Project Report: CX4240

**Introduction and Motivation:**

Music is one of the fundamental way people use to express their emotions and it is no secret that certain faction of people would prefer a certain type of music, whereas another faction would like some other type of music. Schafer and Sedlmeier explored and concluded that there are four main reasons people listen to music; social, emotional, self-related, and arousal-related (Schafer, Sedlmeier, Stadtler, & Huron 2013). Many researchers have tried to establish a relationship between personality type and choice of music and vice-versa. Catell was the first person to explore this topic and developed a list of 16 personality factor and based on the values of these factors during test predicted the personalities of individual. Another attempt to develop personality assessment was done by McCrae and Costa in 1992, during which they used Eysenck Personality Inventory (EPI) and developed a five factors model to assess personality factors in a person. These factors include extraversion, agreeableness, emotional stability conscientiousness and openness to experience. North, in 2008 established that a person’s preference for music is linked to his personality type. In 2009, Nicola Sigg researched on the music preference and personality wellbeing, even though the research failed to establish relationship between music preference, social identity and self esteem yet the study showed a definite relationship between personality traits and music preference. Furthermore, music is also described to have a calming effect on mind (Wells 1985) and enhancing effect on mood of individual.

The way people consume music has changed quite a bit over the past decades. In 2019, most of the music is consumed through internet based streaming services. More than 50 Million people in US alone use these streaming services and the overall size of music streaming services industry today in US is 4.3 Billion USD whereas 11 Billion USD worldwide. As all these streaming services cater to personalized music choices, and since we have established that music choices depend upon the personality traits, so the enormous consumer base of this industry presents a bright prospect to predict the personality types of the consumers. The data on the personality type of individual on one hand can be used to provide an even more personalized music suggestion in the streaming service while on the other hand can be used by the streaming service/shared with the advertisement (platform) to generate more targeted advertisements based on the personality type of individual. The overall size of internet-based advertisement industry is 333 Billion USD and it is projected to further increase to 517 Billion USD by 2023. Even a minor increase in the accuracy of targeted advertisement from music preference can account for billions of revenues for the advertisement industry.

We plan to address this specific area by developing an algorithm that predicts the personality type of individuals based on music choice and then generalize our results so that they could be used to develop targeted ads.

**Sample:**

In order to develop our algorithm, we have used a public dataset on Young People survey in which a total of 23737 participants from various countries were involved. A total of 150 different musical excerpts were used for the test but the participant had to listen to 25 randomly selected excerpts (15 seconds long) out of those 150. The participants and were asked to rate their likeness to those excerpts from 1-9. The genres for these musical excerpts were already labelled. The total number of participants who also took the personality test was 17904. A total of 19 genres were used for testing which were grouped into 5 main Music groups (M= Mellow, U=Unpretentious, S= Sophisticated, I=intense, C=Contemporary). The data on Personality traits analyzes the big 5 personality traits (O= openness, C= Conscientiousness, E= extraversion, A= Agreeableness, N=Neuroticism) and assigns a score for each trait from 1 – 5. Our main objective was to develop a model that can assess the music choices and based on that predict the Personality traits score for the individual.

**Methodology:**

**Preprocessing the Data:**

In order to get valuable insights from the data, it was important to pre-process the data.

* The data was preprocessed by creating labels for scores for each of the personality types. For this we assigned the personality trait a binary 1 condition if it showed a score of equal to or more than 2.5 in that personality type and 0 if scores were lesser than 2.5 in that particular trait.
* The data was clustered using the K Means clustering to study whether we can assign a single personality trait to the individual taking test.
* The data
* Two type of data were created, in the first data type the music preference score was treated as a single feature thereby giving us a total of 25 feature and in the second data type the music preference scores for each genre were averaged for each person thereby resulting in a total of 5 features i.e. average scores for each genre type.
* Two different prediction models were developed using ridge regression and LASSO regression.
* The data was split into 10 equal parts and 90 percent of data was selected to train our model whereas 10 percent data was used to cross validate the results of predictions from the prediction model.

**Results:**

**Conclusion:**