

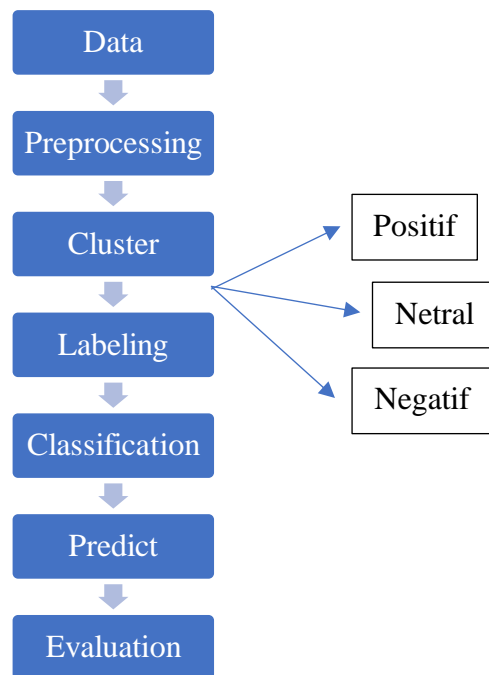


UAS SEMESTER GANJIL TAHUN AKADEMIK 2022/2023

Mata Kuliah : Pembelajaran Mesin

Dosen : Adevian Fairuz Pratama,S.ST.,M.Eng

Kelas : TI 3E, TI 3F



Deteksi Jenis Komentar Pengguna Twitter

Deteksi jenis komentar pada twitter merupakan salah satu permasalahan yang dihadapi pada **Natural Language Processing (NLP)**. Alasannya diantaranya adalah kurangnya dataset berlabel untuk mengklasifikasikan jenis komentar berdasarkan data twitter. Pada Ujian Akhir Semester (UAS) kali ini, tugas utama Anda adalah membuat model yang mumpuni untuk kebutuhan klasifikasi jenis komentar pengguna twitter berdasarkan teks (komentar positif, netral, dan negatif).



Informasi Data

Dataset yang akan digunakan adalah **tweet_emotion.csv**. Berikut merupakan informasi tentang dataset yang dapat membantu Anda.

- Total data: 40000 data
- Terdapat 3 kolom = 'tweet_id', 'sentiment', 'content'

Penilaian UAS

UAS akan dinilai berdasarkan 6 proses yang akan Anda lakukan, yaitu preprocessing data, clustering, labeling, classification (pembuatan model machine learning), prediction, dan evaluasi.

1. Preprocessing Data: Data tweeter yang ada dapatkan merupakan sebuah data mentah, maka beberapa hal dapat Anda lakukan (namun tidak terbatas pada) yaitu,
 - Case Folding
 - Tokenizing
 - Filtering
 - Stemming

CATATAN: PADA DATA TWITTER TERDAPAT MENTION* (@something) YANG ANDA HARUS TANGANI SEBELUM MASUK KE TAHAP SELANJUTNYA

2. Clustering: Pengelompokan data ke dalam beberapa kategori atau cluster, yaitu komentar positif, netral, dan negatif.
3. Labeling: Label merupakan hasil dari pengelompokan example melalui clustering. Sebagai contoh, machine learning yang berfungsi menyaring email spam, melabeli setiap example dengan 'spam' atau 'not spam'.
4. Classification: Anda dibebaskan dalam memilih algoritma klasifikasi. Anda dapat menggunakan algoritma yang telah diajarkan didalam kelas atau yang lain, namun dengan catatan. Berdasarkan asas akuntabilitas pada pengembangan model machine



KEMENTERIAN RISET TEKNOLOGI DAN PENDIDIKAN TINGGI
POLITEKNIK NEGERI MALANG
JURUSAN TEKNOLOGI INFORMASI
Jl. Soekarno Hatta No.9 Malang 65141
Telp (0341) 404424 – 404425 Fax (0341) 404420
Laman://www.polinema.ac.id Email:cs@polinema.ac.id



learning, Anda harus dapat menjelaskan bagaimana model Anda dapat menghasilkan nilai tertentu.

5. Predict: Uji coba apakah machine learning yang Anda buat bekerja dengan baik. Caranya dengan melihat hasil atau prediksi yang dihasilkan. Apakah sesuai dengan *input data*.
6. Evaluasi: Pada proses evaluasi, minimal Anda harus menggunakan metric akurasi. Akan tetapi Anda juga dapat menambahkan metric lain seperti Recall, Precision, F1-Score, detail Confussion Metric, ataupun Area Under Curve (AUC).