Big-Five

Big Five Personality Test

This project attempts to classify individuals into 5 cluster personalities. Based on the Big Five Personality Traits, these clusters are : **E**xtraversion, **A**greeableness, **C**onscientiousness, **N**euroticism and **O**penness.

Two machine learning algorithms were used in this project:

- Unsupervised ***KMeans Clustering*** (n_clusters=5)
- ***Perceptron*** (max_iter=40 was used)

After fitting both models, 4 fictitious characters (Willy, Xavier, Yenny, Zee) were created and the models were employed to predict which cluster their personalities fell under.

Insights

Results from KMeans Clustering yielded the results in the following table :

	Big Five Personalities (KMeans Clustering)				
Labels	***A**greeableness*	***C**onscientiousness*	***E**xtraversion*	***N**euroticism*	***O**penness*
0	128	323	80	351	789
1	1055	526	0	104	491
2	737	219	380	285	433
3	982	200	7	1	954
4	996	79	456	0	856
Total	**3898**	**1347**	**923**	**741**	**3523**
% accuracy	--	--	49%	47%	--

It appears that the clustering for **E** and **N** were the best, while **A** and **O** did not do so well.

For the Perceptron model, scores were 0.79 and 0.77 on the training and testing data respectively.

Labels for Perceptron model were tagged as follows:

Perceptron Labels	**Personality Cluster**
0	***E**xtraversion*
1	***A**greeableness*
2	***C**onscientiousness*
3	***N**euroticism*
4	***O**penness*

After fitting the two models, the predictions for the 4 fictitious characters (Willy, Xavier, Yenny, Zee) were given below.

	Predictions	
Individual	**KMeans Clustering**	**Perceptron**
Willy	***E****xtraversion*	***E****xtraversion*
Xavier	***N****euroticism*	***N****euroticism*
Yenny	***A****greeableness*	***O****penness*
Zee	***A****greebleness*	***A****greeableness*

The results above show alignment in the predictions for Willy and Xavier who were predicted to be ***E*** and ***N*** respectively. We note that this is also compatible with the earlier findings where ***E*** and ***N*** were the most accurately classified by K Means Clustering.

Yenny and Zee were classified as either ***A*** and/or ***O***. This is probably expected as ***A***s and ***O***s make up the majority of the personality types in the sample.

Next Steps

An argument of n_clusters = 5 was used to model KMeans Clustering. In using the Elbow Method for the graph on the optimal number of clusters(k), it would appear that the optimal clusters would be n_clusters = 3 or 4. This could be a result of using a much-reduced sample dataset for this particular project.

Subsequent iterations of this project would involve using the original dataset to carry out further analysis and predictions.

Data Sources

Data is retrieved from Kaggle : https://www.kaggle.com/tunguz/big-five-personality-test

Labels for the individuals were derived from calculations based on the questionnaire appended in this link : https://openpsychometrics.org/printable/big-five-personality-test.pdf. The labels were used for checking against the predictions in KMeans Clustering and for training/testing the Perceptron model.

For expediency purposes, the sample size of the data used in this project was reduced to a smaller subset of 10,432 individuals randomly selected from the original 1M individuals provided.