

# Abstract

Analyzing the emotional tone of songs' texts can give insights into societal trends and this information can be useful especially for recommendation algorithms. This study focuses on developing four Machine Learning models to classify emotions conveyed in English song lyrics at the stanza level. The classification employs Plutchik's eight primary emotions, offering a nuanced understanding of emotional expression in lyrical content. The selected model architectures are:

- **Random Forest**
- **Support Vector Machine (SVM)**
- **One-Dimensional Convolutional Neural Network (1D-CNN)**
- **Recurrent Neural Network (RNN)**

**DA FINIRE CON CONCLUSIONI e far stare nella stessa pagina con l'introduzione**

# Introduction

Songs have the unique ability to engage people in ways that few other mediums can match. Lyrics serve as one of the main foundations of songs, playing a crucial role in expressing feelings in many different ways. The emotional tone of songs can serve various purposes, such as automatized playlist creation or songs' organization, offering an alternative to the more traditional genre-based classification.

The goal of this project is the development of 4 Machine Learning models that perform emotion detection on songs' stanzas. To obtain a deeper understanding of emotional fluctuations within the texts, the models assign emotion labels to individual stanzas instead of full songs. The emotion labels correspond to Robert Plutchik's eight primary emotions (shown in figure 1), providing a comprehensive range for representing various emotional states.



Figure 1: Plutchik's eight primary emotions

This report aims to clearly cover and illustrate various aspects of the work. The *Methods* section provides a detailed explanation of the data and procedures used in the project, with a particular focus on the pipeline implemented to construct the models. Following this, the *Results* chapter presents an overview of the obtained outcomes, using plots and figures to highlight significant findings. These results are further explored in the final sections, *Discussion* and *Conclusions*, which interpret the general findings, recap the primary objectives of the work, and discuss the importance or potential applications of the results.

# 1. Methods

The dataset used in this project is a sampled subset of English-language songs derived from the *Genius Song Lyrics Dataset*<sup>[geniusdataset]</sup>. The original dataset contained numerous attributes; the ones considered relevant for model training are:

- **title:** the song's title;
- **lemmatized\_stanzas:** lyrics of the single stanza;
- **stanza\_number:** identifies the position of the stanza in the song;
- **is\_chorus:** boolean variable that attests whether the stanza is a chorus or not;
- **tag:** represents the genre of the song. For easier handling, this attribute of the original dataset has been one-hot encoded into various boolean variables (`is_country`, `is_pop`, `is_rap`, `is_rb`, `is_rock`);
- **label:** represents the emotional classification of the stanza, assigned by Albert Base v2<sup>[albert-base-v2]</sup> model.

All of these attributes, except for the `title` one, were the result of the preprocessing phase, as will be described in section 1.1. Due to limited computational power, the labeling process was time-intensive, ultimately resulting in a limited dataset consisting of (QUANTE? AGGIUNGEREI NUMERO STROFE).

## 1.1 Preprocessing

The first step in the preprocessing phase of this dataset involved sampling from the original dataset while preserving the proportions of the different genres. This ensured that the genre distribution in the subset remained representative of the full dataset.

The preliminary text cleaning process focused on the `lyrics` attribute, which contained the complete lyrics of each song in string format. Initially, a regular expression (RegEx) was built to remove

noise from the lyrics, specifically targeting words enclosed in square brackets that were irrelevant to the stanza splitting process. Many keywords marking different stanzas were written within square brackets, and removing non-keyword items inside brackets was crucial to avoid potential issues.

The next critical step was stanza splitting. After cleaning the strings from noisy square-bracketed items, the lyrics were split based on various keywords used to denote stanzas (such as "chorus", "verse", "intro", "outro", "refrain", "hook", etc.). The RegEx developed accounted for the different formats in which these keywords appeared, including square brackets, parentheses, or no brackets at all, as well as stanzas separated only by double newline characters. The output of this step was, for each song record, a list of strings representing individual stanzas (each stanza has also a header with the corresponding keyword; this aspect will be discussed in the next paragraph). Next, uninformative strings—such as empty strings or those with fewer than 20 characters—were removed, as they were too short to provide meaningful content. As a result, the output of this preliminary preprocessing phase was a dataset where the records were no longer whole songs but individual stanzas, each numbered according to its position within the song.

A further and more detailed cleaning process on the stanzas involved the creation of the boolean feature `is_chorus`, which was assigned a true value for repeated stanzas within the same song or for stanzas with headers such as "hook", "chorus", "refrain", or "bridge". Next, stanza headers and newline characters between verses were removed to obtain cleaner stanzas. Since choruses, hooks, bridges, and refrains often repeat throughout songs, duplicate stanzas were discarded to avoid redundant data. This resulted in a dataset of cleaned, non-duplicate stanzas, which served as the checkpoint for the labeling step and the starting point for the text lemmatization process.

The subsequent step involved lemmatizing the stanzas using the spaCy library. A list of lemmatized tokens was created by filtering out punctuation and empty words. Lemmatization was chosen over stemming because it produces more accurate and meaningful results, particularly for tasks requiring semantic understanding, such as the one at hand.

Since the dataset was not pre-labeled at the stanza level, the ALBERT Base v2 transformer model

was employed for this task. This transformer model is specifically designed to be fine-tuned on tasks that require an understanding of the entire sentence, such as sequence classification.

## **1.2 Models developement**

## **1.3 Evaluation**

## **2. Static Models**

### **2.1 Random Forest**

### **2.2 SVM**

## **3. Neural Networks**

### **3.1 One-Dimensional Convolutional Neural Network**

### **3.2 Recurrent Neural Network**

## **Key findings and conclusions**



List of figures

1 Plutchik’s eight primary emotions . . . . . 2