BAB III
METODE KERJA

3.1 Waktu dan Tempat

Pelaksanaan program kegiatan Merdeka Belajar – Kampus Merdeka (MBKM) program Studi Independen Bersertifikat (SIB) ini dilakukan pada bulan 7 Februari 2022 hingga 29 Juli 2022 selama kurang lebih 5 bulan yang materi pembelajarannya diberikan langsung oleh PT. Artifisial Intelegensia Indonesia dan dilakukan secara daring.

3.2 Alat dan Bahan (jika ada)

3.2.1 Alat

Alat yang digunakan untuk penyusunan aplikasi berbasis website ini berupa *hardware* dan *software,* antara lain:

a. *Hardware*

   Laptop atau *Personal Computer (PC)*

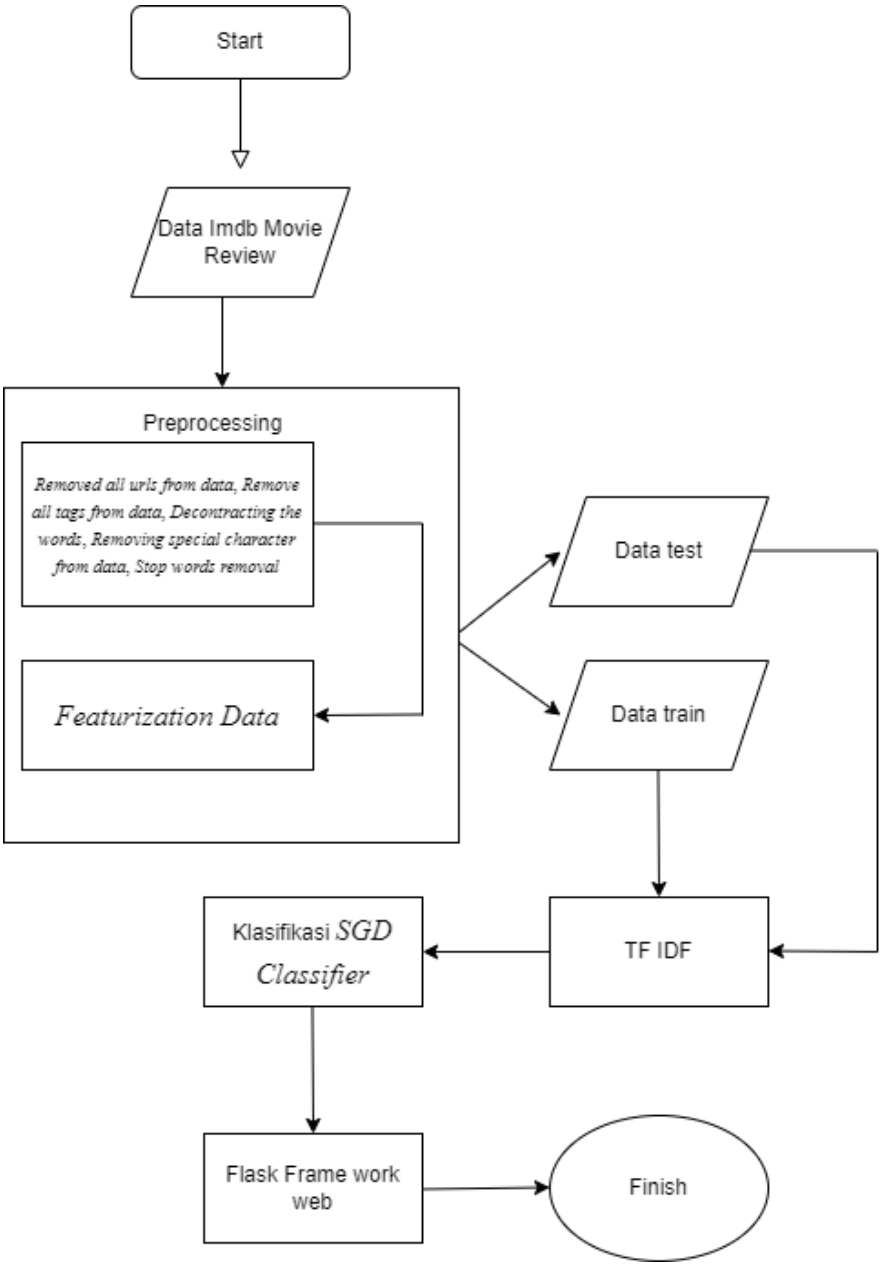
b. *Software*

   Bahasa pemrograman *Python*

3.2.2 Bahan

Bahan yang digunakan untuk penyusunan aplikasi berbasis *website* ini berupa dataset yang di-*download* dari sumber *kaggle*.

3.3 Metode dan Proses Kerja



Gambar 3.1 Flowchart

1. *Preprocessing data* merupakan langkah penting untuk tugas pemrosesan bahasa alami (NLP). Ini mengubah teks menjadi bentuk yang lebih mudah dicerna sehingga algoritme pembelajaran mesin dapat bekerja lebih baik. Dalam hal ini kami melakukan jenis preprocessing berikut.

   Removed all urls from data // Menghapus semua url dari data.

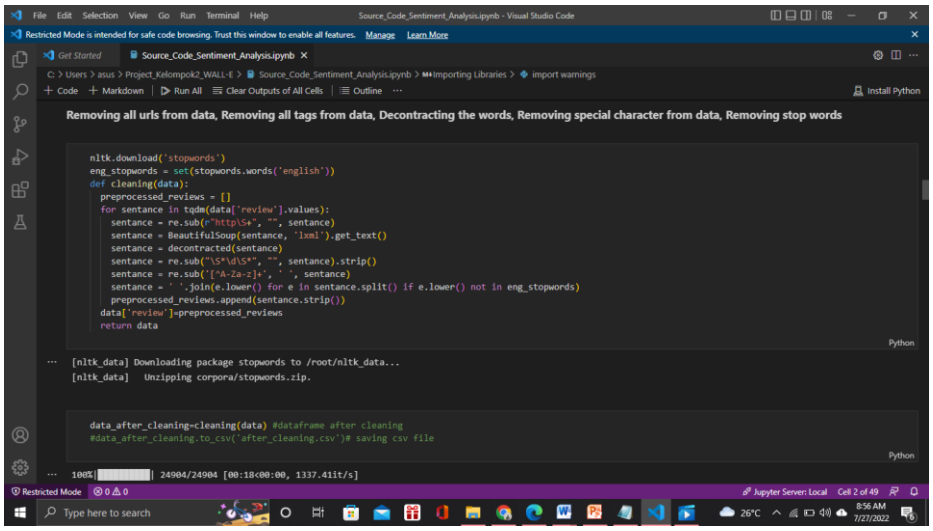
   Remove all tags from data // Hapus semua tag dari data

   Decontracting the words // Mengurai kata-kata

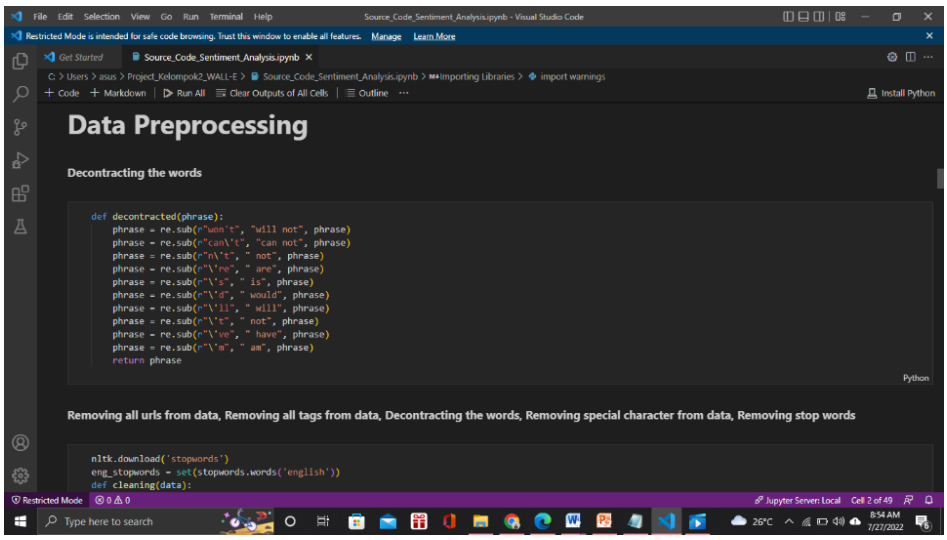
   Removing special character from data // Menghapus karakter khusus dari data

   Stop words removal // Hapus kata berhnti

   Ekstraksi fitur : *TF-IDF Vectorizer*

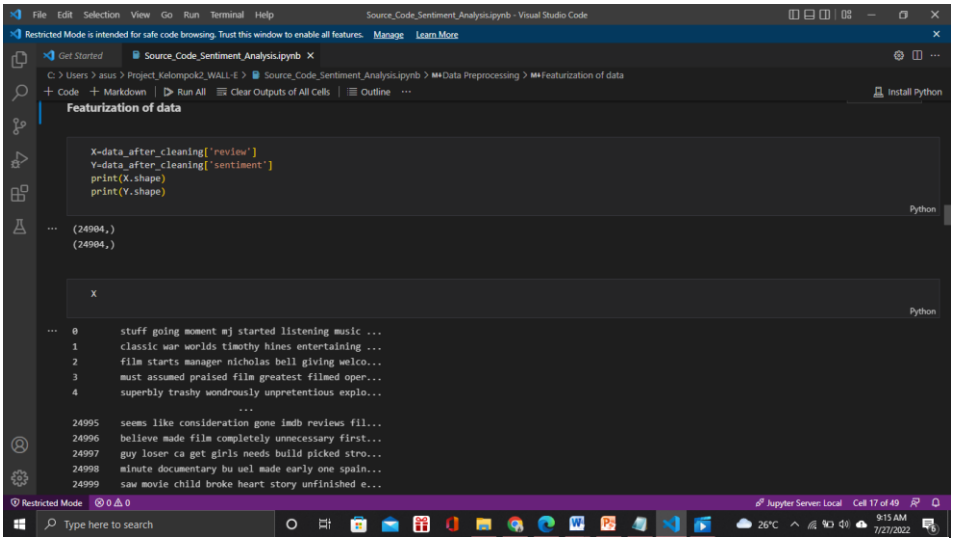


Gambar 3. 2 Source Code Data Processing



Gambar 3.3 *Source Code Removed all urls from data, Remove all tags from data, Decontracting the words, Removing special character from data, Stop words removal.*

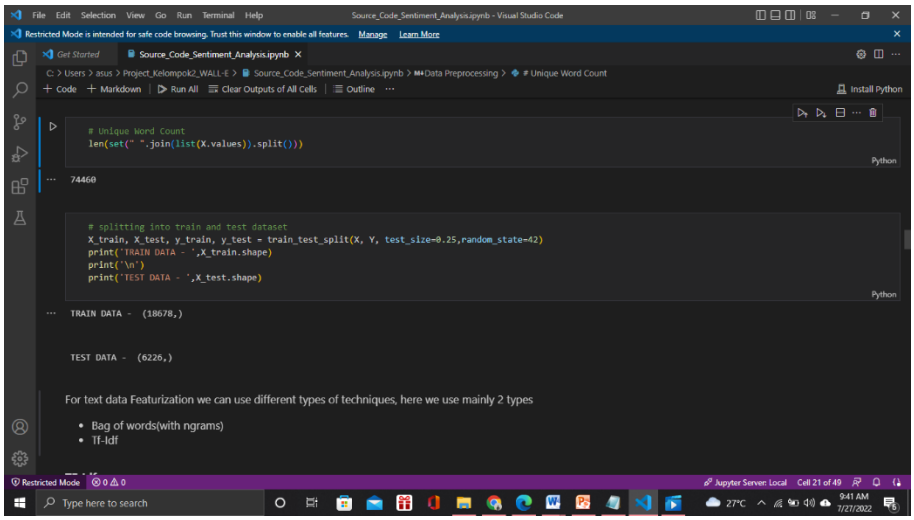
Setelah *data processing* lalu masuk ke *Feturization Data* atau Memisahkan antara *X.Review dan Y.Sentiment* untuk mengetahui skor yang di peroleh <5 menghasilkan skor sentimen 0 berarti *Negative*, dan peringkat >=7 memiliki skor sentimen 1 Berarti *Positive*.



Gambar 3.4 *Source Code Featurization Data*

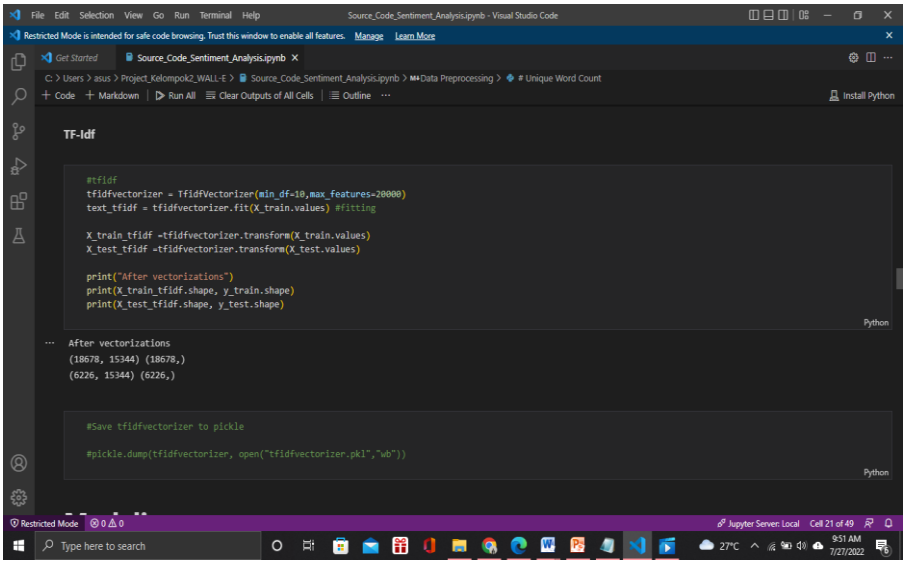
Lalu setelah itu masuk ke *Train Data* dan *Test Data* yaitu langkah atau hasil dari *Machine Learning* Mempelajari data, Jadi disini Train Data yang di

peroleh 18678 atau 75% dari teks dan untuk *Test Data* yang di hasilkan Hanya 25% dari Teks yang di proses oleh mesin.



Gambar 3.5 *Source Code Train Data and Test Data*

Tahap selanjutnya mengekstraksi data yang sudah di proses oleh mesin learning agar bisa di klasifikasi lagi oleh *Machine Learning* di tahap *Modelling*



Gambar 3.6 Source Code TF-idf

2. Tahap *Modelling* yaitu *Dataset Imdb ditrain* menggunakan beberapa algoritma klasifikasi machine learning untuk memprediksi akurasi ketepatan dari sentiment analisis, klasifikasi *machine learning* seperti :

   SVM : 89%

Naive Bayes : 76%

SGD Classifier : 89%

Ridge Classifier : 88%

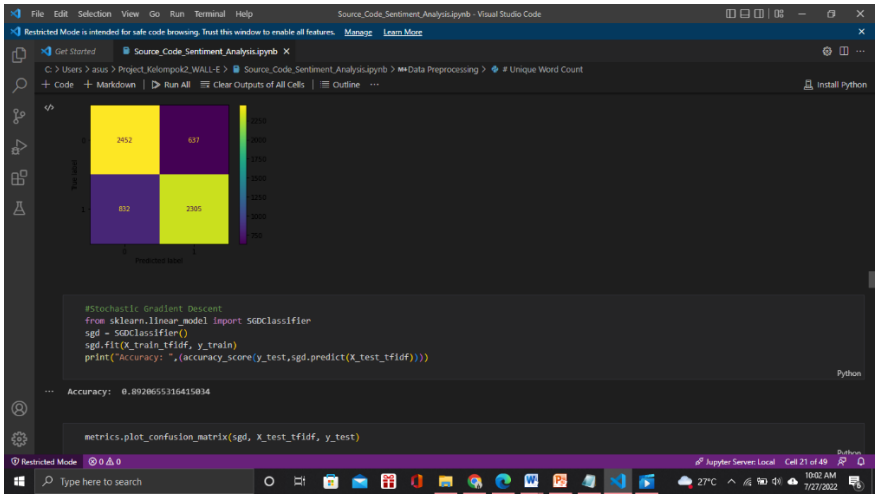
Decision Tree : 71%

Logistic Regression : 88%

Random Forest : 85%

KNN : 81%

Setelah *Ditrain* di tahap *Modelling* sampai menemukan persentasi terbesar atau hampir sempurna dan disini ada 2 Algoritma klasifikasi dengan presentase terbesar yaitu ada Support Vector Machine/ SVM dan Algoritma yang digunakan sebagai model aplikasinya adalah *SGDClassifier* dengan akurasi prediksi 89%.



Gambar 3. 1 *Source Algoritma Machine Learning SGD Classifier*