# NEW YORK UNIVERSITY

**COURSE:**

EL-GY 6143: Introduction to Machine Learning

**PROJECT:**

Sentimental Analysis of Songs from Different Genre

**SUBMITTED TO:** Pei Lui

**SUBMITTED BY:**

Mukta Maheshwari (mm11070)
Amangali Kadirgaliyev (akk8338)

# SUMMARY

This research was aimed to apply NLP techniques to the song lyrics to retrieve sentiment for lyrics. Source of textual data is genius. 5 artists from each genre were selected according to genius site ranking of the greatest artists of all time. To apply machine learning (ML) techniques for sentiment extraction we have gone through 4 fundamental steps of ML working process.

**First step: Data collection -** In this research work we have used 'LyricsGenius' library for lyrics extraction of the songs.

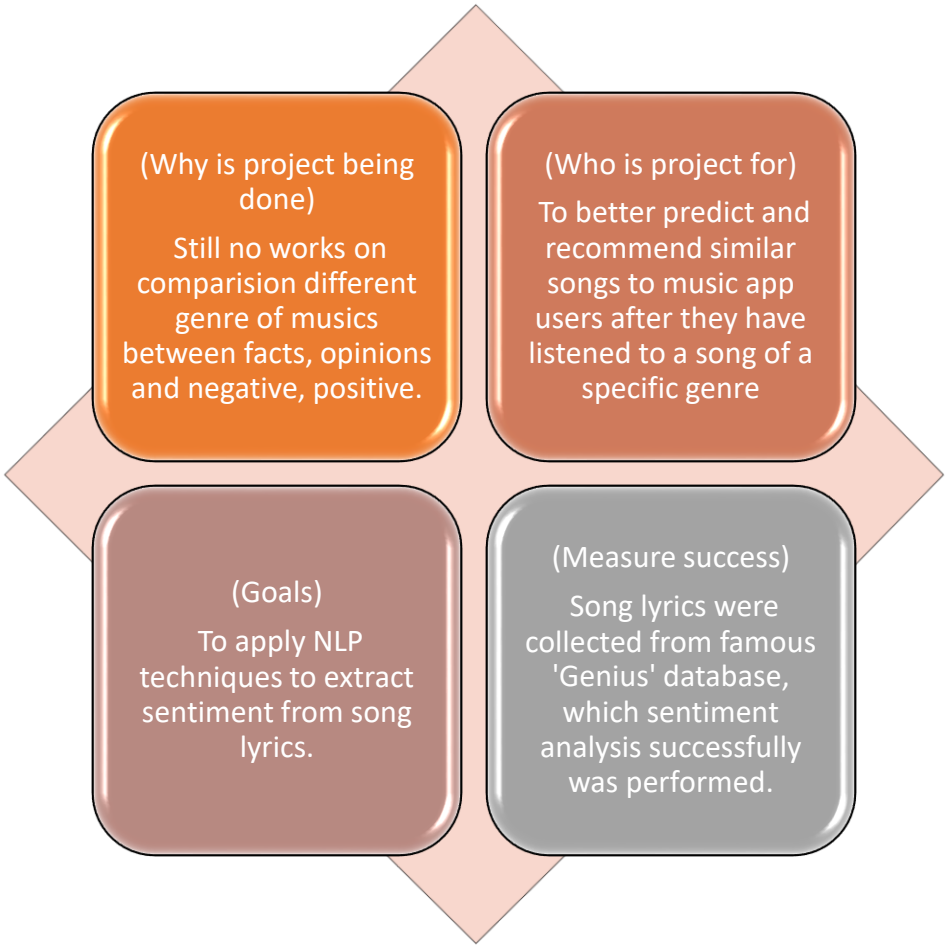
**Second step: Data cleaning -** We have gone through several cleaning rounds to decrease the noise. There were many types of noise. Initially noise was visible, but after several cleaning rounds there were not-so obvious noises. Usually, it was common words which were hiding interesting insights. Cleaning was not only things in this step. There was also organizing data, where raw data must be transformed into a workable format. That format is document term matrix.

**Third step: Exploratory data analysis -** Question was if the collected data makes sense or not. Initially prepared data still had some noise and some trivial data. We went back to the step 2, where we removed noise again. During this process, vocabulary was measured by the number of unique words.

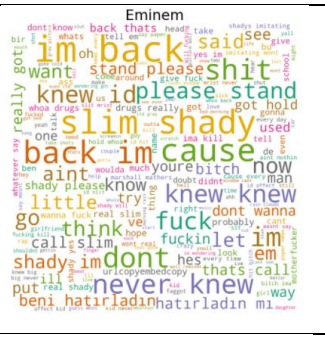
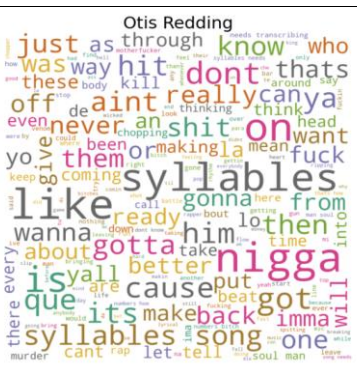
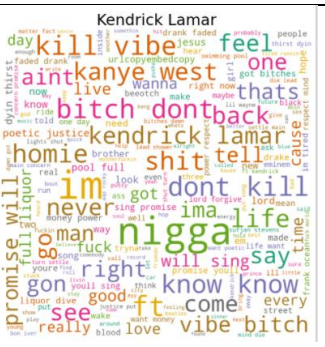
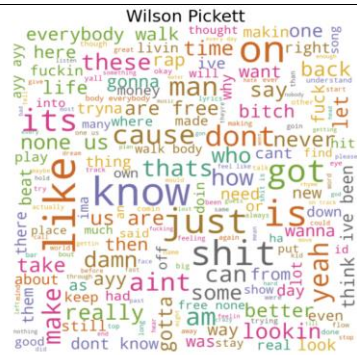
**Fourth step: Sentimental Analysis -** We extracted sentiment for each album with help of TextBlob opensource library.

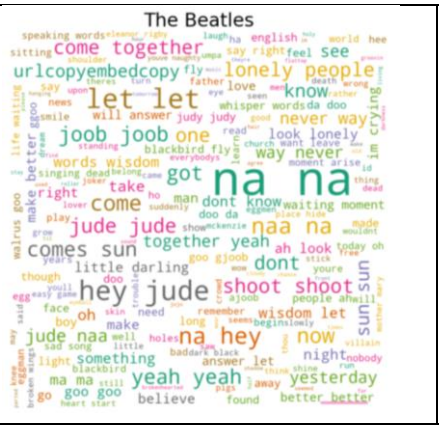
**To conclude**, most of albums were labelled as positive, only two artists were negative. Those artists are Eminem and Micheal Jackson's songs. As I am big fan of Eminem negative scores were as expected them to be. In addition, finding which were derived during the EDA analysis were useful especially number of curse words. Then, we have found that amount of curse words is not the main factor for sentiment being negative. Interesting fact that, all soul music have positive scores and most of them are close to opinion scores. Finally, we reached the goal, which was to apply NLP techniques to extract sentiment from song lyrics.

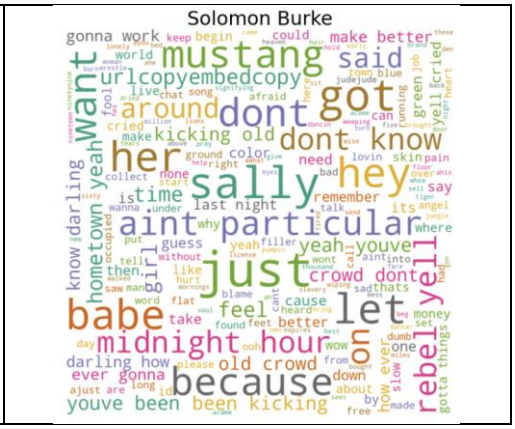
# QUAD PLOT


(Why is project being done)
Still no works on comparision different genre of musics between facts, opinions and negative, positive.
(Who is project for)
To better predict and recommend similar songs to music app users after they have listened to a song of a specific genre
(Goals)
To apply NLP techniques to extract sentiment from song lyrics.
(Measure success)
Song lyrics were collected from famous 'Genius' database, which sentiment analysis successfully was performed.

# EXPLORATORY DATA ANALYSIS

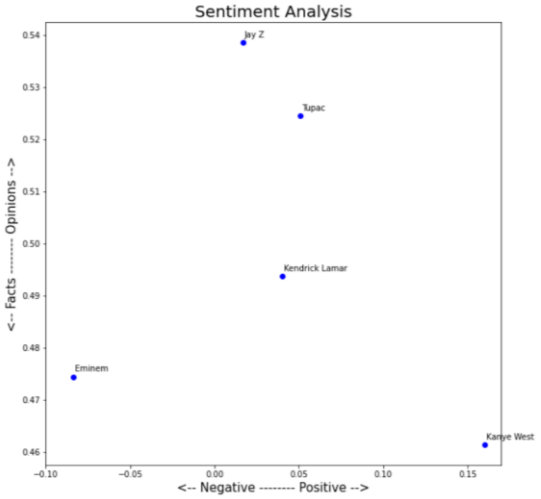
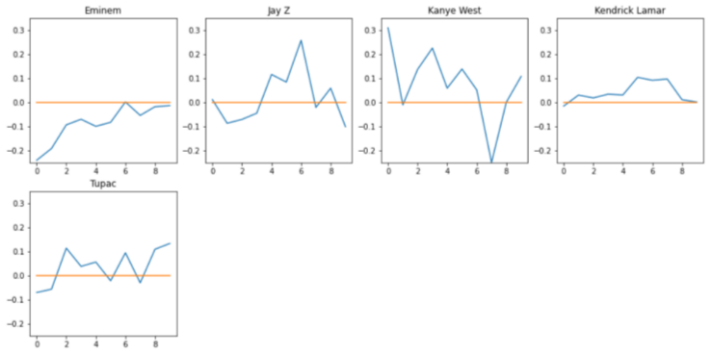
| RAP SONG | POP SONG | SOUL MUSIC |
|---|---|---|
|  | | |
| | | |
| | |  |
|  | |  |

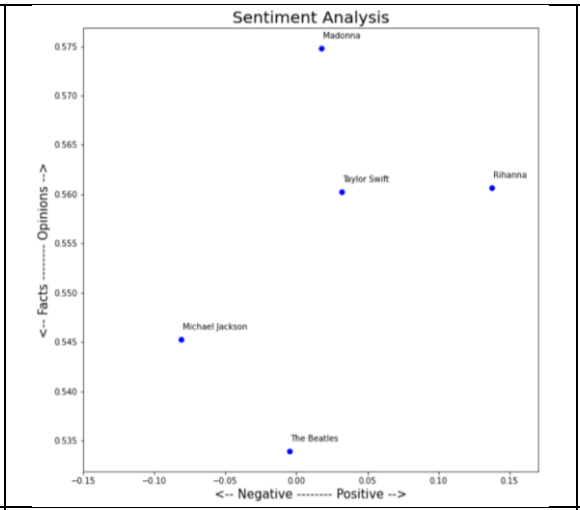
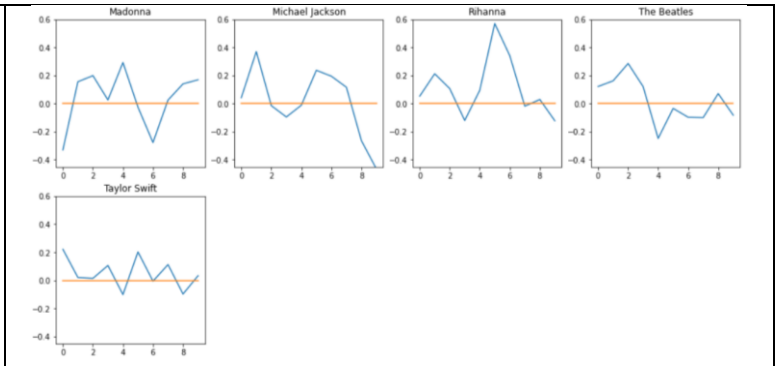
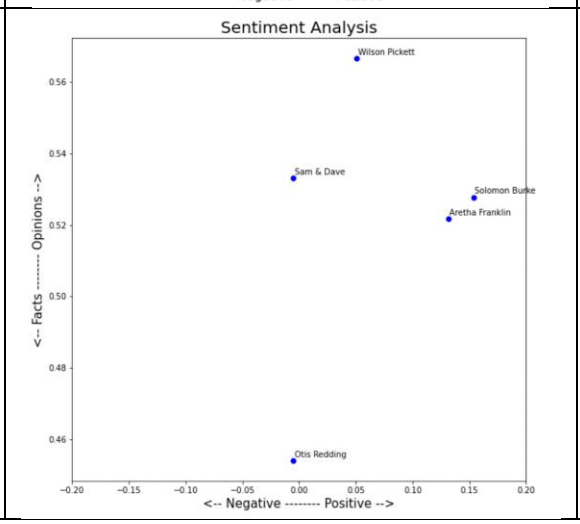
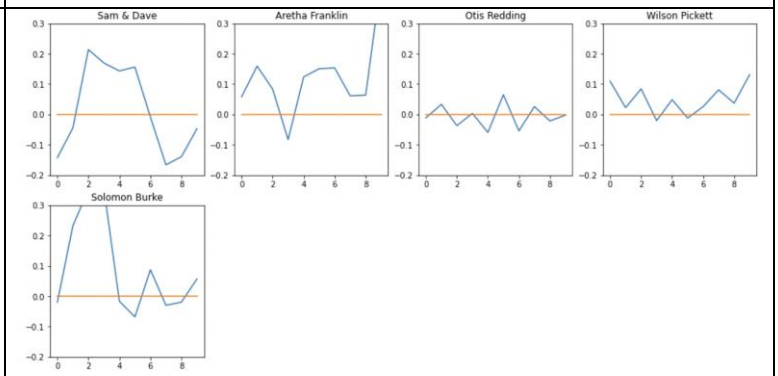



# SENTIMENTAL ANALYSIS

(facts-opinions to negative-positive)

| GENRE | SENTIMENTAL ANALYSIS | POLARITY PLOT |
|---|---|---|
| RAP SONG |  |  |

| | | |
|---|---|---|
| POP SONG |  |  |
| SOUL MUSIC |  |  |