

BAB 1 PENDAHULUAN

Pada bab ini menjelaskan mengenai pendahuluan dari penelitian yang terdiri dari latar belakang, rumusan masalah, tujuan, manfaat dari dilakukannya penelitian Analisa Sentimen Terhadap Varian Covid 19 Omicron menggunakan metode *Gaussian Naïve Bayes Classifier*.

1.1 Latar Belakang

Omicron atau yang lebih dikenal sebagai turunan B.1.1.529 merupakan varian terbaru dari SARS-Cov-2, sebuah virus yang menyebabkan Covid-19. Omicron sendiri pertama kali terdeteksi di Afrika Selatan pada tanggal 24 November 2021 yang dengan sangat cepat menyebar ke negara-negara lain seperti Inggris, Amerika Serikat, India sampai ke Indonesia. Sampai tanggal 14 Februari 2022 sendiri telah tercatat sebanyak 932,9 ribu kasus global terkait dengan Covid-19 Omicron. Negara Inggris menduduki posisi pertama penyebaran Omicron dengan total kasus sebanyak 457,26 ribu diikuti oleh Negara Amerika Serikat dengan total kasus sebanyak 370,86 ribu jiwa (Cindy Mutia Annur, 2022). Indonesia sendiri per tanggal 14 februari tercatat memiliki kasus sebanyak 5.305 kasus yang secara mingguan mengalami pertumbuhan sebesar 40,34%.

Peristiwa ini mengakibatkan dinaikkannya level PPKM dan pengetatan *social distancing* di beberapa daerah di Indonesia. Hal ini juga mengakibatkan terjadinya peningkatan penggunaan media sosial secara global di seluruh dunia (Chakraborty dkk., 2020). Peningkatan penggunaan media sosial ini sendiri terjadi karena media sosial berhasil menghubungkan orang-orang dari berbagai tempat di belahan dunia akibat diberlakukannya *social distancing*. Tidak hanya itu, banyak masyarakat juga yang mengandalkan media sosial untuk dapat memperoleh informasi-informasi lebih lanjut mengenai suatu kejadian (Yulita dkk., 2021). Sebagai contoh kejadian yang baru saja terjadi yaitu munculnya varian terbaru dari Covid-19 yang dinamakan Covid-19 Omicron. Peristiwa ini mengakibatkan banyaknya pendapat-pendapat yang muncul pada media sosial baik yang bersifat pro dan juga kontra terhadap peristiwa ini.

Diantara banyaknya platform media sosial yang terkenal saat ini, *Twitter* mendapatkan perhatian khusus. Hal tersebut dikarenakan para *user*-nya (penggunanya) dapat dengan mudah mendapatkan maupun menyalurkan informasi terkait peristiwa yang sedang terjadi dengan fitur *Twitter* yang bernama *tweet*. *Tweet* yang ada pada *Twitter* sendiri merupakan sebuah cara untuk mengeluarkan pendapat, ekspresi, perasaan sampai opini mengenai hal-hal yang sedang terjadi maupun dialami oleh Penulisnya. *Tweet* sendiri juga menyimpan informasi-informasi yang berkaitan juga dengan penggunanya. Seperti informasi mengenai lokasi dari penggunanya sampai ke *links*, *hashtag* dan *emoticons*. Hal tersebut yang membantu penggunanya untuk dapat mengeluarkan opini dengan lebih baik agar dapat menjadi sumber informasi yang berharga.

Seiring waktu berjalan, informasi-informasi yang berada pada *Twitter* telah digunakan dalam banyak penelitian-penelitian. Seperti penelitian oleh (Ruhayana, 2019) mengenai analisis sentimen terhadap penerapan ganjil/genap pada *twitter* menggunakan metode *Naïve Bayes* dengan akurasi yang didapatkan yaitu 86.67%. Selain itu ada juga penelitian yang dilakukan oleh (Ardiani dkk., 2020) mengenai analisis sentimen tanggapan masyarakat terhadap pembangunan di kota Pontianak yang menggunakan metode *Naïve Bayes* yang menghasilkan nilai akurasi sebesar 72%.

Adapun kekurangan dari kedua paper tersebut adalah dalam melakukan analisis sentimen pada *tweet* di platform media sosial *Twitter*, tidak ada penggunaan ekstraksi fitur *TF-IDF*. *TF-IDF* sendiri berfungsi untuk memboboti setiap kata dalam dokumen berdasarkan tingkat kemunculannya. Sehingga kata yang sering muncul akan memiliki nilai rendah sedangkan kata yang jarang muncul akan memiliki nilai yang tinggi. Hal ini berguna untuk meningkatkan tingkat akurasi dari model menjadi lebih tinggi.

Oleh karena itu penelitian ini dilakukan, penelitian ini dilakukan untuk menganalisis opini publik mengenai varian terbaru dari Covid 19 varian Omicron yang mengakibatkan kontroversi pro dan kontra pada media sosial. Penelitian ini akan mempertimbangkan pesan atau *tweet* yang ada pada media sosial *Twitter*. Data yang dikumpulkan pada penelitian ini dilakukan dengan mengambil data berupa *tweet* yang ada pada *Twitter*. Data set yang berupa *tweet* tersebut

selanjutnya akan dikelompokkan berdasarkan labelnya. Akan digunakan juga metode *Gaussian Naïve bayes* dengan bantuan ekstraksi fitur *TF-IDF* untuk melakukan klasifikasi terhadap sentimen yang didapatkan. Pada akhirnya penelitian ini bertujuan untuk mengetahui apakah penambahan ekstraksi fitur *TF-IDF* berpengaruh terhadap akurasi dari metode yang digunakan. Penelitian ini juga bertujuan untuk mengetahui polaritas opini masyarakat terhadap varian terbaru dari Covid-19 varian Omicron.

1.2 Perumusan Masalah

Berdasarkan latar belakang, maka didapatkan rumusan masalah untuk penelitian ini yaitu :

1. Bagaimana tingkat akurasi ekstraksi fitur *TF-IDF* pada metode *Gaussian Naïve Bayes classifier* dalam melakukan analisa sentimen pada *Twitter* terhadap varian terbaru Covid-19 Omicron.
2. Bagaimana hasil penilaian opini masyarakat pada *Twitter* mengenai Covid-19 varian Omicron menggunakan metode *Gaussian Naïve Bayes Classifier*.

1.3 Batasan Masalah

Batasan masalah untuk penelitian ini yaitu :

1. *Tweet* yang diambil dan akan dianalisis merupakan *tweet* dengan bahasa Indonesia.
2. Sentimen *Twitter* yang diambil dan dianalisis hanya sentimen yang berhubungan dengan Covid-19 khususnya varian Omicron.
3. Metode yang akan digunakan untuk menganalisa sentimen adalah metode *Gaussian Naïve Bayes Classifier*.
4. Media yang digunakan untuk mengambil data yang berupa *tweet* pada platform *Twitter* adalah *Twitter API*.
5. Bahasa pemrograman yang akan digunakan dalam penelitian adalah Python dengan memanfaatkan *tools* dan *library* yang disediakan *Google Collaboratory*.
6. *Tweet* hanya akan dinilai berdasarkan polaritas yang dibagi menjadi 3 yaitu positif, negatif dan netral atau tidak terdefinisi.

7. Pelabelan polaritas pada *Tweet* akan didasarkan pada kamus *lexicon* yang digunakan.

1.4 Tujuan Penelitian

Sesuai dengan rumusan masalah yang diperlihatkan pada sub bab 1.2, maka didapatkan tujuan dari penelitian ini yaitu :

1. Mengetahui tingkat akurasi dari penggunaan ekstraksi fitur *TF-IDF* pada metode *Gaussian Naïve Bayes Classifier* dalam melakukan klasifikasi pada *tweet* terhadap Covid-19 Omicron.
2. Menerapkan analisis sentimen terhadap data sentimen berbahasa Indonesia untuk mendapatkan penilaian opini masyarakat terhadap varian Covid-19 Omicron. Hasil akhir dari penilaian opini akan mengarah ke penilaian yang bersifat positif, negatif dan netral atau tidak terdefinisi.

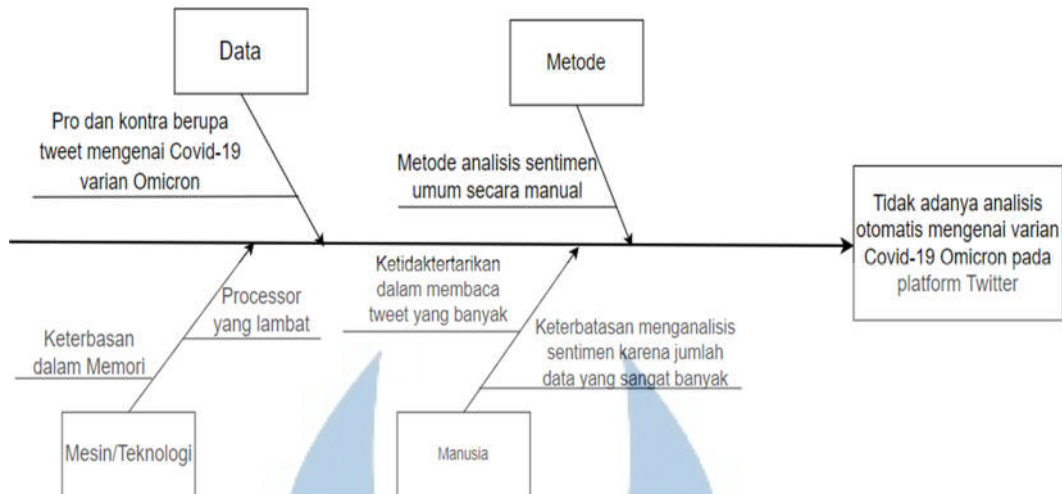
1.5 Manfaat Penelitian

Manfaat yang diharapkan dapat diberikan dari penelitian ini yaitu :

1. Memberikan pengetahuan mengenai akurasi dari ekstraksi fitur *TF-IDF* pada metode *Gaussian Naïve Bayes Classifier* dalam melakukan analisa sentimen terhadap varian terbaru Covid-19 Omicron pada *Twitter*.
2. Membantu menganalisa sentimen berupa *tweet* yang berasal dari *Twitter* menggunakan metode *Gaussian Naïve Bayes Classifier*.
3. Menjadi referensi untuk para Peneliti mendatang yang ingin melakukan penelitian pada tema yang terkait.

1.6 Kerangka Pemikiran

Kerangka pemikiran pada penelitian yang akan dilakukan ini digambarkan dengan diagram Ishikawa atau yang lebih dikenal dengan nama diagram *Fishbone*. Diagram *Fishbone* sendiri merupakan diagram yang menunjukkan penyebab dari sebuah hal atau kejadian yang spesifik. Diagram *Fishbone* ini akan menjelaskan alur pemikiran mengenai penelitian analisis sentimen mengenai Covid-19 varian Omicron pada *tweet* dari platform media sosial *Twitter* dengan menerapkan metode *Gaussian Naïve Bayes Classifier*. Diagram *Fishbone* untuk penelitian ini diperlihatkan pada Gambar 1.1:



Gambar 1. 1 Fishbone Diagram

Pada Gambar 1. 1 dapat dilihat penyebab atau landasan dilakukannya penelitian ini. Terdapat 4 aspek permasalahan yang menjadi landasan utama dari dilakukan penelitian ini, aspek tersebut adalah aspek data, mesin atau teknologi, metode dan terakhir manusia. Aspek yang menjadi landasan pertama adalah aspek material yaitu adanya *tweet* yang bersifat pro maupun kontra mengenai Covid-19 varian Omicron dalam jumlah banyak sehingga dapat memberikan informasi apabila dapat diketahui sentimen dari setiap *tweet* tersebut. Aspek kedua adalah mengenai metode yang dimana metode analisis sentimen masih umum dilakukan secara manual. Aspek ketiga adalah mesin atau teknologi yang dimana adanya keterbatasan memori dan processor yang lambat. Terakhir adalah aspek mengenai manusia yang dimana adanya keterbatasan dari manusia untuk menganalisis sentimen dalam jumlah yang sangat banyak dan ketidaktertarikan manusia untuk membaca *tweet* dalam jumlah banyak. Dari hasil analisis aspek-aspek tersebut maka dihasilkan lah masalah utama yaitu tidak adanya analisis otomatis mengenai varian Covid-19 Omicron pada platform *Twitter*.