Now that you have been equipped with the skills to use different Machine Learning algorithms, you will have the opportunity to practice and apply it on a data set.

In this scenario, you are a Data Scientist working for a college basketball team. Your coaches have asked you to look at historical data to see which team metrics (individually or in combination) make a team more likely to make it into the Final Four. For example, if a team is more efficient defensively, does this have a direct relationship to their ability to get into the Final Four? What about defensively efficiency along with overall wins? Your job is to figure out if there is a combination of metrics that give a team more of a chance of making it into this tournament.

Something to keep in mind is that when trying to predict results of basketball tournaments there are many variables that need to be taken into account. As a result of this creating accurate models is incredibly hard. In the sports betting industry an accuracy rate of anything over 55% is considered good as it indicates profits.

You will load a historical data set from previous seasons, clean the data, and apply different classification algorithms to the data. You are expected to use the following algorithms to build your models:

k-Nearest Neighbour
Decision Tree
Support Vector Machine
Logistic Regression
The results are reported as the accuracy of each classifier, using the following metrics when applicable:

Jaccard index
F1-score
Accuracy
This final project will be graded by your peers who are also completing the course during the same session. This project is worth 25 marks of your total grade, and is distributed as follows:

Review Criteria

Build a KNN model using a value of k equals five, find the accuracy on the validation data (1 mark)
Determine the accuracy for the first 15 values of k the on the validation data:. (1 mark)
Determine the minimum value for the parameter that improves results on validation data. (1 marks)
Building model using Support Vector Machine. (2 marks)
Train a logistic regression model and determine the accuracy of the validation data (set C=0.01) (2 marks)
Calculate the F1 score and Jaccard Similarity score for each model from above. Use the Hyperparameter that performed best on the validation data (2 marks)
Step-By-Step Assignment Instructions:

Step A: Create an account in Watson Studio if you don't have an account already. (If you already have an account, jump to Step B).

Final Project Setup
Step B: Sign into Watson Studio and import your notebook

Sign in to Final Project Setup
Click on "New Project".
Select "Data Science" as type of project.
Give a name to your project and a description for your reference, then set-up your project as follows, then click "Create".
Notice 1: Because you are going to share this project with your peers for evaluation, please make sure you uncheck "Restrict who can be a collaborator".

Notice 2: You must create an IBM Object Storage, if you don't have an IBM Object Storage, you can use the free Lite plan.

5. From the top-right, click on "Add to project" and then select "Notebook".

6. In the "New Notebook" form, click on "From URL" and enter the Notebook URL.

7. Give your notebook a proper name and description and click on "Create Notebook" to initialize the notebook.

Step C: Complete the Notebook

Start running the notebook.
Complete the notebook based on the description in the notebook.
Step D: Share the Notebook

Click on the share icon on the top-right side of your page.
Activate the "Share with anyone who has the link".
Select "All content excluding sensitive code cells".
Copy the link from "Permalink to view notebook".
Submit your Notebook for Grading

Paste the shared link of your Notebook in the provided text box below for peer-review.