

Nama : Diah Mutia choirunnisa

Soal :

Terdapat sekumpulan data mengenai tulisan dalam bentuk *tweet* mengenai sebuah kebijakan. Sekumpulan data tersebut ingin dikelompokkan berdasarkan sentimen dari *tweet* tersebut yaitu sentimen positif dan negatif. Jelaskan algoritma A.I. yang dapat digunakan untuk mengelompokkan *tweet* tersebut beserta alasannya.

Jawaban :

1. Untuk mengelompokkan tweet berdasarkan sentimen (positif atau negatif), Algoritma ini menggunakan berbagai teknik pembelajaran mesin untuk mengklasifikasikan teks ke dalam kategori sentimen yang relevan.

Berikut adalah langkah - langkahnya :

- Pengumpulan Data :
 1. Langkah pertama yaitu mengumpulkan data tweet yang berisi kebijakan yang ingin kita analisis sentimennya.
 2. kedua Data harus dibersihkan dari tanda baca, karakter khusus, dan URL, beserta dikonversi ke huruf kecil agar mempermudah analisis.
- Tokenisasi :
 - Tokenisasi memecah tweet menjadi kata-kata atau token. Ini diperlukan untuk menganalisis teks karena model mesin memproses data dalam bentuk angka dan tokenisasi membantu mengubah teks menjadi representasi numerik.
- Ekstraksi Fitur :
 - Mengubah teks menjadi representasi numerik dalam menganalisis pengelompokan tweet. tf-idf dan word embedding seperti Word2Vec atau GloVe adalah teknik yang kuat untuk menghasilkan vektor fitur yang mewakili setiap tweet. Ini memungkinkan model untuk memahami hubungan antara kata-kata dalam teks.

- Machine Learning :
Setelah fitur ekstraksi, kita akan menggunakan data yang telah diproses untuk melatih model mesin. Beberapa jenis model yang dapat digunakan untuk analisis sentimen meliputi
 - Naive Bayes : Cocok untuk pemula dan cepat dalam pemrosesan teks
 - Support Vector Machine (SVM) : Efektif dalam mengatasi masalah klasifikasi teks
 - Deep Learning (RNN,CNN, atau Transformer) : model yang lebih kompleks dapat menangani teks dengan lebih baik, terutama dalam analisis sentimen yang lebih kompleks
 - Random Forest Classifier : yaitu algoritma ensemble yang bagus untuk analisis sentimen. Ini dapat mengatasi overfitting dan mengatasi masalah fitur yang tidak relevan.
- Evaluasi Model :
Evaluasi model dengan menggunakan data validasi atau uji adalah langkah kunci untuk memastikan bahwa model yang dibangun berkinerja dengan baik. Ini memungkinkan Anda untuk mengukur sejauh mana model dapat memprediksi sentimen dengan akurat.
- Penyaringan Sentimen :
Setelah model terlatih, Kita dapat menggunakan model tersebut untuk mengklasifikasikan tweet baru berdasarkan sentimennya.