

BAB I

Pendahuluan

1.1 Latar Belakang

Beberapa bulan belakangan ini, begitu banyaknya pengguna Intel generasi ke-12 yang menjadi faktor utama semakin berkembangnya zaman dan membuat setiap aplikasi berat yang dipakai oleh setiap pengguna di seluruh dunia mengganti perangkat baik dari komputer maupun laptop dengan generasi yang lebih baik dan dapat meningkatkan kinerja para pekerja, terutama yang bekerja sebagai Data Scientist, Content Creator, Vtuber, Youtuber, dll. Salah satu prosesor terkuat yang dapat mempermudah pekerjaan dan sering digunakan pada zaman ini adalah seri Intel generasi ke-12 yang memiliki performa tercepat bila dibandingkan dengan pendahulunya. Sejak pertama kali keluarnya pembahasan mengenai Intel generasi ke-12 serta ramai diperbincangkan berupa komentar dari media sosial.

Dari komentar-komentar di media sosial tersebut, penulis dapat melihat tingkat emosi dari orang-orang yang sudah berpendapat terhadap Intel generasi ke-12 yang sedang diperbincangkan serta laku di pasaran karena beberapa faktor yang menjadikan hal ini dapat diteliti lebih lanjut dengan analisis sentimen. Dengan analisis sentimen ini, penulis dapat menentukan serta membedakan opini yang positif, ataupun yang negatif.

Terdapat beberapa riset Analisis Sentimen, antara lain adalah pada penelitian [1] yang membahas tentang analisis sentimen dengan klasifikasi Naïve Bayes dan Support Vector Machine pada sentimen yang diambil dari Twitter dengan Accuracy Naïve Bayes sebesar 94% sedangkan Support Vector Machine sebesar 75.5%, penelitian [2] yang membahas tentang analisis sentimen pengguna youtube terhadap program vaksin dengan klasifikasi Support Vector Machine dan hasil yang ditunjukkan memiliki Accuracy sebesar 77%, penelitian [3] menggunakan proses analisis sentimen dengan membandingkan klasifikasi Lexicon Based dan Support Vector Machine pada sentimen yang diambil dari berbagai jenis sosial media dengan Accuracy Lexicon Based sebesar 88% dan Support Vector Machine sebesar 86%., pada penelitian [4] menggunakan proses analisis sentimen dengan klasifikasi Support Vector Machine dan hasil yang ditunjukkan memiliki Accuracy sebesar 90%. Pada paper tersebut, model Machine Learning yang paling sering digunakan untuk analisis sentimen adalah Support Vector Machine. Untuk beberapa Feature yang digunakan, diantaranya adalah Word Embeddings menggunakan GloVe, Word2Vec, dan juga FastText. Adapun kelemahan yang ada dalam riset ini adalah NLP yang hanya dibatasi dengan menggunakan Support Vector Machine Classifier dikarenakan rata-rata accuracy yang dihasilkan lebih tinggi, dari Classifier yang lain, sehingga penggunaan model Machine Learning yang lain jarang ditemukan.

Berdasarkan paper-paper yang telah dijelaskan sebelumnya dan dengan adanya pembicaraan mengenai Intel Generasi ke-12 dari Youtube, serta banyaknya pendapat dari berbagai pengguna media sosial sebagai referensi untuk membuat analisis sentimen. Adapun penggunaan Word Embeddings yang diklasifikasikan dengan Support Vector Machine merupakan kombinasi terbaik secara keseluruhan, dibandingkan kombinasi dengan Random Forest, K-Nearest Neighbors, ataupun Gaussian Naïve Bayes[5], maka dari dalam TA ini diusulkan “Word Embeddings Terbaik dalam Klasifikasi Sentimen Prosesor Intel Generasi ke-12 Dengan Menggunakan Support Vector Machine”.

1.2 Rumusan Masalah

Berdasarkan fenomena yang telah dijelaskan pada latar belakang, dapat ditemukan rumusan masalah sebagai berikut :

1. Bagaimana hasil perbandingan dari penggunaan metode Word Embeddings GloVe, Word2Vec, FastText dengan menggunakan Support Vector Machine?
2. Bagaimana ketiga metode Word Embeddings yang menggunakan Support Vector Machine (SVM) mampu memprediksi sentimen dengan tepat?

1.3 Tujuan

Tujuan yang akan dicapai dalam Tugas Akhir ini berdasarkan fenomena yang telah dijelaskan pada latar belakang adalah sebagai berikut :

1. Mengetahui metode Word Embeddings yang paling baik diantara GloVe, Word2Vec, FastText dengan menggunakan Support Vector Machine.
2. Mengetahui metode yang paling akurat dalam memprediksi sentimen dan mengidentifikasi faktor-faktor yang berkontribusi terhadap akurasi prediksi.

1.4 Batasan Masalah

Berikut adalah batasan masalah yang telah ditentukan agar topik yang akan diteliti tidak keluar dari konteksnya :

1. Data yang dianalisis hanya menggunakan Bahasa Inggris
2. Opini yang diambil hanya terkait Intel Generasi ke-12 dan hanya didapatkan dari Youtube
3. Algoritma Machine Learning yang digunakan untuk mengklasifikasi sentimen adalah Support Vector Machine.