1. INTRODUCTION

This introductory chapter discusses the reasons for choosing the topic of song classification based on subject for the music listeners and how to solve the problem.

1.1 Rationale

The number of songs, especially in Indonesia increases in number and variety, this is evident from the number of singles and even albums launched by a group of music/singers every year. Therefore it needs a system that can categorize songs to help music listeners when searching for songs. Music listeners have different interests in searching songs to listen to the song they want. Among music listeners searches songs based on their favorite artist, genre, and albums, besides that some music listeners search for songs based on the subject of the song they want. In an online survey that has been done in previous research, when searching for songs 33.4% and 17.9% of 427 music listener respondents tend to select songs based on theme/main subject and storyline respectively. The number of respondents was higher than popularity, mood, time period or instrument [1]. This shows that categorizing songs based on the subject is needed for listeners. According to Harimurti Kridalaksana, the subject is part of the clause that marks what is spoken by the speaker [2], in this study the subject of the song is the main topic told by the singer in a song such as love, friendship & nationalism.

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. If the categorization of songs based on themes/subjects is conducted manually, it is very inefficient because it takes effort and a time consuming considering the number of songs in the world are numerous and countless, it is sure because in categorizing the song based on the main subject/ theme, playlist maker must listen to the song one by one and know the meaning of the songs. Therefore, one way to categorize songs based on main theme/ subject songs automatically can be run using machine learning.

The lyric is an important component of a song. Lyrics can be defined as a series of words to express feelings and thoughts of the song writer, so that with the lyrics of the song listeners can find out the message contained in a song [3]. Considering this fact, the lyrics can be used as one of research object in song classification based on subject [4]. In the study of classification of songs based on main subject, there are several other song components that can be used as parameter of song classification, such as user interpretation that has been used in previous research [1]. Based on previous study result [1], the researcher succeeded in proving that in solving poetic lyric problems by adding user interpretation. The system is able to have higher performance with user interpretation data and lyrics by comparing

only using lyrics. In this study, we still use lyric as one of research object. We can't use user interpretations because currently there is no website like songmeanings.com for Indonesian song but we add genre and artist information to overcome the poetic matter in classifying songs based on subjects.

In addition, this study create a system that can categorize songs based on the subject of songs in two stages by using song lyrics, genre, artist as a new concept that has not been done on the previous subject classification research. This two stages concept is needed because some subjects of songs can be divided into more specific subjects. As an example of a love song, ("what kind of love song?") can be divided into falling in love songs (happy) or broken heart songs (sad). Another example, songs with religion subject can be divided into types of religion (Islam, Christian, Hindu, Buddhist). This two stages classification can help song listeners find songs with more specific subjects.

In general, this research consists of several stages: preprocessing data, feature extraction, classifier building, classification and system evaluation. Processing data consists of four stages: case folding, tokenization, stopword removal and stemming. The stemming algorithm used in this preprocessing data is Nazief-Andriani algorithm because this algorithm is capable of achieving 93% success rate [5]. Preprocessing genres and artists are conducted by label encoding process, so the categorical data of artists and genres are converted into numerical data. After preprocessing data, feature extraction is done by using the bag of word representation. The next stage, the two stages classifier model building is done by using Naïve Bayes Method. Furthermore, the classification process is done to test the system that has been built using testing data, where each song lyrics on the test data given a label of the class and subclass. The topic class labels generated in this research are love (fallin in love, Broken Heart), Friendship (Good Friendship, Bad Friendship), Family (Dad, Mom, Child), Nasionalism (Tanah Air Song, Politic), Religion (Islam, Christian, Hindu, Buddhist) And Negative Content (Violence, Pornography, Insulting Religion) that aims to filter songs containing harmful content. Each of these topic has several sub-topics as the more specific topic.The final stage is that the result of this classification is used to evaluate the system by calculating accuracy, precision, recall and f1-measure.

1.2 Problem Formulation

Based on the background, the problems that can be formulated in this research are:

1. How the effect of adding artist and genre features to system performance in classifying songs based on the subject.
2. How does two stages classification affect system performance and time consumtion in classifying songs based on the subject.

1.3 Objective

Based on the formulation of the above problems, the objective of this study are to:

1. Classify songs based on the subject with lyric data and addition of artist, genre features and compare the system performance to the use of different features.
2. Classify songs based on the subject in two stages, and compare it to the single flat classification method used in previous studies.

1.4 Hypotesis

The addition of genre and artist features in classifying songs based on the subject is able to minimize classification errors, so that the system is able to produce higher performance. This is because the system does not only use words in the lyrics to determine the subject class of a song but also by knowing the probability of the artists and genres of the subject class to be produced. The application of the two stages classification concept in this study is able to improve the efficiency and performance of the system compared to the flat classification method.

1.5 Problems Limitation

In order to ensure that the scope of this issue does not extend to the unrelated aspect, the scope limitation of the problem needs to be determined. The scope limits of the problem in this study are as follows:

1. The song documents used are Indonesian songs consisting of lyrics, genres & artists data taken on several websites providing song lyrics (example Lirik.kapanlagi.com) as many as 1149 data.
2. The song document used is a song that only has one class on the class list used in this study, so that this system is not able to produce multi label classification.
3. Instrumental songs are not covered in this study.
4. Data input format in this system uses the Comma Separated Values (.csv)
5. The programming language used is Java and the database used is SQL
6. The topic class labels generated Are Love (Fallin in Love, Broken Heart), Friendship (Good Friendship, Bad Friendship), Family (Dad, Mom, Child), nasionalism(Tanah Air Song, Politic), religion (Islam, Christian, Budha, Hindu), dan Negative Content (Violence, Pornography, Insulting Religion).

1.6 Research Methodology

The methodology used in this research is as follows:

1. Problem Identification
   In the problem identification stage, a literature study was conducted from previous research related to the classification text on the song to find out the problems used as the material in this study. This stage also explored and identified the problems of various song provider applications in categorizing songs and finding solutions.
2. Requirement Identification
   After identifying the problem, the requirement identification is then carried out consisting of research needs and system requirements. Identification of needs related to materials and methods needed in the classification text on song were based on subject research. Identification of system requirements related to software and hardware specifications were needed to support research.

3. System Design
   At this stage system design was conducted to define the stages in creating a system classifying songs in this study.
4. Data Collection and Hand Labelling
   After designing the system, datasets were collected from various websites providing songs such as lirik.kapanlagi.com that provided song lyrics from various artists. Furthermore, the data collected were carried out hand-labeling process consisting of main class and subclass.
5. System Implementation
   At this stage, the implementation of a system previously designed was carried out. In this implementation process, the system used data that had been collected previously, then the data were preprocessed, feature extraction with a representation of bag of words, and the construction of a classifier using the naive bayes method
6. System Testing & System Evaluation
   At this stage, the system built was tested using data testing to determine system performance.
7. Results Analysis
   The last stage, the results of previous tests were analyzed and conclusions from this study were made.

1.7 Systematics Writing

The systematics of the final report writing in this study consists of five chapters as follows:

1. INTRODUCTION
   This chapter describes the background of research, problem formulation, objectives, Hypotesis, problem limitation, research methodology and systematic research writing.
2. LITERATURE REVIEW
   This chapter discusses previous research related to classification texts on song domains, and discusses methods that support this research consisting of preprocessing data, feature extraction, classification methods and evaluation system methods.
3. RESEARCH METHODOLOGY AND SYSTEM DESIGN
   This chapter describes the research methodology and system design that is described by the flow chart with descriptions of each stage.
4. TESTING AND RESULTS ANALYSIS
   This chapter contains the results of system testing and analysis of the results obtained in this study by comparing several testing conditions.
5. CONCLUSIONS AND SUGGESTIONS
   This chapter contains conclusions obtained from this study and suggestions for further research.