

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pada era teknologi yang semakin maju pesat ini menggunakan media sosial saat ini telah menjadi trend dalam komunikasi pemasaran. Dari semua kalangan pada saat ini menggunakan dan menyukai media sosial. Salah satu media yang banyak digunakan adalah twitter. pengguna twitter dapat berbagi antar penggunanya dalam hal berbagi informasi, bersosialisasi dan lain – lainnya, disini publik dapat berkomentar dan menulis apapun. Seringkali pengguna media sosial sulit untuk dapat memahami informasi yang pengguna lain bagikan, karena tulisan yang di tulis oleh pengguna twitter terkadang tidak mempresentasikan opini yang sedang dicurahkan oleh penulis, hal ini yang menyebabkan terjadinya misscommunication, misinpretation, dan banyaknya penyebaran informasi yang tidak benar (hoax)(Dewi 2021). Karena pengguna tidak paham dengan rasa yang ingin disampaikan oleh pengguna lainnya. Banyak pengguna twitter yang menulis (memposting) ekspresi dan pendapat mereka terhadap suatu produk, layanan, isu politik dan hal hal yang sedang viral

Seperti yang telah diberitakan saat ini seorang hacker bernama Bjorka yang namanya telah dibicarakan disemua penjuru Indoneisa karena aksinya yang dapat meretas data-data dari pemerintahan dan masih banyak hal yang telah diretas oleh sosok Bjorka ini. Bjorka bergabung di forum ini sejak 9 Agustus 2022. Karena keahliannya, Bjorka mendapatkan predikat sebutan God atau Dewa. Bersumber dari website tempat dimana Bjorka menjual data tersebut, dia sudah mengunggah milyaran data pribadi yang di klaimnya dari hasil membobol data korporasi hingga lembaga negara (Kurniawan and Syah 2022). Banyak opini dari kalangan masyarakat terhadap aksi yang telah dilakukan oleh hacker Bjorka ini, ada sebagian masyarakat yang mendukung aksi Bjorka ini karena telah membuka aib dari pemerintahan yang selama ini telah ditutup tutupi oleh pemerintah, ada juga sebagian masyarakat yang tidak mendukung aksi yang tidak terpuji tersebut karena dianggap telah memasuki wilayah privasi seseorang, mengambilnya dan menjual data – data pribadi masyarakat untuk

mendapatkan keuntungan berupa uang, maka dari itu banyak penelitian pembelajaran mesin yang dilakukan untuk menghasilkan suatu program yang dapat membantu menganalisa hal tersebut.

Analisis sentiment atau opinion mining merupakan proses memahami, mengekstrak dan mengola data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini (Rozi, Pramono, and Dahlan 2012). Dari penelitian ini mengambil sentimen publik pengguna twitter dengan topik Bjorka. Topik yang diambil kemudian di analisis sentiment dan di klasifikasi data tersebut menjadi tiga bagian. Bagian yang terdapat didalamnya mengandung sentimen positif, negatif, atau netral.

Pada penelitian ini menggunakan metode dari model Naïve Bayes Classifier dengan membandingkan akurasi dari model Multinomial Naïve Bayes, Bernoulli Naïve Bayes, dan Gaussian Naïve Bayes. Model Multinomial dalam klasifikasi Naïve Bayes merupakan model yang memperhatikan jumlah kemunculan kata dalam dokumen, Sedangkan Bernoulli Naïve Bayes adalah suatu model yang menetapkan bahwa dokumen diwakili oleh vektor atribut biner yang menunjukkan kata yang muncul dan tidak muncul dalam dokumen(Wardani, Prahutama, and Kartikasari 2020). Bernoulli Naïve Bayes hanya berfokus pada klasifikasi ya atau tidak, dan Gaussian Naïve Bayes mengansumsikan pendistribusian nilai kontinu yang terkait dengan setiap fitur yang berisi nilai numerik(Putra 2019). Berdasarkan uraian yang telah dijelaskan bahwa metode dari model Naïve Bayes Classifier dapat mengklasifikasi data berdasarkan faktor-faktor prbabilitas. Metode ini merupakan solusi terbaik bagi penelitian ini dengan membandingkan akurasi dari metode algoritma Multinomial, Bernoulli dan Gaussian Naive Bayes. Penelitian ini membandingkan metode dari model Naïve Bayes tersebut karena Naïve Bayes merupakan metode pengklasifikasian yang menggunakan faktor-faktor probabilitas.

Maka dari itu pada penelitian ini adalah bentuk pembuktian metode model Naïve Bayes Classifier yaitu model Multnomial, Bernoulli dan Gaussian Naïve Bayes untuk menganalisis sentiment dengan topik hacker Bjorka pada sosial media twitter

dengan membandingkan dari ketiga model Naïve Bayes Classifier tersebut. Dari hal yang telah dijabarkan tadi, nantinya akan menghasilkan nilai akurasi dari masing-masing model dari Naïve Bayes Classifier. Kemudian membuktikan dari ketiga model tersebut yang memiliki kinerja lebih baik untuk digunakan sentimen analisis hacker bjorka pada media sosial twitter.

1.2. Rumusan Masalah

Dari paparan latar belakang masalah diatas, maka dapat dirumuskan permasalahan yaitu “Bagaimana perbedaan akurasi antara Multinomial, Bernoulli, dan Gaussian Naive Bayes untuk analisis sentimen tweet Bjorka pada sosial media twitter?”. Berdasarkan permasalahan tersebut dapat dijabarkan lebih detail, sebagai berikut :

1. Bagaimana perbandingan hasil penerapan model algoritma Multinomial Naïve Bayes, Bernoulli Naïve Bayes dan Gaussian Naïve Bayes untuk melakukan analisis sentiment pada tweet Bjorka, dilihat dari tingkat akurasi, presisi, recal dan f1 score dari ketiga model?
2. Diantara algoritma Multinomial Naive Bayes, Bernoulli Naive Bayes, dan Gaussian Naive bayes, manakah model yang baik digunakan untuk diimplementasikan pada analisis sentimen tweet Bjorka?

1.3. Batasan Masalah

Banyaknya algoritma yang dapat digunakan untuk mengimplementasikan analisis sentimen sehingga penulis merumuskan batasan permasalahan penelitian agar ketidakpastian yang timbul dapat mengurangi, yakni sebagai berikut :

1. Data yang diambil dari sosial media twitter dipakai sebanyak 1.000 dan berfokus pada satu topik yaitu Bjorka.
2. Penelitian hanya berfokus pada perbedaan performa klasifikasi antara Multinomial Naive Bayes, Bernoulli Naive bayes, dan Gaussian Naive

bayes.

1.4. Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui perbandingan algoritma Multinomial Naive Bayes, Bernoulli Naive bayes, dan Gaussian Naive bayes untuk analisis sentimen. Dengan pengimplementasian masing – masing dapat diketahui tujuan dari penelitian ini sebagai berikut :

1. Membandingkan manakah dari model Multinomial, Bernoulli dan Gaussian Naïve Bayes memiliki kinerja lebih baik digunakan untuk mensentimen analisis hacker Bjorka pada sosial media twitter.
2. Membandingkan manakah dari model Multinomial, Bernoulli dan Gaussian Naïve Bayes memiliki nilai akurasi lebih tinggi diliat dari tingkat akurasi, presisi, recal, dan f1 score dari ketiga model tersebut.

1.5. Manfaat Penelitian

Dengan dilakukannya penelitian ini dapat diambil manfaat yaitu mengetahui akurasi yang paling tinggi diantara algoritma Multinomial Naive Bayes, Bernoulli Naive bayes, dan Gaussian Naive bayes saat diimplementasikan terhadap analisis sentiment tweet Bjorka dengan tujuan menentukan kinerja metode dari model Naïve Bayes Classifier. Manfaat yang lain dari penelitian ini terhadap lingkungan atau publik adalah untuk mengetahui sentiment analisis atau opini publik terhadap suatu produk, organisasi, suatu hal yang viral dan lain-lain agar didapatkan kesimpulan bahwa apakah asumsi publik ini menyukai, membenci, atau biasa saja terhadap suatu yang telah ditentukan, dengan melihat berapa besar ukuran dari kesimpulan opini publik yang menyukainya (positif), membencinya (negatif) atau biasa saja (netral).