# UJIAN AKHIR SEMESTER DATA MINING II

# KLASIFIKASI DATA TWEET TENTANG MINYAK GORENG MENGGUNAKAN NAÏVE BAYES CLASSIFIER

****

Penulis :

Muhammad Zulfikar Reyhan (192103008)

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK DAN TEKNOLOGI INFORMASI**

**UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA**

**2022**

# KATA PENGANTAR

Dengan hormat,

Puji dan syukur penulis panjatkan kepada Allah SWT atas nikmat sehat yang diberikannya penulis dapat menyelesaikan Ujian Akhir Semester pada mata kuliah Data Mining II ini dengan lancar .

Makalah ini dibuat dalam rangka memenuhi salah satu penilaian mata kuliah Data Mining II dan sebagai dosen pengampu yaitu oleh Bapak Muhammad Habibi, S.Kom., M.Cs. di Universitas Jenderal Achmad Yani Yogyakarta pada tahun akademik 2022.

Ujian Akhri Semester ini berisi laporan praktik klasifikikasi data twitter tentang Minyak goreng yang data nya sudah disiapkan oleh dosen, program nya dikerjakan menggunakan Bahasa pemograman Python dan metode nya Naïve Bayes Classifier.

Ucapan terima kasih kepada dosen pengampu yang telah meberikan masukan, ilmu dan juga praktik yang membuat saya mampu mengerjakan ujian akhir semester ini. Penulis menyadari belum sempurna dalam membuat laporan ini, oleh karena itu penulis mengharapkan kritik dan saran dari Dosen pangempu dan teman – teman yang membaca.

Kurang lebih penulis mohon maaf.

Hormat.

Yogyakarta, 8 Juni 2022

Muhammad Zulfikar Reyhan

# DAFTAR ISI

[COVER 1](#_Toc104033118)

[KATA PENGANTAR 2](#_Toc104033119)

[DAFTAR ISI 3](#_Toc104033120)

[DAFTAR GAMBAR 4](#_Toc104033121)

[BAB 1 PENDAHULUAN 5](#_Toc104033122)

[A. Latar belakang 5](#_Toc104033123)

[B. Tujuan 5](#_Toc104033124)

[C. Manfaat 5](#_Toc104033125)

[BAB 2 PEMBAHASAN 7](#_Toc104033126)

[A. Analisis sentiment 7](#_Toc104033127)

[B. Text mining 7](#_Toc104033128)

[C. Text Prepocessing 7](#_Toc104033129)

[D. Klasifikasi 7](#_Toc104033130)

[E. Studi kasus : Perang Rusia ke Ukraina 8](#_Toc104033131)

[BAB 3 KESIMPULAN DAN SARAN 15](#_Toc104033132)

[A. KESIMPULAN 15](#_Toc104033133)

[B. SARAN 15](#_Toc104033134)

[DAFTAR PUSTAKA 16](#_Toc104033135)

# DAFTAR GAMBAR

[Gambar 1. Spreadsheet TAGS 8](#_Toc103614625)

[Gambar 2. Hasil Mining Tweet 9](#_Toc103614626)

[Gambar 3. Program text preprocessing hasil tweet 10](#_Toc103614627)

[Gambar 4. Hasil preprocessing text 10](#_Toc103614628)

[Gambar 5. Pelabelan tweet 11](#_Toc103614629)

[Gambar 6. Data latih/Training 12](#_Toc103614630)

[Gambar 7. Hasil Evaluasi 13](#_Toc103614631)

[Gambar 8. Tweet test 13](#_Toc103614632)

[Gambar 9. Hasil klasifikasi 14](#_Toc103614633)

# BAB 1 PENDAHULUAN

## Latar belakang

Pada masa modern seperti saat ini media social bukan hanya menjadi tempat berkeluh kesah dan mengeskpos masalah pribadi dan masalah yang tidak serius. Namun, sekarang media social bisa menjadi media berita yang digunakan bukan hanya untuk warga biasa bahkan pemimpin dunia.

Belum lama ini dunia di kagetkan dengan penyerangan tantara rusia ke ukraina, walaupun masalah tersebut adalah masalah internal kedua negara, namun semua pemerintah dunia mengkhawatirkan negara mereka masing masing. Termasuk di negara kita Indonesia.

Di dalam negri permasalahan ini bukan hanya menjadi perhatian pemerintah dalam bersikap dan bersiap. Namun masyarakat media social tweeter Indonesia juga sempat tranding dengan pembahasan perang antara dua negara ini.

Melalui makalah ini penulis akan mengklasifikasikan tweet yang Pro dan Kontra tentang masalah peperangan ini. Penulis akan menggunakan naive bayes classifier dalam mengklasifikasi tweet yang di temukan dari spredsheet Tag. Diharapkan makalah ini bisa menjadi pembelajaan untuk penulis dan orang lain yang membacanya.

## Tujuan

Tujuan penulisan makalah ini sebagai berikut :

1. Mengetahui metode Naive Bayes Classifier
2. Mengklasifikasikan Tweet antara Pro dan Kontra
3. Mengetahui Langkah Langkah metode Naive Bayes Classifier
4. Sebagai Tugas Akhir matakuliah Data Mining II

## Manfaat

Manfaat penulisan makalah ini sebagai berikut :

1. Melihat bagaimana Masyarakat indonesia menyikapi Perang tersebut
2. Membedakan tweet yang Pro dan Kontra terhadap perang tersebut menggunakan metode Naive Bayes Classifier.

# BAB 2 PEMBAHASAN

## Analisis sentiment

Analisis Sentimen adalah salah satu cabang sebuah penelitian text mining, yang berkaitan dengan bidang yang lebih luas seperti pengolahan data kegiatan tertentu (Darwis, Siskawati and Abidin, 2021)

Menurut (Samir *et al.*, 2021) Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral.

## Text mining

Menurut (Samir *et al.*, 2021) Teks mining bertujuan menghasilkan informasi dari satu set dokumen. Text Mining mampu menghasilkan informasi melalui pemrosesan, pengelompokan, dan analisis data-data tidak terstruktur dalam jumlah besar.

## Text Prepocessing

Teks preprocessing adalah pembersihan, dan penyederhanaan teks agar dapat diproses lebih lanjut (Purwiantono and Aditya, 2020). Berikut ini merupakan tahapan dari text preprocessing (Rozi, Hamdana and Iqbal Alfahmi, 2018):

1. Case folding
2. Cleaning
3. Tokenizing
4. Stop word.

## Klasifikasi

Data dari Twitter tersebut akan dapat diproses menggunakan text mining, kemudian dilanjutkan dengan mengklasifikasikan Tweet ke dalam dua kelas, Pro dan kontra. Klasifikasi dapat memberikan kemudahan bagi semua pengguna untuk melihat pengguna tweeter yang Pro dan Kontra dengan serangan tersebut.(Darwis, Siskawati and Abidin, 2021)

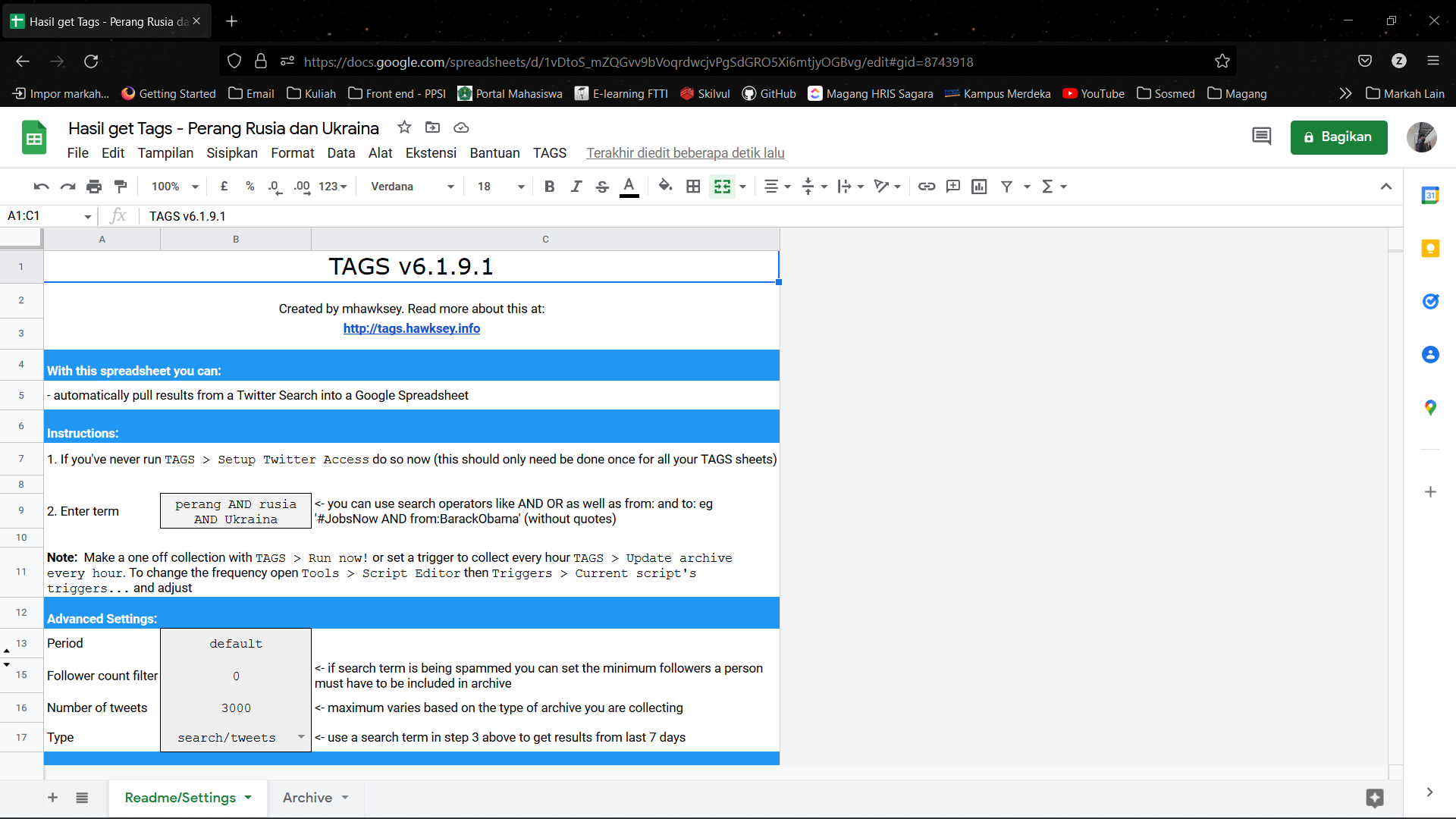
## Studi kasus : Perang Rusia ke Ukraina

Pada studi kasus ini penulis mengakat berita tentang perang yang terjadi di Ukraina akibat dari serangan Rusia yang tidak mengizinkan Ukraina untuk bergabung dengan NATO. Pada kasus ini sentiment yang diambil yaitu Pro dan Kontra.

Proses yang akan dilakukan untuk mengerjakan makalah ini yaitu sebagai berikut:

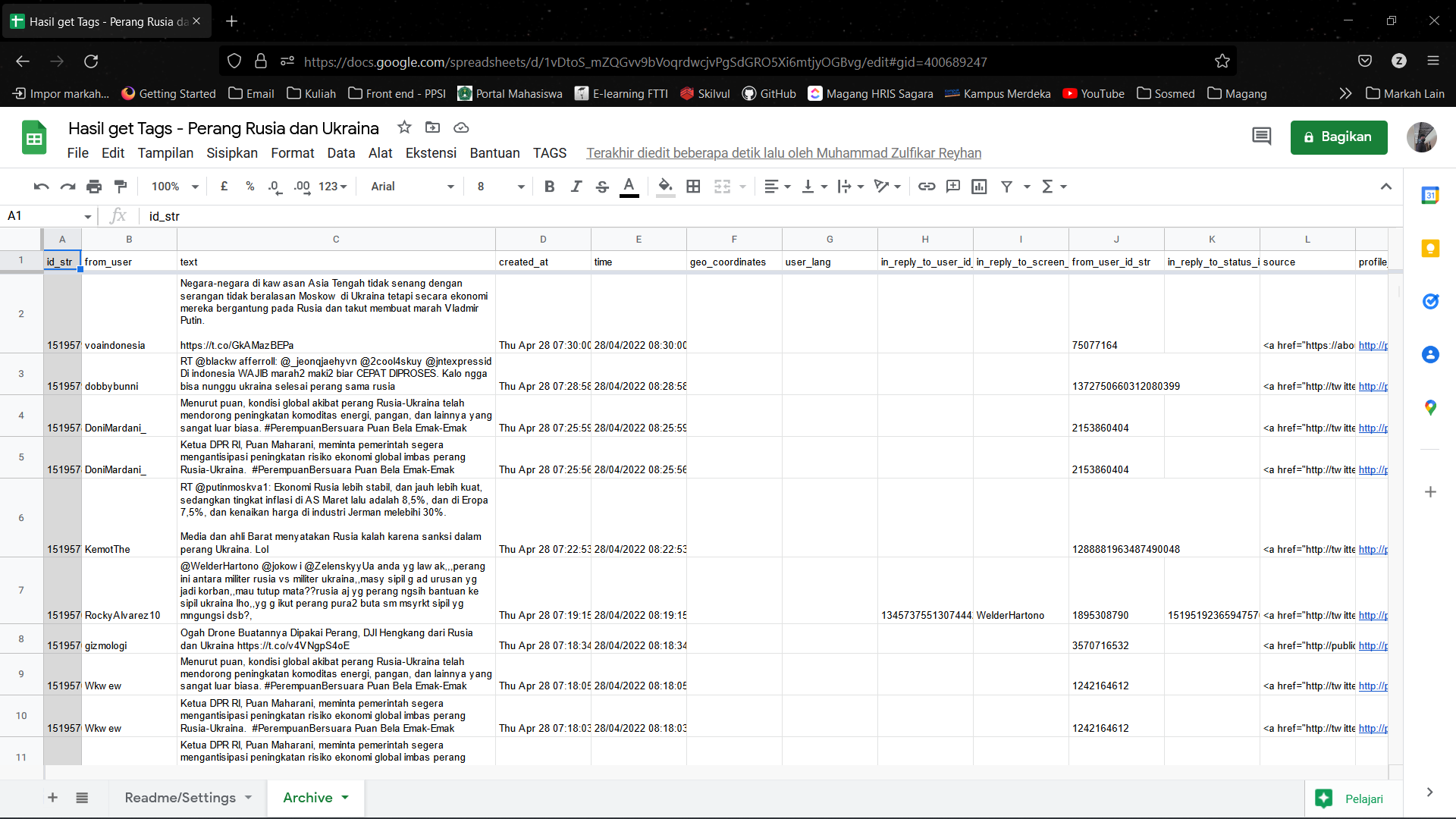
1. Melakukan Text Mining

Dalam melakukan text mining tweet, penulis menggunakan website yang diberikan oleh dosen pengampu yaitu <https://tags.hawksey.info/>. Penulis membuat Salinan ke google spreadsheet.



Gambar 1. Spreadsheet TAGS

Dalam spreadsheet penulis memasukan kata kunci perang, rusia, dan ukraina. Penulis juga menggunakan kata AND agar semua kata kunci yang di cari terpenuhi semua. Supaya tweet yang dicari menggunakan Bahasa Indonesia dan tidak terputus artinya.

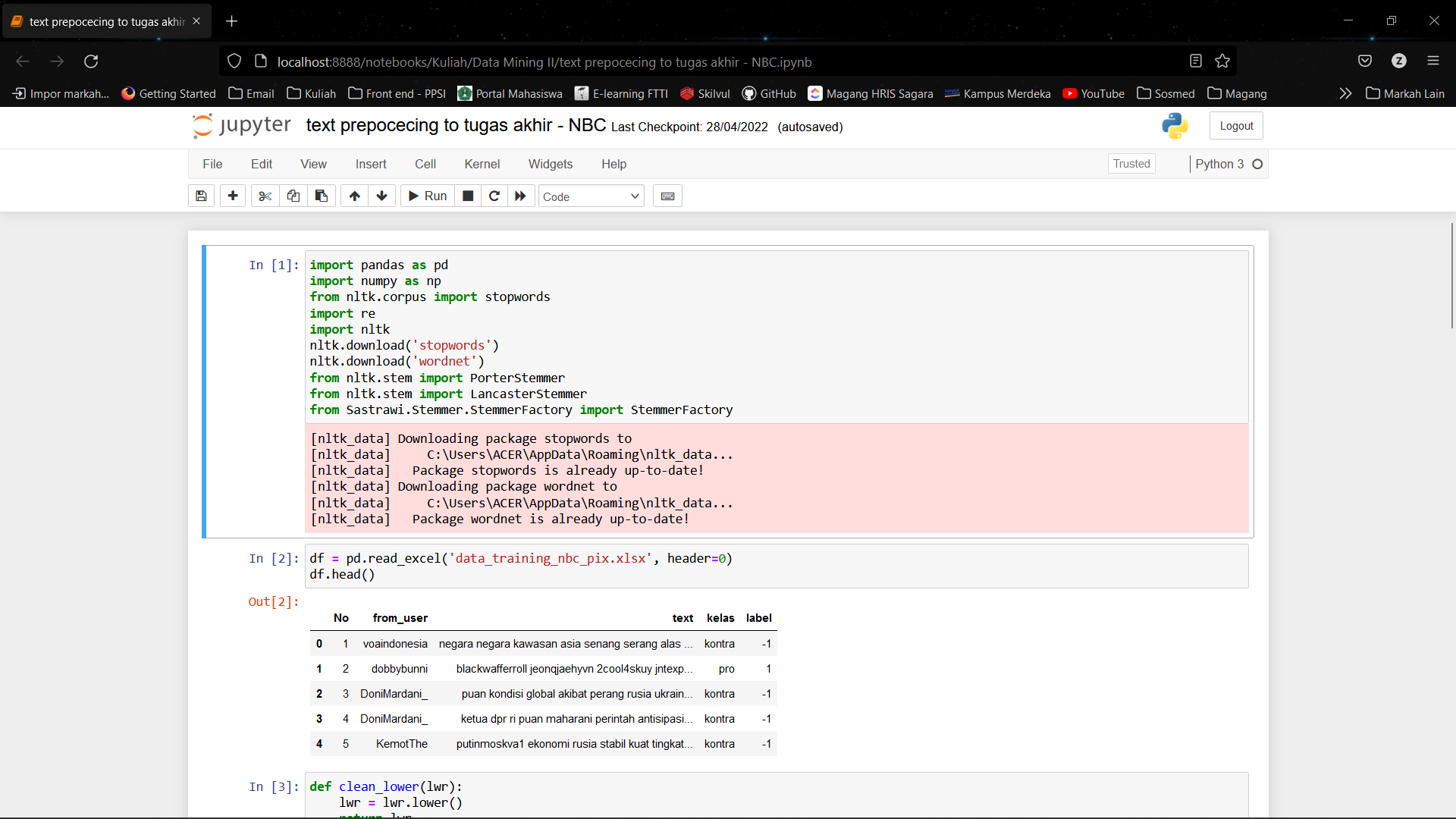


Gambar 2. Hasil Mining Tweet

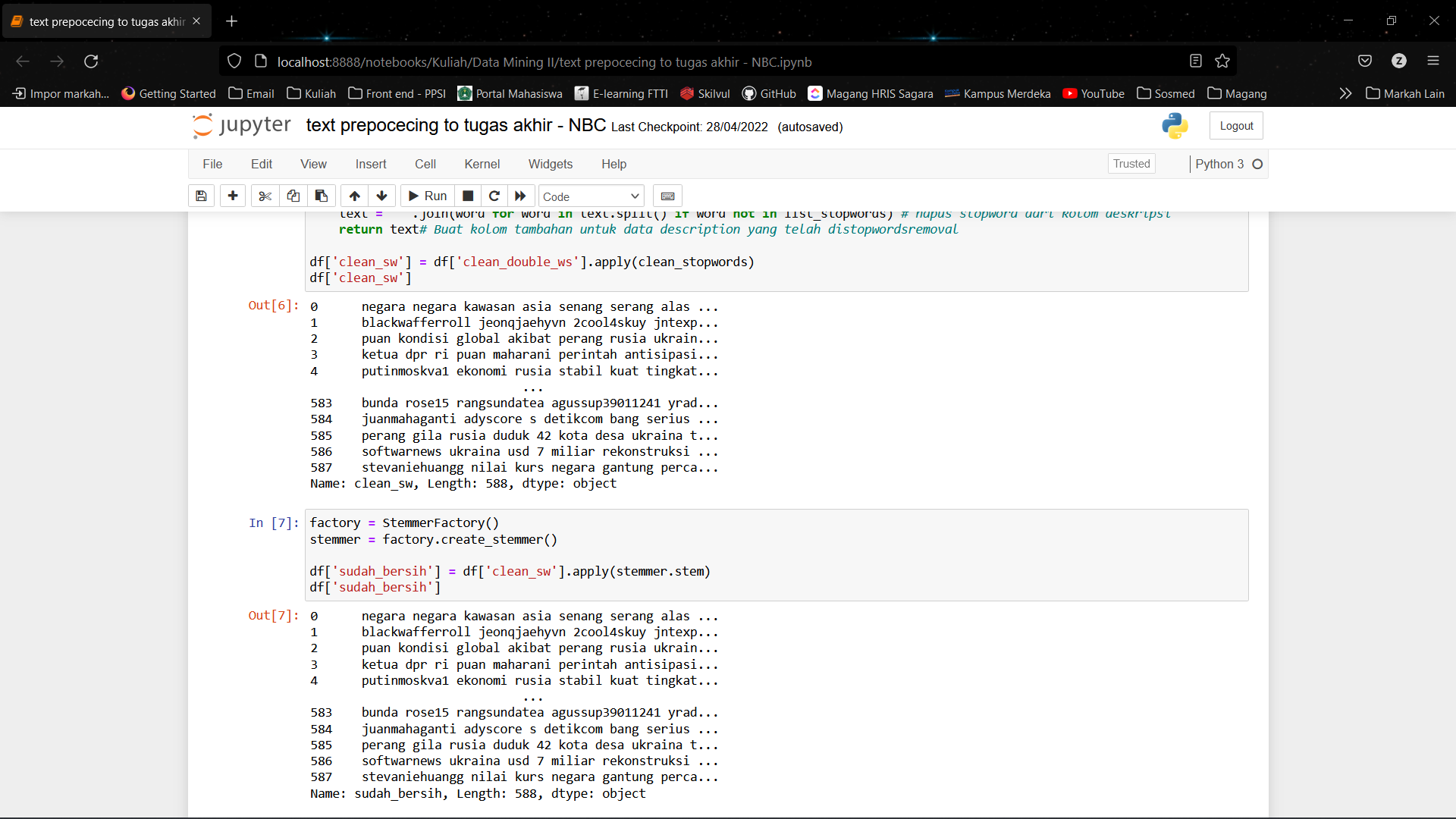
Data yang penulis cari berjumlah 1000 data, namun dari data tersebut beberapa kolom penulis hapus karena tidak sesuai dengan kebutuhan penulis, seperti id\_str, created\_at, time dan lain lain. Penulis tidak hapus hanya from\_user dan text .

1. Preprocessing data

Pada tahap ini penulis memasukan data dalam bentuk excel kedalam program berbahasa pemograman python. Sebelum tugas akhir ini penulis sudah mengerjakan tugas tugas lain yang sesuai dengan kebutuhan pada tugas akhir ini.



Gambar 3. Program text preprocessing hasil tweet



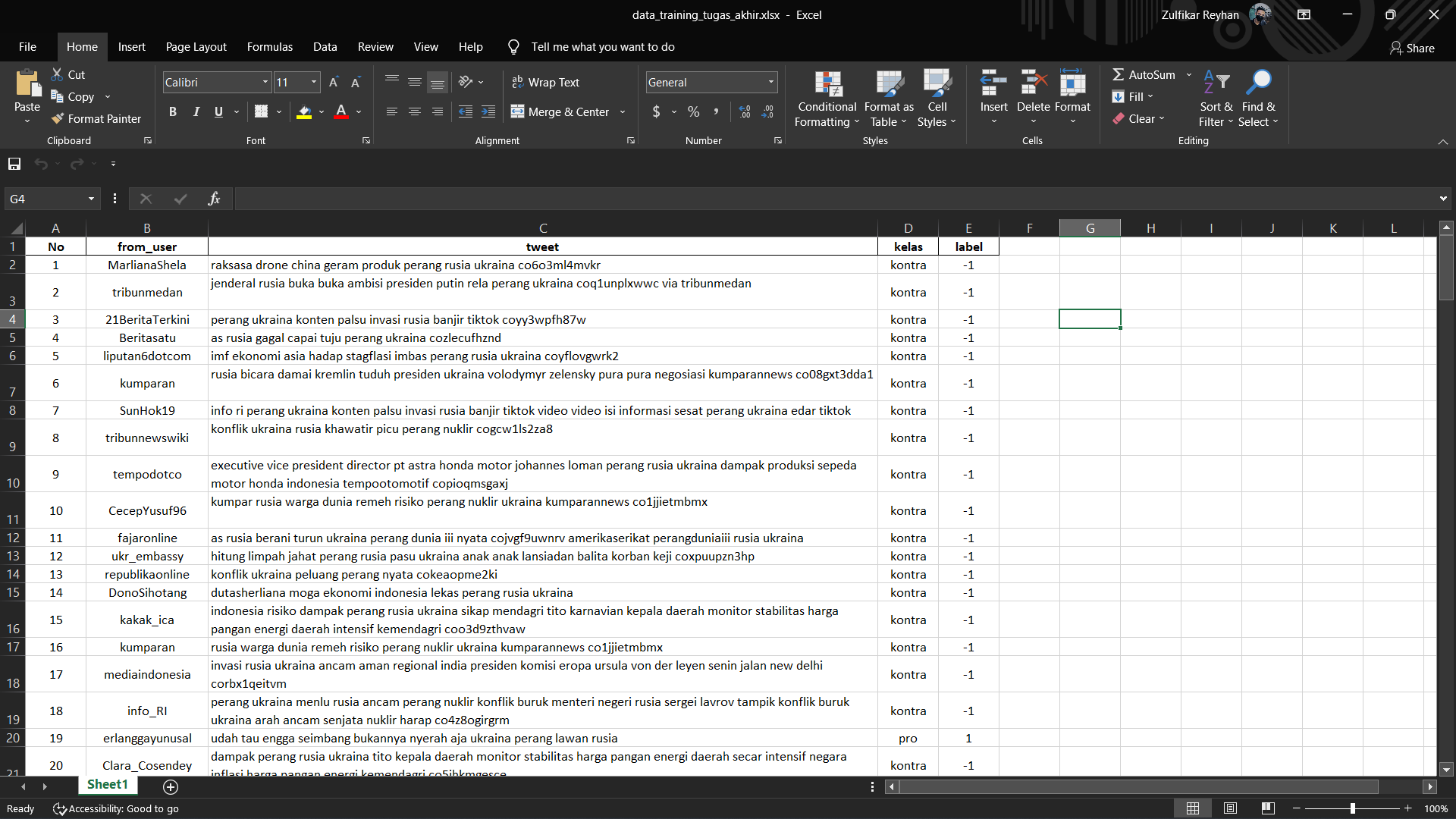
Gambar 4. Hasil preprocessing text

Pertama data di ubah menjadi huruf kecil, lalu data di bersihkan dari angka symbol dan space kosong, setelah itu kata kata yang tidak ada makna dihapus menggunakan stopwords, setelah itu baru data keluarkan lagi dengan format excel.

1. Pelabelan tweet

Pada tahap setelah data dilakukan prepocessing, data kemudian dilabel berdasarkan konten yang terkandung didalamnya, jika konten itu mendukung Tindakan rusia makan akan dilabeli Pro, namun jika sebaliknya maka akan dilabeli Kontra.

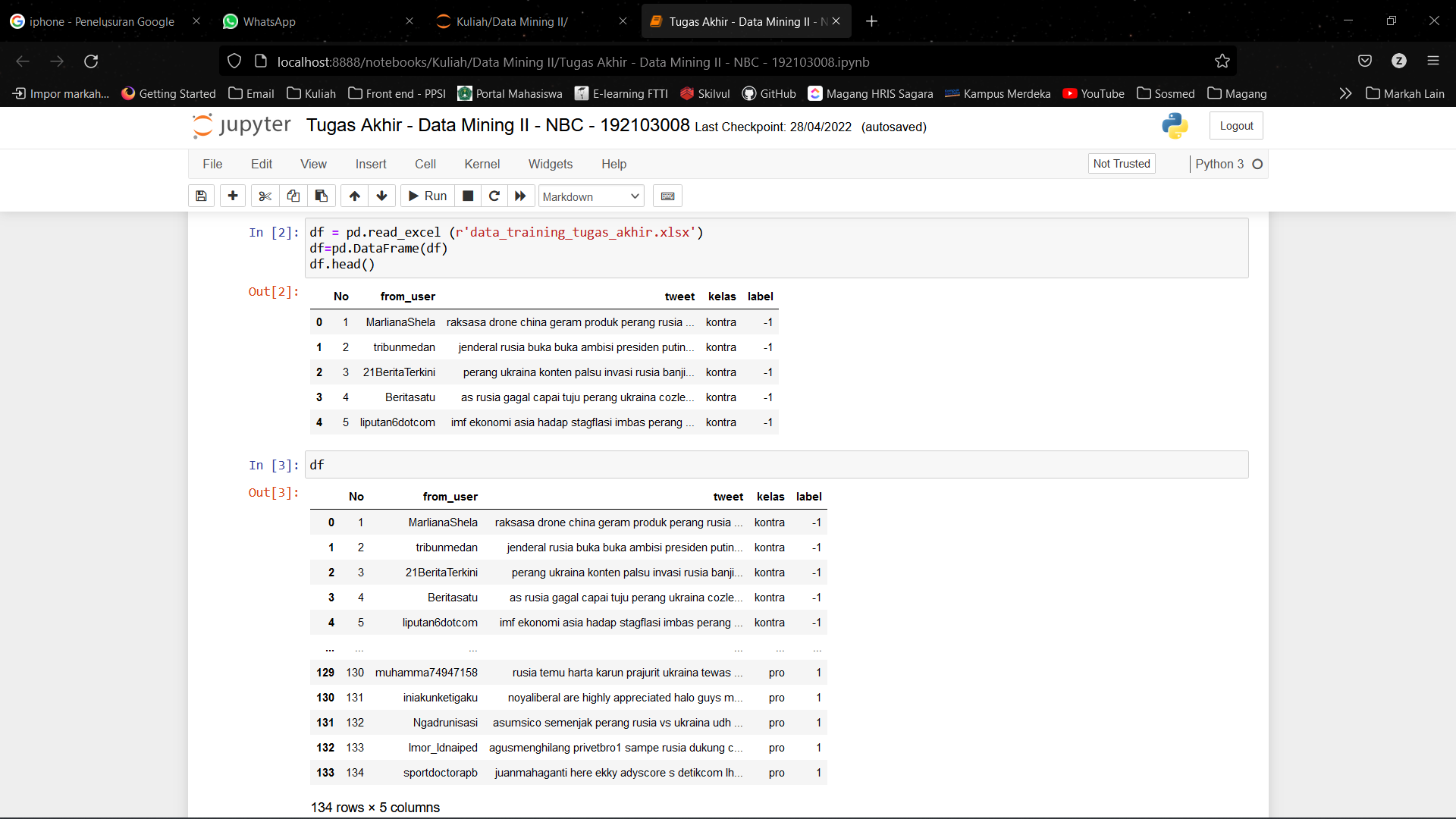
Saat proses pelabelan ditemukan sutu issue dimana ada data tweet yang sama dengan pembuat yang berbeda. Oleh sebab itu penulis melakukan pembersihan lagi data yang duplikasi dengan tools dari excel.



Gambar 5. Pelabelan tweet

1. Pelatihan/Training

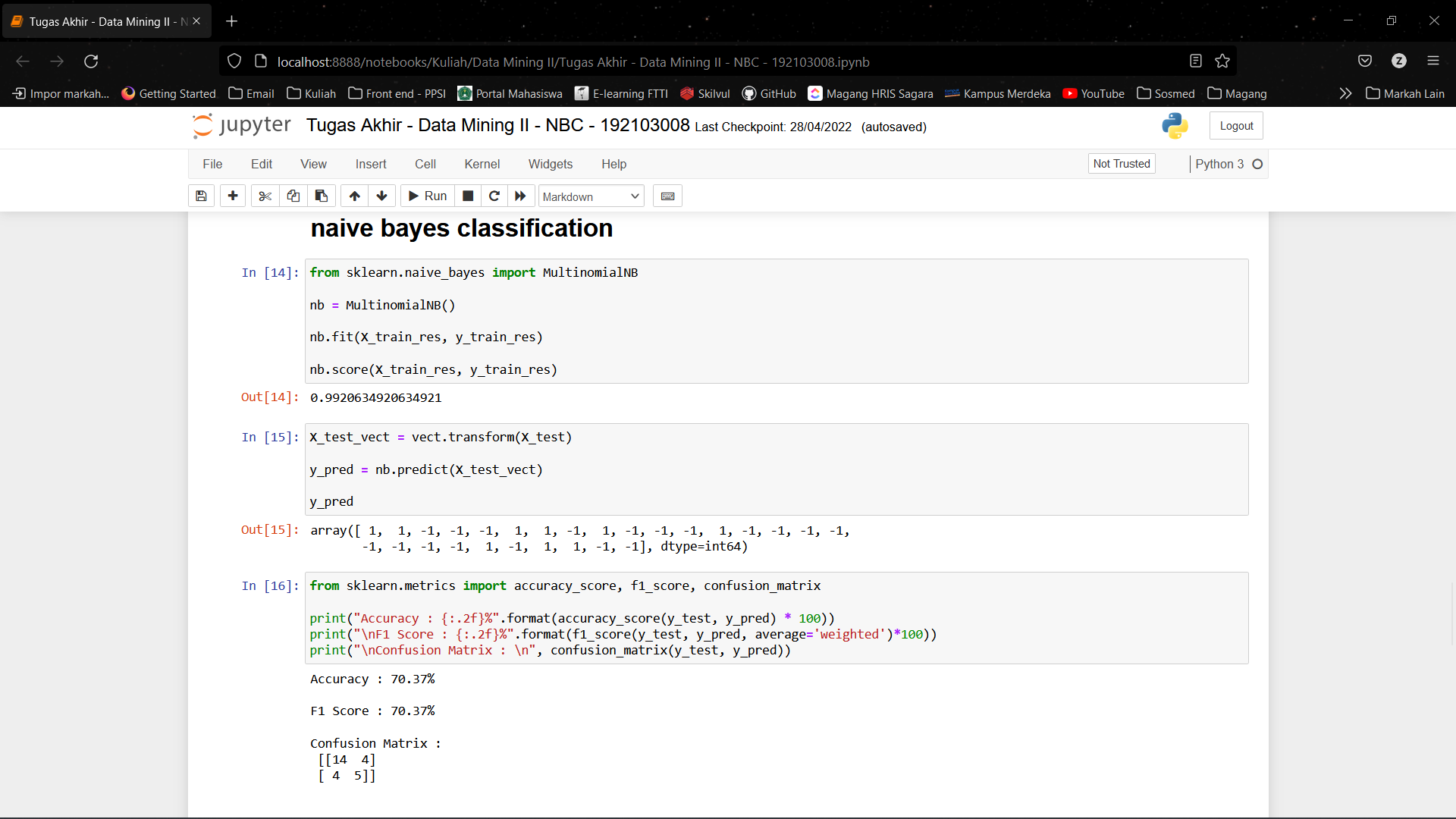
Dari 1000 data yang diambil dari website Tag, hanya 135 data tweet yang digunakan untuk menjadi data training. Karena banyak tweet yang ganda dan banyak tweet yang tidak mengandung topik yang sedang dipecahkan oleh penulis.



Gambar 6. Data latih/Training

1. Nilai Evaluasi

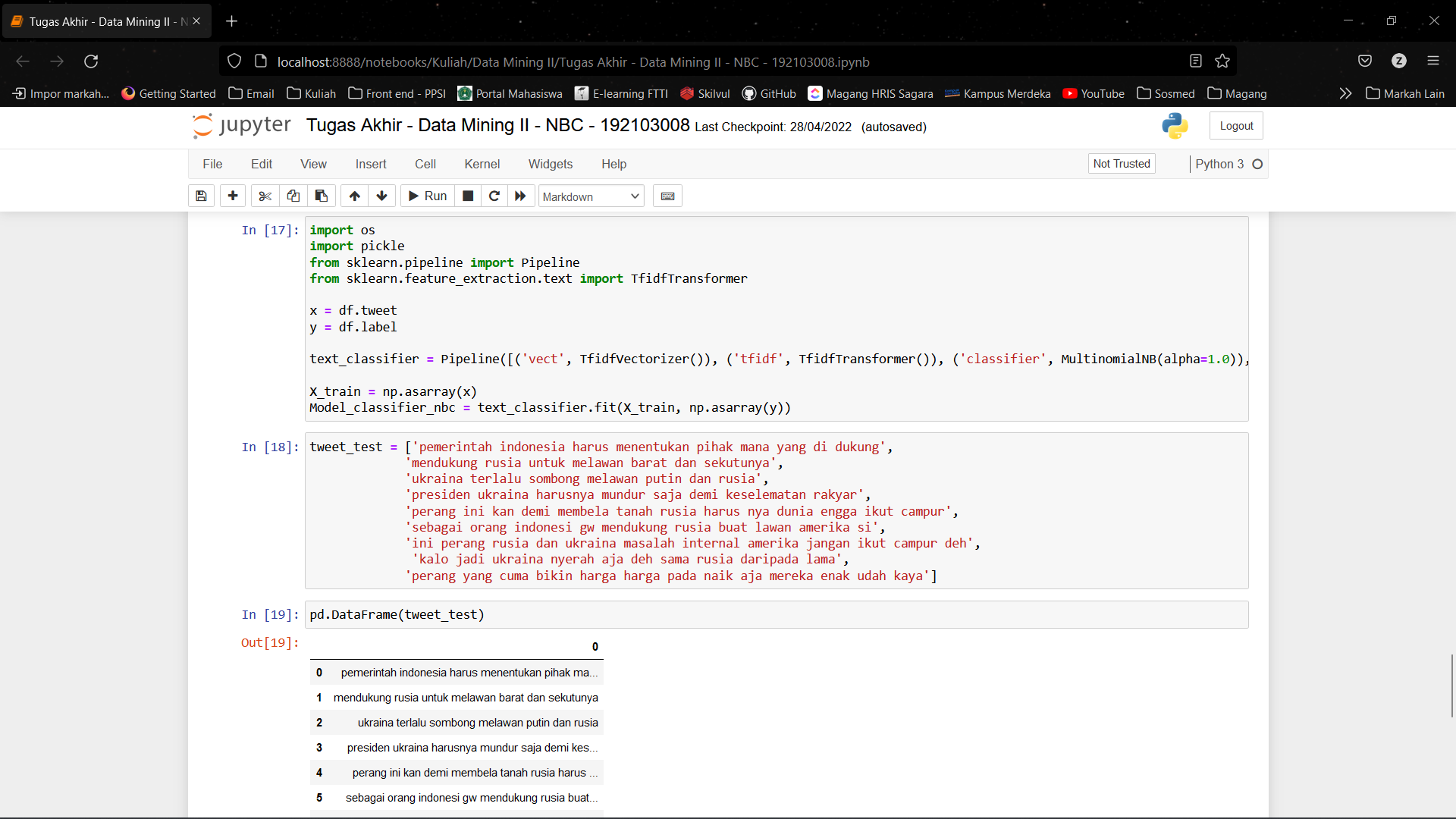
Hasil pengecekan akurasi pada program ini, penulis mendapatkan hasil akurasi yaitu 77,78%. Hasil tersebut di dapatkan dari modul yang di ambil dari sklearn.metric. Lalu untuk pengecekan F1 Score pada program ini didapatakn hasil yaitu 78,41% dimana dihitung dari presisi dan penarikan kembali test.



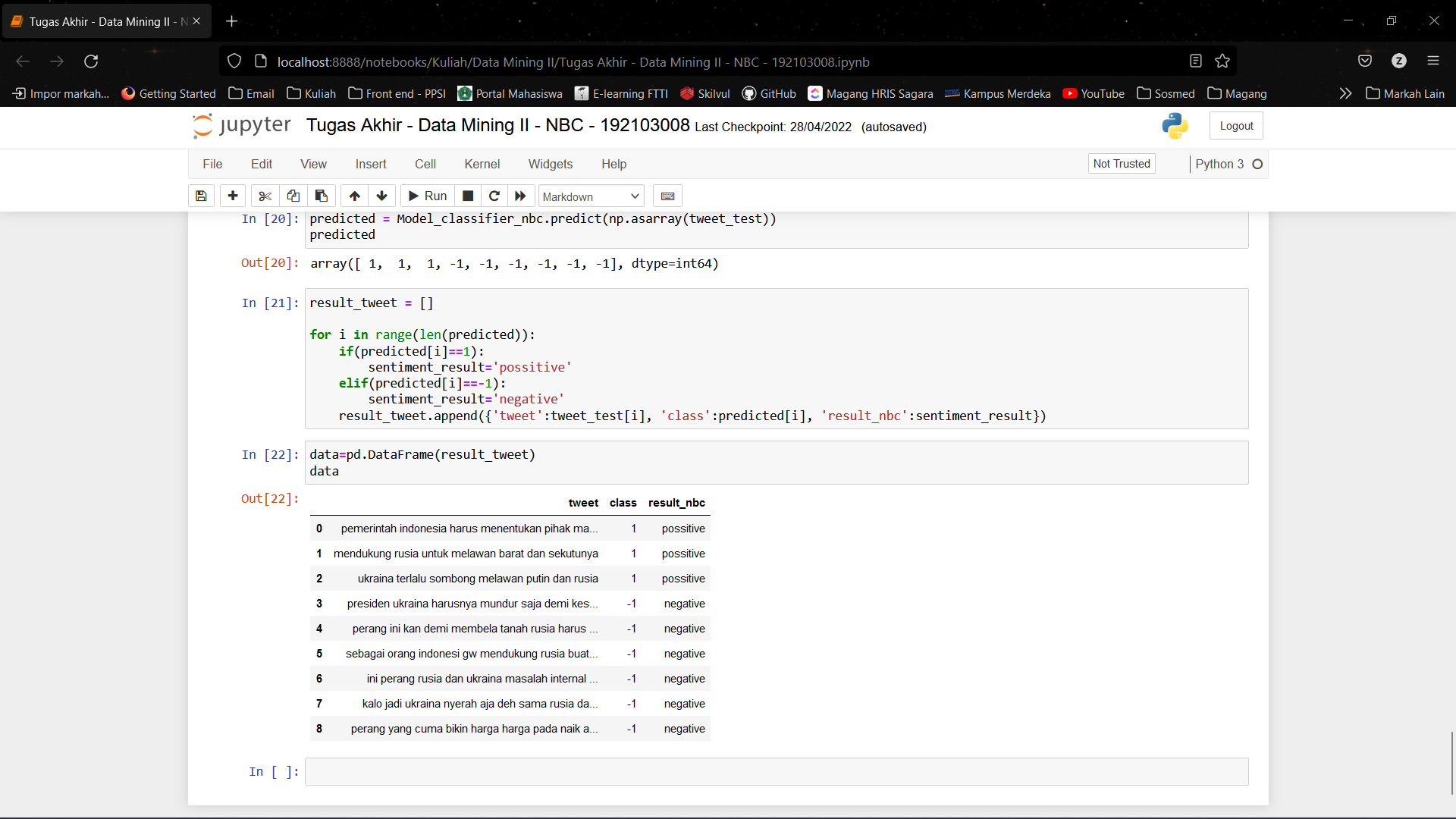
Gambar 7. Hasil Evaluasi

1. Klasifikasi data testing

Tahap terakhir yaitu penulis memasukan tweet baru yang akan di proses oleh program untuk mengetahui klasifikasi dari tweet tersebut secara otomatis dengan hasil dari tweet traing sebelumnya.



Gambar 8. Tweet test



Gambar 9. Hasil klasifikasi

# BAB 3 KESIMPULAN DAN SARAN

## KESIMPULAN

Hasil dari praktikum pengerjaan ini adalah program yang dapat menentukan suatu tweet memiliki stetment yang pro atau kontra dengan topik yang dipilih oleh penulis yaitu Perang Rusia dengan Ukraina.

Program yang penulis rancang ini menghasilkan program yang membantu dalam mengelompokan sebuah tweet termasuk mendukung atau menolak tidakan rusia terhadap ukraina.

## SARAN

Program ini belum sempurna dan belum memiliki akurasi yang sangat tinggi, program ini juga masih jauh dari kata sempurna dan tentu masih banyak kekurangan baik yang disengaja maupun karena keterbatasan dari penulis sendiri.

Oleh karena itu program ini perlu dikembangkan dan diperbaiki lagi, agar dapat menjadi lebih baik lagi. Maka penulis memiliki beberapa saran yaitu sebagai berikut:

1. Dalam menambang data pada tweeter sudah banyak akun akun yang tidak berkompetibel, oleh karena itu perlu dikurasi lagi.
2. Data yang didapat dengan menggunakan kata kunci bisa juga menarik pula tweet yang membicarakan hal lain, oleh karena itu perlu dikurangi lagi.
3. Karena terlalu banyaknya data dabel dengan akun pembuat yang sama maupun berbeda supaya bisa dihapus dulu data dabel tersebut.

# DAFTAR PUSTAKA

Darwis, D., Siskawati, N. and Abidin, Z. (2021) ‘Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional’, *Jurnal Tekno Kompak*, 15(1), p. 131. doi: 10.33365/jtk.v15i1.744.

Purwiantono, F. E. and Aditya, A. (2020) ‘Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text’, *Jurnal Tekno Kompak*, 14(2), p. 68. doi: 10.33365/jtk.v14i2.709.

Rozi, I. F., Hamdana, E. N. and Iqbal Alfahmi, M. B. (2018) ‘PENGEMBANGAN APLIKASI ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER (Studi Kasus SAMSAT Kota Malang)’, *Jurnal Informatika Polinema*, 4(2), p. 149. doi: 10.33795/jip.v4i2.164.

Samir *et al.* (2021) ‘Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes’, *Jurnal Media Informatika Budidarma*, 5(1), p. 149. doi: 10.30865/mib.v5i1.2604.