K.D. GULKO data science analytics

kdgulko@gmail.com 310-963-2535
Github: killerdame101
Linkedin: killerdame

Twitter is where users present themselves to the world, revealing personal details and insights into their lives, leaving behind a wealth of information ripe for mining.

Archetypes are like software of the psyche. Programs that each of us live our lives by. They are present in many aspects of our environment - from literature, philosophy, psychology, mythology, and even pop culture.

This project solves the problem of:

- Finding the right company or team culture fit while hiring.
- Finding creative ways to gather intelligence, discover new trends and discovering new groups to market to.

Project Objective:

I chose 12 well known celebrity Twitter users that I feel depict each of the 12 personality Archetypes. Pulling a maximum of 3200 of the most recent tweets from each user into a Mongo database, I use Natural Language Processing to parse the information into a usable form and then used Machine Learning to train a model. Using that trained model, a new Twitter user handle through the model to predict the personality Archetypes that user is most like.

Goal and Success Metrics

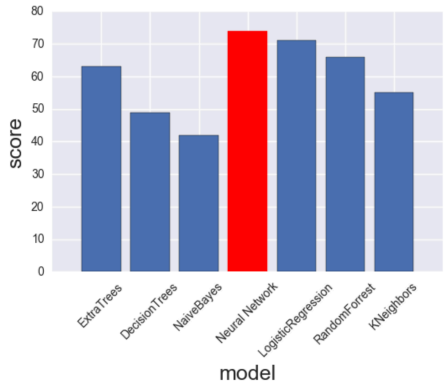
My goal was to identify a model with a level of accuracy and precision greater than 70% that will, when a new twitter account is run through the trained model, produce the archetype probability prediction of the twitter user.

Model Approach

Models Tested:

- Logistic Regression
- ExtraTrees
- DecisionTrees
- NaiveBayes
- Random Forrest
- MLPClassifier

After comparing the results of the 7 different Scikit Learn models, the Neural Network Multi-Layer Perceptron Classifier (MLPClassifier) performed the best with a 74% accuracy rating. KNeighbors.

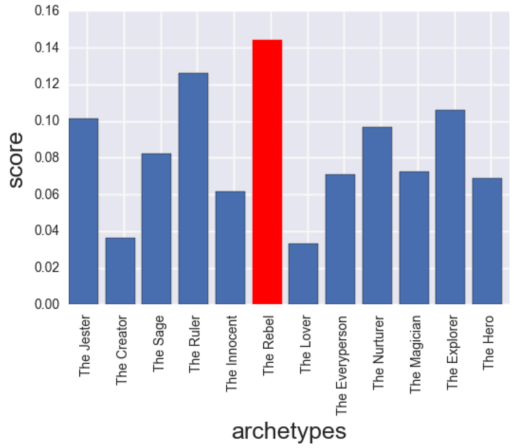


THE ARCHETYPES

The Innocent Jessica Alba
The Lover Madonna
The Explorer Anthony Bourdain
The Every-person Zooey Deschanel
The Creator Martha Stewart
The Rebel Bill Maher
The Hero Capt. Sully Sullenberger
The Jester Seth MacFarlane
The Magician Dr. Oz
The Nurturer Kid President
The Sage Pope Francis
The Ruler Hillary Clinton

Project outcome

Although I was able to predict the twitter Archetype probability with a 75% accuracy rate, as the model is now, only using one Twitter user per Archetype shows similarities to that user more than it does similarities to a particular Archetype.



Github:
https://github.com/killerdame/twitter-personality-prediction
Linkedin:
https://www.linkedin.com/in/killerdame/