Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia

Endah Nofiyanti ¹⁾, Erry Maricha Oki Nur Haryanto ²⁾

1, ²⁾ Program Studi Teknik Informatika, Fakultas Teknik, Universitas Janabadra, Yogyakarta

1) endahnofiyanti@gmail.com, ²⁾ errymaricha@janabadra.ac.id

ABSTRACT

Disasters, whether natural disaster, non-natural disaster or from human error, causes fatalities, environmental damages, property losses, or psychological impact, especially in Indonesia. The National Agency for Disaster Countermeasure (BNPB) is Indonesian board for national natural disaster affairs. For each region, there is Badan Penanggulangan Bencana Daerah (BPBD) in Indonesia as the regional disaster board. According to BNPB, there are 1441 natural disasters occurred in 2021, including Covid-19 pandemic. The disaster management has become frontline agent for those disasters. Meanwhile, social media has become a part of everyday's life for people nowdays to get the newest information, especially about disasters surroundings. The purpose of this research was to find out the public reaction, by classifying positive, neutral or negative opinion to the disaster management in Indonesia from Twitter. A text mining method from Natural Processing Language (NLP) was sentiment analysis which is TextBlob from python as text mining algorithm. Sentiment analysis was applied to analyze data by public opinion as the decisionmaking support. Data collection from Twitter by using API were preprocessed before processing data, analyzing and visualizing. Based on the research, there were 23,53 % positive tweets, 57,35 % neutral tweets and 19,12 % negative tweets. From the result, mostly Indonesian have neutral opinion about the disaster management.

Keywords: sentiment analysis, python, text mining, twitter, bpbd.

I. PENDAHULUAN

Bencana adalah peristiwa atau rangkaian peristiwa yang mengancam dan mengganggu kehidupan dan penghidupan masyarakat yang disebabkan, baik oleh faktor alam dan/atau faktor nonalam maupun faktor manusia sehingga mengakibatkan timbulnya korban jiwa manusia, kerusakan lingkungan, kerugian harta benda, dan dampak psikologis (BNPB, 2007).

Menurut Undang-Undang Republik Indonesia Nomor 24 Tahun 2007 tentang Penanggulangan Bencana Pasal 21 (Pemerintah Republik Indonesia, 2007, p. 12), Badan Penanggulangan Bencana Daerah memiliki tugas usaha penanggulangan bencana yang mencakup pencegahan bencana, penanganan darurat, rehabilitasi, serta rekonstruksi secara adil dan setara.

Seperti diketahui bencana sudah menjadi hal yang acap terjadi di seluruh wilayah Indonesia. Menurut data BNPB, ada 1441 bencana alam yang melanda Indonesia terhitung dari 1 Januari hingga Juni 2021. Bencana yang sering terjadi diantaranya banjir, puting beliung, tanah longsor, kebakaran hutan dan lahan, gempa bumi, gelombang pasang dan abrasi, kekeringan dan lain sebagainya. Terlebih lagi pandemi Covid-19 yang masih melanda. BPBD pun juga berperan menjadi garda terdepan pencegahan dan penanganan Covid-19.

Twitter merupakan platform media sosial yang digunakan penggunanya untuk berkomunikasi dan menyebarkan informasi berupa tweet. Menurut Kominfo, dari 245 juta penduduk Indonesia, pengguna internet di Indonesia mencapai 55 juta orang. Ini menjadikan Indonesia sebagai negara kelima terbesar pengguna Twitter.

Natural Processing Language (NLP) adalah cabang dari Artificial Intelligence yang berhubungan dengan interaksi antara mesin dan manusia menggunakan bahasa natural. Dalam hal ini memanfaatkan salah satu library python yaitu TextBlob untuk mengerjakan algoritma text mining. Salah satu metode dari text mining di bidang NLP adalah dengan pendekatan analisis sentimen. Pendekatan analisis sentimen digunakan untuk menganalisa data berbentuk opini publik sebagai pendukung pengambilan keputusan. Tugas analisis sentimen adalah mengelompokkan kumpulan polaritas dari teks yang berada dalam dokumentasi, kalimat atau fitur entitas dengan tingkat aspek yang bersifat positif, netral atau negatif. (Buntoro, 2017; Fridom Mailo & Lazuardi, 2019; Yasah, 2020).

Berdasarkan uraian di atas, maka penulis akan menerapkan analisis sentimen terhadap penanggulangan bencana, yang dalam hal ini mengacu pada Badan Penanggulangan Bencana Daerah (BPBD), berdasarkan dari opini publik dengan kategori positif, netral atau negatif dari media sosial Twitter, dengan menggunakan *library* Python. Hasil dari analisis sentimen ini dapat digunakan oleh pengemban tugas penanggulangan bencana untuk mengevaluasi penanganan bencana yang terjadi agar selanjutnya bisa lebih baik lagi.

II. TINJAUAN PUSTAKA

2.1. Text Mining

Menurut Ronen Feldman dan James Sanger dalam bukunya "The Mining HandBook : Advanced Approaches in Analyzing Unstructured Data" (Feldman & Sanger, 2007), text mining didefinisikan secara luas sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining yang salah satunya adalah kategorisasi. Text mining bisa dianggap subjek riset yang tergolong baru. Text mining dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian atau pengelompokan dan menganalisa unstructured text dalam jumlah besar.

Salah satu bidang yang terkait dengan *text mining* yaitu *Natural Processing Language* (NLP), yang dalam hal ini dikerjakan oleh salah satu *library* Python, TextBlob. Penelitian yang dilakukan oleh (Fauziyyah, 2020) tentang analisis sentimen pandemi Covid-19 pada *streaming* Twitter dengan *text mining* python, dilakukan dengan memanfaatkan ekstraksi fitur untuk hubungan kata-kata dengan metode Bigram dan Ngram yang selanjutnya digunakan untuk analisis sentimen berdasarkan dari opini dengan mengklasifikasikan perasaan individu seperti sedih, senang, atau panik menghadapi pandemi Covid-19 ke dalam level sentimen yaitu negatif, positif atau netral. Namun hasil *streaming* data tweet masih menggunakan bahasa Inggris. Dari paparan tersebut, penulis ingin mengembangkan penelitian menggunakan bahasa Indonesia dengan memanfaatkan fungsi *translate* dari TextBlob.

Dalam *text mining*, tahapan yang dilakukan diantaranya pengambilan data, *pre-processing* atau praproses sebelum dilakukan klasifikasi, baru selanjutnya data diproses. Data bersih kemudian dapat divisualisasikan (Negara et al., 2016).

2.2. Twitter API

Platfrom media sosial Twitter menyediakan API (*Application Programming Interface*) yang memudahkan setiap orang untuk mengambil data dari Twitter (Pintoko & L., 2018). Pengumpulan data dari twitter digunakan untuk mengetahui opini publik tentang penanggulangan bencana (BPBD) di Indonesia.

2.3. Python

Python merupakan suatu perangkat lunak *open source* yang terkenal karena bahasa pemogramannya yang dinamis. Python yang digunakan yaitu versi 3.9. Ada beberapa

library yang digunakan pada penulisan ini. Untuk menginstall *library*, cukup menggunakan perintah pip.

2.3.1. Tweepy

Tweepy merupakan *library* Python untuk mengakses Twitter API dan mengambil data dari Twitter.

2.3.2. TextBlob

TextBlob merupakan *library* Python yang digunakan untuk memproses data tekstual. Textblob menyediakan API sederhana untuk mengerjakan tugas *Natural Language Processing* (NLP), antara lain : pemberian tag kata, ekstraksi kata benda, analisis sentimen, klasifikasi, terjemahan dan sebagainya (Loria, 2020).

2.3.3. Pandas

Pandas merupakan *library* Python untuk mengolah dan menganalisis data-data yang terstruktur. Untuk mengambil fungsi pandas dengan menuliskan "*import pandas as pd*". Format yang sering digunakan pada Panda yaitu DataFrame.

2.3.4. Numerical Python

Numerical Python atau Numpy merupakan salah satu *library* Python yang berfungsi untuk melakukan proses komputasi dalam operasi vektor dan matriks.

2.3.5. Regular Expression

Regular Expression merupakan *library* Python untuk mendefinisikan pola pencarian dengan pengenalan karakter untuk mendapatkan data yang sudah dimanipulasi sedemikian rupa, sehingga data menjadi bersih.

2.3.6. Matplotlib

Matplotlib merupakan *library* Python untuk menampilkan hasil analisis berupa grafik dan chart.

2.3.7. Wordcloud

Wordcloud merupakan *library* Python untuk menampilan hasil analisis berupa gambar *wordcloud*. Dalam *library* wordcloud juga menyediakan STOPWORDS yang berfungsi untuk menghapus kata-kata yang sering muncul (*common words*) namun kata-kata tersebut tidak memiliki arti untuk hasil analisis data.

2.3.8. Pillow

Pillow merupakan *Python Imaging Library* (PIL) yang berfungsi untuk membuka, menggunakan dan menyimpan gambar.

2.3.9. CSV

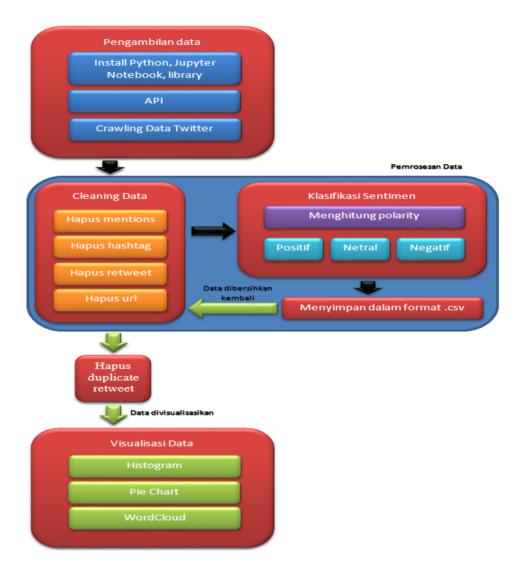
CSV (*Comma Separated Value*) merupakan *library* Python sebagai tempat untuk menyimpan data dengan format ".csv".

2.4. Jupyter Notebook

Jupiter Notebook adalah aplikasi open source sebagai tempat untuk membuat dan berbagi dokumen python. Untuk penginstalannya cukup menggunakan perintah pip. Sedangkan untuk mengakses Jupyter Notebook dengan mengetikkan "jupyter notebook" di Command Prompt.

III. METODE PENELITIAN

Pada penelitian ini, metode penelitian akan dilakukan seperti pada Gambar 1.



Gambar 1. Diagram Alur

3.1. Pengambilan Data

Setelah Python dan *library*-nya serta Jupyter notebook terinstall, proses pengambilan data dilakukan. Pengambilan data tweet dengan metode *crawling* dari media sosial Twitter. Untuk mendapatkan data tweet, dibutuhkan API Twitter, yang terdiri dari *API Key, API Secret Key, Access Token* dan *Access Token Secret*. Data tweet yang diambil dengan kata kunci bpbd, berjumlah 100 dan didefinisikan sebagai tweet berbahasa Indonesia (*lang="id"*). Data dimasukkan ke dalam tiga kolom berupa "tanggal_tweet", "pengguna" dan "isi tweet". Pengambilan data ini dilakukan pada 02 Juni 2021.

3.2. Pemrosesan Data

Sebelum data tweet diproses, data melalui tahap *pre-processing* yaitu *data cleaning*, yang meliputi hapus mentions (@), hastag (#), retweet (RT) dan url (Jarzynski, 2020). Karena textblob hanya bisa bekerja dalam bahasa Inggris maka data diterjemahkan ke dalam bahasa Inggris (Ngodein, 2020), untuk selanjutnya dilakukan proses *analysis.sentimen.polarity* dengan hasil positif, negatif, atau netral, yang disisipkan ke kolom sentimen. Hasil analisis disimpan dalam output.csv.

Selanjutnya memanggil file output.csv. Proses *cleaning* kembali dilakukan. Kemudian dilakukan pula *drop.duplicates* untuk menghapus tweet yang sama. Hasil data bersih selanjutnya divisualisasikan.

3.3. Visualisasi Data

Data bersih kemudian ditampilkan dalam bentuk histogram, diagram lingkaran dan word cloud.

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan analisis sentimen terhadap penanggulangan bencana yang mengacu pada BPBD. Data tweet diambil dan diproses melalui dua program python yaitu ambildata.ipynb dan visualdata.ipynb. Pengambilan data dilakukan pada 02 Juni 2021.

Pada proses ambil data, seperti ditunjukkan pada Gambar 2, pengambilan data dilakukan dengan kata kunci BPBD, berbahasa Indonesia (*lang="id"*), dengan mode tweet *extended*. Data tweet yang diambil berjumlah 100 data.

Gambar 2. Pengumpulan data

Setelah tweet berhasil dikumpulkan, selanjutnya dilakukan pemrosesan data. Pertama, melakukan pra-proses yaitu *cleaning* data seperti yang sudah dijelaskan pada metodologi, berupa hapus mentions (@), hashtag (#), retweet (RT) dan url. Pra-proses ini merupakan salah satu teknik yang digunakan untuk memperbaiki validitas dari data yang diambil. Sehingga, diharapkan data sentimen yang diambil sesuai dengan hasil yang dibutuhkan.

TextBlob hanya memproses data berbahasa Inggris, sehingga data tweet diubah ke dalam bahasa Inggris dengan menggunakan perintah translate(to="en"). Apabila terdapat tweet yang tidak dapat diterjemahkan/diubah atau merupakan tweet yang sudah berbahasa Inggris maka akan muncul notifikasi "Translation API Translation T

Selanjutnya dilakukan proses analisis sentimen dengan menggunakan *analysis.sentiment.polarity*. Apabila nilai polaritas lebih besar dari 0, maka sentimen tersebut positif. Nilai polaritas sama dengan 0, maka sentimen netral. Nilai polaritas lebih kecil dari 0, maka sentimen negatif.

Data kemudian disimpan dengan nama file output.csv, yang berisi empat kolom yaitu Tanggal_Tweet, Pengguna, Isi_Tweet dan Sentimen.

Sebelum menampilkan visualisasi dari hasil analisis, memanggil kembali data yang sudah simpan. Hasilnya seperti pada Gambar 3.

Tanggal_Tweet	Pengguna	lsi_Tweet	Sentimen
0 2021-06-02 04:29:39	BPBDKabBandung	Rabu, 02 Juni 2021\nPendampingan Penyerahan Ba	negatif
1 2021-06-02 04:18:22	ABamburuncing	RT @setro_gm: Banser Tanggap Bencana kab. Tema	netral
2 2021-06-02 04:15:09	Rahman_Nashir	RT @setro_gm: Banser Tanggap Bencana kab. Tema	netral
3 2021-06-02 04:09:48	adtaufiq	RT @setro_gm: Banser Tanggap Bencana kab. Tema	netral
4 2021-06-02 04:01:50	KangSumadi_	RT @setro_gm: Banser Tanggap Bencana kab. Tema	netral
5 2021-06-02 04:01:10	syaltout	RT @setro_gm: Banser Tanggap Bencana kab. Tema	netral

Gambar 3. Data output.csv

Dapat dilihat, masih terdapat tweet yang belum "bersih". Masih terdapat mentions (@), hashtag (#), retweet (RT) dan url. Oleh karenanya, dilakukan *cleaning* data kembali. Selain itu, menghapus tweet yang sama dengan menggunakan *drop.duplicates*. Hasilnya dapat dilihat pada Gambar 4.

Tanggal_Tweet	Pengguna	lsi_Tweet	Sentimen
2021-06-02 04:29:39	BPBDKabBandung	Rabu, 02 Juni 2021\nPendampingan Penyerahan Ba	negatif
2021-06-02 03:55:29	setro_gm	Banser Tanggap Bencana kab. Temanggung bersama	netral
2021-06-02 03:46:57	nusabalicom	Kalaksa BPBD Badung Meninggal, Didiagnosa Sera	netral
2021-06-02 03:43:04	bandungpemkab	Hasil sementara Assesment dari BPBD Kabupten B	netral
2021-06-02 02:40:23	bpbd_kabprob	Sahabat Tangguh berikut disampaikan update seb	positif
2021-06-02 02:31:21	ElshintaBandung	Terjadi banjir akibat jebolnya tanggul di Desa	negatif

Gambar 4. Data tweet bersih

Dapat dilihat dari Gambar 4, tweet dari masing-masing klasifikasi sentimen. Contohnya pada tweet positif berupa "Sahabat tangguh..." lalu pada tweet negatif berisi "Terjadi banjir akibat jebolnya tanggul...". Sementara pada sentimen netral lebih condong kepada tweet yang disertai dengan *link* yang mengarah pada situs atau berita tertentu. Dari hasil tersebut dapat ditarik kesimpulan bahwa data tweet tidak sepenuhnya merupakan kelompok kata dari setiap sentimen. Contoh saja kata "meninggal" tidak dikategorikan ke dalam tweet negatif, begitupun kata "tinggi" yang dalam kalimat netral masuk ke kategori positif.

Dari 100 tweet yang diambil, setelah dilakukan *cleaning* data kembali, terkumpul tweet bersih sejumlah 68 tweet, yang terdiri dari 16 tweet positif, 39 tweet netral dan 13 tweet negatif.

netral 39 positif 16 negatif 13

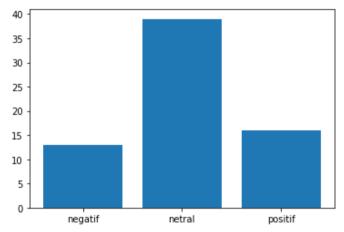
Gambar 5. Jumlah masing-masing sentimen

Apabila ditampilkan dalam bentuk persentase sejumlah 23,53 % tweet positif, 57,35 % tweet netral dan 19,12 % tweet negatif.

23.52941176470588 % tweet positif 57.35294117647059 % tweet netral 19.11764705882353 % tweet negatif

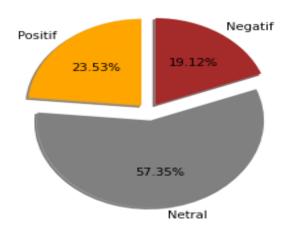
Gambar 6. Sentimen dalam bentuk persentase

Dengan memanfaatkan *library* Matplotlib, data tersebut ditampilkan dalam bentuk histogram, dengan masing-masing kategori positif, netral dan negatif seperti pada Gambar 7.



Gambar 7. Histogram hasil analisis sentimen

Selain dalam bentuk histogram, hasil analisis sentimen juga ditampilkan dalam bentuk diagram lingkaran atau *pie chart* seperti pada Gambar 8.



Gambar 8. Pie Chart hasil analisis sentimen

Data juga ditampilkan dalam bentuk *word cloud. Stop words* merupakan kata-kata yang sering muncul namun tidak memiliki arti pada hasil analisis data. Pada library wordcloud sudah menyediakan STOPWORDS. Namun karena beberapa kata mungkin tidak tercantum pada *stopwords* yang sediakan, maka ditambahkan *stopwords* tersendiri dengan *stopwords.update*. Hasil dari *word cloud* ditampilkan pada Gambar 9.



Gambar 9. word cloud hasil sentimen

Berikutnya menampilkan word cloud berdasarkan masing-masing klasifikasi sentimen. Tampilan word cloud dipadukan dengan library Numpy dan Pillow, yaitu dengan Image.open() dengan mengambil gambar awan. Sehingga, tampilan word cloud akan berbentuk awan. Gambar 10. menampilkan kata-kata yang sering muncul pada tweet positif.



Gambar 10. word cloud tweet positif

Gambar 11. menampilkan kata-kata yang sering muncul pada tweet netral.



Gambar 11. word cloud tweet netral

Gambar 12. menampilkan kata-kata yang sering muncul pada tweet negatif.



Gambar 12. word cloud tweet negatif

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada penelitian ini dilakukan pengumpulan data pada 02 Juni 2021 dengan kata kunci BPBD. Dari analisis sentimen didapat hasil 23,53 % tweet positif, 57,35 % tweet netral dan 19,12 % tweet negatif. Dengan demikian, opini pengguna Twitter tentang BPBD pada waktu tersebut mengarah ke kategori netral. Namun pada penelitian ini masih terdapat kekurangan yaitu pada kelompok kata yang seharusnya ada dalam kategori tweet negatif masuk ke kategori tweet netral dan sebaliknya.

5.2 Saran

Pada penelitian ini, pengambilan data menggunakan *textblob* untuk memproses data pada *text mining* dan melakukan proses klasifikasi sentimen. TextBlob memroses data dalam bahasa Inggris. Pengambilan data pada penelitian ini diatur menggunakan bahasa Indonesia, baru setelahnya diterjemahkan kembali ke bahasa Inggris sebelum dilakukan proses klasifikasi. Namun hasil yang didapat dalam penelitian ini masih belum maksimal.

DAFTAR PUSTAKA

- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Jurnal POROS TEKNIK*, 1(1), 32–41.
- Fauziyyah, A. K. (2020). Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python. *Jurnal Ilmiah SINUS*, *18*(2), 31. https://doi.org/10.30646/sinus.v18i2.491
- Feldman, R., & Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. In *Choice Reviews Online* (Vol. 44, Issue 10). Cambridge university press. https://doi.org/10.5860/choice.44-5684
- Fridom Mailo, F., & Lazuardi, L. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. *Jurnal Sistem Informasi Kesehatan Masyarakat Journal of Information Systems for Public Health*, *4*(1), 28–36. https://jurnal.ugm.ac.id/jisph/article/view/44455
- Jarzynski, P. (2020). Twitter Sentiment Analysis in Python Calculating Subjectivity and Polarity Score of a Twitter Account. Https://Towardsdatascience.Com/. https://towardsdatascience.com/twitter-sentiment-analysis-in-python-1bafebe0b566

- Loria, S. (2020). *TextBlob: Simplified Text Processing TextBlob 0.16.0 documentation*. https://textblob.readthedocs.io/en/dev/
- Negara, E. S., Andryani, R., & Saksono, P. H. (2016). Analisis Data Twitter: Ekstraksi dan Analisis Data G eospasial. *Jurnal INKOM*, *10*(1), 27. https://doi.org/10.14203/j.inkom.433
- Ngodein. (2020). *Tutorial Analisis Sentimen Twitter Python (Full Course)*. Www.Youtube.Com. https://youtu.be/8BWKYniQOis
- Pemerintah Republik Indonesia. (2007). Undang-Undang Nomor 24 Tahun 2007 tentang Penanggulangan Bencana. *Pemerintah Republik Indonesia*, 1–50. http://www.litbang.depkes.go.id/sites/download/regulasi/uu/UU_No._24_Th_2007_tt g Penanggulangan Bencana.pdf
- Pintoko, B. M., & L., K. M. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.
- Yasah, K. S. (2020). Analisa Sentimen Tweet Indonesia Menggunakan Fitur Ekstrasi Dan Teknik Cross Validation Terhadap Model Naïve Bayes. *Jurnal SIGMA*, *10*(4), 189–194.