BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi saat ini berkembang semakin pesat sehingga memudahkan manusia dalam melakukan berbagai macam kegiatan sehari-hari mereka. Teknologi yang perkembangannya begitu pesat saat ini yaitu teknologi informasi dan komunikasi. Salah satunya, pertumbuhan media sosial *online* seperti Twitter yang digunakan untuk menyatakan fakta atau opini dari penggunanya (Ludwianto, 2020). Menurut Kamus Besar Bahasa Indonesia (KBBI), Fakta merupakan hal yang merupakan kenyataan, sedangkan opini merupakan ekspresi yang menggambarkan emosi atau sentiment dari penulis atau pembicaranya.

Pengguna aktif harian Twitter di Indonesia saat ini cukup banyak, pada kuartal ke-3 tahun 2019, Indonesia mengalami peningkatan pengguna sebanyak 17% ke angka 145 juta pengguna (Nistanto, 2019). Selain digunakan untuk memberikan fakta dan opini, media sosial tekstual seperti Twitter seringkali digunakan untuk mengungkapkan emosi, Emosi dapat dikategorikan menjadi emosi positif dan emosi negatif, di mana emosi positif mencakup emosi cinta dan senang, sementara emosi negatif mencakup emosi takut, sedih, dan marah (Firdaus dan Puji, 2020).

Saat ini depresi bukanlah kata yang asing di telinga masyarakat, akan tetapi kata depresi itu sendiri terkadang salah diartikan. Depresi merupakan gangguan emosi yang ditandai dengan kondisi emosi sedih dan muram, depresi merupakan

gejala yang wajar sebagai respon terhadap pengalaman hidup negatif (contoh: kehilangan anggota keluarga). Depresi akan dianggap serius jika emosi sedih dan muram memiliki intensitas yang tinggi dan menetap (Aditomo dan Retnowati, 2015). Menurut Laporan Hasil Riset Kesehatan Dasar (Riskesdas, 2020), pada tahun 2018, 6.1% dari 267.7 juta jiwa populasi orang dewasa menderita gangguan kecemasan dan depresi.

Dengan adanya pemikiran, suasana hati, dan kehidupan sosial yang dituangkan dalam bentuk tekstual pada media sosial Twitter, dapat dilakukan indikasi pengguna mengalami gejala depresi atau tidak melalui bahasa dan emosi yang digunakan dengan cara kategorisasi teks (Choudury, Gamon dkk, 2013). Teks melalui tahap pengambilan data dan prapengolahan yang akan digunakan selanjutnya dengan algoritma *machine learning* dalam mengklasifikasi teks. *Naive Bayes Classifier* merupakan salah satu metode atau algoritma *machine learning* yang sesuai dengan permasalahan klasifikasi teks, *Naïve Bayes* merupakan sebuah teknik sederhana dalam membuat model pengklasifikasi yang menetapkan kelas untuk contoh masalah, yang digambarkan sebagai vektor dari nilai-nilai fitur sebuah data, di mana label kelas tersebut diambil dari beberapa himpunan tak hingga (Hermanto, 2016).

Multinomial Naïve Bayes dan *Complement Naïve Bayes* merupakan salah dua metode spesifik dari metode *Naïve Bayes*. *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* dapat menggunakan *conditional probability*, yaitu dengan menggunakan frekuensi kemunculan kata pada suatu kelas. *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* juga merupakan salah satu *supervised learning* yang terdapat pada *machine learning* yang dapat digunakan untuk

mengklasifikasikan teks dengan menggunakan nilai probabilitas suatu kelas dalam suatu dokumen. Pada penggunaan *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* sering kali ditemukan probabilitas dengan nilai 0, untuk menghindari hal tersebut, digunakan sebuah metode yang dinamakan *Add One* yaitu menambahkan nilai satu agar hasil yang diberikan lebih akurat (Nurfadhila, 2017).

Penjelasan ini juga diperkuat dengan adanya beberapa penelitian serupa, yaitu untuk studi kasus prediksi depresi yang dilakukan untuk mengukur tingkat indikasi depresi pada pengguna Twitter. Penelitian ini dilakukan dengan menggunakan metode *Text Mining* dengan algoritma *Multinomoial Naïve Bayes* yang digunakan untuk mengindikasikan dan menghitung tingkat indikasinya. Aplikasi yang dibuat dengan mengimplementasikan algoritma *Multinomial Naïve Bayes* ini dapat menghasilkan akurasi sebesar 73,86% (Nurfadhila, 2017).

Berikutnya, prediksi identifikasi depresi pada pengguna Twitter Amerika Serikat, algoritma yang dipakai adalah *Naïve Bayes w/ 1-grams, Linear Support Vector Classifier, Decision Trees, Logistic Regression, Naïve Bayes w/ 2-grams, Ridge Classifier* dan nilai akurasi tertinggi dimiliki oleh algoritma *Naïve Bayes w/ 1-grams,* yaitu 86% yang disusul oleh *Linear Support Vector Classifier* 82%, *Naïve Bayes w/ 2-grams* 82%, *Logistic Regression* 82%, dan *Decision Trees* yang bernilai 67%, hasil akurasi juga dikuatkan dengan adanya pembuatan *Receiver Operating Characteristic* (ROC) dan *Area Under the Curve* (AUC) terhadap *Multinomial Naïve Bayes* yang memiliki akurasi bernilai 86% (Nadeem, 2016).

Berikutnya, terdapat penelitian untuk mengidentifikasi emosi dasar manusia terhadap *tweet* yang menggunakan bahasa Indonesia dengan studi kasus tokoh publik menjelang pemilu tahun 2019. Metode yang digunakan adalah *Multinomial*

Naïve Bayes dengan membandingkan hasil dari *k-Fold Cross Validation.* Hasil yang didapatkan oleh akurasi *confusion matrix* tanpa menggunakan *k-Fold Cross Validation* adalah 72.941% dan yang menggunakan *k-Fold Cross Validation* adalah sebesar 71.601% untuk 3-*Fold*, 70.72% untuk *5-Fold*, dan 70.68% untuk *10-Fold* (Harijiatno, 2016).

Penelitian serupa juga dilakukan untuk studi kasus prediksi sentiment analisis *tweet* dengan menggunakan bahasa Indonesia yang dilakukan dengan metode *Deep Belief Network*, penelitian dilakukan guna untuk mendeteksi jumlah sentimen masyarakat Indonesia dengan melakukan kategorisasi, yaitu sentimen positif, negatif, dan netral. Pada penelitian ini dilakukan dengan menggunakan 3 metode yang berbeda untuk mencari metode yang terbaik, metode yang digunakan, antara lain: *Deep Belief Network* (DBN), *Support Vector Machine* (SVM), *Naïve Bayes*. Nilai Akurasi yang dihasilkan oleh DBN bernilai 93.31%, SVM bernilai 92.18%, sementara *Naïve Bayes* memiliki nilai akurasi sebesar 79.10% (Zulfa dan Winarko, 2017).

Berdasarkan beberapa penelitian terkait, dilakukanlah implementasi algoritma *Naïve Bayes* berupa *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* untuk mengklasifikasikan *tweet* dengan indikasi depresi. Dengan penelitian ini, diharapkan dapat membantu dalam meningkatan kewaspadaan masyarakat Indonesia akan penyakit mental, khususnya depresi. Penelitian ini juga diharapkan dapat membantu dalam melakukan indikasi awal penyakit mental (depresi) pada cuatan yang ada di media sosial.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan di atas maka dapat dibuat rumusan masalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* untuk mengklasifikasikan konten Twitter dengan indikasi depresi?
2. Berapa hasil *F1 Score* yang diperoleh dari mengimplementasikan algoritma *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* untuk melakukan klasifikasi *tweet* dengan indikasi depresi?

1.3 Batasan Masalah

Berdasarkan permasalahan yang telah didapatkan, maka saya akan membatasi masalah agar tidak keluar dari topik yang akan difokuskan. Adapun batasan masalah tersebut ialah:

1. *Tweet* yang diambil merupakan data publik yang berasal dari orang-orang dengan kebangsaan Indonesia dengan menggunakan metode *Crawling* dalam pengambilan data dan data divalidasi berdasarkan arahan dari praktisi psikolog (Patricia Astrid Nadia, S.Psi).
2. Media yang digunakan untuk mengambil data *tweet* pada Twitter ialah Twitter *API.*
3. Tolak ukur indikasi depresi yang akan dipakai adalah gejala awal kesehatan mental penulis *tweet*, yaitu emosi sedih yang berlebihan dan berkepanjangan.

4. Sistem akan mengklasifikasikan *tweet* yang memiliki indikasi depresi dan yang tidak memiliki indikasi depresi.
5. Data yang digunakan merupakan data yang diambil pada kisaran waktu 18 November sampai dengan 7 Desember 2020, berjumlah 10081 data dan telah melalui proses normalisasi kata yang dilakukan secara manual.

1.4 Tujuan Penelitian

Tujuan diadakannya penelitian ini yaitu untuk:

1. Mengetahui metode *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* mampu secara otomatis melakukan klasifikasi *tweet* dengan indikasi depresi pada Twitter dengan baik.
2. Mengukur keakuratan sistem dengan menggunakan *metric F1 Score* dengan menggunakan metode *Multinomial Naïve Bayes* dan *Complement Naïve Bayes* dalam melakukan klasifikasi *tweet* dengan indikasi depresi pada Twitter.

1.5 Manfaat Penelitian

Dengan diadakannya penelitian ini, diharapkan dapat membantu praktisi maupun orang awam dalam segi praktis dan teoritis dalam melakukan indikasi depresi.

1.5.1 Manfaat Praktis

Hasil penelitian ini diharapkan dapat memberikan manfaat, diantaranya:

1. Agar dapat dijadikan referensi untuk melakukan identifikasi indikasi depresi pada *tweet* dalam pengembangan ilmu teknologi khususnya pada bidang kedokteran.

2. Diharapkan penelitian ini dapat menjadi referensi bagi peneliti mendatang yang akan melakukan penelitian dengan tema yang serupa.

1.5.2 Manfaat Teoritis

Hasil penelitian ini diharapkan dapat memberikan manfaat bagi dunia pendidikan khususnya ilmu kedokteran dalam melakukan identifikasi indikasi depresi pada pengguna Twitter dan mengurangi *self-diagnosed* oleh pengguna Twitter terhadap kesehatan mental khususnya depresi.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyajian laporan skripsi ini adalah sebagai berikut:

BAB I LATAR BELAKANG

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan yang dilakukan dalam penelitian ini.

BAB II LANDASAN TEORI

Bab ini secara umum menguraikan berbagai macam teori yang mendukung dan menjadi dasar dalam penulisan skripsi ini, diantara lain dengan menjelaskan pengertian sentiment, depresi, *text preprocessing* dalam menambang data *tweet*, dan metode *Naive Bayes* yang digunakan.

BAB III METODE PENELITIAN

Bab ini berisikan metode penelitian dan proses perancangan yang digunakan di dalam pembuatan sistem analisis.

BAB IV IMPLEMENTASI DAN UJI COBA

Bab ini berisikan berbagai macam penjabaran hasil akurasi algoritma yang digunakan, hasil penelitian, dan berbagai macam analisis terhadap penelitian yang dilakukan.

BAB V SIMPULAN DAN SARAN

Bab ini berisikan simpulan dari hasil penelitian yang dilakukan berdasarkan tujuan yang ada dan saran untuk pengembangan lanjut dari penelitian ini.