

INXOL TECHNOLOGIES



AI Intern.

TEAM LEAD

Muhammad Ikrma Sharaf

Submitted by

C1_20_Faiz Raza

faiz.raza.dec@gmail.com

[faizrzaadec/AI-Intern_INXOL \(github.com\)](https://github.com/faizrzaadec/AI-Intern_INXOL)

TASK 1: IMAGE CLASSIFICATION

Situation: AI Gender Classification for Intelligence Gathering

Problem Statement:

In a modern and technologically advanced world, an intelligence agency is tasked with monitoring public spaces and gathering information for various purposes, including security and threat assessment. As part of their efforts, the agency is collecting vast amounts of data, including facial images of individuals, from surveillance cameras, social media, and other sources.

One specific task they are working on is to develop an AI-powered gender classification model. This model will be used to automatically classify the gender of individuals in the collected facial images. The agency believes that this information could provide valuable insights into demographic patterns and potentially aid in identifying persons of interest more efficiently.

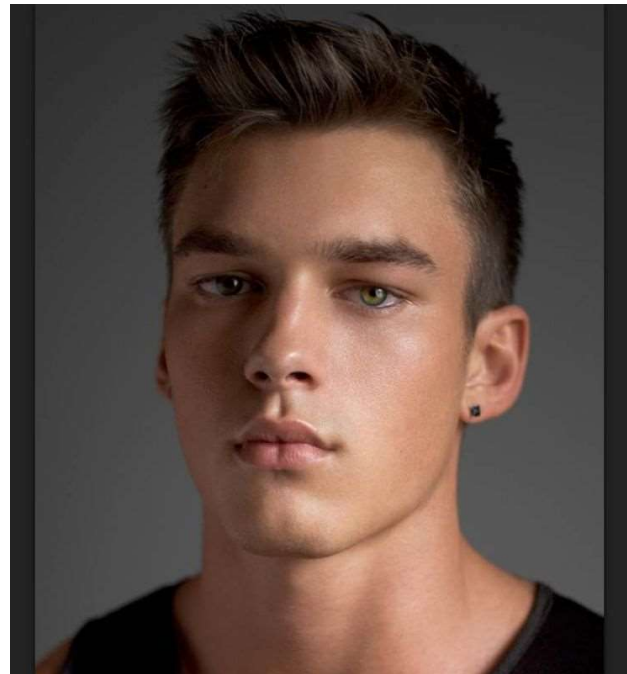
Your Task:

You are supposed to create *Gender Classification* model using *Human Facial* images dataset. Develop gender classification models using provided approaches:

- **Machine Learning Model:** Train a classification model using machine learning. Measure accuracy and computational cost.

Methodology

Data Collection: Data is collected from the Kegg! A collection of Male and Female Faces in a Large Quantity. Sample of the data is below...



Data Preprocessing: In data cleaning, first off, we read the images, after reading we did the following preprocessing steps:

- From the picture we detect the face, if there is any face in the image or not, e.g if we have the image of the person of full body, we don't want to process all the image, we'll just detect the face, so we can just process the face in which we're interested, it'll help us to reduce the computational cost.
- After detecting we extract the pixels of just the face, the rest of the image is not for our use,
- Convert the matrix into data frame after flatten and turning into list.
- We know the max value of a pixel is 255, so we divide all the data frame by 255, it's called normalization, it helps us to increase the accuracy of the model and reduce computational cost of ours.
- Randomize the data frame so the sequence of the labels will be randomized, it'll help us to reduce the effect of biasness in the model!

Dataset Splitting: First we split the data frame into features and target! (x, y). Then the x and y are further splitting into training and the testing. The training part will help us to train the model, testing set will be used to test the model.

Model Selection: This is our Task 1, we have to train the model using Machine Learning Classifiers, so here eight classifiers are used, their names are given below:

- ComplementNB
- DecisionTreeClassifier
- RandomForestClassifier
- LogisticRegression
- SVC
- KNeighborsClassifier
- LinearDiscriminantAnalysis
- QuadraticDiscriminantAnalysis
-

Model Evaluation: In model evaluation, we find the model accuracy on both training and testing sets, classification report (which include the Precision, F1 Score, Recall and more) and Confusion. This step provides insights into the model's real-world performance.

Which Software is used?

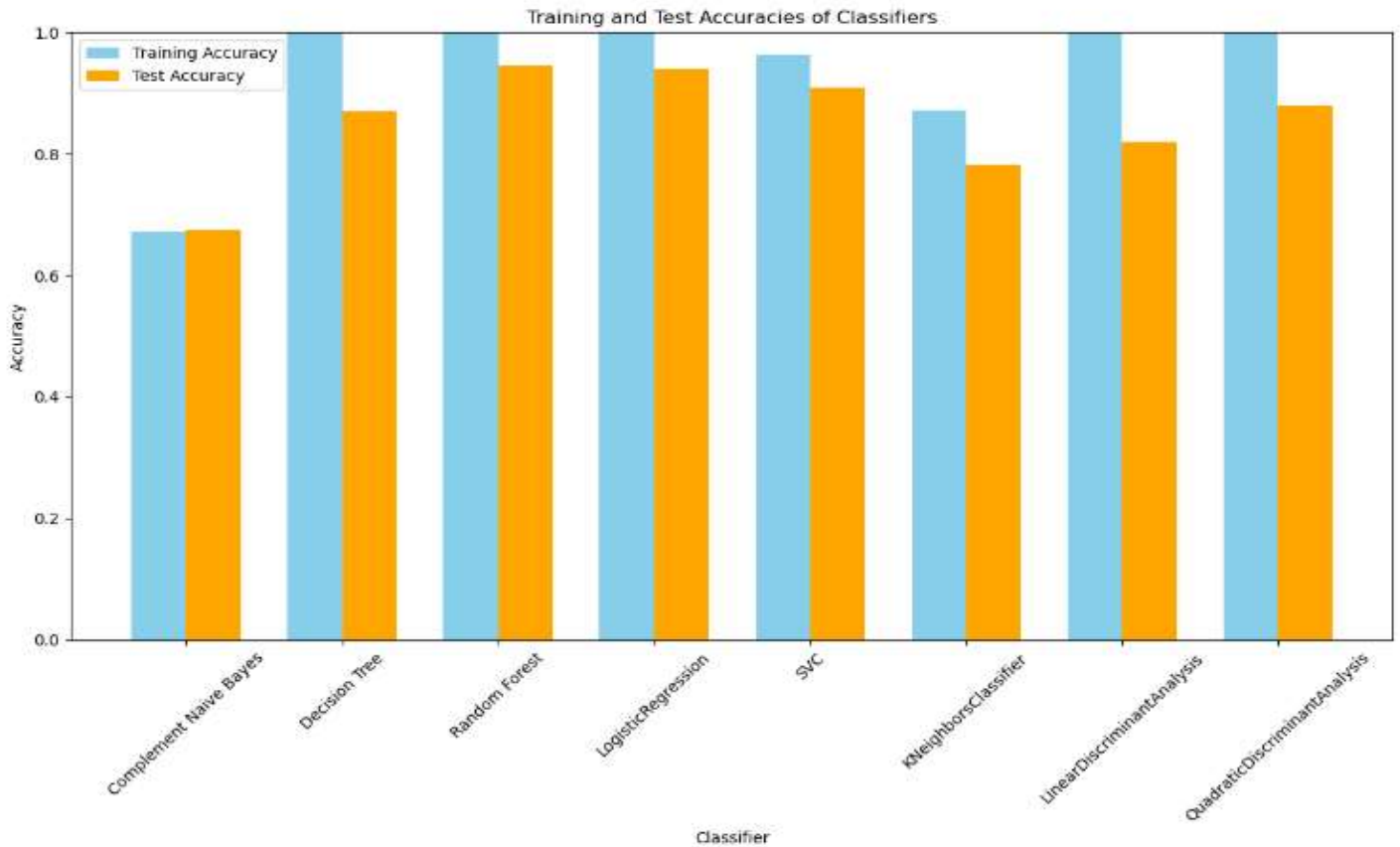
Jupyter Notebook (Anaconda)

System Specs?

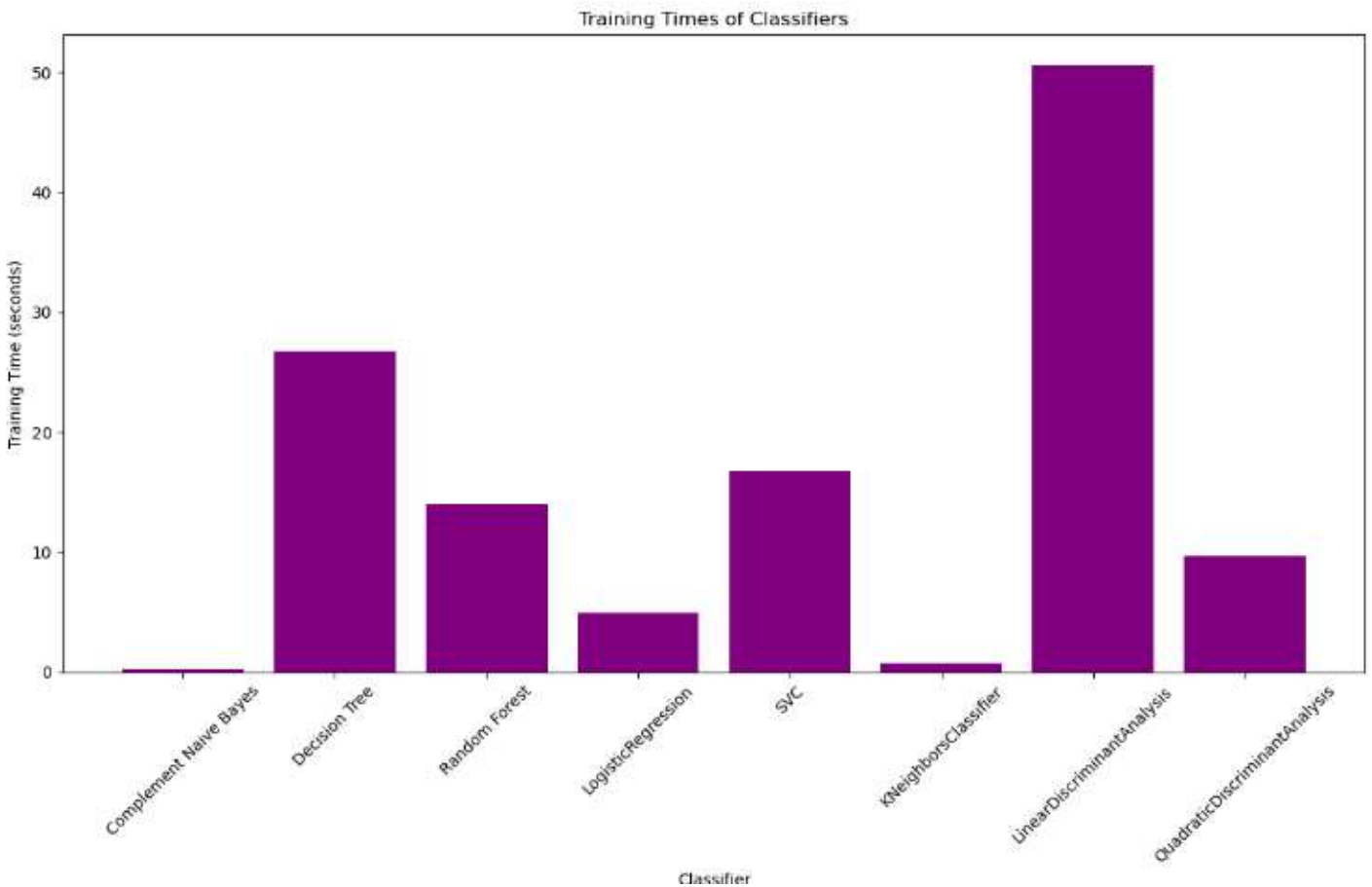
Processor Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz 2.70 GHz
 Installed RAM 12.0 GB

Source Code: See the code at [AI-Intern_INXOL/Task 1 \(Gender Classification\)/Using ML Classifiers/Gender Classification Using ML Different Classifiers.ipynb](https://github.com/faizrzaDec/AI-Intern_INXOL/blob/main/Task%201%20(Gender%20Classification)/Using%20ML%20Classifiers/Gender%20Classification%20Using%20ML%20Different%20Classifiers.ipynb) at main · faizrzaDec/AI-Intern_INXOL (github.com)

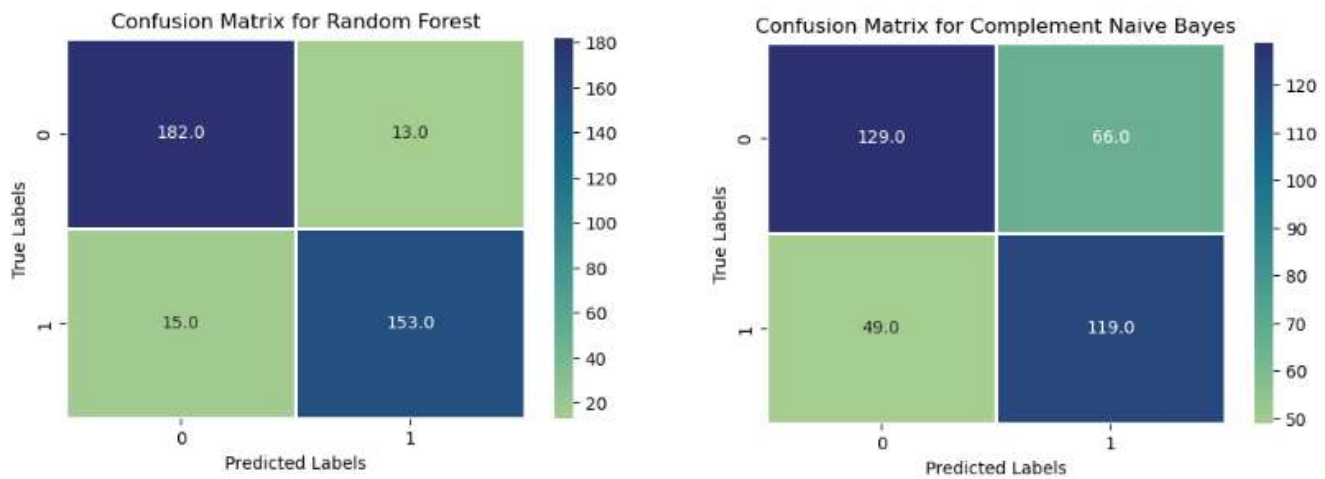
Evaluation of Classifiers: The below Graph is the summary of all the classifiers, that was used. Sky blue bar shows the training accuracy of the respective classifier and yellow orange shows the testing accuracy of the classifier! You can analyze that Random Forest Classifier gives us the greater accuracy than any other classifiers. And lowest accuracy is given by Complement Naïve Bayes.



The given graph is showing the time comparison of the classifiers that is taken by all each classifiers while model fitting. The most time taken was Linear Discriminant Analysis but the accuracy was average, the lowest time taken by Complement Naïve Bayes but the accuracy was also so worst.



Conclusion: In this Task, eight different classifiers are used on the same dataset, although the classifiers are used for classification data, but we got different accuracies and the time taken by each classifier was also different.



As you'll look at the confusion matrix, you'll see that the best prediction are doing by **RandomForestClassifier** and the worst are doing by **ComplementNaïveBayes**. So, according to this report, the **RandomForestClassifier** shows the best result overall!