

## ABSTRAK

Lirik merupakan salah satu media komunikasi yang sering digunakan untuk menyampaikan perasaan dari seseorang. Salah satunya yaitu melalui lirik lagu yang dapat dijumpai oleh seluruh masyarakat melalui media online, buku dan lainnya. Lirik lagu ini digunakan sebagai data yang diperoleh melalui beberapa situs website, lirik lagu yang telah disalin akan disimpan dengan format ekstensi \*.txt agar mampu diolah oleh sistem.

Agar mempermudah dalam pemerolehan informasi data lirik lagu yang dicari, maka dilakukan pengklasifikasian data teks. Penelitian ini bertujuan untuk memudahkan pemilihan lagu yang baik dan tidak baik terhadap anak. Fungsi klasifikasi pada penelitian ini dibagi dalam dua klasifikasi yaitu baik dan tidak baik. Proses yang akan dilalui oleh data diantaranya membaca data, tokenizing, normalization, stop word, stemming, sorting dan grouping. Sistem ini akan menggunakan vektor ciri berupa TF (*Term Frequency*) yang merupakan jumlah kemunculan suatu kata dalam sebuah dokumen.

Proses klasifikasi teks lirik lagu menggunakan algoritma Multinomial Naïve Bayes. Harapannya adalah berdasarkan pengelompokan yang terbentuk melalui proses klasifikasi ini, dapat diketahui lagu yang baik dan lagu yang tidak baik untuk anak. Dengan adanya pengelompokan ini dapat membantu dalam menentukan lagu yang baik dan tidak baik untuk anak.

## ABSTRACT

A song is one of the media that is often used to express feelings for a person. Song texts amongst other can be found by all people online, in books and in other media. The songs used in this research were obtained from various online media and were transcribed and saved with \*.txt extension so that it can be processed through the system.

To facilitate the collection of information on songs that were searched, the text was classified. The purpose of this research is to facilitate the selection of songs that are appropriate and inappropriate for children. The classification function in this research was divided into two groups; appropriate and inappropriate songs for children. The data processing includes reading of the data, tokenizing, normalization, stop word, stemming, calculation of data frekuensi, sorting and grouping. This system uses vector feature Term Frequency, meaning the number of times a certain word appears in certain data.

The classification process of song texts used a Multinomial Naïve Bayes algorithm. The expectation is that the grouping based on this classification will specify which song is appropriate and which song is inappropriate for children. The grouping will assist in deciding which songs are appropriate and inappropriate for children.

