

BAB I

PENDAHULUAN

1.1 Latar Belakang

Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) merupakan Badan Hukum Publik yang bertanggung jawab langsung kepada Presiden dan memiliki tugas untuk menyelenggarakan jaminan Kesehatan Nasional bagi seluruh rakyat Indonesia, terutama untuk Pegawai Negeri Sipil, Penerima Pensiun PNS dan TNI/POLRI, Veteran, Perintis Kemerdekaan beserta keluarganya dan Badan Usaha lainnya ataupun rakyat biasa. BPJS Kesehatan merupakan salah satu dari lima program dalam Sistem Jaminan Sosial Nasional (SJSN), yaitu Jaminan Kesehatan, Jaminan Kecelakaan Kerja, Jaminan Hari Tua, Jaminan Pensiun, dan Jaminan Kematian sebagaimana tercantum dalam Undang-Undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional. BPJS Kesehatan bersama BPJS Ketenagakerjaan (dahulu bernama Jamsostek) merupakan program pemerintah dalam kesatuan Jaminan Kesehatan Nasional (JKN) yang diresmikan pada tanggal 31 Desember 2013. Untuk BPJS Kesehatan mulai beroperasi sejak tanggal 1 Januari 2014, sedangkan BPJS Ketenagakerjaan mulai beroperasi sejak 1 Juli 2015.

Sejak 27 Desember 2019 jumlah penduduk Indonesia yang menggunakan BPJS Kesehatan sebanyak 224,1 Juta atau 83% dari total penduduk Indonesia (Agatha Olivia Victoria, 2020). Hal ini menunjukkan bahwa masyarakat yang menggunakan layanan BPJS cukup banyak. Penyelenggaraan asuransi kesehatan bertujuan untuk dimanfaatkan oleh peserta sebagai bagian dari upaya pemerintah untuk mencegah dan mengentaskan kemiskinan. Melalui program ini pemerintah berupaya untuk memelihara akses penduduk miskin terhadap pelayanan kesehatan yang masih rendah dan cenderung menurun. Selain itu, dengan program ini diharapkan akan lebih tepat sasaran mengingat pada program sebelumnya pernah terjadi penyalahgunaan program kesehatan oleh masyarakat yang mampu. Banyaknya pengguna BPJS di Indonesia menghasilkan banyak opini akan manfaat

penggunannya. Banyak masyarakat yang merasa terbantu atas manfaat BPJS akan tetapi ada juga yg merasa bahwa BPJS tidak sepenuhnya bermanfaat.

Center for Media and Democracy, Lembaga Penelitian, Pendidikan, dan Penerangan Ekonomi dan Sosial (LP3ES) bekerjasama dengan *Drone Emprit* melakukan analisa *big data* untuk menelisik lebih dalam reaksi publik terkait sebagaimana terekam dalam percakapan di media sosial. Direktur *Center for Media and Democracy* LP3ES, Wijayanto mengatakan ada 3 periode yang dianalisa yaitu pada Periode I dari 5 - 13 Mei 2020, atau sejak Perpres itu ditandatangani presiden sampai 8 hari setelahnya. Periode II dari 14 - 25 Mei 2020. Periode III dari 5 - 25 Mei 2020 atau sejak Perpres terbit hingga sekitar 3 minggu setelahnya. Wijayanto menyebut bahwa dari penelitian yang dilakukan ditemukan hal-hal yakni, pertama, sejak dinyatakan kenaikannya pada 5 - 25 Mei 2020, kenaikan premium BPJS mendapat banyak sorotan *netizens* di semua platform media mulai dari media sosial seperti *Twitter*, *facebook*, *youtube* dan *instagram* hingga media daring dengan total 115.599 percakapan. Hasilnya pada *volume* percakapan terbesar terjadi di *Twitter* dengan 101.745 percakapan, dengan hasil sentimen *negatif* sebesar 48%.

Banyaknya *Tweet* dari masyarakat yang membahas tentang kenaikan harga BPJS menjadikan *public sentiment* terhadap BPJS menjadi buruk, sehingga dibutuhkan analisa sentiment mengenai opini masyarakat sebelum dan sesudah kenaikan harga BPJS. Hal tersebut ditujukan untuk mengetahui opini masyarakat secara umum tentang bagaimana pelayanan BPJS kesehatan terutama selama masa *pandemic*, dimana peran BPJS sangat dibutuhkan oleh masyarakat, sehingga dibutuhkan analisis sentimen terhadap pelayanan BPJS Kesehatan. Data yang digunakan merupakan *Tweet* dari masyarakat selama masa *pandemic* di Indonesia.

Analisis sentimen atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen, dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks (Liu, 2012). Dengan menggunakan Analisis sentimen ini dapat mengelompokkan polaritas dari teks dalam kalimat atau dokumen untuk mengetahui apakah opini pada kalimat atau dokumen tersebut apakah termasuk positif atau negatif. Penelitian yang menggunakan *text mining* untuk sentiment analisis, menggunakan dataset positif dan dataset negatif, untuk

analisis sentimen review film menggunakan algoritma *K-Means*. Dari setiap uji coba menggunakan data dengan jumlah berbeda menunjukkan bahwa akurasi mencapai 57.83% didapatkan menggunakan 300 dataset positif dan 300 dataset negatif, sedangkan 700 dataset positif dan 700 dataset negatif menunjukkan akurasi mencapai 56.71%, jika menggunakan 1000 dataset positif dan 1000 dataset negatif akurasinya mencapai 50,40% (Budi, 2017).

Twitter adalah suatu layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan (*Tweet*). *Twitter* didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, *Twitter* telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan "pesan singkat dari Internet. Di *Twitter*, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler (Fonda, Mustika, Sabna, & Irawan, 2020).

Karakteristik teks yang didapat dari *Twitter* tidak berstruktur, sehingga menjadi tantangan dalam ekstraksi fitur pada bidang pemrosesan teks. Arliyanti Nurdin, Bernadus Anggo Seno Aji dkk melakukan penelitian bertujuan untuk membandingkan kinerja dari *word embedding* seperti *Word2Vec*, *GloVe* dan *FastText* dan diklasifikasikan dengan algoritma *Convolutional Neural Network*. Ketiga metode ini dipilih karena dapat menangkap makna semantik, sintatik, dan urutan bahkan konteks di sekitar kata jika dibandingkan dengan feature engineering tradisional seperti *Bag of Words*. Proses *word embedding* dari metode tersebut akan dibandingkan kinerjanya pada klasifikasi berita dari dataset 20 *newsgroup* dan *Reuters Newswire*. Evaluasi kinerja diukur menggunakan *F-measure*. Performa terbaik menunjukkan *FastText* unggul dibanding dua metode *word embedding* lainnya dengan nilai *F-measure* sebesar 0.979 untuk dataset 20 *Newsgroup* dan 0.715 untuk *Reuters*. Namun, perbedaan kinerja antar ketiga *word embedding* tersebut tidak begitu signifikan, menunjukkan bahwa ketiga *word embedding* ini

memiliki kinerja yang kompetitif. Penggunaannya sangat bergantung pada dataset yang digunakan dan permasalahan yang ingin diselesaikan (Nurdin, Bernadus, Bustamin, & Abidin, 2020).

Penelitian mengenai pencegahan insiden atau kecelakaan di kawasan industri diperlukan suatu analisis untuk mengetahui pola-pola yang terbentuk berdasarkan informasi dari laporan kejadian (Syaani, Irhamah, Mukarromah, & Fithriasari, 2021). Tujuan perusahaan adalah untuk dapat mengidentifikasi dan mengambil tindakan untuk menangani dengan cepat jika beberapa insiden terjadi. *Text mining* dilakukan dengan menerapkan teknik bigram untuk membentuk pola kejadian. Metode analisis yang digunakan adalah *K-Means* Klastering dan Hierarchical Klastering. Selain itu, seleksi fitur juga digunakan dengan menerapkan metode Algoritma Genetika untuk mendapatkan fitur yang optimal. Hasil yang diperoleh menyatakan bahwa proses seleksi fitur sangat berpengaruh terhadap pembentukan kluster insiden. Jika dibandingkan, metode *K-Means* Klastering dan Hierarchical Klastering memiliki efek yang berbeda pada masing-masing sektor gudang. Hasil terbaik pada Sektor Gaya Hidup dan Sektor Teknologi dibentuk dengan menggunakan metode *K-Means* Klastering sedangkan pada Sektor Konsumer, Ritel dan SPL hasil Kluster terbaik diperoleh berdasarkan metode Hierarchical Klastering.

Pandemi Covid-19 adalah peristiwa menyebarnya Penyakit koronavirus 2019 (Bahasa Inggris: Coronavirus disease 2019, disingkat Covid-19) di seluruh dunia untuk semua Negara. Penyakit ini disebabkan oleh korona virus jenis baru yang diberi nama SARS-CoV-2. Wabah Covid-19 pertama kali dideteksi di kota Wuhan, Hubei, Tiongkok pada tanggal 31 Desember 2019, dan ditetapkan sebagai pandemi oleh Organisasi Kesehatan Dunia (WHO) pada tanggal 11 Maret 2020.

Seperti yang diketahui bahwa BPJS memiliki peran yang sangat besar dalam dunia kesehatan di Indonesia. Terlebih selama terjadinya pandemi covid 19 ini. Keberadaan BPJS sangat dibutuhkan bagi rakyat. Hal ini tentunya menghasilkan beberapa opini rakyat akan peran BPJS selama pandemi di Indonesia. Dalam penelitian ini akan *mining* opini masyarakat yang ada di *Twitter*

menggunakan *library Phyton* yang bernama *Twint* yang nantinya akan diubah menjadi *vector* perkata melalui algoritma *Word2Vec* dan *GloVe*. Hasil dari proses vektorisasi tersebut nantinya akan diklaster menggunakan algoritma *K-Means* dan *Ward-hierarchical* Klustering, sehingga dari Klaster tersebut akan terlihat bagaimana sentiment masyarakat terhadap pelayanan BPJS Kesehatan selama masa *Pandemic*.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana hasil analisis sentimen opini masyarakat terhadap pelayanan BPJS Kesehatan dengan menggunakan *K-Means* dan *Ward*?
2. Bagaimana perbandingan hasil kerja algoritma *Word2Vec* dan *GloVe* jika dipadukan dengan *K-Means* dan *Ward*?

1.3 Tujuan

Tujuan dari penelitian ini adalah :

1. Menganalisa sentiment opini masyarakat tentang pelayanan BPJS Kesehatan dengan menggunakan algoritma *K-Means* dan *Ward*.
2. Membandingkan hasil kerja algoritma dari *Word2Vec* & *GloVe* jika dipadukan dengan algoritma klasterisasi K – Means dan *Ward*

1.4 Manfaat

- a. Rekomendasi algoritma yang optimal untuk proses klasterisasi data *twitter*.
- b. Mendapatkan informasi tentang persepsi pengguna terhadap layanan BPJS berdasarkan hasil analisis klaster
- c. Mempermudah dalam pengecekan maupun perbaikan layanan karena dapat mengelompokkan langsung keluhan dan saran dari pengguna.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini ialah:

1. Dataset yang digunakan adalah data *Tweet* dan komentar yang menggunakan bahasa Indonesia yang dibuat selama Januari 2020 – September 2021.

2. *Keyword* yang digunakan adalah “bpjs kesehatan until:2021-09-30 since:2020-01-01”
3. Membandingkan performa dengan algoritma *Word2Vec* & *GloVe* untuk *word embedding*.
4. Menggunakan Algoritma *K – Means* dan *Hirarcy Ward* untuk klasterisasinya.
5. Pemilihan Algoritma optimal menggunakan *Sum of Squares Error*, *Klaster Silhouette* dan *Davies Bouldini Index*.
6. Pembangunan program untuk klasterisasi menggunakan *library* dari *Phyton*, dan *Jupyter* sebagai Text Editor.