­­­­Football Player Price Prediction

**Problem Statement:**

In the English Premier League, May - July represents a lull period due to the lack of club football. What makes up for it, is the intense transfer speculation that surrounds all major player transfers today. An important part of negotiations is predicting the fair market price for a player. You are tasked with predicting this Market Value of a player using the data provided below

The attached data set consists of the following attributes:

* **name**: Name of the player
* **club**: Club of the player
* **age** : Age of the player
* **position** : The usual position on the pitch
* **position\_cat** :
  + 1 for attackers
  + 2 for midfielders
  + 3 for defenders
  + 4 for goalkeepers
* **market\_value** : As on transfermrkt.com on July 20th, 2017
* **page\_views** : Average daily Wikipedia page views from September 1, 2016 to May 1, 2017
* **fpl\_value** : Value in Fantasy Premier League as on July 20th, 2017
* **fpl\_sel** : % of FPL players who have selected that player in their team
* **fpl\_points** : FPL points accumulated over the previous season
* **region**:
  + 1 for England
  + 2 for EU
  + 3 for Americas
  + 4 for Rest of World
* **nationality**
* **new\_foreign** : Whether a new signing from a different league, for 2017/18 (till 20th July)
* **age\_cat**
* **club\_id**
* **big\_club**: Whether one of the Top 6 clubs
* **new\_signing**: Whether a new signing for 2017/18 (till 20th July)

You have learned about a number of regression algorithms in your course: Linear Regression, Lasso Regression, Ridge Regression, Nearest Neighbour Regression, Support Vector Regression, Tree Regression, Random Forest Regression and Gradient Boosted Regression.

Your ask is:

* Use Seaborn to investigate the data and present your findings (20 marks)
* Build models using all the algorithms above to predict market\_value (15 marks)
* Tune the hyperparameters and build the most accurate model (20 marks)
* Use model selection approaches discussed in class to choose the best model (10 marks)
* Implement a Genetic Algorithm for learning attribute weights for the Nearest Neighbour Algorithm. Implement at least one mechanism for maintaining Diversity within the Population (25 marks)
* Deploy your model as a RESTful Web Service (10 marks)

**Submission Requirement:**

Code and Report (100%) –

• A short report, no more than 3 pages long, providing key observations from Exploratory Data Analysis and a comparative analysis of the different models and hyper parameters chosen

• the iPython Notebook or Python Code

• Deployed RESTful Web Service of you winning model