

ANALISIS SENTIMEN UNTUK IDENTIFIKASI BANTUAN KORBAN BENCANA ALAM BERDASARKAN DATA DI TWITTER MENGGUNAKAN METODE K-MEANS DAN NAÏVE BAYES

Vincentius Riandaru Prasetyo^{*1}, Gatum Erlangga², Delta Ardy Prima³

^{1,2,3}Universitas Surabaya, Surabaya

Email: ¹vincent@staff.ubaya.ac.id, ²s160417054@student.ubaya.ac.id, ³delta@staff.ubaya.ac.id

*Penulis Korespondensi

(Naskah masuk: 15 Februari 2023, diterima untuk diterbitkan: 26 September 2023)

Abstrak

Media sosial telah menjadi sarana yang umum bagi orang untuk mengekspresikan diri dan meminta bantuan ketika mereka mengalami musibah. Banyak korban bencana alam di Indonesia menggunakan Twitter untuk meminta bantuan seperti makanan, air bersih, dan lainnya. Penelitian ini bertujuan untuk melakukan analisis sentimen dari data *Twitter* untuk menentukan bantuan bagi korban bencana alam di Indonesia. Pada penelitian ini, metode *K-Means* dan *Naïve Bayes* dikombinasikan untuk melakukan analisis sentimen. Dalam penelitian ini, bantuan yang akan ditemukan adalah pakaian, makanan, air bersih, dan obat. Metode *K-Means* dipilih karena mudah digunakan dan mudah diimplementasikan, sementara metode *Naïve Bayes* digunakan karena menghasilkan nilai akurasi yang baik dalam klasifikasi. Hasil uji coba memperlihatkan bahwa kombinasi *K-Means* dan *Naïve Bayes* menghasilkan akurasi sebesar 76,46%, di mana akurasi tersebut lebih tinggi daripada implementasi *Naïve Bayes* saja, dengan akurasi sebesar 74,65%. Berdasarkan validasi yang dilakukan dengan Kepala Badan Penanggulangan Bencana Daerah (BPBD) di Kota Tarakan, sistem ini dapat membantu BPBD Kota Tarakan dalam memberikan bantuan yang tepat ke lokasi bencana.

Kata kunci: *k-means, naïve bayes, bantuan bencana, analisis sentimen*

SENTIMENT ANALYSIS FOR IDENTIFICATION OF NATURAL DISASTER VICTIMS BASED ON DATA ON TWITTER USING K-MEANS AND NAÏVE BAYES METHODS

Abstract

Social media has become a common place for people to express themselves and ask for help when they are going through a calamity. Many victims of natural disasters in Indonesia use Twitter to request assistance such as food, clean water, and others. Therefore, this study aims to conduct sentiment analysis from Twitter data to determine aid for victims of natural disasters in Indonesia. In this research, *K-Means* and *Naïve Bayes* methods will be combined for sentiment analysis. In this study, the assistance that will be found is clothing, food, clean water, and medicine. The *K-Means* method was chosen because it is easy to use and easy to implement, while the *Naïve Bayes* method was chosen because it has a good level of accuracy in classification. The results showed that the combination of *K-Means* and *Naïve Bayes* had a higher accuracy rate of 76.46%, compared to the use of *Naïve Bayes* alone, which was 74.65%. Based on the validation conducted with the Head of the Regional Disaster Management Agency (BPBD) in Tarakan City, this system can assist the Tarakan City BPBD in providing appropriate assistance to disaster locations.

Keywords: *k-means, naïve bayes, disaster relief, sentiment analysis*

1. PENDAHULUAN

Menurut Undang-Undang Nomor 24 Tahun 2007 Tentang Penanggulangan Bencana, bencana alam merupakan kejadian yang terjadi karena peristiwa alam. Peristiwa alam ini bisa berupa gunung meletus, gempa bumi, tsunami, dan peristiwa alam lainnya. Indonesia berpotensi lebih tinggi mengalami

bencana alam seperti gempa bumi, tsunami dan gunung meletus karena memiliki banyak gunung berapi (Rosyida, et al., 2019). Dampak yang dihasilkan bisa sangat merusak bahkan menyebabkan bencana lain. Kerugian yang dididapatkan dari bencana alam dapat terbilang tidak sedikit. Kerugian dari bencana banjir yang terjadi di Bekasi mencapai

150 milyar rupiah (Hakim, 2019). Selain itu, berdasarkan statistik yang ada, korban meninggal akibat bencana alam pada tahun 2018 mencapai angka 4.000 korban meninggal (Rosyida, et al., 2019). Dampak bencana alam juga dirasakan oleh pemerintah yang menyebabkan sistem pemerintahan lumpuh untuk sementara (Hakim, 2019).

Berbagai upaya pemerintah telah dilakukan untuk mengatasi bencana alam ini. Salah satunya memberikan bantuan kepada masyarakat yang menjadi korban. Tidak hanya pemerintah, relawan dari berbagai kota juga memberikan bantuan dengan melakukan aksi dana kemudian membelikan kebutuhan korban atau bisa langsung menyerahkan dana tersebut. Akan tetapi, tidak jarang bantuan yang diberikan tidak sesuai dengan kebutuhan dari para korban. Akibatnya barang yang tidak digunakan akan dibuang. Dana yang telah dialokasikan juga menjadi tidak tepat. Contohnya pada kasus gempa di Lombok Barat, ditemukan dana sebesar 800 juta rupiah dari bantuan yang diduga tidak tepat sasaran (Suarantb, 2019).

Pada zaman milenial saat ini, informasi dibagikan dengan sangat cepat melalui internet. Contohnya, banyak orang membagikan posting media sosial tentang bencana alam. Mereka membagikan kondisi saat atau sesudah terjadi bencana alam. Berdasarkan dari survei yang dilakukan *Waresosial*, 56% dari jumlah penduduk Indonesia merupakan pengguna aktif media sosial (Katadata, 2019). Dalam kaitannya dengan bencana alam, tak sedikit juga orang yang membagikan pengalaman yang mereka alami saat musibah. Contohnya adalah *tweet* atau post yang berisi tentang permintaan pertolongan kepada pemerintah. Contoh *tweet* dapat dilihat pada gambar 1. Para korban membagikan apa saja yang mereka butuhkan melalui media sosial. Informasi ini bisa sangat bermanfaat untuk menentukan bantuan yang tepat bagi korban bencana. Masalahnya adalah diperlukan waktu yang banyak untuk menganalisa dan mengetahui kebutuhan korban secara manual pada sosial media. Dengan teknik tertentu, data dari media sosial bisa dimanfaatkan untuk mengidentifikasi kebutuhan dari korban bencana alam. Dataset yang digunakan pada penelitian ini berasal dari media sosial *Twitter*. Alasan penggunaan *twitter* adalah karena *twitter* menyediakan fasilitas untuk mengambil data *tweet* dari pengguna secara legal dengan menggunakan *Twitter API*. Oleh sebab itu, diperlukan sebuah cara untuk membantu pemerintah atau relawan dalam menentukan atau mengidentifikasi bantuan bagi para korban bencana alam dengan tepat melalui informasi yang tersebar di sosial media, di mana metode yang dipakai adalah *K-Means* dan *Naïve Bayes*.



Gambar 1. Contoh *Tweet* Meminta Bantuan

K-Means dan *Naïve Bayes* merupakan sebagian metode dari *data mining*. Kedua metode ini dipilih karena sederhana, namun akurasi yang didapatkan juga baik. Masalah ketika mengolah data pada media sosial adalah *noise*, sehingga diperlukan metode yang bisa mengelompokkan data dengan baik. Kelebihan dari metode *K-Means* adalah metode ini bisa mengelompokkan data tanpa label serta fleksibel karena jumlah *cluster* yang bisa disesuaikan dengan kebutuhan (Orkphol & Yang, 2019). Pada penelitian ini, pengelompokkan data *tweet* dari korban bencana alam akan mengimplementasi metode *K-Means*. Misalkan, *tweet* pertama dan kedua membahas kebutuhan makanan dan minuman, maka *tweet* pertama dan kedua akan masuk kedalam kelompok yang sama. Dengan metode ini, diharapkan kebutuhan umum dari para korban bencana akan didapatkan, seperti pangan, sandang, dan lain-lain. Pada penelitian lainnya, metode *K-means* dapat dikombinasikan dengan metode lain dan mendapatkan akurasi hingga 87% (Zul, et al., 2018). Pada penelitian ini, *K-Means* dan metode *Naïve Bayes* dikombinasikan untuk mengklasifikasi jenis bantuan yang lebih spesifik dari korban. Metode *Naïve Bayes* dipilih karena merupakan metode klasifikasi probabilistik yang terkenal. Metode ini juga merupakan metode yang sederhana, namun efisien dengan beragam aplikasi dunia nyata, khususnya klasifikasi teks (Abbas, et al., 2019).

Terdapat penelitian sebelumnya yang juga berfokus pada penggunaan analisis sentimen sebagai penentu bantuan bagi korban bencana alam. Beigi, et al pada penelitiannya, melakukan analisis sentimen untuk membantu pihak berwajib agar dapat menentukan tindakan yang akan diambil berdasarkan analisa yang telah dilakukan di media sosial. Hasil penelitian tersebut sangat membantu dibandingkan dengan survei yang dilakukan secara manual kepada daerah yang cukup sulit untuk dijangkau (Beigi, et al., 2016).

Penelitian lainnya dari Li, et al pada tahun 2018, berusaha mengidentifikasi *Tweet* yang relevan untuk menentukan bantuan bagi korban bencana alam sesuai dengan bencana yang dialami. Identifikasi dilakukan dengan menerapkan metode *Naïve Bayes*. Tahap *preprocessing* yang ditambahkan pada

penelitian ini adalah mengolah *URL*, *hashtag*, *username* serta data yang kembar. Hal tersebut berdampak pada akurasi yang didapatkan semakin meningkat, karena *noise* pada data berkurang (Li, et al., 2018).

Berbeda dengan penelitian sebelumnya, Muppidi, et al., pada tahun 2019 menganalisis tweet untuk mengetahui daerah yang paling terpengaruh oleh bencana alam yang terjadi. Pada penelitian ini, Muppidi, et al., membandingkan beberapa metode machine learning seperti *Naïve Bayes*, *Logistic Regression*, *KNN*, dan *Random Forest* untuk menentukan metode klasifikasi terbaik (Muppidi, et al., 2019).

Selain penelitian Muppidi, et al., penelitian yang dilakukan oleh Karami, et al pada tahun 2020 juga menggunakan data *Tweet* untuk melakukan analisis sentimen sebagai bahan pertimbangan untuk membantu meningkatkan *situational awareness*. Pada penelitian tersebut, sistem yang dikembangkan mencari topik-topik yang sedang dibahas pada media sosial untuk melihat besar kerusakan, kerugian, serta situasi yang diakibatkan oleh bencana alam (Karami, et al., 2020).

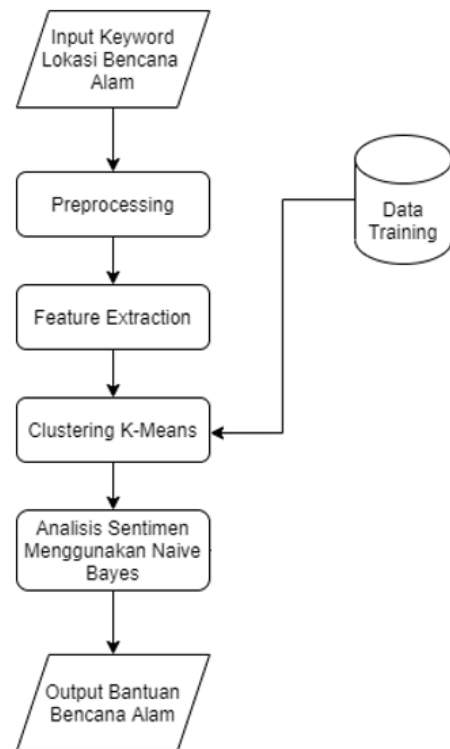
Penelitian lainnya yang dilakukan oleh Ruz, et al., pada tahun 2020 melakukan analisis sentimen dengan menggunakan data *tweet* terkait gempa di Chili pada tahun 2010. Penelitian ini bertujuan untuk menguji efektifitas metode *SVM* dan *Random Forest* untuk analisis sentimen pada data *tweet* berbahasa Spanyol. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi hubungan antar kata untuk klasifikasi sentimen di Twitter menggunakan jaringan Bayesian (Ruz, et al., 2020).

Berbeda dengan penelitian yang dilakukan oleh Beigi, et al., Muppidi, et al., Karami, et al., dan Ruz, et al., penelitian ini bertujuan untuk mengkombinasikan metode *K-Means* dan *Naïve Bayes* untuk membantu pihak terkait dalam mengidentifikasi bantuan korban bencana alam berdasarkan data *tweet*. Penelitian yang dilakukan oleh Li, et al juga menggunakan data *tweet* untuk menentukan bantuan bagi korban bencana alam. Akan tetapi, perbedaan dengan penelitian ini adalah pada penelitian Li, et al., luaran yang dihasilkan yaitu apakah suatu *tweet* relevan dengan bantuan bencana alam atau tidak. Sedangkan untuk penentuan jenis bantuannya masih dilakukan secara manual oleh pengguna. Berbeda dengan penelitian Li, et al., pada penelitian ini luaran yang dihasilkan adalah suatu *tweet* termasuk dalam kategori yang membutuhkan bantuan makanan, pakaian, air bersih, obat-obatan, atau lain-lainnya.

2. METODE PENELITIAN

Secara umum, sistem yang dikembangkan pada penelitian ini dibagi menjadi beberapa tahapan, seperti yang ditunjukkan pada gambar 1. Sistem menerima *input* dari pengguna berupa nama bencana dan lokasi dari bencana tersebut. Selanjutnya sistem

secara otomatis melakukan pencarian data *Tweet* bencana alam sesuai dengan *input* dari *user* menggunakan teknik *crawling*.



Gambar 2. Tahapan Penelitian

Sebelum data diolah menggunakan metode *K-Means* dan *Naïve Bayes*, data akan mengalami *preprocessing* yang meliputi penghapusan *URL*, penghapusan *username*, pemrosesan *hashtag*, *stemming*, *stopword removal*, dan penghapusan duplikasi data. Setelah data mengalami *preprocessing*, sistem akan melakukan *feature extraction* dengan mengimplementasikan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*).

Hasil *feature extraction* yang didapatkan, kemudian dikelompokkan menggunakan metode *K-Means* untuk membantu dalam proses pelabelan *class*. Setelah itu, data yang sudah dikelompokkan, akan diklasifikasikan menggunakan metode *Naïve Bayes* bersama *data training* yang sudah tersedia sebelumnya untuk mendapatkan hasil yang lebih spesifik, yaitu bantuan yang cocok bagi korban bencana. *Output* yang dihasilkan berupa rekomendasi bantuan untuk korban bencana alam. Pada penelitian ini, landasan teori yang digunakan antara lain: analisis sentimen, *crawling*, *preprocessing*, ekstraksi fitur, *K-Means*, dan *Naïve Bayes*.

2.1. Analisis Sentimen

Analisis sentimen atau bisa disebut *opinion mining*, adalah salah satu bidang ilmu *Natural Language Processing* (NLP) yang bertujuan untuk

menganalisis sentimen dan opini dari suatu teks. Analisis sentimen berusaha menganalisis opini, sentimen, sikap, persepsi dari seseorang, terhadap suatu topik, produk, layanan, dan sebagainya. Evolusi cepat aplikasi berbasis Internet seperti situs web, jejaring sosial, dan blog, mengarahkan orang untuk menghasilkan banyak sekali opini dan ulasan tentang produk, layanan, dan aktivitas sehari-hari. Analisis sentimen berfungsi sebagai alat yang ampuh bagi bisnis, pemerintah, dan peneliti untuk mengekstraksi dan menganalisis pandangan publik, untuk mendapatkan informasi, dan bahan pertimbangan untuk pengambilan keputusan. Metode yang bisa digunakan untuk melakukan analisis sentiment dapat beraneka ragam. Ada juga peneliti yang menggabungkan beberapa metode untuk mendapatkan akurasi yang lebih baik (Birjali, et al., 2021).

2.2. Crawling

Crawling atau *web scrapping* adalah sebuah proses yang bertujuan untuk mendapatkan data secara otomatis yang digunakan untuk analisis lebih lanjut. Data bisa didapatkan dari *website* yang bersifat publik atau dapat diakses secara bebas. Setelah masuk ke dalam *web* tersebut, aplikasi *crawling* mengambil data pada struktur HTML tertentu (Prasetyo, et al., 2022). *Library* yang dapat digunakan untuk melakukan *crawling* salah satunya adalah *Simple HTML DOM*.

Simple HTML DOM merupakan *library* yang diterapkan dalam penelitian ini. *Library* ini dapat dijalankan menggunakan bahasa pemrograman PHP. Untuk menggunakannya, pengguna perlu mengetahui target tujuan *website* yang akan diambil datanya. Selanjutnya, pengguna juga harus melihat bagian apa yang akan diambil. Bagian yang akan diambil bisa terdapat pada *tag html* tertentu, misalkan `<div>` atau yang lainnya. Agar hasil yang didapatkan lebih detail, maka *class* atau *id* dari *tag* yang berisi informasi tersebut dapat disertakan (Gunawan, et al., 2018).

2.3. Preprocessing

Langkah *preprocessing* bertujuan untuk memfilter informasi yang tidak relevan dalam suatu data teks. Dalam penelitian ini, proses *preprocessing* melibatkan tiga tahap, yaitu: penghapusan simbol dan URL, proses *stemming*, dan penghapusan *stopword* dari suatu *tweet*. Langkah pertama adalah menghapus simbol dan URL yang terdapat dalam suatu kalimat. Tujuannya adalah untuk membuang karakter non verbal yang dianggap tidak penting bagi proses selanjutnya, seperti tanda baca dan simbol-simbol lain, seperti (&), atau (/), dan "@" (Prasetyo, et al., 2022).

Langkah kedua adalah proses *stemming*, yaitu membuat kembali kata menjadi bentuk aslinya, sehingga meminimalisasi duplikasi kata yang memiliki makna yang sama. Algoritma Nazief dan

Andriani digunakan dalam penelitian ini untuk mengubah kata menjadi akar kata. Algoritma ini sering digunakan untuk memproses data, terutama data teks yang berbahasa Indonesia (Prasetyo, et al., 2019). Langkah terakhir adalah penghapusan *stopword*, yaitu membuang kata-kata yang kurang berpengaruh pada pemrosesan data selanjutnya, seperti konjungsi, preposisi, dan kata tanya. (Prasetyo & Samudra, 2021).

2.4. Ekstraksi Fitur

Feature extraction adalah proses untuk mengambil *term* atau atribut dari suatu data, dalam hal ini adalah dokumen teks. *Term* atau atribut dari setiap dokumen teks adalah kata. Proses ekstraksi fitur ini diawali dengan melakukan tokenisasi atau memecah dokumen, dalam hal ini adalah *tweet*, menjadi beberapa kata. Setelah dilakukan tokenisasi, kata-kata tersebut akan diubah menjadi suatu nilai yang dapat dihitung secara statistik. *Term Frequency – Inverse Document Frequency* (TF-IDF) merupakan salah satu metode yang umum digunakan untuk melakukan proses ekstraksi fitur (Jimmy & Prasetyo, 2021).

Selain itu, TF-IDF juga merupakan salah satu cara untuk menentukan kepentingan sebuah kata dalam sebuah dokumen dalam kumpulan dokumen atau korpus. Metode tersebut memperhitungkan frekuensi kata dalam dokumen yang ditargetkan, serta frekuensi kata-kata yang muncul di dokumen lain dalam kumpulan tersebut. Persamaan (1) memperlihatkan rumus perhitungan dari metode TF-IDF (Dang, et al., 2020).

$$Tf_{t,d}.Idf_t = Tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

Di mana *Tf* merupakan jumlah kemunculan suatu kata atau *term* pada suatu dokumen. Untuk *df* adalah banyaknya dokumen yang terdapat suatu kata atau *term* tertentu, sedangkan *N* merupakan jumlah dokumen pada dataset.

2.5. K-Means

K-Means merupakan algoritma atau metode yang digunakan untuk mengklasterkan data berdasarkan kedekatan yang dimiliki. Data yang dikelompokkan adalah data yang belum mempunyai kelompok atau kelas tertentu. Metode ini menerima *input*, yaitu jumlah kelas atau kelompok (*k*) dan data. Hal ini membuat *k-means* dapat bersifat fleksibel dalam menentukan jumlah kelas sesuai dengan kebutuhan (Prasetyo, 2018).

Cara kerja *k-means* diawali dengan menentukan *centroid*, kemudian menghitung jarak *centroid* dengan setiap data, dan memberikan nama kelas. Selanjutnya adalah menghitung *mean* dari anggota kelas dan menggeser *centroid* berdasarkan *mean* dari kelas tersebut. Proses tersebut akan dilakukan berkali-kali hingga tidak ada lagi perubahan *centroid*

(Prasetyo, et al., 2022). Untuk perhitungan jarak *centroid* dengan setiap data, metode yang umum digunakan adalah *Eulidean* yang dihitung dengan persamaan (2).

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Di mana x merupakan TF-IDF dari *centroid* dan y adalah nilai TF-IDF dari data yang ingin dihitung jaraknya. Metode *Euclidean* dipilih karena mudah dalam hal implementasi dan memiliki akurasi yang baik (Patel & Upadhyay, 2020). Pada penelitian ini, nilai K yang digunakan pada metode K-Means adalah 5. Nilai K tersebut ditentukan berdasarkan hasil diskusi dengan Bapak Ir. Kadjat Prasetyo Widodo, M.P. selaku Kepala Bidang Kedaruratan, Logistik, Rehabilitasi dan Rekonstruksi BPBD Kota Tarakan. Dari hasil diskusi tersebut, terdapat 5 kategori bantuan bencana yang akan diidentifikasi antara lain: makanan, pakaian, air bersih, obat-obatan, dan lain-lainnya. Oleh karena itu, nilai K yang digunakan pada implementasi metode K-Means juga 5.

Pada penelitian ini, metode K-Means digunakan untuk membantu proses pelabelan *tweet* ke dalam lima kategori yang telah disebutkan sebelumnya. Hasil dari proses *clustering* dengan K-Means akan menghasilkan lima *cluster* yang berbeda, di mana dari tiap *cluster* tersebut akan diambil kata kunci dari tiap *tweet* anggota *cluster*, untuk ditentukan label dari *cluster* tersebut.

Sebagai contoh, *cluster* pertama memiliki 100 *tweet* sebagai anggotanya, kemudian dari 100 *tweet* tersebut akan diambil kata kunci untuk tiap *tweet* untuk selanjutnya dikategorikan ke salah satu dari kelima kategori yang ada. Sebagai contoh, apabila terdapat kata kunci seperti “makanan”, “lapar” dan “kelaparan” yang mengindikasikan kebutuhan makanan, maka *tweet* tersebut dikategorikan ke dalam “makanan”. Selanjutnya, apabila *tweet* mengandung kata kunci “air bersih” dan “mandi” atau kata kunci lain yang mengindikasikan kebutuhan air bersih, maka *tweet* akan dikategorikan ke dalam “air bersih”, dan seterusnya. Setelah 100 *tweet* tersebut mendapatkan kategori, sistem akan menentukan kategori mana yang paling dominan dari 100 *tweet* tersebut. Misalkan 70 dari 100 *tweet* dikategorikan ke dalam “makanan”, maka semua *tweet* yang ada pada *cluster* pertama akan dilabeli dengan “makanan”. Proses ini berlaku untuk *cluster-cluster* lainnya. Hasil pelabelan data dengan metode K-Means inilah yang akan digunakan oleh *Naïve Bayes* sebagai data training untuk membentuk model klasifikasi.

2.6. Naïve Bayes

Naïve Bayes merupakan salah satu metode yang termasuk dalam *supervised learning*, di mana metode tersebut menggunakan teorema *Bayes*. Untuk mengklasifikasikan sebuah data baru, diperlukan data

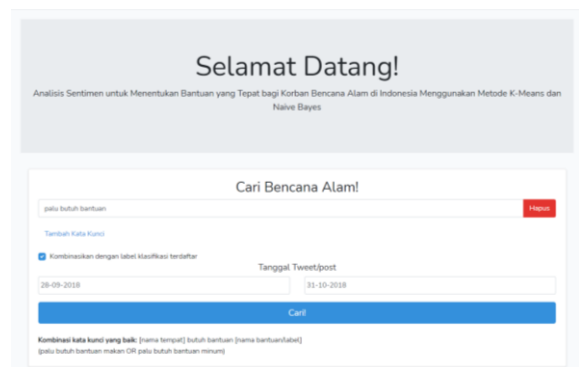
training yang sudah diberi *label* sebelumnya. Metode ini menggunakan konsep probabilitas untuk mengklasifikasikan suatu dokumen. Terdapat 2 cara untuk melakukan klasifikasi berdasarkan jenis data pada suatu atribut, yaitu numerik dan non-numerik. Pada penelitian ini, metode *Naïve Bayes* akan menghitung data numerik, dalam hal ini adalah menggunakan nilai TF-IDF yang didapatkan dari proses *feature extraction* (Negara, et al., 2020).

Untuk menghitung data numerik, ada dua data yang diperlukan, yaitu rata-rata fitur tiap kelas (μ_{C_i}) beserta standar deviasinya (σ_{C_i}). Setelah itu, suatu data akan ke dalam kelas tertentu ($P(x|C_i)$) ditentukan melalui persamaan (3) (Ramadhan & Adhinata, 2022).

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

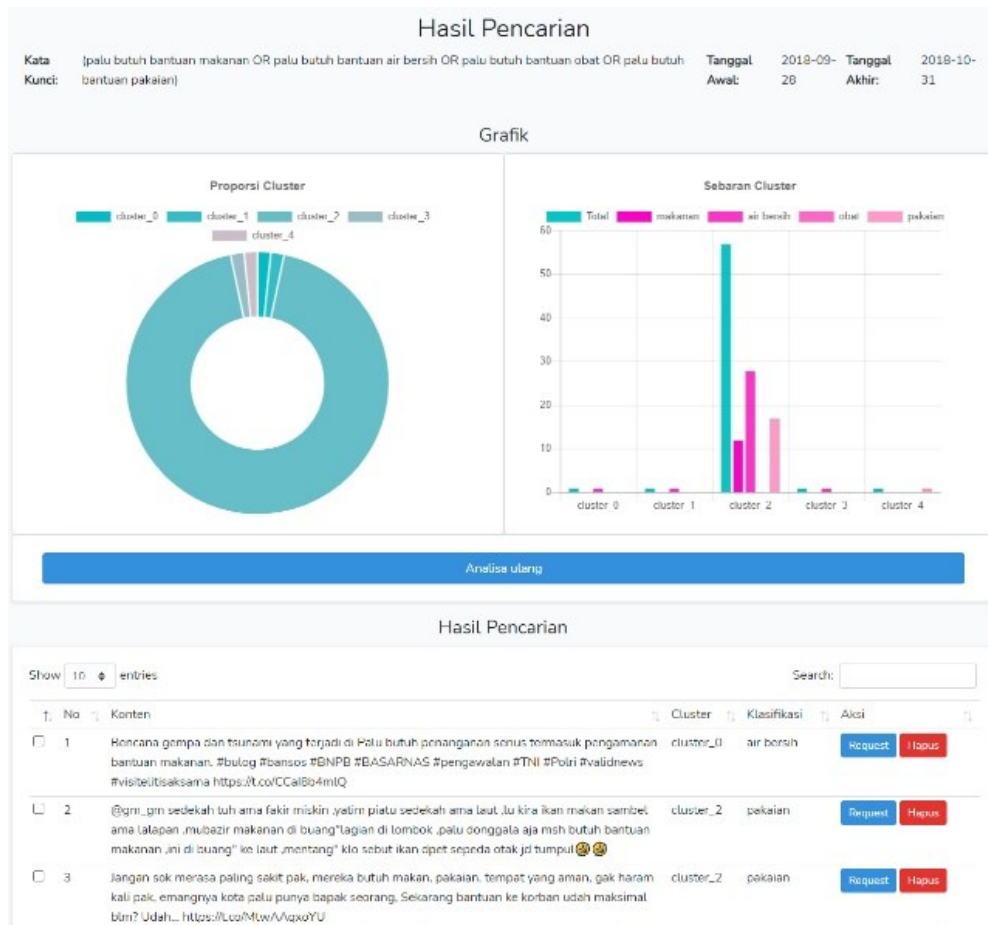
3. HASIL DAN PEMBAHASAN

Sub bab ini dijabarkan uji coba dan evaluasi yang dilakukan pada sistem yang telah dikembangkan. Gambar 3 menunjukkan tampilan awal dari sistem yang dikembangkan. Pada awal penggunaan sistem, pengguna harus memasukan beberapa informasi yaitu kata kunci bencana alam, *checkbox* melakukan kombinasi kata kunci dengan label klasifikasi, dan yang terakhir adalah tanggal *tweet*. Tanggal dan *checkbox* kombinasi bersifat opsional. Apabila pengguna tidak memasukan tanggal, maka *tweet* yang akan dicari adalah *tweet* yang memiliki tanggal terdekat.



Gambar 3. Tampilan Awal Sistem

Pada gambar 3 dapat dilihat bahwa contoh kata kunci yang dimasukkan adalah “palu butuh bantuan”. Untuk mendapatkan hasil yang lebih spesifik, tanggal bencana di Palu juga dimasukkan, yaitu 28 September 2018 dan mencari hingga tanggal 31 Oktober 2018. Setelah tombol “Cari” ditekan, sistem akan melakukan *crawling* data *tweet* dan analisis sentimen secara otomatis. Apabila proses *crawling* dan analisis selesai dilakukan, maka tampilan yang disajikan ke *user* dapat dilihat seperti contoh pada gambar 4.



Gambar 4. Tampilan Hasil Analisis Sentimen

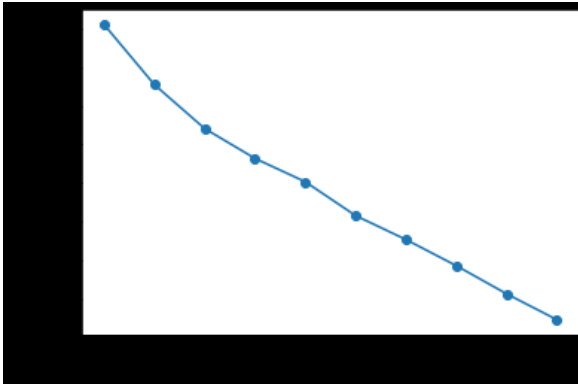
Selain melakukan uji coba fungsionalitas sistem, penelitian ini juga mengukur akurasi dari kombinasi metode *K-Means* dan *Naïve Bayes* yang diimplementasikan pada sistem. Selain itu, akurasi dari kombinasi tersebut juga dibandingkan dengan penerapan metode *Naïve Bayes* sendiri, tanpa menggunakan *K-Means* dalam melakukan analisis sentimen. Metode yang digunakan untuk mengukur akurasi tersebut adalah *10-fold cross validation*. Metode ini akan mengelompokkan sampel asli secara acak menjadi 10 subsampel berdasarkan ukuran yang sama. *Dataset* uji akan dianggap sebagai *dataset* yang tidak diketahui kelasnya, dan *dataset training* mewakili *dataset* yang sudah memiliki label *class*. Proses *cross-validation* akan diulangi sebanyak 10 kali untuk mendapatkan rata-rata akurasi dari model klasifikasi yang berbeda-beda (Jung, et al., 2020). Tabel 1 menunjukkan hasil perbandingan dari pengukuran akurasi yang dilakukan dengan *10-fold cross validation*.

Berdasarkan hasil perbandingan akurasi pada tabel 1, dapat dilihat bahwa kombinasi antara metode *K-Means* dengan *Naïve Bayes* mendapatkan akurasi yang lebih tinggi untuk analisis sentimen, dibandingkan dengan penerapan metode *Naïve Bayes* sendiri. Selain itu, untuk membuktikan pengaruh nilai *K* yang optimal pada metode *K-Means*, akan digunakan metode *elbow*.

Tabel 1. Hasil Perbandingan Akurasi

Iterasi	Naïve Bayes	K-Means + Naïve Bayes
1	66,67%	91,32%
2	69,70%	62,89%
3	77,78%	66,41%
4	77,78%	64,79%
5	77,78%	94,21%
6	77,78%	69,47%
7	66,67%	67,48%
8	74,74%	83,68%
9	78,79%	69,79%
10	78,79%	94,53%
Rata-rata	74,65%	76,46%

Metode *elbow* adalah metode yang umum digunakan untuk mencari nilai *K* terbaik pada metode *K-Means*. Dengan metode ini, dapat diketahui seberapa jauh titik pada setiap kluster terhadap pusat klasternya, di mana semakin kecil jarak tersebut, semakin baik kluster yang dibentuk (Ramduge & Rajhans, 2023). Pada kasus ini, gambar 5 menunjukkan hasil perhitungan *within cluster sum of squares* atau *WCSS*. Percobaan dilakukan dengan beberapa nilai *K* yang terbentang dari 2 hingga 10. Kluster kemudian dibentuk pada data *training* yang berjumlah 355 dan memiliki total fitur sebanyak 2315 sesuai dengan masing-masing *K* yang ada.



Gambar 5. Hasil Perhitungan WCSS Pada Setiap Kluster

Berdasarkan hasil percobaan yang dilakukan sebelumnya, tidak ditemukan nilai K yang optimal, karena tidak terbentuk sudut siku pada gambar 5. Hal ini disebabkan oleh tingginya dimensi atau jumlah fitur yang ada. Jumlah fitur yang besar membuat algoritma K -Means semakin sulit untuk mencari data yang relevan saat melakukan pengelompokan (Xu et al., 2020). Tingginya jumlah fitur disebabkan oleh masih ditemukan *slangword* dan singkatan (akronim) yang bervariasi dan belum bisa diproses oleh sistem, contohnya adalah kata “makanan” dapat menjadi “mkn” dan “mkanan”.

Selain melakukan pengukuran akurasi, pada penelitian ini juga dilakukan evaluasi bersama Bapak Ir. Kadjat Prasetyo Widodo, M.P. selaku Kepala Bidang Kedaruratan, Logistik, Rehabilitasi dan Rekonstruksi BPBD Kota Tarakan sebagai narasumber. Proses evaluasi dilakukan dengan mendemokan penggunaan aplikasi secara langsung, kemudian narasumber akan memberikan penilaian dan komentar mengenai sistem yang dibangun. Narasumber merasa sistem yang dibangun sudah cukup baik untuk mendapatkan gambaran data secara umum bantuan yang dibutuhkan. Pada kasus ini, bantuan yang bisa diklasifikasikan adalah makanan, pakaian, air bersih dan obat-obatan.

Selain itu, narasumber memberikan saran yang dapat digunakan untuk mengembangkan sistem yang telah dibuat. Saran pertama, sistem diharapkan dapat memilah informasi dengan lebih baik lagi, untuk memisahkan informasi palsu atau tidak. Saran berikutnya, perlu adanya label klasifikasi tambahan lainnya, sebagai contoh adalah tempat ibadah, kebutuhan psikologi, dan lainnya.

4. KESIMPULAN

Berdasarkan hasil evaluasi dan uji coba yang dilakukan, dapat ditarik kesimpulan bahwa implementasi metode K -Means dikombinasikan dengan metode *Naïve Bayes*, menghasilkan akurasi lebih baik daripada implementasi metode *Naïve Bayes* saja, dengan akurasi sebesar 76,46%. Selain itu, menurut Kepala Bidang Kedaruratan, Logistik, Rehabilitasi dan Rekonstruksi BPBD Kota Tarakan, sistem yang dikembangkan sudah cukup baik dalam membantu pihak BPBD untuk mendapatkan

informasi awal terkait bantuan yang harus diberikan kepada para korban bencana.

DAFTAR PUSTAKA

- ABBAS, M., ALI, K., JAMALI, A. A., MEMON, S. DAN AHMED, A., 2019. Multinomial Naive Bayes Classification Model for Sentiment Analysis. *International Journal of Computer Science and Network Security*, 19(3), pp. 62-67.
- BEIGI, G., HU, X., MACIEJEWSKI, R. dan LIU, H., 2016. An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief. *Sentiment Analysis and Ontology Engineering*, Volume 639, pp. 313-340.
- BIRJALI, M., KASRI, M. dan HSSANE, A. B., 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, Volume 226.
- DANG, N. C., MORENO-GARCÍA, M. dan PRIETA, F. D. L., 2020. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, Volume 9, pp. 1-29.
- GUNAWAN, R., RAHMATULLOH, A., DARMAWAN, I. dan FIRDAUS, F., 2018. *Comparison of Web Scraping Techniques : Regular Expression, HTML DOM and Xpath*. Yogyakarta, Atlantis Press.
- HAKIM, L., 2019. Kerangka Kerja Kesiapan Menghadapi Bencana. *Jurnal Dialog Penanggulangan Bencana*, 10(1), pp. 1-11.
- JIMMY dan PRASETYO, V. R., 2021. *Sentiment Analysis on Feedback of Higher Education Teaching Conduct An Empirical Evaluation of Methods*. Surabaya, AIP Publishing.
- JUNG, K., BAE, D. H., UM, M. J., KIM, S., JEON, S. DAN PARK, D., 2020. Evaluation of Nitrate Load Estimations Using Neural Networks and Canonical Correlation Analysis with K-Fold Cross-Validation. *Sustainability*, 12(1), pp. 1-17.
- KARAMI, A., SHAH, V., VAEZI, R. dan BANSAL, A., 2020. Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science*, 46(3), pp. 313-324.
- KATADATA, 2019. *Databooks*. [Online] Available at: <https://databoks.katadata.co.id/datapublish/2019/02/08/berapa-pengguna-media-sosial-indonesia> [Accessed 15 Januari 2023].
- LI, H., CARAGEA, D., CARAGEA, C. dan HERNDON, N., 2018. Disaster response aided by tweet classification with a domain adaptation approach. *Journal of Contingencies and Crisis Management*, Volume 26, pp. 16-27.
- MUPPIDI, S., RAO, P. S. dan MURTHY, M. R. K., 2019. Identification of Natural Disaster

- Affected Area Using Twitter. *Advances in Decision Sciences, Image Processing, Security and Computer Vision*, Volume 1, pp. 792-801
- NEGARA, A. B. P., MUHARDI, H. dan PUTRI, I. M., 2020. Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), pp. 599-606.
- ORKPHOL, K. dan YANG, W., 2019. Sentiment Analysis on Microblogging with K-Means Clustering and Artificial Bee Colony. *International Journal of Computational Intelligence and Applications*, 18(3).
- PATEL, S. P. dan UPADHYAY, S. H., 2020. Euclidean distance based feature ranking and subset selection for bearing fault diagnosis. *Expert Systems With Applications*, Volume 154, pp. 1-16.
- PRASETYO, V. R., 2018. *Searching cheapest product on three different e-commerce using k-means algorithm*. Bali, IEEE Xplore.
- PRASETYO, V. R., AXEL, S., SOEBROTO, J. T., SUGIARTO, D., WINATAN, S. A. DAN NJUDANG, S. D., 2022. Prediksi Harga Emas Berdasarkan Data gold.org menggunakan Metode Long Short Term Memory. *Jurnal SISTEMASI*, 11(3), pp. 623-629.
- PRASETYO, V. R., HARTANTO, B. dan MULYONO, A. A., 2019. Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient. *Jurnal TEKNIKA*, 8(1), pp. 44-51.
- PRASETYO, V. R., MIRANTI, F. A. dan LIMANTO, S., 2022. *Implementation of Feature Selection to Reduce the Number of Features in Determining the Initial Centroid of K-Means Algorithm*. Yogyakarta, IEEE Xplore.
- PRASETYO, V. R. dan SAMUDRA, A. H., 2021. *Hate speech content detection system on Twitter using K-nearest neighbor method*. Surabaya, AIP Publishing.
- PRASETYO, V. R., WIDIASRI, M. dan ANGKIRIWANG, M. M., 2022. Sistem Berbasis Web Untuk Koreksi Soal Esai Dengan Association Rules. *Jurnal TEKNIKA*, 11(1), pp. 62-68.
- RAMADHAN, N. G. dan ADHINATA, F. D., 2022. Sentiment analysis on vaccine COVID-19 using word count and Gaussian Naïve Bayes. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(3), pp. 1765-1772.
- RAMDUGE, K. D. dan RAJHANS, N. R., 2023. K-means clustering for optimization of spare parts delivery. *Management Science Letters*, 13, pp. 235-240.
- ROSYIDA, A., NURMASARI, R. dan SUPRAPTO, 2019. Analisis Perbandingan Dampak Kejadian Bencana Hidrometeorologi Dan Geologi Di Indonesia Dilihat Dari Jumlah Korban Dan Kerusakan (Studi: Data Kejadian Bencana Indonesia 2018). *Jurnal Dialog Penanggulangan Bencana*, 10(1), pp. 12-21.
- RUZ, G. A., HENRIQUEZ, P. A. dan MASCARENO, A., 2020. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, Volume 106, pp. 92-104.
- SUARANTB, 2019. *SUARANTB.com*. [Online] Available at: <https://www.suarantb.com/2019/05/08/ratus-an-juta-bantuan-gempa-di-lobar-diduga-salah-sasaran/> [Accessed 15 Januari 2023].
- XU, H., YAO, S., LI, Q. dan YE, Z., 2020. *An Improved K-means Clustering Algorithm*. Dortmund, IEEE Explore.
- ZUL, M. I., YULIA, F. dan NURMALASATI, D., 2018. *Social Media Sentiment Analysis Using K-Means and Naïve Bayes Algorithm*. Batam, IEEE Explore.