**LAPORAN PROYEK AKHIR**

**PRAKTIKUM DATA SCIENCE**

**ANALISIS SENTIMEN TERHADAP SEBUAH RESTAURANT**

****

**DISUSUN OLEH :**

|  |  |
| --- | --- |
| 1. **LISA MELIANA** | **123190025** |
| 1. **ANDRE CHRISTIAN SURBAKTI** | **123190029** |

### PROGRAM STUDI INFORMATIKA

### JURUSAN TEKNIK INFORMATIKA

### FAKULTAS TEKNIK INDUSTRI

### UNIVERSITAS PEMBANGUNAN NASIONAL ”VETERAN”

#### YOGYAKARTA

#### 2021

# KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa yang senantiasa mencurahkan rahmat dan hidayah-Nya sehingga kami dapat menyelesaikan laporan proyek akhir Praktikum Data Science yang berjudul Analisis Sentimen Terhadap sebuah Restaurant. Adapun laporan ini berisi tentang proyek akhir yang kami pilih dari hasil pembelajaran berlangsung.

Tidak lupa ucapan terima kasih kepada dosen dan Asisten laboratorium yang selalu membimbing dan mengajari kami dalam melaksanakan praktikum dan dalam menyusun laporan ini. Laporan ini masih sangat jauh dari kesempurnaan, oleh karena itu kritik serta saran yang membangun kami harapkan untuk menyempurnakan laporan akhir ini.

Atas perhatian dari semua pihak yang membantu penulisan ini, kami ucapkan terima kasih. Semoga laporan ini dapat dipergunakan seperlunya.

Lampung, 15 November 2021

Penyusun

# DAFTAR ISI

[KATA PENGANTAR ii](#_Toc90151780)

[DAFTAR ISI iii](#_Toc90151781)

[BAB I PENDAHULUAN 1](#_Toc90151782)

[1.1 Latar Belakang 1](#_Toc90151783)

[1.2 Tujuan 2](#_Toc90151784)

[BAB II PEMBAHASAN 3](#_Toc90151785)

[2.1 Dasar Teori 3](#_Toc90151786)

[2.1.1 Bahasa Pemrograman R 3](#_Toc90151787)

[2.1.2 Metode 4](#_Toc90151789)

[2.2 Deskripsi Umum Proyek Akhir 5](#_Toc90151790)

[2.3 Inti Pembahasan 5](#_Toc90151791)

[2.3.1 Listing Program 5](#_Toc90151792)

[2.3.2 Output Program 8](#_Toc90151793)

[BAB III JADWAL PENGERJAAN DAN PEMBAGIAN TUGAS 13](#_Toc90151794)

[3.1 Jadwal Pengerjaan 13](#_Toc90151795)

[3.2 Pembagian Tugas 13](#_Toc90151796)

[BAB VI KESIMPULAN DAN SARAN 14](#_Toc90151797)

[4.1 Kesimpulan 14](#_Toc90151798)

[4.2 Saran 14](#_Toc90151799)

# BAB I PENDAHULUAN

## Latar Belakang

Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, sekarang umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap (Lee, 2010). Selain itu terdapatnya aplikasi yang memudahkan para pecinta kuliner untuk mencari lokasi dan makanan yang sesuai harapan, membuat pecinta kuliner dengan mudah menemukan tempat dan makanan yang diinginkan dimanapun mereka berada. Aplikasi yang ada menyediakan foto, lokasi, ulasan konsumen dan rating yang diberikan oleh konsumen. Review dan rating yang diberikan membantu pelanggan lain untuk mengevaluasi usaha atau servis yang di berikan dan mereka dapat menentukan pilihan (Govindarajan, 2014). Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan cara otomatis mengklasifikasikan ulasan pengguna menjadi pendapat positif atau negatif (Zhang, Ye, Zhang, & Yijun, 2011).

Sudah ada beberapa penelitian analisis sentimen yang dilakukan untuk mengklasifikasi review ataupun opini restoran yang ada di internet di antaranya, klasifikasi sentimen pada review restoran di internet yang ditulis dalam bahasa Canton menggunakan pengklasifikasi Naïve Bayes dan Support Vector Machine yang dilakukan oleh Z.Zhang pada tahun 2011. Analisis sentimen multiclass pada review restoran yang dilakukan oleh Moontae Lee pada tahun 2010. Kategorisasi sentimen pada review restoran yang dilakukan oleh Amir Ghazvinian pada tahun 2010. Pada penelitiannya di tahun 2011, Zhang dan kawan-kawan meneliti pengunaan fitur n-gram dalam menggabungkan kata agar dapat dilihat perbedaan sentimen dari tiap gabungan kata.Rupanya penggunaan fitur n-gram berpengaruh dengan akurasi yang dihasilkan pengklasifikasi. Jika fitur yang digunakan 2-gram (gabungan 2 kata) atau 3-gram (gabungan 3 kata), maka otomatis jumlah kata yang diproses akan semakin banyak dan bisa mempengaruhi akurasi klasifikasi.

## 1.2 Tujuan

#### Tujuan dari proyek ini, yang berjudul Analisis Sentiment Terhadap Sebuah Restaurant bertujuan meningkatkan akurasi pada analisis sentimen review restoran. Serta menyelesaikan tugas akhir dari mata kuliah Praktikum Data Science.

# BAB II PEMBAHASAN

## Dasar Teori

### Bahasa Pemrograman R

### R merupakan bahasa pemrograman untuk analisis statistik yang paling banyak digunakan, karena dengan menggunakan R memungkinkan kamu untuk melakukan import data dari berbagai sumber database dan format yang berbeda. R juga memiliki 7000+ packages yang gratis untuk digunakan. Packages ini memungkinkan kamu untuk melakukan:

* Analisa statistik kompleks
* Econometrics
* Optimisasi
* Pembuatan model machine learning
* Pembuatan teknik simulasi

Software pemrograman R yang paling umum digunakan untuk meningkatkan produktivitas developer adalah RStudio. RStudio merupakan user-friendly interface yang memiliki:

* *Console*
* *R Script editor*
* *R Environment*
* *Graphical Output*

Tidak hanya itu, RStudio juga memiliki tools untuk melakukan plotting, pengolahan data histori, debugging, dan manajemen workspace.  Secara lebih spesifik, RStudio juga memiliki:

* Dokumentasi R Help yang terintegrasi
* Kemampuan untuk mengelola multiple working directories menggunakan projects
* Workspace browser dan data viewer
* Debugger interaktif untuk melakukan diagnosa dan memperbaiki error dengan cepat.

Dengan belajar bahasa pemrograman R dan RStudio kamu dapat:

* Melakukan anotasi kode dengan lebih lengkap
* Melakukan debugging dengan lebih mudah
* Melihat data dalam format spreadsheet
* Melihat seluruh objek yang didefinisikan dalam suatu sesi

Lebih dari itu, RStudio memungkinkan kamu untuk menggunakan packages untuk melakukan otomasi pembuatan laporan dengan format word, pdf, atau ppt. Dan juga memiliki plugin visualisasi yang powerful bernama Shiny. Shiny memungkinkan pengembangan visualisasi data yang inetraktif dan fleksibel yang kemudian dapat dijadikan sebuah dashboard dan dapat diakses oleh non-programmers menggunakan internet browsers. Dengan belajar bahasa pemrograman R kamu dapat mengotomasi dan menggunakanya untuk mengimplementasi berbagai tugas manajemen data, analisis statistik, dan pembuatan model machine learning.

### Metode

Metode penelitian yang penulis lakukan adalah metode eksperimen, dengan tahapan berikut:

* 1. Pengumpulan Data

Pengumpulan data menggunakan data review restaurant yang didapat dari kaggle yang lebih dari 100 data data review.

* 1. Pengolahan Data

Dataset ini terdapat dalam tahap preprocessing yang harus melalui beberapa tahapan yaitu proses dimana data yang memiliki tanda baca dan symbol yang bukan huruf akan dihilangkan, kemudian dilakukan penghapusan dan pembuangan kata-kata yang sering ditampilkan dalam document, dan juga dilakukan pembuangan prefix dan suffix sehingga membentuk menjadi kata dasar.

* 1. Metode yang diusulkan

Untuk mengetahui pengaruh pengklasivikasi Naive Bayes menggunakan Bahasa pemrograman R dalam analisis sentiment review restoran.

## Deskripsi Umum Proyek Akhir

Data yang digunakan dalam pengklasifikasian teks diambil dari kagle yang terdiri dari review positif dan review negative. Sebelum diklasifikasi data tersebut harus melawati beberapa tahapan proses agar dapat supaya dapat diklasifikasikan ke proses selanjutnya, tahapan itu terdiri dari pengumpulan data, pengolahan awal data, kemudian klasifikasi. Dimana klasifikasi ini dalam proses review restoran dapat menentukan class untuk setiap reviewnya, class review tebagi menjadi dua yaitu class positif dan class negative.

Data yang diujikan menggunakan beberapa fold cross validation. Dimana dengan menggunakan teknik ini dengan membagi secara acak ke dalam tiap bagian dimana terdiri dari beberapa bagian an setiap bagian akan dilakukan proses klasifikasi. Model klasifikasi teks pada review restoran membuat pembaca dapat dengan mudah mengidentifikasi mana review yang postif dan negatif. kata-kata dipisahkan lalu diberikan bobot pada masing-masing kata untuk review restoran yang sudah ada. Kata yang ada dilihat dari hubungan dengan sentimen yang sering muncul dan memiliki bobot paling tinggi untuk setiap dokumennya. Dengan demikian dapat diketahui review tersebut positif atau negatif.

## Inti Pembahasan

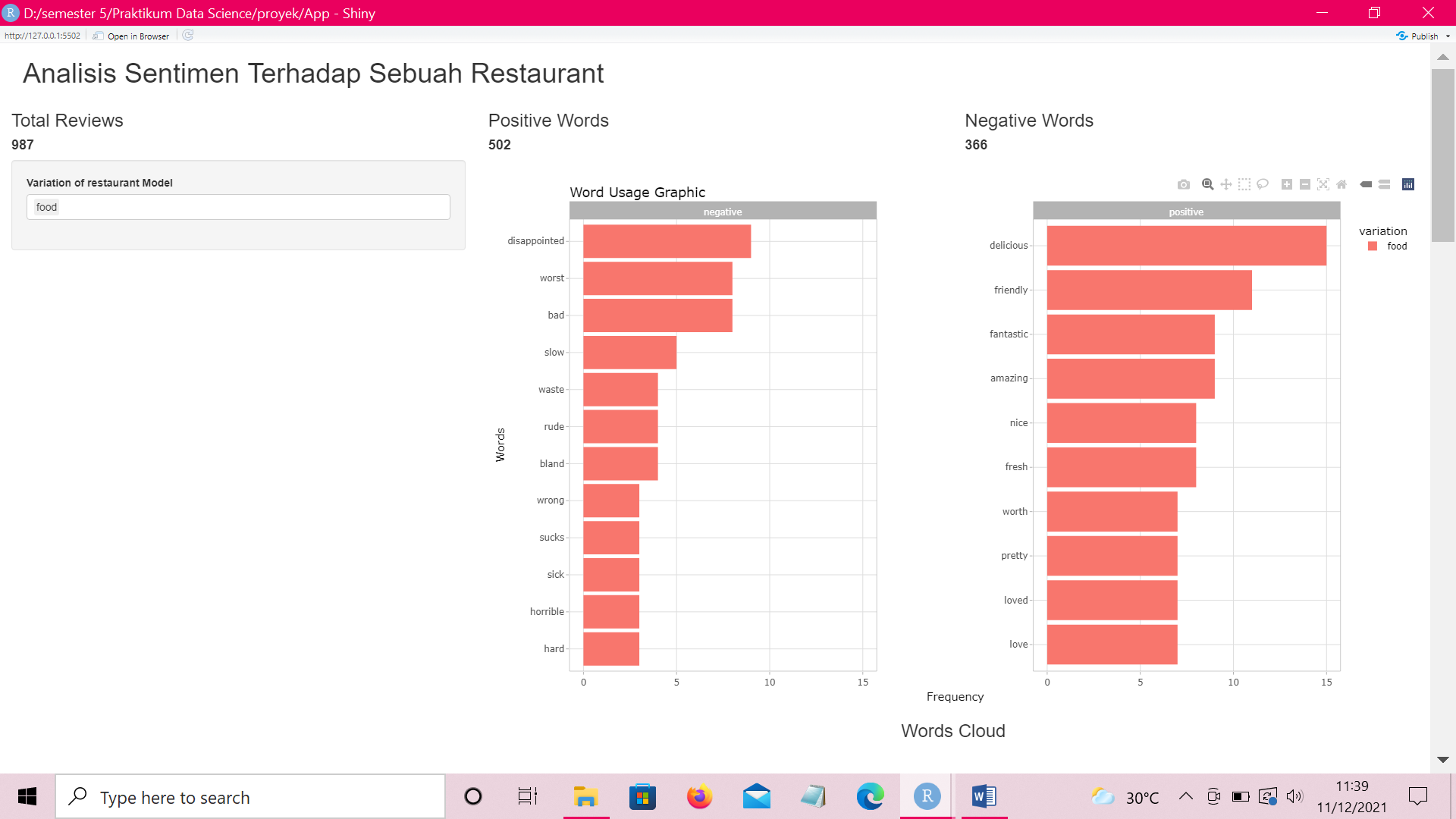
### Listing Program

|  |
| --- |
| library(dplyr)  library(vroom)  library(here)  library(tidyverse)  library(ggplot2)  library(plotly)  library(tidytext)  library(wordcloud)  library(wordcloud2)  library(reshape2)  library(shiny)  library(tm)  library(memoise)  restaurant = vroom(here("restaurant\_reviews.tsv"), delim = "\t")  option\_variation = unique(restaurant$variation)  count\_reviews = function() {  restaurant %>%  nrow()  }  count\_sentiments = function(x) {  restaurant %>%  unnest\_tokens(word, verified\_reviews) %>%  anti\_join(stop\_words) %>%  inner\_join(get\_sentiments("bing")) %>%  count(sentiment) %>%  filter(sentiment == x)  }  table\_restaurant = function() {  restaurant %>%  mutate(feedback = case\_when(  feedback == 1 ~ "Positive",  TRUE ~ "Negative"  )) %>%  select(variation, verified\_reviews, feedback) %>%  head(50)  }  ui = fluidPage(  title = "Sentiment Analysis of Restaurant Review",  headerPanel("Sentiment Analysis of Restaurant Review"),    fluidRow(  column(  4,  h3("Total Reviews"),  h4(strong(textOutput(outputId = "total\_reviews")))  ),  column(  4,  h3("Positive Words"),  h4(strong(textOutput(outputId = "total\_positive")))  ),  column(  4,  h3("Negative Words"),  h4(strong(textOutput(outputId = "total\_negative")))  )  ),    sidebarLayout(  sidebarPanel(  selectInput(  inputId = "variation",  label = "Variation of restaurant Model",  choices = option\_variation,  multiple = TRUE,  selected = option\_variation[[1]]  )  ),  mainPanel(  br(),  plotlyOutput(outputId = "plot\_word\_usage", height = "700px"),  h3("Words Cloud", align = "center"),  plotOutput(outputId = "plot\_word\_cloud", height = "1200px"),  h3("Table Reviews"),  tableOutput(outputId = "plot\_reviews")  )  )  )  server = function(input, output, session) {  plot\_word\_freq = reactive({  restaurant %>%  group\_by(variation) %>%  unnest\_tokens(word, verified\_reviews) %>%  group\_by(variation) %>%  anti\_join(stop\_words) %>%  count(word, sort = T) %>%  na.omit() %>%  filter(n >= 30) %>%  ggplot(aes(x = reorder(word, n), y = n, fill = variation)) +  geom\_bar(stat = "identity") +  coord\_flip() +  labs(  x = "Words",  y = "Frequency",  title = "Word Frequency Graphic"  ) +  theme\_light()  })    output$plot\_word\_freq = renderPlotly({  ggplotly(plot\_word\_freq())  })    plot\_word\_usage = reactive({  restaurant %>%  filter(variation %in% input$variation) %>%  unnest\_tokens(word, verified\_reviews) %>%  anti\_join(stop\_words) %>%  inner\_join(get\_sentiments("bing")) %>%  group\_by(sentiment, variation) %>%  count(word) %>%  top\_n(10) %>%  ggplot(aes(x = reorder(word, n), y = n, fill = variation)) +  geom\_col(show.legend = T) +  coord\_flip() +  facet\_wrap(~sentiment, scales = "free\_y") +  labs(  x = "Words",  y = "Frequency",  title = "Word Usage Graphic"  ) +  theme\_light()  })    output$plot\_word\_usage = renderPlotly({  ggplotly(plot\_word\_usage())  })    output$plot\_word\_cloud = renderPlot({  restaurant %>%  filter(variation %in% input$variation) %>%  unnest\_tokens(word, verified\_reviews) %>%  anti\_join(stop\_words) %>%  inner\_join(get\_sentiments("bing")) %>%  count(word, sentiment) %>%  acast(word~sentiment, value.var = "n", fill = 0) %>%  comparison.cloud(colors = c("#1b2a49", "#00909e"), max.words = 200, scale = c(4,1))  })    output$total\_reviews = renderText({  count\_reviews()  })    output$total\_positive = renderText({  count\_sentiments("positive")$n  })    output$total\_negative = renderText({  count\_sentiments("negative")$n  })    output$plot\_reviews = renderTable({  table\_restaurant()  })  }  shinyApp(ui = ui, server = server) |

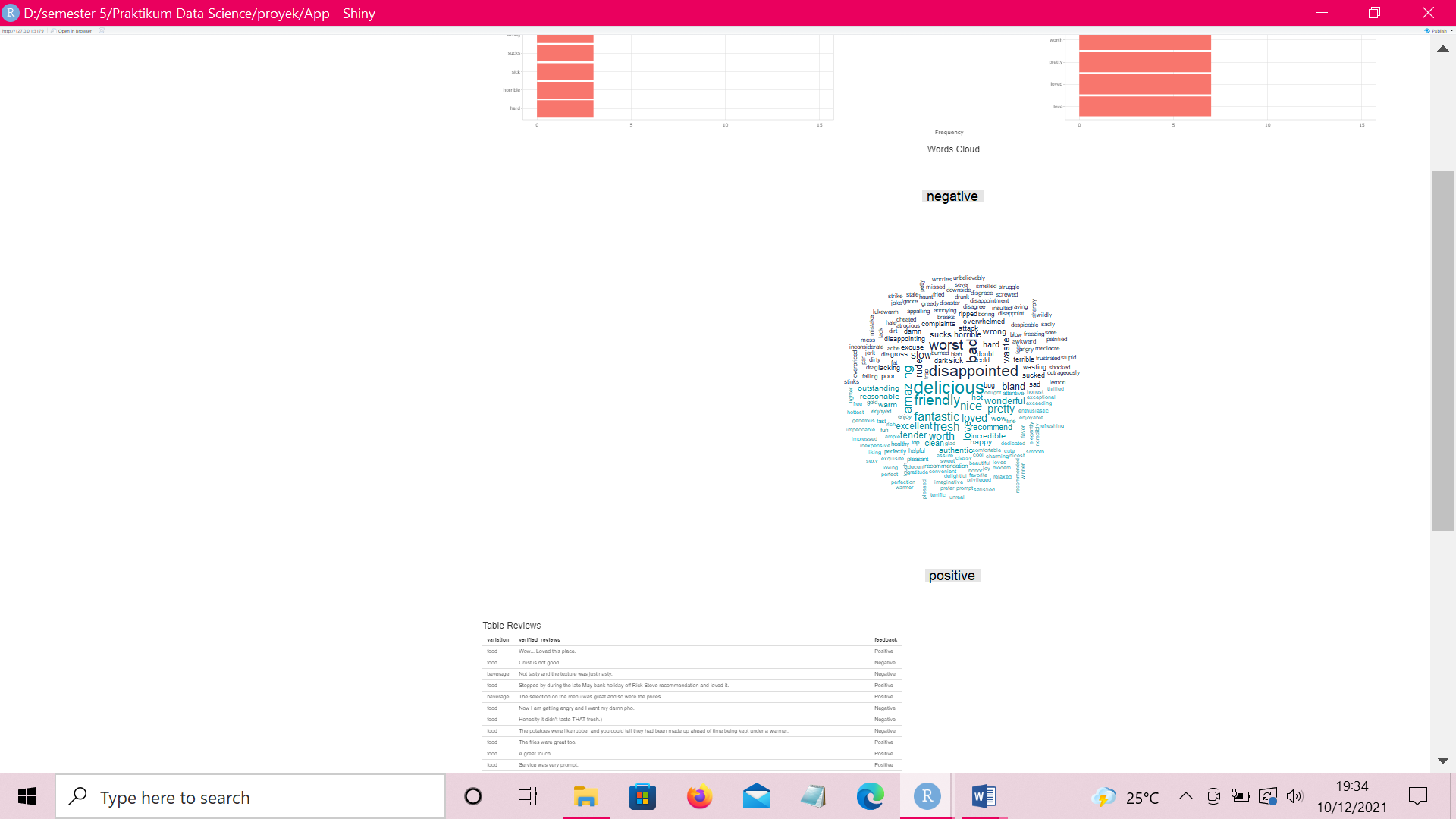
**Listing 2.1** Listing Program

### Output Program

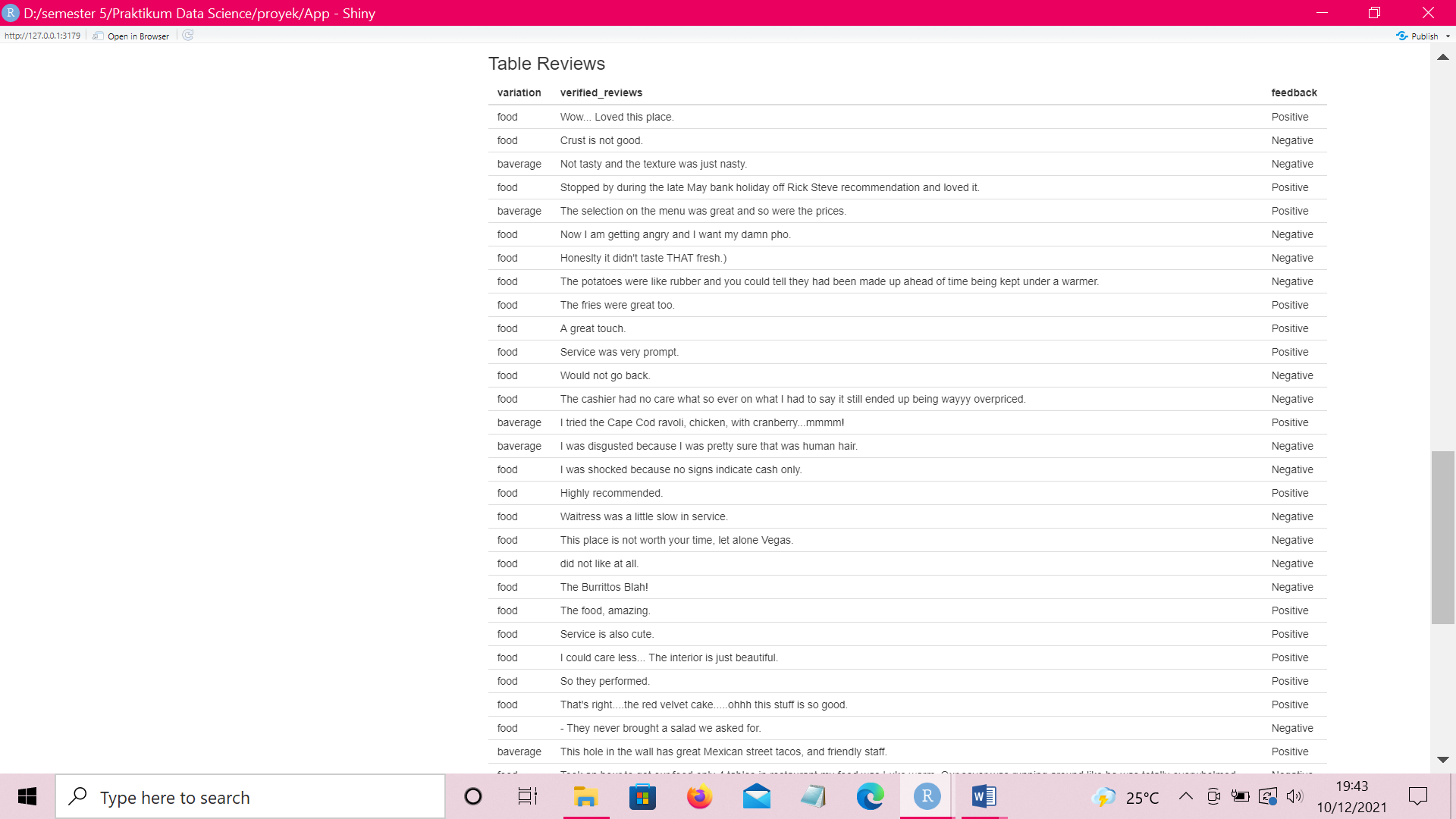
**Tampilan Food**



**Gambar 2.1** Tampilan validation dan Graphic

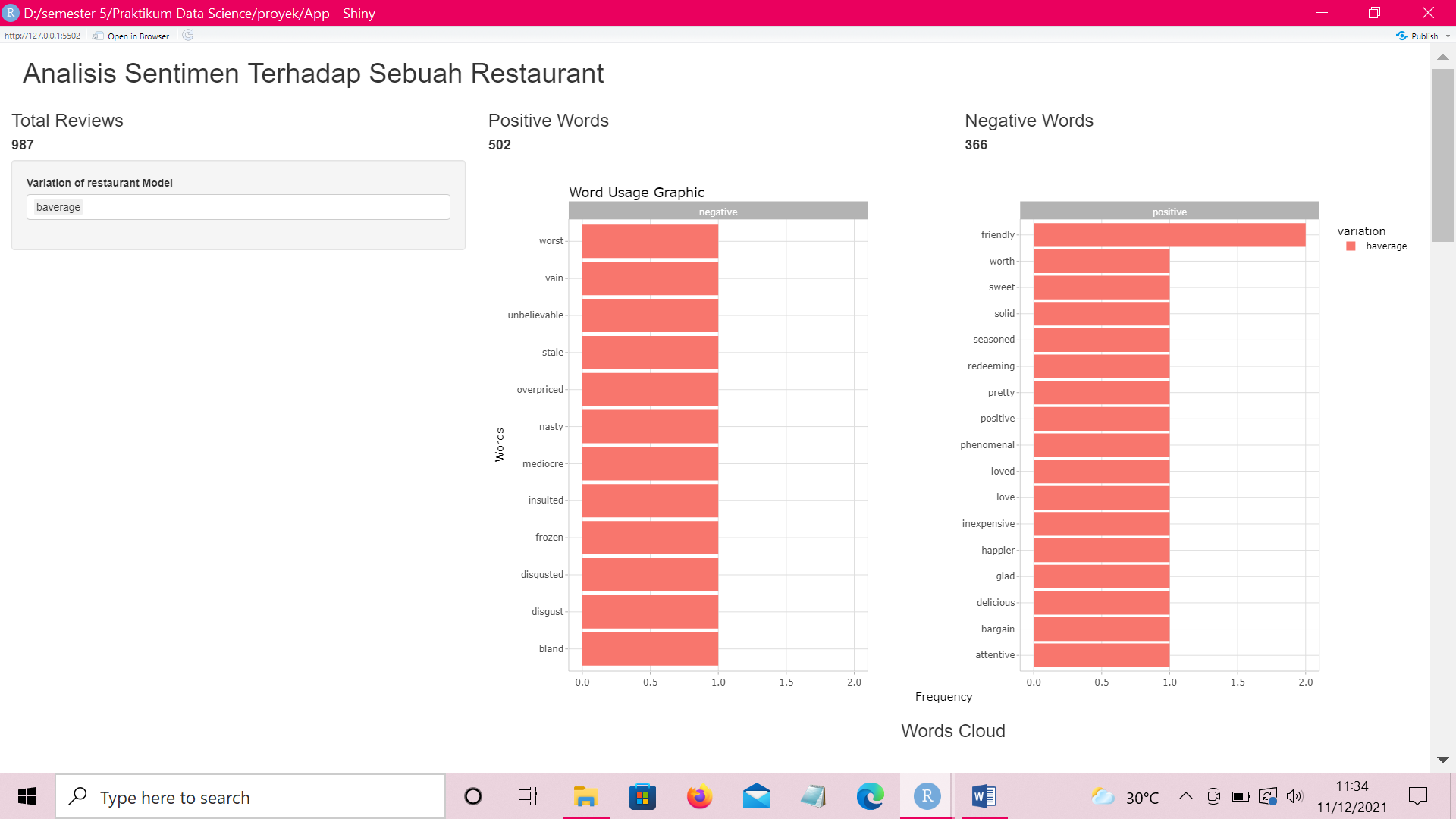


**Gambar 2.2** Tampilan Words Cloud

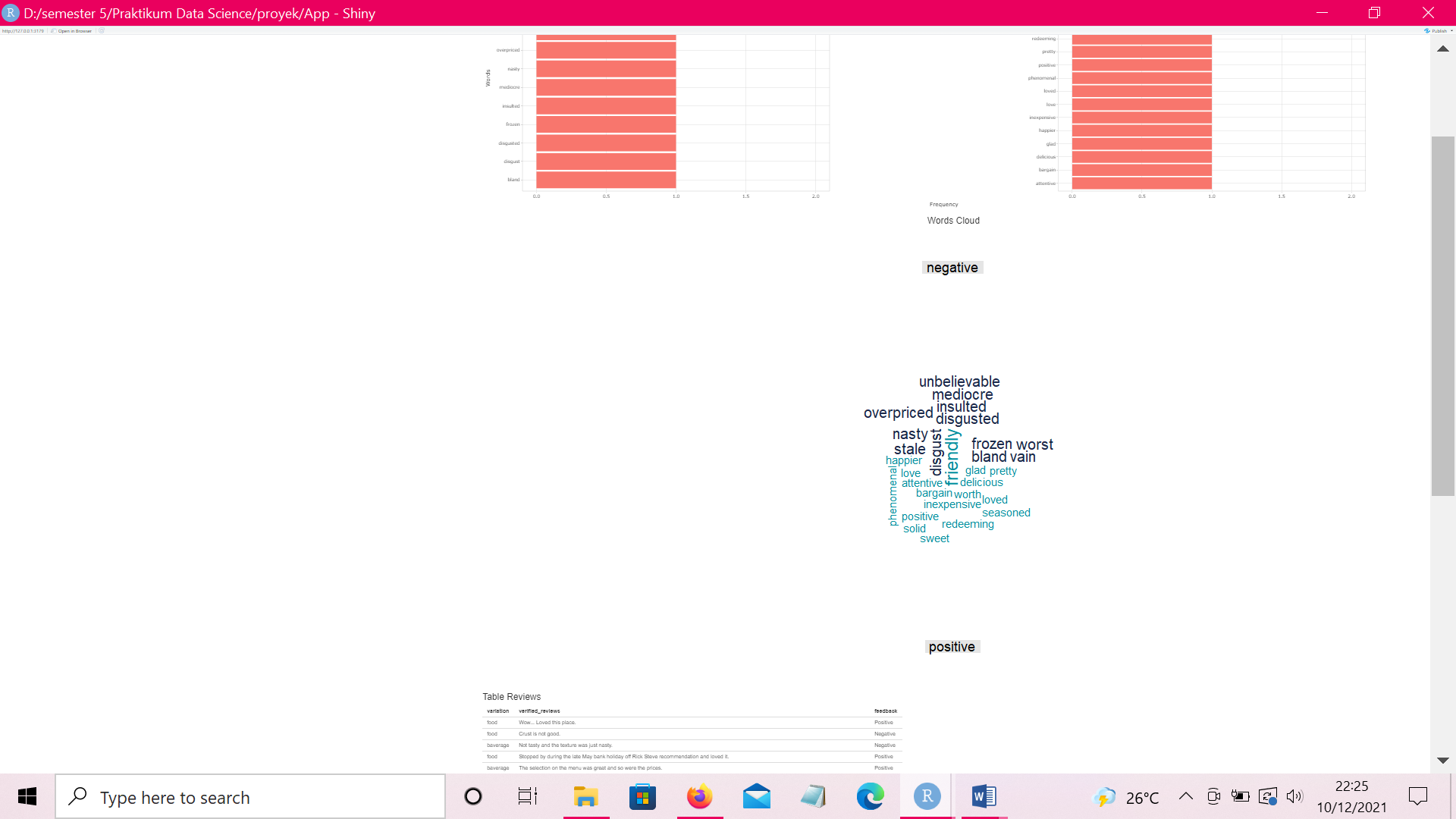


**Gambar 2.3** Tampilan Table Reviews

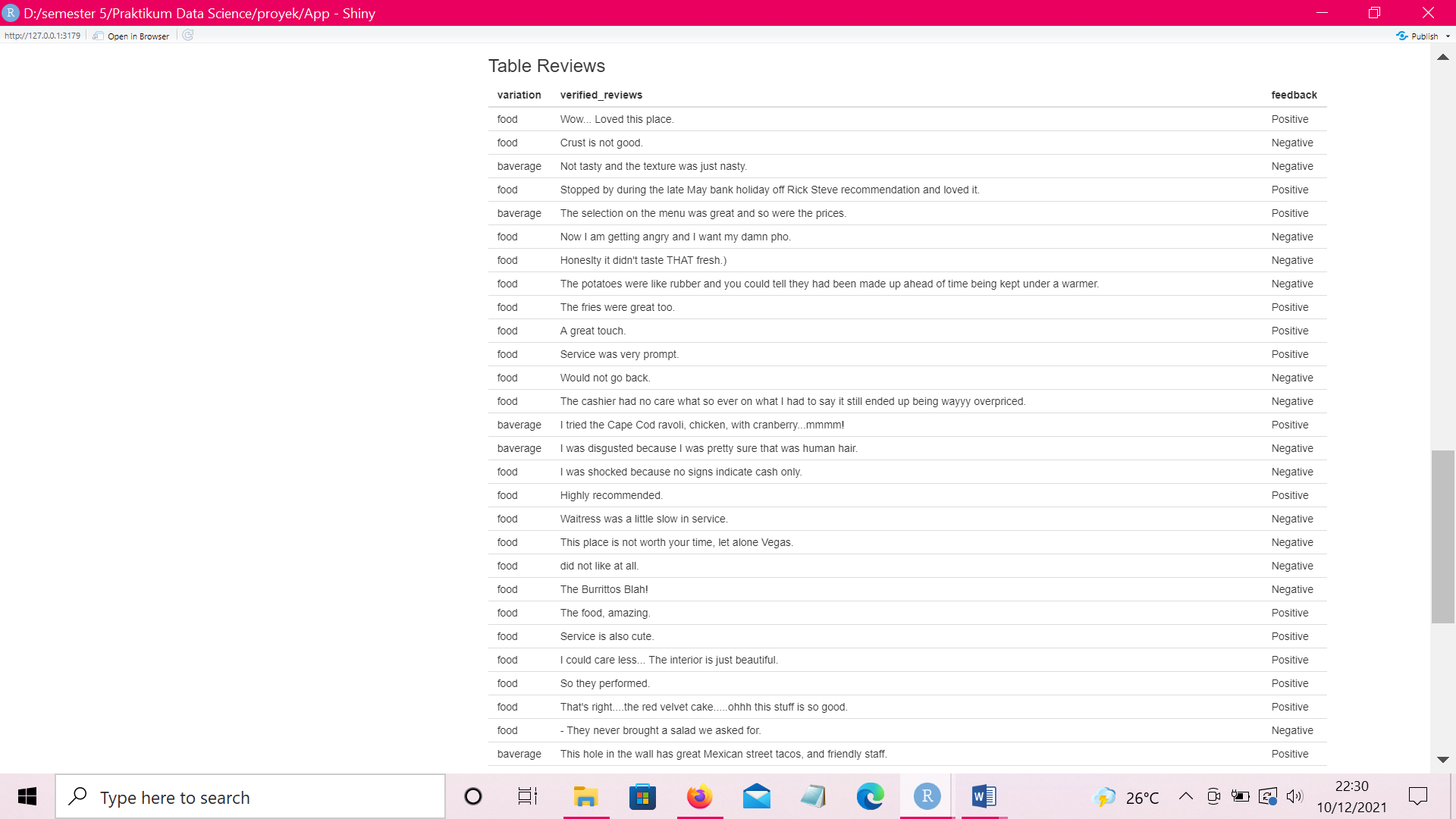
**Tampilan Baverage**



**Gambar 2.4** Tampilan validation dan Graphic

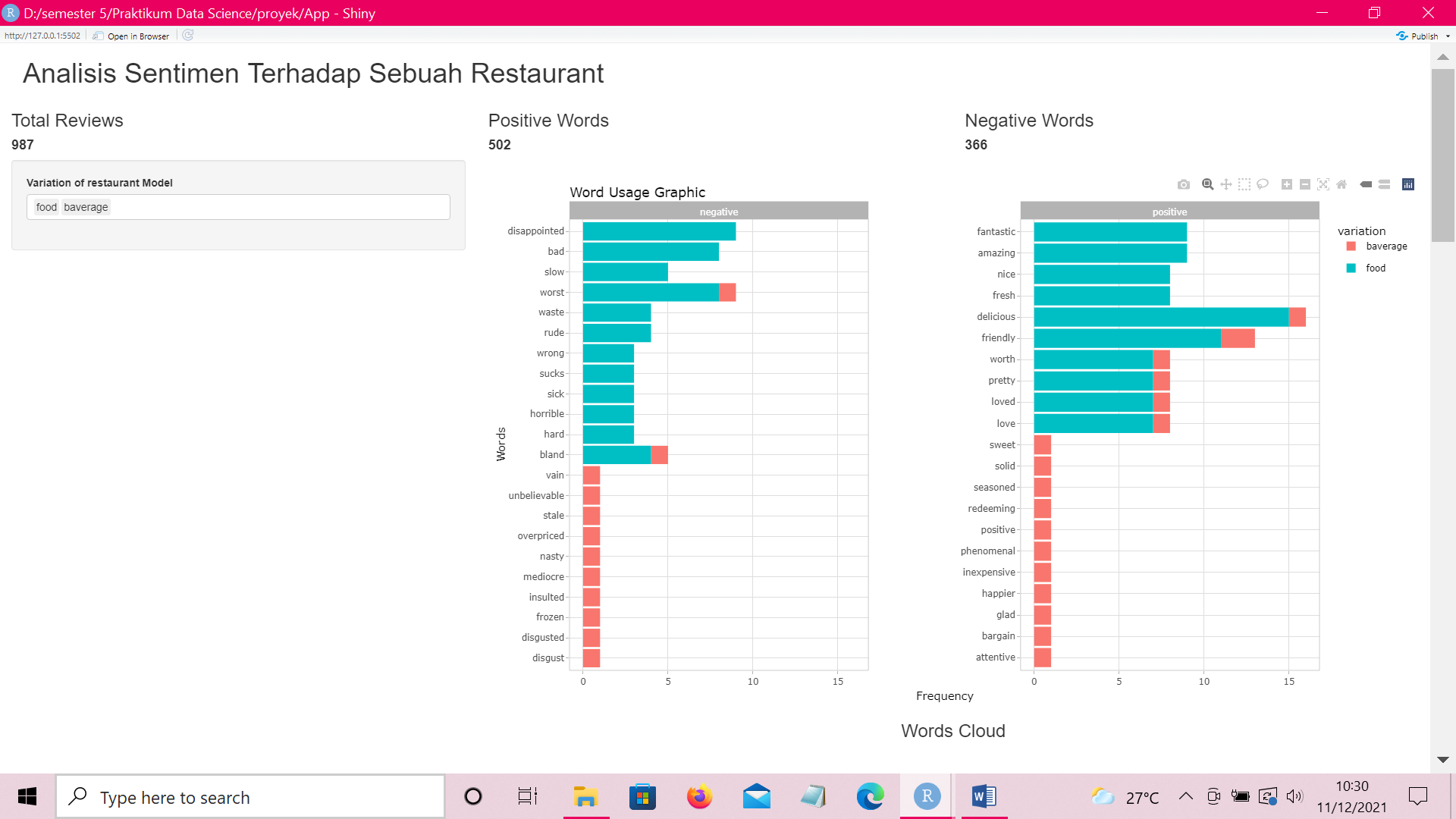


**Gambar 2.5** Tampilan Words Cloud

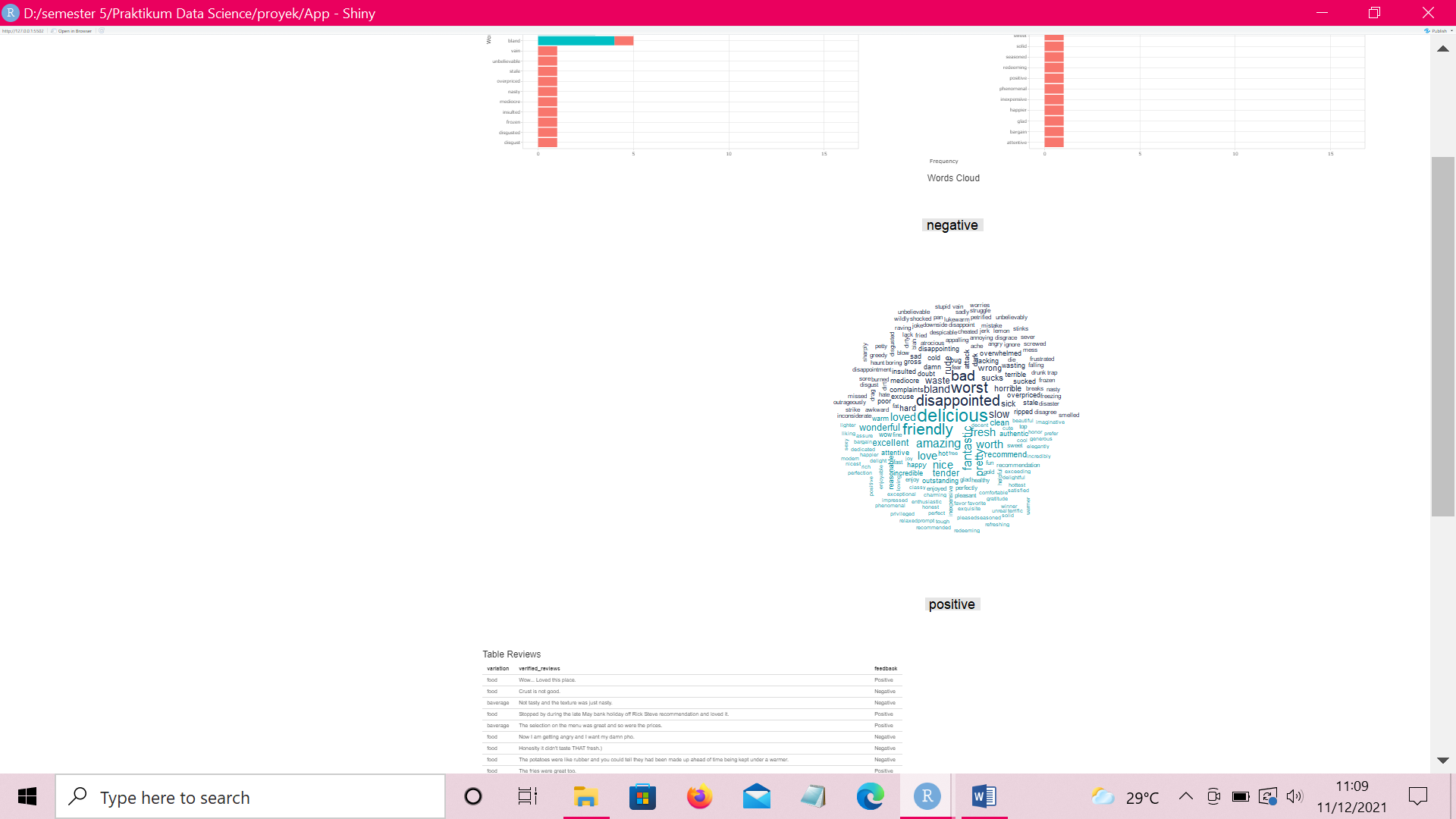


**Gambar 2.6** Tampilan Table Reviews

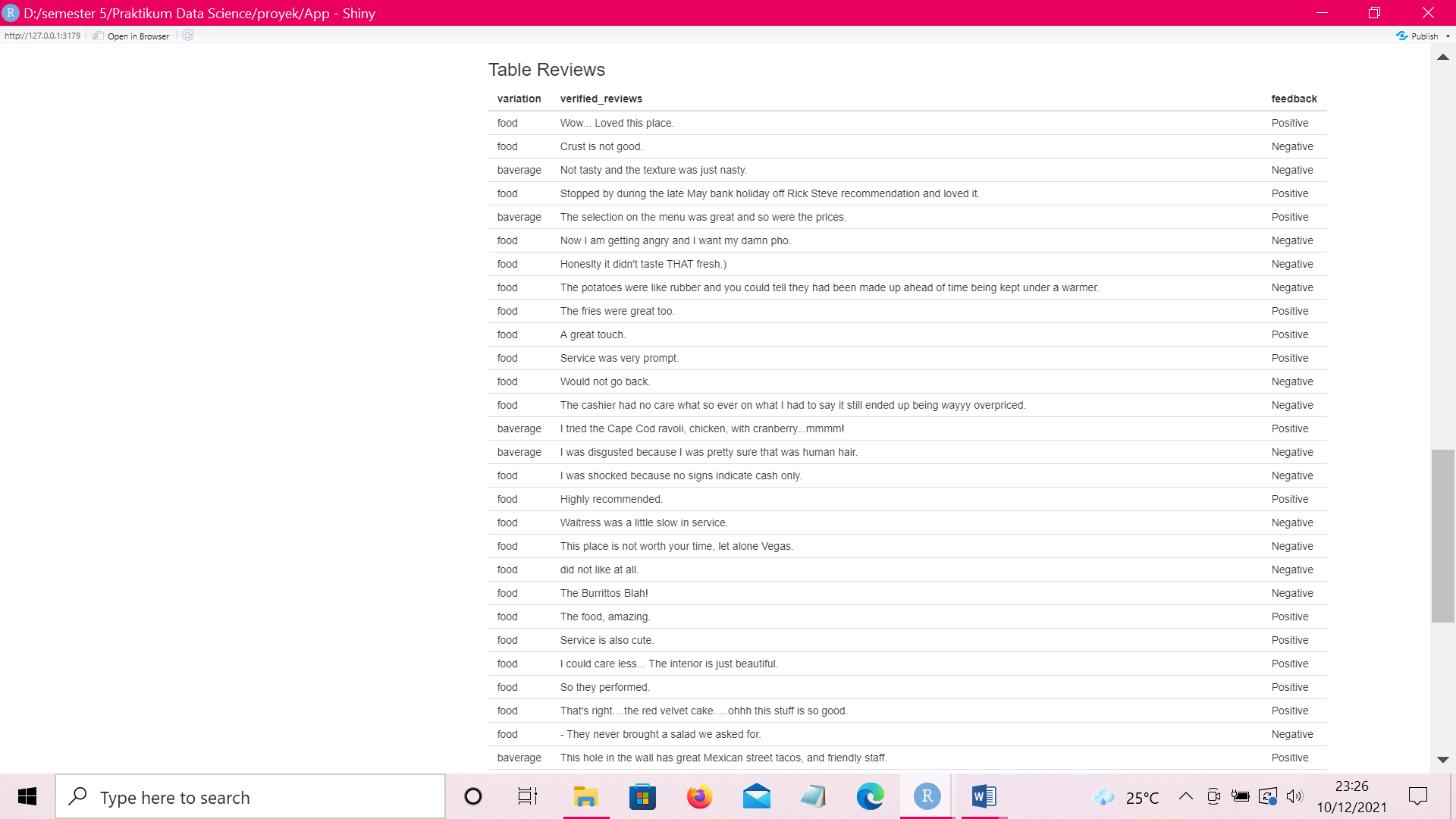
**Tampilan Food dan Baverage**



**Gambar 2.7** Tampilan validation dan Graphi



**Gambar 2.8** Tampilan Words Cloud



**Gambar 2.9** Tampilan Table Reviews

# BAB III JADWAL PENGERJAAN DAN PEMBAGIAN TUGAS

## 3.1 Jadwal Pengerjaan

Jadwal pengerjaan proyek Praktikum Data Science Pemilihan Beras ini sebagai berikut:

**Tabel 3.1** Tabel Jadwal Pengerjaan

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| no | Nama Tugas | Minggu ke- | | | |
| November | | Desember | |
| 3 | 4 | 1 | 2 |
| 1 | Menentukan Judul Proyek |  |  |  |  |
| 2 | Perancangan Konsep Proyek |  |  |  |  |
| 3 | Desain Skenario Program |  |  |  |  |
| 4 | App |  |  |  |  |
| 6 | Pembuatan Laporan |  |  |  |  |

## 3.2 Pembagian Tugas

Pembagian tugas proyek Praktikum Data Science Pemilihan Beras ini sebagai berikut:

**Tabel 3.2** Tabel Pembagian Tugas

|  |  |  |
| --- | --- | --- |
| no | aktivitas | penanggung jawab |
| 1 | Pembahasan Judul | Andre & Lisa |
| 2 | Perancangan Konsep | Andre & Lisa |
| 3 | Desain Skenario | Andre |
| 4 | App | Andre & Lisa |
| 6 | Pembuatan Laporan | Lisa |

# BAB VI KESIMPULAN DAN SARAN

## 4.1 Kesimpulan

Berdasarkan permasalahan yang dibahas, disini kita menggunakan metode eksperimen dengan tahapan pengumpuan data, pengolahan data dan metode yang diusulkan. Untuk data yang digunakan dalam pengklasifikasian teks ini diambil dari kagle yang terdiri dari review positif dan review negative. Sebelum diklasifikasi data tersebut harus melawati beberapa tahapan proses agar dapat supaya dapat diklasifikasikan ke proses selanjutnya, tahapan itu terdiri dari pengumpulan data, pengolahan awal data, kemudian klasifikasi. Dimana klasifikasi ini dalam proses review restoran dapat menentukan class untuk setiap reviewnya, class review tebagi menjadi dua yaitu class positif dan class negative. Degan menggunakannya metode ini maka pembaca dapat dengan mudah membaca review negative dan positive

## 4.2 Saran

Adapun saran yang dapat dipertimbangkan untuk penyempurnaan dan pengembangan program analisis sentiment terhadap restoran lebih lanjut adalah sebagai berikut:

1. Sistem diharapkan dapat bekerja dengan lebih optimal dengan menyediakan fitur-fitur maupun menu-menu yang lebih banyak serta lebih kompleks, sehingga *user* dapat lebih nyaman dalam melakukan aktivitas.
2. Penambahan fitur pada apliaksi supaya bisa di buka di berbagai media dan fitur, sehingga tidak hanya bisa di buka di PC saja