Mini Project Report on

Gender Recognition by Voice Analysis

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Shashank Kothari **University Roll No.2017020**

Under the Mentorship of
Dr. Vishan Gupta
Professor

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand
January 2023

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"Gender Recognition By Voice Analysis"** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Vishon Gupta, Professor,** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

SHASHANK KOTHARI 2017020

Table of Contents

Chapter No.	Description	Page No.
Chapter 1	Introduction	**1-2**
Chapter 2	Literature Survey	**3-4**
Chapter 3	Methodology	**5-7**
Chapter 4	Result and Discussion	**8-9**
Chapter 5	Conclusion and Future Work	**10**
	References	**11**

Chapter 1

Introduction

In the following sections, a brief introduction and the problem statement for the work has been included.

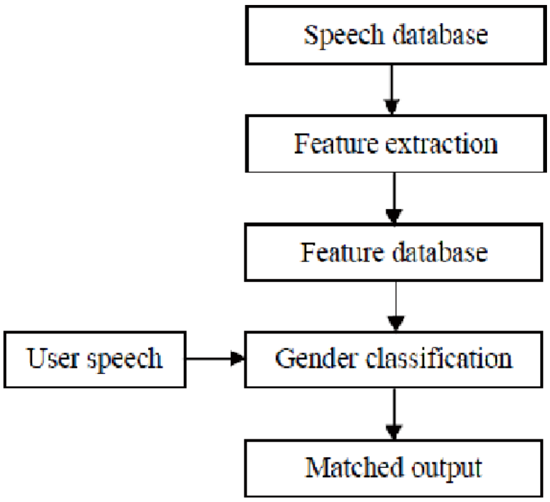
1.1 Introduction

The voice of human speech is an effective communication method consisting of unique semantic linguistic and paralinguistic features such as gender, age, language, accent, and emotional state. The sound waves consisting of human voice are unique among all creatures producing sound since every single wave carries a different frequency. Identifying human gender based on voice has been a challenging task for voice and sound analysts who deploy numerous applications.

Gender recognition is a technique which is often utilized to determine the gender category of a speaker by processing speech signals. Speech signals taken from a recorded speech can be used to acquire acoustic attributes such as duration, intensity, frequency and filtering . Some applications where gender recognition can be useful are speech emotion recognition, human to machine interaction, sorting of telephone calls by gender categorization, automatic salutations, muting sounds for a gender and audio/video categorization.

There are a set of features used for recognizing the voice gender. Among the most common features utilized for voice gender recognition are Mel-scaled power spectrogram (Mel), Mel-frequency cepstral coefficients (MFCCs), power spectrogram chroma (Chroma), spectral contrast (Contrast), and tonal centroid features . By getting the extracted features combined with the gender label as a form of a training set, ML techniques are used to build a high-quality model for recognizing the voice gender. In particular, each classification technique is used to build a set of hypothesis models and selects the most optimal one. This model classifies the unknown voice label by receiving the voice features and categorizing the voice gender and shows accuracy of model.

Speech recognition has various applications including human to machine interaction, sorting of telephone calls by gender categorization, video categorization with tagging and so on. Currently, machine learning is a popular trend which has been widely utilized in various fields and applications, exploiting the recent development in digital technologies and the advantage of storage capabilities from electronic media. Recently, research focuses on the combination of ensemble learning techniques with the semi-supervised learning framework aiming to build more accurate classifiers. In this , we focus on gender recognition by voice utilizing various ensemble semi-supervised self-labeled algorithm and test accuracy of these algorithms.



Chapter 2

Literature Survey

During the last decades, machine learning models and data mining techniques have been widely utilized for gender recognition by voice. These prediction models can identify the gender of a person by utilizing various features such as the length of the vocal folds, gait and speech. More specifically, the acoustic properties acquired from voice and speech signals like duration, intensity and frequency can be used as features to recognize the gender of the speaker. A number of studies have been carried out in recent years; some useful outcomes of them are briefly presented below.

Maka, T.; Dziurzanski, used 630 speakers, 438 males and 192 females in their experiments for the gender identification problem in different acoustical environments (indoor and outdoor auditory scenes). In addition, for the evaluation stage each sentence has been mixed with several types of background noise In their results, they found out that non-linear smoothing increases the classification accuracy by 2% and the recognition accuracy obtained was 99.4%

Pˇribil, J.; Pˇribilová, A.; Matoušek, J. proposed a two level Gaussian Mixture Model (GMM) algorithm to recognize age and gender. Their proposed classifier was first verified for detection of four age categories (child, young, adult, senior) and for recognizing the gender for all but children's voices in Czech and Slovak languages. The prediction accuracy on gender identification was above 90%. In a similar work, Pˇribil, J.; Pˇribilová, A.; Matoušek, J.]developed a two level GMM classifier to detect age and gender. The classification accuracy achieved on gender recognition was 97.5%. Furthermore, the obtained gender and age classification accuracy results were compared with the results achieved by the conventional listening test which is an evaluation method of the quality of the synthetic speech .Yukiyamas et al. , utilized a multilayer perceptron deep learning model using the acoustic properties of the voices and speech to identify the voice gender. The dataset they utilized for their experiments consisted of 3168 recorded samples of human voices. Their classification model managed to achieve 96.74% accuracy. Additionally, they have designed a web page to detect the gender of voice by utilizing the obtained model. Zvarevashe et al. Zvarevashe, K.; Olugbara , proposed a gender voice recognition technique utilizing feature selection through the random forest recursive feature elimination with gradient boosting machines (GBMs)algorithm for gender classification. Acoustic features were collected from a public gender voice dataset including 1584 males and 1584 females. The GBMs algorithm had obtained an accuracy of 97.58% without feature selection while by applying feature selection it almost achieved 100%.

Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A.; Frà, C.; Valla, M. developed an android Speech Processing Platform as smartphone Application (SPECTRA) for gender, speaker and language recognition by utilizing multiple unsupervised support vector machine classifiers. An interesting and innovative point in this work is the dynamic training with the features extracted from every user, having SPECTRA installed on his personal android smartphone. This can lead on building more robust classifiers with higher classification accuracy, resulting in better recognition performances. Pahwa et al. proposed a recognition system to determine the gender using speech samples of 46 speakers. In particular, they extracted one of the most dominant and most researched speech feature, Mel coefficients and the first and second order derivatives. Their proposed model consists of a support vector machine and neural network classifier using a stacking methodology. The classification accuracy obtained from their numerical experiments was 93.48%.

Chapter 3

Methodology

The tools and algorithms used in this project are:-

1.Python

Python is a widely used general-purpose, high level programming language. It was created by Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently.

2. Jupyter Notebook

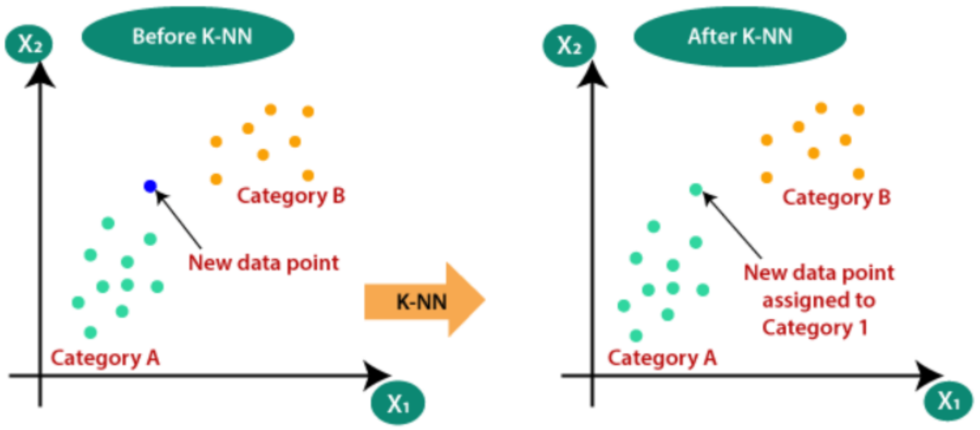
The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter. Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

3. Naive Bayes

Naive Bayes is a classification algorithm that is suitable for binary and multiclass classification. It is a supervised classification technique used to classify future objects by assigning class labels to instances/records using conditional probability. In supervised classification, training data is already labeled with a class. For example, if fraudulent transactions are already flagged in transactional data and if we want to classify future transactions into fraudulent/non-fraudulent, then that type of classification would be called supervised. The Naive Bayes classifier assumes that every feature/predictor is independent, which is not always the case, so it is important to understand the type of data you are analyzing before choosing this or any other analytical technique

4. K-Nearest Neighbour

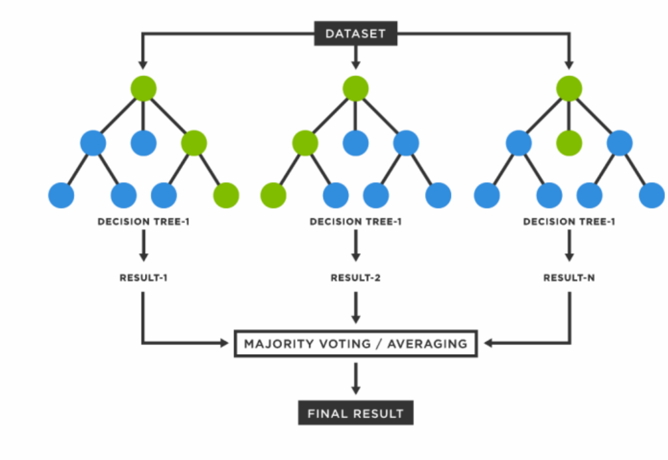
K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.



4. Random Forest

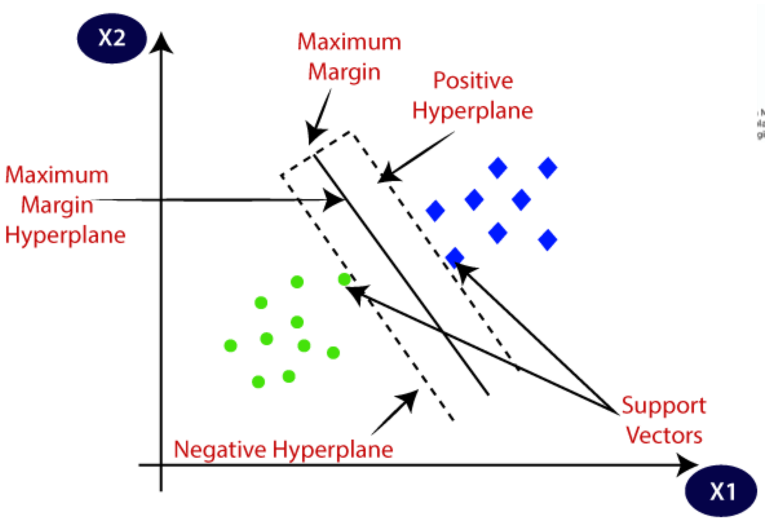
Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of *combining* multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset**.**" Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



5. Support vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.



Chapter 4

Result and Discussion

Classification Score

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
rand_forest = DecisionTreeClassifier(random_state=50)
rand_forest.fit(x_train , y_train)

print('Random Forest Classifier Score :  ',rand_forest.score(x_test,y_test))

algo_names.append("Random Forest")
algo_scores.append(rand_forest.score(x_test,y_test))
p2=rand_forest.predict(x_test)
```

```
Random Forest Classifier Score :    0.9511041009463722
```

SVM

```
from sklearn.svm import SVC
svm = SVC(random_state=50)
svm.fit(x_train , y_train)

print('SVM Classifier Score :  ',format(svm.score(x_test,y_test)))

algo_names.append("SVM")
algo_scores.append(svm.score(x_test,y_test))
p3=svm.predict(x_test)
```

```
SVM Classifier Score :    0.6750788643533123
```

Naive Bayes

```
from sklearn.naive_bayes import GaussianNB
naive_bayes = GaussianNB()
naive_bayes.fit(x_train , y_train)

print('Naive Bayes Classifier Score :  ',format(naive_bayes.score(x_test,y_test)))

algo_names.append("Naive Bayes")
algo_scores.append(naive_bayes.score(x_test,y_test))
p4=naive_bayes.predict(x_test)
```

```
Naive Bayes Classifier Score :    0.8943217665615142
```

K-Nearest Neighbour

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(x_train , y_train)

print('KNN Classifier Score :  ',format(knn.score(x_test,y_test)))

algo_names.append("KNN")
algo_scores.append(knn.score(x_test,y_test))
p5=knn.predict(x_test)
```

```
KNN Classifier Score :    0.7208201892744479
```

Classification Report

Random forest Classification Report

```
report_ran_forest=classification_report(y_test,p2)
print(report_ran_forest)
```

```
              precision    recall  f1-score   support

           0       0.95      0.95      0.95       310
           1       0.95      0.95      0.95       324

    accuracy                           0.95       634
   macro avg       0.95      0.95      0.95       634
weighted avg       0.95      0.95      0.95       634
```

Svm Classification Report

```
report_svm=classification_report(y_test,p3)
print(report_svm)
```

```
              precision    recall  f1-score   support

           0       0.63      0.83      0.71       310
           1       0.77      0.52      0.62       324

    accuracy                           0.68       634
   macro avg       0.70      0.68      0.67       634
weighted avg       0.70      0.68      0.67       634
```

Naive bayes

```
report_nav_bayes=classification_report(y_test,p4)
print(report_nav_bayes)
```

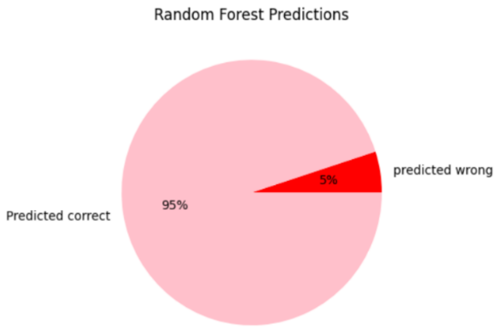
```
              precision    recall  f1-score   support

           0       0.89      0.90      0.89       310
           1       0.90      0.89      0.90       324

    accuracy                           0.89       634
   macro avg       0.89      0.89      0.89       634
weighted avg       0.89      0.89      0.89       634
```

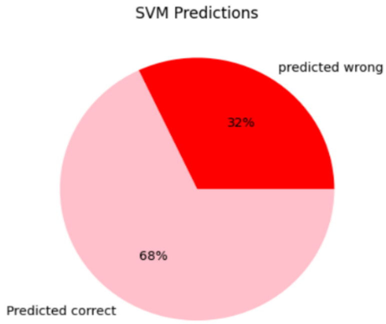
KNN classification Report

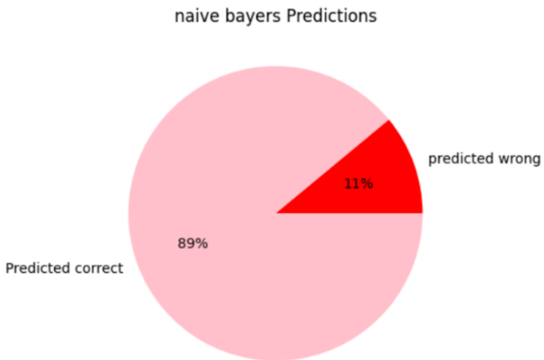
```
report_knn=classification_report(y_test,p5)
print(report_knn)
```

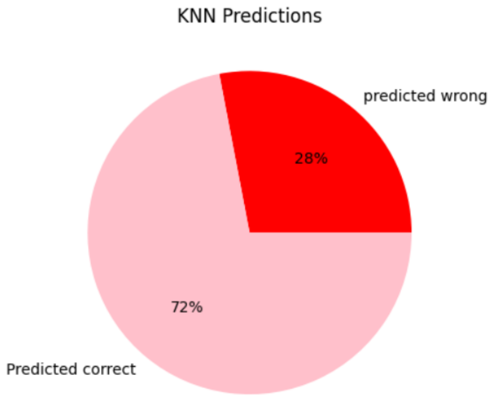
```
              precision    recall  f1-score   support

           0       0.70      0.75      0.72       310
           1       0.74      0.69      0.72       324

    accuracy                           0.72       634
   macro avg       0.72      0.72      0.72       634
weighted avg       0.72      0.72      0.72       634
```

Pie Chart

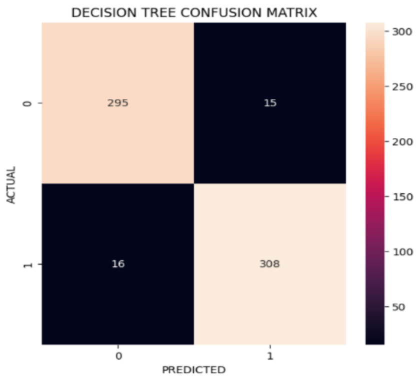


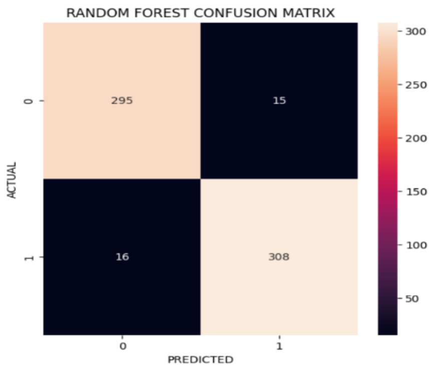


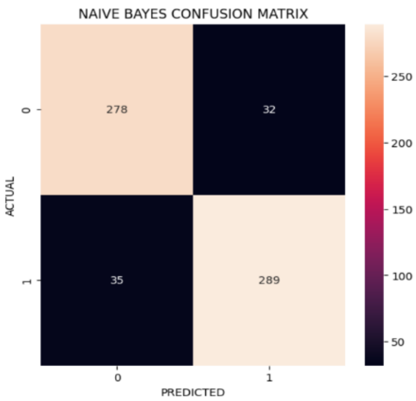


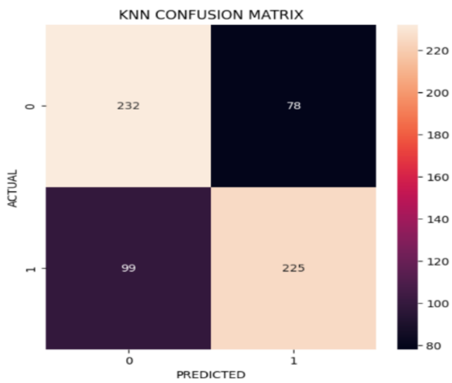


Confusion matrix









Chapter 5

Conclusion and Future Work

This work has proposed the use of acoustics to predict Gender of human beings using their voice input. The work has also analyzed the use of above-mentioned speech features with different Machine Learning models to provide the prediction with most accuracy.

In this paper, we can see that based on precision, F-score, recall values ,pie chart, accuracy score and confusion matrix we can say that among all the above algorithms RANDOM FOREST CLASSIFICATION is best suited for this problem with 98% accuracy. Support vector Machine is lest suitable based on same analysis.

Recognizing the gender of human voice has been con.sidered one of the challenging tasks because of its importance in various applications

In the future work, more experiments are being conducted to use many feature categories, ML techniques, and other natural feature selection techniques. Furthermore, the proposed techniques are being examined on different datasets. Different methods and techniques are being developed to improve these algorithms to use in various applications .

References

[1] International Journal of Innovative Science and Research Technology ISSN No: - 2456-2165

[2] M. Grimaldi and F. Cummins, "Speaker identification using instantaneous frequencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6, pp. 1097–1111, 2008.

[3] R. Djemili, H. Bourouba, and M. C. A. Korba, "A speech signal based gender identification system using four classifiers," in *Proceedings of the 2012 International Conference on Multimedia Computing and Systems*, pp. 184–187, Tangiers, Morocco, May 2012.

[4] Eliathamby Ambikairajah. Speaker verification - the present and future of voiceprint based security. http://www.apsipa.org/doc/APSIPA, 21.10.2013. last accessed on 10.04.2022.

[5] Gender Recognition by Voice using an Improved Self-Labeled Algorithm Ioannis E. Livieris , Emmanuel Pintelas and Panagiotis Pintelas. Published: 5 March 2019

[6] Pˇribil, J.; Pˇribilová, A.; Matoušek, J. GMM-based speaker gender and age classification after voice conversion. In Proceedings of the 2016 First International Workshop on Sensing, Processing and Learning for Intelligent. Machines (SPLINE), Aalborg, Denmark, 6–8 July 2016; pp. 1–5.

[7] Pˇribil, J.; Pˇribilová, A.; Matoušek, J. GMM-based speaker age and gender classification in Czech and Slovak.J. Electr. Eng. 2017,68, 3–12.