Lab 7

**Step1**: Import all libraries

• Numpy

• Pandas

• Matplotlib, Pyplot

• Seaborn

• SVC

• Train\_test\_split

• Confusion\_matrix, classification\_report, accuracy\_score, balanced\_accuracy\_score

**We will be creating our own dataset from the three given csv files. We will do that in the next 4 steps:**

**Step**2: Load all the data from the three given csv files for ODI, T20 and test match types.

**Step3**: Concatenate runs from all the match type and store in **Runs**. Similarly do for strike rate and store in **SR** and Grounds in **ground**.

**Step4**: Create a new **match\_type** variable by storing info on the type of match for each datapoint. This can be extracted from the csv filename for each type of match. The 3 categories will be: ***ODI, T20, TEST*.**

**Step5**: Create a data frame **data\_cricketer** and store Runs, SR, **match\_type** and **grounds** in it.

**Step6**: From **data\_cricketer**, store **Runs** and **SR** in **X** and **match\_type** in **Y.**

**Step7**: Split the data X and Y in **x\_train, x\_test, y\_train , y\_test.** Use **test size=0.3** and **random state=0**

**Step8:** Use sklearn to perform SVM classification on this data. Try different kernels to find out the best one for your data.

**Step 9**: Fit a SVM classification model with a Polynomial kernel with degree 6.

**Step10**: For this model, compute confusion matrix, accuracy and balanced accuracy

***Step11: Plot confusion matrix***

**Step12:** Also print **classification report** that includes precision, recall, f1 score and accuracy.

**Step13**: Create a scatter plot for Runs scored on x-axis and strike rate on y-axis. Group each point by match type (by assigning a different colour for each match type on the plot)

Note: See why we are using polynomial degree of 6th order. If you can visualize it, plot this graph.

Submission the following:

\* Submit Code pdf, code editable link (if colab), code file.

\* All the graphs and metrics calculated need to be plotted / printed.