Developing Music Recommendation System by Integrating an MGC with Deep Learning Techniques

Md. Omar Faruk RIAD
BGC Trust University

Subhasish GHOSH
BGC Trust University

Abstract: In the modern scenario, everyone uses the internet to find music, movies, products, services and other commodities on a regular basis to make their lives easier. Because of a lot of data on millions of music, movie, products and services on websites, we need a recommender system very much to assist people in making decisions more quickly and easily. In this research study, we have developed an intelligent music recommendation system by integrating a Music Genre Classification (MGC) with different types of Deep Learning Techniques such as RNN-LSTM, GRU and CNN. We have used the GTZAN Genre dataset to training our system. We have extracted the features from GTZAN dataset by the help of Mel Frequency Cepstral Coefficients (MFCCs) then pass the MFCCs into our deep learning networks. After classifying the appropriate music genre, recommended the music from particular genre from the labelled database which has been classified by our system. From our proposed models the GRU, CNN and RNN-LSTM produced about 71%, 72% and 74% respectively in our testing accuracy. The RNN-LSTM has achieved the best accuracy result (74%) among all of our proposed models.

Keywords: Music recommendation system, MFCCs, CNN, GRU, RNN-LSTM

Introduction

With the widespread use of the internet, the music industry has seen tremendous change due to the internet, along with other types of development. Due to the development of music streaming services, consumers can now listen to music anytime, anywhere, and through a variety of platforms like Spotify, YouTube, Soundcloud, and many others. A major amount of the world's population now regularly downloads and buys music through online music stores. Users frequently categorize their tastes by genre, such as rock, metal, blues, hip hop, pop, or disco etc. But the majority of the songs that are now accessible are not automatically classified into genre. These systems are primarily focused on the classification of music genres for the purposes of mining music listening and music tagging, music recommendations to boost sales profits, and music copyright management for writers. Particularly through music streaming and broadcasting services like Spotify, last FM, Fizzy, etc., consumers have access to millions of songs at any time, from anywhere. Music Genre classification is crucial for organizing, searching, retrieving, and recommending music due to the huge amount of current collections.

Due to the gradual reduction in the complexity of music production in recent years, a large number of people now produce music and upload it to streaming services. People now spend a lot of time searching for particular music because of the massive music streaming industry. So, the ability to quickly classify musical genres is crucial in today's society. So, we want to build a Music Genre Classification System (MGCS) and its recommendation system so the music mining music listening and music tagging, music recommendations to boost sales profits, and music copyright management for writers.

So far, several researchers have built variety of Music Recommendation Systems in a variety of approaches. We are now going to discuss some of the researches about Music Recommendation System. An automated machine learning model that recommends potential candidate's resumes to the HR department based on the provided job description has been proposed by Roy et al. (2020). The suggested approach operated in two stages: first, it divided the resumes into various groups. Second, it suggests resumes based on how well their contents match the job description. The suggested method accurately captured resume insights and their semantics, and it produced a Linear SVM classifier accuracy of 78.53 % (Roy et al., 2020). Phaneendra et al. (2022) present a similarity metric that is simply based on raw audio content. Using this metric, songs that are similar to the user's collection are then recommended. A Siamese Neural Network (SNN) was trained on a dataset of similar and dissimilar song pairs to produce the similarity metric. To create a bitmap representation of each song, a Mel-Spectrogram was initially created. The Mel-Spectrogram bitmap of each song pair is sent to two identical CNNs that make up the SNN. The SNN gains the ability to function as a similarity metric between songs based on the raw audio information by being trained on this dataset of song picture pairs (Phaneendra et al., 2022). Using convolutional neural networks (CNN) as the foundation, Chang et al. (2018) offer the PMRS, or customized music recommendation system. The CNN method divides music into several genres based on the audio signal rhythms. The output of the CNN and the log files are combined in PMRS to create a collaborative filtering (CF) recommendation system that suggests music to the user. The history of each user who uses the PMRS is recorded in the log file. From the log file, the PMRS extracts the user's history and provides genre-specific music recommendations. They assess the PMRS using the million-song dataset (MSD). They created a mobile application to demonstrate how the PMRS functions (an Android version). To evaluate the effectiveness of the PMRS, they used the confidence score metrics for various musical genres(Chang et al., 2018). By examining the history records of user tracks in the data sets, Wang et al. (2018) technique RTCF (Tag-driven Collaborative Filtering Recommendation System) authors created a fair scoring system for users based on the training statistical language model (Good-Turing Estimate). Furthermore, they discovered that each tag has a linked count when describing a song or an artist, meaning that various tags can define the tracks and the artist's attributes to differing degrees. The tags work in conjunction with their analysis of each user's listening history to not only widen the music's content but also create individualized music recommendations for each user (Wang et al., 2018). A method for making branching playlists that allow music recommendations based on listener preferences is proposed. They discovered that branching playlists raised listeners' levels of contentment, familiarity, and interest by analyzing the method's efficacy. They also put the suggested technique into practice by implementing "reco.mu," a web-based music recommendation system. They discovered that while making a branching playlist, the creator is more aware of the recommender (Nonaka & Nakamura, 2021).

The key challenge is to recommend any things to users that they will enjoy and grade highly to match their expectations. Many researchers process of a recommender system makes use of data mining techniques like clustering to group related items together in order to find similarities between them when we choose an item as the centroid of the cluster or a user as the centroid of the cluster, as well as association rules to find hidden patterns and uncover new connections between products in order to boost sales as part of e-commerce. There are various methods used by recommender systems:

a) Content-based filtering: this type of filtering relies on examining the textual data's content and identifying patterns in the specifications of the items.
b) Collaborative filtering techniques: Depending on how similarly users rate products on the site, additional users may be recommended to like the person.
c) Hybrid collaborative filtering combines content-based filtering and collaborative filtering in order to maximize benefits and produce the best recommended items.

When Music Genre Classification (MGC) comes to human-machine interaction, it becomes a major topic. Due to the selection and extraction of appropriate audio elements, classifying music is considered as a very difficult process. While unlabeled data is readily available music tracks with appropriate genre tags is very less. Feature extraction and categorization are the two fundamental phases in the classification of musical genres. At 1st stage, the waveform is initially processed to extract various features. At 2nd stage, the features generated from the training data are used to construct a classifier. The features to be extracted from the music have been determined as zero crossing rate, spectral centroid, spectral contrast, spectral bandwidth, spectral rolloff and Mel-frequency Cepstral Coefficients-MFCC.

In this research study, it is aimed to develop an Intelligent Music Recommendation system by integrating a Music Genre Classification (MGC) with different types of Deep Learning Techniques such as RNN-LSTM (Recurrent Neural Networks-Long-Short Term Memory), GRU (Gated Recurrent Unit) & CNN (Convolutional

Neural Network) and we will also compare the experimental results for each of these from our experimentation. We are using the GTZAN Genre Dataset training our system based on genres.

Literature Review

An automated machine learning model that recommends potential candidate's resumes to the HR department based on the provided job description has been proposed by Roy et al. (2020). The suggested approach operated in two stages: first, it divided the resumes into various groups. Second, it suggests resumes based on how well their contents match the job description. The suggested method accurately captured resume insights and their semantics, and it produced a Linear SVM classifier accuracy of 78.53 %. Utilizing deep learning models such as Convolutional Neural Network, Recurrent Neural Network, Long-Short Term Memory, and others may improve the performance of the model. The proposed approach can be used to create an industry-specific model if a large number of resumes are provided by the industry. A more accurate model might be created by including domain experts like HR professionals, and the HR professional's feedback could then be used to iteratively enhance the model (Roy et al., 2020).

This study solves the problem of recommending new songs and artists by basing recommendations entirely on the audio content rather than extraneous metadata like the artist or genre or user listening histories. The authors Phaneendra et al. (2022) present a similarity metric that is simply based on raw audio content. Using this metric, songs that are similar to the user's collection are then recommended. A Siamese Neural Network (SNN) was trained on a dataset of similar and dissimilar song pairs to produce the similarity metric. To create a bitmap representation of each song, a Mel-Spectrogram was initially created. The Mel-Spectrogram bitmap of each song pair is sent to two identical CNNs that make up the SNN. The SNN gains the ability to function as a similarity metric between songs based on the raw audio information by being trained on this dataset of song picture pairs. On the test set, the SNN model scored 81.64 % accurate. It was built a query-by-multiple-examples based music recommendation system that uses the generated similarity metric to recommend music. A survey website was created to gauge how well their suggested system performed. Participants in the survey first compile a selection of music they enjoy, after which they score the suggestions. The user then reviews a single list of music that was generated using both the naive genre-based baseline system and the recommendations generated by both systems. Because the survey is a blind-use study, potential biases are reduced because participants are unaware that two systems are being utilized. The results revealed that more participants preferred the suggested system's choices, which also scored much higher than the baseline system even though it suggested less well-known songs than the baseline system did (Phaneendra et al., 2020).

Using convolutional neural networks (CNN) as the foundation, Chang et al. (2018) offer the PMRS, or customized music recommendation system. The CNN method divides music into several genres based on the audio signal rhythms. The output of the CNN and the log files are combined in PMRS to create a collaborative filtering (CF) recommendation system that suggests music to the user. The history of each user who uses the PMRS is recorded in the log file. From the log file, the PMRS extracts the user's history and provides genre-specific music recommendations. They assess the PMRS using the million-song dataset (MSD). They created a mobile application to demonstrate how the PMRS functions (an Android version). To evaluate the effectiveness of the PMRS, they used the confidence score metrics for various musical genres (Chang et al., 2018).

The melody, rhythm, timbre, and other significant aspects of music are challenging to extract and process; as a result, most personalized music recommendation systems for users do not fully take into account the content features of music itself, leaving them unsatisfied with the music recommendation. By examining the history records of user tracks in the data sets, Wang et al. (2018) proposed the technique RTCF (Tag-driven Collaborative Filtering Recommendation System) created a fair scoring system for users based on the training statistical language model (Good-Turing Estimate). Furthermore, they discovered that each tag has a linked count when describing a song or an artist, meaning that various tags can define the tracks and the artist's attributes to differing degrees. The tags work in conjunction with their analysis of each user's listening history to not only widen the music's content but also create individualized music recommendations for each user. Users' experimental findings on the tagged dataset suggest that our proposal method has a significant advantage over existing tag-based approaches for producing individualized recommended performance (Wang et al., 2018).

A method for making branching playlists that allow music recommendations based on listener preferences is proposed. They discovered that branching playlists raised listeners' levels of contentment, familiarity, and interest by analyzing the method's efficacy. They also put the suggested technique into practice by implementing

"reco.mu," a web-based music recommendation system. They discovered that while making a branching playlist, the creator is more aware of the recommender (Nonaka & Nakamura, 2021).

A mood-based music player is developed in the proposed system, which does real-time mood recognition and proposes songs in accordance with detected mood. This adds a new feature to the classic music player apps that are already pre-installed on our mobile devices. Customer satisfaction is a significant advantage of mood detection. This system aims to analyze a user's appearance, anticipate their facial expression, and provide music that fits their mood. Seven moods—anger, disgust, fear, joyful, sad, surprise, and neutral—can be correctly detected by their model, which has an accuracy rate of about 75%. Our Android application may then play music that is appropriate for the mood (Mahadik et al., 2018).

One of the top collaborative filtering-based neighbor algorithms is KNN. The KNN recommendation algorithm's error rate is higher when user ratings have changed, though. The shortcomings of the KNN algorithm, which are significantly impacted by the rating, are highlighted by Li and Zhang (2018) in this study. Then they suggest the KNN-Improved method, which is built on the KNN algorithm and uses the Baseline algorithm's mean value's concept, as well as increasing the rating's standard deviation. These steps can successfully lessen the effects of excessively high or excessively low user ratings, lower the algorithm's error rate, and produce better recommendation outcomes. Finally, performance of various recommendation algorithms is evaluated, and then improved algorithms are used (Li & Zhang, 2018).

This research focused on content-based music recommendation systems. The primary uniqueness of our study is the created recommender system's acoustic similarity-based approach to musical composition. This research examines two methods for developing a content-based music recommendation system. The first strategy, which makes use of auditory feature analysis, is highly popular. The second strategy uses computer vision and deep learning techniques to enhance the performance of the recommender system (Niyazov et al., 2021).

The sound browser's present implementation supports fundamental search and filtering features but lacks a method for sound discovery, such as a recommendation system. Users have generally chosen a limited range of high-frequency sounds as a result, which has resulted in less compositional diversity. In this study Smith et al. (2019) develop a recommendation system that uses audio attributes and collaborative filtering to suggest sounds for the EarSketch sound browser (Smith et al., 2019).

The usage of a content-based recommendation system that examines the rhythmic, melodic, and chordal aspects of the music will be extremely beneficial for classical music because these features help establish a user's musical taste. As a result, Cruz and Coronel (2020) describe a method for content-based recommendation that makes use of high-level musical qualities to compare classical music. The preliminary findings show that these features and methods can be used to build a content-based classical music recommender(Cruz & Coronel, 2020).

An information filtering technology called a music recommendation system (MRS) is used to manage the vast volume of digital music that is made available through online platforms. One of the most used methods in MRS is collaborative filtering (CF). The CF approaches are effective at recommending well-liked music but fall short of including less well-liked tunes. This essay tries to discuss longtail songs, or less well-known tunes. This research suggests an adaptive clustering method to include longtail tunes to recommendations. To find longtail tunes, the suggested approach is contrasted with user-based and item-based CF models. The metric called Tail-P has been proposed and contrasted with user-based and item-based CF models for the adaptive clustering approach. Results show that adaptive clustering outperformed CF models in terms of identifying longtail songs (Sunitha et al., 2022).

Due to the current world's quick expansion of music tracks both offline and online, greater access to automatically classify a song's genre and recommendations for related songs will affect the user's experience greatly. In order to perform genre categorization on the song, this study proposes a system that employs deep learning approach. Based on the results, the song will be recommended using word2vec. It's vital to compile a sizable collection for classification, index the songs according to their genres, and use the skip gram model to discover songs with comparable context that should be recommended. For a fantastic user experience, the proposed system functions as a whole music recommendation system (Budhrani et al., 2020).

This article proposes an innovative 15 deep learning-based multi-criteria collaborative filtering model. Their approach consists of two parts: the first portion gets the features of the users and the items and feeds them into a deep neural network that predicts the criteria ratings. These criteria ratings serve as the input for the second component, a deep neural network that forecasts the overall rating. Our suggested model outperformed existing

state-of-the-art methods in experiments using real-world data, which offers 20 evidence that deep learning and multi-criteria techniques may be successfully applied in recommendation systems (Nassar et al., 2020).

This study introduces a system for recommending music to users based on their current moods, activities, and demographic data like age, gender, and ethnicity. Additionally, voice commands and hand gestures can be used to operate the device. In order to make music recommendations based on the user's emotions and demographic information, unsupervised learning techniques were applied. The crucial concept is to make music recommendations based on the entire user's information, including demographics, emotions, and activities. The overall system performance was manually tested and evaluated with a group of people, and the results showed a 70% satisfaction rate for the recommendation. In addition, supporting models like demographic identification, emotion identification, and hand gesture identification have a higher percentage of accuracy rates, which has contributed to the success of the research. Unlike other systems, ours makes music recommendations while taking into account the entire user's data (Wijekoon et al., 2021).

MRS has experienced a growth in recent years. Its primary goal is to provide appropriate, meaningful suggestions to people for particular things according on the user's mood and interest in such items. The two most well-liked recommendation systems are: Collaborative filtering and content-based filtering are two examples. The Content Based technique suggests music based on user data, whereas the Collaborative method uses user ratings and content sharing to suggest music. We use a content-based approach to examine subjective music qualities like speechiness, loudness, and acoustness in order to generate music recommendations. For new users, cold-start is the most typical issue. Users are suggested the most well-liked tracks to solve it here (Dutta & Vishwakarma, 2021).

In this paper, Yi et al. (2019) offer a deep matrix factorization (DMF) based collaborative filtering (CF) framework that can easily and efficiently incorporate any sort of side information. To immediately produce latent factors of users and items from a variety of input information, two feature transformation functions are included into the DMF. Implicit feedback embedding (IFE) is suggested for the implicit feedback that is frequently used as input for recommendation algorithms. IFE transforms the high-dimensional, sparse implicit feedback data into a low-dimensional, real-valued vector while preserving the essential characteristics. IFE might significantly minimize the scale of model parameters and improve the effectiveness of model training. In terms of quantitative evaluations, experimental results on five public databases show that the proposed method outperforms the most advanced DL-based recommendation algorithms in terms of accuracy and training effectiveness (Yi et al., 2019).

This research presents a feeling-based music recommendation system that incorporates a user's mood from signals obtained via wearable physiological sensors. In particular, a wearable figure device that includes a galvanic skin reaction (GSR) and photograph plethysmography (PPG) physiological sensor (OR) makes a customer's feeling more tasteful. Any community-focused or content-based suggestion engine receives this emotional data as vital input. In this way, this knowledge can be used to expand already-existing proposed motor exhibitions. Their suggested system for acknowledging feelings is viewed as an issue of excitement and valence forecast from multiple physiological signals. Depending on the user's mood, we'll automatically play some music (Lakshmi et al., 2019).

A hybrid recommendation method based on the profile expansion technique is proposed by Tahmasebi et.al. (2021) address the cold start issue in recommender systems. In order to improve the user base in the area, they also take into account demographic information about users (such as age, gender, and occupation) in addition to user ratings. In particular, two distinct tactics are employed to offer some more ratings to individuals in order to enhance their rating profiles. The performance of recommender systems can be significantly improved, especially when they are dealing with the cold start problem, thanks to the proposed rating profile expansion technique. This assertion is supported by the fact that the suggested mechanism adds some more ratings to the original user-item rating matrix, making it denser than the original one. Naturally, giving a target user a rating profile with additional ratings helps recommender systems avoid the cold start issue. To estimate user similarity and forecast items that haven't yet been seen, enlarged rating profiles are used. Experiment findings show that the suggested strategy performs better than the other recommendation methods in terms of accuracy and rate coverage measures (Tahmasebi et al., 2021).

In this study, Mohamed et al. (2020) develop two novel algorithms to address the recommendation system's sparsity, accuracy, and performance issues. First, they employed clustering along with association rule mining to uncover a hidden pattern, measure the number of songs played every transaction, and compute similarities through cosine vector similarity to give recommendations to consumers. Second, to decrease dimensionality,

improve efficiency, and address accuracy and sparsity issues, they combined K-means clustering techniques with SVD. Their experiments are conducted on recent FM music datasets and movie-lens datasets with implicit and explicit feedback. They compare their new algorithms with k-means collaborative filtering using RMSE (root mean square error) to demonstrate the performance and accuracy of movie lens and demonstrate the accuracy between the two new algorithms and basic collaborative filtering by measuring accuracy using precision, recall, and F-measure. This experiment demonstrates that when combined with SVD and fundamental collaborative filtering, association rule outperforms improved k-means. However, their new k-means plus SVD algorithm outperforms random collaborative filtering K-means in terms of performance (Mohamed et al., 2020).

The authors of this paper Dhahri et al. (2018) propose an autonomous and adaptive recommendation system that uses the user's mood and implicit feedback to suggest songs without knowing the user's preferences in advance. Their approach uses relationships between users, songs, users' moods, and song emotions to automatically create a latent factor model from the online data of numerous users (generic song map per mood). The basic song map for each mood is customized using a mix of the Page-Hinkley (PH) test and the Reinforcement Learning (RL) framework, taking into account the implicit rewards of the user. In order to demonstrate the impact of mood on music recommendation and how the suggested solution might outperform other traditional solutions over time in terms of hit rate and F1 score, the researchers conduct a number of tests using the LiveJournal two-million (LJ2M) dataset (Dhahri et al., 2018).

There are two significant drawbacks to the current content-based recommendation techniques. First, the recommendation outcomes are relatively limited as a result of the item flaws and user model matching algorithms. Second, the scenario is not given much thought therefore the suggestion algorithm is not context-aware. Enhancing user pleasure through superior suggestion is crucial. Two cutting-edge techniques are examined and expanded in this research in order to improve recommendation performance. The context-aware recommender, which incorporates context information into the suggestion process, is the first approach. The second approach uses a recommender system that uses semantic analysis and takes domain semantics into account. The challenge is to put them together in a way that will completely utilize their potential despite the fact that they are compatible. This research suggests an enhanced content-based model that incorporates semantics and context. Context-aware suggestion is used to increase context awareness. To solve the narrowness issue, semantic relevance-based instance similarity is computed. The suggested recommendation system is assessed using metrics (such as the recall meter) and contrasted with the existing content-based techniques. Results show that the proposed system is superior in terms of accuracy (Yang, 2018).

In this research, a system is shown for gathering user musical preferences and musical textual characteristics on YouTube, as well as for generating collaboratively filtered music suggestions. Additional attributes for the item are required because collaborative filtering suffers greatly from cold start and data sparsity. They provide a technique to boost YouTube recommendation performance by combining implicit feedback and textual features. The results of the experiments show that this system's recommended playlists meet both individual and group preferences (Wang et al., 2020).

This study outlines the topics that require attention and thought and covers the key concepts and techniques employed in contemporary recommendation systems. The study seeks to develop a directed tag-based personalized music recommendation system that can offer consumers basic music services and push them personalized music suggestion lists based on these algorithms, user history data information, and music data information currently available. Next, the tag-based collaborative filtering algorithm is shown. Typically, this method employs discrete tags and juxtaposes and levels user tags and music tags, which does not accurately reflect the significance and ranking order relationship of each tag or the cognitive sequence of users as they listen to and annotate music. The user-tag and music-tag data are associated through the tag sequence of tag and music-tag data are correlated and analytically modeled, and feature directed graphs are produced in order to solve this issue and improve recommendation accuracy (Huo, 2021).

This essay conducts a thorough and in-depth analysis of the qualities of music. A proposed intelligent recommendation algorithm for modern popular music is knowledge graph-based. In this study, user-defined tags are referred to as the free DNA of music, facilitating the analysis of user behavior and the identification of user interests. This algorithm's recommendation quality has been confirmed to be relatively high, and it offers a fresh development route for enhancing the speed of looking for health information services (Zhang, 2022).

Different filtering methods are applied in recommendation systems. Along with such filtering techniques, some soft computing techniques can be incorporated to improve its performance. This article provides a summary of evaluation criteria, stages, difficulties, and how soft computing approaches are combined with filtering

techniques. The popularity of various filtering approaches and soft computing methods employed in recommendation systems is also statistically analyzed in this research (Das et al., 2019).

This research suggests using latent variables to suggest a playlist of songs for a specific film. The suggestion is based on the proposed scoring function, which uses a weighted average of the latent elements for both the video and the music. Additionally, pairwise ranking is used to create the objective function, and stochastic gradient descent is used to optimize the suggested objective function. In the studies, they lay up two hypotheses and create a number of tests to evaluate the performance and efficacy of the suggested algorithm from several angles, such as accuracy, quantitative research, and qualitative research. The experimental findings show that the suggested model has promise for accuracy and quantitative analysis. Additionally, this paper offers a thorough analysis to look at whether the system's suggested background music is appropriate through subject interviews (Liu & Chen, 2018).

This study introduces a music recommendation system for offline song collections that uses reinforcement learning ideas to produce acceptable recommendations based on the many taken into account content-based characteristics. Parallel instances of single-play multi-arm bandit algorithms are maintained to gain insights into the efficacy of the provided recommendations. By assuming that the environment's reward-generating process is non-stationary and stochastic, the ideas of Bayesian learning are then considered to describe the user preferences. Within the limitations of the input data capabilities, the system is intended to be straightforward, straightforward to deploy, and at-par with user pleasure (Bharadwaj et al., 2022).

This work provides a customized music recommendation system based on multidimensional time-series analysis, which can enhance the effect of music suggestion by appropriately considering user midterm behavior. This approach represents the user's behavior as a multidimensional time series, analyzes the series to better forecast the usage of music users' behavior preferences, and then expresses each song as the probability of belonging to numerous hidden themes. Then, a method for recommending music is put forth that takes into account users' long-term, medium-term, and real-time behaviors as well as the dynamic adjustment of the influence weight of the three behaviors. This method makes use of advanced long short time memory (LSTM) technology and aims to further enhance the effect of music recommendation. The practicality of the suggested method is initially verified through the use of the prototype system (Shi, 2021).

This article proposes a system for product recommendations that makes use of an autoencoder built on a collaborative filtering technique. The findings section includes a comparison of this model with the Singular Value Decomposition. Given that the recommendations given to users are in line with their interests and are unaffected by the data sparsity problem because the datasets are extremely sparse, their experiment has a very low Root Mean Squared Error (RMSE) value, which is 0.996. The outcomes are very encouraging, with an RMSE value of 0.029 in the first dataset and 0.010 in the second (Ferreira et al., 2020).

Dataset

Description

For a very long time, experts have worked to understand sound and what makes one music different from another. How is sound visualized? How it makes one tone different from another? Hopefully, a suitable dataset will allow for the opportunity to do the study. The GTZAN dataset, which may be found in at least 100 published articles, is the most often used public dataset for analysis in machine listening research for music genre recognition (MGR). For train-up our Music Genre Classification (MGC) technology, we use the GTZAN Genre Dataset.

The most popular publicly available dataset for testing in machine listening research for music genre identification is the GTZAN dataset (MGR). The files were gathered between 2000 and 2001 from various sources, including as personal CDs, radio, and microphone recordings, in order to illustrate a range of recording situations.

A total of 1000 audio tracks with duration of 30 seconds are included in the GTZAN Genre Dataset. The dataset is organized into a total of 10 classes, each of which has 100 tracks and represents a different musical genre. The tracks are all 22050Hz Mono 16-bit.wav audio files.

GTZAN genre Dataset Structure

GTZAN genre Data Fields

- audio: a tensor containing audio.
- genre: a class label tensor to classify audios in 10 classes

GTZAN Genre Categories

The classes or genres are:

1) blues
2) classical
3) country
4) disco
5) hiphop
6) jazz
7) metal
8) pop
9) reggae
10) rock

GTZAN genre Data Splits

- GTZAN genre contains a single split with 10000 audio tracks

Data Pre-processing

Table 1. Audio pre-processing setting

Hyperparameter	Value
Sample rate	22050 Hz
FFT	2048
Hop length	512
Number of segments	5
Number of MFCCs coefficients	13

Wave Form
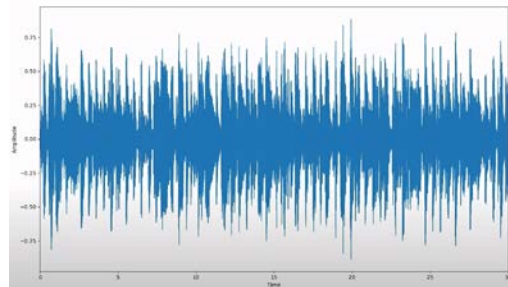
Figure 1. 30 sec of a wave from GTZAN dataset

Each wave form audio tracks with a 30-second duration and 22050Hz Mono 16-bit audio files in .wav format. Here X axis as Time domain and Y axis is Amplitude in the Wave form.

FFT (Power Spectrum)

In Fast Fourier Transform the normal wave are converted from Time to Frequency Domain and also Amplitude to Magnitude
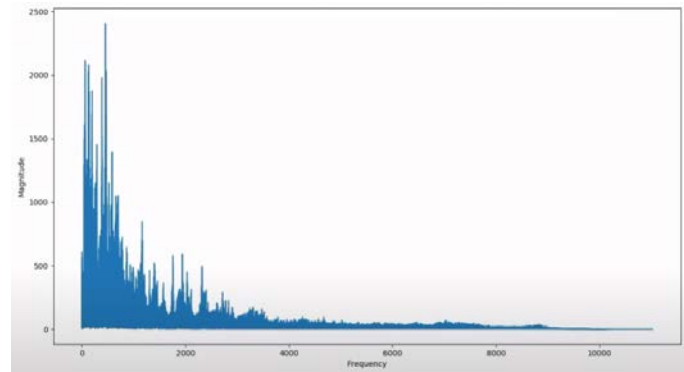
Figure 2. Power spectrum

MFCCs (Mel-Frequency Cepstral Coefficients)

The modeling of audio sounds and music by other writers has also used MFCCs. For instance, Foote (1997) uses a cepstral representation of sounds to build a retrieval system. One of the aspects in the retrieval system described by Blum et al. (1999) includes MFCCs (Logan, 2000). There is a description of a system for summarizing music using cepstral features (Logan, 2000). However, these works do not thoroughly examine the assumptions given; they just use cepstral features because they have been so successful for voice recognition.

Mel Frequency Cepstral Coefficients (MFCCs) were initially employed in a number of audio or voice processing approaches, but as the area of Music Information Retrieval (MIR) advanced alongside machine learning, it was discovered that MFCCs could effectively represent tone. The following is the fundamental process for developing MFCCs:

- Divide signal into frame
- Convert from Hertz to Mel Scale
- Take logarithm of Mel representation of audio
- Take logarithmic magnitude and use Discrete Cosine Transformation
- This result creates a spectrum over Mel frequencies as opposed to time, thus creating MFCCs
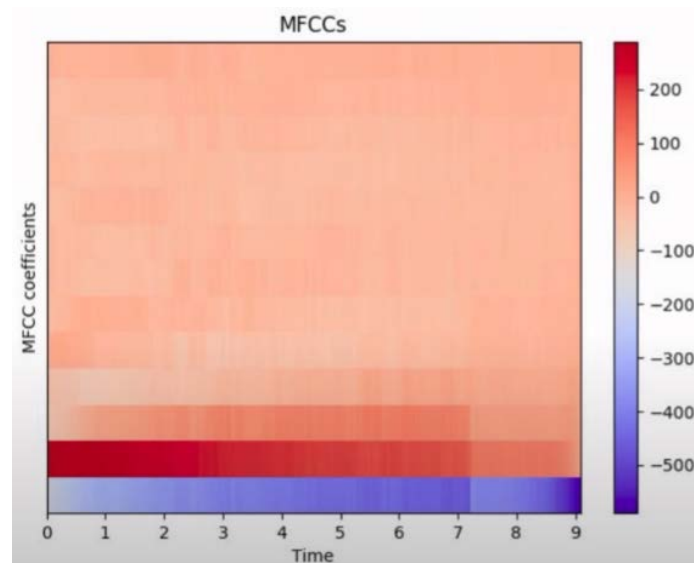

Figure 3. Visual representation of MFCCs

The GTZAN dataset which has 10 classes of music genre and 1000 of music wav files have been used in this system. Pre-process the audio data with the help of MFCCs to generate the features vectors. After extracted the features vectors using Mel Frequency Cepstral Coefficients, we pass the MFCCs into our deep learning networks.
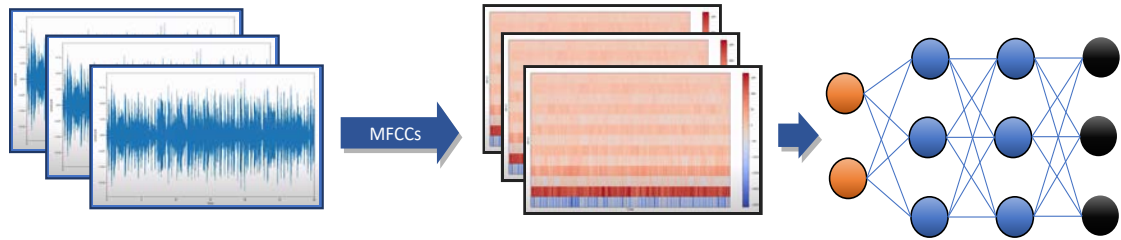
Figure 4. Pre-processing Pipeline

Methodology

Firstly, set the sample rate of music as 22050 Hz, set the window size or number of intervals for FFT as 2048 and hop length is 512. Here, the number of segments is 5 and the number of MFCC is 13. We have pre-processed the audio data by using the Mel Frequency Cepstral Coefficients (MFCCs) technique, then split the MFCCs feature data or Mel-Spectrogram for training & testing those data. Our network topology is a sequential model. Our network has 2 LSTM layers. In the 1st layer of LSTM, the input unit is 64 and the return sequence is True. That means it is a sequence-to-sequence layer. In the 2nd layer of LSTM, the input unit is also 64, set dropout and recurrent dropout is 0.4 (40%). This layer is a sequence to vector.
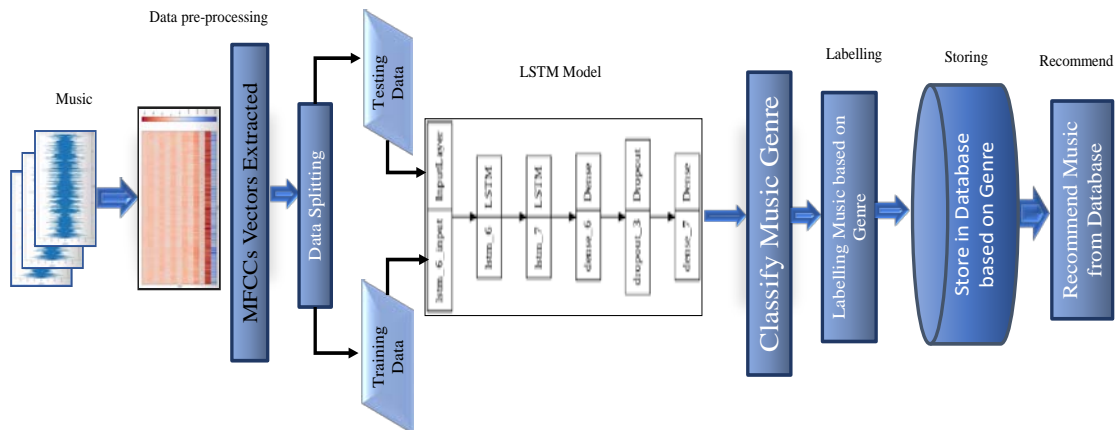

Figure 5. Methodology of music genre classification & recommendation

In the Dense Layer, the number of units is 64 and its activation function is Rectified Linear Unit (ReLU). In the Dropout Layer set dropout probability is 0.3 (30%) to avoid overfitting. In the Output Layer is set activation function as SoftMax which a classifier and set 10 neurons. Here 10 neurons represent the 10 Musical Genres such as blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock. This layer is finally classified the music genres. To compile the model, we use the Adam optimizer and the learning rate is 0.0001. We use 100 epochs and the batch size is 32 to train our model. After classifying the appropriate music genre, the particular audio gets labelled. Then labelled music are store in a database. Then finally, recommended music from particular genre from the labelled database which has been classified by our system.
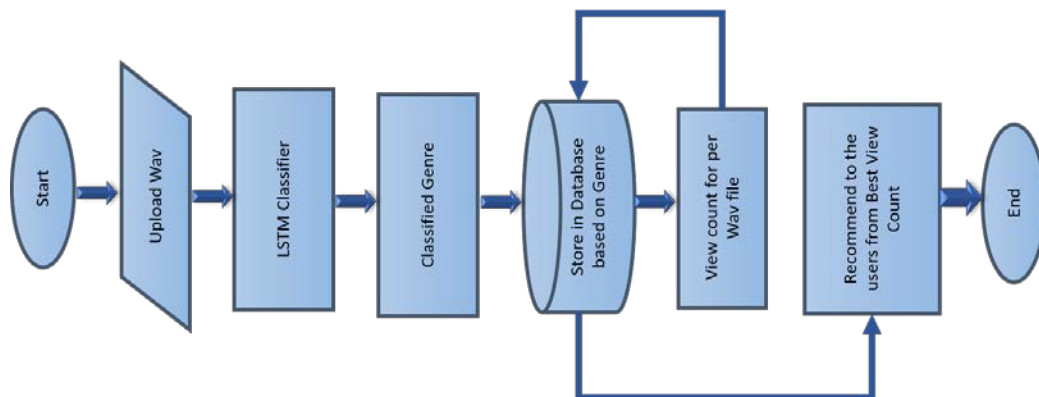

Figure 6. Flow chart of recommendation process

Steps of the flow chart of our recommendation system:

1) Start the process.
2) Upload Wav Music.
3) Classifying the Music Genre by Deep Learning   Model (RNN-LSTM/GRU/CNN)
4) Store the classified music in the database that has been categorized according to the genre.
5) Count the view of each wav file and also store the view count in the database in descending order.
6) Recommend the top 10 music (based on view count) from same categories genre from the database.

Experiments

TensorFlow is an artificial intelligence framework created by Google, and Colaboratory is a development tool. Today Google has made Colaboratory free for the general public to use since 2017, and TensorFlow is open-sourced. Google Colab or just Colab are the current names for Collaboratory. We use google Colab environment in our all experimentation. Users can run Python programs through a browser easily through Colaboratory. For those who enjoy machine learning and data science, this is very handy. The best feature of Google Colab is the fact that it offers free access to powerful computing tools like GPUs and TPUs. It costs nothing to utilize Google Colab.

Our network topology is a sequential model. Our network has 2 LSTM layers. In the 1st layer of LSTM, the input unit is 64 and the return sequence is True. That means it is a sequence-to-sequence layer. In the 2nd layer of LSTM, the input unit is also 64, set dropout and recurrent dropout is 0.4 (40%). This layer is a sequence to vector. In the Dense Layer, the number of units is 64 and its activation function is Rectified Linear Unit (ReLU). In the Dropout Layer set dropout probability is 0.3 (30%) to avoid overfitting. In the Output Layer is set activation function as SoftMax which a classifier and set 10 neurons. Here 10 neurons represent the 10 Musical Genres such as blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock. This layer is finally classified the music genres. To compile the model, we use the Adam optimizer and the learning rate is 0.0001. We use 100 epochs and the batch size is 32 to train our model. After classifying the appropriate music genre, the particular audio gets labelled. Then labelled music are store in a database. Then finally, recommended music from particular genre from the labelled database which has been classified by our system

Experiments Results

In this research study of "Developing music recommendation system by integrating an MGC with Deep Learning Techniques", we have used 3 different Deep Learning Techniques in our system such as RNN-LSTM (Recurrent Neural Networks-Long-Short Term Memory), CNN (Convolutional Neural Network) and GRU (Gated Recurrent Unit) model. The accuracy that we have gained by using those 3 models are given in the below.

Table 2. Comparative analysis of accuracy of the baselines and proposed model on GTZAN dataset

Models	Accuracy
CNN	0.72
GRU	0.71
LSTM	0.74

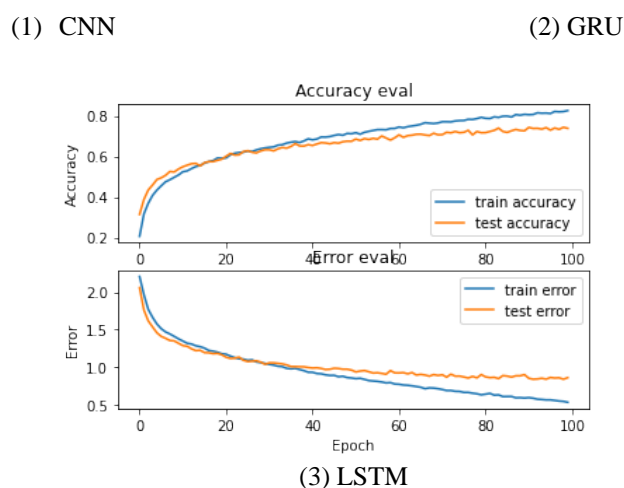(1) CNN   (2) GRU

(3) LSTM
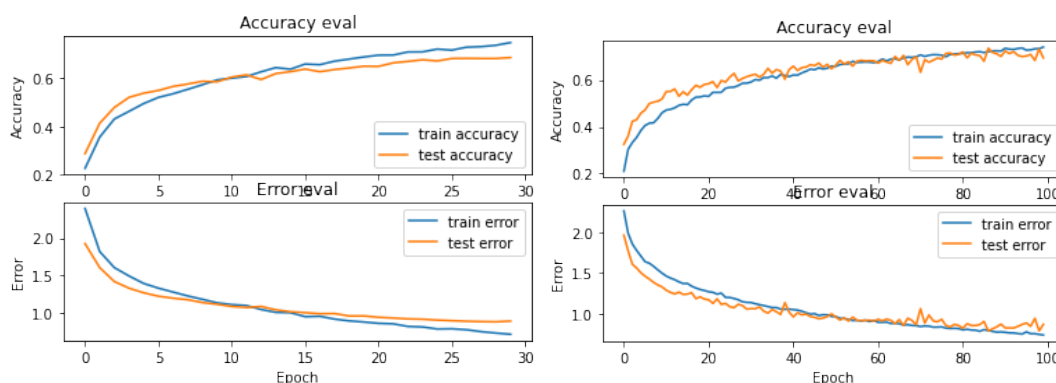
Figure 7. Train and test accuracy per epoch - CNN, GRU, LSTM model

By Gated Recurrent Unit (GRU) Model we have gained about 71% accuracy in our testing. On other hand, Convolutional Neural Network (CNN) have produced about 72% of accuracy in our testing which is 1% better than GRU Model. The Recurrent Neural Networks-Long-Short Term Memory (RNN-LSTM) have produced about 74% of accuracy in our testing experiment which is 2% better than CNN and 3% better than GRU Model. From all of our proposed models and experimentation, we see that the RNN-LSTM has achieved the best accuracy result (74%) among all of our proposed models.

Conclusion

Identifying a specific genre of music is a very challenging task, and the results depend on the accuracy of the feature and an appropriate model. In our research study, we have developed an Intelligent Music Recommendation System by integrating an MGC (Music Genre Classification) with Deep Learning Techniques such as CNN (Convolutional Neural Network), RNN-LSTM (Recurrent Neural Networks-Long-Short Term Memory), GRU (Gated Recurrent Unit) Model. We have used the GTZAN Dataset to training our system. We have extracted features by using of Mel Frequency Cepstral Coefficients (MFCCs) from GTZAN Dataset. Then pass the features into our neural networks. In our experiment, our neural networks GRU, CNN and RNN-LSTM models have gained the accuracy are about 71%, 72% and 74% respectively. From all of our proposed models and experimentation, the RNN-LSTM has achieved the best accuracy result (about 74%) among all of our proposed models. After classified music according to the genre, recommend the top 10 music (based on view count) from same categories genre music from the labelled database.

Recommendations

We will try to achieve more accuracy and try to reduce the classification time of music genre so that recommendation of music is shown quicker. Additionally, we will expand the dataset and integrate a variety of additional models into our system to improve accuracy and produce more precise results. We also develop an android Intelligent Music Player application with more accurate music recommendation system.

Scientific Ethics Declaration

The authors declare that they are responsible for the scientific, ethical, and legal aspects of the paper published in EPSTEM.

References

Bharadwaj, B., Selvanambi, R., Karuppiah, M., & Poonia, R. C. (2022). Content-based music recommendation using non-Stationary Bayesian reinforcement learning. *International Journal of Social Ecology and Sustainable Development (IJSESD)*, *13*(9), 1-18.

Blum, T. L., Keislar, D. F., Wheaton, J. A., & Wold, E. H. (1999). *U.S. Patent No. 5,918,223*. Washington, DC: U.S. Patent and Trademark Office.

Budhrani, A., Patel, A., & Ribadiya, S. (2020). Music2vec: music genre classification and recommendation system. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1406-1411). IEEE.

Chang, S. H., Abdul, A., Chen, J., & Liao, H. Y. (2018). A personalized music recommendation system using convolutional neural networks approach. *2018 IEEE International Conference on Applied System Invention (ICASI)* (pp. 47-49). IEEE.

Cruz, A. F. T., & Coronel, A. D. (2020). Towards developing a content-based recommendation system for classical music. In: Kim, K.J.,& Kim, H.-Y. (Ed.), *Information Science and Applications* (pp. 451-462). Singapore: Springer.

Das, S., Mishra, B. S. P., Mishra, M. K., Mishra, S., & Moharana, S. C. (2019). Soft-computing based recommendation system: A comparative study. *International Journal of Innovative Technology Exploring Engineering (IJITEE)*, *8*(8), 131-139.

Dhahri, C., Matsumoto, K., & Hoashi, K. (2018). Mood-aware music recommendation via adaptive song embedding. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 135-138). IEEE.

Dutta, A., & Vishwakarma, D. K. (2021). Personalized Music Recommendation System based on Streamer Streaming Trends. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

Ferreira, D., Silva, S., Abelha, A. and Machado, J. (2020). Recommendation system using autoencoders. *Applied Sciences, 10*(16), p.5510.

Foote, J. T. (1997, October). Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II* (Vol. 3229, pp. 138-147). SPIE.

Huo, Y. (2021). Music personalized label clustering and recommendation visualization. *Complexity*, *2021*, 1-8.

Lakshmi, D., Keerthana, K., & Harshavardhini, N. (2019). Feeling based music recommendation system using sensors. *International Journal of Research in Engineering, Science and Management*, *2*(3), 672–676.

Li, G., & Zhang, J. (2018). Music personalized recommendation system based on improved KNN algorithm. *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (pp. 777-781). IEEE.

Liu, C. L., & Chen, Y. C. (2018). Background music recommendation based on latent factors and moods. *Knowledge-Based Systems*, *159,* 158-170.

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *In International Symposium on Music Information Retrieval*.

Mahadik, A., Milgir, S., Patel, J., Jagan, V. B., & Kavathekar, V. (2021). Mood based music Recommendation System. *International Journal of Engineering research & Technology (IJERT),10*.

Mohamed, M. H., Khafagy, M. H., & Ibrahim, M. H. (2020). Two recommendation system algorithms used SVD and association rule on implicit and explicit data sets. *Inernational. Journal of Scientific & Technology Res*erach, *9*(11), 508-515.

Nassar, N., Jafar, A., & Rahhal, Y. (2020). A novel deep multi-criteria collaborative filtering model for recommendation system. *Knowledge-Based Systems*, *187*, 104811.

Niyazov, A., Mikhailova, E., & Egorova, O. (2021). Content-based music recommendation system. In *2021 29th Conference of Open Innovations Association (FRUCT)* (pp. 274-279). IEEE.

Nonaka, K., & Nakamura, S. (2021). reco. mu : A music recommendation system depending on listener's preference by creating a branching playlist. *International Conference on Entertainment Computing* (pp. 252-263). Cham: Springer.

Phaneendra, A., Muduli, M., Reddy, S. L., & Veenasree, R.(2022). EMUSE–An emotion based music recommendation system. *International Research Journal of Modernization in Engineering Technology and Science*, *4*(5), 4159-4163.

Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A machine learning approach for automation of resume recommendation system. *Procedia Computer Science*, *167*, 2318-2327.

Shi, J. (2021). Music recommendation algorithm based on multidimensional time-series model analysis. *Complexity*, Article ID 5579086. https://doi.org/10.1155/2021/5579086

Smith, J., Weeks, D., Jacob, M., Freeman, J., & Magerko, B. (2019). *Towards a hybrid recommendation system for a sound library*. IUI Workshops.

Sunitha, M., Adilakshmi, T., Ravi Teja, G., & Noel, A. (2022). Addressing longtail problem using adaptive clustering for music recommendation system. *Smart Intelligent Computing and Applications,* (pp. 331-338). (1th ed.). Singapore : Springer.

Tahmasebi, F., Meghdadi, M., Ahmadian, S., & Valiallahi, K. (2021). A hybrid recommendation system based on profile expansion technique to alleviate cold start problem. *Multimedia Tools and Applications*, *80*(2), 2339-2354.

Wang, M., Xiao, Y., Zheng, W., Jiao, X., & Hsu, C. H. (2018). Tag-based personalized music recommendation. In *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)* (pp. 201-208). IEEE.

Wang, Y. C., Yang, P. L., Sou, S. I., & Hsieh, H. P. (2020, December). The MuTube dataset for music listening history retrieval and recommendation system. In *2020 International Computer Symposium (ICS)* (pp. 55-60). IEEE.

Wijekoon, R., Ekanayaka, D., Wijekoon, M., Perera, D., Samarasinghe, P., Seneweera, O., & Peiris, A. (2021). Optimum music: Gesture controlled, personalized music recommendation system. In *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)* (pp. 23-28). IEEE.

Yang, Q. (2018). A novel recommendation system based on semantics and context awareness. *Computing*, *100*(8), 809-823.

Yi, B., Shen, X., Liu, H., Zhang, Z., Zhang, W., Liu, S., & Xiong, N. (2019). Deep matrix factorization with implicit feedback embedding for recommendation system. *IEEE Transactions on Industrial Informatics*, *15*(8), 4591-4601.

Zhang, Y. (2022). Intelligent recommendation model of contemporary pop music based on knowledge map. *Computational Intelligence and Neuroscience*, Article ID: 1756585 https://doi.org/10.1155/2022/1756585

Author Information

Md. Omar Faruk Riad
BGC Trust University Bangladesh
Chittagong, Bangladesh
Contact E-mail: *omarfaruk@bgctub.ac.bd*

Subhasish Ghosh
BGC Trust University Bangladesh
Chittagong, Bangladesh

To cite this article: