[[1]](#footnote-0)***Abstract —***

***Voice biometrics is one of the most secure system till now. The Project aims to develop a software which takes in voice inputs and on basis of machine learning modelling it predicts whether the speaker is authenticated user or not. This will bring great innovation and convenience in everyday life of common people. Like every person has his/her own unique iris, fingerprint, voice of every person is also unique.***

***The project implements voice biometrics and also consists of an interface which takes voice commands by the user to perform certain actions.***

# ***INTRODUCTION:***

Since the traditional lock and key has flaws, researchers are trying to build innovative security systems including voice biometrics. But since no machine learning model can reach 100% accuracy, it may give wrong results at times which could be irritating and it is what we call machine learning. I have tried to increase the accuracy of the model as much as possible.

The project features text dependent speaker recognition since text dependent model performs well than text independent giving it a plus point. The User will be required to speak a phrase or word written on the screen in the time limit given. The model will try to classify if the person speaking is its owner or not based on the data fed while training the model.

As displayed in the science fiction movies which include iron man, the project also focuses on giving user a taste of what it feels to be iron man i.e. the further part of the project is a software which takes in audio input from the owner which it triggers only after identifying the owner and perform some basic tasks. The tasks could be opening an application, creating new text file, searching on google, youtube search, search on wikipedia,etc.

***Objective*:** The main motive of this research is to gain enough eligibility to replace the traditional pattern lock systems or even character password systems with the modern voice biometrics security systems. Since machine learning models also make mistakes voice biometrics is currently not in extensive use. With time, these researches will prove to be helpful in attaining the aim.

# ***Dataset:***

The dataset contains 40 rows of valid users and 100 rows of unknown speakers. I had taken 20 data points for each unknown speaker i.e. 5 unknown speakers each having 20 voice samples. Each datapoint has 40 features which are MFCCs of the audio samples.Target variable i.e. speaker column is in integer datatype in which 1 means authorized person and 0 means unknown person. Here as an authorized person I have registered myself. Each speaker was asked to utter the same phrase “Jai Jinendra” which is a greeting phrase in Jainism religion.

# ***Results:***

The following are the comparative results of various machine learning models with and after applying some changes.

**The following are the scores with simply defaults parameters**.

Score of MLPClassifier: 0.6071

Score for decision tree classifier: 0.8928

Score for Random forest classifier: 1.0

Score for KNN classifier: 0.9642

Score for gaussian NB classifier: 0.8928

Score for bernoulli NB classifier: 0.9181

**After changing “sahil” & “Unknown” Y variables to 1 and 0 respectively.**

Score of MLPClassifier: 0.8571  
Score for decision tree classifier: 0.8928  
Score for Random forest classifier: 1.0  
Score for KNN classifier: 0.9642  
Score for gaussian NB classifier: 0.8928  
Score for bernoulli NB classifier: 0.8928  
Score for SVM classifier: 0.6071

**Taking equal data points of both the classes.**

Score of MLPClassifier: 0.5625   
Score for decision tree classifier: 0.8125  
Score for Random forest classifier: 0.9375   
Score for KNN classifier: 0.9022  
Score for gaussian NB classifier: 0.9375  
Score for bernoulli NB classifier: 0.875  
Score for SVM classifier: 0.7521

**Taking different train and test samples sizes**

| Test Size | MLP | Tree | RF | KNN | NB gaussian | NB bernoulli | conclusion |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0.2 | 85.7 | 89.2 | 100 | 96.4 | 89.2 | 89.2 | RF |
| 0.1 | 92.85 | 100 | 100 | 100 | 92.85 | 92.85 | Tree,forest,knn |
| 0.25 | 85.71 | 91.42 | 97.14 | 97.14 | 88.57 | 88.57 | RF, Knn |
| 0.3 | 83.3 | 92.85 | 95.23 | 95.23 | 88.09 | 88.09 | RF, Knn |
| 0.15 | 85.71 | 95.23 | 100 | 100 | 90.47 | 90.47 | RF, Knn |

**MLP:**

Only solver and activation were dominant parameters in MLP. Adam and sgd solver have more or less accuracies near each other while lbfgs had serious issues in convergence in many combinations of the parameters. The activation “logistic” showed good results with adam solver while activation “relu” showed fine results with “sgd” and “lbfgs”.

**Random Forest:**

all the combinations of parameters viz. Criterion, Max\_features, Max\_depth and n\_estimators performed well while train and test phase. But while cross validating Random forest there was a visible drop in accuracy. In real time RF performed affordable when criterion was set to “entropy” and Max\_features set to “auto”.

# Conclusion

The research done in the project gave us an understanding about the topic speaker recognition or speaker identification. The whole project is done in python language. Sample voices of all the speakers which were used to train the model were recorded in a silent environment having minimal noise of some high frequency sounds in background. So one of the limitations of the working system is that it should be operated in a silent environment which itself is a major issue. But with effective silence removal techniques the objective of the research of using the software for security of our computers can be achieved. The models tested were Neural network, decision tree, random forest, KNN, Gaussian Naive Bayes, Bernoulli Naive Bayes and SVM. Except MLP and random forest every other model went saturated and gave similar accuracies. While with some more trials MLP and Random forest started working better and now giving right results given that there is no high background noise. Although the train and test accuracies were increasing with some changes, the accuracies in real time testing with the hardware were important. So those accuracies went from 60% which was highest initially to above 95%. The former is of MLP and the latter was given by MLP,Random forest, Bernoulli Naive Bayes.Another disadvantage of using voice biometrics security systems is that the voice of a user could alter with time, or may change simply because of cold. So at that time it reduces the true positive answers and eventually decreases accuracy. This happened during the final testing process of the selected model. I got a viral infection due to which my throat was sore for 3 days. During that time the model failed to recognize myself and the research came to a halt.  
 Finally the future scope of the project will consist of developing the input and voice processing techniques, making more features count to make the decisions, implementing silence removal techniques and noise removal techniques so that the machine focuses only on the important part of the audio and lastly adding a feature of having more than one authorized speaker with the point of view that the software could be of great use in a conference room.

References

1. <https://www.researchgate.net/publication/326632358_REVIEW_SPEAKER_RECOGNITION_USING_AUTOMATED_SYSTEMS>
2. <https://www.ijert.org/research/a-review-on-speaker-recognition-approaches-and-challenges-IJERTV2IS50845.pdf>
3. <https://medium.com/@alexandro.ramr777/audio-files-to-dataset-by-feature-extraction-with-librosa-d87adafe5b64>
4. <http://blog.dominodatalab.com/building-a-speaker-recognition-model>
5. <https://www.researchgate.net/publication/302734183_Speaker_identification>

1. [↑](#footnote-ref-0)