Text Analysis approach to Predict Recommendation system

Roshan Shaik
rshaik2@gmu.edu
George Mason University

***Abstract*—We plan a novel technique for consequently creating a playlist of suggested tunes in the famous social music sharing application Spotify that are preferred with high likelihood by a client. Our strategy utilizes various seed specialists as an info that are acquired through any semblance of craftsmen and the listening history of melodies of a Spotify client. In the first place, we develop an information vector involving every one of the craftsmen that the client likes and tunes in to in Spotify. At that point, we look for different craftsmen and groups identified with them . We appoint a score to each craftsman in the along these lines got assortment, in light of the recurrence of his/her appearance. At last, we develop a playlist including arbitrarily chosen well known tunes related with the most as often as possible refered to craftsmen. We analyze the suggestion execution of our calculation by registering its WTF score (part of loathed melodies) and curiosity factor (part of new preferred tunes) on playlists produced for various seed input sizes.**

***Index Terms*—Machine Learning, Selection, Data, NLP**

I. Introduction

Spotify is a digital music, podcast, and video streaming service that gives you access to millions of songs and other content from artists all over the world. Basic functions such as playing music are totally free, but you can also choose to upgrade to Spotify Premium [1] [2] [3]. Either way, you can:

- Choose what you want to listen to with browse and search.
- Get recommendations from personalized features, such as discover weekly, release radar, and daily mix.
- Build collections of music.
- See what friends, artists, and celebrities listen to.
- Create your own Radio station.

Spotify is available across a range of devices, including computers, phones, tablets, speakers, TVs, and cars, and you can easily transition from one to another with Spotify Connect. (What is Spotify?, n.d.)

Spotify is one of the newest innovations to have come to audio listening and experience with over 125 million subscribers. Though the service has recently begun it dominates Apple Music and Amazon music in the audio streaming market. From music, they have extended the audio service to Podcasts, Audiobooks, and so on. Spotify Trends helps any content creator/musician in order to understand what listeners prefer and how to compete in this immensely growing market.

As the web moved from an owner model to an open freely supporting model and empowered people to contribute energetically, it saw a remarkable climb in the proportion of substance available, which was something to be appreciative [1] [2], [4], [5] [3]. Be that as it may, this incited two essential issues: Aggregation: The proportion of information ended up being enormous to the point that it motivated outrageous to direct it while at this point having the ability to run a web advantage that was reachable to all pieces of the world. This issue was handled by building generally substance movement and scattering frameworks, helped by the climb of NoSQL Database structures and reducing accumulating costs. Searching: The second huge issue was the methods by which to ensure that the information is inside the extent of the customer and that the customer doesn't get stirred up in the enormous data dumps open. This ended up being an altogether more concerning issue than gathering since the data stores are colossal and each customer conveys close by him/her an exceptional perspective and therefore a one of a sort pursuit plan. We are at this point endeavoring to deal with this issue today and are far from achieving an ideal response for it. This is the spot recommender systems become conceivably the main factor. Recommender frameworks need data for working, information about a specific client. This specific information can be brought straightforwardly or by implication. Straightforwardly gathering information implies that client of a specific assistance gives criticism and survey of the thing. By implication implies that framework will dissect the clients connection with the specific assistance comprising of history and present administration.

II. Building the recommendation system

We will build our recommendation system based on the genre we get and the distance from numerical features. The logics are:

1) We filter songs with the same genre cluster.
2) If there is only one song, we return this song.
3) If there is more than one song, we choose the one with smallest distance.

III. Song Prediction

Collaborative filtering and NLP models [5]–[15]can be used to suggest songs, but the same data can also determine if two songs are similar. By modeling the audio itself, Spotify can tell if two songs are similar in their very composition. It can parse out things like timbre, pitch, or a song's loudness curve. Spotify predicts your playlist by seeing what people like you enjoy, researching how the internet discusses songs you like,

and listening to songs that are musically comparable to your favorites. (Davis, 2020). we use Song list in amazon prime for collecting data recommendation [11], [16], [17].

IV. Literature Review

Playlists have become a significant part of our music listening experience today. There are over three billion of these on Spotify alone. There are playlists for every moment, every mood, every season, and so on. With millions of songs at their fingertips, users today have grown accustomed to:

- Immediate attainment of their music demands.
- An extended experience. While recommendation engines service the first aspect, playlists handle the second aspect of this changing behavior, making playlist recommendation extremely important, both for the users and music companies.

(i) Expansive taste: People whose melodic learning are particularly wide. They contributed 7 percent of absolute division.

(ii) Aficionados: There are part of individuals in this world who accepts that music is life, and they are insane for music. Undoubtedly, music is the most loosening up thing in this world.

(iii) Casual music audience members: People who nonchalantly tune in to music in their spare energy incorporate 32 percent of this division.

(iv) Unconcerned: They have distinctive mentality about music and including 40 percent of this age gathering. According to an examination each individual requires exceptional arrangement of ideas. Scholastics is incredibly pressing and are thusly the most inconvenient crowd individuals to give ideas to. They require perilous and keen proposition instead of renowned ones. Darlings on the other hand Lovers of course esteem an amicability between captivating, dark, and typical proposition. Casuals and unconcerned, who address 72% of the general population, don't need befuddled recommendations and popular standard music that they can without a very remarkable stretch identify with would oblige their melodic requirements.

There are playlists for every moment, every mood, every season, and so on. With millions of songs at their fingertips, users today have grown accustomed to:

- Immediate attainment of their music demands.
- An extended experience. While recommendation engines service the first aspect, playlists handle the second aspect of this changing behavior, making playlist recommendation extremely important, both for the users and music companies.

V. Questions we want to answer with our data here:

1) Which Attributes/features will likely lead a song to be more popular in Application?
2) What are the most Popular tracks which people are listening to?

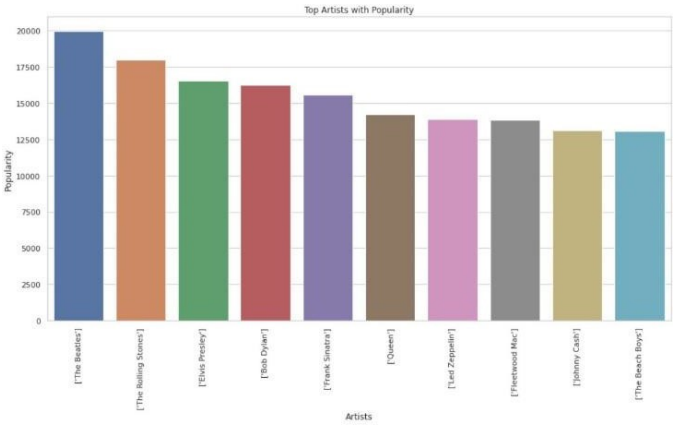

Fig. 1. **Top 10 songs by duration:**When checked for which song has the highest duration, we have sorted the top 10 songs by duration and found out that the song Brown Noise which has the highest duration of 90 minutes and followed by the song Brown Noise for Sleep stands second with the duration of around 70 minutes.

3) How to build a Song Recommendation System according to our interest?

VI. Proposed Approach

The first step is data cleaning and preprocessing for emotion and sentiment analysis [18], [19], [19]–[25] since they could affect the song play list. Data cleaning is utilized to allude to a wide range of undertakings and exercises to recognize and fix mistakes in the data. There are numerous sorts of blunders that exist in a dataset, albeit probably the easiest blunders incorporate segments that don't contain a lot of data and copied columns. Data preprocessing is a necessary advance in Machine Learning as the nature of information and the helpful data that can be gotten from it straightforwardly influences the capacity of our model to learn; thusly, it is critical that we preprocess our information prior to taking care of it into our model. The exploratory analysis will be performed on the data set to know the insights. By observing the insights, the feature selection will be made on the data set. When constructing an machine learning model, in actuality, it's practically uncommon that every one of the factors in the dataset are helpful to fabricate a model. Adding repetitive factors lessens the speculation capacity of the model and may likewise decrease the general precision of a classifier. Moreover, adding an ever increasing number of factors to a model expands the general intricacy of the model. We would like to build a model according to the observations such as decision tree and random forest, also the recommendation system.

VII. Methodology

We are planning on using various methods and techniques to implement analytical methods which analyzes the data and cleans it. Later, we will use visualizations to show the data and obtain various analysis to improve certain outcomes.

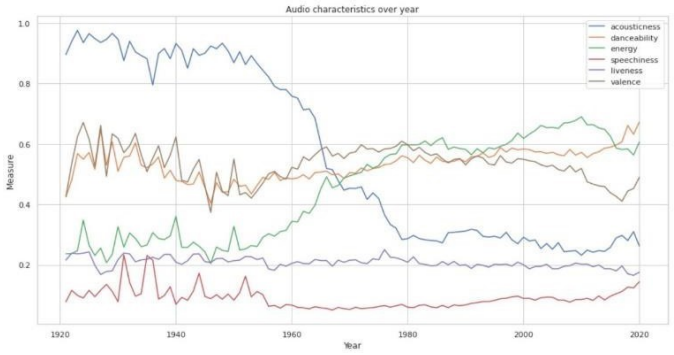

Fig. 2. **Top artists:** When the analysis is performed to know which artist is popular, we found out that Artist: The Beatles have the highest popularity followed by 'The Rolling Stones' and 'Elvis Presley'.

REFERENCES

[1] A. Miglani, "Coronavirus tweets nlp - text classification." https://www.census.gov/popclock/, 2020.

[2] Y. Song, S. Liu, and W. Ji, "Research on personalized hybrid recommendation system," in *2017 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 133–137, 2017.

[3] R. Gao, J. Li, B. Du, X. Li, J. Chang, C. Song, and D. Liu, "Exploiting geo-social correlations to improve pairwise ranking for point-of-interest recommendation," *China Communications*, vol. 15, no. 7, pp. 180–201, 2018.

[4] J. Wang, H. Song, and X. Zhou, "A collaborative filtering recommendation algorithm based on biclustering," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 803–807, 2015.

[5] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[6] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[7] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[8] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[9] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[10] S. Lin, C. Liu, and Z.-K. Zhang, "Multi-tasking link prediction on coupled networks via the factor graph model," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5570–5574, 2017.

[11] Y. Chu, F. Huang, H. Wang, G. Li, and X. Song, "Short-term recommendation with recurrent neural networks," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 927–932, 2017.

[12] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network to recommend airline tickets by using online reviews," in *IEEE 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2020*, 2020.

[13] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[14] Z. Liao, Y. Song, Y. Huang, L.-w. He, and Q. He, "Task trail: An effective segmentation of user search behavior," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3090–3102, 2014.

[15] C.-Y. Chi, Y.-S. Wu, W.-r. Chu, D. C. Wu, J. Y.-j. Hsu, and R. T.-H. Tsai, "The power of words: Enhancing music mood estimation with textual input of lyrics," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6, 2009.

[16] C. Yang, X. Chen, T. Song, B. Jiang, and Q. Liu, "A hybrid recommendation algorithm based on heuristic similarity and trust measure," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 1413–1418, 2018.

[17] M. Heidari, J. H. J. Jones, and O. Uzuner, "Misinformation detection model to prevent spread of the covid-19 virus during the pandemic," 2021.

[18] S. Ji and J. Liu, "Interpersonal ties and the social link recommendation problem," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 456–462, 2019.

[19] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[20] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[21] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[22] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[23] H. Yang, C. He, H. Zhu, and W. Song, "Prediction of slant path rain attenuation based on artificial neural network," in *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 152–155 vol.1, 2000.

[24] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[25] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.