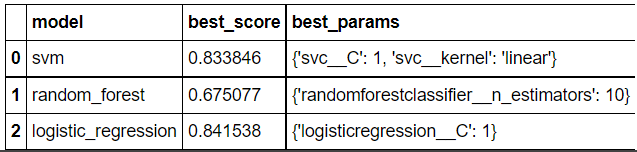
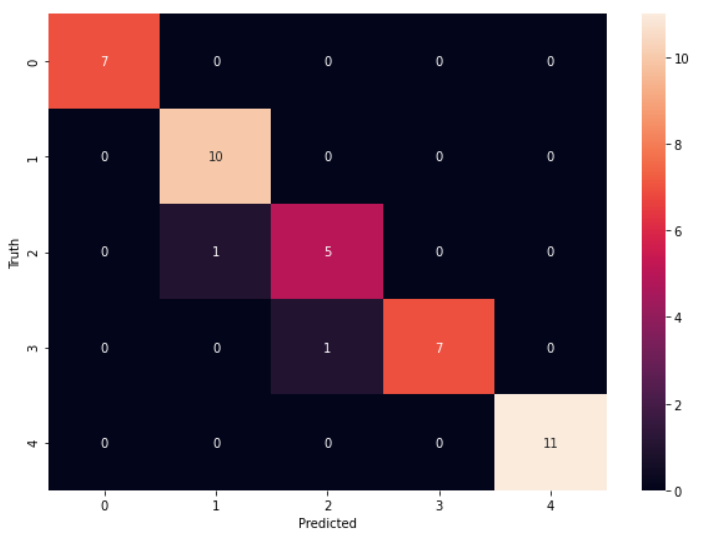
In this data science and machine learning project, I have classified sports personalities. I have restricted classification to only 5 people,

1. Maria Sharapova
2. Serena Williams
3. Virat Kohli
4. Roger Federer
5. Lionel Messi

## Technologies used in this project,

1. Python 3.8
2. Numpy and OpenCV for data cleaning
3. Matplotlib & Seaborn for data visualization
4. Sklearn for model building
5. Jupyter notebook, visual studio code and pycharm as IDE
6. Python flask for http server
7. HTML/CSS/Javascript for UI

## Model Selection and Methodology

[](https://github.com/ritwik4690/Image-Classifier/blob/master/model_performance.png)  
In this project, OpenCV haarcascade was used to detect face and two eyes. Wavelet transform was used to extract these features from the images. Finally after processing and cleaning the data, it was given as input to models. Various models were tried using GridSearchCV and **Logistic Regression** was found to give the best resutls with a score of **84.15%**.  
  
[](https://github.com/ritwik4690/Image-Classifier/blob/master/heatmap.png)  
Above is the heatmap showing the results of Logistic regression on test data. For test images, Logistic regression was found to give a score of 95%. The model accuracy could be improved by using a bigger dataset. Also deep learning may give better results for image classificaion