**LAPORAN PROYEK PENGOLAHAN DATA BESAR**

**Sentiment Analysis of Music Topic Review in Twitter Application With Text Blob**



**Disusun Oleh:**

|  |  |  |
| --- | --- | --- |
| 1 | 12S18003 | Citra Hutajulu |
| 2 | 12S18004 | Rosalia Pane |
| 3 | 12S18017 | Putri Yohana Panjaitan |
| 4 | 12S18030 | Unedo Ignasius Manalu |

**PROGRAM STUDI SARJANA SISTEM INFORMASI  
FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO  
INSTITUT TEKNOLOGI DEL  
MEI 2022**

**DAFTAR ISI**

**DAFTAR GAMBAR**

**DAFTAR TABEL**

**BAB 1**

**PENDAHULUAN**

**1.1 Latar Belakang**

*Twitter* merupakan salah satu sosial media yang sangat populer di masyarakat. Pada *twitter*, setiap pengguna berinteraksi dengan pengguna lainnya melalui sebuah pesan yang biasa disebut *tweet.*  Bentuk interaksi pada twitter adalah singkat dan jelas. Hal itu memungkinkan penggunanya lebih mudah untuk mendapatkan dan mengetahui informasi[1]. Twitter juga digunakan untuk mengungkapkan emosi mengenai sesuatu hal, baik untuk memuji ataupun mencela dalam bentuk emosi.

Twitter juga menyediakan api twitter yang bertujuan agar para pengembang dapat dengan mudah untuk melakukan analisis terhadap data pada twitter. Analisis yang pada umumnya digunakan adalah sentimen analisis. Sentimen Analisis *(opinion mining*) merupakan suatu implementasi yang diperoleh ketika penambangan teks ( *text mining* ) digunakan untuk mengklasifikasikan informasi berupa teks berdasarkan sentimen/pandangan subjektif pada teks tersebut. Analisis sentimen dilakukan secara otomatis menggunakan algoritma pembelajaran mesin (*machine learning*). Algoritma ini akan mendeteksi kecenderungan opini yang terkandung dalam suatu teks dan akan diklasifikasikan menjadi kelas positif, negatif, ataupun netral. Tujuan utama dilakukan sentimen analisis adalah untuk mengklasifikasikan polaritas informasi dalam dokumen baik kalimat maupun kata-kata. Analisis sentimen bekerja secara sistematis untuk mengidentifikasi, mengekstrak, dan mempelajari kondisi dan informasi subjektif yang secara luas diterapkan pada analisis pendapat konsumen, ulasan, tanggapan survey, dan media sosial.

Data yang digunakan pada sentimen analisis adalah data yang di streaming lewat twitter menggunakan apache spark. Apache spark berguna untuk membangun jalur pemrosesan data dengan mudah dan dapat mengakses data dari berbagai sumber. Apache spark dapat juga menyediakan data yang kemudian akan dilakukan text mining dan klasifikasi terhadap review atas data pada twitter sehingga mendapat informasi baik buruknya review tersebut.

Salah satu fitur yang ada pada twitter yaitu topic mengenai musik, dimana ketika topik musik sedang trending maka pendengar dapat memberikan komentar berupa review untuk musik tertentu. Pendengar musik tersebut akan menilai baik dan buruk nya musik tersebut.Oleh karena itu, dalam penelitian ini, peneliti ingin melakukan sentimen analisis terhadap salah satu topik di twitter yakni musik. Hal ini dilakukan peneliti agar mendapatkan seberapa banyak review positif dan negatif pada topik music dengan menggunakan text mining dan support vector machine.

**1.3 Rumusan Masalah**

Adapun rumusan masalah yang dilakukan pada proyek ini adalah:

1. Bagaimana cara memperoleh data pada *streaming* data pada *twitter*
2. Bagaimana cara untuk mengklasifikasikan informasi *review* musik pada *twitter* untuk mengetahui baik dan buruknya sebuah *review*

**1.2 Tujuan**

Adapun tujuan dilakukannya proyek pengolahan data besar ini yaitu:

1. Menerapkan *sentiment analysis* pada *streaming* data *twitter* untuk *tweet* yang berisi review musik tertentu.
2. Mengklasifikasikan informasi mengenai *review* musik pada *twitter* dengan menggunakan *Text Mining* dan *Support Vector Machine* untuk mengetahui baik buruknya sebuah *review*.

**1.3 Manfaat**

Manfaat dari pengerjaan proyek adalah sebagai berikut:

1. Mengetahui cara dan proses analisis menggunakan *sentiment analysis* pada *streaming* data *twitter.*
2. Mengetahui seberapa banyak *review* positif dan negatif dari *review* musik *twitter* berdasarkan klasifikasi menggunakan *Text Mining* dan *Support Vector Machine.*

**1.4 Ruang Lingkup**

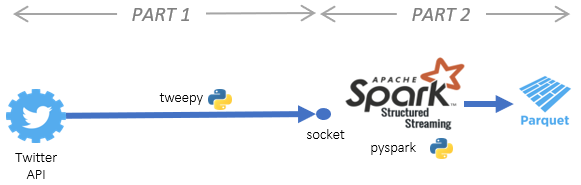
Ruang lingkup dalam pengerjaan proyek ini yaitu:

1. Algoritma yang digunakan dalam pengerjaan proyek ini adalah *support vector machine* dan *text mining* untuk mengklasifikasikan data.
2. Data yang digunakan adalah review positif dan negatif pada topik music yang ada pada *twitter* dengan melakukan API streaming.

# **BAB 2 SOLUSI**

## **2.1 Big Data Architecture**

Berikut ini merupakan big data architecture yang akan digunakan sebagai solusi dari permasalahan pada rumusan masalah.



Dataset yang digunakan dalam pengerjaan proyek ini adalah data\_tweet\_music dalam format CSV dan untuk meload data yaitu menggunakan apache spark.