Personality analyzer and affinity finder

07.04.2020

CS314b(Machine Learning)

—

Saumya Mishra | Kuhu Pyasi | Rushikesh Ghule
2017230   2017327   2017218

Methodology

We are aiming to develop a web application that would be able to analyze personality traits of people and based on this should be able to get insight about their inclinations, choices and commonality with the other people associated with them. The solution can leverage social media platforms to access publicly accessible information or use authentication from legal users for access to information available on social networking platforms.

Dataset

We are using following REST API's -

> ➢ Twitter : tweepy python wrapper to collect tweets
> ➢ Reddit : Python Reddit API Wrapper (PRAW) to get submissions by user
> ➢ Stackexchange : StackAPI python wrapper for stackexchange API to get users top questions and answers rated by votes
> ➢ Youtube : youtube API to get description and content of playlists created by user

Technologies / Platforms

> ➢ Python
> ➢ Tone Analyzer
> ➢ Natural Language Processing
> ➢ Linguistic Analysis
> ➢ Node K-Means Clustering Analysis
> ➢ Data Visualization
> ➢ Docker
> ➢ Nodejs
> ➢ Flask
> ➢ Native MongoDB
> ➢ Twitter, YouTube, Stack, Reddit APIs

Linguistic Analysis

It will be used to detect emotional and language tones in written text. We will be using it to get the nature of tone in a written text. By this way , we can analyze tone at both the document and sentence levels.

Natural Language Processing

Natural language processing is the parsing and semantic interpretation of text, allowing computers to learn, analyze, and understand human language. With NLP comes a subset of tools– tools that can slice data into many different angles. NLP can provide insights on the entities and concepts within an article, or sentiment and emotion from a tweet, or even a classification from a support ticket. Hundreds of types of information can be extracted from textual data, and enterprises can leverage this information to better understand customer behavior and improve internal efficiency. It will be used to get personal interests of a person.

BigFive

Big Five is one of the most studied of the personality models that were developed by psychologists (Costa and McCrae, 1992, and Norman, 1963). It is the most widely used personality model to describe how a person generally engages with the world. The service computes the five dimensions and thirty facets of the model. The dimensions are often referred to by the mnemonic *OCEAN*, where *O* stands for Openness, *C* for Conscientiousness, *E* for Extraversion, *A* for Agreeableness, and *N* for Neuroticism.

Data Visualization

It will be used to analyse the personality as well as for calculation of Friends Affinity using A Novel, Complex and Logical Method.

K-Means Clustering Analysis

It will be used to find out various categories of people. As Clustering allows us to identify which observations are alike, and potentially categorize them therein. K-means clustering is

the simplest and the most commonly used clustering method for splitting a dataset into a set of k groups.

Data Flow Diagram