Music Recommendation System using LSTM

Arvind Raman, Dhanush Raj, Prajakta More
Department of Computer Science.
University of Texas at Arlington

Abstract

In this paper, an enhanced Long Short-Term Memory (LSTM) neural network model for forecasting music popularity and producing personalized song recommendations is proposed. Our LSTM model performs exceptionally well using a broad dataset from Spotify that includes both textual (genre) and numerical (audio characteristics) data. The implemented method demonstrates how well combining the two kinds of information may result in a comprehensive music recommendation system. In the constantly changing world of music streaming services, the study describes the dataset, the design of the LSTM model, and shows its noteworthy accomplishments in predicting and suggesting popular songs.

1 Introduction

The emergence of music streaming services in the modern day has drastically changed how people consume music. in so many channels available and a vast song collection constantly expanding, streaming companies and artists alike are increasingly concerned in forecasting the popularity of music. This research explores the dynamic interplay between technology and music analytics, focusing on the effectiveness of a model based on Long Short-Term Memory (LSTM) in predicting song popularity in the setting of a large and varied dataset from Spotify.

This study is primarily motivated by the growing significance of predictive models in the music business, which is a result of the requirement to predict user preferences and increase the exposure of songs that connect with a variety of listeners. Because it is so good at capturing temporal correlations, the LSTM architecture is used to extract the complex patterns from music data, which helps with song popularity forecasts.

This study expands its scope to include the recommendation domain in addition to the prediction area. Acknowledging the mutually beneficial association between user comments and well-liked music, the study offers a method for selecting recommendations that closely correspond with user tastes. Through the incorporation of user feedback into the recommendation process, the model adjusts to the ever-changing preferences of users, resulting in a more customized and captivating music streaming encounter.

The robustness of the LSTM-based model is demonstrated by the evaluation measures used in this work. Superior results are found in terms of accuracy, precision, recall, and F1-score when compared to reference articles. This success highlights the model's ability to identify subtleties in the Spotify dataset, which makes it a useful tool for services looking to improve song recommendations and predict user engagement patterns.

Essentially, this study fills a requirement that is relevant to the music industry as well as adding to the growing subject of music analytics. Accurately predicting and recommending songs becomes essential as music streaming services fight for users' attention. The study's conclusions offer insightful information to platforms and artists who want to improve their comprehension of audience preferences and streamline their methods for selecting and curating material. The LSTM-based model's versatility and predictive power make it a potential tool in the competitive and dynamic world of music streaming services, especially as the landscape of music consumption changes.

2 Dataset Description

With its extensive collection of music-related variables, the Spotify dataset utilized in this work offers a solid basis for popularity prediction and customized recommendation generation. The essential characteristics consist of:

artist_name: The identity of the artist who composed the piece.

track_name: The track's title, which serves as a unique identification.

track_id: A unique number assigned to every song on the Spotify service.

popularity: A number that represents the track's popularity and is used as the prediction's target variable.

year: The track's year of release, which adds to the dataset's temporal component.

genre: The musical genre, or classification, providing useful textual data for the model.

danceability: A musical element-based assessment of a track's suitability for dancing.

energy: The track's energy level, reflects its activity and intensity.

key: The track's key signature, which sheds light on the musical arrangement of the piece.

loudness: The track's total loudness is expressed in decibels, which affects how loud it is perceived to be.

mode: Indicates the track's modality, highlighting major and small differences.

speechiness: Indicates how many spoken words are in a track and provides information about the vocal content.

acousticness: How much of the song is made up of acoustic components rather than electronic noises.

instrumentalness: An assessment of a track's instrumentality based on the lack of vocals.

liveness: Captures the dynamic quality of the song and suggests that it was performed live.

valence: Sums up the track's good musical aspects, which can range from mournful to exuberant.

tempo: The beats per minute (bpm) of the track, which affects how quickly and rhythmically it is perceived.

duration_ms: The track's temporal length expressed in milliseconds, indicating its duration.

time_signature: Adds to the track's rhythmic framework by indicating how many beats there are in each bar.

Our LSTM-based music recommendation engine is built on this varied combination of textual and numerical information from the dataset, which allows for precise popularity predictions and a more sophisticated knowledge of each tune.

3 Project Description

The main goal of this research project is to create and assess a novel Long Short-Term Memory (LSTM) neural network model that may be used to forecast music popularity and provide tailored song recommendations. The model incorporates textual and numerical information that are taken from the large Spotify dataset. These elements include popularity, genre, track title, artist name, and other audio properties. A comprehensive approach to music analytics is provided by the LSTM architecture, which is made to process textual (genre) and numerical (audio characteristics) data simultaneously. The model is trained across ten epochs once the dataset has been preprocessed and divided into training and validation sets. The resulting model shows remarkable recall, accuracy, precision, and F1-score metrics, demonstrating its effectiveness in predicting well-known songs.
Furthermore, a distinctive feature of the project is that it uses user input to provide suggestions for the best songs. This shows how adaptable the model is to accommodating user preferences in the ever-changing world of music streaming platforms.

3.1 Description

With its many layers, the Long Short-Term Memory (LSTM) model is carefully designed to handle both textual (genre) and numerical (audio characteristics) input at the same time, resulting in a single forecast of music popularity. The model is optimized through 10 training epochs on a split dataset using the Adam optimizer and mean squared error loss, maximizing its ability to identify patterns and correlations. The model's solid predictive performance in a dynamic music streaming landscape is attributed to its capacity to capture complex features of music through independent processing of textual and numerical data.

3.2 Main references used for your project

The main reference guiding our project is the insightful study "Emotion-Based Music Recommendation System Using LSTM - CNN Architecture". It's work examines the changing field of music recommendation systems in the modern media and technological landscape, with a special emphasis on the complex relationship between music and emotions. The authors investigate a number of models, such as CNN-LSTM, LSTM-CNN Architectures, Long Short-Term Memory (LSTM), and Convolution Neural Network (CNN), by combining Natural Language Processing with Deep Learning approaches. Their focus on emotion detection—which includes feelings like grief, love, wrath, and happiness—is in line with our main goal of comprehending consumer preferences. Moreover, the use of facial expression analysis via a CNN model highlights their inventive methodology. This reference strengthens the foundation of our project and influences our own LSTM-based music popularity forecast and recommendation system by offering priceless insights into utilizing deep learning for emotion-based music suggestions.

3.3 Difference in APPROACH/METHOD between your project and the main projects of your references

Our experiment takes a different approach by focusing on the prediction of music popularity, whereas the reference work primarily uses a combination of LSTM and CNN architectures for emotion-based music suggestions. Our LSTM model provides a comprehensive view of song properties by simultaneously processing textual (genre) and numerical (audio aspects) input. Our model's adaptability is demonstrated by its ability to predict binary popularity, in contrast to the reference's emotive categorization. Furthermore, our model integration captures a wider range of elements impacting music liking, going beyond emotions. This shift in emphasis and use of features highlights our special technique for using LSTM for music analytics, making a noteworthy contribution to the discipline.

3.4 Difference in ACCURACY/PERFORMANCE between your project and the main projects of your references

Our model outperforms the LSTM-CNN model in the reference paper in a comparison test, obtaining an astounding 95.56% versus their stated 93.88%. Our method shows improved recall, precision, and F1-score metrics, confirming our LSTM-based music popularity model's higher predictive power. The robustness of our model's performance was enhanced by the notable extension of its training to 10 epochs. The difference between our results and those of the cited LSTM-CNN model, which was trained across five epochs, highlights the effectiveness of our method in identifying complex patterns within the music dataset and yields more precise forecasts of song popularity.

4 Analysis

4.1 What did we do well?

Model Performance: Exhibiting remarkable accuracy (95.56%), precision (93.91%), recall (97.34%), and F1-score (95.59%), the LSTM-based music popularity prediction and recommendation system is quite successful. The combination of textual and numerical information improves the model's capacity to represent a variety of musical elements.
Creative Combination: One noteworthy strength that contributes to the model's strong prediction capacity is the combination of LSTM with both numerical and textual input. The architecture of the model and the addition of different characteristics are in line with the complexity of the dataset.The model's exceptional performance can be ascribed to the careful incorporation of many features and the ten training epochs, which enable the model to identify subtle patterns in the data.

4.2 What could we have done better?

Fine-tuning Hyperparameters: More research into fine-tuning these parameters may help the model function at its best. Changing up the optimizer, learning rate, and dropout rate may reveal setups that improve generalization and accuracy.
Handling Skewed Data: Using methods like oversampling or under sampling could enhance model performance on minority classes if the dataset shows class imbalance in popularity labels.

4.3 What is left for future work?

Enhancement of User Interaction: Adding real-time user comments and preferences to the recommendation system could make it more personalized.
Dynamic Model Training: To guarantee adaptability to new musical trends, techniques for ongoing model training on dynamic datasets should be put into practice.
Cross-Dataset Generalization: Evaluating the model's effectiveness on other datasets outside Spotify may confirm its ability to generalize across various musical settings.
Interpretable Models: Creating techniques to decipher model forecasts and comprehend how particular attributes affect song popularity will improve openness and user confidence.

5 Conclusion

In conclusion, this study offers a strong LSTM-based system for predicting and recommending song popularity that uses both textual and numerical variables from a sizable Spotify dataset. The model's remarkable recall, accuracy, precision, and F1 score highlight how well it captures the nuances of musical preferences. Its capacity for prediction is enhanced by the creative way in which LSTM is integrated with various elements. Despite the noteworthy achievement, there is room for improvement, particularly in resolving class imbalance and investigating hyper parameter adjustment for additional optimization. The study represents a substantial advancement in music analytics by providing a flexible method for popularity prediction and tailored recommendations.
The model is positioned as a valuable asset in the dynamic convergence of technology and music consumption, given its adaptability and potential for user-centric enhancements as the music streaming industry continues to change. Upcoming projects seek to improve user engagement, interpretability, and cross-dataset generalization in order to provide a more thorough and sophisticated knowledge of music recommendation algorithms.

References

[1] M. G. Galety, R. Thiagarajan, R. Sangeetha, L. K. B. Vignesh, S. Arun and R. Krishnamoorthy, "Personalized Music Recommendation model based on Machine Learning," *2022 8th International Conference on Smart Structures and Systems (ICSSS)*, Chennai, India, 2022, pp. 1-6, doi: 10.1109/ICSSS54381.2022.9782288.

[2] J. S. Joel, B. Ernest Thompson, S. R. Thomas, T. Revanth Kumar, S. Prince and D. Bini, "Emotion based Music Recommendation System using Deep Learning Model," *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023, pp. 227-232, doi: 10.1109/ICICT57646.2023.10134389

[3] S. Joshi, T. Jain and N. Nair, "Emotion Based Music Recommendation System Using LSTM - CNN Architecture," *2021 12th International Conference on Computing*

Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 01-06, doi: 10.1109/ICCCNT51525.2021.9579813

[4] M. Jiang, Z. Yang and C. Zhao, "What to play next? A RNN-based music recommendation system," *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA

[5] J. Singh, M. Sajid, C. S. Yadav, S. S. Singh and M. Saini, "A Novel Deep Neural-based Music Recommendation Method considering User and Song Data," *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2022, pp. 1-7, doi: 10.1109/ICOEI53556.2022.9776660.