Pintelas, 2019)

In an experiential study by Mucahit Buyukyilmaz and Ali Osman Cibikdiken, authors have developed a multiple layer perceptron deep learning model to acknowledge gender by voice. On Kaggle dataset author attained an accuracy of 96.74%. (Büyükyılmaz and Çıbıkdiken, 2016)

For recognizing the gender using voice Zvarevashe et al. performed feature collection through the random forest recursive feature removal with gradient boosting machines. Excluding feature selection, the gradient boosting model reached an accuracy of 97.58% and including feature selection, it touched an accuracy of almost 100%. (Zvarevashe and Olugbara, 2018)

To accomplish the task of gender acknowledgement by voice, Harb and Chen used a set of Gaussian modeling features, pitch related features, and contrasting acoustic properties with dissimilar classifiers. They used a neural network for constructing the classifier. (Harb and Chen, 2007)

To recognise gender of humans a new simpler method is proposed by Mridula, Abhishek and Aditya. (Mridula Korde, 2022)

3. METHODOLOGY

3.1 DATASET

Our research aims to identify gender based on voice. For this, we have taken our dataset from Kaggle. It comprises of 3168 occurrences of recorded voice of which 1584 are males and 1584 are females. Figure1 represents the workflow of the research work for achieving the end goal of recognizing the gender using voice.

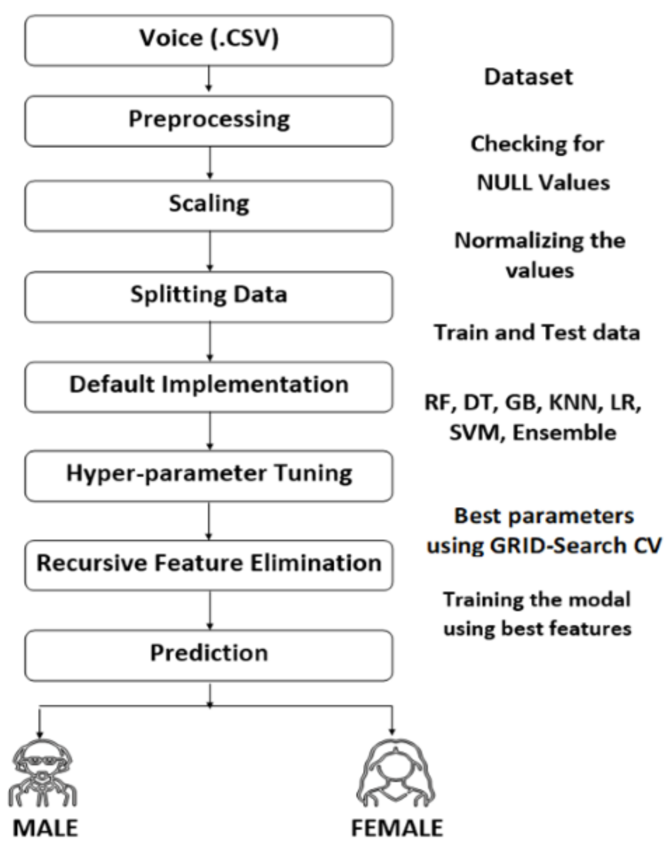

Figure 1: Workflow representation

3.2 PRE-PROCESSING OF DATASET

First, we have pre-processed dataset from Kaggle. This helps in assuring the maximum accuracy of the models which are later applied to this dataset.

—Checking for NULL values
—Checking for Numeric and Non-Numeric Features
—Checking for Dependent and Independent Features

3.3 SCALING

While training a few of the models it is necessary to do scaling. Machine learning algorithms only deal with numbers and have no knowledge about what these numbers represent. So, it may happen that numbers having higher values will be given more importance than other numbers, which can directly affect our output and degrade performance.

3.4 SPLITTING TRAINING AND TESTING DATA

Before training, for avoiding overfitting of data and to see how the model will perform with the test data we have split the corpus into training and testing corpus. The data corpus is split into two chunks in a 2:8 ratio i.e., 20% of the dataset for testing and 80% of the dataset for training.

3.5 DEFAULT MODEL

First, we have trained and tested the models using the default parameters of the algorithm using all 20 features. But this is not optimal and time-consuming and can also cause overfitting. Here, Hyperparameter tuning comes in handy as it helps in maximizing the performance of the model without causing the before stated issues.

3.6 HYPERPARAMETER TUNING

To find the best parameters, we have done hyperparameter tuning using GridSearchCV for all the models. Then we trained our model on the hyperparameters that we got. In some algorithms, we got better accuracy after doing hyperparameter tuning.

3.7 RECURSIVE FEATURE ELIMINATION

This RFE step help us to reduce the training time of the model without affecting the performance of the model. So, using RFE from sklearn.feature_selection we can achieve this. We have done the recursive feature elimination for all the models.

3.8 MACHINE LEARNING MODELS

We have trained our model with the following algorithms.

(1) Random Forest : RF is referred to as a collaborative machine learning algorithm. We first trained the model using default parameters and then we did hyperparameter tuning.
(2) Support Vector Machine : SVM can be used for solving both regression as well as categorization problems. In SVM there are many different types of kernel. Out of which we have implemented three kernels: linear, RBF, and polynomial.
(3) Gradient Boosting : Initially, we trained the model using the Gradient Boosting Classifier module of Python's Scikit Learn library. Then we applied the Grid Search CV method on n_estimators, learning_rate and max_depth parameters for a bunch of values and found out the best values for the model to be able to give the most accurate results. Next, we found out the most related features using recursive feature elimination.
(4) Decision Tree : ID3 algorithm is implemented for the categorization of male and female occurrences. It works by choosing the features with maximum information gain. This algorithm is a simple and powerful learning model for categorization of the data.

(5) K-Nearest Neighbors : First, we implemented the KNN algorithm on default parameters using the KNeighbor Classifier module from scikit learn library. After that Grid Search CV was performed.

(6) Logistic Regression : We have implemented logistic regression using the sklearn linear model. There are different parameters in logistic regression like solvers, penalty, C, max_iter, and many others.

(7) Ensemble : This is an ensemble model which will be using three different classifiers KNN, SVM, and Logistic regression to predict the final output. This will help us to improve the accuracy. All these three classifiers will be predicting the output, then using mode final output will be predicted.

4. RESULTS AND DISCUSSIONS

The results obtained from all the models before and after applying hyperparameter tuning are summarized in Table 1 as shown below.

Models		**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC**
Decision Tree	Before	95.11	95.43	94.51	94.97	95.09
Decision Tree	After	96.05	97.65	94.19	95.89	96.01
Random Forest	Before	97.47	98.03	96.77	97.40	97.46
Random Forest	After	96.68	97.37	95.80	96.58	96.66
Gradient Boosting	Before	97.00	97.39	96.45	96.92	96.99
Gradient Boosting	After	97.31	97.71	96.77	97.24	97.30
Support Vector Machine	Before	98.26	98.51	98.21	98.36	98.26
Support Vector Machine	After	97.79	97.64	98.21	97.92	97.76
K Nearest Neighbours	Before	97.94	97.92	98.24	98.07	97.93
K Nearest Neighbours	After	98.10	98.21	98.21	98.35	97.93
Logistic Regression	Before	98.10	98.50	97.92	98.21	98.11
Logistic Regression	After	98.10	98.50	97.92	98.21	98.13
Ensemble Model	Before	98.73	99.10	98.51	98.80	98.75
Ensemble Model	After	NA	NA	NA	NA	NA

Table I: Result of Hyperparameter Tuning

For some models, results were better after performing hyperparameter tuning and for others default, algorithm results were better. After performing recursive elimination, we were able to optimize the time required to train the model as we were using the best 10 features to train the model instead of using all 20 features from the dataset.

Table 2. Elucidate the values for accuracy, precision, f1-score and recall of all the ML models after performing GridSearchCv and recursive feature elimination.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Decision Tree	95.542	96.989	93.548	95.238
Random Forest	97.476	98.039	96.774	97.402
Gradient Boosting	97.318	97.719	96.774	97.244
Logistic Regression	98.107	98.507	97.922	98.214
Ensemble	98.738	98.507	98.516	98.809
K Nearest Neighbours	98.107	98.219	98.219	98.219
Support Vector Machine	98.264	98.511	98.219	98.365

Table II: Result of all Models

From Table 2, we can witness that the overall performance of the ensemble algorithm built using SVM, Logistic Regression, and KNN is preeminent followed by SVM. Generally, Ensemble algorithms perform better than other models as they are built using a number of classifiers but in our case SVM performed better than random forest and gradient boosting may be because of the use of weak classifiers.

Using the confusion matrix, we investigated the performance of different classification models. Table3 depicts outcomes from the confusion matrix.

Model	True Negative	False Negative	True Positive	False Positive
Decision Tree	315	20	290	9
Random Forest	318	10	300	6
Gradient Boosting	317	10	300	7
Logistic Regression	292	7	330	5
Ensemble	294	5	332	3
K Nearest Neighbours	290	6	331	7
Support Vector Machine	293	7	330	4

Table III: Confusion matrix summary for all models

From Table 3, it can be perceived that the Ensemble algorithm built using KNN, SVM and logistic regression performed the preeminent as it shows minimum number of misclassifications. The decision tree model depicts highest number of misclassification. SVM classifier also performed well with little number of misclassification. Figure2 presents a comparative analysis of various machine learning models developed in our research study.

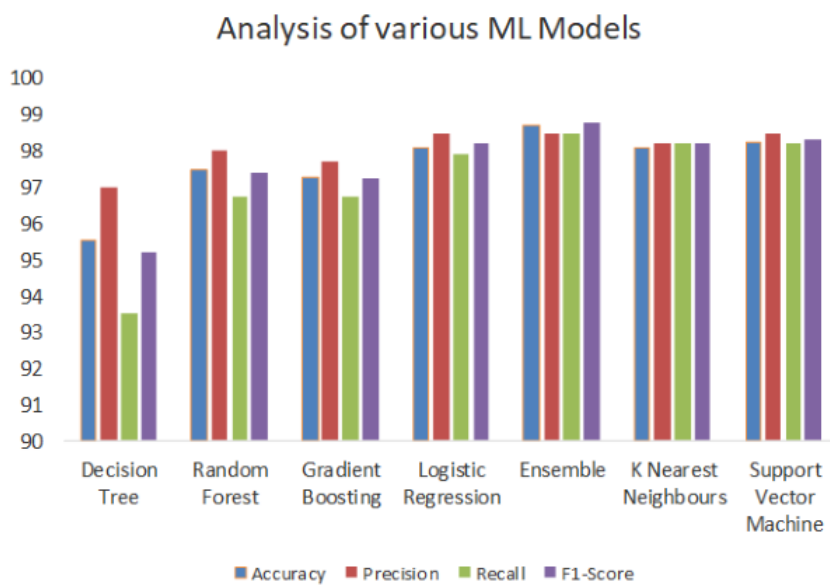

Figure 2: Analysis of various ML Models

5. CONCLUSION

The presented research work elucidate a system for gender prediction by voice using various machine learning models. We successfully build the model using the voice dataset on different machine learning algorithms and also performed GridSearchCV for hyperparameter tuning. We found that the overall performance of ensemble model build using KNN, SVM, and logistic regression were finest both in terms of time and accuracy. It gave us the uppermost accuracy of 98.731%. Gender prediction by voice is still a challenging area of research as real time voice may suffers from noise, voice quality, and pre-processing of real time input recordings. Real time voice based gender recognition can open numerous doors of application in criminal gender investigation and product recommendation.

References

Büyükyilmaz, M. and Çibikdiken, A. 2016. Voice gender recognition using deep learning. In *Conference: Modeling, Simulation and Optimization Technologies and Applications.*

Harb, H. and Chen, L. 2007. A general audio classifier based on human perception motivated model. *Multimedia Tools Appl. 34,* 3 (09), 375–395.

Livieris, I., Pintelas, E., and Pintelas, P. 2019. Gender recognition by voice using an improved self-labeled algorithm. *Machine Learning and Knowledge Extraction Vol.1*, 492–503.

Mridula Korde, Abhishek Joshi, A. S. 2022. Human recognition model using cnn. *International Journal of Next-Generation Computing (IJNGC) 13*.

Zvarevashe, K. and Olugbara, O. O. 2018. Gender voice recognition using random forest recursive feature elimination with gradient boosting machines. *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, 1–6.

Prof. Sweta V. Jain received the Masters in Technology in Computer Science and Engineering from Nagpur University in 2009 as a first merit holder and currently pursuing Ph.D. in biomedical image processing. She is an Assistant professor in Computer Science and Engineering department at Shri Ramdeobaba College of Engineering and Management Nagpur. Her research interests include Pattern Recognition, Digital Image Processing and Machine Learning.
E-mail: jains@rknec.edu.

Ms Neha Pandey is a Student at Shri Ramdeobaba College of Engineering and Management,Nagpur.Currently pursuing bachelor's degree in computer science.She is very fascinated about the field of machine learning, artificial intelligence and all the emerging new technologies.
E-mail: pandeynj@rknec.edu.

Ms Vaidehi Choudhari is a Student at Shri Ramdeobaba College of Engineering and Management,Nagpur.Currently pursuing bachelor's degree in computer science along side being enrolled in honors program with an aspiration to instigate in the field of machine learning and data science which are the most prominent technology to be used in future. She has completed entire schooling from nagpur and look forward to make a break through in the world of tech. Her area of interests include machine learning, big data, cloud computing and all the emerging new technologies.
E-mail: choudharivd@rknec.edu.

Mr Pratik Yawalkar Bachelor of Engineering (B.E.) Graduate in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management. He is a final year student having avidity towards Cloud Computing, Artificial Intelligence and Cyber Security. His area of research are Machine Learning and Digital Forensics. He has published a research paper on "Artificial Intelligence : New Backbone to Education" during his academic career.
E-mail: yawalkarpy@rknec.edu.

Mr Amey Kishor Admane, Bachelor of Engineering (B.E.) Graduate in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management. He is a final year student with an interest in field of Artificial intelligence, Machine Learning and Cyber security .
E-mail: admaneak@rknec.edu.