

Nama : Muhammad Aimar Rizki Utama
Kelas : Golang - A
Tugas : 25- Algoritma Ai

Soal

1. 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, kita dapat menggunakan algoritma analisis sentimen berbasis machine learning, yang melibatkan beberapa tahap pemrosesan, Berikut untuk langkah langkahnya :
 - Pengumpulan Data : Pertama kita harus mengumpulkan data twwet yang berhubungan dengan kebijakan yang ingin kita analisis sentimennya, data ini harus berisi tweet beserta label sentimen yang sesuai (positif,negatif,netral)
 - Preprocessing Data : Data sering kali memiliki berbagai macam karakteristik dan kekacauan seperti tanda baca, huruf besar/kecil campur, dan kata-kata yang tidak relevan. Proses preprocessing ini akan membersihkan data dan mengubahnya menjadi format yang mudah dianalisis. Tahap preprocessing melibatkan :
 - Tokenization : memisahkan teks menjadi kata kata individu
 - Pembersihan teks : menghapus tanda baca, tautan, dan karakter khusus lainnya
 - Stemming atay Lemmatization : mengubah kata kata menjadi bentuk dasarnya
 - Stop word removal : menghapus kata-kata umum yang tidak memiliki makna signifikan
 - Ekstraksi Fitur : selanjutnya kita harus mengubah teks tweet menjadi representasi numeruk yang dapat digunakan oleh model pembelajaran mesin salah satu teknik yang umum digunakan adalah "tf-idf" (Term Frequence-Inverse Document Frequency) atau word embedding seperti Word2Vec atau GloVe. Ini akan menghasilkan vektor fitur untuk setiap tweet
 - Machine Learning : Setelag fitur ekstraksi, kita akan menggunakan data yang telah diproses untuk melatih model mesih pembelajaran. 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
- Evaluasi Model : Setelah melatih model, kita harus mengevaluasi kinerjanya menggunakan data validasi atau pengujian yang tidak digunakan selama pelatihan. Metrik evaluasi yang umum digunakan meliputi akurasi, presisi, recall, F1-score, dan matriks kebingungan (confusion matrix).
- Penyaringan Sentimen : Setelah model terlatih, Kita dapat menggunakan model tersebut untuk mengklasifikasikan tweet baru berdasarkan sentimennya.

Alasan menggunakan algoritma yang saya jelaskan di atas :

1. Algoritma ini cocok untuk analisis sentimen teks, karena dapat memahami konteks dan nuansa dalam teks.
2. Memungkinkan analisis dalam skala besar, sehingga Anda dapat mengelompokkan ribuan atau jutaan tweet dalam waktu yang relatif cepat.
3. Dengan penggunaan teknik-teknik preprocessing yang sesuai, Anda dapat mengatasi masalah yang sering muncul dalam teks sosial media, seperti singkatan, slang, dan bahasa yang tidak baku.
4. Model mesin pembelajaran dapat ditingkatkan seiring berjalannya waktu dengan penambahan data baru dan penyesuaian parameter.