Project Proposal

Team Member	Team Member	Team Member	Team Member
Dhaval Sonawane	Jignesh Darji	Rohit Mandge	Saurabh sakpal
dsonawa@ncsu.edu	jndarji@ncsu.edu	rnmandge@ncsu.edu	ssakpal@ncsu.edu

1 Title& Data

Music Rating Prediction System
https://www.kaggle.com/c/MusicHackathon/data

2 Project idea

Music rating prediction system intends to predict how much rating a user would give on a scale of 1-100 for an EMI company music track they are made to hear. This helps the company know how a particular track or an artist would be received upon launch and also helps them to design a music recommendation system by predicting songs for which the user will give a high rating.

This prediction has to be made considering various attributes of users and artists. The data available in the data set about various users, artists and track ratings will be merged into a single data frame and will act as an input to various regression models. We also intend to use following models for our project. Linear regression, Gradient Boosting Model and Random Forest.

The software that will be used primarily is R. The plan is also to use python if that will be better for data cleansing and manipulation and Tableau for data visualization. Various packages like glm and RandomForest will be used.

3 Division of work

Dhaval Sonawane, Jignesh Darji, Rohit Mandge and Saurabh Sakpal will be working together in the project. Dhaval will be primarily taking care of the data cleansing, preparation and merging. Jignesh and Rohit be will be focusing on implementing modeling techniques like Linear Modeling, Gradient Boosting and Random Forest etc. and Saurabh will mainly focus on visualization of these models and predicting the outcome. There also will be collaborative work while making the report and finalizing the results.

4 Midterm milestone

On the midterm, this team would have completed data pre-processing. For the remaining part, the team will look to optimize the model, generate visualizations, and suggest recommendations to improve the prediction rates.

5 Papers to read

[1] Majumdar, Abhishek, Arvind Kumar, and Sriram Manohar. "Music Recommendations Based on Implicit Feedback and Social Circles: The Last FM Data Set." N.p., n.d. Web

[2] Schneider, Astrid, Gerhard Hommel, and Maria Blettner. "Linear Regression Analysis." N.p., 5 Nov. 2010. Web

[3] Natekin, Alexey, and Alois Knoll. "Gradient Boosting Machines." N.p., 4 Dec. 2013. Web