

**LAPORAN PROYEK AKHIR  
PRAKTIKUM DATA SCIENCE**

**Analisis Sentimen Bencana Alam**



EKKY RAMADHAN  
REGA MUHAMMAD AGASSI

123190124  
123190138

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

## **1. PENDAHULUAN**

Bencana merupakan kejadian yang disebabkan oleh alam maupun oleh kelalaian manusia. Tanah longsor, gempa bumi, puting beliung, tsunami, banjir dan tanah longsor, letusan gunung merapi, kekeringan serta gelombang pasang adalah bencana yang disebabkan oleh alam. Sementara itu aksi teror, konflik, kecelakaan industri, kecelakaan transportasi, dan kebakaran hutan merupakan bencana akibat kelalaian manusia. Bencana yang disebabkan oleh alam dan kelalaian manusia samasama menimbulkan kerugian terhadap lingkungan dan perekonomian

Twitter merupakan salah satu aplikasi yang marak digunakan untuk membuat ulasan tentang apapun apa yang sedang terjadi di sekitar kita. Twitter telah dimanfaatkan untuk banyak hal misalnya kampanye, media belajar, berbagi pendapat/pandangan, media protes, dan yang lainnya. Penggunaan Twitter cenderung akan melonjak ketika terjadi peristiwa trending/populer, Twitter mampu menjadi layanan jejaring sosial dengan fitur posting pesan singkat atau status dan beberapa fitur menarik lainnya (Twitter 2019)

Aplikasi ini juga menyediakan ulasan-ulasan dari pengguna yang mengulas tentang berbagai informasi, berita atau pendapat yang sedang terjadi disekitar pengguna tersebut dan dapat dibaca oleh pengguna lain. Data ulasan yang ada di aplikasi tersebut dapat dimanfaatkan bagi pengguna twitter, berdasarkan hasil analisis sentimen yang kelas sentimen positif, negatif, atau netral. Hasil analisis opini itu selanjutnya dapat digunakan sebagai analisis emosi saat ada bencana yang menimpa atau sedang trending.

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, dan wordcloud.

## **2. METODE**

Kami melakukan pengumpulan data yang didapat dari website *twitter.com* untuk melihat tweet yang telah dipublikasi oleh pemilik akun. Pengumpulan data ini terbagi menjadi dua yaitu, observasi dan teknik pengolahan data. Observasi digunakan untuk menentukan objek yang nantinya ingin diteliti yaitu, data cuitan pengguna twitter yang positif maupun yang negatif. Teknik pengolahan data yaitu mengambil tweet

pengguna twitter yang memuat kata positif dan negatif dari website twitter dengan data yang selalu terbaru jika kita mengambil kembali pada twitter dengan keyword #flood, #volcano, dan #earthquake.

Metodologi yang digunakan untuk melakukan analisis sentimen menentukan emosi atau perasaan pengguna twitter dalam ulasannya menggunakan metode *k-means Clustering*. *K-means* adalah metode pengklasteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. *K-means* mampu meminimalkan rata-rata jarak setiap data ke klasternya.

Pada penelitian ini, algoritma *k-means clustering* memiliki beberapa tahapan seperti berikut:

1. Menentukan *k* sebagai jumlah cluster yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat cluster awal (centroid) sebanyak *k*.
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean (Euclidean Distance) hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

Keterangan:

$x_i$  : data kriteria

$\mu_j$  : centroid pada cluster ke-*j*

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Memperbaharui nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus:

$$\mu_j(t+1) = 1/N_{sj} \sum_{x_j \in S_j} x_j$$

Keterangan:

$\mu_j(t+1)$  : centroid baru pada iterasi ke (*t*+1)  $N_{sj}$

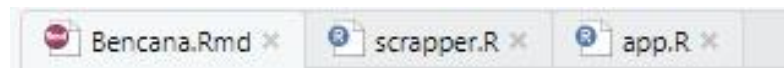
: banyak data pada cluster *S<sub>j</sub>*.

6. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap cluster tidak ada yang berubah. Jika langkah 6 telah terpenuhi, maka nilai pusat

cluster ( $\mu_j$ ) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

Di proyek ini kami menggunakan library shiny, here, vroom, dplyr, ggplot2, plotly, topicmodels, tidyverse, tidytext, wordcloud, syuzhet, lubridate, scales, reshape2, tm.

Pembuatannya menggunakan 3 file .R yaitu untuk tampilan scrapper, bencana, app (**Gambar 2.2**). File scapper.R berisi codingan mengambil data dari twitter.com, bencana.Rmd digunakan untuk membersihkan data, membuat wordcloud, menentukan sentiment emosi pengguna dan mengirim data ke app.R, dan app.R digunakan untuk tampilan web dan UI..



**Gambar 2.2** Persiapan File

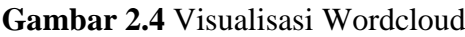
```
k <- 7 # diusahakan ganjil, 7 yang paling tinggi

# dicari 7 data terdekat
predict_knn <- cosine %>%
  inner_join(train, by = "id") %>%
  select(id, hasil_cosine, keyword, text) %>%
  head(k)

view(predict_knn)
```

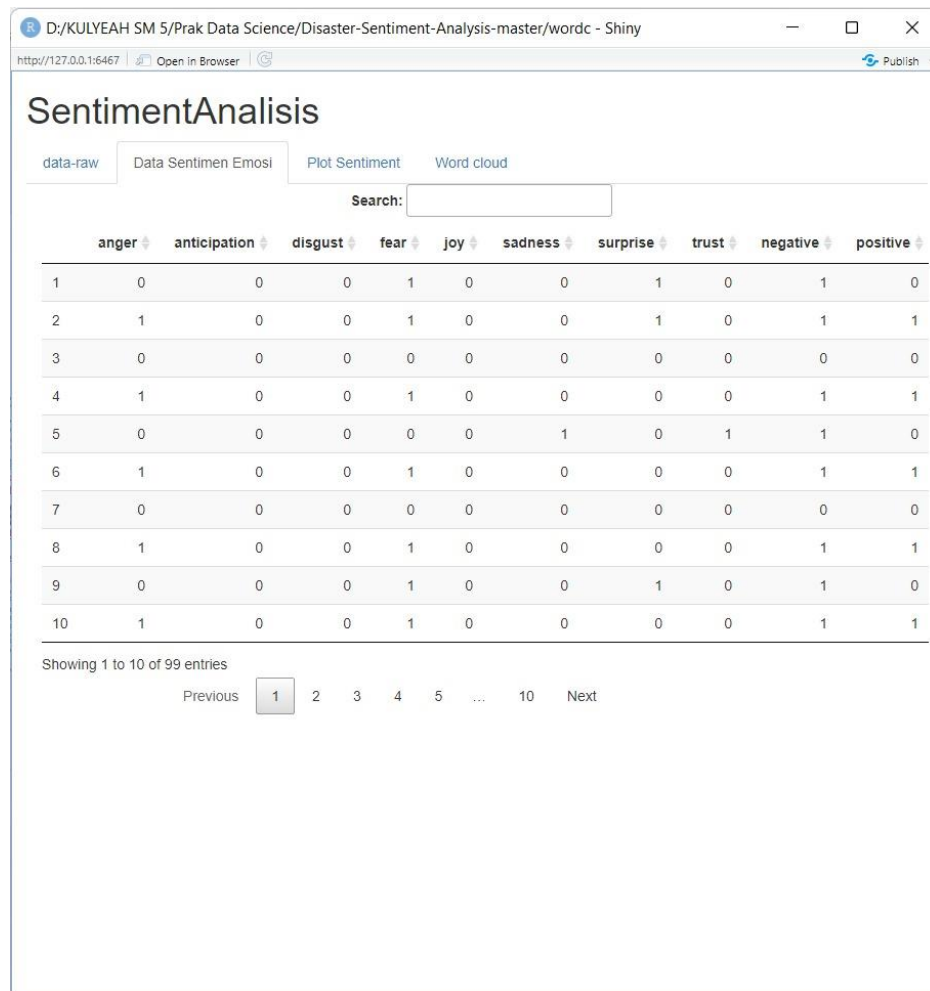
**Tabel 2.1** Menentukan k-Mean

Berdasarkan **Tabel 2.1** digunakan untuk proses pemisahan kata-kata yang berada di twitter.com. Setelah melakukan pengclusteran dengan k-Mean data akan di visualisasi dengan menggunakan wordcloud, yang akan menghasilkan kata yang paling sering muncul dalam tweet pengguna.



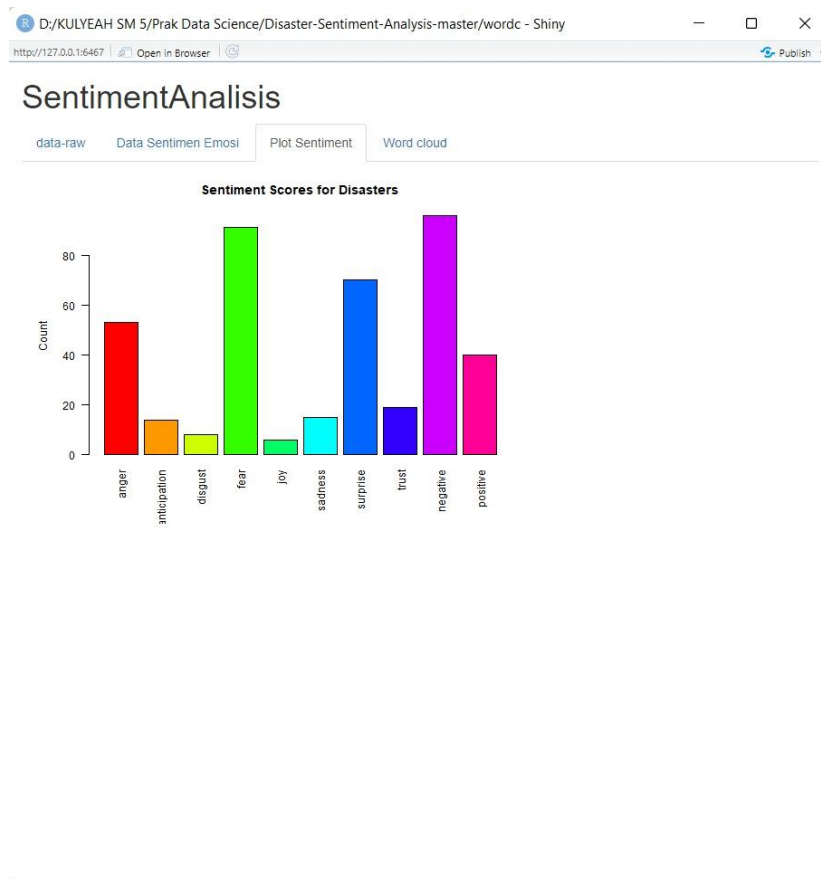
### Gambar 3.1 Tampilan Awal

Tampilan awal halaman ketika program dijalankan terdapat empat fitur disebelah kiri yang digunakan untuk mendapatkan informasi dari data tweet pengguna twitter yang berada di twitter.com yaitu data raw, data sentiment emosi, plot sentiment, dan wordcloud.



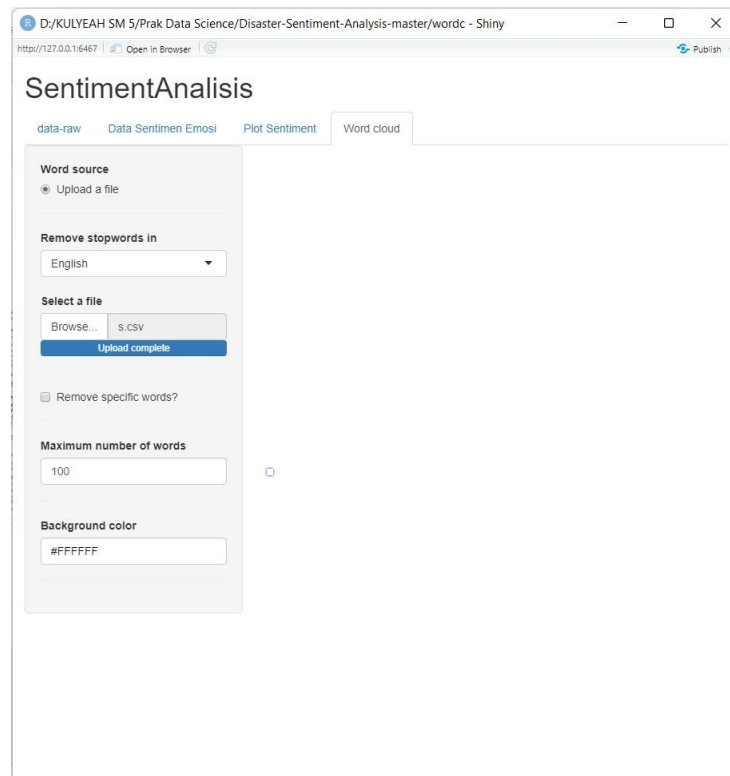
**Gambar 3.2** Data Sentimen Emosi

Tampilan selanjutnya, pembagian emosi yang dihasilkan dari data tweet di raw data dapat dilihat di **Gambar 3.2** terdapat grafik yang digunakan untuk memvisualisasikan sentiment emosi,. Kemudian, di **Gambar 3.3** Plot Sentimen Emosi.



**Gambar 3.3** Plot Sentimen Emosi

Tampilan selanjutnya, plot sentiment emosi dapat dilihat di **Gambar 3.2** tabel yang menjadi dasar dari grafik sentiment emosi ini.



**Gambar 3.4** Wordcloud

Tampilan selanjutnya, memunculkan huruf yang sering keluar di kolom tweet dengan menggunakan wordcloud. Disini kita mengupload file data yang telah dibersihkan baru akan muncul wordcloud bencananya.

#### **4. KESIMPULAN**

Penelitian ini melakukan pengklasifikasi teks dari twitter.com dengan pemisahan kata melalui metode K-Nearest Neighbor, menggunakan 1000 data tweet yang diambil langsung dari twitter. Serta emosi yang telah dianalisis menyimpulkan banyak yang merasa takut dan bersikap negative terhadap bencana alam, yang lainnya ada perasaan sedih, marah dll.