

Depression Analysis using Machine Learning Based on Musical Habits

Suyoga Srinivas, Naveen N Bhat, Yashwanth Venkat Chandolu

Abstract: *Depression has been a main cause of mental illness. Depression results in vital impairment in lifestyle. A significant reason for suicidal cerebration is observed to be depression. Music varies the intensity of emotional experience by captivating the neurotransmitters and brain anatomy, including the brain's dopaminergic projections. The popularity of using Regression Models in data analysis in both research and industry has driven the development of an array of prediction models. It relies on independent variables and can provide the prediction for the dependent variable. The paper outlines the development of a Regression model to get the depression score of a person based on the music the user listens to. A regression model is used to predict the depression score depending upon the data obtained from a varied span of individuals and the genre of music they have listened to. We generate a suitable report based on the depression score. The doctor can then use the report to give the necessary treatment to the depressed patient. With our research, we have obtained variance and r^2 score of over 0.95.*

Keywords : Multivariate Linear Regression, Music, Principal Component Analysis, Support Vector Regression.

I. INTRODUCTION

Depression is a common and severe mental syndrome that adversely changes the way you feel, the way you behave and your thought process. Many a time, being depressed makes you feel like carrying a very heavy burden. Though depression cannot be detected easily, it has extensive and diverse impacts [1]. Some of the observed symptoms of depression are: guilt and worthlessness feel, feeling restlessness and irritable, lack of concentration and poor decision-making skills, problems with sleep pattern, change for appetite and weight, suicidal thoughts. Music plays a very significant role in enhancing an individual's life as it is an important means of entertainment [2]. We can observe that people extensively listen to music on their smartphones, computers or other electronic gadgets. Platforms like Spotify, Jio Savan, YouTube, SoundCloud, Wynk, etc help users listen to music. Sandra L(2006)[3], Dave Miranda and Michel Claes(2009) [4] and Dianna Kenny and Bronwen Ackermann(2013) [5] have cited in their paper that listening to music is related to a person's mood. Depending upon the individual's mood, we can decide whether or not an individual is depressed. We can classify

Music based on different genres, namely rock, hip-hop, pop, country music, soul music, funk, jazz, etc. However, these genres can be grouped into happy songs and patho songs. In this paper, we generate a depression score based on how long a person listens to a particular genre of music. The depression score determines the extent of a person's depression. Regression has been utilized as a method to imitate and inspect the correlation amongst the variables and the extent of their contribution in building a specific conclusion collectively. The values of the response variables can be estimated from the observed values of the explanatory variables. In this paper, we've used multivariate regression. We have used a statistical linear technique to represent the correlation in betwixt a explained variable and many explanatory variables.

In Multivariate Linear Regression, we effectuate a correlation among various input predictor independent variables and a dependent output response variable. It is a technique that provides the extent to which multiple predictor and multiple response have been linearly correlated. This technique is widely utilized to anticipate the behaviour among the explained variables connected with deviations in the explanatory variables. We used multivariate regression to determine the depression score based on how long the person listened to a particular music. This research paper is structured as follows: Part II represents the procedure involved in depression detection. Part III incorporates the methodology utilized in this paper. The results and conclusion of the research is provided in part IV and part V, respectively.

II. PROCEDURE

User listens to the songs:

The user has to listen to six songs, three each from happy and patho genre. The six songs are randomly selected from the database containing a collection of relevant songs.

B. Score Generation:

The user is asked to listen to the song for thirty seconds at least. However, if the user does not like the song, can skip to the next song. The user is graded based on the length of the song listened to. This is explained in detail in later parts of the paper.

Revised Manuscript Received on December 12, 2019.

Suyoga Srinivas *, Information Science Department, PES university ECC, Bangalore, India. Email: suyoga06031999@gmail.com

Naveen N Bhat, Computer Science Department, PES university ECC, Bangalore, India. Email: naveennbhat22@gmail.com

Yashwanth Venkat Chandolu, Computer Science Department, RNSIT, Bangalore, India. Email: ychandolu0131@gmail.com

C. Training and Fitting the model:

Part III of this paper, explains the various models we have used during our research. Once the model is fit, we predict the depression score of the person which ranges between -100 to +100.

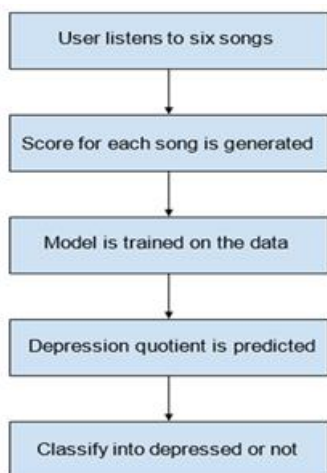


Figure 1. Flow model for procedure

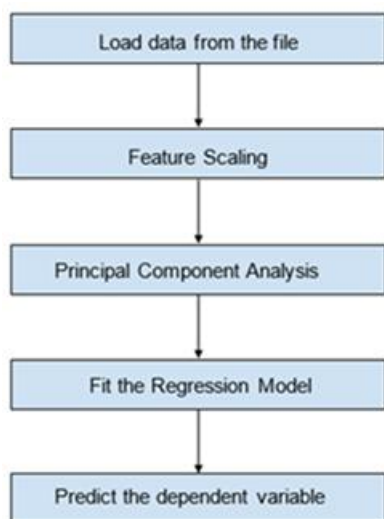


Figure 2. Flow model for Methodology

D. Classify whether the person is depressed or not:

Once we get the depression score, we generate a report about the person's depression and categorize them based on the extent of depression. This can be used for treating the person, if and only if they required it.

III. METHODOLOGY

A. Data Collection

The first step of any machine learning task is to gather relevant data. In this paper, we have collected data from 2025 different individuals. We have six independent variables and a dependent variable. Each independent variable represents the score obtained after listening to the song.

Table 1 represents scoring criteria.

Score	Length of the song listened to
0	0 to 30 seconds
1	30 seconds or more but not the entire song
2	Entire song
3	Listened to song multiple times

Table 1. Scoring criteria

The dependent variable represents the depression quotient of the patient. The depression quotient varies between -100 to +100. A depression quotient of a positive integer suggested that the person is happy, whereas a depression quotient of a negative integer suggested that the person was unhappy. In simple words, the mood of the person was directly proportional to the depression quotient.

B. Data Preprocessing

The data is read from the file, and these preprocessing steps are effectuated on the data: 1. Feature scaling: when a model is being trained using many disparate features, some of them are off the scale in their magnitude. In such cases, we obtain results which will be dominated by these features. We have used min max scaling to normalise the range of independent variables. This technique yields a value between 0 and 1. The shape of the distribution is retained so information described by original data is preserved. The MinMaxScaler is shown in equation 1, where X_{sc} represents scaling value, $\min(x)$ and $\max(x)$ represents the minimum and maximum values of x in all iterations.

$$X_{sc} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad \text{--- (1)}$$

2. Principal Component Analysis: A mathematical mechanism which utilizes orthogonal transformation for transforming features which may have correlation to give features which are linearly uncorrelated called principal components. PCA is a dimensionality reduction technique in which, the higher principal component will have larger possible variance compared to the succeeding components.

C. Training

A good training model is necessary for an accurate

prediction. The data after processing is trained to fit into the

model. In this paper, we train the model in two ways, linear

regression and Support Vector Regression.

1. Linear Regression:

It is a supervised machine learning technique, used to obtain the correlation among the explanatory variables and predicting the explained variables.

These models can be differentiated using the type of correlation among explained and explanatory variables, which are being considered and the number of explanatory variables being used. In this paper, we have used a multivariate regression with six independent variables. The equation of a multivariate regression is given by equation 2:

$$Y_{(m)} = a + b_1 X_{(m)} + b_2 X_{(m)}^2 + \dots + b_n X_{(m)}^n \quad (2)$$

$Y_{(m)}$ gives the approximation of m^{th} component of the explained variable Y which has n explanatory variables.

$X_{(m)}^j$ gives the m^{th} component of the j^{th} independent explanatory variable.

2. Support Vector Regression:

It is one of a regression method that maintains all the main features that characterise the algorithm. SVR is based on the computation which maps input data from a linear function of a higher dimensional feature space through a nonlinear function. [6] The idea behind using two different regression models is to compare the results obtained and to use the model with a better accuracy.

D. Testing

These steps are followed for Testing:

1. Load the pre trained model: The pre trained regression models are loaded.
2. Regression score prediction using test data points: model considers frequency and recurrence of each data point and obtains a best possible fit into the data.
3. Validation of result: Evaluation metrics such as explained variance score, $r2_score$ are calculated and predicted line is verified.

IV. RESULTS

The model is validated by calculating $r2_score$ and Explained_variance_score. Explained_variance_score tells us how much a statistical model explains the extent to which dispersion of the data. The r^2 score is the coefficient of determination, which represents the goodness of the fit and percentage of variance which has been elucidated by the independent variables in the model.

Comparison of performance metrics		
Name	Variance score	$r2_score$
Multivariate Linear Regression	0.95306	0.95238
Support vector Regression	0.94954	0.94906

Table 2. Result

V. CONCLUSION

After applying regression to the data, we have got a higher value of explained variation and r -squared score, with which we can conclude that it is possible to determine the depression quotient based on the length of music the person listened to. The depression quotient is obtained, which is used to determine whether an individual is depressed or not and categorise them based on the extent of depression.

REFERENCES

1. Benjamin L, Ana M, Pei Chen, Brian Mullin, Sherry Hou and Enrique "Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms", 2016
2. Hemanth P, Adarsh, Aswani C, Ajith P, Veena A, "EMO PLAYER", 2018
3. Sandra "Effect of music on depression", 2006
4. Dianna, Bronwen, "Performance-related musculoskeletal pain, depression and music performance anxiety in professional orchestral musicians", 2013
5. Dave, Michel, "Music listening, coping, peer affiliation and depression in adolescence", 2009
6. Debasish, Srimanta and Dipak "Support Vector Regression", 2007
7. Mandar Deshpande, Vignesh Rao "Depression Detection using Emotion Artificial Intelligence", 2017

AUTHORS PROFILE



Mr. Suyoga Srinivas PES University ECC, Department of Computer Science and Engineering, Bangalore, Karnataka, India Publication: Simulation tool to determine presence of impurities in an organic compound using IR spectroscopy, International Journal of scientific Research in Science and Technology.
Email: suyoga06031999@gmail.com



Mr. Naveen N Bhat PES University ECC, Department of Computer Science and Engineering, Bangalore, Karnataka, India
Email: naveennbhat22@gmail.com



Mr. Yashwanth Venkat Chandolu RNSIT, Department of Computer Science and Engineering, Bangalore, Karnataka, India
Email: ychandolu0131@gmail.com