**LAPORAN PROYEK AKHIR**

**PRAKTIKUM DATA SCIENCE**

**Analisis Sentimen Review Bisnis Pakaian Wanita Menggunakan Metode Naïve Bayes**



|  |  |
| --- | --- |
| NUREVA FUSFA DEWI | 123190027 |
| M. SAYYID QUTHUB BB | 123190016 |

**PROGRAM STUDI INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK INDUSTRI**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” YOGYAKARTA**

**2021**

# **PENDAHULUAN**

Pembelian dan penjualan secara online menjadi suatu perkembangan dalam dunia perdagangan.Karena masyarakat dapat membeli baju atau menjual baju tanpa harus bertatap muka. Dengan banyaknya website atau e-commerce yang menyediakan produk baju ataupun lainnya dalam jual beli. Maka dari transaksi tersebut akan terekap dalam dataset.Kumpulan dataset tersebut dapat digunakan sebagai informasi atau pengetahuan. Transaksi jual beli yang sudah terekap dan tersimpan dalam database dapat diolah. Data mining adalah proses mengolah dan menggali informasi dan pengetahuan dari kumpulan dataset.

Produk pakaian wanita yang beredar di pasaran saat ini sangat beragam, Namun tidak semua pakaian wanita memiliki kualitas yang baik sesuai kebutuhan konsumen dan hal ini yang harus diperhatikan oleh para konsumen. Sebelum konsumen memutuskan untuk membeli pakaian wanita sebaiknya konsumen mengetahui dengan detail produk yang akan dibeli, hal ini dapat dipelajari dari testimoni dan opini atau hasil review dari konsumen yang sudah membeli dan menggunakan produk tersebut sebelumnya.

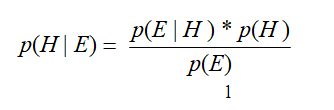
Beberapa review tentang produk pakaian wanita dapat membantu konsumen dalam mengetahui kualitas brand pakaian wanita tersebut layak atau tidak untuk digunakan .Banyaknya website untuk review menyebabkan ketersediaan data review sangat banyak. Saat ini konsumen yang menulis opini dan pengalaman secara online semakin meningkat.Review produk dapat memberikan manfaat besar bagi konsumen karena dengan membaca review, konsumen mendapat informasi yaitu dapat mengetahui kualitas suatu produk dari review konsumen lain yang membagikan pengalamannya tentang produk terkait.

Di era saat ini sentimen dapat menjadi potensi besar bagi perusahaan industri pakaian wanita yang ingin mengetahui umpan balik dari masyarakat terhadap brand pakaian wanita mereka, selain itu masyarakat yang baru ingin mencoba berbagai produk dari pakaian wanita tersebut dapat melihat review produk mana yang baik dan tidak dari klasifikasi review pakaian wanita . Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan review pengguna menjadi opini positif atau negatif.

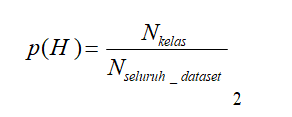
Ketersediaan data review inilah yang menjadi alasan penelitian ini menggunakan sumber data berupa review produk untuk Analisa Sentimen Terhadap Review Pakaian Wanita Menggunakan Metode Naive Bayes.

# **METODE**

Naive bayes suatu algoritma yang menerapkan nilai probabilitas dari suatu kemungkinan antara data query terhadap dataset. Di mana data query akan dihitung nilai probabilitasnnya berdasarkan dataset. Proses menghitung nilai probabilitas dengan Persamaan 1.



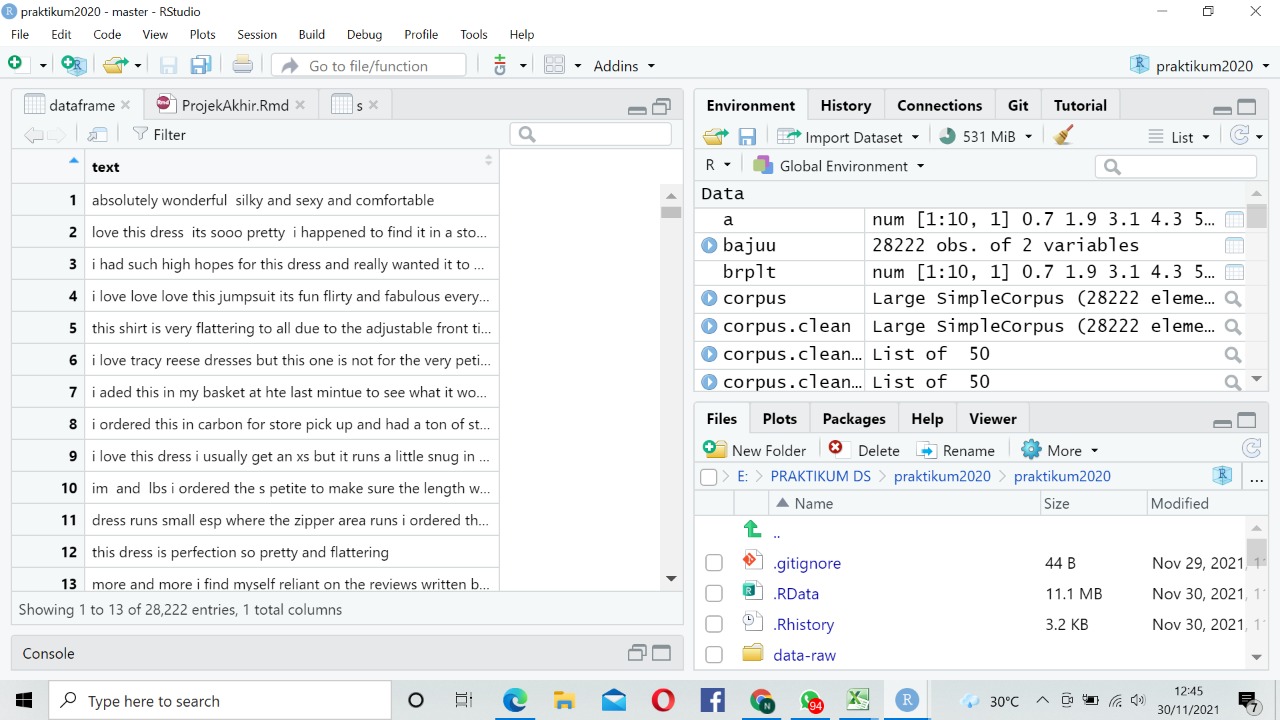
Di mana p(H|E) adalah nilai probabilitas hipotesis terhadap evidance/variabel. Dan p(E|H) adalah nilai probabilitas evidance query terhadap semua kelas hipotesis. Dan p(H) adalah nilai probabilitas semua kelas terhadap semua dataset. P(E) adalah nilai probabilitas evidance (semua variabel input). Dan untuk menghitung probabilitas hipotesis dalam Persamaan 2



Algoritma dari data mining yang juga dapat digunakan dalam jual beli online atau e-commerce adalah association Rule. Di mana algoritma association rule yang digunakan dalam penelitian Arif Ismail dan kawan-kawan adalah Apriori. Algoritma Apriori digunakan untuk strategi menyusun etalase penjualan pakaian. Etalase berjualan secara online perlu disusun secara strategis agar dapat meningkatkan penjualan[3]. Hasil dari penelitian Arif Ismail mengatakan bahwa adanya hubungan antara baju yang dibeli bersamaan terhadap penyusunan etalase[3]. Market Analisis juga salah satu algoritma data mining association rule, di mana market basket analisis dapat digunakan untuk menyusun tata letak jualan pada Minimarket Aira. Salah satu contoh penyusunan tata letak jualannya, yaitu antara menyusun snack kacang dengan minuman kopi. Ditunjukkan dari hasil market basket analisis konsumen yang membeli snack kacang maka 97,2% akan membeli snack lainnya, dan 95% akan membeli kopi.

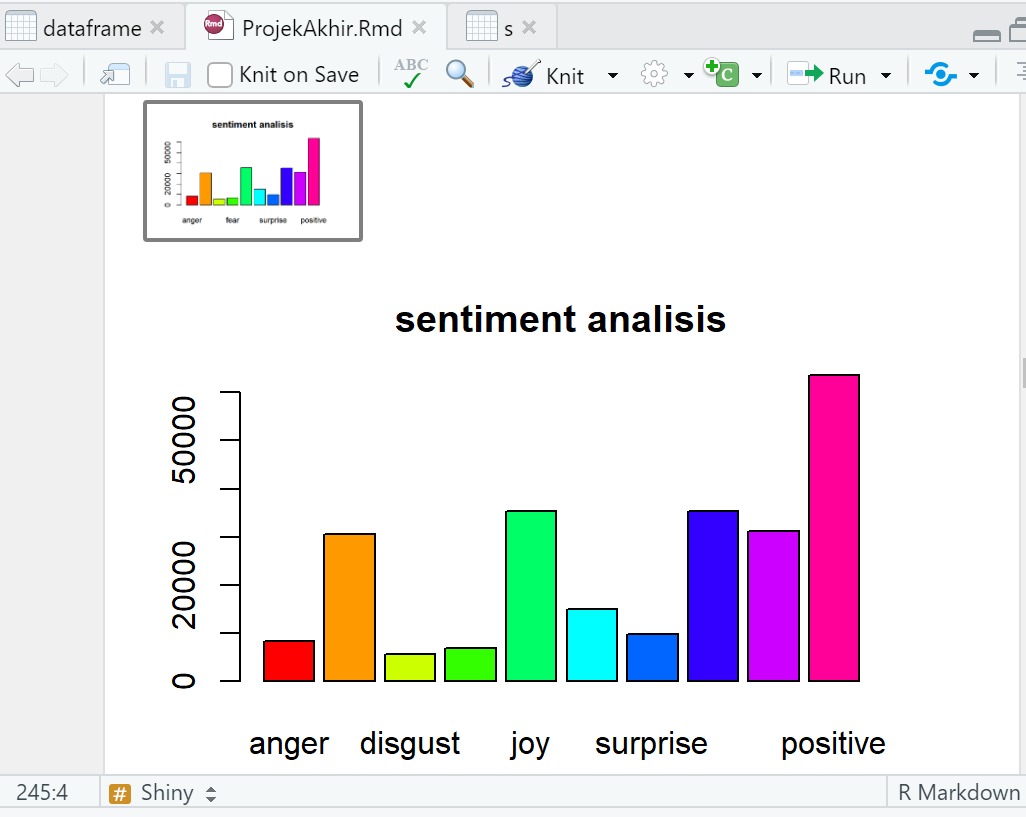
Sistem rekomendasi review pakaian wanita menggunakan metode Naive bayes mempunyai beberapa tahapan. Langkah pertama yang dilakukan dalam metode Naïve bayes yaitu menghitung nilai probabilitas P(H) dengan cara menghitung jumlah item berdasarkan class name kemudian dibagi dengan jumlah dataset keseluruhan seperti dalam Persamaan 2.Langkah kedua menghitung data item query yang diinputkan dibagi dengan nilai jumlah tiap kelas tersebut P(E|H).Kemudian mengkalikan nilai P(E|H) berdasarkan classname nya masing-masing, langkah terakhir yaitu mengkalikan hasil nilai P(E|H) dengan nilai probabilitas setiap kelas P(H).Sehingga akan didapatkan hasil dari item rekomendasi untuk konsumen masuk kedalam class name yang nilai probabilitas tertinggi. Dalam proses perhitungan naive baiyes dibutuhkan dataset,berikut link dataset yang digunakan <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>

# **HASIL DAN PEMBAHASAN**

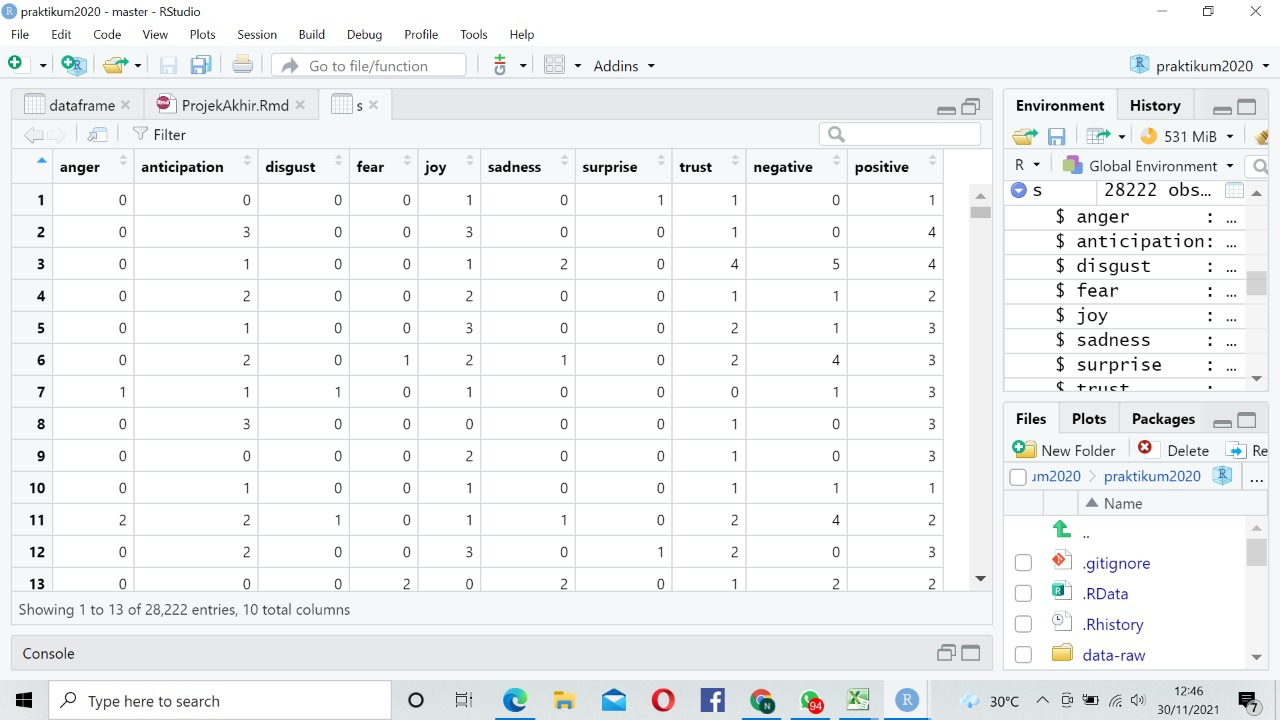


Gambar 3.1

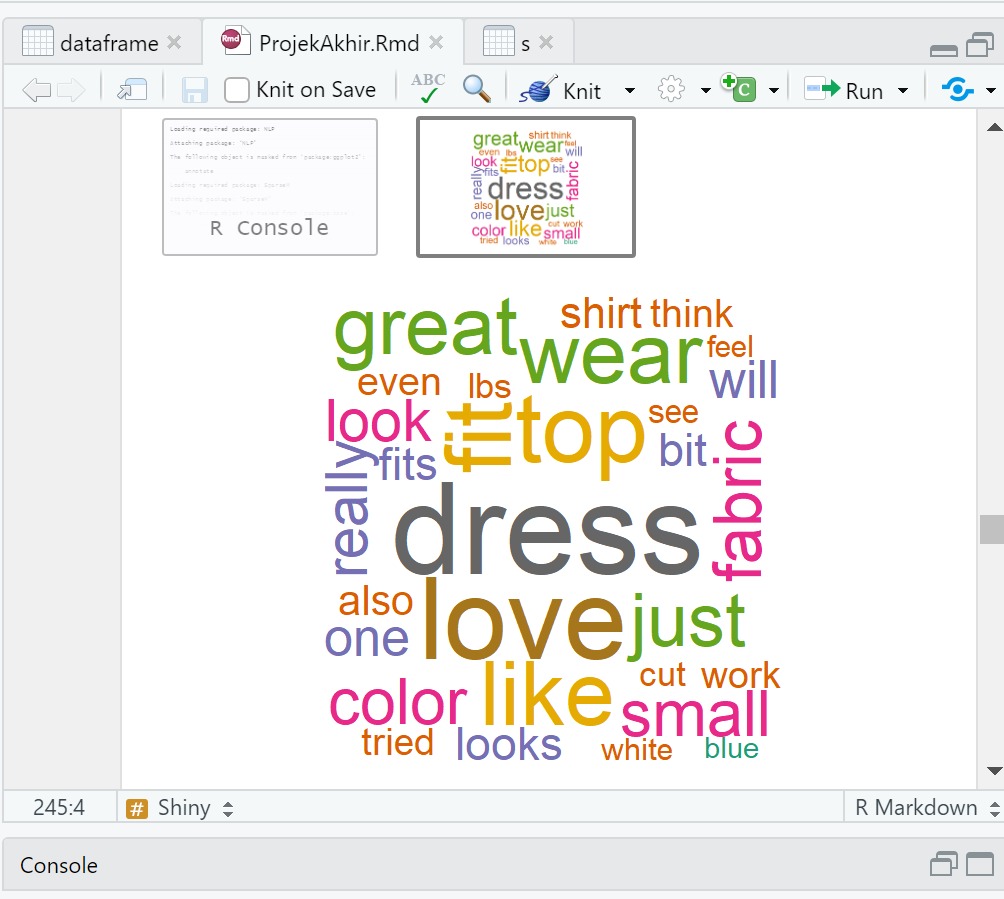
Pada gambar 3.1 merupakan halaman ketika program sudah dijalankan yaitu berupa kumpulan data yang sudah dibersihkan dan disimpan kedalam data frame dan dibuat file baru.



Gambar 3.2

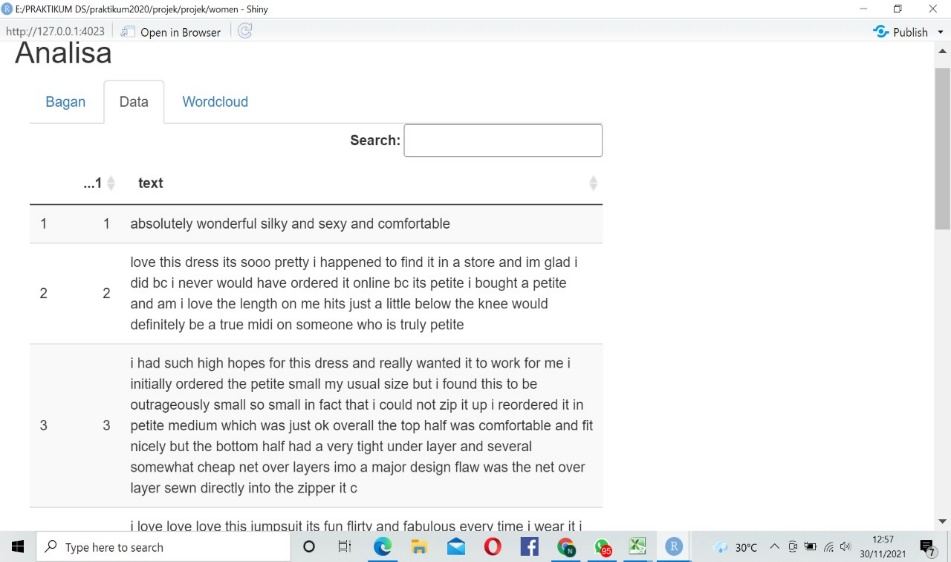
Gambar 3.3

Pada gambar 3.2 hasil dari program bagan/scarletplot,pada bagian ini memerlukan library untuk metode naïve bayes,klasifikasi data dll.Kemudian data nya dirubah dari text menjadi char dan kemudian melakukan klasifikasi data dan membuat barplot dan disimpan di variable a,keterangan dengan bentuk diagram batang menjelaskan presentasi positive,negatif dll secara sederhana.Dan pada gambar 3.3 merupakan data analisis sentiment tersebut ada positif,negative dll.

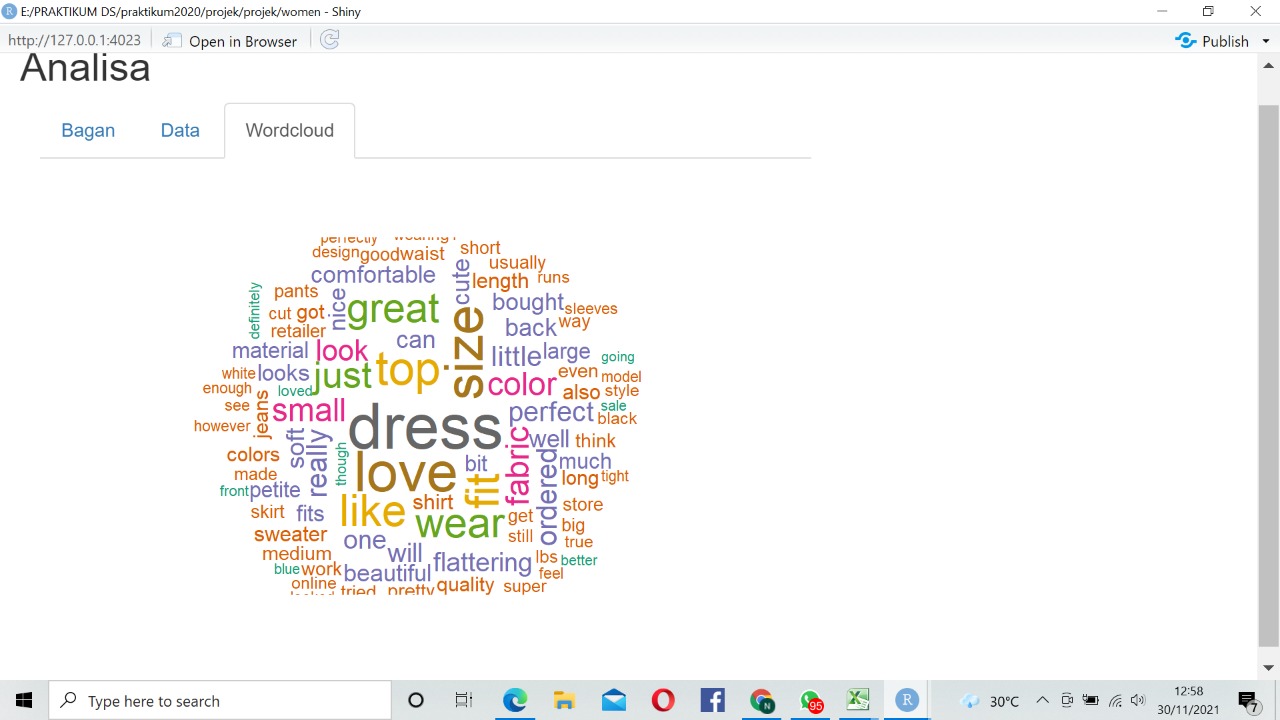


Gambar 3.4

Pada gambar 3.4 merupakan wordcloud yaitu kata” yang sering muncul pada sebuah data.



Gambar 3.5 Gambar 3.6



Gambar 3.7

Gambar 3.5,3.6 dan 3.7 merupakan halaman pada web pada bagian bagan, data dan wordcloud.

# **KESIMPULAN**

Dari review data tentang pakaian wanita memiliki banyak opini yang berbeda dari customer baik itu positif maupun negative. Dari data yang dikumpulkan oleh praktikan menyimpulkan bahwa sebagian besar data terdiri atas opini positive yang diberikan oleh para customer pakaian wanita. Dengan program yang praktikan buat kali ini memudahkan user membaca data tanpa menggunakan waktu yang lama dan dapat menentukan kualitas suatu pakaian melalui data tersebut.