Department of Computer Science and Engineering

Amrita School of Engineering, Coimbatore

Amrita Vishwa Vidyapeetham

End Semester Examination – November 2020

Seventh Semester

15CSE481 Machine Learning and Data Mining Laboratory

**SET-1**

*Rubric:*

CO1: Python library usage: 5 marks

CO2: EDA and visualization: 7 marks

CO4: Implementation and Results: 8 marks

*Question:*

For this problem you will experiment with association rule mining using **Apriori** algorithm.

The data set you will use is based on a music playlist data set obtained from [**Yes.com**](http://api.yes.com/). We will only use a portion of this data. The provided data archive contains two files. The file "playlists.txt" contains on each line a sequence of songs played as part of one playlist. The songs are represented by integer values. The file "song\_names.txt" contains the mapping between the integer codes and song tiles and artists (format of the song names is [song title]::[artist]). You will need both of these files to generate association rules.

Your tasks in this problem are the following:

1. Load the playlist data and songlist separately.
2. Visualize the popularity of songs.
3. Visualize the popularity of artists.
4. Run the Apriori algorithm on the playlist data using a min-support value of 0.002. (Note: For not less than 2000 samples.)
5. Generate rules using confidence as metric. you may set the min-confidence threshold to about 0.5.
6. Identify 3-4 rules and map the resultant rules from integer codes to song titles:artist. Explain why you think they are interesting.