Mood-Based Music Recommendation System: Optimal GMM Clustering

1st Siphesihle Luthando Khayalethu Ndhlovu
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa
1599603@students.wits.ac.za

2nd Ritesh Ajoodah
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa
Ritesh.Ajoodah@wits.ac.za

***Abstract*—Some music recommendation systems do not consider the uniqueness of song items well enough to be able to classify these by the general mood and feel they may give off. The purpose of this report is to thus create a system that attempts to solve this problem and create a model to categorise each musical item by mood using their content-based features. The model of choice is the Gaussian Maximization Model since it allows the creation of inferences modelled in a computationally efficient way. The dataset to be explored is the publicly available Spotify dataset which comes with pre-processed content-based features of different songs from the era 1920-2020. Once this is done, I will then compare the different silhouette scores and Bayesian Inference Criterion (BIC) score obtained by this model with the score obtained by KMeans clustering algorithm which serves as the benchmark model we hope to improve on.**

***Index Terms*—music recommendation system, Content-based features, mood clustering, Gaussian Mixture Models, spotify-dataset-19212020-160k-tracks, Silhouette score, BIC score**

I. Introduction

Music acts as a vice for people to relate the way that they feel (or want to feel) both instrumentally and/or lyrically. Whether you need a song to listen to on the bus on the way to work, or one that will give you that extra pump at the gym or even to help you relax after a long hard day, music has become super accessible to each and every single one of us for any possible need we might have right at our fingertips. Music streaming platforms such as Youtube Music, Spotify, Apple Music and many more have planted themselves as the mediums of choice for accessing music and have made it possible for us, the general public, to store a unique music library within our profiles [1]. Most recommendation systems today commonly classify music based on their metadata (i.e. the name of the artist, album, year of release, genre etc.). Such a method has its conveniences however, the disadvantage is also obvious in that it ignores the users' direct emotional needs and the type of music that may be fitting to accompany those needs. This is linked to music's audio content i.e the music's vocality, melody, rhythm etc. [2] of songs which is what gives song pieces an overall theme/feel. This forms the basis of the motivation behind using content-based features of song pieces within a music recommendation system.

The principal objective of conducting this research is to classify music by mood using their content-based features. The songs used in this study was from the publicly available dataset provided by the popular music streaming service, Spotify. I will be modelling this content as a parametric probability distribution problem with the hypothesized outcome being that this music recommendation system - using content-based features obtained directly from Spotify - can be used to recommend similar sounding songs to a user.

II. Related Work

A. Music Emotion Models

Paper/Author	Model
Hevner [3]	Hevner's categorization of emotions: the adjective circle
Russell [4]	Russell's Arousal-Valence 2D Plane
Yang et al. [5]	Thayer's Arousal-Valence 2D Plane
Panda et al. [6]	MIREX Mood Classification Task

TABLE I
TABLE SHOWING KEY PAPERS FEATURED IN THIS SECTION

The relationship between music and emotions is a long-standing study by psychologists across the world since emotions are as complex as humans because of the differences in the ways that we each interpret and resonate with the world around us. Models that have been suggested can be split into two distinctive groups [6] :

Categorical Models: These typically refer to adjectives that can be used to classify emotions into groups for example(joyful, melancholic, anxious etc)

Dimensional Models: These on the other hand, consider multidimensional spaces where each space consists of clusters with synonymous words to classify emotions.

A famous example is the 2D geometric model called 'Russell's Circumplex Model of emotion' analogous to a compass [4] in which the vertical plane measures the level of arousal (intensity of the associated emotional state) from low to high and the horizontal plane describes the level of valence (the extent to which an emotion is positive or negative). An example of interpretation is that happiness would be interpreted as +arousal +valence, anger as +arousal -valence, depression as -arousal -valence and calmness as -arousal +valence. The figure can be found at figure 1.

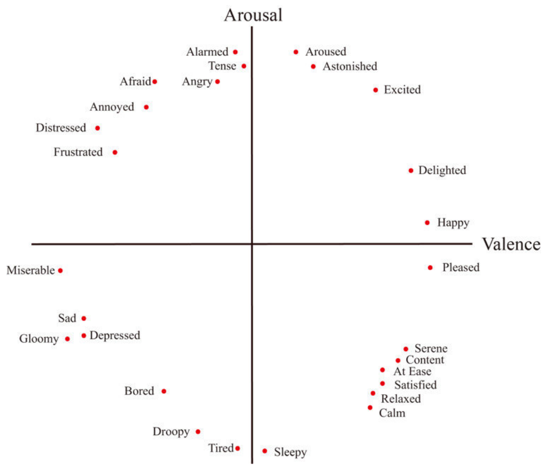

Fig. 1. Russell's Circumplex Model of emotion [7]

Another well-known example is in the renowned paper of researcher Kate Hevner [3] whereby a survey was conducted in which subjects were to write emotional adjectives that they felt in relation to the musical piece that was playing. Hevner then grouped descriptive adjectives into 8 clusters and laid them out in a circle as seen in figure 2.

Fig. 2. Kate Hevner's Adjective circle [3]

Thayer's model of music is an application of the above mentioned circumplex model in that it shows music's energy and stress as measures analogous to arousal and valence. Energy refers to the volume and/or intensity of the musical piece and Stress refers to the tonality and tempo of music. They can be divided into four groups: Exuberance (high energy, low stress), Anxious (High stress, high energy), Contentment(low energy, low stress), and Depression(low energy, high stress).

Similarly, utilising another taxonomy model of mood [5] influenced by Thayer shows a 'Music Emotion Variation Detection' scheme which basically translates results obtained from the fuzzy vectors used in the classification model into values representing valence and arousal as described above.

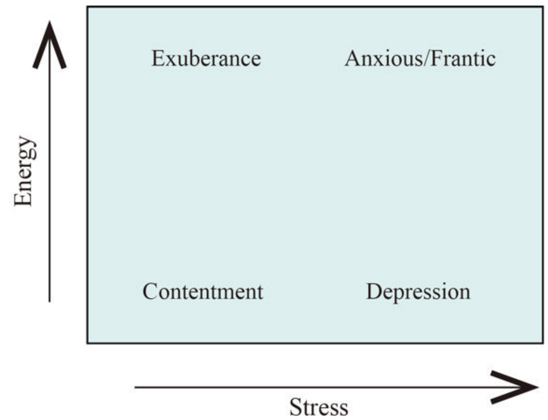

Fig. 3. Thayers Model [7]

B. Feature Analysis

Paper/Author	Data
Li et al. [8]	GTZAN dataset
Ajoodha et al. [9]	
Ajoodha [10]	GTZAN Genres
Panda et al. [6]	Mirex-like dataset
Model	**Accuracy**
CNN	84%
Naive Bayes	53.2%
K-nearest neighbours	72.80%
Multilayer perceptron	75.2%
Support vector machines	75.4%
Random forests	75.7 %
Linear logistic regression models	81.00%
Naive Bayes	53.2%
Support vector machines	32.50%
Multilayer perceptron	63.70%
Random forests	66.7 %
K-nearest neighbours	72.80%
Linear logistic regression models	81.00%
NaiveBayes	38.3%
NaiveBayes*	44.8%
C4.5	55.9%
C4.5*	57.3%
K-nearest neighbours	41.7%
K-nearest neighbours*	56.7 %
Support vector machines	52.8%
Support vector machines*	64.0%
Results in terms of F-measure and utilises all features as well as selected features (denoted by *)	64.0%

TABLE II
TABLE SHOWING KEY PAPERS FEATURED IN THIS SECTION

Music is both social and psychological [11] and front-end data (artist, album, genre etc) is no longer enough for the characterization and assembling of similar songs. In order to create a successful music classification model, it is crucial to extract reliable audio signals from musical data. An observation was made [8] that symbolic musical data [1] provides information that is usually readily available - these include

[1]Music stored in a notation-based format (e.g. MIDI (Musical Instrument Digital Interface)), which contains information about data of musical piece such as note onsets, pitch etc on individual tracks

the note onsets, note offset, intensity, and many more precise musicological descriptions of the songs -. For audio music, however, only the recorded sound signal is available. There are vast techniques that can be used to extract music data from various sources. Some of these include: content-based acoustic feature extraction,symbolic feature extraction, text-based feature extraction [10] with content-based techniques being the best and most note-worthy approach in extracting features that are not only relevant to the context of the signal of the audio piece, but also to genre classification problems. Content-based feature extraction methods make for better classification accuracy which subsequently maximises the speed at which music information is retrieved. The following list shows some of the relevant content-based features that can be extracted [6] which control the mood perceived by a listening ear. Namely:

- Timbre: spectral and/or harmonic richness
- Pitch: high/low
- Vibrato: extent and/or speed
- Timing: tempo and its variation
- Melody: range (small/large)
- Mode: major/minor
- Dynamics: crescendo/decrescendo
- Musical form: complexity and/or repetition
- Interval: small/large
- Harmony: consonant/complex-dissonant
- Rhythm: regular-smooth/firm/flowing-fluent/irregular-rough
- Articulation: overall staccato/legato,
- Loudness:high/low volume
- Tonality: chromatic-atonal/key-oriented

These features, although relevant, are not always easily and readily accessible and researchers have made use of manual feature selection methods [5] which subsequently improved classification accuracy as it removes , iteratively, the worst/least helpful features ensuring that the most optimal number of features remain.

The main goal of performing a similarity search is to allow the collection of music that is similar in sound and overall ambience, together. The utilisation of sound signals is therefore, justified since it is found that digital or analog audio signals of the same genre share similar characteristics such as comprising of similar instruments, rhythmic patterns and a similar pitch distribution [11].

C. Existing Music Emotion Retrieval Models

The key to successful music recommendation systems are those that provide each individual user tailored music recommendations. One way to do this is by giving them the ability to select context-specific moods. These user-specific mood tags prove to be relevant when considering Music Information Retrieval (MIR) and Park et al. [16] notes this importance for accurate suggestions of music that the person might enjoy. A Bayesian Fuzzy Network was the classification method of choice in which Park et al. [16] used this as a means of context inference. This idea was therefore creatively extended by authors such as Kim et al. [12] in which they showed the relationship between mood tags and the Arousal-Valence plane by means of a K-Means Clustering algorithm and frequency histograms for each mood class.

Paper/Author	**Data**
Kim et al. [12]	Independent Survey and program
Wang and Wang [2]	The Echo Nest Taste Profile Subset
Yang et al. [5]	Independent Collection of popular western,s Chinese and Japanese song
Liu et al. [13]	Music Spectrogram
Li and Ogihara [11]	Jazz and Classical Music Dataset
Kaur and Kumar [14]	
Koolagudi et al. [15]	Speech Audio Spectral Features Model
Model	**Accuracy**
K-Means Clustering	-
Deep Belief	-
Network (DBN)	-
Fuzzy KNN	68.2%
FNM	71.34%
	-
CNN	70.9%
SVM	70 - 83%
AEM	
Gaussian Mixture	-

TABLE III
TABLE SHOWING KEY PAPERS FEATURED IN THIS SECTION

Problems that typically arise is the fact that content-based classification methods usually rely on what is known as collaborative filtering(songs recommended to users based on what other users who enjoyed the same music listened to next) which has the downfall of not being inclusive of less ”popular” music when creating these recommendations [16] [2]. A way of mitigating this is perhaps joining collaborative filtering with content features of audio in order to improve and maximize the performance of the recommendation system [2] with results that show a significant improvement in accuracy using deep learning techniques (namely, a Deep Belief Network(also a PGM)) which excludes the exclusive reliance on collaborative filtering that previous classifiers required. Extending the use of neural networks in model training, is the use of Convolutional Neural Networks with a SOFTMAX classifier as a means to analyse the mood of the song [13]. Features were extracted from music spectrogram with the output being the same dimension space as emotion tags that the audio is tagged with.

Another publication proposed to characterize music emotion detection as a set of binary classification problems due to the fact that a single music sound can have more than one label i.e. it could be classified as both “happy” and “exciting” (multi-label classification) [11] . Each binary problem is then developed with a classifier that is based on the projection of the training data. These binary classifiers are then run on the test data individually and are selected according to some predefined threshold.

Gaussian Mixture Models (GMM) was researched to have its benefits when performing this task [15] in that an Emotion Retrieval Model on the spectral features of speech audio was developed which are then sorted into vectors. These vectors are then split into training and testing data for the GMM in which the output becomes the emotion recognised by the system. This model works by utilising an Expectation Maximisation Algorithm which can be used to estimate the parameters of class labels since these are unknown and is therefore known as the latent variable of the problem [15]. The algorithm initialises the set of parameters arbitrarily and in what is known as the ”E-Step”, utilises these arbitrary values to estimate the missing output. In the ”M-Step”, these estimates are used to re-estimate the parameters. This process is continues until convergence.

III. METHODOLOGY

As explored from the above literature, this paper looks at utilising an unsupervised machine learning clustering model. In particular, a probabilistic algorithm known as the Gaussian mixture model is used to train the dataset as the outcome is unknown. This algorithm is a generative process whereby a mixture of probability distributions are used in order to predict unknown parameters from this specific dataset. Specific insights of this methodology are explored further in the relative subsections that follow.

A. Dataset Description

The dataset of choice is from a popular music streaming service , *Spotify*. This platform has made public the audio features of approximately 160000 songs that were released within the years 1920 - 2021 housed on the platform *Kaggle*. This removes the extra step of performing feature extraction on different sounds as this has already been done. The feature-set is described in Table IV each ranging from 0.0-1.0 where applicable.

B. Models

The aforementioned features provided will be normalized, to ensure that all values are processed in the the same way within the same range to avoid corrupt results. The feature set will then be used for training and testing the model and will be split into 2 groups: 70% for training and the remaining 30% for testing if the model gives the right output.

1) Gaussian Mixture Model: This clustering technique is trained to work with the specific dataset in question where it takes subsets of the data and assess the normal distribution of each group. It then sorts data points that belong to the same distribution, together. In order to estimate the parameters , the classifier makes use of an iterative Expectation Maximization(EM) algorithm used for each Gaussian component and the mixture weight. This 2 step formula is written out as follows:

E-Step:

Features Maintained	**Description**
valence	’positiveness’ of a song
acousticness	how acoustic the song is
artists	song arist
danceability	tempo, beat strength, rhythm, stability, regularity
energy	intensity and activity of song
instrumentalness	amount of vocals in song
liveness	probability of song taking place live
loudness	how sound is perceived (quiet - loud)
name	song title
speechiness	spoken bits of a song
tempo	speed of song
Features Removed	**Descrption**
year	Song release year
duration ms	song duration
explicit	how explicit the song is
mode	notes used in a song
key	major/minor scale of song
popularity	song popularity
release date	date song released

TABLE IV

TABLE SHOWING THE FEATURES KEPT (UPPER PORTION) AND THE FEATURES ELIMINATED (LOWER PORTION). COLUMN TWO DESCRIBES EACH FEATURE RESPECTIVELY

$$r_{ic} = \frac{\pi_c N(\boldsymbol{x_i} \mid \boldsymbol{\mu_c}, \boldsymbol{\Sigma_c})}{\Sigma_{k=1}^{K} \pi_k N(\boldsymbol{x_i} \mid \boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})}$$

and

$$N(\boldsymbol{x_i}, \boldsymbol{\mu_c}, \boldsymbol{\Sigma_c}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_c|^{\frac{1}{2}}} exp(-\frac{1}{2}(\boldsymbol{x_i} - \boldsymbol{\mu_c})^T \boldsymbol{\Sigma_c^{-1}}(\boldsymbol{x_i} - \boldsymbol{\mu_c}))$$

Where,

- c and g: Each cluster ’c’ is represented by gaussian ’g’
- $\mu_c, \sum_c$: The mean and co-variance of that cluster, initialized by model
- r_{ic} : Probability that a data point x_i belongs to cluster c
- N(x$|\mu_c, \sum_c$): describes the multivariate gaussian

M-Step:

Where,

- m_c: fraction of points that assigned to cluster c

Based on the features, the classifier(GMM) takes a decision on the values to set the parameters for the latent variables [2]

[2]Variables not directly observed but inferred from previous features. These are the values used as labels for our dataset

$$m_c = \Sigma_i r_{ic}$$

$$\pi_c = \frac{m_c}{m}$$

$$\boldsymbol{\mu_c} = \frac{1}{m_c} \Sigma_i r_{ic} \boldsymbol{x_i}$$

$$\boldsymbol{\Sigma_c} = \Sigma_i r_{ic} (\boldsymbol{x_i} - \boldsymbol{\mu_c})^T (\boldsymbol{x_i} - \boldsymbol{\mu_c})$$

as this is an unlabelled dataset.

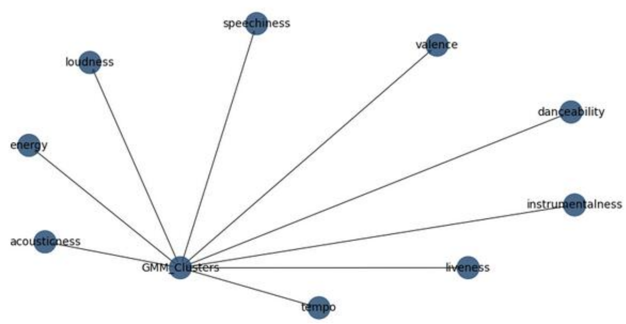

Fig. 4. Bayesian Graph Model Showing Causal Relationship Between GMM Clases and feature-sets Used to Generate Them

2) Bayesian Graph Model: As an extension to the above algorithm of creating different clusters for similar songs, Bayesian Graph Models will provide a computationally efficient way of creating inferences on the learned model of the exact probability values for song queries. A significant improvement is eminent when utilising graph models [17] if there are rich relationships that exist between songs. In this case, the causal relationship that exists between each class is modelled with each content-based feature of the songs that belongs to that cluster as edges. Since Bayesian Models produced higher accuracy [17] against other graph models, this made it a plausible option to improve this model. Each music item will have some probability of belonging to each cluster where the greatest probability generated infers the belonging of that song into that particular class. Figure 4 shows this structure.

C. Evaluation Techniques

Since the ground-truth of the different categories that each song may belong to is unknown, it is important to see how well/ poorly this clustering technique performs. Therefore, a comparison between two accuracy techniques will be compared in this study: Silhouette Coefficient and Bayesian Information Criterion (BIC) to obtain the most accurate number of mood clusters for our particular dataset.

1) Silhouette Coefficient: This technique measures the goodness of fit of the GMM model with values ranging from -1 to 1 where we can interpret the values as follows:

1: Means that the clusters are well apart from each other and we can easily distinguish each cluster from the next
0: Means that there isn't a significant distance between each cluster
-1: Means that the GMM clusters were not correctly assigned.

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Fig. 5. Silhouette Coefficient formula used in the evaluation of the goodness of fit of the GMM

The formula for this technique shown in figure 5 where b(i) is calculated as the average intra-cluster distance i.e distance of i (a point) to other points found in each respective cluster formed and a(i) is calculated as the average inter-cluster distance i.e the average distance between each cluster.

2) Bayesian Information Criterion: This scoring method works well on models that utilizes maximum likelihood estimation. The aim is to get as low a score as possible. The formula for this technique is shown in 6

$$BIC = (k^* ln(n)) — (2ln(L))$$

Fig. 6. BIC formula which acts as an evaluation technique for a given model

Instead of randomly choosing the number of classes that would be appropriate for classification, tests were run in this study where various numbers of cluster components were chosen along with the type of co-variance(spherical, tied, diagonal or full) the GMM model has. The number of these components which generated the best score was then selected as the optimal size.

D. Time Plan

The completion of this research will be determined by the completion of 2 phases, each phase to be completed in each semester of the year 2021. The projected time plan of this is thus as follows:

Phase I (February2021-July2021): Preamble

During this phase of research, background knowledge and related info pertaining to the proposed research is collected as follows

February 2021- March 2021: Topic finalisation and selection
March2021-April2021: Annotated Bibliography
April2021-May2021: Literature Review/ Related Works

May2021-June2021: Research Proposal

Phase II (July 2021- November2021): Software Development and testing and submission of final research report

During this phase of research, proposed models are applied to the dataset and evaluation takes place on the performance of our model. Full project is estimated to be completed by **28 November 2021**.

IV. Results and Discussion

	Covariance_Type	N_Components	BIC_Score
0	spherical	2	0.305337
1	spherical	3	0.312650
2	spherical	4	0.289264
3	spherical	5	0.289975
4	spherical	6	0.311544
...	...	...	...
379	full	25	-174738.411403
380	full	26	-173858.534084
381	full	27	-175521.126131
382	full	28	-175023.658338
383	full	29	-174684.397330

384 rows × 3 columns

TABLE V
BIC Table showing score generated using different sized clusters until convergence

	Covariance_Type	N_Components	Silhouette_Score
0	spherical	2	0.305337
1	spherical	3	0.312650
2	spherical	4	0.289264
3	spherical	5	0.289975
4	spherical	6	0.311544
...	...	...	...
147	full	35	-0.079418
148	full	36	-0.072755
149	full	37	-0.083967
150	full	38	-0.060680
151	full	39	-0.083865

152 rows × 3 columns

TABLE VI
Silhouette Table showing score generated using different sized clusters run until convergence

Figures 7 and 8 show some of the visual discernment of the scores calculated from the data when different sized components were tested. The Silhouette Coefficient shows the greatest score of 0.312... when the number of clusters is set to 4. The BIC score, however, shows the most optimal score when the number of cluster is set to 27. The GMM models were then generated using both of these results separately

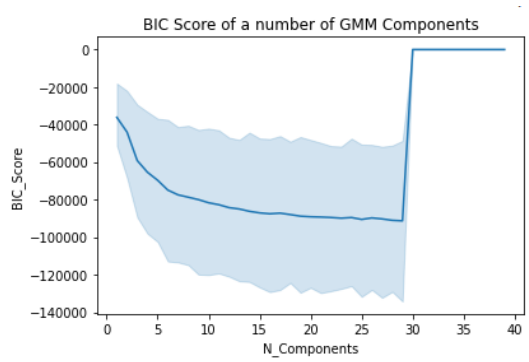

Fig. 7. BIC Graph of Convergence

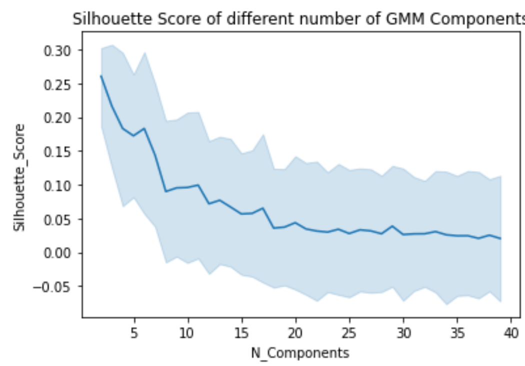

Fig. 8. Silhouette Graph of Convergence for GMM model

for further investigation. As seen on figures 9 and 10, the model did not produce a clean and simple representation of the data points sorted into their respective clusters. However, an important factor to acknowledge is that this was in fact a multi-variable dataset with 9 features used for classification. This means that this is a very compact view of the generated clusters and that higher dimensions would be required for better visualizations. Nonetheless, the algorithm managed to successfully group similar songs together.

Model	No. of Clusters	Best Silhouette Score
GMM	4	0.312...
K-Means	2	0.315...

TABLE VII
Table summarizing best silhouette score evaluation obtained by each model when tested against various cluster sizes.

Model	No. of Clusters	Best BIC Score
GMM	27	-175156.605672
K-Means	N/A	N/A

TABLE VIII
Table summarizing best BIC score evaluation obtained by each model when tested against various cluster sizes.

Since we do not have the ground truth of the correct number of mood clusters of this recommendation system using content-based features, an explicit measurement on how the

GMM model performed could not be obtained. This is why different iterations of the number of clusters was performed so that the best score could then be taken as the one to represent this dataset. The inferred classes generated by the training data was used to classify the testing data and the accuracy achieved by the models GMM and K-Means was then calculated. The latter model was used as the bench mark by which GMM was expected to perform better against. Table VIII shows a summary of the figures obtained.

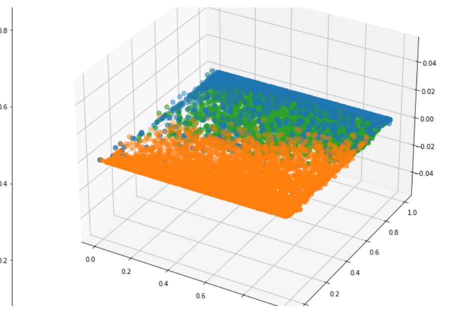

Fig. 9. Graph Showing BIC-Driven GMM Model

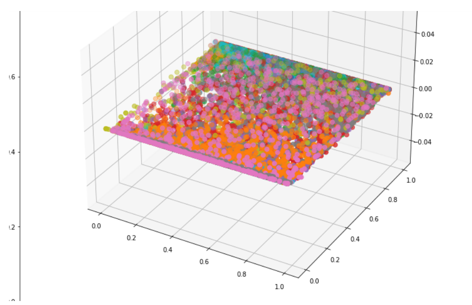

Fig. 10. Graph Showing Silhouette-Driven GMM model

V. Conclusion and Future Work

The aim of this paper was to create a music recommendation system that uses content based features to classify songs by mood. The Spotify Music dataset which is publicly available was used. Since the ground truth of the different mood clusters was not given, this was treated as an unsupervised machine learning problem using Gaussian Mixture Models utilizing the Expectation-Minimization Algorithm. The Bayesian Infrence Criterion score was calculated on different sized GMM clusters in an attempt to get clusters that are as accurate as possible. It was found for this particular dataset, songs can be classified into 1 of 27 moods.

A possible extension of this research would be to integrate this model within Spotify's already-setup APIs whereby a user specific playlist can be generated automatically. This can be done by perhaps calculating the Euclidean Distance between each song - the shorter the distance, the similar the songs may be which can then be sorted in descending order and be given as an output playlist. The top 50 (or any number to the users preference) would then be picked and returned in a tracklist.

VI. Acknowledgement

I would like to extend my thanks to my supervisor, Dr Ritesh Ajoodah, who made the completion of this work possible as his guidance assisted me throughout all the technical stages of compiling this report. I would also like to thank my lecturer, Dr Helen Roberts, for teaching me the skills necessary to compile my first computer science research project. Finally, I would like to express my utmost gratitude to God who carried me throughout this journey by giving me the strength, wisdom and good health that has allowed me to remain steadfast in my goals and my faith not only in Him but in myself too.

VII. Appendix

The Python code of the model, as well as the dataset is presented in the github repository: https://github.com/Siphesihle13/MoodRecommendationSystem

References

[1] R. E. S. Panda and R. Paiva, "Automatic mood tracking in audio music," *Master in Informatics Engineering, M. Sc. Thesis, panda@ student. dei. uc. pt Thesis Supervisor: Professor Rui Pedro Paiva July*, 2010.

[2] X. Wang and Y. Wang, "Improving content-based and hybrid music recommendation using deep learning," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 627–636.

[3] K. Hevner, "Experimental studies of the elements of expression in music," *The American Journal of Psychology*, vol. 48, no. 2, pp. 246–268, 1936.

[4] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[5] Y.-H. Yang, C.-C. Liu, and H. H. Chen, "Music emotion classification: A fuzzy approach," in *Proceedings of the 14th ACM international conference on Multimedia*, 2006, pp. 81–84.

[6] R. Panda, B. Rocha, and R. P. Paiva, "Music emotion recognition with standard and melodic audio features," *Applied Artificial Intelligence*, vol. 29, no. 4, pp. 313–334, 2015.

[7] Y.-S. Seo and J.-H. Huh, "Automatic emotion-based music classification for supporting intelligent iot applications," *Electronics*, vol. 8, no. 2, p. 164, 2019.

[8] T. L. Li, A. B. Chan, and A. H. Chun, "Automatic musical pattern feature extraction using convolutional neural network," *Genre*, vol. 10, p. 1x1, 2010.

[9] R. Ajoodha, R. Klein, and B. Rosman, "Single-labelled music genre classification using content-based features," in *2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*. IEEE, 2015, pp. 66–71.

[10] R. Ajoodha, "Automatic music genre classification," Ph.D. dissertation, 2015.

[11] T. Li and M. Ogihara, "Content-based music similarity search and emotion detection," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, 2004, pp. V–705.

[12] J. Kim, S. Lee, S. Kim, and W. Y. Yoo, "Music mood classification model based on arousal-valence values," in *13th International Conference on Advanced Communication Technology (ICACT2011)*, 2011, pp. 292–295.

[13] X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, "Cnn based music emotion classification," *arXiv preprint arXiv:1704.05665*, 2017.

[14] C. Kaur and R. Kumar, "Study and analysis of feature based automatic music genre classification using gaussian mixture model," in *2017 International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2017, pp. 465–468.

[15] S. G. Koolagudi, A. Barthwal, S. Devliyal, and K. S. Rao, "Real life emotion classification using spectral features and gaussian mixture models," *Procedia engineering*, vol. 38, pp. 3892–3899, 2012.

[16] H.-S. Park, J.-O. Yoo, and S.-B. Cho, "A context-aware music recommendation system using fuzzy bayesian networks with utility theory," in *International conference on Fuzzy systems and knowledge discovery*. Springer, 2006, pp. 970–979.

[17] T.-T. Dang and K. Shirai, "Machine learning approaches for mood classification of songs toward music search engine," in *2009 International Conference on Knowledge and Systems Engineering*, 2009, pp. 144–149.