

ABSTRACT

Music listeners have different interests in searching songs to listen to the song. They search songs based on artists, genres, and popular albums, besides some music listeners search for songs based on the theme/subject of the song. Searching songs based on themes/subjects is the most favored by listeners of music. This has been proven in online surveys in the previous studies. Basically, many music applications are able to categorize songs by genres, artists and albums. This is reasonable because in the audio file there is information about the artists, genres and albums so that music applications can automatically create playlists. However, to categorize songs based on themes/subjects requires a process to find out the theme/subject of the song. Machine learning is one of the solutions to categorize songs based on themes/subjects as has been done in previous research using lyrics as the object of research. In this study, an automatic classification system based on the subject using lyric, genre & artist data sources is created. As a result, it is found that the system performance with the addition of genre & artist information outperformed compared to using lyrics only. This study also tries to apply the two stages concept in the classification process and compares it to the single flat classification method implemented in previous research. The results indicate that the classification process with two stages classification method outperformed compared to single flat classification method both in the system performance and the efficiency of running time of the classification process. The system performance produced in this study using the Naïve Bayes method is able to produce an average value of 94.03% accuracy, 71.19% Precision, Recall 64.42, and F1-Measure 67.85%.

Keywords: text classification, machine learning, Song, Naïve Bayes