**LAPORAN AKHIR PROJEK**

**‘STORE REVIEW SENTIMENT ANALYSIS’**



**Disusun Oleh:**

**KELOMPOK 5**

Hendra Cahyono 22031554029

Ahmad Rafi Syaifuddin 22031554030

Filda Zamardah 22031554047

**S1 SAINS DATA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS NEGERI SURABAYA**

**2023**

**PENDAHULUAN**

**1.2 Latar Belakang**

Dalam era digital yang semakin berkembang perusahaan tidak hanya berinteraksi oleh pelanggan tetapi juga dihadapkan oleh ulasan dan umpan balik yang diberikan oleh pelanggan melalui berbagai platform atas pelayanan yang diberikan. Ulasan tersebut dapat memberi dampak yang banyak bagi perusahaan tersebut. Sudah banyak platform bagi para pelanggan untuk memberikan ulasannya terhadap perusahaan baik itu ulasan yang bersifat positif, negative, maupun netral. Ulasan yang diberikan pelanggan tidak hanya bentuk interaksi antara pelanggan kepada perusahaan tetapi dapat berpengaruh membentuk persepsi publik terhadap perusahaan tersebut.

Dengan begitu sangat penting bagi perusahaan untuk mengamati, memantau, dan memberi respons terhadap ulasan yang telah pelanggan berikan. Dengan memperhatikan padangann, kritik, kebutuhan yang dibutuhkan pelanggan dalam pelayanan perusahaan tersebut akan berdampak untuk memperbaiki kinerja dalam perusahaan tersebut, meningkatkan pelayanan, dan terjalin hubungan yang baik oleh pelanggan.

**1.2 Rumusan Masalah**

1. Bagaimana pengaruh ulasan yang diberikan pelanggan terhadap perusahaan
2. Bagaimana mengatur strategi perusahaan berdasarkan data ulasan yang diberikan pelanggan
3. Bagaimana respons yang diberikan perusahaan terhadap ulasan pelanggan
4. Mengapa kata yang tidak ada di model otomatis menjadi very positif

**1.3 Tujuan dan Manfaat**

Dari pembuatan projek ini yaitu membantu perusahaan untuk menganalisis sentiment yang diberikan pelanggan sehingga dapat digunakan untuk memproses dan memahami makna yang diberikan oleh pelanggan. Dari analisis sentiment tersebut Perusahaan dapat mengidentifikasi tren yang terdapat pada ulasan, mengklasifikasikan ulasan menjadi tiga jenis yaitu positif, perusahaan, dan netral, perusahaan juga dapat mengambil perusahaan yang tepat dengan mempertimbangkan hasil ulasan yang diberikan pelanggan.

Manfaat dari projek ini yaitu dapat mengetahui apa yang dirasakan oleh pelanggan seperti apa yang pelanggan suka dan tidak suka, dapat memperbaiki dan meningkatkan pelayanan yang ada di perusahaan, dan dapat menjadi dasar dalam mengatur strategi untuk perkembangan bisnis perusahaan tersebut.

**METODE**

* 1. **PREPROCESSING**

Pada preprocessing dilakukan beberapa tahapan diawali dengan Remove Puntuation yaitu menghapus tanda baca yang terdapat pada teks. Selanjutnya dilakukan Lower Case yang digunakan untuk mengubah huruf yang terdapat pada teks menjadi huruf kecil semua. Setelah itu dilakukan Remove Stopwords yang digunakan untuk menghapus kata yang sering muncul pada teks tetapi kata tersebut tidak memiliki makna yang penting. Setelah teks tersebut sudah menjadi teks yang berisi informasi penting saja maka dilakukan Lemmatizer untuk mengelompokan berbagai bentuk kata yang ada pada teks sehingga kata-kata tersebut dapat dianalisis oleh lemma kata atau bentuk kamus.

* 1. **Ekstraksi Fitur**
* TF-IDF merupakan proses yang dapat dilakukan untuk transformasi data dari bentuk data tekstual dirubah menjadi data numerik yang digunakan untuk pembobotan pada setiap kata atau fitur. TF-IDF memeriksa frekuensi kemunculan kata pada dokumen dan menunjukan seberapa penting kata dalam dokumen tersebut [2].
* Skip N-Gram merupakan teknik yang digunakan dalam bidang pemrosesan ucapan dengan bentuk bigram, tri-gram, dan masih bayak lagi. Selain memungkinkan untuk rangkaian kata, Teknik ini juga mengizinkan token untuk dilewati [3].
* Naïve Bayes merupakan suatu metode yang digunakan untuk klasifikasi yang berakar dari teorema bayes. Metode naïve bayes ini menggunakan metode probabilitas dan statistic untuk memprediksi suatu peluang di masa depan berdasarkan pengalaman yang telah didapat sebelumnya [4].
* SVM merupakan algoritma pada machine learning yang berprinsip pada structural risk minimization (SRM) yang memiliki tujuan untuk menemukan hyperplane terbaik yang dapat memisahkan dua class pada input space [5].

**HASIL DAN ANALISIS**

**3.1 Ekstraksi Fitur**

Pada tahap ekstraksi fitur dilakukan dengan menggunakan dua cara. Yang pertama dengan mengimplementasikan ekstraksi fitur pada model menggunakan metode Skip N-Gram. Hasil implementasi ekstraksi fitur tersebut menghasilkan akurasi sebesar 60.16%. Selanjutnya yaitu ekstraksi fitur pada model dengan menggunakan metode TF-IDF. Hasil implementasi ekstraksi fitur tersebut meghasilkan akurasi sebesar 70.00%.

**3.2 Akurasi**

Untuk melakukan pemodelan dilakukan dengan empat cara yaitu yang pertama menggunakan ekstraksi fitur TF-IDF dan model SVM, lalu yang kedua yaitu TF-IDF dan Naive Bayes, selanjutnya ekstraksi fitur menggunakan Skip N-Gram dan model SVM, dan yang terakhir Skip N-Gram dan model Naive Bayes.

|  |  |
| --- | --- |
| Metode | Akurasi |
| TF-IDF dengan SVM | 93% |
| TF-IDF dengan Naive Bayes | 78% |
| Skip N-Gram dengan SVM | 89% |
| Skip N-Gram dengan Naive Bayes | 83% |

Dari tabel tersebut menunjukan bahwa pemodelan menggunakan ekstraksi fitur dengan TF-IDF dan model SVM memiliki akurasi paling tinggi dari pemodelan yang lain yaitu 92% sehingga dalam melakukan prediksi model TF-IDF dan SVM yang akan digunakan.

**3.3 Kata Yang Tidak Ada di Model**

Terjadi kesalahan yang dapat diatasi dengan mengelompokan rating lalu melakukan preprocessing teks setelah itu menangani ulasan yang kosong atau tidak memiliki kata periksa juga vocabulary yang digunakan untuk pemodelan apakah sudah benar, dan yang terakhir memastikan pemodelan yang digunakan untuk sentiment sudah dilatih dengan benar.

**KESIMPULAN**

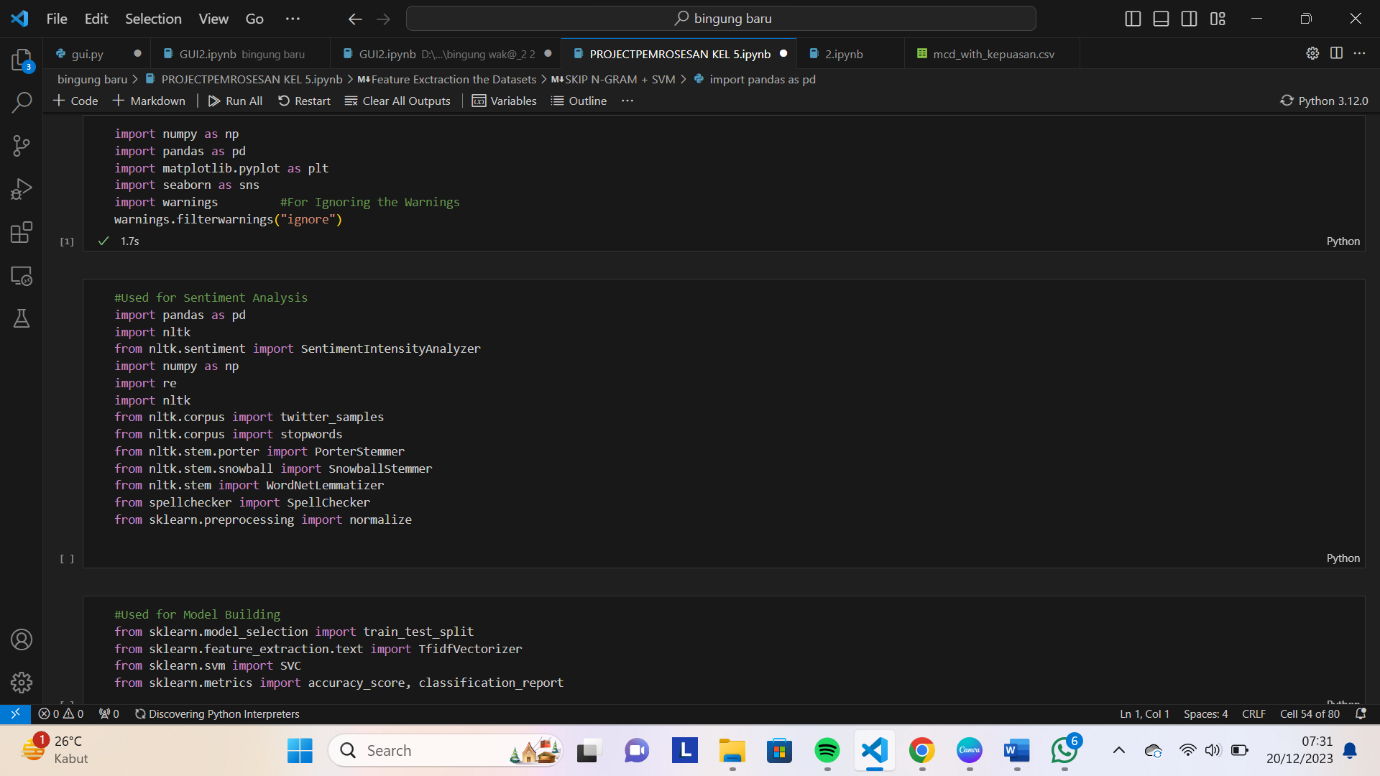
Dalam hal ini TF-IDF menghasilkan akurasi yang paling tinggi dibandingkan metode lainnya. TF-IDF sangat cocok untuk digunakan dalam klasifikasi teks seperti memprediksi sentimen dari ulasan yang pelanggan berikan. Dibandingkan metode lain TF-IDF memiliki keunggulan yang terletak dari kemampuan untuk menilai pentingnya suatu kata dalam sebuah dokumen dengan mempertimbangkan frekuensi kemunculan suatu kata dalam sebuah dokumen. Dengan menggunakan TF-IDF perusahaan dapat lebih tepat untuk mengklasifikasikan ulasan pelanggan yang bersifat positif, negatif, atau netral. Selain itu perusahaan dapat meningkatkan kualitas layanan dan mengatur strategi untuk meningkatkan pemasaran.

**REFERENSI**

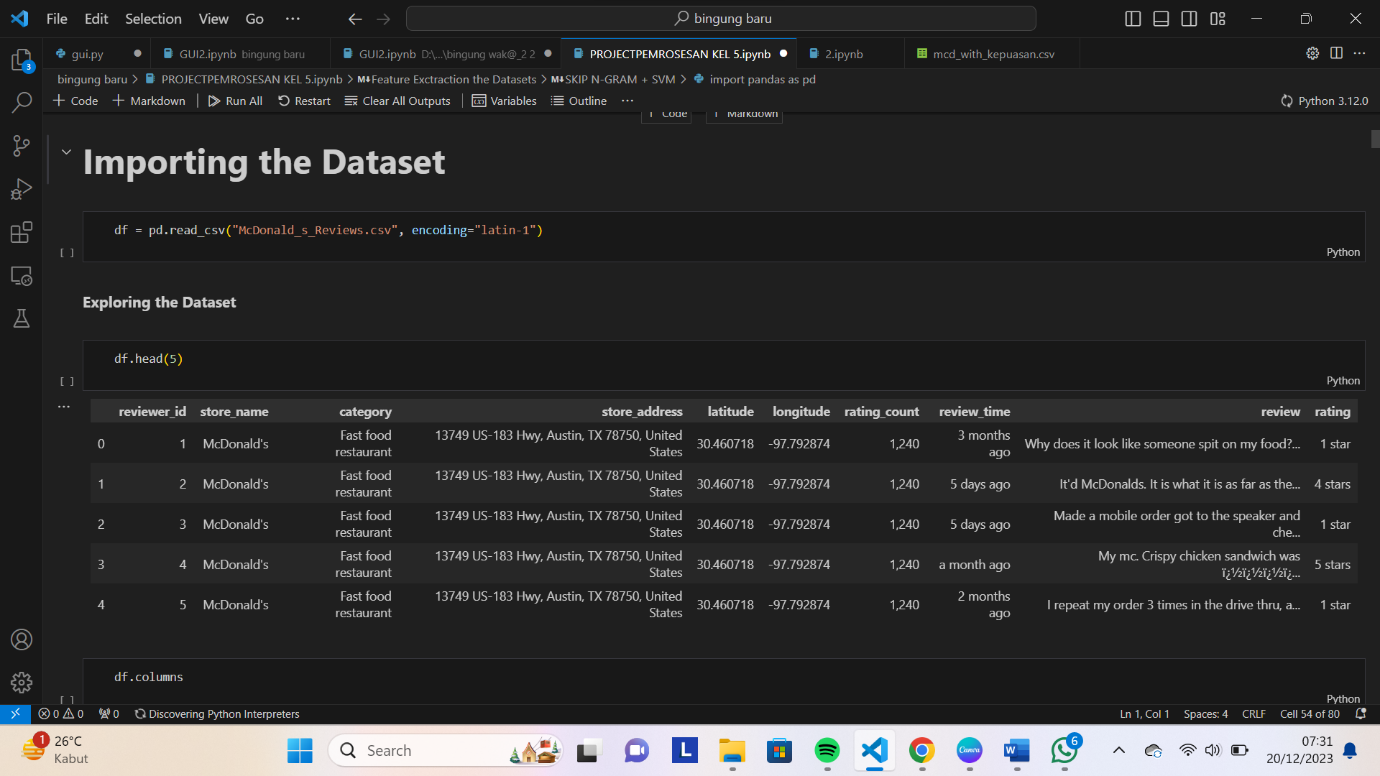
1. J. Sangeetha., U. kumaran. (2023). “ Sentiment Analysis Of Amazon User Reviews Using a Hybrid Approach”. Measurement: sensors. <https://doi.org/10.11591/ijece.v9il.pp245-254>
2. Septian A. s. dkk., (2019). “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. Journal of Intelligent System and Computation.
3. Guthrie. A., *et al.* (2006). “A closer look at skip-gram modelling. Proceedings of the 5th International Conference on Language Resources and Evaluation
4. Watratan A. F., Puspita A. Moeis D. (2020). “Implementasi Algoritma Naïve Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. Journal of Applied Computer Science and Technology.
5. Irawan. D., dkk. (2021). “Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naïve Bayes Classifier, Random Forest dan Bagging Classifier. Jurnal Sisfokom (Sistem Informasi dan Komputer).

**LAMPIRAN**

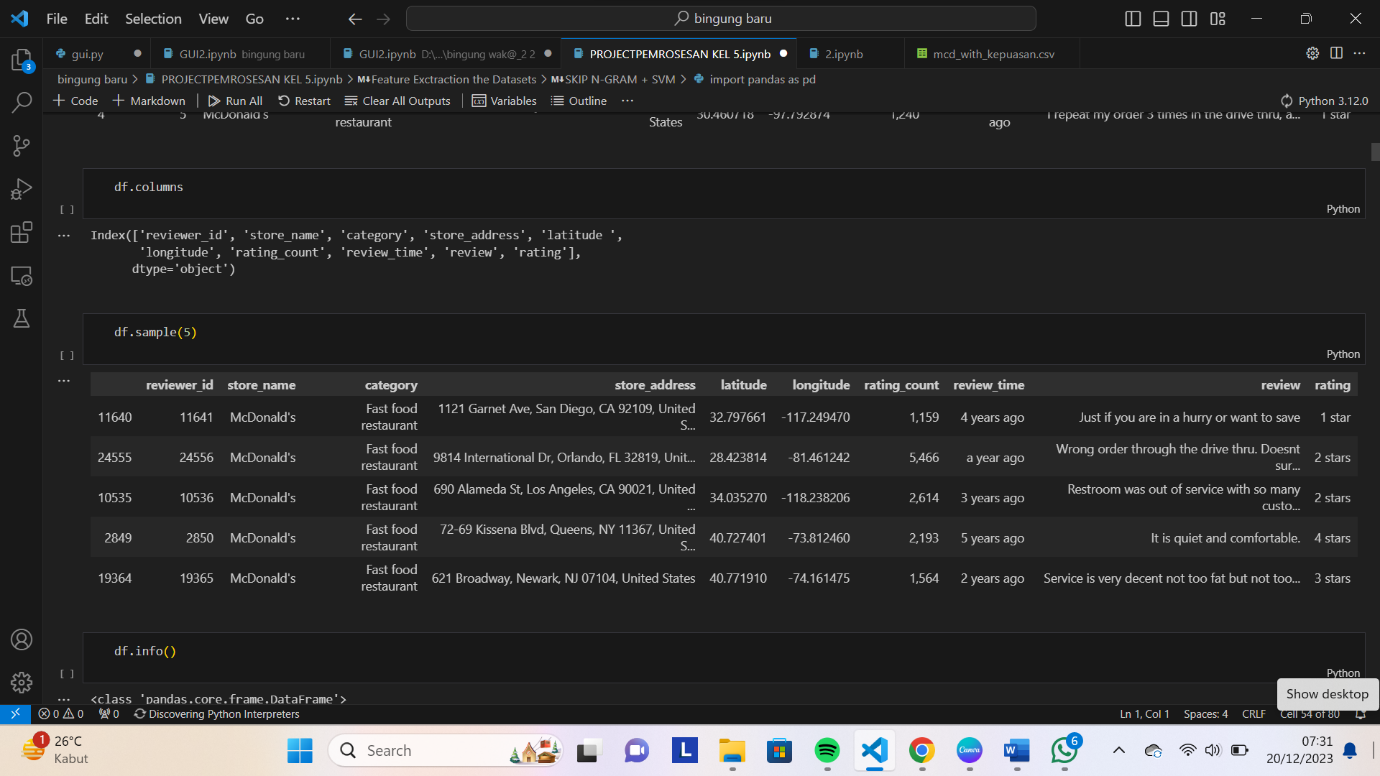
[1]. Library yang digunakan dalam pembuatan projek.



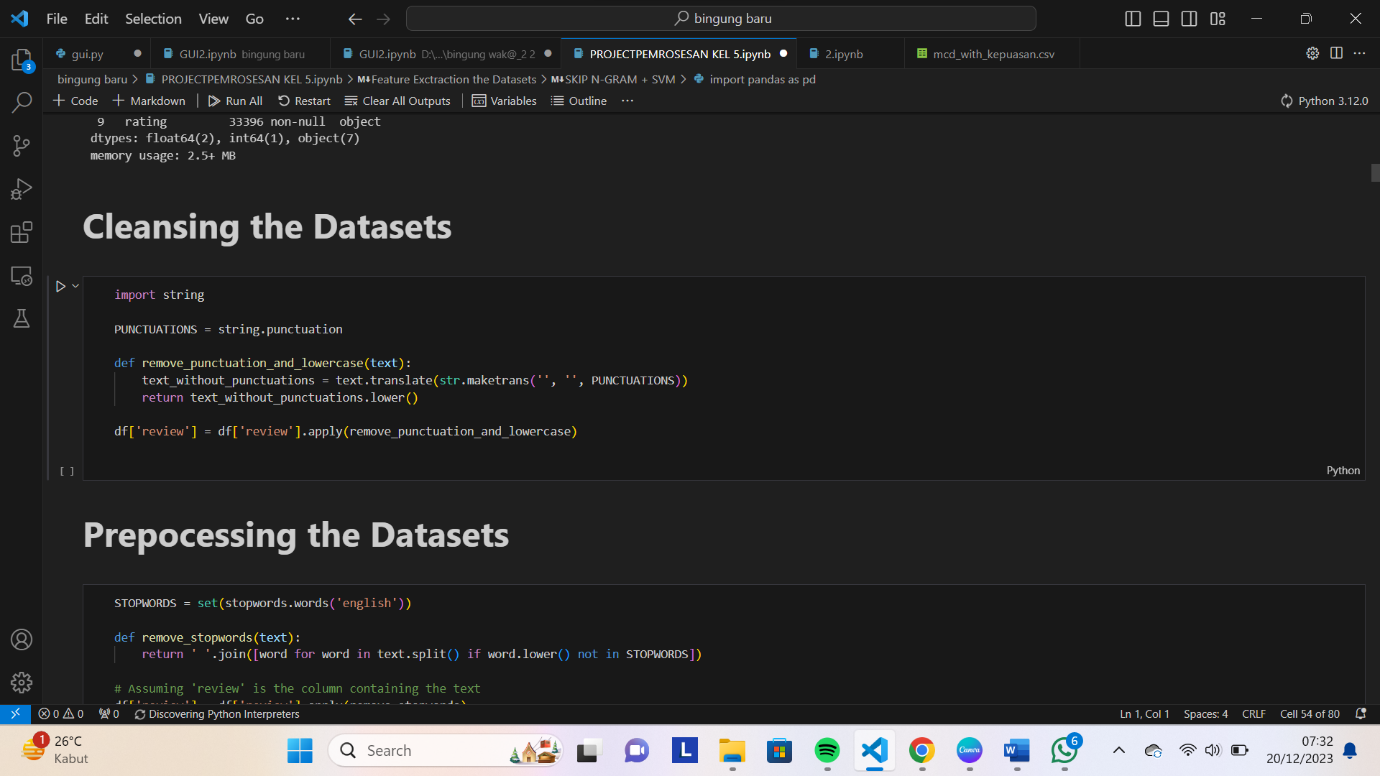
[2]. Memanggil dataset yang akan diproses, disini kami menggunakan data McD Store.

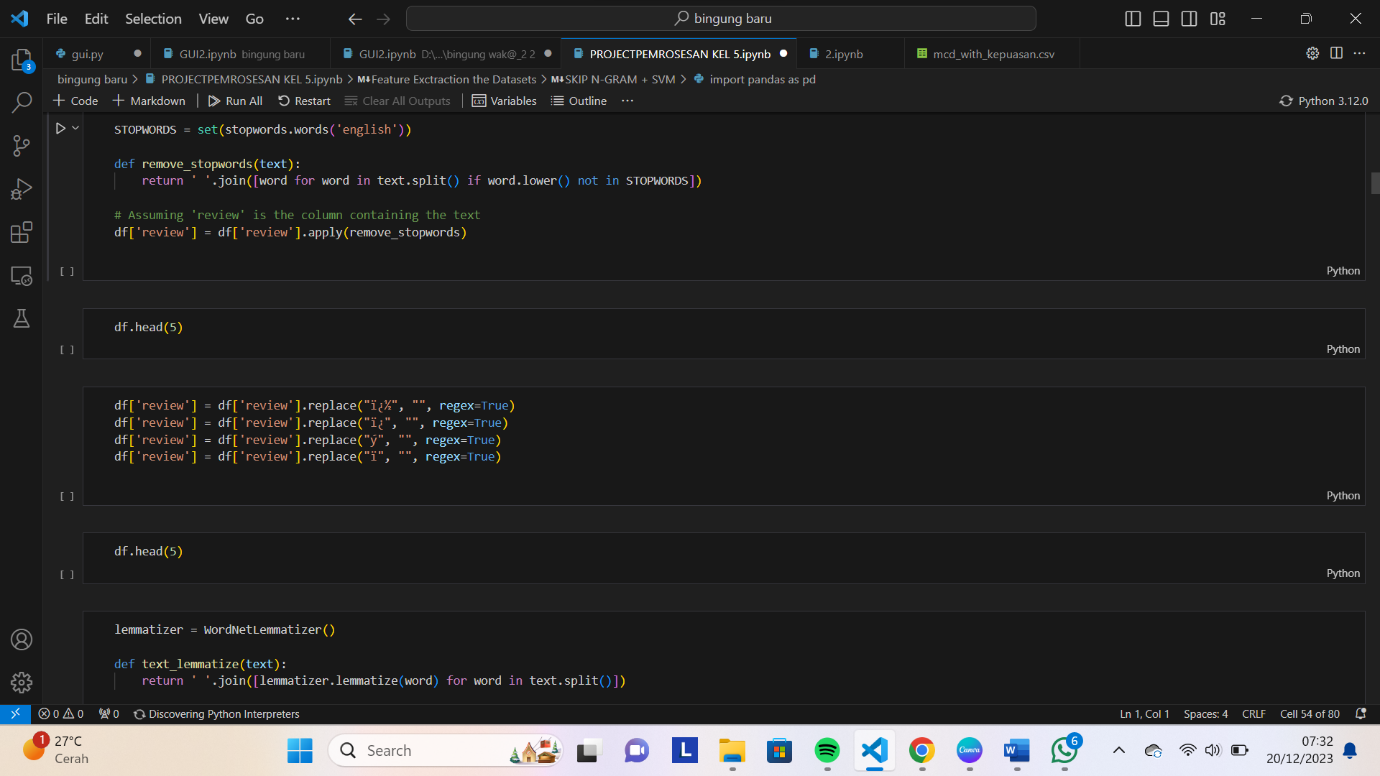


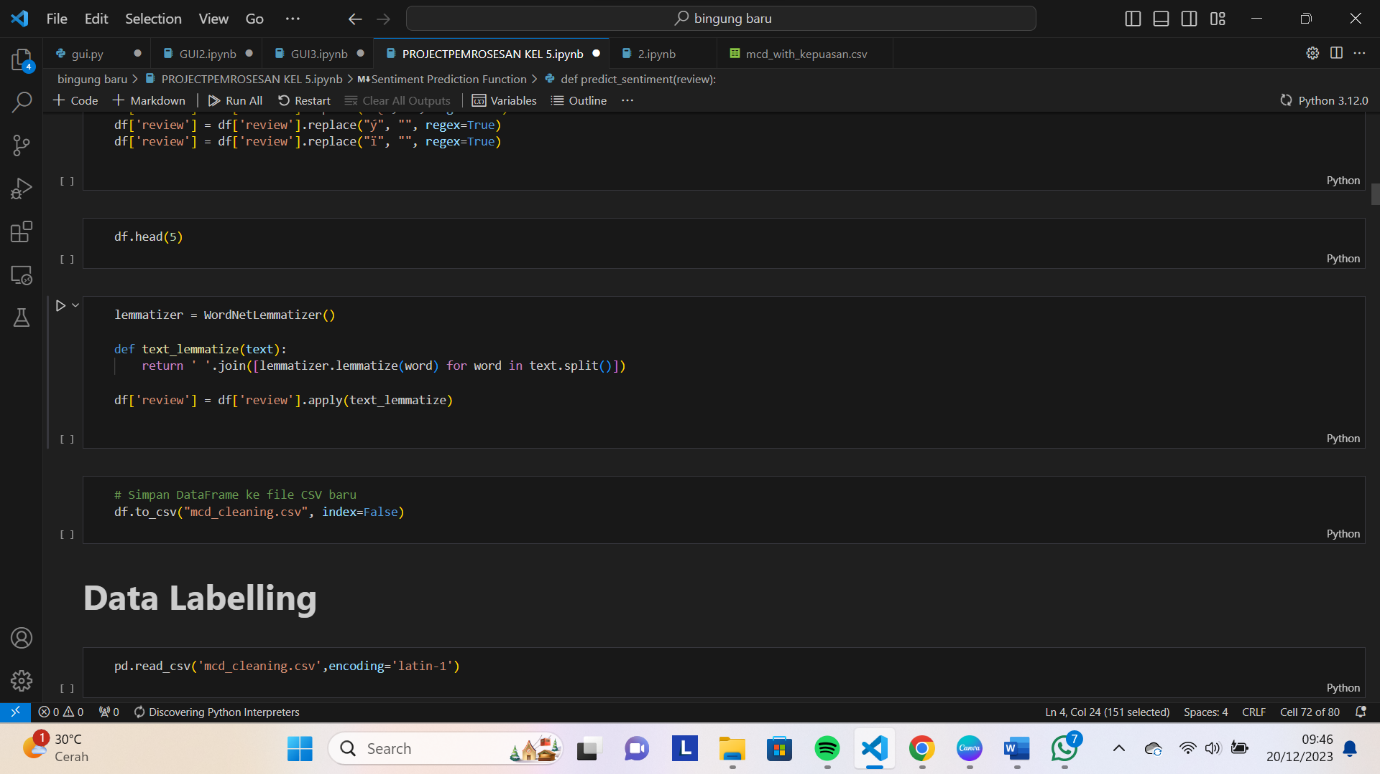
[3]. Menampilakan semua kolom yang ada dalam data yang berisi : review\_id, store\_name, category, store\_address, latitude, longitude, rating\_count, review\_time, review, dan rating.



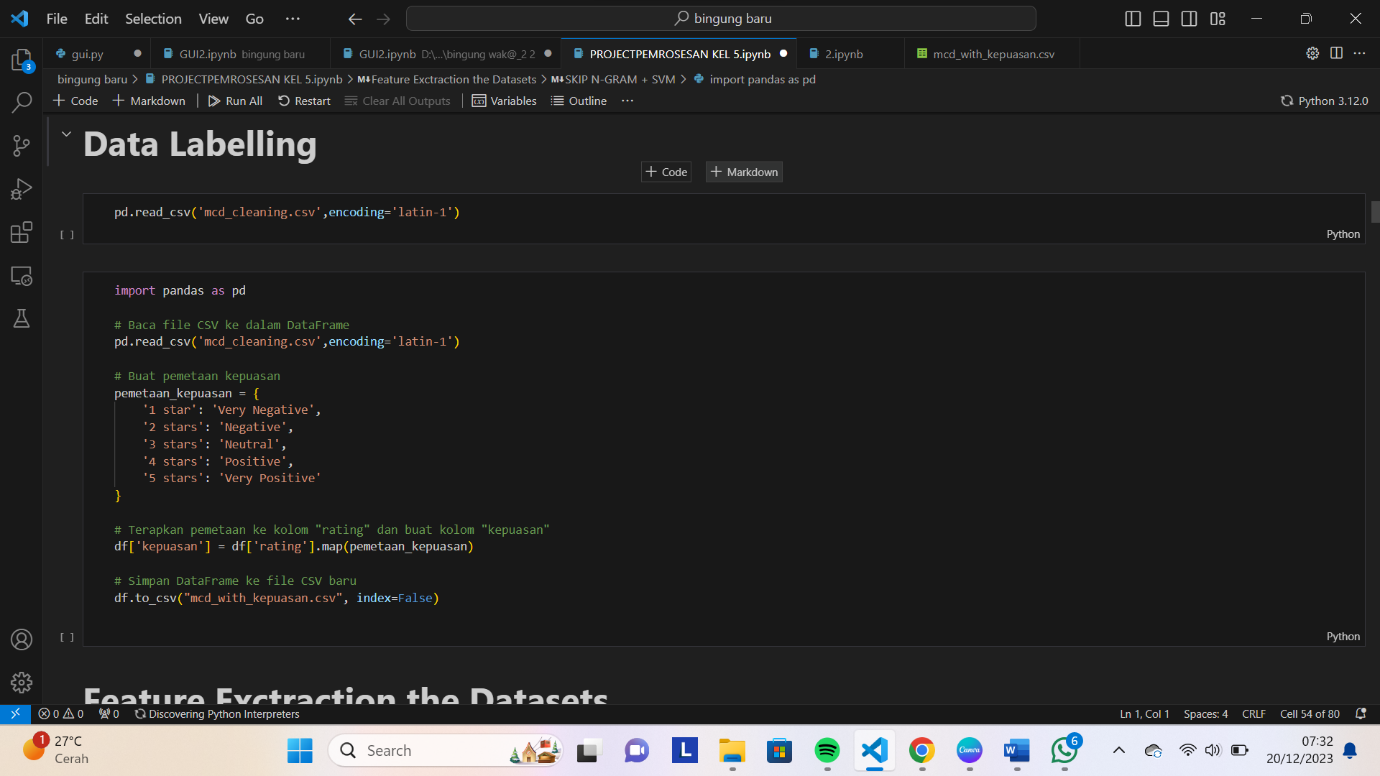
[4]. Membersihkan data dari tanda baca dan mengubah huruf yang ada pada teks menjadi huruf kecil semua. Lalu ada remove stopwords yaitu digunakan untuk menghapus kata yang tidak bermakna tetapi sering muncul, dan melakukan lemmatizer untuk mengelompokan kata.



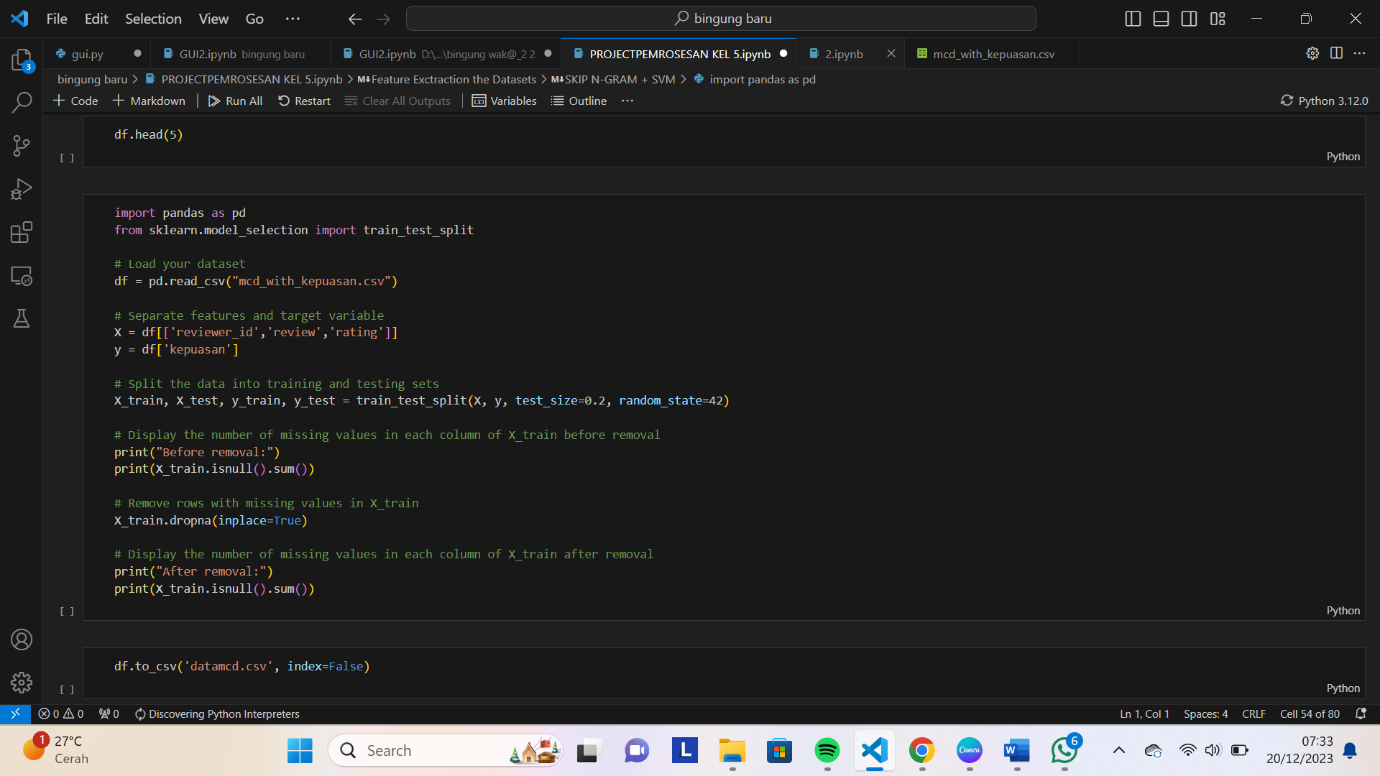




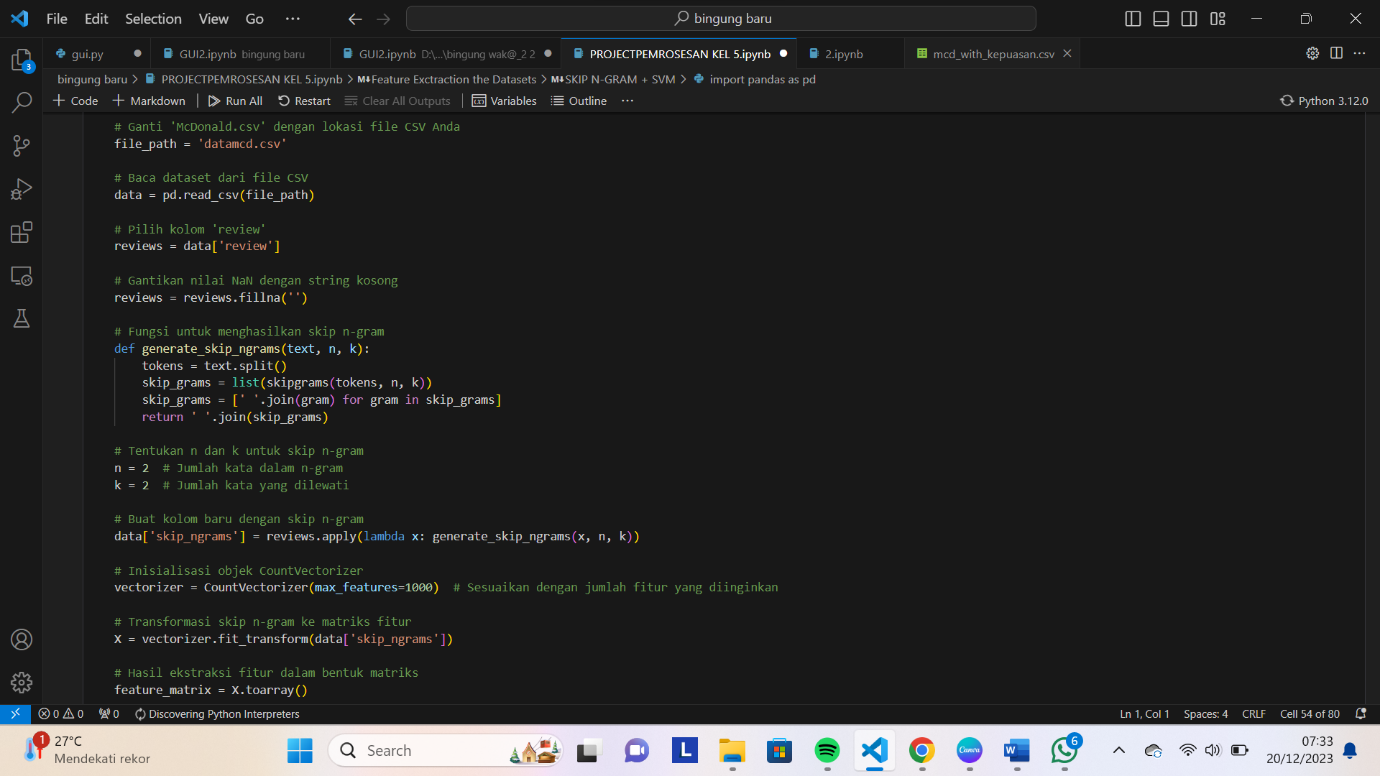
[6]. Memberi label pada data dengan lima jenis yaitu : very negative, negative, netral, positis, very positif.



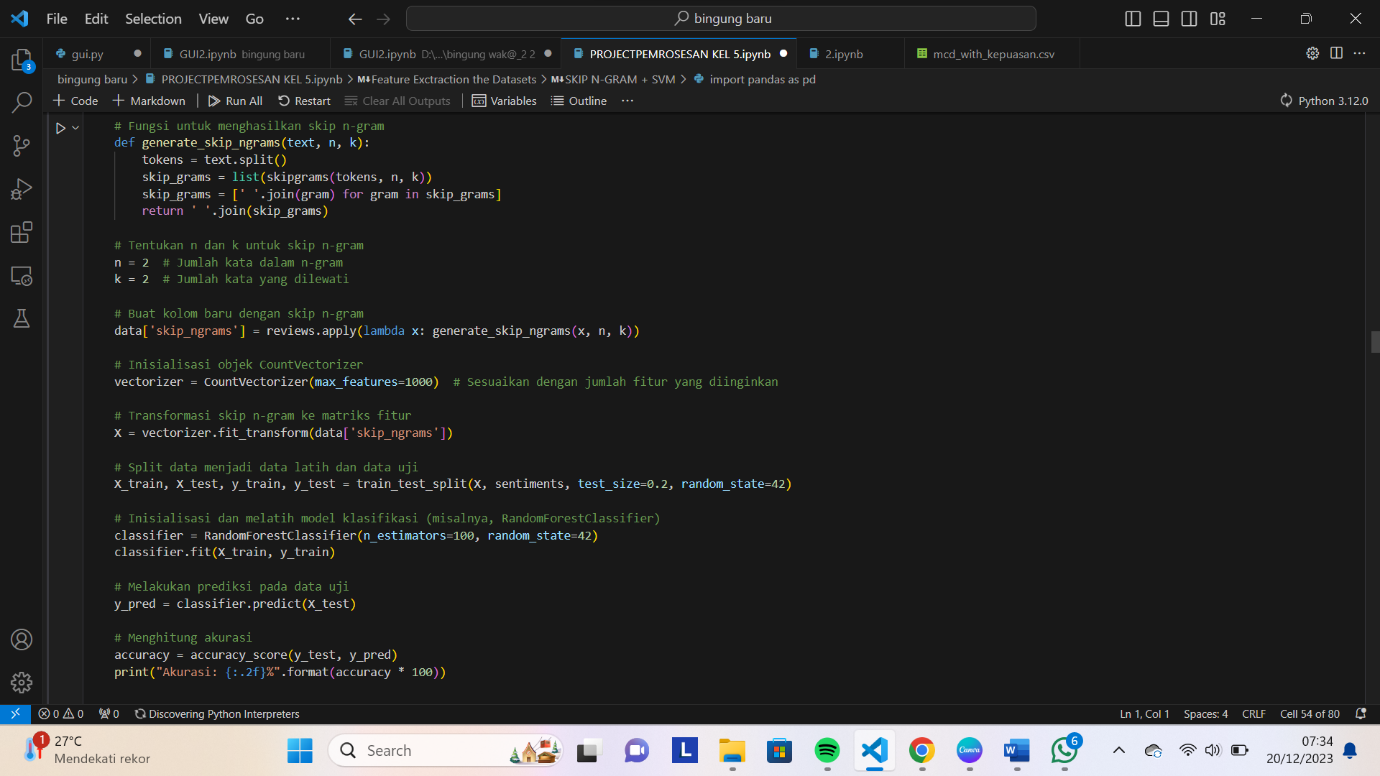
[7]. Melakukan seleksi model yang dibagi menjadi dua yaitu train dan test.



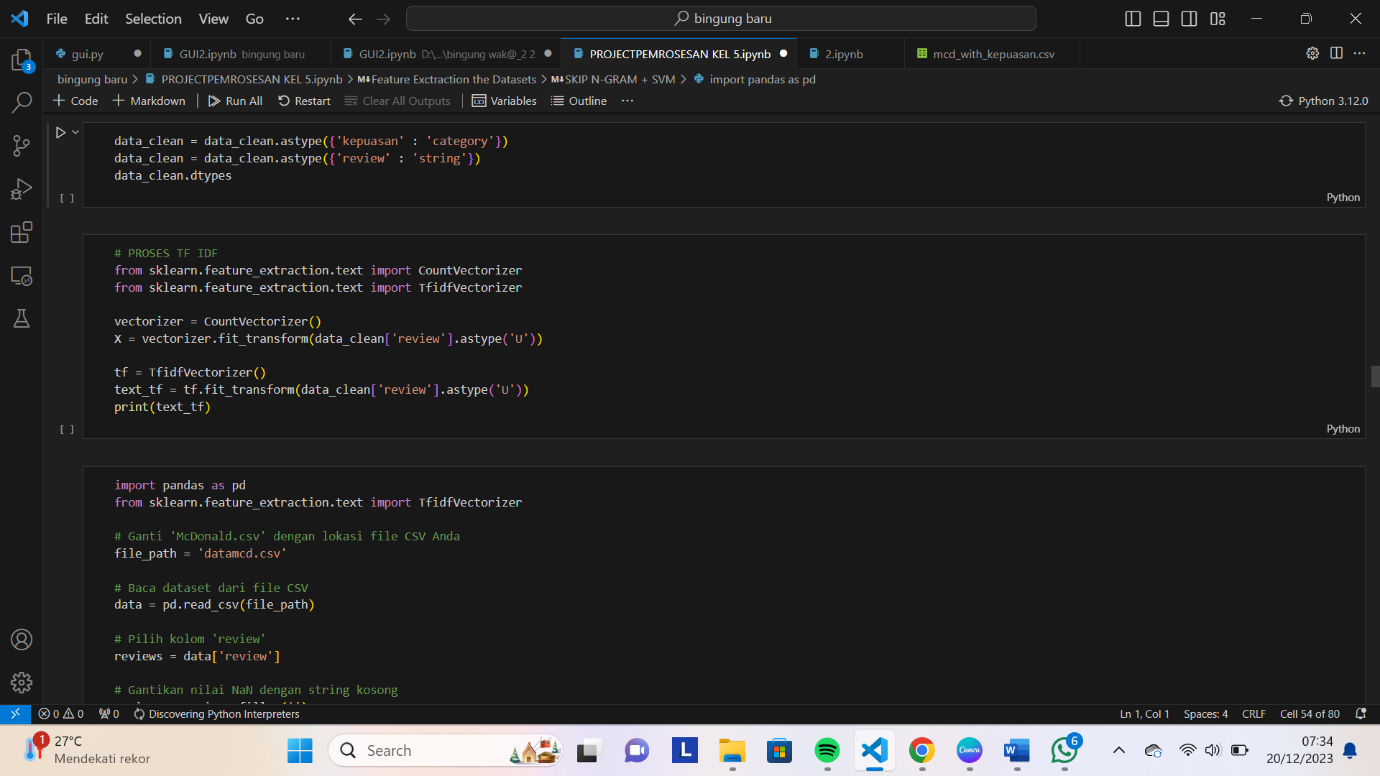
[8]. Melakukan ekstraksi fitur menggunakan metode Skip N-Gram.

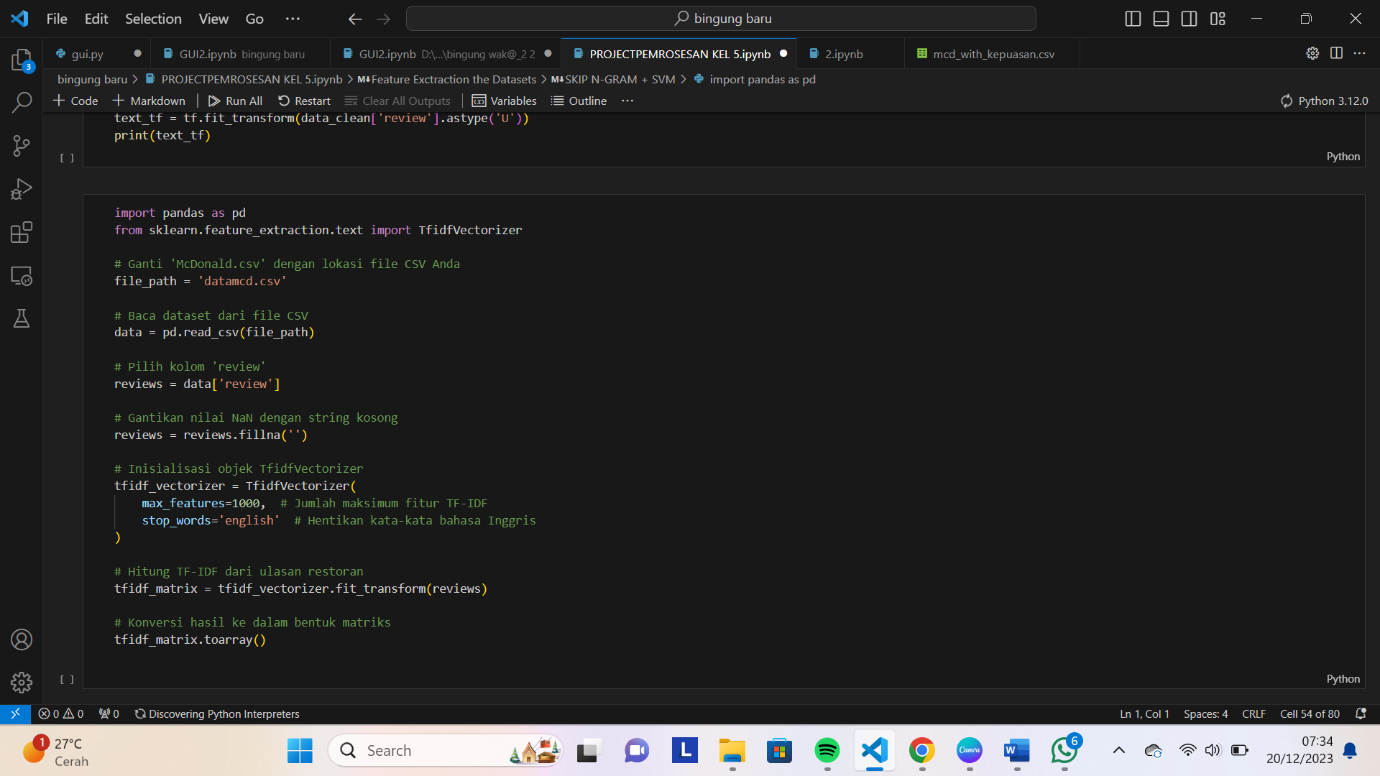


[9]. Menghitung akurasi pada Skip N-Gram.

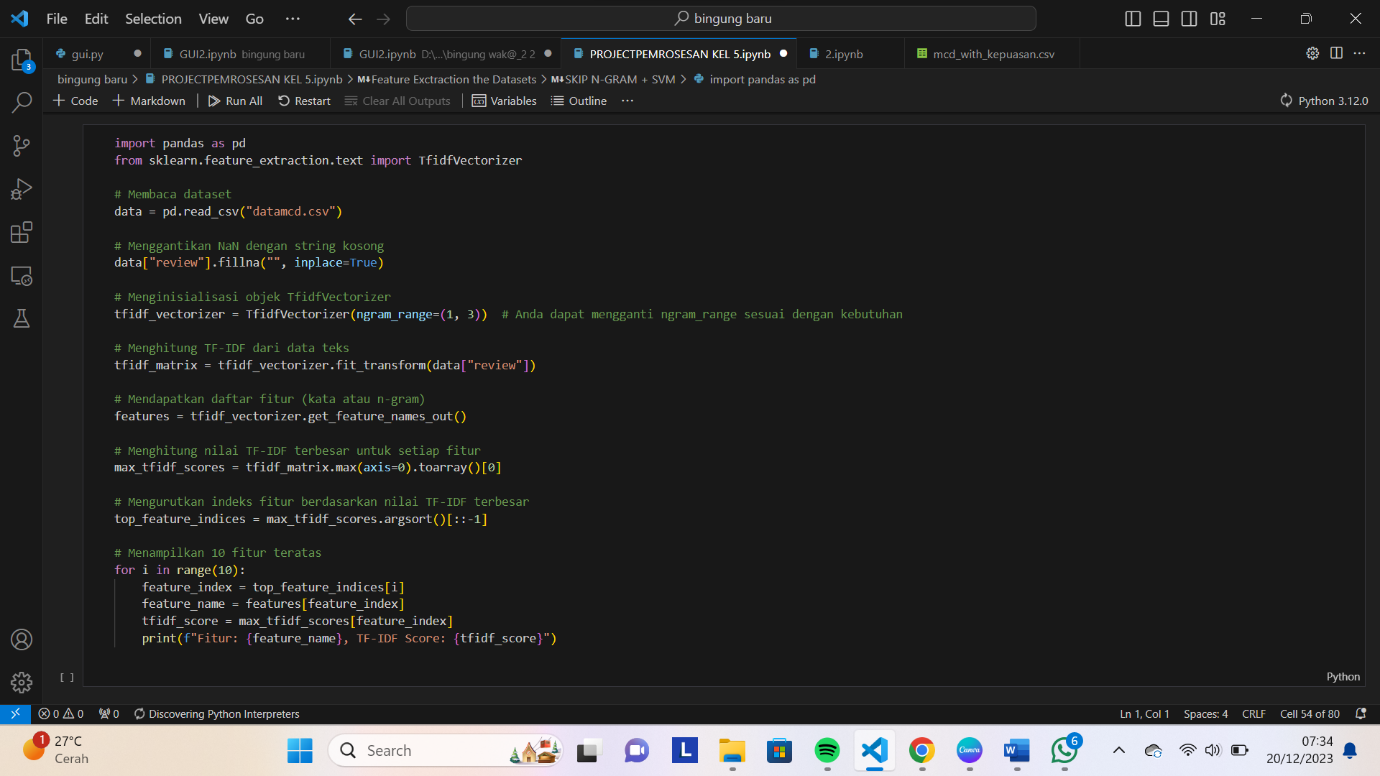


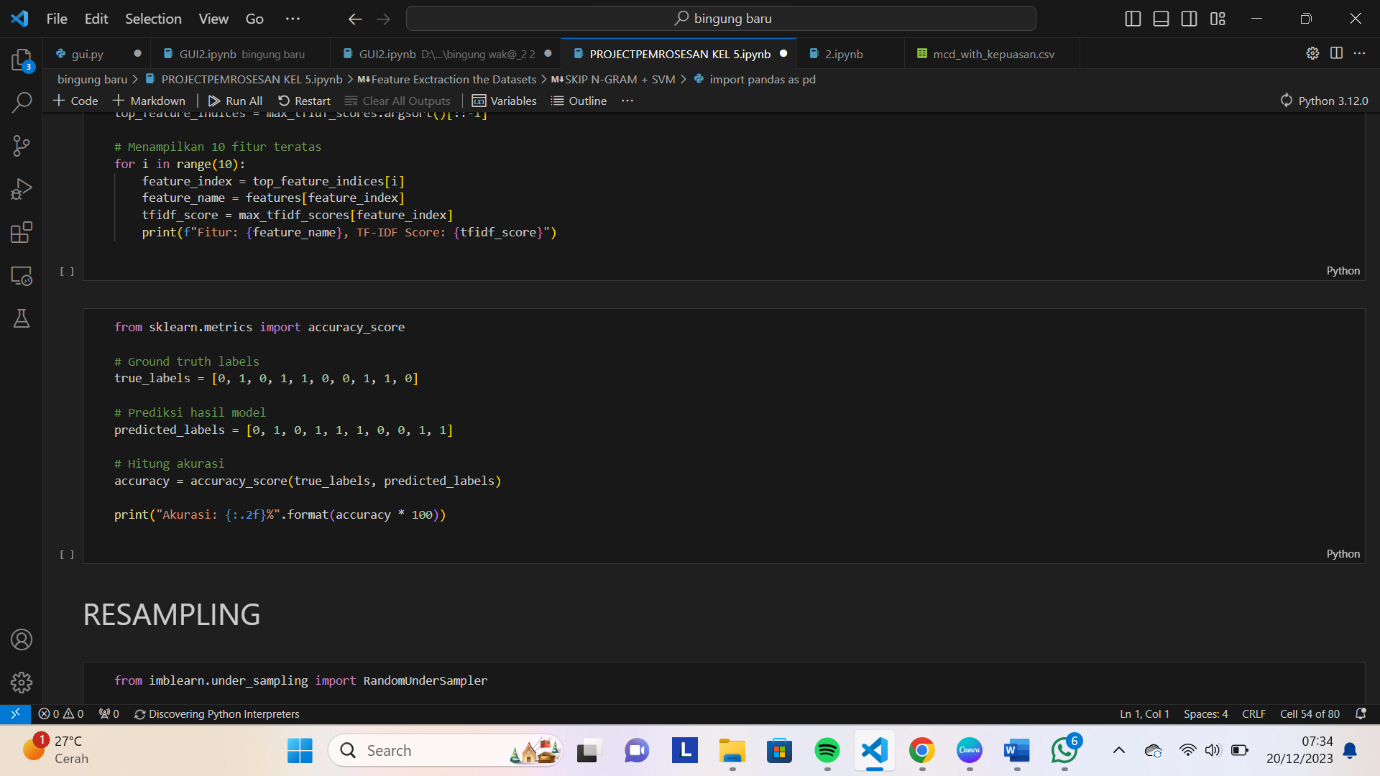
[10]. Melakukan ekstraksi fitur dengan menggunakan metode TF-IDF.



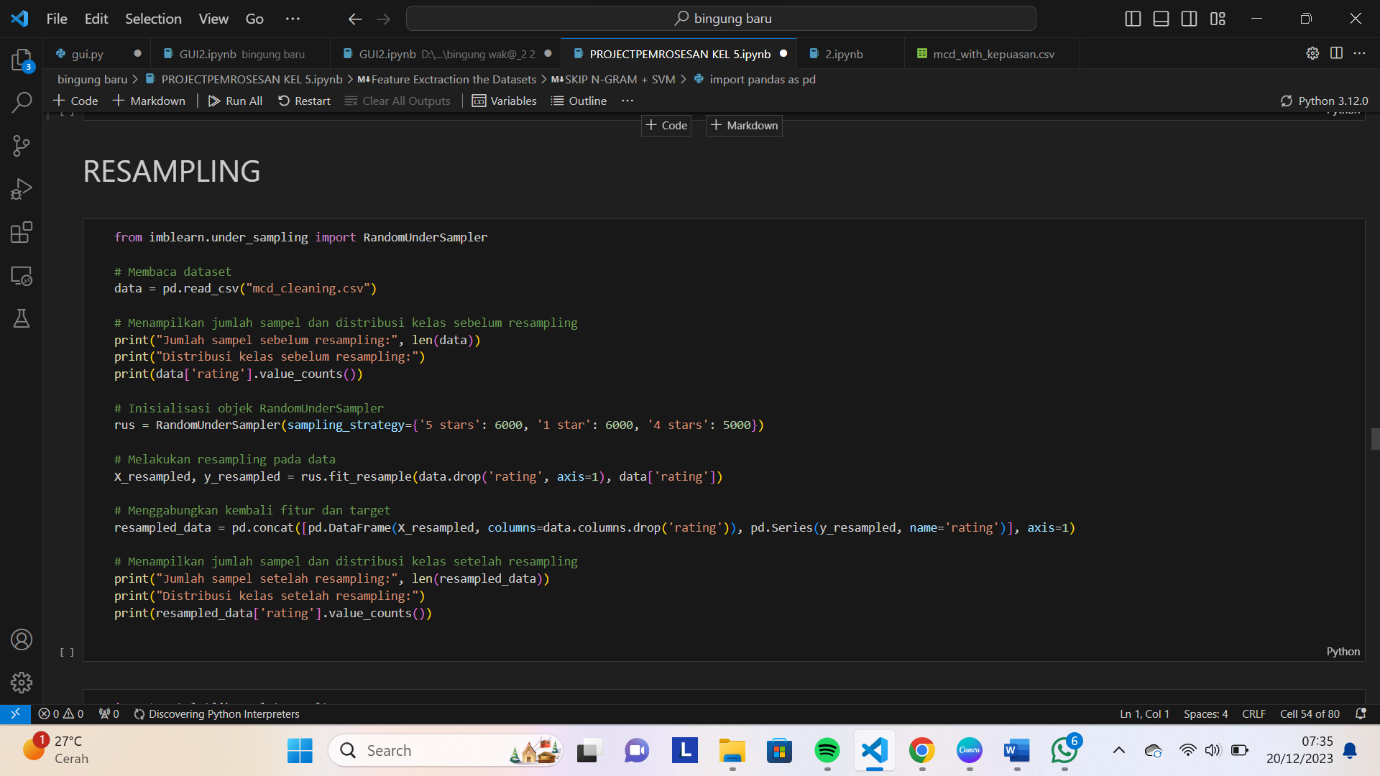


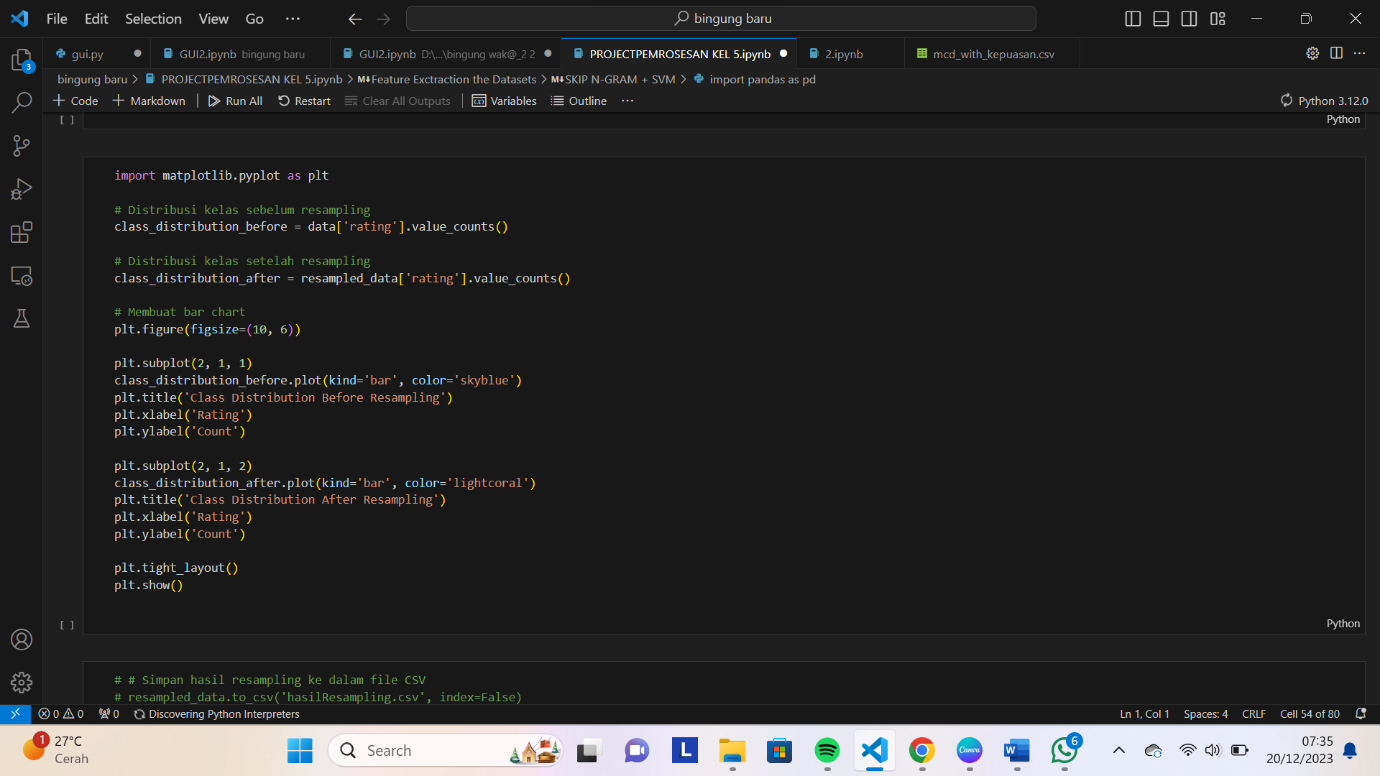
[11]. Menghitung akurasi pada ekstraksi fitur dengan menggunakan TF-IDF.



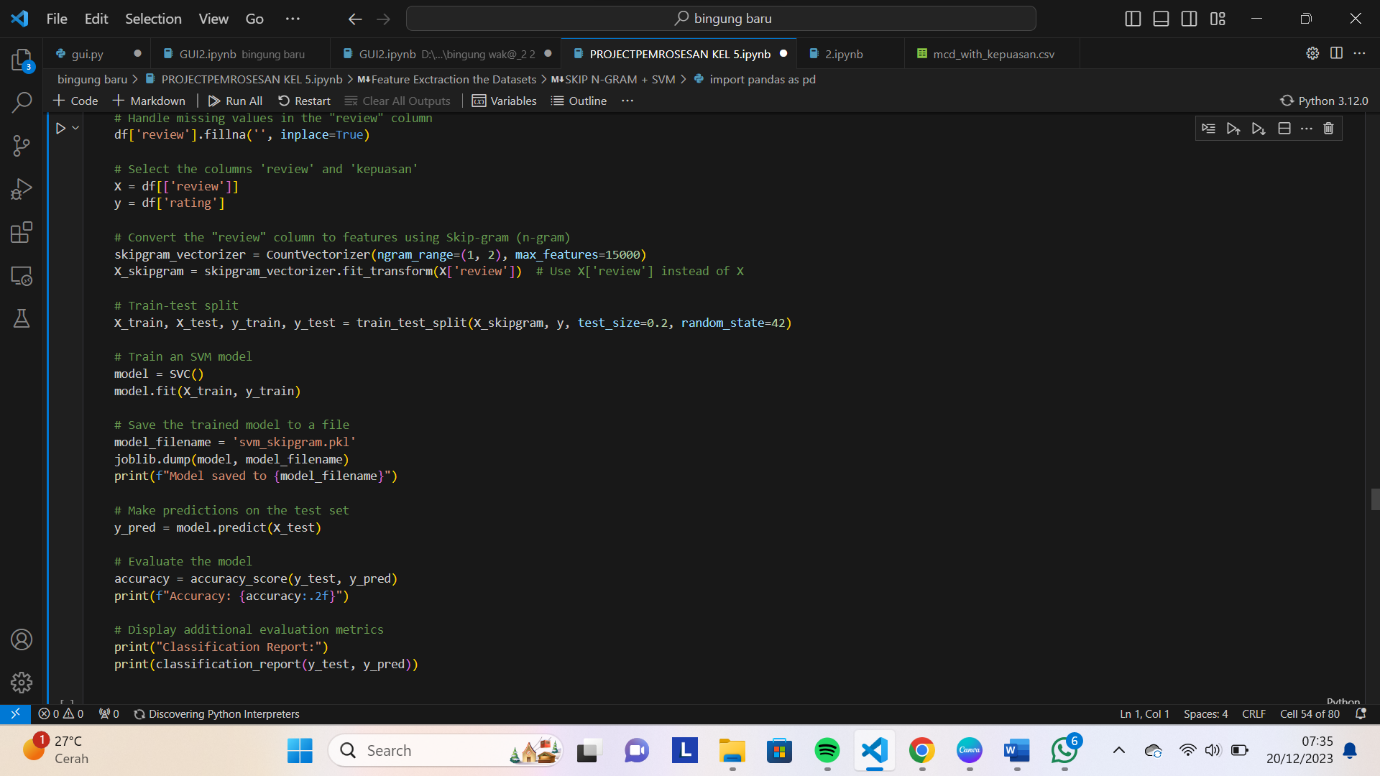


[12]. Melakukan resampling yang bertujuan untuk menyeimbangkan agar model tidak condong ke class mayoritas dan memperhatikan class minoritas.

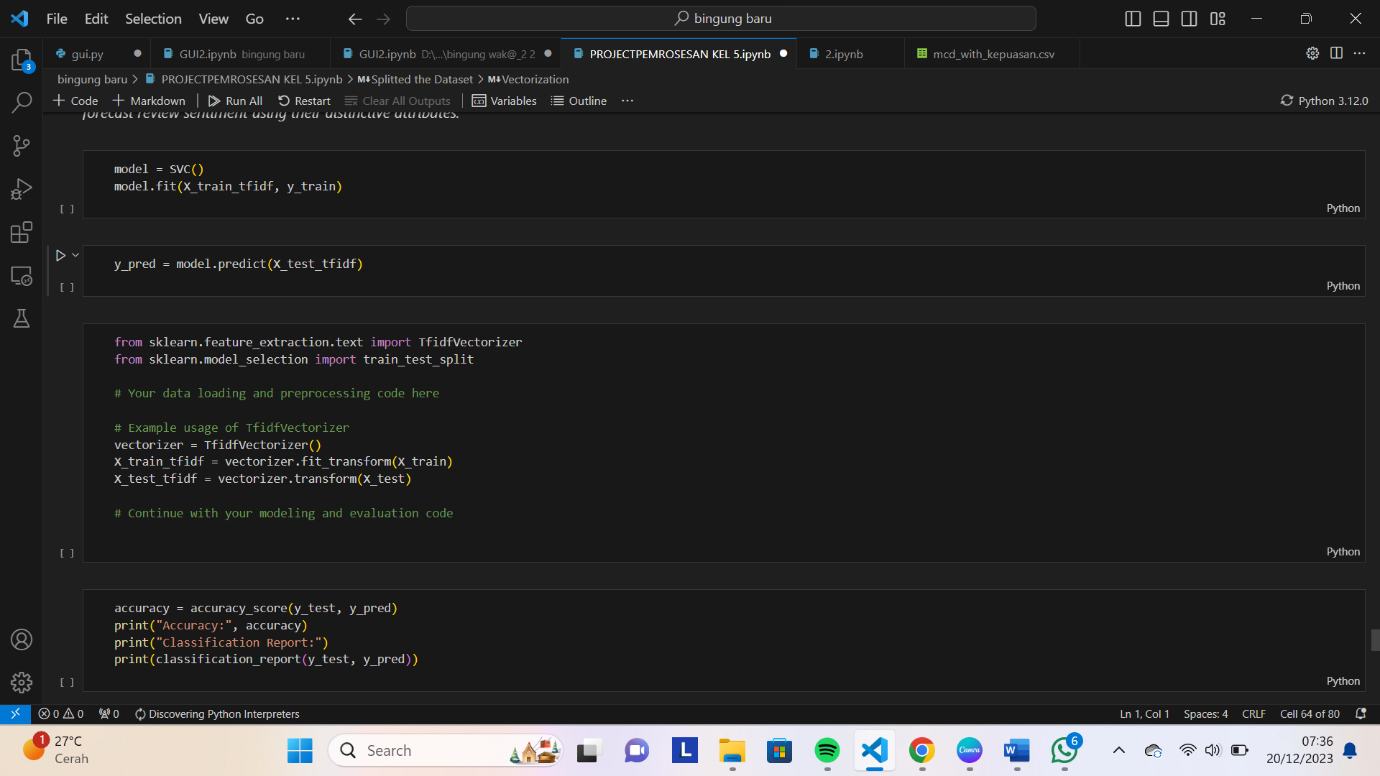


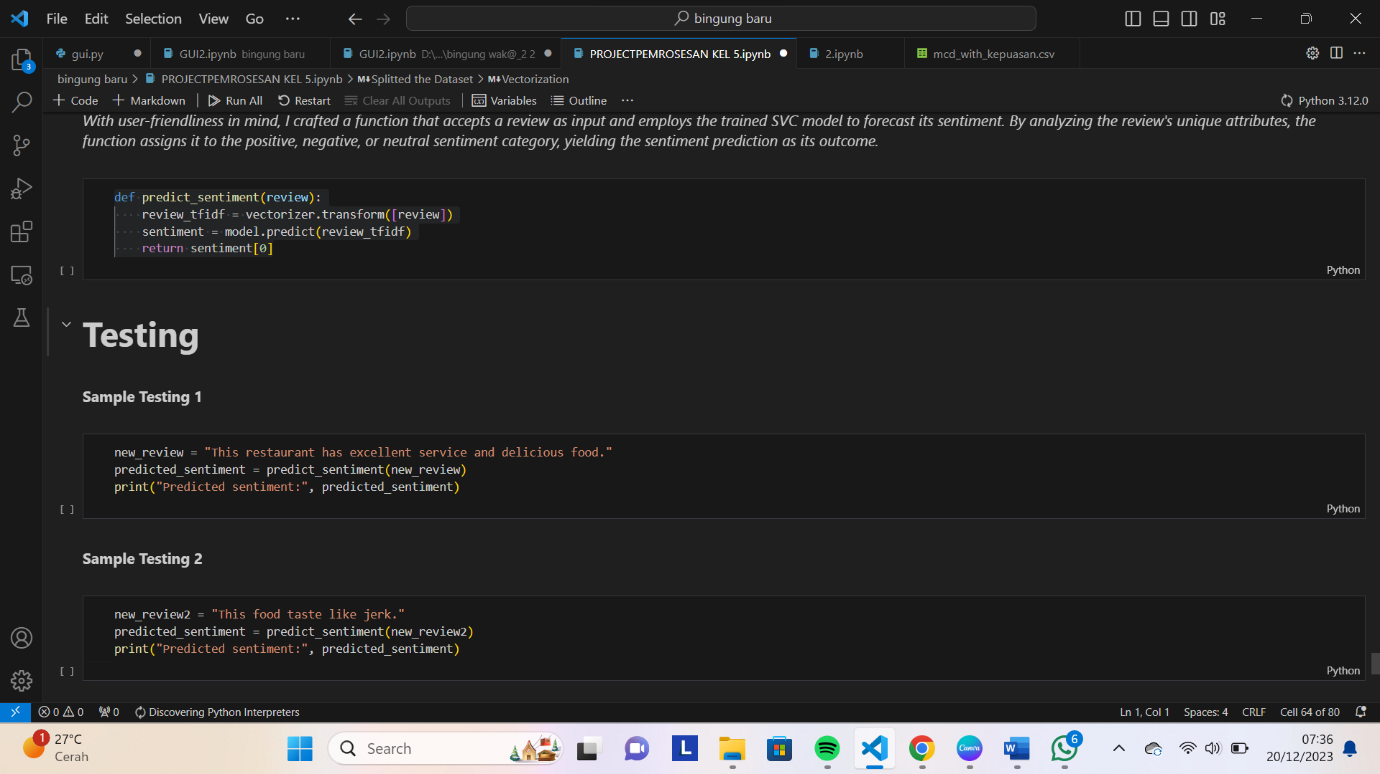


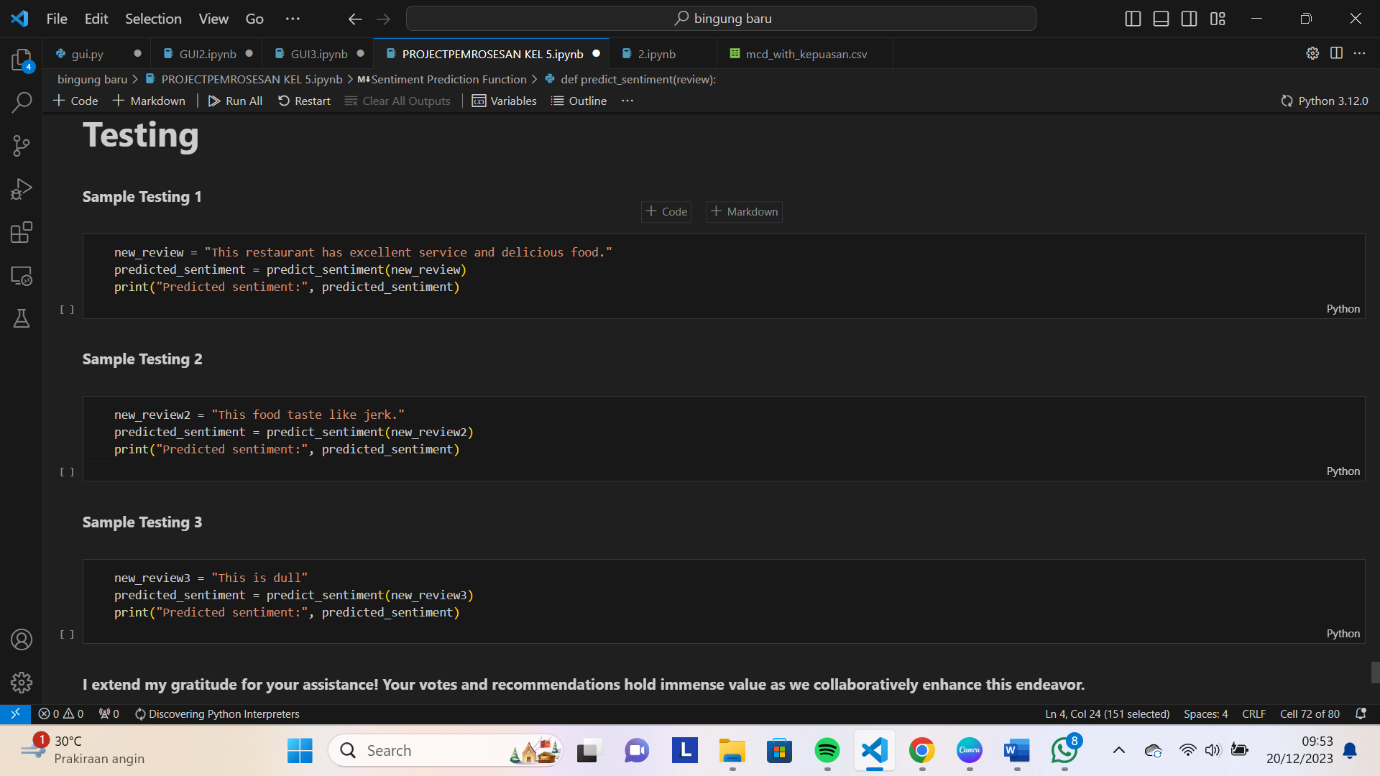
[13]. Menghitung akurasi pada metode Skip N-Gram yang sudah dilakukan resampling.



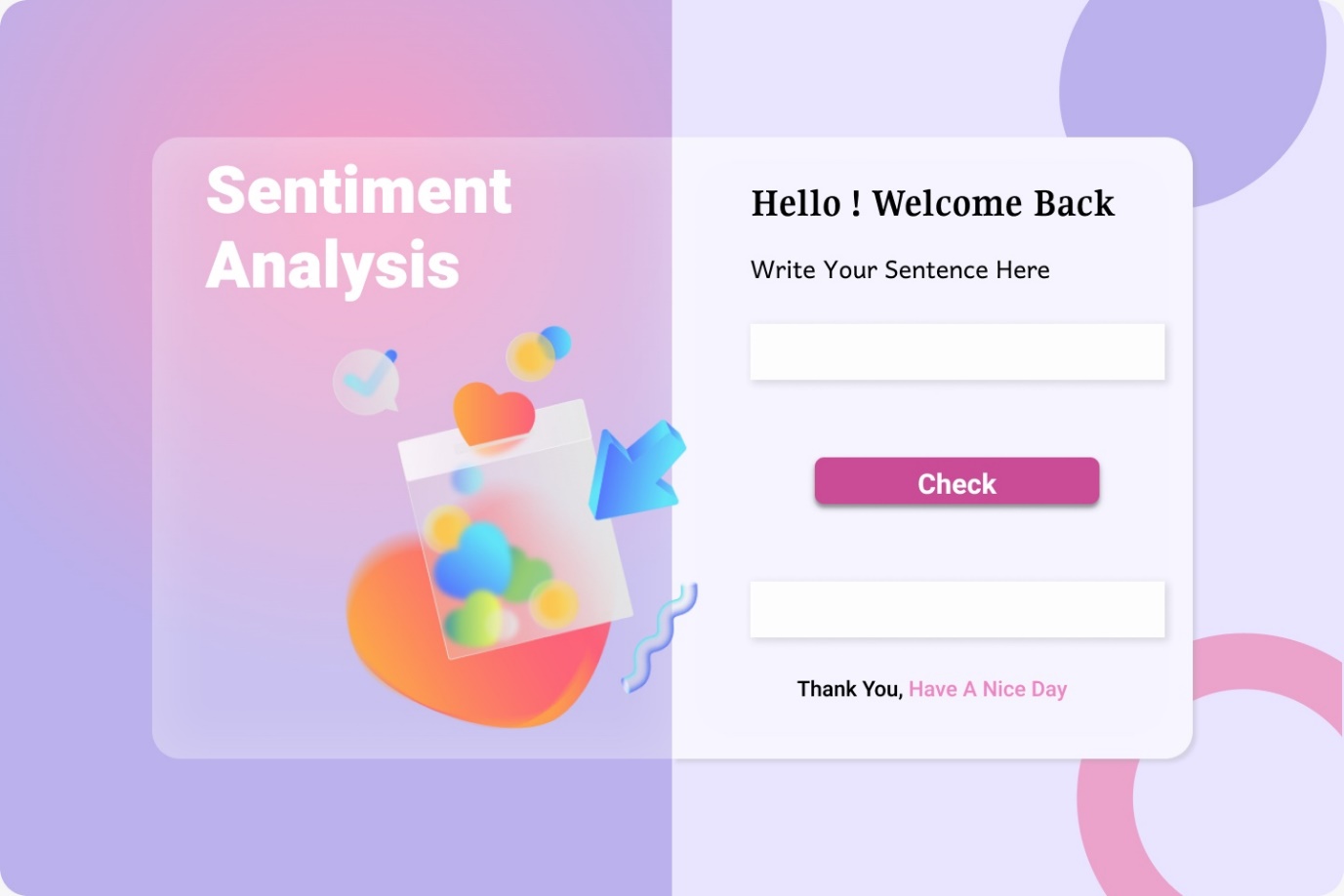
[14]. Melakukan split data dengan membagi menjadi dua yaitu train dan tes untuk menguji model. Setelah itu dilakukan testing pada model apakah model tersebut sudah dapat memprediksi sentiment melalui review yang diberikan. Hasilnya sudah dapat melakukan prediksi pada review yang diberikan.



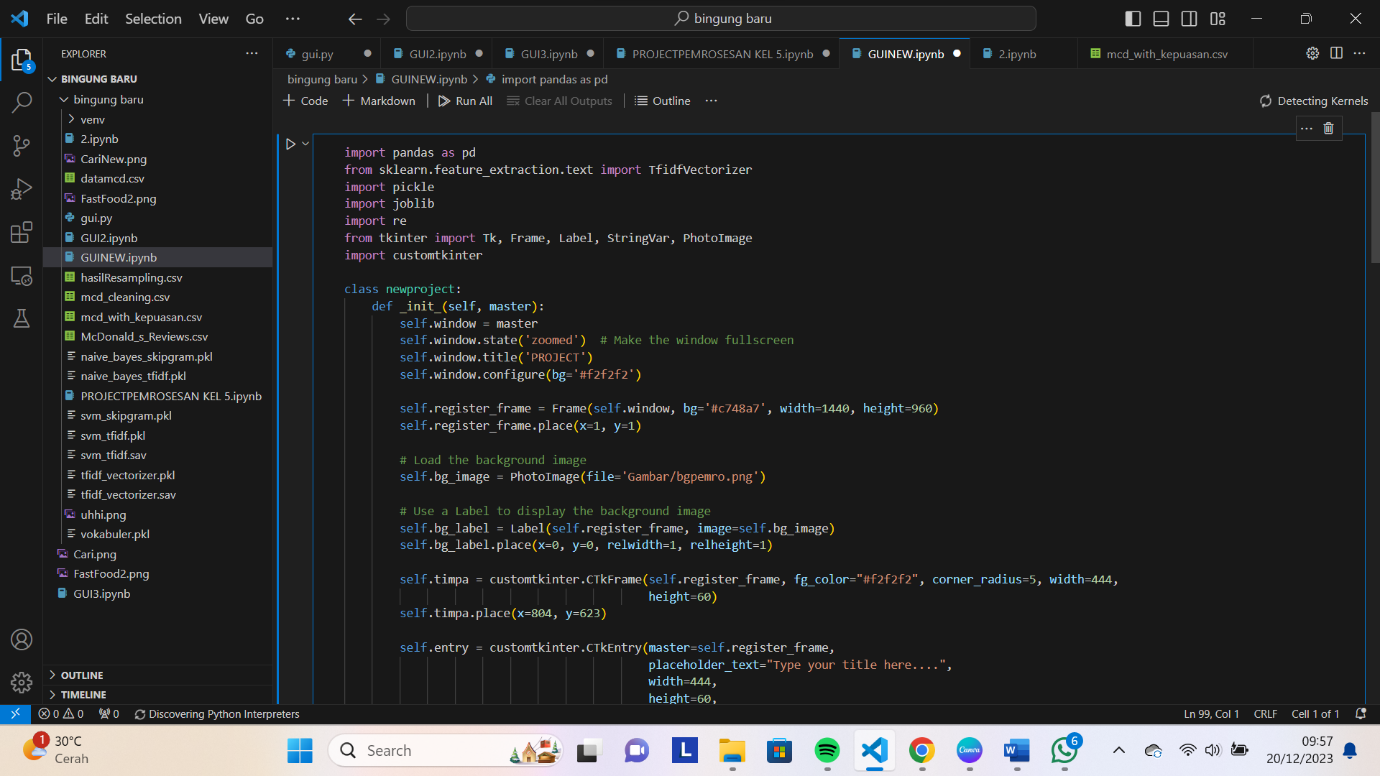




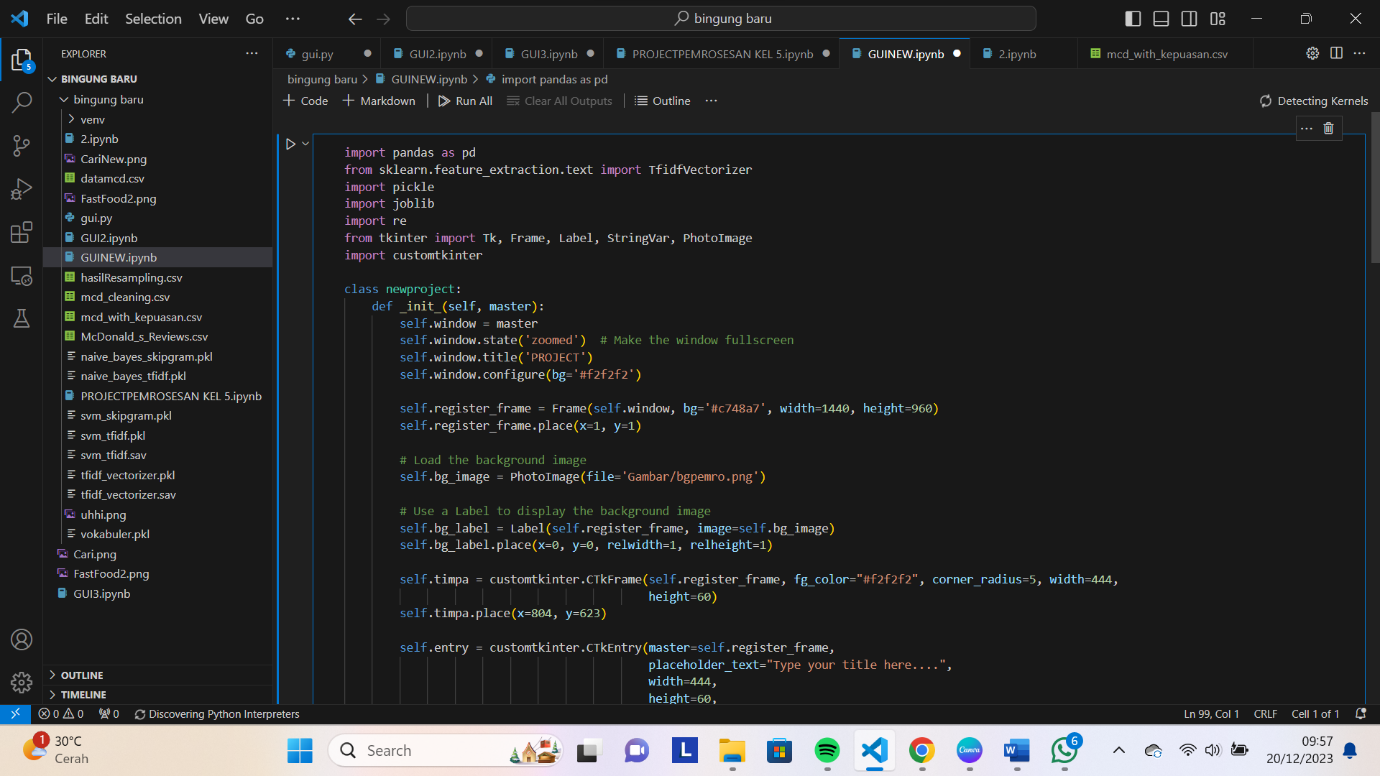
**GUI**

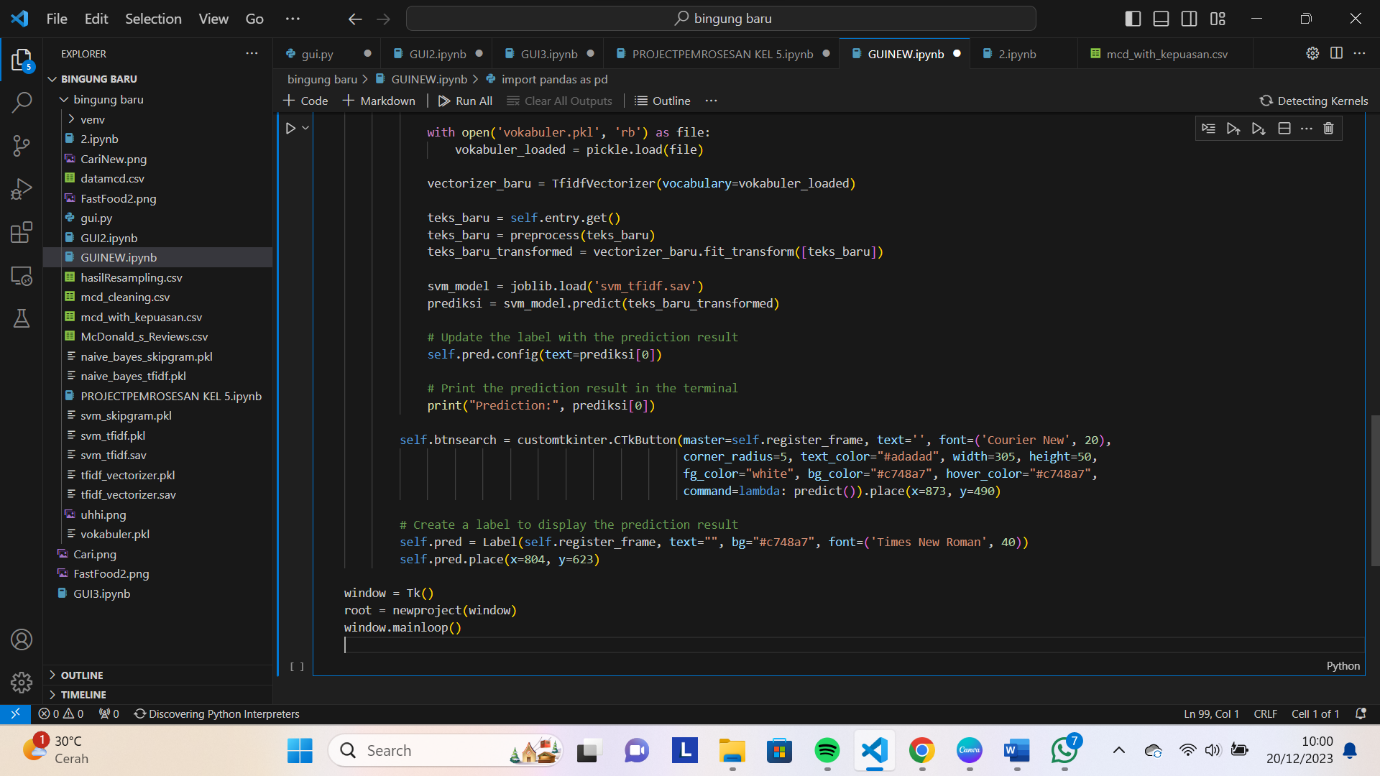


[1]. Library yang digunakan untuk membuat GUI.



[2]. Diawali dengan user yang menginputkan review lalu nantinya review tersebut akan melalui tahap preprocessing setelah itu di ekstraksi fitur dan di cocokan oleh model yang sudah disimpan sehingga menghasilkan output hasil prediksi.

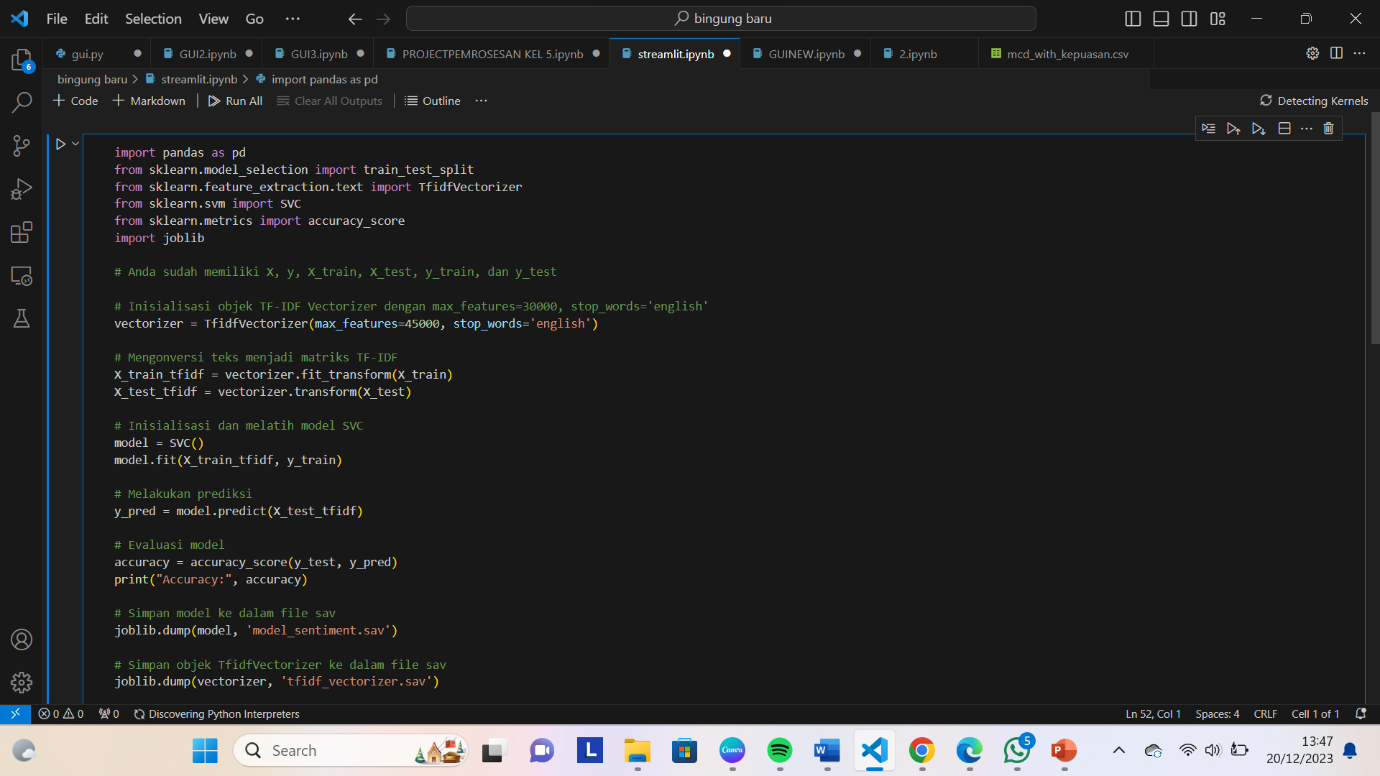


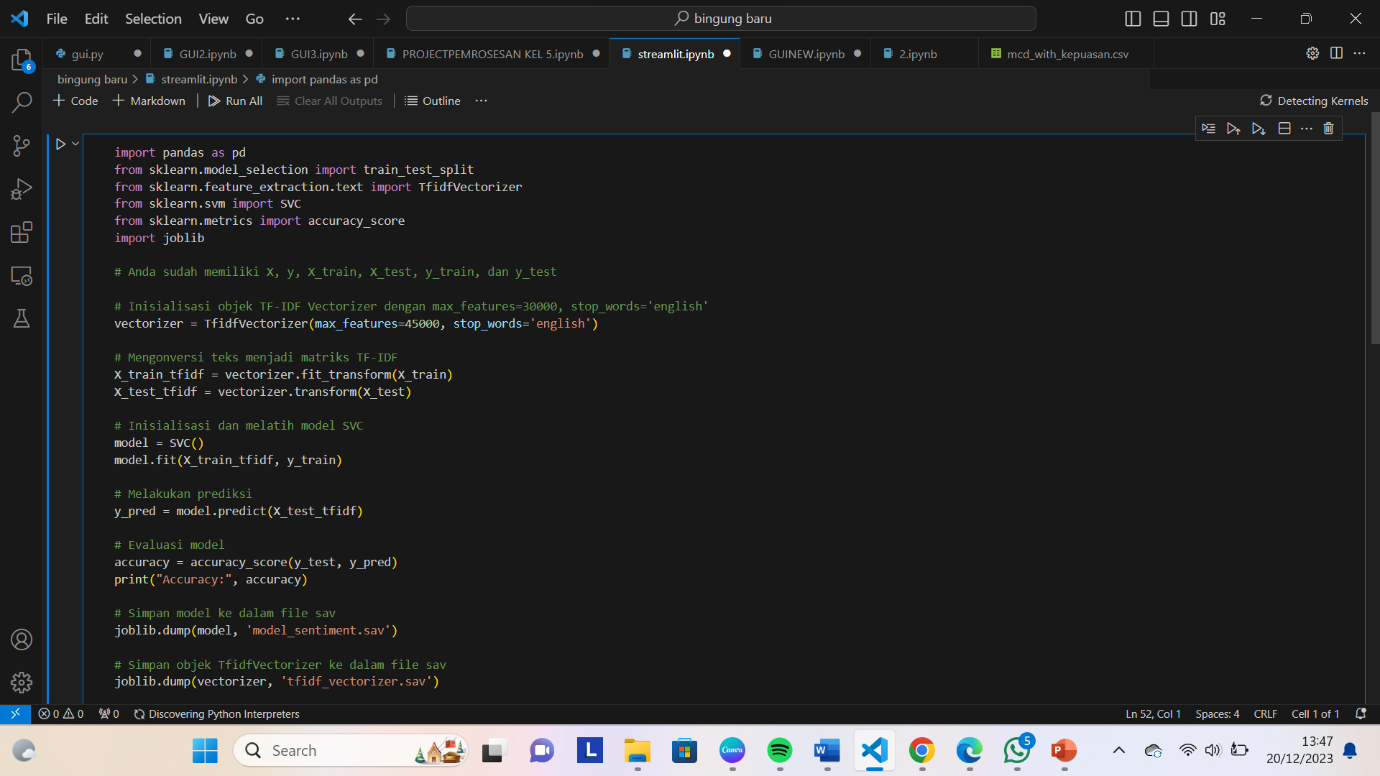


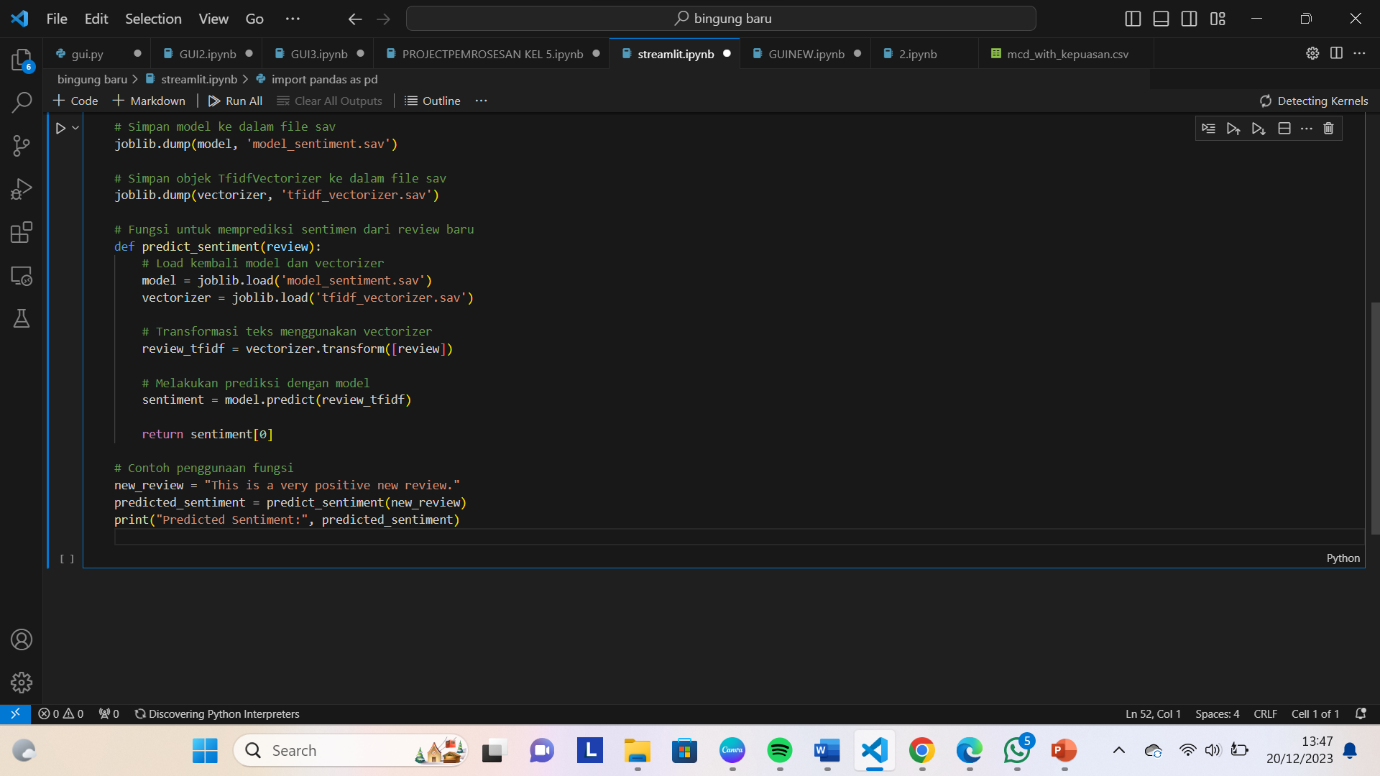
**STREAMLIT**



**[1]. Library yang digunakan dalam membuat streamlit.**

****

****

****