

LAPORAN AKHIR PROYEK PENGOLAHAN DATA BESAR

SENTIMENT ANALYSIS PADA

STREAMING DATA TWITTER UNTUK GAME

MOBILE LEGEND



DISUSUN OLEH :

12S18001 Cindy A. Siregar

12S18014 Giovanni Situmorang

12S18021 Lastri S. N. Marbun

12S18061 Angela F. Simamora

PROGRAM STUDI SARJANA SISTEM INFORMASI

FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO

INSTITUT TEKNOLOGI DEL

2022

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

I. PENDAHULUAN

Pada bab ini akan membahas mengenai latar belakang, tujuan serta ruang lingkup dari pengerjaan proyek yang akan dilakukan.

1.1 Latar Belakang

Analisis Sentimen adalah proses untuk mengidentifikasi dan menganalisis sebuah opini sehingga pengguna lebih mudah memahami maksud dari suatu data dengan menggunakan teknik pemrosesan data. Dengan menggunakan bantuan teknologi ini kita bisa mengidentifikasi lebih banyak jenis sentiment seperti tweet, review hotel, customer review dan masih banyak lagi yang nantinya dapat dimanfaatkan.

Analisis sentimen menyatukan berbagai bidang penelitian seperti pemrosesan bahasa alami, penambangan data, penambangan teks, serta penting bagi organisasi karena mereka berusaha untuk mengintegrasikan metode *computational intelligence* ke dalam operasi mereka, dan upaya untuk menjelaskan lebih lanjut, meningkatkan, produk dan layanan mereka. Analisis sentimen juga memiliki aplikasi di mesin pencari yang berorientasi pada ulasan, ringkasan ulasan, dan untuk memperbaiki kesalahan dalam peringkat pengguna (seperti untuk kasus di mana pengguna secara tidak sengaja memilih peringkat yang salah ketika ulasan mereka menunjukkan evaluasi yang berbeda). Dalam analisis sentimen, tujuannya yaitu untuk menemukan pendapat orang yang diungkapkan dalam bahasa tertulis (teks). Sentimen dalam istilah berarti "apa yang dirasakan seseorang tentang sesuatu", "pengalaman pribadi, perasaan sendiri", "sikap terhadap sesuatu" atau "pendapat".

Twitter merupakan salah satu media sosial yang memiliki konsep sederhana dan cepat. Karena merupakan pesan singkat, berita atau informasi di Twitter bisa lebih mudah dicerna. Media sosial ini juga banyak digunakan sebagai objek para peneliti atau industri untuk melakukan analisis sentimen di bidang sosial, ekonomi, politik, atau bidang lainnya. Opinion mining atau juga biasa disebut dengan analisis sentimen adalah proses menganalisis teks untuk mendapatkan informasi tertentu dalam sebuah kalimat berupa opini. Analisis sentimen merupakan salah satu cabang dari ilmu text mining. Penambangan teks adalah teknik pemrosesan bahasa alami dan metode analisis yang diterapkan pada data teks untuk mendapatkan informasi yang relevan. Opini atau sentimen publik di twitter sangat dinamis dan cepat berubah; sistem analisis sentimen real-time diperlukan. Secara otomatis diperbarui terus menerus sehingga perubahannya selalu dapat dipantau, kapan saja dan di mana saja.

Penelitian ini membangun sebuah sistem untuk menganalisis sentimen dari twitter secara real-time, otomatis, dan berkelanjutan. Hasil uji coba sistem berhasil mengambil data, melakukan analisis sentimen, dan menampilkannya dalam grafik dan website real-time, dan diperbarui secara otomatis. Selanjutnya, penelitian ini akan dikembangkan dengan fokus pada keakuratan algoritma yang digunakan dalam melakukan proses analisis sentimen.

Permasalahan proyek Pengolahan Data Besar yang akan diselesaikan adalah melakukan klasifikasi analisis sentimen pada *streaming* data twitter. Sumber data yang akan digunakan merupakan data mentah yang diperoleh dari Twitter dengan memanfaatkan layanan API Stream yang disediakan oleh Twitter. Data mentah diproses sampai analisis sentimen yang diharapkan dihasilkan. Sebelum dilakukan analisis sentimen terhadap data yang diperoleh dari Twitter, dilakukan pre-processing. Pada tahap ini, pembersihan data tweet dilakukan dengan langkah-langkah sebagai berikut: *folding case*, penghapusan simbol, tokenisasi, konversi slang word, dan penghapusan stopword.

1.2 Tujuan

Adapun tujuan dari pengerjaan proyek ialah untuk melakukan analisis sentimen terhadap dataset *live streaming twitter* dengan menggunakan *keyword game mobile legend*.

1.3 Rumusan Masalah

Bagaimana suatu sistem dapat mengklasifikasikan suatu *game* berdasarkan *review* yang diberikan *gamers* sebagai ulasan yang bernilai positif atau ulasan yang negatif.

1.4 Scope

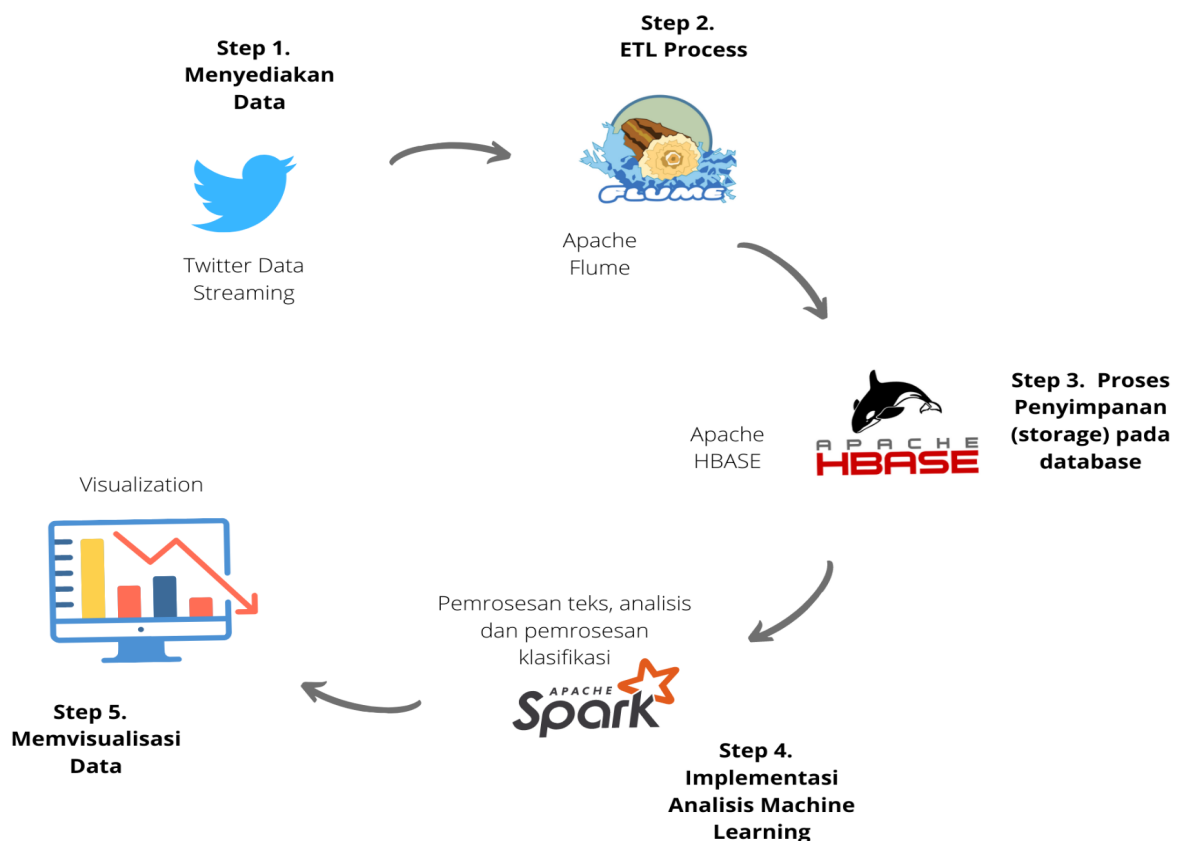
Adapun yang menjadi batasan dari pengerjaan proyek sentimen analisis pada dataset *live streaming twitter* sebagai berikut.

1. Algoritma yang digunakan dalam melakukan analisis sentimen terhadap *Games Mobile Legend Streaming Data Twitter* adalah Naive Bayes
2. Data yang digunakan adalah data *games mobile legend* yang diperoleh dari Twitter dengan memanfaatkan layanan API Stream yang disediakan oleh Twitter

II. SOLUSI

Pada bab ini diberikan sebuah arsitektur sistem yang dapat menjadi solusi yang digunakan dalam masalah pengklasifikasian sentiment analysis *Live Streaming Data Twitter*. Saat ini *games mobile legend* merupakan sebuah produk aplikasi yang sangat populer. Maraknya perkembangan *game mobile legend* juga menimbulkan tantangan seperti banyaknya pendapat yang diberikan oleh pengguna *mobile legend*. Analisis Sentimen dapat digunakan untuk memberikan informasi sentimen yang terjadi pada komentar-komentar mengenai *games mobile legend* dari streaming data twitter. Informasi yang diberikan yaitu berupa klasifikasi sentimen positif dan negatif serta metode yang digunakan yaitu *Naive Bayes Classifier*.

Arsitektur big data merupakan suatu struktur keseluruhan yang mempresentasikan dari logical dan physical sistem dari big data yang akan disimpan, diakses serta diolah.



Gambar 1. Arsitektur Sistem

Pada gambar diatas dijelaskan alur arsitektur sistem yang akan digunakan dalam melakukan klasifikasi *sentiment analysis* pada *Live Streaming Twitter*. Arsitektur sistem dibentuk agar

dapat mengolah data yang terstruktur dan tidak terstruktur. Selanjutnya adalah menggunakan spark untuk melakukan implementasi analisis *Machine Learning*.

1. Menyediakan data

Hal pertama yang harus dipersiapkan adalah data. Data terdiri dua jenis yaitu terstruktur dan tidak terstruktur. Data yang terstruktur dapat diketahui apabila data tersebut telah tersimpan dalam sebuah format file. Untuk data yang tidak terstruktur merupakan data yang tidak mengikuti suatu susunan format tertentu sebagai contoh data yang berasal dari sebuah media sosial atau website.

2. ETL Process

Setelah menentukan jenis data yang digunakan, tahap selanjutnya adalah ETL Process. ETL atau Extraction, Transfer, Loading merupakan sebuah proses sebuah fase pemrosesan data dari sumber data ke dalam satu penyimpanan yang konsisten dan dimuat ke dalam gudang data.

3. Proses penyimpanan (*storage*) pada database

Setelah data diproses kemudian data akan disimpan dalam HBase. Hbase merupakan database terdistribusi yang berorientasi pada kolom. HBase adalah program yang berjalan diatas Hadoop Distributed File System (HDFS) yang mampu memproses data dalam skala besar secara interaktif.

4. Implementasi analisis *Machine Learning*

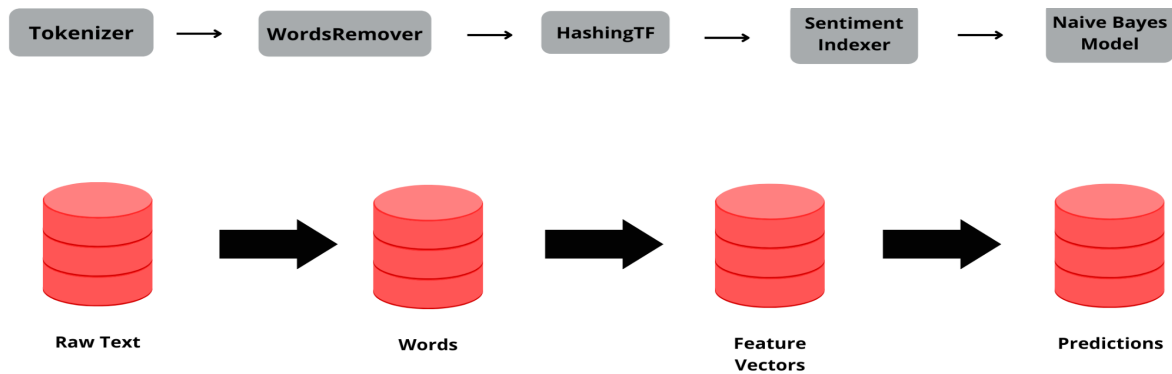
Pada tahap ini, data yang disimpan akan dilakukan pemrosesan data dimana berguna untuk menghilangkan data yang noise, kemudian data yang telah dilakukan preprocessing, akan dipakai kembali untuk melakukan klasifikasi analisis sentimen pada Live Streaming Data Twitter

5. Melakukan visualisasi data

Pada tahap ini, hasil klasifikasi sentimen analisis pada Live Streaming Data Twitter akan diolah kembali untuk divisualisasikan berdasarkan hasil klasifikasi sentimen analisis untuk mempermudah dalam memahami hasil pengolahan data.

III. MACHINE LEARNING PIPELINE

Pada bab ini akan dijelaskan Machine Learning Pipeline yang digunakan dalam kasus *Sentiment Analysis - Game Mobile Legend Streaming Data Twitter*.



Gambar 2. Machine Learning Pipeline

Berikut tahapan Machine Learning Pipeline yang akan digunakan sebagai berikut:

1. Tokenizer

Tokenizer perasi memisahkan teks menjadi potongan-potongan berupa token, bisa berupa potongan huruf, kata, atau kalimat, sebelum dianalisis lebih lanjut. Entitas yang bisa disebut sebagai token misalnya kata, angka, simbol, tanda baca, dan lain sebagainya. Tujuannya untuk mengeksplorasi kata-kata dalam sebuah kalimat. Daftar token menjadi masukan untuk diproses lebih lanjut.

2. Stop Words Removal

Terdapat banyak kata dalam dokumen yang sering kali muncul pada dasarnya tidak memiliki arti karena digunakan untuk menggabungkan kata dalam kalimat. Stop Word Removal merupakan penghapusan kata-kata pada dokumen yang tidak memiliki arti tersebut.

3. Hashing TF

HashingTF merupakan proses untuk melakukan transformer yang nantinya akan mengambil kata dan mengubah kata tersebut menjadi suatu vektor dengan panjang yang tepat.

4. Naive Bayes Model

Naive Bayes Model dibangun dengan menggunakan algoritma Naive Bayes. algoritma Naive Bayes merupakan algoritma yang digunakan untuk klasifikasi multi

kelas. Naive Bayes menerapkan fungsi statistik sederhana berdasarkan teorema bayes dengan asumsi keberadaan dari suatu fitur tertentu terhadap suatu kelas yang tidak berhubungan dengan fitur lainnya.

IV. IMPLEMENTASI

On Progress

V. HASIL DAN PEMBAHASAN

On Progress

VI. KESIMPULAN

On Progress