

BABI

PENDAHULUAN

1.1. Latar Belakang

Depresi merupakan kondisi di mana seseorang merasa kecewa atau sedih ketika mengalami kegagalan, kehilangan, atau perubahan dan menjadi tidak wajar ketika mengalami ketidakmampuan untuk beradaptasi (Rosyanti et al., 2018). Stres kronis dan peristiwa kehidupan yang penuh tekanan di awal kehidupan sangat kuat prediktor proksimal dari timbulnya depresi (Shadrina et al., 2018). Gangguan depresi mayor adalah penyakit yang kompleks dan heterogen dengan etiologi yang didasarkan pada banyak faktor yang dapat bertindak pada tingkat yang berbeda, termasuk psikologis, biologis, genetik, dan sosial. Dijelaskan dengan berkurangnya minat atau kesenangan pada kegiatan yang biasa dilakukan sebelumnya (Rosyanti et al., 2018).

Menurut WHO atau yang biasa kita sebut Organisasi Kesehatan Dunia, depresi adalah masalah paling umum keempat di bumi. Kurang lebih 20% perempuan dan 12% laki-laki mengalami depresi di beberapa titik dalam hidup mereka. Diperkirakan kurang lebih 300 juta orang, atau 4,4% populasi dunia, menderita depresi pada tahun 2015 (Priya et al., 2020). Tinggal di Asia Tenggara sekitar setengah (27%) dari mereka memiliki beresiko (Tadesse et al., 2019). Saat ini diperkirakan sekitar 15,6 juta orang di Indonesia menderita depresi. Angkanya terus meningkat, menjadikan depresi sebagai penyakit kedua terbanyak setelah penyakit jantung. Di Indonesia diperkirakan akan terus meningkat sepanjang tahun. Depresi yang terus-menerus dapat menyebabkan bunuh diri. Indonesia dianggap sebagai negara dengan jumlah bunuh diri tertinggi di Asia Tenggara. Tentu saja, depresi dapat menyerang siapa saja tanpa kecuali, jadi Anda harus memperhatikan fakta ini.

Perkembangan teknologi informasi dan komunikasi terus meningkat seiring dengan meningkatnya kebutuhan manusia, dan tidak jarang para pengguna media sosial membagikan informasi tentang dirinya dan keluhan yang dialaminya di media sosial. Pada titik ini, deteksi dapat didasarkan pada aktivitas pengguna media sosial itu sendiri. Pengguna media sosial twitter sering membagikan

informasi tentang diri mereka dan keluhan yang mereka alami di Twitter. Oleh karena itu, Twitter merupakan pilihan untuk mengidentifikasi tingkat kesehatan mental seseorang.

Twitter adalah layanan microblogging yang memungkinkan penggunanya untuk memperbarui informasi dan berbagi pendapat tentang berbagai topik dan isu yang muncul (Sholikha et al., 2020). Pengguna aktif twitter pada tahun 2016 lebih dari empat ratus juta. Twitter memiliki Fitur utama cuitan atau tweet, yang memungkinkan pengguna menulis catatan mengenai keadaan yang sedang terjadi. Sifat Twitter yang real-time menjadikannya sosial media memiliki daya tarik untuk digunakan metode pendeteksian peristiwa. Tujuan utama analisis sentimen adalah untuk mendeteksi depresi pada pengguna Twitter. Dengan menganalisis tweet, penulis dapat menentukan seberapa besar kemungkinan seseorang menderita salah satu penyakit di atas. Data ini sangat membantu para dokter dalam merawat pasiennya. Selain itu, ahli forensik dapat menggunakan data ini untuk mengetahui apakah seseorang melakukan bunuh diri atau tidak. Selain itu, data ini dapat berfungsi sebagai peringatan bagi keluarga yang terkena dampak. Lebih banyak postingan negatif dan lebih sedikit aktivitas sosial adalah faktor kunci bagi orang-orang seperti itu (Tiwari et al., 2021).

Algoritma *Random Forest* adalah salah satu algoritma pembelajaran berkinerja terbaik, lebih akurat memperkirakan tingkat kesalahan dibandingkan dengan *decision tree*. Karena tingkat kesalahan telah terbukti secara matematis (Schonlau & Zou, 2020). Model Random Forest adalah kumpulan *decision tree*, digunakan untuk klasifikasi atau regresi. Dalam kasus klasifikasi digunakan untuk prediksi, didasarkan pada suara terbanyak dari nilai prediksi menggunakan *decision tree*, dan dalam kasus regresi, hasilnya adalah rata-rata dari hasil *tree* (Resende & Drummond, 2018).

Algoritma *Extreme Gradient Boosting (XGBoost)* adalah salah satu algoritma berkinerja terbaik yang digunakan untuk *supervised learning*. Dapat digunakan untuk masalah regresi dan klasifikasi. *Xgboost* disukai oleh ilmuwan karena kecepatan eksekusinya yang tinggi (Ibrahim Ahmed Osman et al., 2021). Ide utama algoritma *boosting* adalah menggabungkan output pelajar yang lemah

secara berurutan untuk mencapai kinerja yang lebih baik. *XGBoost* menggunakan banyak pohon klasifikasi dan regresi dan mengintegrasikannya menggunakan metode peningkatan gradien (Can et al., 2021).

Berdasarkan dari permasalahan yang ada, tugas akhir ini akan mengimplementasikan algoritma klasifikasi untuk mendeteksi depresi pada pengguna Twitter. Ia menggunakan algoritma *random forest* dan *extreme gradient boosting (XGBoost)*. Kedua algoritma ini diusulkan karena pendekatan dua arah dapat mengekstraksi informasi kontekstual lebih cepat sehingga tidak kehilangan makna dan konteks kalimat. Studi ini diharapkan dapat meningkatkan penyembuhan bagi pengguna media sosial Twitter yang depresi.

1.2. Rumusan Masalah

Adapun rumusan masalah dalam makalah ini dapat kita ambil dari latar belakang proposal ini, yaitu sebagai berikut:

1. Bagaimana membuat sebuah sistem yang mampu mendeteksi depresi pada pengguna twitter?
2. Bagaimana hasil pengujian sistem Algoritma *Random Forest* dan Algoritma *XGBoost* menggunakan *confussion matrix* dan *K-Fold cross validation*?

1.3. Batasan Masalah

Berdasarkan batasan masalah di atas, maka permasalahan yang diuraikan dibatasi pada batasan masalah sebagai berikut:

1. Ambil data menggunakan API KEY penulis menggunakan antarmuka pemrograman aplikasi (API).
2. Penelitian berfokus pada analisis kicauan pengguna Twitter berbahasa Indonesia tanpa menganalisis aktivitas Twitter yaitu gambar, video, dan tautan.
3. Rentang waktu *crawling* data yaitu 19 Agustus 2022 sampai dengan 28 Agustus 2022, dengan menggunakan kata kunci Depresi
4. Menggunakan algoritma *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*

5. Menggunakan bahasa pemrograman *Python*
6. Analisis sentimen dengan pelabelan otomatis menggunakan *Valence Aware Dictionary and sEntiment Reasoner (Vader)*
7. Menggunakan *Library Tweepy* untuk *crawling* data
8. Menggunakan *Library Satrawi* untuk melakukan *stemming* bahasa Indonesia.
9. Menggunakan *TF-IDF* untuk pembobotan kata
10. Menggunakan *Confussion matrix* untuk evaluasi kinerja sistem

1.4. Tujuan

Tujuan dari penelitian ini adalah untuk menjawab gambaran masalah yang telah dikemukakan sehingga penelitian dapat dilakukan berdasarkan tujuan yang telah ditetapkan. Tujuan yang dicapai dalam studi ini adalah:

1. Menganalisis Algoritma Kalsifikasi *Random Forest* dan *Extreme Gradient Boosting (XGBoost)* untuk Mendeteksi Depresi pada Pengguna Twitter
2. Menguji perfomansi klasifikasi dalam penerapan algoritma *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*.

1.5. Manfaat

Dengan tercapainya tujuan penelitian di atas, maka diharapkan akan bermanfaat bagi semua pihak, manfaat dari penelitian ini adalah:

1. Memberikan informasi mengenai deteksi depresi pada pengguna twitter.
2. Dapat mempermudah pengguna twitter dalam mendeteksi depresi sehingga dapat ditangani dengan tepat.