VOICE-TO-TEXT-BASED MUSIC RECOMMENDATION USING LSTM MODEL

[1]Mr P. RADHAKRISHNAN, [2]MULUKUNTLA NAVYA SRI, [3]JANGA SRI CHANDANA, [4]MOOLA KARTHIK, [5]GADDAM SHIVA KUMAR, [6]CHINTHIREDDY SANATH REDDY

[1.] Assistant Professor, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: rksiva13@gmail.com

[2.] B. Tech Student, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: navyasrinavya702@gmail.com

[3.] B. Tech Student, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: srichandanajanga@gmail.com

[4.] B. Tech Student, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: karthikmoola143@gmail.com

[5.] B. Tech Student, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: shivagaddam.sr@gmail.com

[6.] B. Tech Student, Department of Computer Science and Engineering, SR University, Ananthasagar, Warangal Urban (TS). India.

Email-: sanathchinthireddy@gmail.com

ABSTRACT

A floaty, newfangled manner to listen to music is done by this research using a Long Short-Term Memory (LSTM) model incorporated with a Voice-to-Text system. The system changes the user suggestions into text through speech recognition, after which it uses the LSTM model to detect the mood of the user. Depending on the predicted mood, we query a database of songs associated with different moods and recommend suitable ones to play back to an individual. Also, Pygame has been integrated into it for sound reproduction and offers seamless interaction through a friendly GUI (Graphical User Interface). It provides empirical evidence that show how efficient this proposal can be in accurately predicting users' feelings and recommending them appropriate music. User experience improves from including voice inputs as they provide easy ways of interacting with music recommendation systems. This study advances personalized music recommendation systems and advocates combining voice recognition techniques with machine learning approaches in order to improve users' engagement in musical applications.

I. INTRODUCTION

Human emotions and wellbeing are impacted by music and also bring about changes in moods, behaviours and even physiological responses. The challenge of properly recommending music that resonates with individual preferences has become more important with the advancement of digital music libraries and streaming platforms. Traditional systems for recommending music rely on user-interactions such as ratings and browsing history to generate personalized recommendations. However, it often fails to capture the immediate mood or emotional state of the user.

This limitation can be addressed by integrating voice recognition technology into machine learning models so as to enhance music recommendation systems. By utilizing speech recognition algorithms, users will communicate with natural language to the system hence making it more intuitive and seamless. Moreover, machine learning models such as Long Short-Term Memory (LSTM) networks have been shown to be effective in analyzing sequential data as well as capturing temporal dependencies thus making them suitable for modeling human emotions dynamic nature.

In this research, we propose a Voice-to-Text based Music Recommendation System that uses LSTM model to predict user mood through their spoken input. The system understands the text transcript of the user's speech to guess what emotion is being felt at the time and suggests music correspondingly. Through combining speech recognition, machine learning, and audio playback functionalities; we intend to deliver an individualized and participatory music recommendation encounter.

Key contributions include:

- Development of a voice-to-text interface for music interaction, where users can express their preferences using natural language.
- Applying LSTM model trained on textual data to predict user mood from spoken input thus improving system's ability to provide personalized recommendations.
- Integration of a song database categorized by mood enabling it to recommend songs based on current emotional situation.
- Evaluating the system's performance in terms of empirical studies on mood prediction accuracy and user satisfaction.

As a whole, this research project seeks to prove that voice recognition and machine learning techniques can be used to improve music recommendation systems. Our system aims at allowing users to express themselves through preferences so as to come up with more engaging and personalized music discovery experience.

II. LITERATURE SURVEY

1. A Survey of Music Recommendation Systems.

Author: Xavier Serra and Jordi Janer

This comprehensive survey affords a top level view of numerous methods and strategies hired in song advice systems. The authors classify recommendation strategies into content material-based totally, collaborative filtering, and hybrid structures, discussing their benefits and barriers. They additionally discover emerging tendencies inclusive of context-aware recommendation and the integration of social and cultural factors into music advice algorithms. The survey serves as a valuable aid for expertise the evolution of track recommendation structures and figuring out areas for future research and development.

2. **Deep Learning Techniques for Music Generation.**

Author: Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang

This survey investigates the software of deep mastering strategies in tune era responsibilities, focusing on generative fashions along with recurrent neural networks (RNNs), convolutional neural networks (CNNs), and generative antagonistic networks (GANs). The authors offer an overview of different deep learning architectures and methodologies used for generating song, discussing their skills in capturing temporal dependencies and high-degree musical structures. Furthermore, they look at challenges and possibilities inside the subject, including problems related to statistics scarcity, model interpretability, and the combination of person choices in music era structures.

3. **Music Emotion Recognition.**

Author: Yi-Hsuan Yang, Homer H. Chen, and Yeh-Shen Chen

This evaluate article surveys the research landscape of track emotion recognition, that specialize in techniques and methodologies for automatically studying and categorizing the emotional content material of tune. The authors speak various features used for representing musical emotions, which include acoustic functions, lyrics, and symbolic representations. They also evaluation extraordinary machine studying algorithms and models hired in music emotion popularity responsibilities, which include aid vector machines (SVMs), decision timber, and deep neural networks. Additionally, the overview highlights challenges in the discipline, which include the subjective nature of emotional perception and the shortage of standardized datasets for comparing emotion reputation algorithms.

4. **Voice User Interface (VUI) Design.**

Author: James Giangola, Jennifer Balogh, and Michael H. Cohen

This e-book presents a comprehensive manual to designing voice user interfaces (VUIs) for interactive systems, protecting principles, methodologies, and fine practices in VUI layout. The authors discuss essential concepts including speech reputation, herbal language understanding, and conversation control, presenting insights into designing intuitive and user-pleasant voice interfaces. The ebook additionally

addresses realistic concerns in VUI improvement, consisting of usability trying out, blunders managing, and designing for various user demographics. With case research and real-international examples, this e-book serves as a valuable resource for designers and builders seeking to create powerful voice-driven applications and structures.

5. Recurrent Neural Networks for Audio Recognition.

Author: Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng

This paper explores the software of recurrent neural networks (RNNs) for audio popularity duties, which includes speech popularity and track category. The authors introduce a singular structure referred to as Deep Speech, which utilizes RNNs with deep bidirectional LSTM layers to at once transcribe speech audio into text. They speak the layout alternatives and schooling techniques employed in Deep Speech, demonstrating its effectiveness in accomplishing present day overall performance on speech recognition benchmarks. Additionally, the paper highlights the potential of RNNs in handling sequential statistics and modeling temporal dependencies in audio indicators, paving the way for advancements in audio recognition technology.

III. EXISTING SYSTEM:

The present day panorama of tune advice systems encompasses a lot of methods and technologies aimed at helping customers in coming across tune that aligns with their choices and temper. Traditional advice systems usually rely upon collaborative filtering strategies, which examine consumer interactions which includes ratings, playlists, and listening records to generate customized guidelines. These structures often suffer from the bloodless-begin hassle, wherein new users or objects have restrained statistics to be had for correct recommendation.

To deal with the limitations of collaborative filtering, content-primarily based recommendation systems leverage capabilities extracted from song metadata, together with genre, artist, and tempo, to suggest similar objects to the ones previously favored by using the user. However, content-based totally techniques may additionally conflict to seize the

nuanced characteristics of song that make a contribution to emotional responses and mood.

In latest years, hybrid recommendation systems have emerged as a promising solution, combining collaborative filtering and content-based totally methods to enhance advice accuracy and coverage. These structures leverage both consumer choices and item characteristics to provide greater strong and diverse guidelines.

Furthermore, advancements in system studying and herbal language processing have paved the manner for revolutionary strategies to song recommendation. Voice-controlled interfaces, powered with the aid of speech reputation algorithms, allow customers to engage with advice systems the usage of natural language, improving person enjoy and accessibility. Additionally, deep learning techniques, which include convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have proven promise in modeling complex patterns in music statistics and generating customized pointers based on audio functions and user options.

Despite these improvements, demanding situations stay in growing tune advice systems that as it should be seize person possibilities, adapt to evolving tastes, and offer various and serendipitous hints. Future research efforts may additionally recognition on addressing issues along with statistics sparsity, model interpretability, and incorporating contextual records to further decorate the overall performance and usefulness of track advice structures.

IV. DISADVANTAGES OF EXISTING SYSTEM:

The existing tune advice structures face several negative aspects. Firstly, collaborative filtering methods are prone to the cold-start problem, struggling to offer accurate pointers for brand spanking new users or gadgets with restrained statistics. Secondly, content material-based totally processes may additionally fail to capture the nuanced emotional attributes of tune, main to much less personalised guidelines based totally totally on metadata. Additionally, hybrid systems, whilst combining collaborative filtering and content material-primarily based strategies, may additionally still be afflicted by

scalability problems and shortage adaptability to changing consumer choices. Moreover, traditional structures frequently depend upon express consumer remarks, which can be sparse and biased, restricting the effectiveness of recommendation algorithms. Lastly, the lack of interpretability in a few machine getting to know fashions hinders users' understanding of why sure recommendations are made, lowering accept as true with and person engagement.

IV. PROPOSED SYSTEM:

The proposed device ambitions to revolutionize tune advice through integrating voice recognition technology with a Long Short-Term Memory (LSTM) version. This progressive approach lets in users to engage with the advice machine the usage of natural language, converting spoken input into textual content transcripts. The LSTM model then analyzes those transcripts to are expecting the consumer's mood or emotional nation, leveraging the version's capability to capture temporal dependencies and contextual data. Based at the predicted mood, the device recommends song from a database classified via temper, enhancing the personalization and relevance of hints.

Key features of the proposed system consist of:

Enhanced User Interaction: By allowing voice-primarily based enter, the device offers a more intuitive and seamless interaction revel in, putting off the need for guide textual content input or selection.

Mood Prediction with LSTM: Leveraging the capabilities of LSTM models, the gadget correctly predicts the consumer's mood from their spoken input, taking into account greater customized and context-conscious guidelines.

Mood-Based Music Recommendations: Utilizing a database of songs labeled by temper, the device recommends tune that aligns with the person's emotional nation, enhancing the relevance and engagement of pointers.

Real-time Recommendation: The system presents actual-time guidelines, permitting customers to acquire personalized song recommendations right away based totally on their current temper or preferences.

Integration of Audio Playback: Seamless integration with audio playback capability permits customers to concentrate to encouraged music at once in the utility, enhancing the general person revel in.

Overall, the proposed device gives a singular and efficient approach to song advice, leveraging voice reputation and LSTM modeling to provide personalized and contextually applicable hints tailor-made to the user's temper. By bridging the space between user expression and track recommendation, the system pursuits to beautify user pleasure and engagement in tune discovery and intake.

VI. ADVANTAGES OF PROPOSED SYSTEM:

The proposed Voice-to-Text Based Music Recommendation System making use of an LSTM version gives several blessings:

Natural Interaction: By incorporating voice recognition era, the device permits users to engage with the advice machine the usage of natural language, enhancing usability and accessibility.

Personalized Recommendations: Leveraging the LSTM version's capability to capture temporal dependencies and contextual data, the machine predicts the user's temper from spoken input, permitting personalized music guidelines tailor-made to the consumer's emotional kingdom.

Context-Aware Recommendations: The device considers the consumer's cutting-edge temper in recommending music, supplying context-conscious recommendations that align with the person's emotional possibilities at the time of interaction.

Real-Time Response: The system gives real-time guidelines, permitting users to receive personalised music guidelines immediately based totally on their spoken input, improving consumer engagement and pride.

Enhanced User Experience: Seamless integration with audio playback functionality permits customers to pay attention to recommended song immediately inside the software, supplying a cohesive and immersive user enjoy.

Adaptability and Learning: LSTM fashions can adapt to evolving person alternatives through the years, constantly enhancing the accuracy and relevance of guidelines as greater information is accumulated and analyzed.

Reduced Manual Effort: The device gets rid of the need for guide textual content input or selection, streamlining the advice procedure and decreasing consumer attempt in coming across new music.

Innovative Approach: The integration of voice popularity with LSTM modeling represents an revolutionary approach to music advice, showcasing the capacity of mixing natural language processing and system mastering techniques for better consumer interaction and personalization.

VII. RESULTS:

Fig.7.1 This is GUI where we give input and get output.

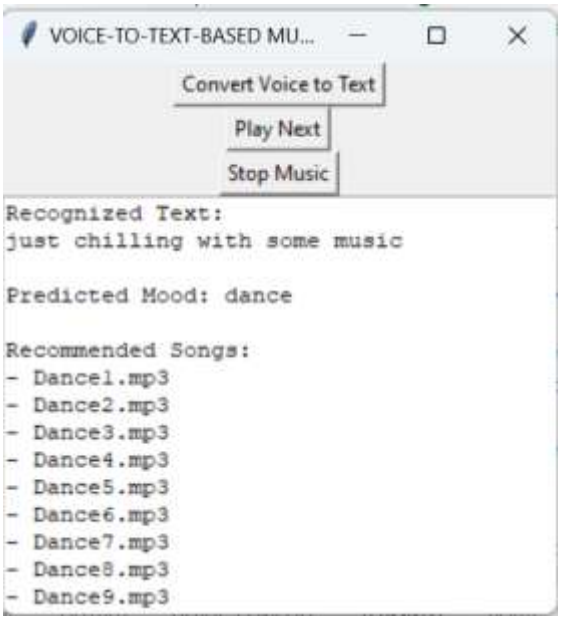

Fig.7.2 Click Convert voice to text button to give input and then input your voice. Then it predicts mood and recommends songs.

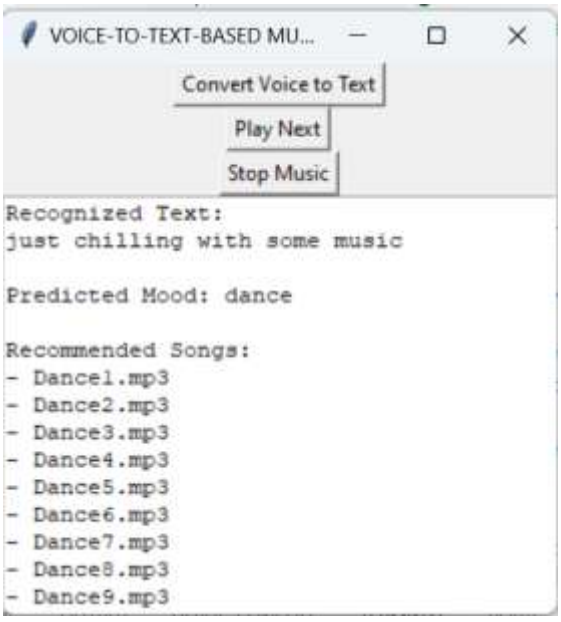

Fig.7.3 If you want to play next song, then click PLAY NEXT button.

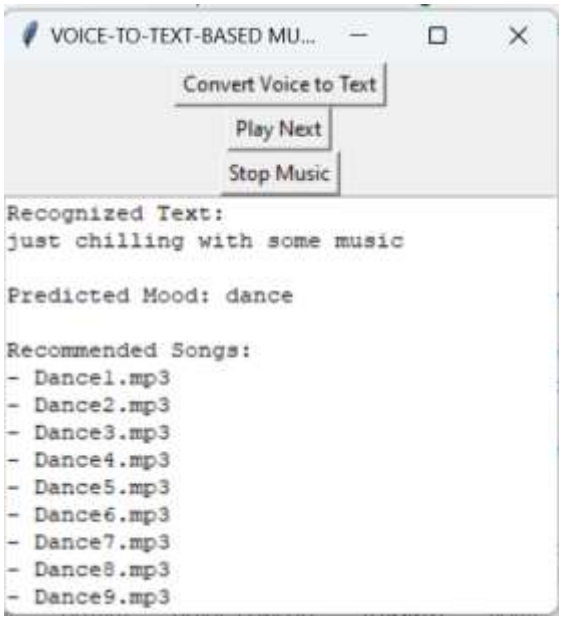

Fig.7.4 If you want to stop the music while it is playing, then click STOP MUSIC button.

VIII. CONCLUSION

In end, the proposed Voice-to-Text Based Music Recommendation System using an LSTM version represents a promising development in the discipline of track advice. By leveraging voice recognition generation and LSTM modeling, the device gives a natural and intuitive interface for users to have interaction with the recommendation system, allowing customized and context-aware song recommendations based totally at the consumer's mood. With its real-time reaction, adaptability, and seamless integration of audio playback, the system enhances the consumer revel in, streamlines the advice procedure, and reduces guide effort. Overall, the proposed gadget showcases the ability of combining voice reputation and system learning strategies to create innovative and user-centric answers in music discovery and intake.

IX. REFERENCES

[1]. A. S. Abdullah and R. R. Rajalaxmi, ‘‘A data mining model for predicting the coronary heart disease using random forest classifier,’’ in Proc. Int. Conf.Recent Trends Comput.Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[2]. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ‘‘Using PSO algorithm for producing best rules in diagnosis of heart disease,’’ in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.

[3]. N. Al-milli, ‘‘Backpropogation neural network for prediction of heart disease,’’ J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[4]. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and

R. Deepika, ‘‘Analysis of neural networks based heart disease prediction system,’’ in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

[5]. P. K. Anooj, ‘‘Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27– 40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002

[6]. L. Baccour, ‘‘Amended fused TOPSISVIKOR for classification (ATOVIC) applied to some UCI data sets,’’ Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[7]. C.-A. Cheng and H.-W. Chiu, ‘‘An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-widedatabase,’’ in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566– 2569

[8]. H. A. Esfahani and M. Ghazanfari,

‘‘Cardiovascular disease detection using a new ensemble classifier,’’ in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.