**LAPORAN AKHIR SEMESTER MACHINE LEARNING**

**DETEKSI EMOSI PENGGUNA TWEETER**

Diajukan Untuk Memenuhi Tugas

Mata Kuliah Machine Learning

Yang diampu oleh:

**Ibu Adevian Fairuz Pratama, S.S.T, M.Eng.**

Semester Genap Tahun Akademik 2022/2023

****

**Disusun Oleh:**

**Iqri Mannisa’ Buchori (2041720066 / 12)**

**Wazir Qorni Abud (2041720124 / 21)**

**PROGRAM STUDI D-IV TEKNIK INFORMATIKA**

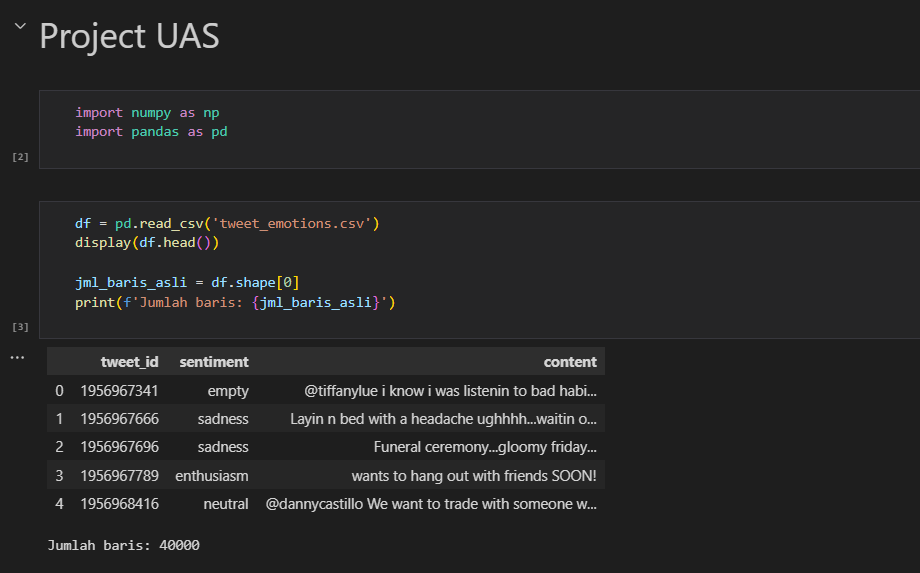
**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

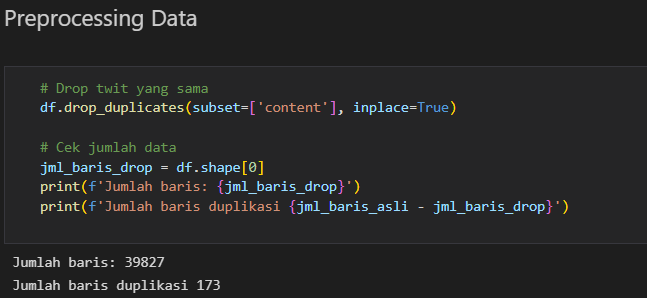
**2022**

1. **Prepocessing Data**

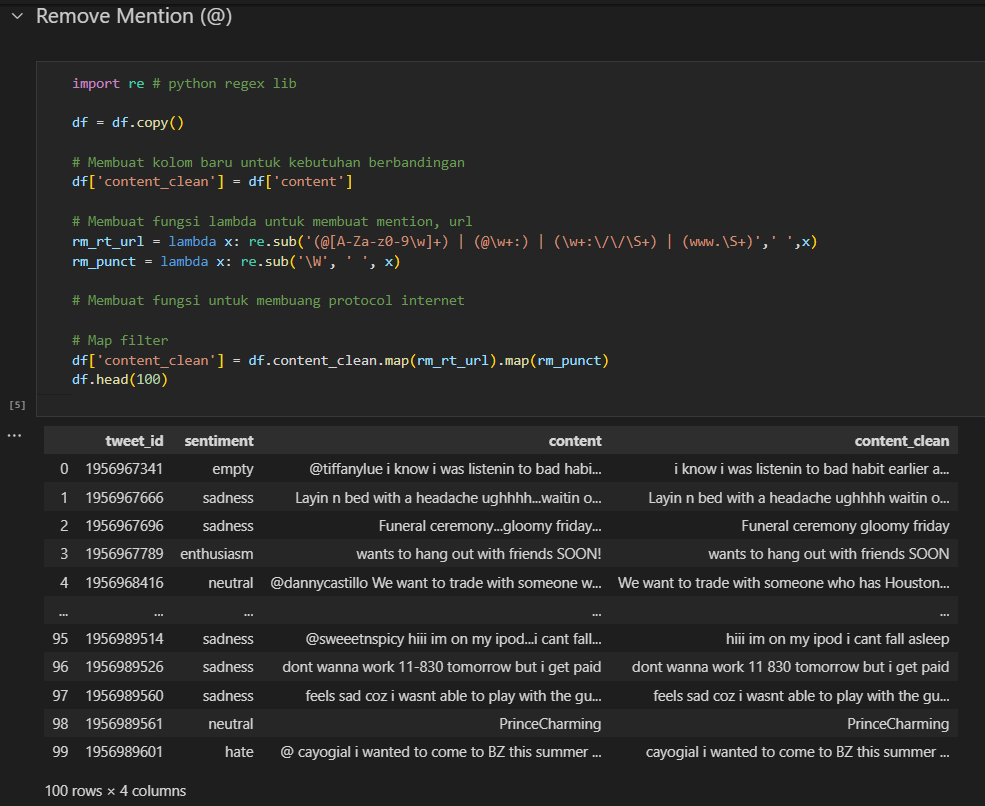
Data yang akan diolah untuk project UAS kali ini adalah data *tweet\_emotions.csv.* Tahap pertama yang dapat dilakukan adalah load dataset kedalam Dataframe menggunakan Pandas.



Setelah data terbaca oleh fungsi python, didapati jumlah baris yang terdapat pada data *tweet\_emotions.csv* berjumlah40.000. Kemudian masuk pada tahap Prepocessing Data dimana akan dilakukan inisialisasi terlebih dahulu apakah didapati data yang terduplikasi dari 40.000 data yang ada. Data yang terduplikasi akan dihapus dengan menggunakan method drop\_duplicates.

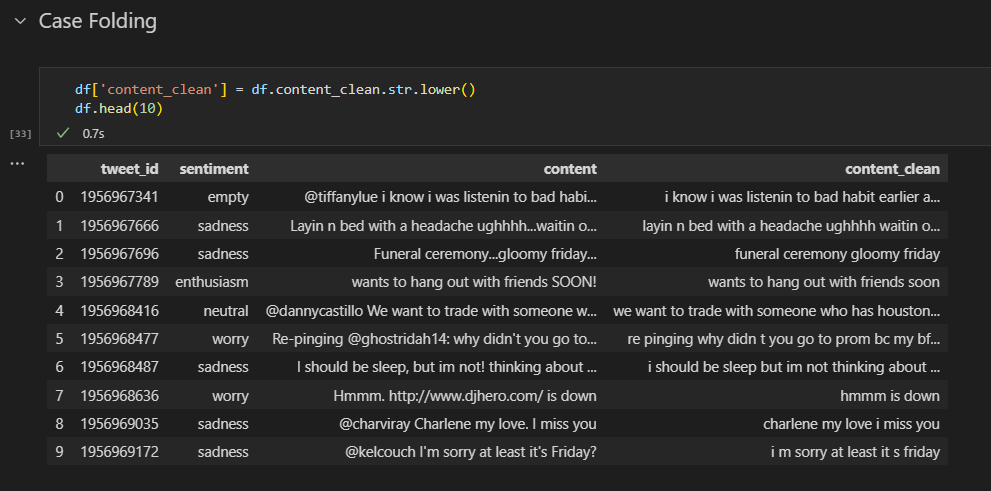


Setelah menghapus data yang terduplikasi masuk pada tahap operasi dasar yang digunakan pada tahap pra pengolahan data adalah Case Folding, Tokenizing, Filtering, dan Stemming. Namun sebelum itu hal yang perlu diperhatikan lagi adalah menghapus mention @.



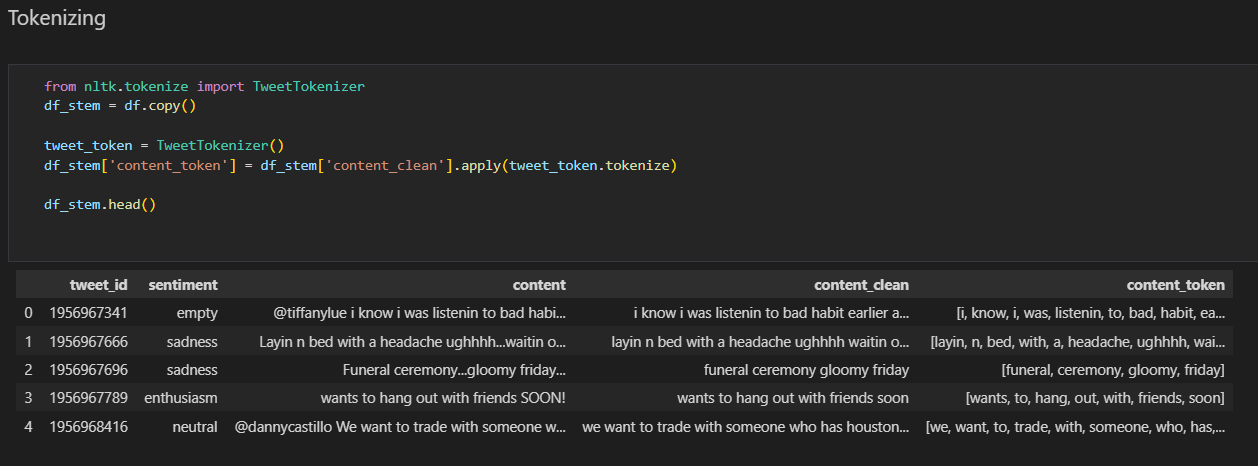
1. Case Folding

Case Folding digunakan untuk mengubah semua bentuk huruf dalam sebuah teks atau mengubah isi dokumen menjadi huruf kecil semua. Sementara itu, karakter lain yang bukan termasuk huruf dan angka, seperti tanda baca dan spasi dianggap sebagai delimiter. Delimiter ini bisa juga dihapus atau diabaikan dengan menggunakan perintah yang ada di Python.



1. Tokenizing

Ditahap ini akan dilakukan proses number removal, whitecase removal, puntuation removal dan word\_tokenize() untuk memecah string kedalam tokens. Pandas Dataframe atau Series mampu menjalankan function external untuk di terapkan pada kolom atau baris dengan menggunakan fungsi .*apply*().



1. Filtering

Pada tahap ini kita akan menggunakan stopword bahasa English yang didapatkan dari library NLTK untuk filtering terhadap Dataframe. Tahapan filtering yang digunakan untuk mengambil kata-kata yang penting dari hasil token tadi. Kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan stopword.

1. Stemming

Stemming adalah proses mengurangi infleksi kata-kata ke bentuk akarnya, seperti memetakan sekelompok kata ke batang yang sama, bahkan jika batang itu sendiri bukan kata yang valid dalam Bahasa.

