

BAB I

PENDAHULUAN

1.1 Latar belakang

Covid-19 pertama kali terdeteksi di Kota Wuhan, Provinsi Hubei, China pada akhir Desember 2019. Pandemi Covid-19 muncul ketika virus ini diketahui menyebar dari orang ke orang dalam waktu singkat dan dengan gejala seperti demam tinggi, batuk, sesak, tidak nafsu makan dan lemas (Utami et al., 2020). Virus covid-19 berdampak sangat luas, bersifat masif dan global bukan saja mempengaruhi tingkat kesehatan masyarakat, tapi juga mempengaruhi aktivitas pendidikan, sosial, ekonomi, budaya, sosial, psikologis, olahraga, politik, pemerintahan, agama dan lain-lain (Umara & Hasanah, 2020). Peningkatan jumlah orang yang menderita Covid-19 mengkhawatirkan semua orang yang terlibat. Pemerintah mengeluarkan berbagai kebijakan salah satunya, pelaksanaan kebijakan pendidikan dalam masa darurat penyebaran *Coronavirus Disease* atau Covid-19 yaitu memuat arahan tentang proses belajar dari rumah (Arifa, n.d.). Instalasi sekolah dari taman kanak-kanak hingga perguruan tinggi memberlakukan pembelajaran secara daring karena pandemi Covid-19. Namun 2021 pemerintah mengeluarkan kebijakan pembelajaran tatap muka (PTT) terbatas yang jelas tidak sama dengan pembelajaran tatap muka seperti biasanya karena jumlah waktu yang dihabiskan siswa dan guru menjadi terbatas.

Saat ini kita memasuki era industri 5.0 yang ditandai dengan makin pesatnya digitalisasi, menyebabkan bertambahnya jumlah pengguna media sosial. Twitter

merupakan salah satu platform yang populer di Indonesia (Krisdiyanto, 2021). Media sosial Twitter seringkali digunakan untuk penyampaian pesan berupa pendapat atau opini masyarakat tentang bahasan tertentu serta bahasan yang menjadi trending topik terkini (Mentari et al., n.d.) Salah satu topik yang menjadi perbincangan oleh masyarakat saat ini diantaranya adalah kebijakan pemerintah mengenai pemberlakuan kembali pembelajaran tatap muka terbatas. Diberlakukannya kembali belajar tatap muka terbatas pada masa Covid-19 saat ini menimbulkan pro kontra dari masyarakat, Kita dapat melihat respon masyarakat terhadap pembelajaran tatap muka terbatas melalui berbagai sarana dan media.

Opinion mining atau Analisis Sentimen adalah proses mendalami, mengekstrak dan mengolah data tekstual secara otomatis bertujuan memperoleh informasi sentimen yang terkandung dalam suatu kalimat opini (Buntoro, n.d.). Analisis sentimen dapat digunakan untuk mengukur opini publik tentang isu-isu seperti pemberlakuan kembali pembelajaran tatap muka. Saat ini banyak dilakukan penelitian dalam analisis sentimen karena topik ini sangat menarik untuk dibahas. Ada banyak teknik yang dapat digunakan dalam mengelompokkan dokumen antara lain *K-Nearest Neighbor*, *Support Vector Machine*, *Naive Bayes*, *Artificial Neural Network* dan masih banyak lagi. Penelitian terkait pernah diteliti oleh (Furqan et al., 2022) yang berjudul "Analisis Sentimen Menggunakan *K-Nearest Neighbor* Terhadap *New Normal* Masa Covid-19 Di Indonesia" penelitian ini menggunakan 1000 tweet dan mengklasifikasikan sentimen ke dalam kelas negatif dan kelas positif, lalu dilakukan *preprocessing* data dengan tahapan *cleaning*, *normalisasi*, *stemming*,

case folding, filtering, dan tokenizing. Tahap pembobotan kata menggunakan metode *TFIDF*. Hasil pelabelan data yang didapat yaitu 811 tweet positif dan 189 tweet negatif. *Klasifikasi K-NN* dengan nilai $k=1$ menghasilkan nilai *accuracy* terbaik 100%. Penelitian lain yang menggunakan metode pembobotan kata dengan *TFIDF* dilakukan oleh Jeremy Andre Septian dan kawan-kawan dengan k terbaik $k=23$ menghasilkan nilai *accuracy* 79.99%.

Penelitian lain yang menggunakan metode *K-Nearest Neighbor* yaitu penelitian yang dilakukan oleh (Onantya, n.d.) dengan judul “Analisis Sentimen Pada Ulasan Aplikasi BCA *Mobile* Menggunakan *BM25* Dan *Improved K-Nearest Neighbor*”. Penelitian ini memperoleh data dari penilaian atau ulasan pengguna aplikasi BCA *Mobile*, data akan dibagi menjadi kelas positif dan negatif. Tahap *preprocessing* terdiri dari *Case Folding, Tokenizing, Filtering* dan *Stemming*. Metode *BM25* digunakan untuk pembobotan kata, dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor* dengan $k=10$ menghasilkan *accuracy* terbaik 94%. (Mahardhika & Zuliarso, 2018) melakukan penelitian dengan judul “Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma *Naives Bayes Classifier*”. Pada penelitian ini akan dilakukan proses analisis opini masyarakat mengenai pemerintahan joko widodo, Data yang digunakan berjumlah 400 tweet yang akan diklasifikasikan menjadi tweet kelas positif dan kelas negatif, yang didapat dari media sosial twitter dengan jumlah data latih 300 dan data uji 100. Klasifikasi dilakukan dengan menggunakan metode *Naïve Bayes Classifiers*. Penelitian ini menghasilkan bahwa nilai *accuracy* terbaik sebesar 97%, Untuk nilai akurasi tiap

sentimennya yaitu 98% untuk sentimen positif dan 96% untuk sentimen negatif.

Metode *Naïve Bayes Classifiers* juga digunakan oleh (Nugroho et al., 2016) yang berjudul “Analisis Sentimen Pada Jasa Ojek *Online* Menggunakan Metode *Naïve Bayes*”. Hasil penelitian menunjukkan bahwa data yang diperoleh sebanyak 3000 data yang terdiri dari 1000 kalimat netral, 1000 kalimat negatif dan kalimat positif 1000. Tingkat *accuracy* terbaik menggunakan metode *Naïve Bayes* yaitu 88%. Pada penelitian ini akan membuat sistem yang mampu mengklasifikasikan sentimen positif, negatif dan netral memperoleh tingkat ketepatan *accuracy* 80%. Penelitian sebelumnya juga melakukan penelitian menggunakan Metode *Naive Bayes* dan *Support Vector Machine (SVM)* dalam melakukan perbandingan kinerja metode untuk mengetahui sentimen masyarakat khususnya *netizen* Twitter terhadap calon Gubernur DKI Jakarta 2017 (Nurhuda et al., 2016). Teknik pengumpulan data dalam penelitian adalah dengan cara *crawling* sebanyak 300 tweet dengan kata kunci Ahok, AHY, Anies. untuk *preprocessing* data menggunakan *tokenisasi*, *cleansing* dan *filtering*. Pelabelan data dilakukan secara otomatis menggunakan metode *Lexicon Based*. Nilai *accuracy* yang dihasilkan rata-rata 95%.

Berdasarkan penelitian sebelumnya, dapat dijadikan acuan dalam melakukan penelitian ini maka telah diputuskan bahwa beberapa hal dari penelitian sebelumnya penulis akan menerapkan di penelitian penulis sehingga penelitian penulis berbeda dan memiliki keistimewaan dari penelitian lainnya. Pada penelitian ini akan dilakukan analisis mengenai tweet yang di kumpulkan dengan 4 *keyword* sehingga data yang diperoleh akan lebih banyak diantaranya “tatap muka”, “luring”, “kuliah *offline*”, dan “sekolah *offline*” selama Covid-19. Setelah memperoleh Data dari proses *crawling* selanjutnya dilakukan tahap *preprocessing*. Pada tahap pelabelan penulis akan melakukan pelabelan secara manual dan otomatis menggunakan *library*

python textblob sehingga akan dilakukan komparasi, lalu diklasifikasikan menjadi sentimen positif, negatif dan netral. Algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini yaitu *Naives Bayes Classifier* dan *K-Nearest Neighbor* dan dilakukan perbandingan yang diharapkan dapat menghasilkan analisis data yang sesuai kategori dengan akurasi yang terbaik.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang penulis telah sebutkan, permasalahan yang menjadi fokus pada penelitian ini yaitu mengklasifikasikan sentimen terhadap kesiapan belajar tatap muka dan bagaimana perbandingan metode *Naive Bayes* dan *K-Nearest Neighbor* saat dipakai untuk melakukan klasifikasi menganalisis sentimen.

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah mengklasifikasi opini masyarakat di media sosial yang mengandung nilai positif, netral dan negatif mengenai kesiapan pembelajaran tatap muka menggunakan metode *Naive Bayes* dan KNN (*K-Nearest Neighbor*).

1.4 Batasan Masalah

Agar penelitian ini tidak keluar dari pokok permasalahan dan mengakibatkan permasalahan, maka Penelitian ini membatasi ruang masalah sebagai berikut :

- a. Data sentimen berasal dari opini publik di media sosial twitter yang di ambil dengan kata kunci “tatap muka”, “luring”, “kuliah *offline*”, dan “sekolah *offline*” berbahasa indonesia dengan jumlah data yang tidak

dibatasi dengan teknik *crawling* yang diambil dari tanggal 9 Mei 2022 sampai dengan 4 Juni 2022.

- b. Kumpulan data berupa text dan tidak memakai simbol, angka, tanda baca, dan *emoticon*.
- c. Mengklasifikasikan emosi kedalam kemungkinan positif, negatif atau netral.
- d. Metode yang digunakan dalam pengklasifikasian ini adalah *naive bayes* dan *K-Nearest Neighbor*.
- e. Untuk menyelesaikan penelitian penulis memakai bahasa pemrograman *python*.

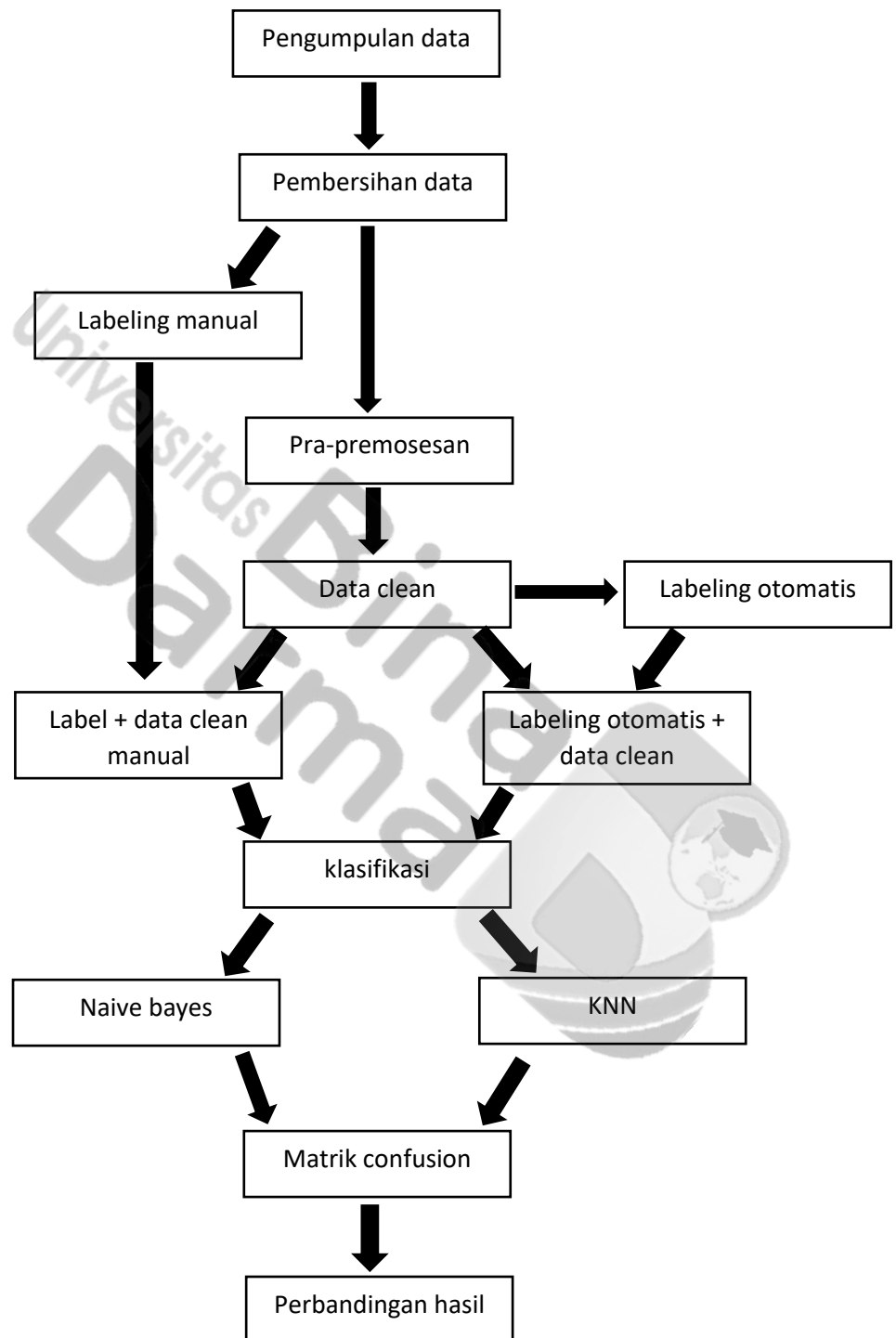
1.5 Manfaat Penelitian

Berdasarkan latar belakang dapat disimpulkan bahwa manfaat penelitian adalah

- a. Dari data yang besar akan dilakukan klasifikasi tweet ke dalam kelas negatif, positif dan netral dengan cepat.
- b. Menguji tingkat akurasi terhadap dua metode yaitu *naive bayes* dan *K-Nearest Neighbor*.
- c. Menjadi sistem labeling secara otomatis untuk penelitian selanjutnya.

1.6 Metode Penelitian

Agar mempermudah pelaksanaan penelitian maka dibuat model penelitian yang berfungsi untuk alur perjalanan penelitian. Tahapan yang dilakukan dalam model penelitian bisa dilihat pada gambar 1.



Gambar 1. 1 Model Penelitian

Dalam penelitian ini data yang digunakan berupa tanggapan masyarakat pada masa covid-19 dari tanggal 9 Mei 2022 sampai dengan 4 Juni 2022, tentang kebijakan pembelajaran tatap muka. Data diperoleh dengan metode

crawling pada media twitter menggunakan twitter API dengan memanfaatkan fitur dari twitter yang diperuntukan untuk *developer*. Jumlah data tweet yang di ambil tidak dibatasi dan merupakan data tweet berbahasa indonesia dengan kata kunci “tatap muka”, “luring”, “kuliah *offline*”, dan “sekolah *offline*”. Tahap selanjutnya pembersihan data dilanjut label secara manual dan proses *preprocessing* yaitu *tokenizing, case folding, normalization, filtering, stemming*. Sehingga menghasilkan data clean lalu dilakukan labeling secara otomatis. Dilanjutkan dengan klasifikasi data menggunakan metode *naive bayes* dan *K-Nearest Neighbor* sehingga dapat diketahui secara jelas berapa nilai akurasi, *recall, precision* dari setiap model klasifikasi. Terakhir dilakukan perbandingan dari model klasifikasi *naive bayes* dan *K-Nearest Neighbor* mana yang paling akurat dalam mengklasifikasikan data tweet kesiapan belajar tatap muka.

1.6.1 Waktu dan Tempat Penelitian

Pelaksanaan penelitian dilakukan sejak semester genap 2021/2022 pada fakultas ilmu komputer program studi sistem informasi universitas Bina Darma Palembang, dan diperkirakan penelitian berakhir pada bulan agustus 2022.

1.6.2 Alat Dan Bahan Penelitian

1.6.2.1 Bahan Penelitian

Untuk mendukung penelitian ini maka penulis membutuhkan beberapa alat dan bahan. Pada penelitian ini membutuhkan data opini masyarakat

berbahasa indonesia yang berasal dari media sosial twitter. Tidak hanya itu informasi lain juga didapat dari internet, studi literatur.

1.6.2.2 Alat penelitian

Adapun alat yang digunakan penulis untuk melakukan penelitian ini yaitu sebagai berikut :

1. Perangkat Keras
 - a. Laptop Asus ViviBookMax X441U
2. Perangkat Lunak
 - a. Twitter
 - b. Browser Internet (Google Chrome atau Mozilla Firefox)
 - c. Microsoft Office 2010
 - d. Sistem Operasi Windows 10 64 Bit
 - e. Bahasa Pemrograman Python 3.7
 - f. Zotero

1.6.3 Metode pengumpulan data

Metode pengumpulan data pada skripsi ini adalah :

1. (observasi) adalah cara atau cara untuk menghadapi memperoleh informasi penting dengan memperhatikan secara langsung objek informasi tersebut. Pendekatan persepsi dapat disusun menjadi persepsi perilaku dan persepsi nonsosial. Siklus dilakukan dengan

memperhatikan perasaan atau penilaian populer yang diberikan pada hiburan virtual Twitter terkait pembelajaran dari dekat dan personal.

2. Studi Literatur (*Literature Research*) adalah Penulisan Penelitian Interaksi ini dilakukan dengan mengumpulkan semua eksplorasi yang berhubungan dengan judul penelitian, khususnya penelitian investigasi opini. Menulis berkonsentrasi pada itu akan membuat informasi opsional.

1.6.4 Jadwal Penelitian

Jadwal pelaksanaan penelitian akan dilaksanakan dalam waktu 5 bulan terhitung dari bulan April 2022.

Tabel 1. 1 Jadwal Penelitian

| No | Kegiatan | Waktu Penelitian | | | | |
|----|---|------------------|-----|------|------|---------|
| | | April | Mei | Juni | Juli | Agustus |
| 1 | Studi literatur | ■ | ■ | | | |
| 2 | Pengumpulan data | | ■ | | | |
| 3 | Pelabelan data | | | ■ | | |
| 4 | Pembuatan model <i>Naive Bayes & K-Nearest Neighbor</i> | | | ■ | ■ | |
| 5 | Pengujian model | | | | ■ | |
| 6 | Proses hasil klasifikasi | | | | | ■ |
| 7 | Penulisan skripsi | ■ | ■ | ■ | ■ | |