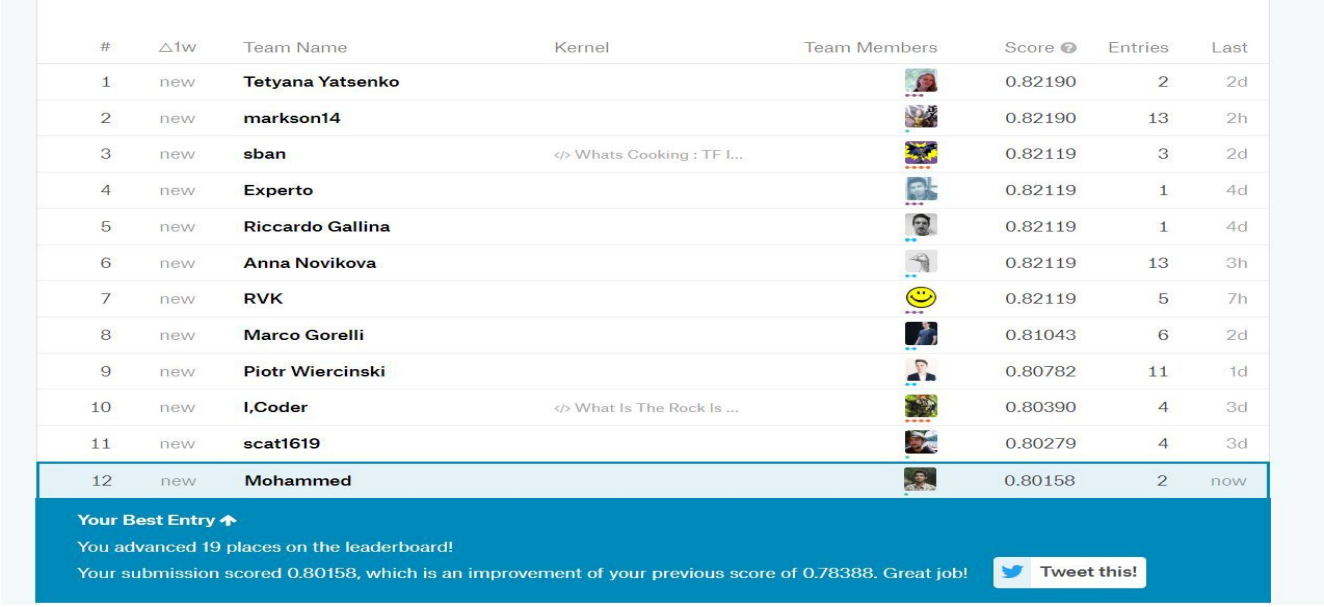
Big Data 4X Group work

Topic - What's Cooking ?

Submitted by - Patricia Londono, Fernanda De Oliveira Guimaraes,
Abraham Chandy, Mohammed Topiwalla



#	△1w	Team Name	Kernel	Team Members	Score	Entries	Last
1	new	Tetyana Yatsenko			0.82190	2	2d
2	new	markson14			0.82190	13	2h
3	new	sban	</> Whats Cooking : TF I...		0.82119	3	2d
4	new	Experto			0.82119	1	4d
5	new	Riccardo Gallina			0.82119	1	4d
6	new	Anna Novikova			0.82119	13	3h
7	new	RVK			0.82119	5	7h
8	new	Marco Gorelli			0.81043	6	2d
9	new	Piotr Wiercinski			0.80782	11	1d
10	new	I,Coder	</> What Is The Rock Is ...		0.80390	4	3d
11	new	scat1619			0.80279	4	3d
12	new	Mohammed			0.80158	2	now

Your Best Entry
You advanced 19 places on the leaderboard!
Your submission scored 0.80158, which is an improvement of your previous score of 0.78388. Great job! Tweet this!

What is the task ?

Given a set of ingredients, can you determine the cuisine of the dish? We have 20 different cuisines and 3384 different dishes. What makes this difficult is that every dish can have

different number of ingredients and text mining techniques would need to be applied to carry on before using machine learning techniques.

[See the Competition on Kaggle]

What we did ?

We ran different text mining techniques like countvectorizer and inverse document frequencies technique. Ultimately the winner was IDF. After this we started iterating and trying our luck out with

- Logistic Regression and parameter C tuning - 79%
- Random forest with hyper parameter tuning - 72%
- XGBOOST with hyper parameter tuning - 73%
- SVM with hyper parameter tuning - 81%

You guessed it wrong, the winner here was not SVM but neural network on kaggle with these parameters. We landed with this rank on Kaggle -12/56

Technology we used?

- Python Kaggle jupyter notebook
- Sklearn, nltk idf vectorizer
- Packages for xgboost, random forest , tensorflow, svm