

BAB I

PENDAHULUAN

1.1 Latar Belakang

COVID-19 (*Corona Virus Disease 2019*) saat ini menjadi perhatian khusus di berbagai negara termasuk negara Indonesia. Coronavirus merupakan sekumpulan virus bagian dari subfamili *Othocoronvirinae* keluarga *Coronaviridae* dan ordo *Nidovirales* [1]. Di Indonesia sampai tanggal 1 September 2021 kasus terkonfirmasi COVID-19 mencapai 4.089.801 kasus, meninggal 133.023 kasus [2]. Untuk menekan angka kasus terkonfirmasi maupun meninggal akibat COVID-19 maka diperlukan sebuah kebijakan yang tepat dalam menanggulangi permasalahan ini. Pada tanggal 31 Maret 2020, Presiden Joko Widodo melakukan jumpa pers untuk memberitahukan kepada masyarakat tentang kebijakan yang dipilih guna menindaklanjuti pandemi COVID-19. Pada jumpa pers tersebut, Presiden Joko Widodo menyatakan kebijakan Pembatasan Sosial Berskala Besar (PSBB) adalah kebijakan yang diambil untuk menindaklanjuti kedaruratan dalam bidang kesehatan [3]. Kebijakan PSBB membuat fasilitas umum ditutup seperti tempat bermain, tempat wisata dan mall namun, beberapa sektor penting layaknya fasilitas kesehatan, pasar atau minimarket tetap dapat beroperasi saat masa PSBB berlangsung.

Kasus terkonfirmasi COVID-19 semakin melonjak dikarenakan banyaknya varian baru virus COVID-19 yang masuk ke Indonesia, tercatat 2.156.465 orang terkonfirmasi COVID-19 pada tanggal 30 Juni 2021 [4]. Hal tersebut membuat pemerintah segera mengupayakan kebijakan baru karena situasi di Indonesia semakin memburuk. Dengan pertimbangan matang akhirnya pada tanggal 03 Juli 2021 pemerintah meresmikan kebijakan baru yaitu PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) darurat [5]. Sejak diberlakukannya kebijakan PPKM darurat banyak sekali respon dan tanggapan dari masyarakat melalui media sosial terutama twitter.

Twitter adalah salah satu platform media sosial terkenal yang dibangun oleh Jack Dorsey dimana fungsinya adalah untuk mengirim pesan atau biasa disebut dengan *tweet* [6]. Dengan adanya kemudahan dalam mengirim tulisan (*tweet*), memberikan komentar, dan adanya fitur *retweet* membuat media sosial twitter banyak digunakan oleh masyarakat Indonesia untuk meluapkan opini mereka. Salah satu topik hangat yang sedang diperbincangkan saat ini adalah mengenai kebijakan pemerintah yaitu PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat). Opini masyarakat mengenai PPKM sangat beragam mulai dari respon positif seperti memberikan dukungan dan persetujuan hingga respon negatif yang memberikan kritikan bahkan cacian. Selain itu, muncul beberapa *hashtag* yang berhubungan dengan kebijakan PPKM seperti *#ppkmlevel4* dan *#ppkm*. Dengan banyaknya respon masyarakat di media sosial twitter dapat digunakan sebagai

pertimbangan pemerintah dengan cara membandingkan jumlah respon positif dan respon negatif. Salah satu cara untuk mengklasifikasikan *tweet* di twitter termasuk opini positif atau negatif adalah dengan melakukan analisis sentimen.

Analisis sentimen adalah bidang studi yang berkaitan dengan menganalisis pendapat orang secara tertulis [7]. Analisis sentimen juga disebut sebagai *opinion mining* karena berfokus pada opini yang menyatakan positif atau negatif. Pada penelitian ini analisis sentimen dilakukan untuk menganalisis suatu dokumen, *review*, komentar kemudian mengkategorikannya sebagai sentimen positif, negatif atau netral [8]. Saat ini, banyak peneliti dalam bidang ilmu komputer yang tertarik dengan penelitian seperti analisis sentimen. Dalam melakukan analisis sentimen harus menggunakan sebuah model atau metode perhitungan probabilitas yang dapat membantu dalam proses pengklasifikasian tulisan. Salah satu metode yang dapat digunakan dalam analisis sentimen adalah *naive bayes*.

Naive bayes adalah algoritma *machine learning* yang menggunakan teorema bayes untuk masalah klasifikasi [9]. Algoritma *naive bayes* dapat dikatakan sebagai algoritma klasifikasi sederhana namun memiliki kecepatan tinggi dalam melakukan klasifikasi data dengan jumlah yang sangat besar [9]. Data yang berbentuk teks dari twitter akan diolah terlebih dahulu sebelum dilakukan perhitungan algoritma *naive bayes*. Metode TF-IDF merupakan metode yang dapat dimanfaatkan untuk ekstraksi fitur dan pembobotan kata. TF-IDF merupakan metode ekstraksi fitur dan pembobotan kata yang digunakan untuk mendapatkan nilai dari kata dalam dokumen atau kalimat [10]. Rumus umum TF-IDF berasal dari penggabungan formula perhitungan raw TF dan formula IDF dengan cara mengalikan nilai *Term Frequency* (TF) dan nilai *Inverse Document Frequency* (IDF) [11]. Fitur yang dihasilkan dari metode TF-IDF akan sangat banyak karena data yang digunakan berupa data teks. Oleh karena itu, diperlukan proses seleksi fitur untuk memilih fitur atau atribut yang dianggap penting. Pada penelitian ini metode *Principal Component Analysis* dipilih untuk proses seleksi fitur. *Principal Component Analysis* mampu digunakan dalam meneliti data dengan variabel yang cukup kompleks. Selain itu, *Principal Component Analysis* juga dapat mereduksi dimensi data tanpa kehilangan informasi penting dari data yang digunakan [12].

Algoritma *naive bayes* digunakan dalam penelitian ini didasarkan pada banyaknya data *tweet* dari twitter yang akan ditentukan kategori kelasnya. Karena diketahui bahwa fungsi dari algoritma *naive bayes* sendiri adalah algoritma untuk mengklasifikasikan data. Algoritma *naive bayes* sendiri memiliki tingkat akurasi tinggi dan baik dalam perhitungan data dengan jumlah besar dibandingkan algoritma klasifikasi lainnya. Pada penelitian sebelumnya, yang dilakukan oleh Dedi, Neri, dan Zaenal pada tahun 2021 mengenai analisis sentimen pada twitter studi kasus

BMKG nasional menghasilkan tingkat akurasi *naive bayes* sebesar 69,97% [13]. Penelitian lainnya dilakukan oleh Ariel, Agung, dan Eri pada tahun 2022 mengenai analisis sentimen KRI Nanggala 402 dengan menggunakan data *tweet* pada twitter menghasilkan nilai akurasi sebesar 73% [14]. Dengan menggunakan kombinasi antara metode *naive bayes* dan metode seleksi fitur *Principal Component Analysis* pada penelitian ini diharapkan mampu memberikan informasi yang akurat tentang pendapat pengguna twitter terhadap kebijakan PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat). Diharapkan dengan adanya penelitian ini dapat menghasilkan informasi akurat dan bermanfaat tentang pendapat pengguna twitter terhadap kebijakan PPKM.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Dengan diresmikannya Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang menuai banyak pendapat di media sosial twitter maka, diperlukan analisis sentimen untuk mengetahui jumlah sentimen positif dan sentimen negatif dari masyarakat. Hal itu dikarenakan belum diketahui persepsi masyarakat mengenai kebijakan PPKM dan rendahnya hasil akurasi *Naive Bayes Classifier*. Sehingga penelitian ini dilakukan untuk mengetahui pengaruh metode *Principal Component Analysis* terhadap hasil akurasi apabila dikombinasikan dengan metode *Naive Bayes Classifier* dalam melakukan analisis sentimen dengan studi kasus PPKM ini. Berikut adalah permasalahan dari penelitian ini:

1. Praktis : Belum diketahui gambaran persepsi dari masyarakat mengenai kebijakan PPKM.
2. Teoritis : Rendahnya hasil akurasi menggunakan metode *Naive Bayes Classifier*.

1.2.2 Metode Usulan

Dalam penelitian ini mengusulkan metode *Naive Bayes Classifier* sebagai algoritma klasifikasi data, metode TF-IDF sebagai metode untuk pembobotan kata serta ekstraksi fitur, metode *Principal Component Analysis* untuk seleksi fitur.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang diatas maka dapat dirumuskan beberapa pertanyaan pada penelitian ini yaitu:

1. Berapa akurasi klasifikasi yang dihasilkan menggunakan metode *Naive Bayes Classifier*?
2. Seberapa besar pengaruh *Principal Component Analysis* terhadap akurasi?
3. Berapa lama waktu eksekusi program yang menggunakan *Principal Component Analysis* dan program yang tidak menggunakan metode *Principal Componenten Analysis*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan

1. Mendapatkan hasil akurasi klasifikasi data menggunakan metode *Naive Bayes Classifier*.
2. Mengetahui pengaruh *Principal Component Analysis* terhadap hasil akurasi.
3. Mengetahui lama waktu eksekusi program yang menggunakan *Principal Component Analysis* dan program tanpa menggunakan *Principal Component Analysis*.

1.3.2 Manfaat

1. Memperoleh informasi tentang bagaimana opini masyarakat pengguna twitter terhadap kebijakan PPKM.
2. Dapat digunakan sebagai literasi untuk penelitian yang berkaitan dengan analisis sentimen, metode *naive bayes*, TF-IDF, dan *Principal Component Analysis*.

1.4 Batasan Masalah

Penentuan batasan masalah dilakukan agar tidak terjadi pelebaran maupun penyimpangan pembahasan pokok masalah penelitian sehingga penelitian menjadi lebih terfokus, lebih mudah dalam pembahasan, dan tujuan penelitian akan berhasil. Berikut beberapa batasan masalah pada penelitian ini:

1. Luas lingkup penelitian hanya pada analisis sentimen pengguna twitter terhadap kebijakan PPKM dengan kata kunci “ppkm”, “ppkmlevel4”, “PSBB”, “vaksin booster”, “ppkmmikro”, “WFH”.
2. Perhitungan TF-IDF, *Principal Component Analysis*, dan *Naive Bayes Classifier* dilakukan dengan bantuan *library* python.
3. Banyak data yang digunakan dalam penelitian ini adalah data sebanyak 2850 data *tweet*, dan data diambil pada tanggal 13 Maret – 26 Maret 2022.

1.5 Metodologi Penulisan Proposal Skripsi

Dalam penulisan skripsi ini telah disesuaikan dengan format penulisan skripsi untuk Program Studi Sistem Informasi Fakultas Teknik Universitas Trunojoyo Madura. Berikut dijelaskan sistematika penulisan yang digunakan:

BAB I. PENDAHULUAN

Pada bab ini terdapat latar belakang penelitian, tujuan dan manfaat penelitian, rumusan masalah, serta batasan masalah dan sistematika penulisan.

BAB II. TINJAUAN PUSTAKA

Pada bab ini membahas teori-teori yang berhubungan dengan penelitian mengenai analisis sentimen pada media sosial twitter mengenai topik tertentu yang menggunakan metode *Naive Bayes Classifier*.

BAB III. METODOLOGI PENELITIAN

Pada bab ini membahas mengenai rancangan dari sistem terkait alur untuk melakukan pengumpulan data dari twitter, alur untuk melakukan *preprocessing* data, alur untuk melakukan seleksi fitur menggunakan metode *Principal Component Analysis*, alur untuk melakukan pembobotan menggunakan TF-IDF, alur untuk melakukan klasifikasi dengan metode *Naive Bayes Classifier* selain itu juga terdapat analisis kebutuhan *software*, *hardware* untuk membuat implementasi program.

BAB IV. HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai implementasi sistem yang dikerjakan sesuai dengan rancangan yang dibuat pada bab sebelumnya. Dimana terdapat skenario pengujian, hasil dari tahap *crawling* data pada twitter, hasil dari tahap pelabelan manual, hasil dari tahap *preprocessing* data (*tokenizing*, *cleansing*, *case folding*, konversi *slangword*, *stemming*, *stopword*), hasil dari tahap pembobotan dengan TF-IDF, hasil dari tahap seleksi fitur dengan metode *Principal Componen Analysis*, hasil dari tahap *Naive Bayes Classifier*, hasil dari tahap evaluasi model menggunakan *confusion matrix* dan *k-fold cross validation*, terakhir terdapat hasil dari uji skenario.

BAB V. PENUTUP

Pada bab ini dijelaskan tentang kesimpulan dari hasil penelitian yang meliputi hasil *accuracy* yang didapatkan, serta pengaruh metode *Principal Component Analysis* terhadap hasil *accuracy*. Selain itu juga terdapat saran bagi penelitian berikutnya yang akan mengangkat topik yang sama yaitu analisis sentimen.