

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tercatat pada tanggal 12 maret 2022 pemerintah Indonesia telah melaporkan data terkait penambahan kasus harian *Coronavirus Disease 2019* (COVID-19) sebanyak 14.900 kasus baru dan sembuh sebanyak 33.733. Data tersebut dipublikasikan oleh humas Badan Nasional Penanggulangan Bencana (BNPB), yang diperbarui setiap hari dengan waktu *cut off* pukul 12.00 WIB. Dengan penambahan kasus tersebut, jumlah kasus COVID-19 di Indonesia total sejak maret 2020 menjadi 5.878.910 kasus dengan 357.380 di antaranya adalah kasus aktif yaitu pasien dinyatakan masih positif COVID-19. Pemerintah Indonesia tidak hentinya mengimbau warga untuk selalu menaati protokol kesehatan pencegahan COVID-19 yaitu dengan selalu memakai masker, mencuci tangan menggunakan sabun serta menjaga jarak. Program untuk vaksinasi COVID-19 juga sudah dilakukan agar terciptanya kekebalan komunal (*herd immunity*) (Fadhil, 2022).

Salah satu solusi yang diterapkan pemerintah Indonesia untuk pengendalian dan pelacakan kasus COVID-19 adalah dengan menggunakan aplikasi PeduliLindungi. Dilansir dari *website* pedulilindungi.id aplikasi PeduliLindungi merupakan aplikasi yang dikembangkan guna membantu instansi pemerintah dalam melakukan proses pelacakan untuk menghentikan penyebaran COVID-19. Aplikasi PeduliLindungi dalam melakukan pelacakan sangat mengandalkan partisipasi dari masyarakat dengan cara membaca data lokasi

masyarakat saat bepergian sehingga dapat dilakukan penelusuran riwayat kontak dengan penderita COVID-19 (Pedulilindungi.id, 2022).

Sudah hampir satu tahun aplikasi PeduliLindungi digunakan masyarakat dalam melakukan aktivitas, meskipun demikian masih banyak permasalahan yang bersumber dari aplikasi tersebut. Permasalahan yang ada antara lain masih banyak data vaksinasi yang belum muncul di aplikasi, padahal *developer* menjanjikan kemudahan bagi pengguna untuk mengakses sertifikat vaksin yang dibutuhkan pengguna ketika bepergian. Kemudian terdapat juga kasus pencurian data yang terjadi melalui aplikasi PeduliLindungi yang mana data tersebut merupakan data penting pribadi masyarakat yang telah diunggah di aplikasi PeduliLindungi. Aplikasi PeduliLindungi yang dianggap sebagai solusi untuk proses pengendalian dan juga pelacakan kasus COVID-19 masih belum mampu untuk melindungi masyarakat ketika terjadi kecurangan. Terlebih lagi ketika angka penderita COVID-19 mengalami peningkatan masih ada penderita COVID-19 yang dapat memasuki ruang publik yang seharusnya memerlukan izin menggunakan aplikasi PeduliLindungi (Anantatama, 2022).

Pada layanan *Google Play Store*, aplikasi PeduliLindungi memiliki total *download* yang tinggi yaitu lebih dari 50 juta *download*. *Google Play Store* sendiri merupakan sebuah layanan konten digital milik dari *Google* yang berisi produk *online* seperti aplikasi, musik atau lagu, buku, *game* maupun pemutar media berbasis *cloud* (Saputra et al., 2019). *Google Play Store* dilengkapi dengan fitur ulasan bagi pengguna terhadap suatu aplikasi, dimana ulasan atau *review* pengguna tersebut dapat digunakan sebagai media untuk menemukan informasi

terkait suatu aplikasi bagi calon pengguna baru. Ulasan pengguna terhadap aplikasi PeduliLindungi yang ada di *Google Play Store* dapat dianalisis dengan teknik khusus yaitu analisis sentimen, sehingga dapat diklasifikasikan mengenai ulasan masyarakat terhadap aplikasi PeduliLindungi apakah bersifat positif atau negatif. Data ulasan pengguna di *Google Play Store* dapat diambil dengan mudah dan juga data yang tersedia begitu banyak, oleh karena itu pada penelitian kali ini diputuskan untuk menggunakan data ulasan pengguna dari *Google Play Store* terkait aplikasi PeduliLindungi.

Ada banyak metode yang bisa digunakan untuk melakukan proses analisis sentimen. Metode yang bisa digunakan untuk proses analisis sentimen antara lain berbasis pembelajaran mesin atau *machine learning*, berbasis *lexicon* dan metode pendekatan *hybrid* (Sharma et al., 2020). Untuk metode dengan berbasis pembelajaran mesin sendiri dibagi menjadi *supervised learning* dan *unsupervised learning* (Medhat et al., 2014). Proses klasifikasi pada analisis sentimen pada dasarnya adalah permasalahan pada klasifikasi teks, oleh karena itu metode *supervised learning* dapat langsung digunakan dan diimplementasikan seperti dengan menggunakan metode *Support Vector Machine (SVM)* atau *Naïve Bayes Classifier (NBC)* (Liu, 2015). Metode yang telah dikembangkan sebelumnya berguna untuk membantu dalam menganalisis dan juga memprediksi data sentimen teks baik dari data *twitter*, data ulasan pada *Google Play Store* dan data teks lain yang ada di internet sehingga menghasilkan hasil prediksi yang akurat.

Salah satu algoritma yang dapat digunakan adalah *Multinomial Naïve Bayes* yang merupakan variasi lain dari *Naïve Bayes*, *Multinomial Naïve Bayes*

merupakan metode *Probabilistic Reasoning* yang bertujuan untuk melakukan klasifikasi data pada kelas tertentu. Dengan menggunakan metode *Multinomial Naïve Bayes* dapat melakukan proses klasifikasi dengan baik dan memiliki rata-rata tingkat akurasi yang cukup tinggi (Bunga et al., 2018). Seperti pada penelitian yang telah dilakukan terkait klasifikasi artikel *online* tentang gempa di Indonesia dengan menggunakan metode *Multinomial Naïve Bayes* didapatkan akurasi yang tinggi yaitu dengan nilai *f-measure* mencapai 95,20% dan standar deviasi sebesar 1,58% (Sabrani et al., 2020). Kemudian diperkuat dengan penelitian lainnya yaitu terkait analisis sentimen pemindahan ibu kota baru dengan membandingkan metode klasifikasi *Bernoulli Naïve Bayes* dan *Multinomial Naïve Bayes*, dengan nilai performa *Multinomial Naïve Bayes* lebih tinggi dibanding *Bernoulli Naïve Bayes* yaitu 93,45% untuk *Multinomial Naïve Bayes* dan 90,19% untuk *Bernoulli Naïve Bayes* (Wardani et al., 2020).

Algoritma lain yang dapat digunakan untuk melakukan analisis sentimen adalah *Decision Tree*. *Decision Tree* merupakan suatu representasi sederhana untuk klasifikasi, proses pada *Decision Tree* adalah mengubah data dalam bentuk tabel menjadi bentuk pohon atau *tree* lalu mengubah bentuk pohon tersebut menjadi aturan atau *rule* (Rizkia et al., 2019). Dilihat pada beberapa hasil penelitian sebelumnya penerapan *Decision Tree* memiliki nilai akurasi yang tinggi dibandingkan algoritma lainnya. Seperti pada penelitian terkait analisis sentimen efek PSBB pada *twitter*, *Decision Tree* memiliki nilai akurasi tertinggi dibandingkan dengan *KNN* dan *Naïve Bayes* yaitu dengan nilai akurasi 83,3 % untuk *Decision Tree*, 80,80% untuk *KNN* dan 80,03% untuk *Naïve Bayes*

(Syarifuddin, 2020). Diperkuat dengan penelitian tentang perbandingan *KNN*, *Decision Tree* dan *Naïve Bayes* terhadap analisis sentimen pengguna layanan BPJS, dengan hasil nilai tingkat akurasi *Decision Tree* lebih tinggi dari algoritma lainnya yaitu dengan nilai 96,13% untuk *Decision Tree*, 95,58% untuk *KNN* dan 89,14% untuk *Naïve Bayes* (Puspita & Widodo, 2021).

Permasalahan dalam *machine learning* menurut Sommerville yaitu terdapat pada tingkat keakuratan yang relatif rendah (Adhitya et al., 2015). Pada penelitian terkait analisis sentimen terhadap ulasan pengguna aplikasi informasi COVID-19 yaitu PeduliLindungi dengan menggunakan metode *Naïve Bayes Classifier*, SVM dan KNN menghasilkan nilai akurasi untuk SVM sebesar 76,5%, diikuti oleh NBC 72,3%, dan KNN memiliki akurasi 59,1%, sehingga perlu metode lain dalam melakukan analisis sentimen dan klasifikasi lebih lanjut untuk meningkatkan hasil akurasi (Salma & Silfianti, 2021).

Untuk meningkatkan akurasi hasil pengujian dan juga mengurangi bias harus ditambahkan sebuah metode, salah satu metode yang dapat digunakan adalah metode *ensemble* yaitu dengan menggunakan *AdaBoost (Adaptive Boosting)* (Hamzah, 2021). Penggunaan *AdaBoost* untuk meningkatkan akurasi pada algoritma klasifikasi sudah terbukti, seperti pada penelitian terkait analisis sentimen *review* restoran dengan menggunakan algoritma *Naïve Bayes* yang awalnya hanya memiliki nilai akurasi 70% meningkat menjadi 99,5% setelah menggunakan *AdaBoost* (Dini et al., 2015).

Berdasarkan uraian sebelumnya, maka akan dilakukan analisis sentimen dengan menggunakan algoritma *Multinomial Naïve Bayes* dan *Decision Tree*

dengan menerapkan *boosting AdaBoost* untuk meningkatkan akurasi dan membandingkan hasil dari kedua algoritma sebelum dan setelah penerapan *AdaBoost* dengan mengambil judul “Perbandingan Akurasi Algoritma *Multinomial Naïve Bayes* Dan *Decision Tree* Dengan Menerapkan *Adaboost* Pada Analisis Sentimen Ulasan Aplikasi Pedulilindungi”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, adapun rumusan masalah pada penelitian ini adalah:

1. Bagaimana akurasi algoritma *Multinomial Naïve Bayes* sebelum dan setelah penerapan *AdaBoost*?
2. Bagaimana akurasi algoritma *Decision Tree* sebelum dan setelah penerapan *AdaBoost*?
3. Bagaimana akurasi algoritma *Multinomial Naïve Bayes* dan *Decision Tree* berdasarkan jumlah $n_estimators$ pada *AdaBoost*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Algoritma yang digunakan dalam penelitian ini yaitu algoritma *Multinomial Naïve Bayes* dan *Decision Tree* untuk menganalisis sentimen terhadap ulasan aplikasi PeduliLindungi
2. Menerapkan *AdaBoost* pada algoritma *Multinomial Naïve Bayes* dan *Decision Tree*.

3. Klasifikasi sentimen dikelompokkan menjadi sentimen positif, netral dan negatif.
4. Data yang digunakan berasal dari ulasan pengguna aplikasi PeduliLindungi pada *Google Play Store* dengan kriteria paling relevan.
5. Pelabelan kelas sentimen dilakukan secara otomatis dengan menggunakan *library vaderSentiment*.
6. Bahasa pemrograman yang digunakan untuk mengimplementasikan algoritma sentimen terhadap ulasan aplikasi PeduliLindungi adalah *Python*.

1.4 Tujuan Penelitian

Tujuan pada penelitian ini adalah:

1. Mengukur akurasi algoritma *Multinomial Naïve Bayes* sebelum dan setelah penerapan *AdaBoost*.
2. Mengukur akurasi algoritma *Decision Tree* sebelum dan setelah penerapan *AdaBoost*.
3. Mengukur akurasi algoritma *Multinomial Naïve Bayes* dan *Decision Tree* berdasarkan jumlah *n_estimators* pada *AdaBoost*.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Dapat memberikan pengetahuan bagi pembaca mengenai apakah penerapan *AdaBoost* dapat meningkatkan akurasi dari algoritma *Multinomial Naïve Bayes* dan *Decision Tree*.

2. Dapat memberikan pengetahuan bagi pembaca mengenai perbandingan algoritma *Multinomial Naïve Bayes* dengan *Decision Tree* sebelum dan setelah penerapan *AdaBoost*.
3. Dapat memberikan pengetahuan bagi pembaca terkait akurasi dari algoritma *Multinomial Naïve Bayes* dan *Decision Tree* berdasarkan jumlah *n_estimators* pada *AdaBoost*.
4. Dapat menjadi referensi ilmiah bagi peneliti di masa yang akan datang, khususnya untuk penelitian pada bidang *machine learning* dan analisis sentimen.
5. Dapat memberikan informasi bagi pembaca terkait kecenderungan sentimen masyarakat terhadap aplikasi PeduliLindungi.
6. Hasil dari analisis sentimen dapat digunakan oleh pemerintah sebagai bahan evaluasi terkait kebijakan pengendalian dan pelacakan kasus COVID-19 di Indonesia menggunakan aplikasi PeduliLindungi.

1.6 Metodologi Penelitian

Metodologi penelitian menjelaskan mengenai tahapan-tahapan penelitian yang dilalui dalam proses penelitian. Tahapan-tahapan tersebut terdiri dari:

1. Studi Literatur

Pada Tahap ini mempelajari semua hal yang berkaitan dengan analisis sentimen dan penerapan menggunakan bahasa pemrograman *python* dari berbagai sumber pustaka berupa buku, jurnal, tugas akhir, surat kabar digital serta hasil pencarian pustaka di internet.

2. *Data Collecting*

Data Collecting adalah tahap pengambilan data ulasan pengguna aplikasi PeduliLindungi di *Google Play Store* dengan menggunakan *library Google Play Scraper*.

3. *Data Processing*

Data Processing adalah proses untuk mengolah data yang didapat dari tahap *Data Collecting* dengan tujuan untuk melakukan pengontrolan ukuran daftar kata dan diharapkan dapat meningkatkan performa pada proses klasifikasi yang dilakukan.

4. *Feature Extraction*

Pada tahap ini dilakukan proses ekstraksi fitur dengan dua proses yaitu *CountVectorizer* dan *TF-IDF*.

5. *Text Classification*

Pada proses *Text Classification* dilakukan proses *data training* dan *data testing* pada data ulasan aplikasi PeduliLindungi. Pengujian yang dilakukan menggunakan *K-Folds Cross Validation* dengan nilai k yang digunakan adalah 5.

6. Analisis Hasil

Pada tahap ini dilakukan evaluasi terhadap model yang dibuat, kemudian dilakukan perbandingan berdasarkan nilai akurasi tertinggi dan perbandingan nilai akurasi berdasarkan nilai $n_estimators$ pada model yang digunakan dalam penelitian ini.

1.7 Sistematika Penulisan

Sistematika penulisan yang dilakukan dalam penulisan tugas akhir ini dapat diuraikan sebagai berikut:

BAB I : PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini membahas tentang uraian penelitian terkait serta teori-teori dan konsep-konsep yang berhubungan dan mendukung penelitian yang dilakukan. Sumber referensi yang menjadi acuan adalah buku, jurnal dan media elektronik.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas tentang metodologi penelitian dan langkah-langkah selama mengerjakan penelitian. Metodologi penelitian terdiri dari studi literatur, *data collecting*, *data processing*, *feature extraction*, *text classification* dan analisis hasil.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini membahas hasil dari proses analisis yang telah dilakukan serta pembahasan yang meliputi data yang akan dianalisis dan juga penerapan metode yang digunakan.

BAB V : KESIMPULAN DAN SARAN

Bab ini membahas mengenai kesimpulan dari hasil penelitian dan saran untuk penelitian di masa mendatang.