# Classifying Artist's lyrics using Supervised Machine Learning

## Steps:

1. In this project I am extracting lyrics of 6 artists ['Eminem','Jay-Z','Justin-Timberlake','50-Cent','Bob-Marley','Michael-Jackson'] from lyrics.com (# of artist can be any length) and classify using Supervised Machine Learning Models.
2. In this project I have two python scripts ‘cli\_get\_lyrics.py’and‘Lyrics\_Classification\_RFC\_25Dec2021’
3. Libraries used:
   1. cli\_get\_lyrics.py (web scraping protocols)
      * + - Requests
          - re
          - os
          - pandas
          - argparser
          - sys
          - BeautifulSoup
   2. Lyrics\_Classification\_RFC\_25Dec2021.ipynb (ML model)
      * + - Pandas
          - Numpy
          - Sklearn
          - Nltk
          - String
          - re
          - os
4. For scraping lyrics please run 'cli\_get\_lyrics.py' on your command line interface.

Eg: type the below line on your command line interfaces with path where your python file is located.

* + - python cli\_get\_lyrics.py 6 --list Eminem Jay-Z Justin-Timeberlake 50-Cent Bob-Marley Michael Jackson
    - If you just run the file without giving any input arguments, then the code would function with its default values stored

default val: number of artist is "1" and name of the artist is "Akon".

* + - the file 'cli\_get\_lyrics.py' required two input arguments to function.
      * number of artists to be scrapped (in my case I have 6)
      * name of the artists as mentioned on the lyrics.com (--list Eminem Jay-Z Justin-Timeberlake 50-Cent Bob-Marley Michael Jackson)
    - Once scrapped, the program will store all the lyrics as .txt file in individual folder with artist names
      * eg: I scrapped 100 songs from Akon, so there will be a folder with name Akon and 100 lyrics stored in it.
    - the program takes cares of all the duplicate file name and repeated lyrics.
    - to abort the program in between press ctrl + c to interrupt.

1. Once all the lyrics are scrapped! Next, open Lyrics\_Classification\_RFC\_25Dec2021.ipynb
   * + this code required two input arguments:
       - folders = ['Eminem','Jay-Z','Justin-Timberlake','50-Cent','Bob-Marley','Michael-Jackson']
       - base\_url = r"Paste your file path here"
2. It is important to know that this model is optimized for the above artists only. Feel free to to optimize while using different artists.