SASYA - Mental Health Analysis based on the Music choices of the individual
CS550 - Machine Learning - Major Project REPORT

Kata Lakshmi Lasya[1], M. V. Kiran Sooraj[2], and Pingili Sameeksha Reddy[3]

[1]12140880, BTech, CSE, Third year
[2]12141000, BTech, CSE, Third year
[2]12141240, BTech, EE, Honours in CSE, Third year

Keywords:
Decision Trees
GaussianNB
K- Nearest Neighbours
Neural Networks
Support Vector Machine
Linear Discriminant Analysis

Abstract This project is driven by the motivation to explore the untapped potential of personal musical preferences as a non-invasive and accessible criterion for assessing an individual's mental and emotional well-being. In a world where comprehensive mental health assessments can be daunting and often underutilized, this research aims to harness the power of music - a universal language that resonates with people from all walks of life. The project delves into the myriad ways in which music reflects and influences human emotions, delving into emotional expression, regulation, and personal associations. It also considers how cultural and social influences shape musical preferences and the role of music as a coping mechanism in the face of stress and anxiety. The project recognizes that the choice of music can reveal not just emotional states but also a person's artistic sensitivity and openness to new experiences, all of which are crucial facets of psychological well-being. With this foundation, the project seeks to develop a practical tool for mental health assessment based on musical preferences, providing an accessible and engaging means for individuals to self-assess their mental health. By highlighting the potential of this innovative approach, the project aims to inspire individuals to seek professional help, promoting timely intervention and improving mental health outcomes. This research represents a fresh perspective on mental health assessment, one that capitalizes on everyday choices to empower individuals in their mental health journey.

 Submitted: as the

1. Introduction and Problem Motivation

Personal preferences, including those related to daily aspects like food, art, and music, can serve as mirrors reflecting the state of an individual's heart and mind. While comprehensive mental health assessments have their place, sometimes, one can gain valuable insights into a person's psychological well-being without the need for extensive procedures. A major concern in this regard is determining the criteria to focus on when analyzing these preferences.

Music stands out as a particularly apt criterion due to its inherent diversity and the scientific validation of its impact on the human brain. Different genres, rhythms, and melodies can evoke a wide range of emotions. The type of music someone enjoys and the emotional response it elicits can provide significant clues about their mental state. Music also serves as a unique outlet for emotional expression and regulation.

Furthermore, personal associations with specific songs or genres can offer glimpses into one's past experiences and how these shape their present emotions. Music preferences can reflect cultural and social influences and reveal an individual's coping mechanisms for stress and anxiety. An appreciation for various musical styles can also signify a person's artistic sensitivity and openness to new experiences, all of which contribute to their psychological well-being.

In a world where mental health sensitivity often falls short, analyzing daily choices, such as musical preferences, can motivate individuals to seek professional help without undergoing lengthy and complex tests.

2. Datasets

1. GITZAN Audio Dataset

In the pursuit of music genre prediction, the renowned GTZAN audio dataset is employed, which is widely acknowledged in the field. This dataset consists of ten distinct music genres, with each genre being composed of precisely one hundred audio tracks. These audio tracks are uniformly encoded in the .wav format, exhibiting a consistent audio quality characterized by a sampling rate of 22050Hz, mono-channel audio, and the use of 16-bit encoding. The music genres encompassed within this dataset include blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. Collectively, the GTZAN dataset comprises one thousand audio tracks, with each track precisely lasting 30 seconds.

2. MXMH Survey Results Dataset

For the mental health analysis model, the dataset of interest is sourced from Kaggle, specifically the "mxmh survey results" dataset. This dataset provides a comprehensive repository of data capturing diverse factors that influence an individual's mental health. It encompasses a wide array of attributes, such as age, the number of hours dedicated to music consumption, the frequency of exposure to different music genres, and the individual's preferred music genre. The dataset's primary purpose is to predict the levels of anxiety, depression, obsessive-compulsive disorder (OCD), and insomnia that may be experienced by an individual. These mental health indicators are quantified on a scale that spans from 0 to 10. It's a valuable resource for the development of models aimed at predicting an individual's mental health status.

3. Solution Approach

3.1 Input from user

- A form asking basic questions about the user's lifestyle: Age, Hours per day, listening to music while working, whether or not an instrumentalist or composer, favourite music genre, Exploratory or not about music, foreign languages, Beats per minute of heart.
- Five audio clips of the user's current favourite music.

3.2 Music Genre Prediction model

The initial usage of the 30 second audio clips directly was disregarded when the dataset was expanded from 999 to almost 10000 instances of 3 second audio clips. CNN model gave good results with the new dataset whose architecture is as follows.

```
model_n = models.Sequential([
    layers.Dense(256, activation='relu',
    input_shape=(X_train_1.shape[1],)),
    layers.Dropout(0.5),

    layers.Dense(128, activation='relu'),
    layers.Dropout(0.5),

    layers.Dense(64, activation='relu'),
    layers.Dropout(0.5),

    layers.Dense(10, activation='softmax')
])
```

The first layer is a dense layer with 256 units and a rectified linear unit (ReLU) activation function. It specifies an input shape compatible with the shape of the training data (`X_train_1`). A dropout layer with a dropout rate of 0.5 is applied after the first dense layer, serving as a regularization technique to prevent overfitting. Subsequently, two additional pairs of dense and dropout layers follow: a dense layer with 128 units and ReLU activation, followed by a dropout layer, and a dense layer with 64 units and ReLU activation, followed by another dropout layer. The final layer is a dense layer with 10 units and a softmax activation function, which is commonly used for multi-class classification problems. This architecture suggests a deep neural network with multiple hidden layers, incorporating dropout regularization to enhance generalization performance during training.
The accuracy was validated to 89.66 percentage.

Each of the music clips is evaluated by the music genre prediction model to give a probabilistic genre prediction spectrum. The probability of each of the genre in all the cases are averaged to derive a single probability for each genre. For instance, let the probabilities of the audio file belonging to pop genre be p_1, p_2, p_3, p_4, p_5.
The overall probability of the audio file being pop is evaluated to be:

$$p = \frac{p_1+p_2+p_3+p_4+p_5}{5}$$

The inference from this is derived to be that if p is the probability of the liked songs to belong to pop genre, then p is the probability that the user likes pop songs.

3.3 Mental Health Prediction model

The input of degree of liking towards a genre for the mental health prediction model is mapped as 3-very frequently, 2-sometimes, 1-rarely, 0-never. The output of the music prediction model is thus mapped as, for a probability of p, p = (1 - 0.75: 3 , 0.75 - 0.5: 2 , 0.5 - 0.25: 1, <0.25: 0)
To integrate the models, the genres of music prediction model and those in the mental health prediction model are manipulated, compressed, deleted or modified accordingly.
The mental health prediction model finally uses different models to predict the probabilty of different mental conditions, depending on the best accuracy results, as follows.

- Depression - KNN
- Anxiety - Decision tree
- insomnia - Decision tree
- OCD - SVM

```
=======K-Nearest Neighbors (KNN)=======

from sklearn.neighbors import KNeighborsClassifier

# Create instance with KNN function and parameters
knn = KNeighborsClassifier
(n_neighbors=15, weights='distance')

# Fit the model
knn.fit(X_train, y_train)

=======Support Vector Machine=======

from sklearn import svm
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import classification_report

# Create an SVM classifier object
svc = svm.SVC()

# Define the parameter grid to search over
param_grid = {
    'C': [0.1, 1, 10],
    'kernel': ['linear', 'rbf'],
    'gamma': [0.1, 1, 10]
}

# Create a GridSearchCV object with the
SVM classifier and parameter grid
grid_search = GridSearchCV(svc, param_grid,
scoring='f1_macro', cv=5)

# Fit the GridSearchCV object on the training data
grid_search.fit(X_train, y_train)

# Get the best parameters and best score
best_params = grid_search.best_params_
best_score = grid_search.best_score_

# Create an SVM classifier with the best parameters
svc = svm.SVC(**best_params)

# Fit the model on the training data using
the best parameters
svc.fit(X_train, y_train)
```

```
=======Decision Tree=======

from sklearn.tree import DecisionTreeClassifier

# Initialize the Decision Tree classifier
clf = DecisionTreeClassifier()

# Fit the classifier to the training data
clf.fit(X_train, y_train)
```

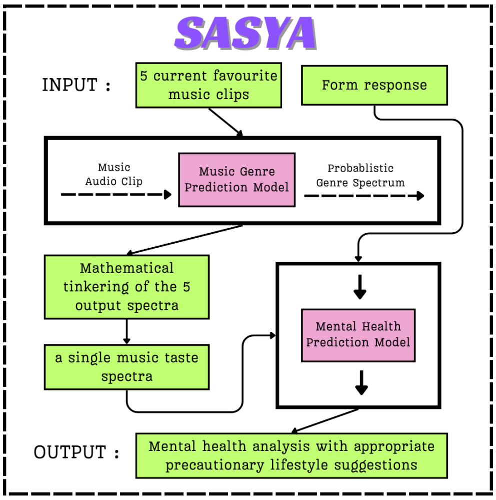

FIGURE 1. Schematic diagram of SASYA

The results of these models in case of different mental health disorders are given as predictions in the output that the user might be having that particular disorder. **It is to be noted that no output of the model should be quoted as a formal prescription but strictly as a prediction only. The users are expected to cross verify the output about their mental health with a trained professional in the field.**

4. Uniqueness

While there have been many studies to understand the impact of music on mental health, the solution improves the accuracy of the prediction of mental health of the user by expanding the scope to analysing the current favourite songs of the user along with the answers manually given based on the user's self awareness.

5. Models used in training and comparision

5.1 Music Genre Prediction Model

Models used were - KNN, SVM, Ensemble of SVM, CNN

5.1.1 Complexity, Ease of interpretation

CNN is the most complex model but giving best results. The general interpretation of CNN is sparse because of several unknown hidden layers.

5.1.2 Accuracy or other relevant performance metrics

SVM: 64 percent accuracy
Achieved an accuracy of 64 percentage, indicating moderate predictive performance.
Ensemble: 63 percent accuracy
When Ensemble models are employed, the accuracy is expected to rise but when it was applied with SVM, it gave degraded results.
KNN: 87.47 percent accuracy
CNN: 89.66 percent accuracy
3 CNN models were used. With increasing hidden layer count, accuracy reduced. Hence the best model, when considered had a 91.08 percent accuracy.

5.1.3 Variance

There is an average variance of 2 percentage in all the models of SVM, Ensemble, KNN and CNN

5.1.4 Training time

KNN, SVM, Ensemble of SVM - negligible time in milliseconds CNN - around 90 secs

5.2 Mental Health Prediction Model

Models used were - KNN, Gaussian Naive bayes, Decision tree, SVM, Linear Classifier

5.2.1 Complexity, Ease of interpretation

K-Nearest Neighbors (KNN) -
Theoretical Complexity: KNN is non-parametric with low training complexity, dependent on dataset size and k. Prediction cost can be high for large datasets.
Ease of Interpretation: Conceptually simple, predictions based on majority neighbors. High interpretability, especially with small k.

Decision Tree -
Theoretical Complexity: Non-linear, hierarchical model. Complexity linked to tree depth, risking overfitting.
Ease of Interpretation: Highly interpretable, decision rules in a clear tree structure.

Support Vector Machine (SVM)
Theoretical Complexity: Handles linear and non-linear boundaries. Complexity linked to kernel choice and regularization. Potentially computationally expensive.
Ease of Interpretation: Less intuitive, especially with non-linear kernels. Interpretability can be challenging, particularly in high dimensions.

5.2.2 Accuracy or other relevant performance metrics

For Depression - 64.09 percent (KNN)
For Anxiety - 57 percent (Decision tree)
For Insomnia - 72.93 percent (Decision tree)
For OCD - 84.53 percent (SVM)
Incase of OCD model, there is a competition between SVM and GaussianNB model in terms of accuracy. In some cases either of them are giving higher percentage of accuracy. SVM is chosen after observing pattern of majority turnups.

5.2.3 Variance

There is a variance of 0 percent - depression 1.05 percent - anxiety, 0.06 percent - insomnia, 0 percent - OCD in the models after multiple trails in case of mental health prediction model.

5.2.4 Training time

All the models: KNN, SVM, Decision trees, have taken almost less than a minute to train.

6. Scope of improvement

The overall improvement in the validation accuracy of the main model: mental health prediction model has only been two percentage, in comparision to the reference model. The plausible explanations for it could be that- the modification of dataset of mental health prediction model according to the results of music genre prediction model and the ED performed to eliminate the outliers. Tinkering with hyperparameter tuning further may yield better outcomes in mental health prediction model.
The project can be improved by devising better mathematical strategies to retain accuracy of music taste derived and analysed as a combined output from genre prediction model. Also, better datasets need to be collected as the one available currently for mental health prediction model is very limited.

7. References:

1. https://cs229.stanford.edu/proj2018/report/21.pdf
In music genre classification, the work by Derek A. Huang, Arianna A. Serafini, and Eli J. Pugh was referenced. The utilization of machine learning techniques for music genre classification was explored in their paper, with an emphasis on the effectiveness of mel-spectrograms and convolutional neural networks. Their insights and findings were leveraged to inform our research in this domain.

2. https://macuriels.com/project/predicting-depression-with-music/
For mental health analysis model, the blog "Predicting Depression Based on Musical Taste" by Miguel Curiel, served as a good reference for our project. It delves into machine learning models, including K-Nearest Neighbors and Naive Bayes, to predict depression from music and mental health survey data. Key technical steps involve data preprocessing, feature selection, and model evaluation using metrics like Macro F1 score. The blog highlights the nuances of applying machine learning to a complex topic and emphasizes the importance of holistic variables and controlled, randomized sampling for improved model accuracy.