Details:

Kaggle UserName: Nabilahmed Patel (nabilahmed.patel@mavs.uta.edu)

Kaggle Rank: 1874

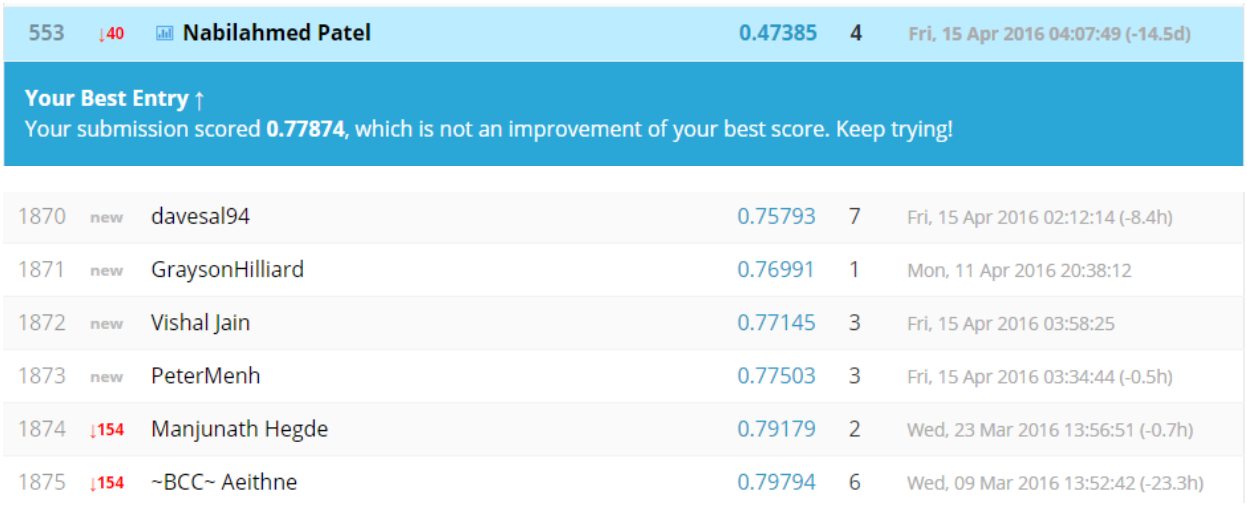
Kaggle Score: 0.77874

I had done submission few days ago for just checking using file from kaggle itself and I got the rank 543.

Now I am trying to submit my own file, It shows score 0.77874. But, it doesn't show my rank.

But from my score I deduced that my rank would be 1874.

Thank you.



553	↓40	**Nabilahmed Patel**	0.47385	4	Fri, 15 Apr 2016 04:07:49 (-14.5d)

Your Best Entry ↑
Your submission scored **0.77874**, which is not an improvement of your best score. Keep trying!

1870	new	davesal94	0.75793	7	Fri, 15 Apr 2016 02:12:14 (-8.4h)
1871	new	GraysonHilliard	0.76991	1	Mon, 11 Apr 2016 20:38:12
1872	new	Vishal Jain	0.77145	3	Fri, 15 Apr 2016 03:58:25
1873	new	PeterMenh	0.77503	3	Fri, 15 Apr 2016 03:34:44 (-0.5h)
1874	↓154	Manjunath Hegde	0.79179	2	Wed, 23 Mar 2016 13:56:51 (-0.7h)
1875	↓154	~BCC~ Aeithne	0.79794	6	Wed, 09 Mar 2016 13:52:42 (-23.3h)

Design and Implementation:

I have tried to solve this problem using **"Naïve Bayes"** method of classification. I used "Naïve Bayes", because it is the one of the most appropriate method for classification method. The training set is given as **"train.csv"** and testing set is given as **"test.csv"**. If someone closely takes look at "train.csv", he/she can easily find that there are exactly 13 different values for relevance **{1.0,1.25,1.33,1.5,1.67,1.75,2.0,2.25,2.33,2.5,2.67,2.75,3.0}**. Therefore, one can deduce that there is **13 classes** for relevance.

To apply **"Naïve Bayes"** classification, I have first created some features/attributes for training data using **"train.csv"**, **"product_descriptions.csv"** and **"attributes.csv"**. I have created attributes like **"product_title", "product_description", "product_brand" and "product_material"** using the method **"create_features ()"**. After creating features, I have

developed "**Naïve Bayes Model"** using method like **"naive_bayes_model ()"** which will in turn call **"probability_of_attributes ()"**, **"probability_of_class ()", and "naïve bayes Calculation ()".**

After getting this **"Naïve Bayes Model",** I have created same features/attributes for testing data using **"test.csv"**, **"product_descriptions.csv"** and **"attributes.csv"**. Same 4 attributes are created for testing data also using method **"create_features ()".** By using the ***model*** and ***features,*** calculation for relevance of testing data has been done. To find relevance of testing data, I have created method **"calculate_result ()",** which will count the relevance for all rows in ***test.csv.*** The result of **"calculate_result ()"** is converted to ***submission.csv*** file.

To implement, this whole functionalities, I have used packages like ***pandas*** (which is used to access .csv input files and to produce .csv output file). ***Pandas*** is read .csv file into ***DataFrame*** Datastructure.

Create_features (): Accept the details of training data or testing data with product details and return attribute/feature.

probability_of_class (): Accept the count of class and calculate the probability of all classes.

probability_of_attributes (): Accept the features and count of class and return the probability of attributes.