BABI

PENDAHULUAN

1.1. Latar Belakang

YouTube adalah situs yang paling banyak digunakan untuk berbagi video. Tidak hanya untuk melihat video hiburan, tetapi juga sebagai sarana untuk belajar[1]. Kepopuleran tersebut sangatlah terasa di Indonesia. YouTube dapat digunakan sebagai sumber penghasilan dengan menjadi konten creator. Selain itu, banyak penyanyi menggunakan YouTube sebagai situs untuk berbagi video musik mereka[2]. Selain melihat video pengguna bisa memberikan opini, saran, atau kritik pada kolom komentar. Komentar pada video YouTube dapat menjadi sumber data yang digunakan untuk kebutuhan analisis sebuah emosi atau pesan yang ingin disampaikan para pengguna. Para pengguna dapat memberikan pendapat mereka dengan cara menyampaikan opini, cerita, dan emosi pada kolom komentar secara bebas. Komentar pada YouTube dapat diolah menggunakan *machine learning* agar menghasilkan sebuah informasi yang nantinya informasi tersebut dapat digunakan sebagai ilmu pengetahuan[3].

Memahami pemaknaan pada pesan yang tersampaikan di kolom komentar pada video YouTube, sangat penting bagi konten kreator untuk bisa mengukur dan membuat video yang bisa diterima oleh pengguna. Dalam proses membuat sebuah informasi dari data akan diproses menggunakan pendekatan text mining dengan Natural Language Processing[4]. Penggunaan metode clustering akan memudahkan pengelompokkan data yang nantinya menjadi informasi. Setelah mendapatkan informasi tersebut cluster tersebut akan dicari pemaknaanya.

Dari uraian di atas menunjukkan bahwa perlu dilakukan penelitian "Pemaknaan Semantik Berbasis *Text Mining* Pada Komentar Video Klip Lagu Kali Kedua". Penelitian ini bertujuan untuk mengetahui makna pesan dari para penonton video musik Raisa – Kali Kedua yang telah berkomentar, sehingga

dapat diketahui relasi dari makna komentar dengan musik video tersebut.

1.2. Rumusan Masalah

Berdasarkan latar belakang, maka dapat diketahui rumusan masalah dalam penelitian ini yaitu.

- 1. Bagaimana *clustering* menggunakan metode *K-means*?
- 2. Bagaimana word embedding menggunakan Word2vec?
- 3. Bagaimana makna dari komentar yang telah di *clustering* pada setiap *cluster*?

1.3. Batasan Masalah

Berdasarkan permasalahan yang sudah dijelaskan, terdapat beberapa batasan tertentu dalam melakukan penelitian, yaitu.

- 1. Penelitian ini menggunakan bahasa pemrogaman python.
- Data diambil dari kolom komentar YouTube video klip musik Raisa Kali kedua pada tanggal 2 Agustus 2021 dengan alamat web https://www.youtube.com/watch?v=SHj2kJzVi g.
- 3. Word embedding menggunakan metode Word2vec.
- 4. Pencarian jumlah *cluster* menggunakan *Elbow Method*.
- Pemaknaan menggunakan pendekatan teori semantik segitiga Ogden dan Richard's.

1.4. Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah, maka tujuan penelitian ini adalah sebagai berikut.

- 1. Mengetahui implementasi *Clustering* pada data teks komentar dari YouTube menggunakan metode *K-means*.
- 2. Mengetahui implementasi *word embedding* pada data teks komentar dari YouTube menggunakan metode *Word2vector*.
- 3. Mengetahui implementasi pemaknaan suatu pesan menggunakan pendekatan teori semantik segitiga dari Ogden dan Richard's.

1.5. Metode Penelitian

1. Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset yang diambil dari kolom komentar pada video musik Raisa – Kali Kedua. Pengambilan data menggunakan YouTube API v3. Pengambilan data dilakukan pada tanggal 2 Agustus 2021. Pengambilan data tersebut menggunakan script. Setelah script berlangsung maka data akan didapatkan dengan format file CSV (Comma Separated Value). Dataset yang telah diambil memiliki lima kolom dan 6.480 baris.

Dataset yang telah diambil memiliki lima kolom dan 6.480 baris. Kolom tersebut adalah Name, Comment, Time, Likes, dan Reply Count. Dari lima kolom yang telah didapatkan peneliti memilih data pada kolom Comment sebagai bahan penelitan.

2. Pemilihan *Library*

Pemilihan *library* dibutuhkan untuk mendukung penelitian agar memudahkan peneliti untuk mengolah data. Peneliti memilih beberapa *library* untuk proses *data preprocessing*, *data explorarion*, *clustering* dan evaluasi hasil dari penelitian. *Library* yang digunakan berbasis bahasa pemrogaman *Python*. *Library* sendiri berasal dari pengembangan *Project Open Source*.

3. Data Preprocessing

Data Preprocessing merupakan tahap pembersihan data. Pada proses ini data tersebut diolah dari raw data atau data mentah menjadi data yang lebih bersih. Proses ini dapat dikatakan sebagai langkah awal untuk mengambil suatu informasi yang tersedia. Setelah proses ini, data yang dihasilkan akan siap ke tahap selanjutnya.

4. Text Processing

Setelah data melalui tahap *preprocessing* data yang bersih akan diolah lagi menjadi data yang siap untuk proses. Pada tahap ini, data akan

di proses dengan langkah - langkah sebagai berikut.

a. Corpus Tokenization

Pada tahap ini, kumpulan teks data akan dipecah menjadi token.

b. Stopwords Removal

Proses pembuangan *stopwords* adalah proses pembuangan *term* yang tidak memiliki arti atau tidak relevan. *Term* diperoleh dari tahap *tokenization* akan diperiksa dalam suatu *library* yang memuat daftar *stopwords*, jika kata tersebut ada di dalam daftar *stopwords*, maka kata itu akan dihapus atau tidak akan diproses lebih lanjut.

c. Noise Removal

Noise Removal merupakan proses penghapusan kata-kata non-ASCII (American Standard Code for Information Interchange) dan angka.

d. Stemming

Proses *Stemming* adalah proses pembentukan kata dasar. Proses ini bertujuan untuk memperoleh kata dasar agar dapat mendapatkan informasi yang relevan.

5. Data Exploration

Data Exploration adalah eksplorasi data, setelah data telah melalui tahap preprocessing. Tahap ini bertujuan untuk mengetahui atau memahami isi dan komponen data. Proses ini penting sebelum penentuan model untuk tahap selanjutnya.

6. Word embedding

Word embedding adalah fungsi parameter yang memetakan kata ke dalam vector. Word embedding sendiri tidak perlu membutuhkan anotasi yang berarti langsung diturunkan dari corpus. Word embedding dapat dibuat langsung atau menggunakan pre-trained yang berarti word embedding yang telah dilatih menggunakan data set yang besar pada permasalahan tertentu.

7. Cluster Analysis

Clustering adalah proses pengelompokkan berdasarkan adanya kemiripan satu sama lain. Setelah data melewati proses clustering data

akan divisualisasikan. Visualisasi data yang telah diolah akan dianalisis dan menjadi suatu informasi.

8. Pemaknaan Pesan

Setelah data sudah menjadi sebuah informasi, selanjutnya *cluster* yang telah diproses tersebut akan diteliti pemaknaanya dengan pendekatan teori semantik segitiga dari Ogden dan Richard's. Pemaknaan tersebut telah divalidasi *expert judgement* yang dilakukan oleh Ike Devi Sulistyaningtyas S.Sos., M.Si.

1.6. Sistematika Penulisan

Sistematika dalam penulisan tugas akhir "Pemaknaan Semantik Berbasis Text Mining Pada Komentar Video Klip Lagu Kali Kedua" akan ditulis menggunakan susunan seperti berikut:

BAB I PENDAHULUAN

Bab pendahuluan berisi tentang latar belakang masalah yang akan diselesaikan. Selain latar belakang, pada bab ini berisi rumusan masalah, batasan penelitian, tujuan penelitian, dan metode penelitian. Rumusan masalah berisi tentang tulisan singkat berupa pertanyaan yang nantinya akan terjawab setelah penelitian ini selesai. Batasan penelitian berisi tentang batasan yang harus ditentukan agar tidak meluas, sehingga dapat membatasi permasalahan yang ada. Metode penelitian berisi tentang metode yang digunakan untuk menyelesaikan penelitian ini.

BAB II TINJAUAN

Bab tinjauan pustakan berisi refrensi yang digunakan penulis dalam pembuatan penelitian agar hasil penelitian lebih baik dari pada eksperimen sebelumnya. Pada refrensi yang ada, penulis dapat menjadikannya sebagai tolak ukur dalam pembuatan penelitian yang ada.

BAB III LANDASAN TEORI

Bab landasan teori berisi tentang teori yang akan digunakan untuk penelitian ini. Terdapat penjelasan tentang teori dan tools sebagai penunjang pembuatan penelitian ini. Selain penjelasan teori dan *tools*, terdapat beberapa tentang data yang akan digunakan untuk penelitian.

BAB IV DATASET DAN PENGEMBANGAN MODEL

Bab *dataset* dan pengembangan model berisi tentang deskripsi masalah yang akan diselesaikan dari *dataset* yang digunakan, analisis *dataset*, data *preprocessing*, pengembangan model yang digunakan dalam penelitian, hasil dan evaluasi penelitian.

BAB V HASIL EKSPERIMEN DAN PEMBAHASAN

Bab hasil eksperimen dan pembahasan berisi tentang hasil eksperimen yang dibuat. Setelah hasil tersebut diketahui maka akan dibahas yang berisi tentang deskripsi tujuan eksperimen dan pembahasan tentang eksperimen yang ada.

BAB VI HASIL EKSPERIMEN DAN PEMBAHASAN

Pada bab penutup, terdapat kesimpulan dan saran yang diberikan oleh penulis untuk eksperimen yang akan dilakukan pada masa mendatang.