

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
SENTIMEN ANALISIS PRODUK BERDASARKAN
REVIEW PADA MARKETPLACE TOKOPEDIA**



NIKEN SAPTA WULAN 123190050

YESSIKA NUR AGNI MILLANIA 123190139

**PROGRAM STUDI INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
YOGYAKARTA
2021**

1. PENDAHULUAN

Data science adalah ilmu gabungan dari matematika, statistika, dan ilmu komputer. Data science merupakan metode analisis data dari suatu himpunan data yang sangat banyak atau biasa disebut *Big Data*, dengan pengaplikasian algoritma tertentu untuk memperoleh suatu informasi yang digunakan untuk menjawab permasalahan tertentu.

Dewasa ini banyak bisnis memanfaatkan data science untuk membantu mengolah data. Para pelaku bisnis memerlukan bantuan data science untuk mengetahui performa bisnis mereka. Data yang tak kalah penting untuk diolah adalah data rating pelanggan. Data rating pelanggan merupakan penilaian langsung kepada suatu produk yang telah mereka beli..

Didasari alasan tersebut, kami mencoba membuat program untuk mengolah data rating. Data rating ini akan diolah dan untuk memberi label positif, negatif, dan netral pada varian produk yang dijual pada sebuah toko di marketplace Tokopedia berdasarkan review yang diberikan oleh pelanggan. Informasi yang didapat nantinya digunakan untuk mengetahui perbandingan dari review yang telah diolah. Sehingga kita dapat mengetahui apakah produk tersebut memiliki penilaian yang baik atau tidak.

2. METODE

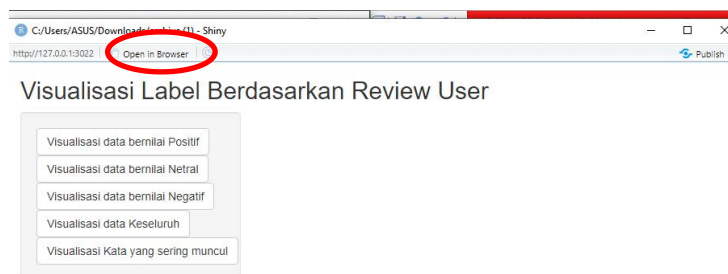
Pada proyek kali ini, metode yang kami gunakan adalah text preprocessing dengan natural language processing. Text preprocessing merupakan tahap awal dari natural language processing untuk dokumen yang berupa teks. Text preprocessing mengolah data yang tidak terstruktur menjadi data yang siap diolah. Sedangkan Natural Language Processing (NLP) adalah salah satu cabang dari AI yang berfokus pada pengolahan bahasa natural. Bahasa natural adalah bahasa yang digunakan oleh manusia untuk berkomunikasi dengan sesama. Maka dari itu, agar dapat dimengerti oleh komputer, bahasa natural harus diolah terlebih dahulu dengan metode NLP.

Pengerjaan dilakukan dengan membaca file data terlebih dahulu. Dikarenakan data yang dipakai sudah tidy sehingga dapat langsung digunakan untuk menganalisis label dari data tersebut. Data yang sudah di import selanjutnya akan dipakai bagian Review untuk dibuat menjadi Corpus lalu diubah wujudnya kedalam matriks untuk selanjutnya dibuat menjadi dataframe. Data review nantinya akan digunakan untuk mengetahui seberapa sering kemunculan dari suatu kata.

Selanjutnya data akan dipisah berdasarkan sentiment nya. Terdapat Tiga jenis sentiment dalam data ini yaitu sentiment positif, netral, dan negatif. Sentiment - sentiment ini akan di visualisasikan sehingga kita dapat mengetahui sentiment mana yang paling dominan dalam produk ini.

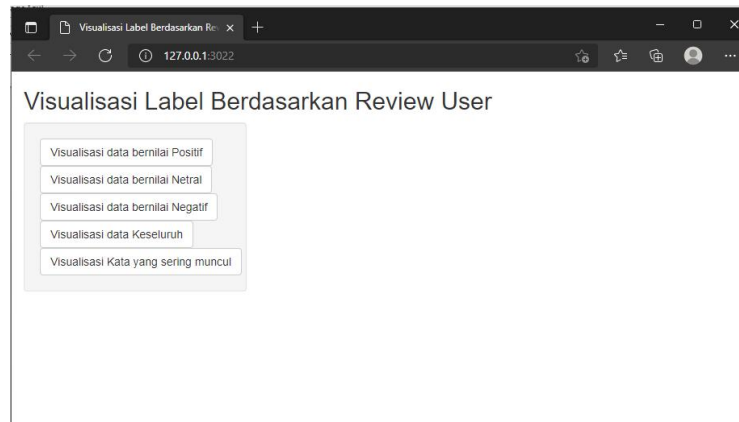
3. HASIL DAN PEMBAHASAN

Hasil running program akan seperti ini.



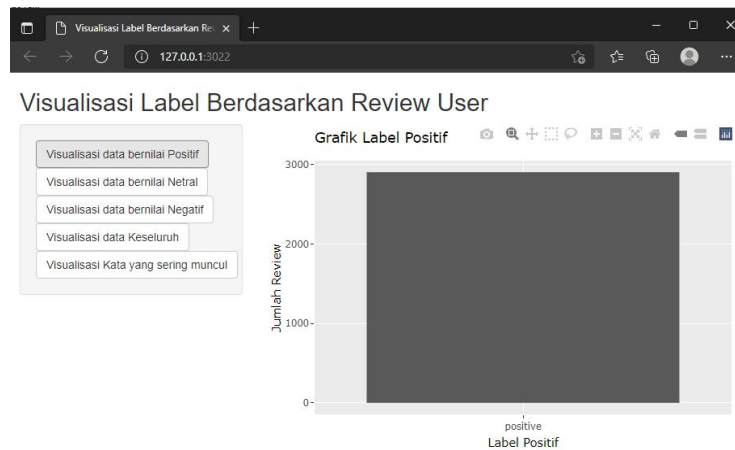
Gambar 3.1 Tampilan R Studio Shiny

Kemudian kita bisa masuk ke tampilan web dengan cara, klik pada “Open in Browser” dan akan muncul tampilan sebagai berikut.



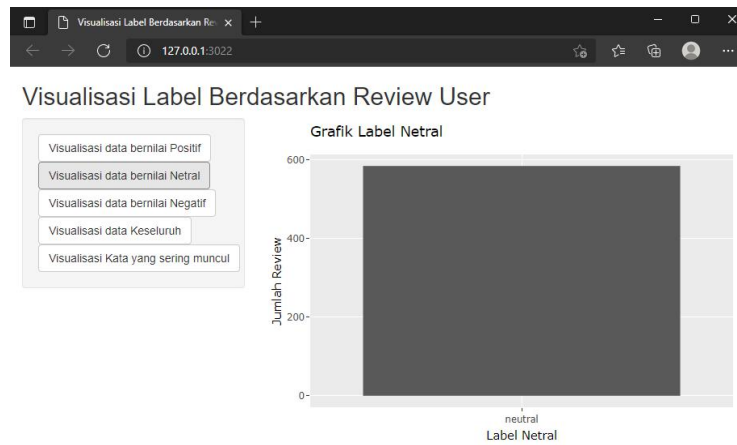
Gambar 3.2 Tampilan pada Browser

Pada menu “Visualisasi data bernilai Positif” akan mem-*filter* dan menampilkan jumlah sentiment yang bernilai positif dari database.



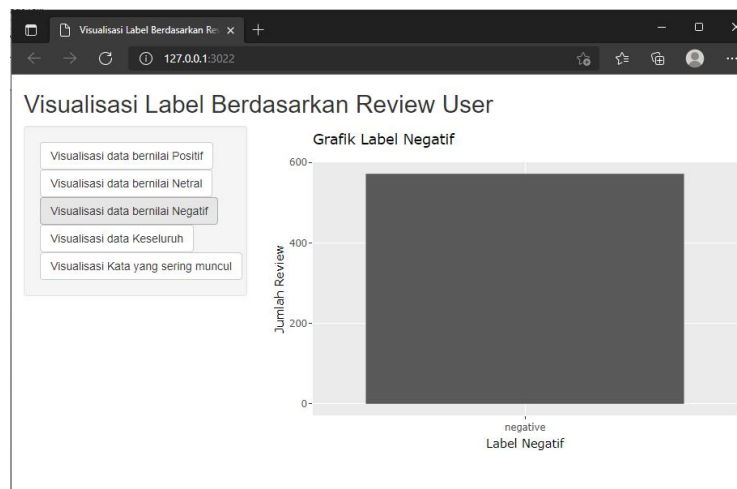
Gambar 3.3 Grafik Label Positif

Pada menu “Visualisasi data bernilai Netral” akan mem-*filter* dan menampilkan jumlah sentiment yang bernilai netral dari database.



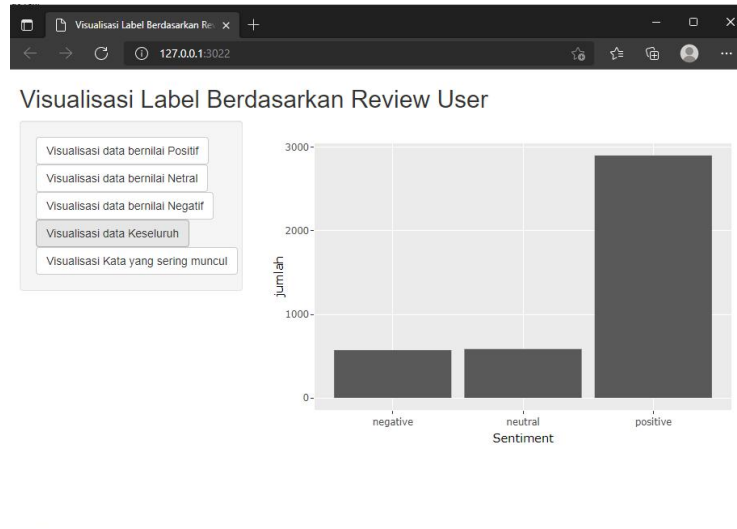
Gambar 3.4 Grafik Label Netral

Pada menu “Visualisasi data bernilai Negatif” akan mem-*filter* dan menampilkan jumlah sentiment yang bernilai negatif dari database.



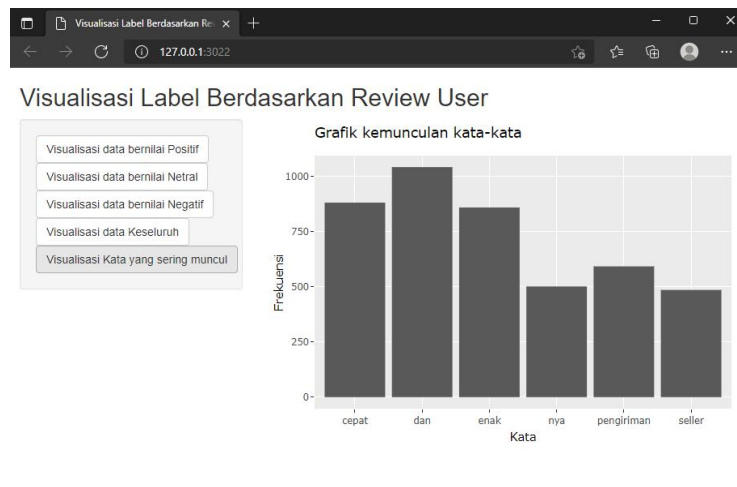
Gambar 3.5 Grafik Label Negatif

Pada menu “Visualisasi data Keseluruhan” akan mem-*filter* dan menampilkan jumlah dari masing-masing sentiment yang bernilai positif, netral, dan negatif.



Gambar 3.6 Grafik Label Keseluruhan

Pada menu “Visualisasi Kata yang sering muncul” akan mem-*filter* kata-kata yang sering muncul, kemudian menampilkan frekuensi jumlah kemunculan kata-kata tersebut.



Gambar 3.7 Grafik Kemunculan Kata

4. KESIMPULAN

Dengan menggunakan text preprocessing kita dapat mengolah data dengan cepat. Data - data yang telah diolah inilah yang akan menjadi pertimbangan kita dalam menilai sebuah produk. Produk yang memiliki ribuan review akan dilakukan penglabelan sehingga setiap review akan memiliki nilainya masing-masing. Nilai yang dimaksud adalah review dengan nilai positif, negatif, dan netral. Dengan adanya label ini kita dapat melakukan pengelompokan sentiment berdasarkan labelnya untuk selanjutnya di visualisasikan agar mempermudah memahami hasil akhir dari data yang diperoleh. Dari visualisasi inilah kita akan mengetahui label yang mana yang paling dominan dalam produk, untuk nantinya dapat ditarik kesimpulan apakah produk termasuk kedalam kategori bagus atau buruk.