

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE**

**Analisis Sentimen Untuk Menentukan Kata-Kata Paling Negatif dan
Positif Dari Review Hotel**



ANNAS ADHARUQUDNI 123190014
FRISKA EKA KHOIRUNISA 123190031

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

1. PENDAHULUAN

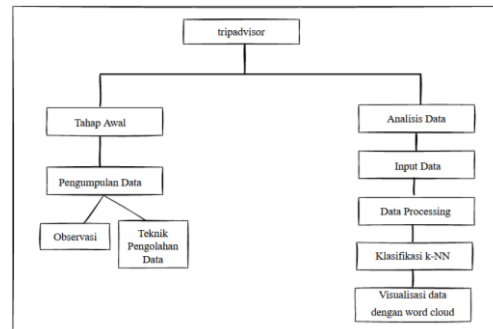
Salah satu layanan online pariwisata yang banyak digunakan masyarakat dunia saat ini adalah TripAdvisor. TripAdvisor adalah salah satu situs wisata terbesar di dunia yang membantu wisatawan mengoptimalkan potensi setiap perjalanan. TripAdvisor menawarkan sarana dari jutaan wisatawan serta berbagai pilihan dan fitur perencanaan wisata dengan link praktis ke alat bantu pemesanan yang memeriksa ratusan situs web untuk menemukan harga hotel terbaik. Situs web TripAdvisor merupakan komunitas wisata terbesar di dunia, menjangkau rata-rata 390 juta pengunjung untuk setiap bulannya, serta menampilkan 435 juta ulasan dan opini tentang 6,8 juta akomodasi, restoran, dan objek wisata yang beroperasi di 49 pasar di seluruh dunia (TripAdvisor, 2016).

Situs ini juga menyediakan kolom komentar dari pengguna yang mengulas tentang objek wisata maupun kuliner yang telah dikunjungi dan dapat dibaca oleh wisatawan lain. Data komentar yang ada di situs web tersebut dapat dimanfaatkan bagi pengelola hotel, berdasarkan hasil analisis sentiment yang kelas sentiment positif, negatif, atau netral. Hasil analisis opini itu selanjutnya dapat digunakan sebagai pertimbangan wisatawan dalam penentuan keputusan memilih hotel yang akan dikunjungi.

Untuk membuat sebuah analisis sentimen banyak hal yang harus dipersiapkan terlebih dahulu, salah satunya dengan memilih classifier yang akan digunakan. Classifier metode yang dapat melakukan klasifikasi data menjadi beberapa kelas. Dalam penelitian ini, classifier yang dipilih adalah K-means, topic modelling, dan wordcloud.

2. METODE

Metodologi yang digunakan untuk melakukan analisis sentimen menentukan komentar positif dan negative dari aplikasi tripadvisor menggunakan metode k-nearest neighbor(k-NN) dibagi menjadi tiga tahapan yaitu tahap awal dan analisis data



Gambar 2.1 Metodologi Penelitian

2.1 Tahap Awal

Di tahap awal ini kami melakukan studi literatur untuk menentukan studi kasus serta metode yang akan digunakan ketika melakukan penelitian. Studi literature ini melakukan pengumpulan dan menganalisa jurnal yang sudah dibuat sebelumnya.

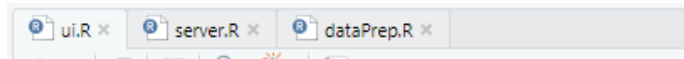
Kemudian, melakukan pengumpulan data yang didapat dari website *kaggle* untuk melihat komentar pelanggan di Tripadvisor. Pengumpulan data ini terbagi menjadi dua yaitu, observasi dan teknik pengolahan data. Observasi digunakan untuk menentukan objek yang nantinya ingin diteliti yaitu, kata komentar pelanggan baik yang positif maupun yang negatif. Teknik pengolahan data yaitu mengambil komentar pelanggan yang memuat kata positif dan negative dari aplikasi Tripadvisor dengan total data 20.491komentar.

2.2 Analisis Data

Tahapan klasifikasi K-Nearest Neighbor (k-NN). Metode k-NN merupakan model supervised learning yang dimana database populasi yang digunakan diidentifikasi terlebih dahulu ke dalam kelas tertentu. K-NN memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih.

Di proyek ini kami menggunakan library shiny, shinycssloaders, shinydashboard, tidyverse, ggtech, tidytext, widyr, stm, e1071, caret, syuzhet, tm, dplyr, wordcloud, cluster, factoextra.

Pembuatannya menggunakan 3 file .R yaitu untuk tampilan interface, untuk persialan data, dan server (**Gambar 2.2**). File ui.R berisi codingan untuk antarmuka di website, file server.R digunakan untuk mengirim data ke ui.R, dan dataPrep.R digunakan untuk persiapan data yang akan diolah.

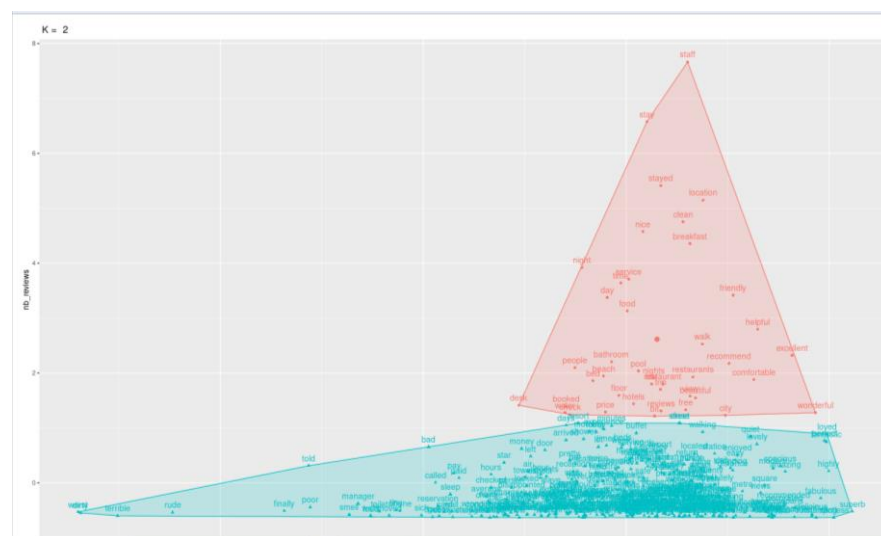


Gambar 2.2 Persiapan File

```
output$kmeans <- renderPlot({
  k2 <- kmeans(indexed, centers =
    input$sizeKmeans)
  fviz_cluster(k2, data = indexed) +
  ggtitle(paste(" K = ",input$sizeKmeans))
})
```

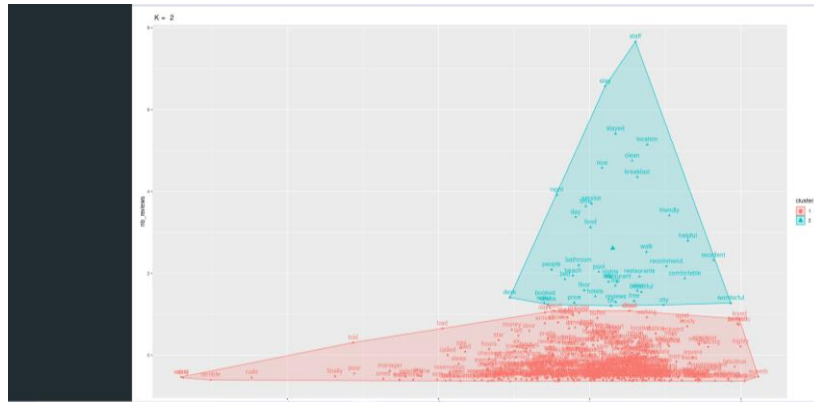
Tabel 2.1 Menentukan k-NN

Berdasarkan **Tabel 2.1** nilai k yang direkomendasikan adalah 2 digunakan ntuk melihat pengelompokkan komentar yang berada di aplikasi Tripadvisor. Dengan hasil seperti **Gambar 2.3**.

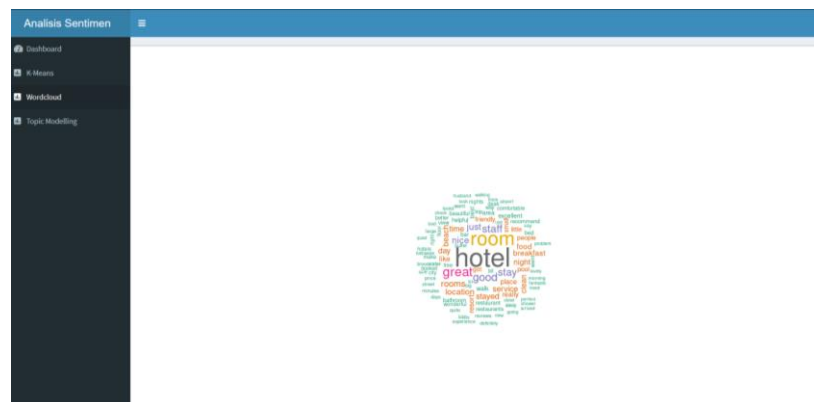


Gambar 2.3 Rekomendasi k-NN

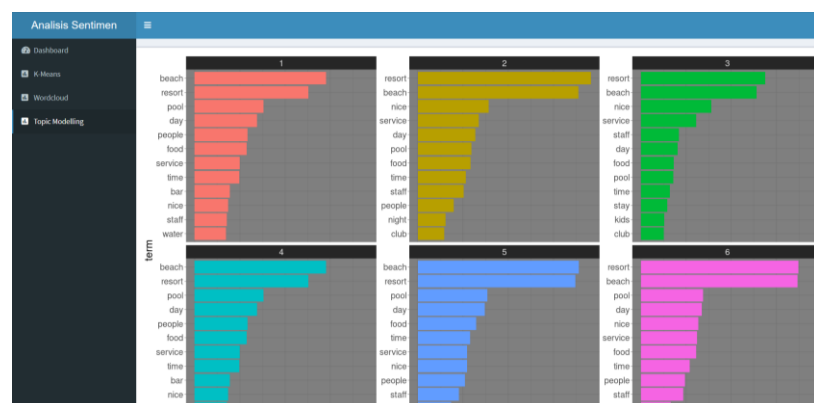
Setelah melakukan pengklasifikasian dengan k-NN data akan di visualisasi dengan menggunakan wordcloud, yang akan menghasilkan kata yang paling sering muncul dalam komentar di Tripadvisor.



Tampilan selanjutnya, mengelompokkan kata dengan k-NN dapat dilihat di **Gambar 3.2** terdapat grafik yang digunakan untuk merekomendasi k-NN yang digunakan yaitu, 2. Kemudian, di **Gambar 3.3** terdapat pengelompokkan dengan k-NN 2.



Tampilan selanjutnya, memunculkan huruf yang sering keluar di kolom komentar dengan menggunakan wordcloud. Terdapat kata hotel, staff, good, great, stay, dll.



Tampilan terakhir dari adalah topic modeling yaitu Topic modeling atau pemodelan topik merupakan metode clustering yang termasuk dalam unsupervised learning. Dalam unsupervised learning tidak ada label untuk suatu objek.

4. KESIMPULAN

Penelitian ini melakukan pengklasifikasi teks dari komentar trivadsitor dengan pengklasifikasi K-Nearest Neighbor, menggunakan 20.491 komentar. Serta kata yang sering muncul dalam komentar adalah sick, moved, loud, walls, smell, toilet, dll.