

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada aktivitas kehidupan pada zaman sekarang sudah dikelilingi oleh berbagai kesuksesan teknologi yang kian semakin pesat tentunya akan meningkatkan kebutuhan manusia sehingga akan mengeluarkan beraneka ragam media terkini. Media terkini merupakan sebuah sarana teknologi telekomunikasi terkait dengan digitalisasi dan dapat diakses dengan secara luas sebagai pengirim dan penerima pesan interaktif yang mampu dijangkau dimana pun dan kapan pun (McQuail, 2011). Media terkini hadir dengan berbagai pembaharuan berdasarkan media lama yang sudah tidak bermakna lagi dengan pembesaran teknologi zaman sekarang ini. Tumbuhnya berbagai media terkini dengan perincian fungsional yang bertentangan dengan media lama seakan sukses menyediakan berbagai kebutuhan manusia yang semakin canggih. Pada zaman dahulu media audio terdiri atas radio, *tape recorder*, dan *Compact Disc* (CD) dan lainnya namun media audio tersebut tidak terkubur dengan begitu saja, tetapi diolah kemudian diadaptasikan menjadi media terkini yang baru. Sehingga sekarang media audio telah menjadi sebuah aplikasi yang bisa di pakai kapan pun dan dimana pun dengan hanya melakukan download aplikasi pada Google Play Store. Dengan beragam pilihan media audio yang dapat kita gunakan sekarang ini seringkali membingungkan pengguna media audio untuk dapat mengetahui manakah media audio yang berkualitas. Mungkin pengguna dapat melihat pada ulasan yang tersedia pada Google Play Store aplikasi yang bersangkutan, seperti Spotify, JOOX Music, Resso Musik, SoundCloud:Play Music & Song, dan masih banyak lagi.

Spotify merupakan salah satu media audio *online* yang peminatnya sangat banyak. Keadaan tersebut dapat didukung dari jumlah masyarakat yang mengunduh aplikasi Spotify pada Google Play Store telah memperoleh 1 Miliar unduhan sampai pada tanggal 20 September 2022. Dengan kondisi ketatnya persaingan antara aplikasi Spotify dengan beberapa media audio online yang lain telah menjadikan perhatian utama dalam melakukan peningkatan dan penyempurnaan sebuah aplikasi.

Sehubungan hadirnya ulasan Google Play dari pemakai aplikasi Spotify bisa meluncurkan pandangan-pandangan dan juga opini berdasarkan penilaian pemakai terhadap aplikasi tersebut. Ulasan merupakan sebuah kalimat atau teks yang berisi komentar atau penilaian terhadap aplikasi tersebut. Namun beberapa pengguna seringkali susah buat mengerti mengenai informasi aktual yang pemakai lain berikan dikarenakan informasi yang diberikan pemakai dengan bentuk teks. Dimana teks tersebut harus mampus memberitahukan pandangan yang diberikan oleh penulis ulasan, oleh karena itu sering mengakibatkan kesenjangan informasi karena seseorang belum tentu bisa benar-benar memahami rasa yang telah disampaikan oleh pemakai lain. Oleh karena itu, banyak penelitian telah dilakukan dengan pembahasan pembelajaran mesin untuk mengembangkan perangkat lunak yang dapat membantu analisis.

Analisis sentimen termasuk ke dalam bagian pengolahan data tekstual yang melaksanakan analisis dengan berdasarkan evaluasi, sentimen, opini, dan perilaku, serta emosi seseorang yang bisa menjadi sebagai sumber untuk di evaluasi. Analisis sentimen bisa dilaksanakan dengan menggunakan berbagai macam algoritma, diantaranya Deep Neural Network, Support Vector Machine, Naïve Bayes, Recurrent Neural Network, dan lainnya masih banyak lagi algoritma pembelajaran mesin. Berdasarkan dari penelitian yang telah dilaksanakan oleh Nurdin et al., (2020) terdapat berbagai macam variasi yang bisa digunakan ketika melaksanakan sebuah analisis, dengan satu contoh diantaranya dengan menerapkan metode *word embedding* didalam analisis tersebut. Word embedding dapat diartikan sebagai sebuah metode penyisipan kata, dimana lebih spesifiknya mengubah sebuah kata menjadi sebuah vektor dengan panjang yang telah ditentukan, sehingga tidak akan dibatasi dengan kosa kata yang lebih banyak. Pada model *word embedding* ini juga membolehkan kita untuk mempertemukan kata-kata yang mirip berdasarkan pengukuran terhadap jarak semantik antara vektor yang disematkan dalam kata tersebut. (Rhman et al., 2021). Menurut penelitian yang dilaksanakan oleh Nurdin et al., (2020), *feature engineering* tradisional pada data tekstual yang umum dipakai biasa disebut dengan *Bag of Word* yang merepresentasikan teks ke dalam kawatan kata yang diwakili oleh frekuensi kemunculan kata pada dokumen, diantaranya *term frequencies* (TF), *term frequency-inverse document frequency* (TF-IDF), serta *n-grams*.

Kelemahan dari *Bag of Words* adalah ketidakmampuan menghasilkan informasi mengenai struktur, urutan, konteks, dan semantik disekitar kata pada dokumen. *Word embedding* menggambarkan setiap kata pada dokumen ke dalam bentuk dense vektor. Dimana vektor mewakili proyeksi kata yang berada pada ruang vektor. Posisi kata dipelajari berasal teks atau kata-kata disekitarnya. Word embedding memungkinkan untuk mengambil makna semantik dan sintaksis dari kata tersebut.

Adapun beberapa penelitian terdahulu yang telah dilakukan dimana membahas mengenai analisis sentimen. Ekspansi fitur pada analisis sentimen *twitter* dengan pendekatan metode Word2Vec yang diimplementasikan dengan diawali dengan pengambilan data pada *twitter* dengan cara crawling data sebanyak 16.571 tweet yang disimpan kedalam database, selanjutnya dilakukan pelabelan data yang terbagi menjadi dua yaitu *positive* (1) dan *negative* (-1). Setelah itu dilaksanakan proses data *preprocessing* dengan tahapan-tahapan filtering, case folding, tokenization, stopword removal, dan stemming yang selanjutnya masuk kedalam word embedding dengan menggunakan metode Word2Vec dan dilakukan pada dua algoritma. Penelitian ini memperoleh hasil akurasi untuk setiap algoritmanya, sebagai berikut K-Nearest Neighbor (KNN) 83% dan Support Vector Machine (SVM) 83.7% (Naufal & Setiawan, 2021). Pada penelitian lain oleh Sreya dan Setiawan menggunakan data yang diambil dari twitter dengan cara crawling data sebanyak 142.544 data dari beberapa media yaitu Republika, CNNIndonesia, SindoNews, Kompas, Detik, dan Liputan6 yang selanjutnya dilakukan pelabelan data yang terbagi menjadi dua yaitu *positive* (1) dan *negative* (-1) setelah itu dilakukan data *preprocessing* dengan melakukan case folding, tokenization, stopword removal, dan stemming. Dilanjutkan dengan metode Global Vector (GloVe) dengan menggunakan dua algoritma. Penelitian ini memperoleh hasil akurasi untuk setiap algoritmanya, sebagai berikut Support Vector Machine (SVM) 83.23% dan Naïve Bayes (NB) 77.86% (Sreya & Setiawan, 2022). Penelitian lain yang dibuat oleh Fransiska et al., (2020) menggunakan data yang berbeda dibandingkan dengan penelitian diatas dengan menggunakan data yang diambil dari ulasan aplikasi by.U pada Google Play Store dengan bahasa pemrograman Python dan memanfaatkan library *google-play-scraper* sebanyak 10.000 ulasan yang selanjutnya dilakukan pelabelan data yang terbagi menjadi dua yaitu *positive* (rating 4 dan 5) dan *negative* (rating 1 dan 2) serta *neutral*

(rating 3) setelah itu dilakukan data *preprocessing* dengan melakukan data cleaning, normalization data, stopword removal, dan negation handling. Kemudian dilanjutkan dengan klasifikasi sentiment menggunakan Support Vector Machine (SVM) dan TF-IDF sebagai ekstraksi fitur dengan 5-fold validation menghasilkan akurasi cukup bagus dengan akurasi rata-rata 84.7%, presisi 84.9%, recall 84.7%, dan f-measure 84.8% serta hasil akurasi tertinggi terdapat pada fold 2 dengan 86.1%.

Berdasarkan penelitian terdahulu, peneliti akan melakukan penerapan dengan perbandingan dari metode *word embedding* *Word2Vec*, *GloVe*, dan *FastText* yang akan diaplikasikan pada algoritma *Support Vector Machine* terhadap ulasan penggunaan aplikasi Spotify pada Google Play Store dan kemudian akan memberikan informasi terbaru yaitu melakukan klasifikasi ulasan tersebut ke dalam kelas *positive*, dan *negative*. Metode *word embedding* yang akan digunakan adalah metode yang telah diterapkan pada penelitian terdahulu dan memberikan hasil yang baik namun berdasarkan beberapa penelitian terdahulu mengenai topik analisis sentimen dengan *word embedding* mempergunakan algoritma *Support Vector Machine* telah menghasilkan nilai akurasi tinggi, maka dari itu pada penelitian ini peneliti akan melakukan penerapan klasifikasi dengan algoritma SVM dimana menggunakan input berupa vektor yang dihasilkan dari ketiga metode *word embedding* tersebut. Ketiga metode *word embedding* yang digunakan ini pada dasarnya akan menghasilkan sebuah vektor, namun fokus pada penelitian ini yaitu nilai yang akan dihasilkan oleh ketiga vektor tersebut. Setelah itu, peneliti akan mengambil kesimpulan untuk membuktikan metode *word embedding* manakah yang memiliki tingkat akurasi, presisi, *recall*, dan *f1-score* tertinggi.

Sehingga dari penjelasan diatas, peneliti akan melaksanakan penelitian dengan judul “Perbandingan Akurasi *Word Embedding* *Word2Vec*, *GloVe*, *FastText* Menggunakan *Support Vector Machine* Untuk Analisis Sentimen Ulasan Aplikasi *Spotify*”

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, peneliti akan merumuskan beberapa permasalahan penelitian perihal analisis sentimen aplikasi menggunakan

algoritma *Support Vector Machine* dengan metode *word embedding Word2Vec*, *GloVe*, dan *FastText*. Dengan rumusan masalah utama, yaitu “Bagaimana perbandingan nilai akurasi yang dihasilkan dari antara *word embedding* dengan metode *Word2Vec*, *GloVe*, dan *FastText* pada *Support Vector Machine* yang bertujuan analisis sentimen ulasan pengguna aplikasi Spotify pada *Google Play Store*?”. Dengan permasalahan utama tersebut akan dideskripsikan lebih rinci sebagai berikut:

1. Bagaimana penerapan metode *word embedding Word2Vec*, *GloVe*, dan *FastText* untuk melakukan analisis sentimen pada ulasan aplikasi Spotify menggunakan *Support Vector Machine*?
2. Bagaimana analisis perbandingan nilai akurasi hasil penerapan metode *word embedding Word2Vec*, *GloVe*, dan *FastText* untuk melakukan analisis sentimen pada ulasan aplikasi Spotify menggunakan *Support Vector Machine* yang nantinya dilihat dari tingkat akurasi, presisi, *recall*, dan *f1-score* pada ketiga model?

### 1.3 Ruang Lingkup Penelitian

Banyaknya algoritma yang digunakan dalam implementasi bahasan analisis sentimen, sehingga peneliti membuat rumusan ruang lingkup penelitian agar ketidakjelasan yang akan muncul nantinya dapat dihindari. Berikut ruang lingkup penelitian:

1. Data yang digunakan berasal dari ulasan mengenai aplikasi Spotify pada *Google Play Store*. Data yang dipakai terdiri dari ulasan dengan jumlah data sebesar 1512 ulasan dalam rentang waktu 09 Oktober 2022 – 17 Oktober 2022 dan 01 April 2023 – 16 April 2023
2. Penelitian berfokus kedalam perbandingan implementasi *word embedding* dengan metode *Word2Vec*, *GloVe*, dan *FastText*.
3. Klasifikasi hanya menggunakan satu algoritma, yaitu *Support Vector Machine* (SVM).
4. Penelitian dilakukan hingga ditingkat pembuktian perbedaan akurasi pada setiap metode *word embedding* yang digunakan

#### 1.4 Tujuan Penelitian

Berdasarkan latar belakang serta rumusan masalah yang telah diuraikan, garis besar tujuan dalam penelitian yaitu mengetahui hasil perbandingan kinerja *word embedding* dengan metode *Word2Vec*, *GloVe*, dan *FastText* pada *Support Vector Machine* untuk analisis sentimen. Dengan begitu dapat dideskripsikan tujuan pada penelitian ini dengan sebagai berikut:

1. Menerapkan metode *word embedding* *Word2Vec*, *GloVe*, dan *FastText* untuk melakukan analisis sentimen pada ulasan aplikasi Spotify menggunakan *Support Vector Machine*.
2. Menganalisis perbandingan nilai akurasi hasil implementasi metode *word embedding* *Word2Vec*, *GloVe*, dan *FastText* untuk melakukan analisis sentimen pada ulasan aplikasi Spotify menggunakan *Support Vector Machine* dengan melakukan perbandingan hasil dari akurasi, presisi, *recall*, dan *f1-score* pada ketiga model.

#### 1.5 Manfaat Penelitian

Melalui penelitian tentang analisis sentimen menggunakan *Support Vector Machine* dengan tiga metode yang berbeda-beda, beroleh manfaat yaitu mengetahui metode *word embedding* manakah dari antara *Word2Vec*, *GloVe*, dan *FastText* yang dapat menghasilkan nilai akurasi tertinggi saat diimplementasikan kedalam analisis sentimen ulasan aplikasi Spotify pada *Google Play Store* menggunakan *Support Vector Machine*.

#### 1.6 Luaran yang Diharapkan

Berdasarkan uraian tujuan dan manfaat penelitian diatas, luaran yang diharapkan oleh peneliti dalam penelitian ini adalah penerapan metode *word embedding* (*Word2Vec*, *GloVe*, dan *FastText*) yang dapat menghasilkan nilai akurasi tertinggi saat diimplementasikan kedalam analisis sentimen ulasan aplikasi Spotify pada *Google Play Store* menggunakan *Support Vector Machine*.

## 1.7 Sistematika Penelitian

Sistematika penelitian dalam penelitian ini berisikan mengenai gambaran dalam setiap bab yang akan dibahas dengan tujuan untuk menjelaskan secara singkat dan jelas isi dari setiap bab sehingga dapat mempermudah pembaca untuk memahami penelitian ini secara keseluruhan. Sistematika penelitian dapat diuraikan sebagai berikut:

### BAB I PENDAHULUAN

Dalam Bab I ini menjelaskan dengan singkat dan jelas mengenai latar belakang, rumusan masalah, ruang lingkup, tujuan, manfaat, luaran yang diharapkan, dan sistematika penelitian yang nantinya dapat digunakan oleh peneliti sebagai acuan dalam penulisan penelitian.

### BAB II TINJAUAN PUSTAKA

Dalam Bab II ini menjelaskan mengenai konsep-konsep yang berisi teori-teori yang mendasar dan memaparkan penelitian-penelitian terdahulu yang terkait dengan topik penelitian ini. Pada bab ini berisikan definisi, model, algoritma, proses yang selaras dengan masalah penelitian.

### BAB III METODOLOGI PENELITIAN

Dalam Bab III ini menjelaskan mengenai alur penelitian mulai dari sumber dan metode-metode seperti metode pengumpulan data, metode pengolahan data, dan metode analisis data yang akan dijalani untuk mencapai tujuan dalam penelitian ini,

### BAB IV HASIL DAN PEMBAHASAN

Dalam Bab IV ini akan membahas mengenai pengumpulan data, preprocessing data, pembobotan data, ekstraksi fitur, klasifikasi data, permodelan data, dan

pengujian dengan confusion matrix serta membahas mengenai analisis perbandingan dari hasil pengujian data yang telah diolah dalam penelitian.

## BAB V PENUTUP

Dalam Bab V ini membahas mengenai kesimpulan yang akan didapatkan dari hasil penelitian serta saran yang peneliti dapat berikan sehingga bisa digunakan pembaca sebagai acuan untuk penelitian selanjutnya.

## DAFTAR PUSTAKA

Daftar Pustaka memuat informasi mengenai daftar bacaan yang menjadi sumber referensi dalam penelitian yang dilaksanakan.