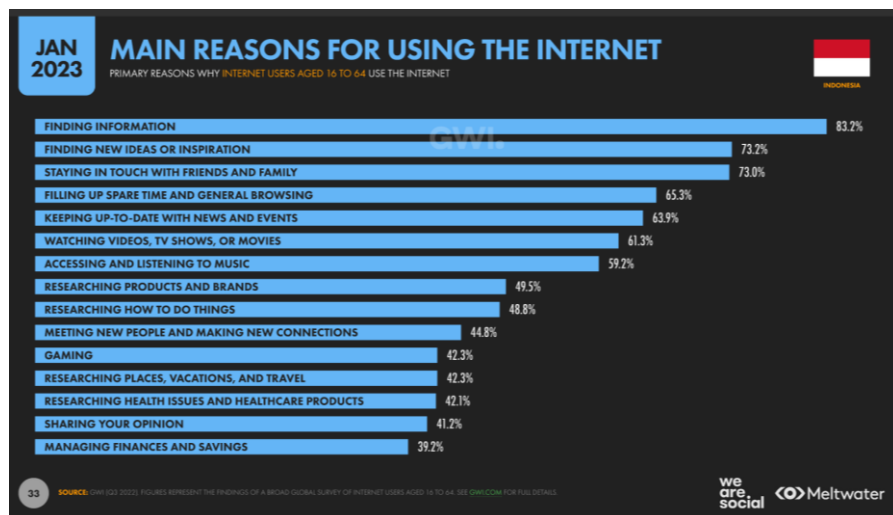


BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam Kamus Besar Bahasa Indonesia, teknologi adalah sarana dalam mempersiapkan suatu barang yang dibutuhkan untuk keberlangsungan hidup manusia. Teknologi informasi merupakan salah satu cabang dari teknologi. Pada zaman teknologi informasi masa kini, manusia menggunakan media internet untuk berbagai tujuan. Salah satu negara yang memiliki jumlah pengguna internet aktif cukup banyak di dunia adalah Indonesia yaitu sebanyak 212.9 juta pengguna atau sekitar 77% dari jumlah total populasi di Indonesia pada tahun 2023 [1]. Para pengguna tersebut menggunakan internet karena internet adalah media yang mudah untuk mendapatkan sebuah informasi dikarenakan perkembangannya bersamaan dengan teknologi yang juga semakin berkembang. Seperti pada Gambar 1.1 terdapat berbagai alasan utama para pengguna menggunakan internet, beberapa diantaranya yaitu untuk mencari sebuah informasi, berita terkini, dan membagikan opini atau pendapat kita [2]. Dari ketiga alasan utama tersebut masing-masing memiliki jumlah yaitu diangka 83.2%, 63.9%, dan 41.2% [2].



Gambar 1.1 Alasan Utama Menggunakan Internet

Media sosial merupakan sebuah alat komunikasi yang dapat menghasilkan sebuah interaksi antar pengguna secara online seperti berbagi dan menciptakan suatu hal serta dapat digunakan oleh masyarakat di seluruh dunia [3]. Selain menggunakan internet browser, di media sosial pengguna juga dapat mencari

sebuah informasi, berita terkini, maupun menulis opini pribadi mereka terhadap suatu hal. Para pengguna dapat menyampaikan pendapat atau opini mereka terhadap suatu hal yang dia rasakan dan yang terjadi di masyarakat. Dengan semakin banyaknya pengguna media sosial dan internet browser, dapat memberikan dampak positif maupun negatif pada informasi serta opini publik. Para pengguna akan mendapatkan informasi maupun berita terkini yang mereka butuhkan dengan serba cepat dan tepat. Untuk mencari beberapa hal tersebut para pengguna dapat menggunakan berbagai macam platform media sosial seperti Instagram, Tiktok, Facebook, Youtube, Whatsapp, dan salah satunya yaitu Twitter.

Twitter adalah salah satu platform media sosial yang dimanfaatkan oleh manusia untuk sarana komunikasi dan penyebaran informasi. Berdasarkan laporan dari *Data Reportal We Are Social*, pada bulan Januari tahun 2023 Indonesia menempati urutan ke-5 yang memiliki pengguna aktif di Twitter sebanyak 24 juta pengguna aktif [4]. Pada media sosial Twitter, terdapat pesan yang dapat diunggah ke halaman Twitter atau profil pengguna yang dapat berisikan foto, video, link atau tautan, dan teks yang disebut juga dengan *Tweets*. *Tweets* ini memiliki batas maksimal penggunaan karakter dalam sekali unggah yaitu sebanyak 280 karakter. *Tweets* ini dapat digunakan oleh pengguna untuk menyampaikan informasi, mengungkapkan opini pribadi mereka terhadap suatu hal maupun perasaan atau emosi yang sedang mereka rasakan baik itu rasa senang, cinta, sedih, marah, maupun rasa ingin memuji atau mencela [5]. Salah satu contoh *tweets* mengenai opini pribadi terhadap suatu produk atau layanan yaitu tentang kualitas layanan jaringan internet yang diberikan oleh provider tertentu. *Tweets* ini dapat berupa penyampaian keluhan ataupun menyanjung dari produk tersebut. Dari *tweets* ini dapat digunakan sebagai data untuk penelitian yang menggunakan topik analisis sentimen.

Analisis sentimen merupakan sebuah proses yang bertujuan untuk memahami, mengolah informasi yang berbentuk teks, menghasilkan sebuah teks yang dilakukan secara otomatis untuk memperoleh informasi yang didalamnya terdapat kandungan sentimen pada teks tersebut [6]. Analisis sentimen ini dilakukan untuk mengetahui tentang opini pribadi masyarakat tersebut lebih cenderung kepada opini yang positif atau opini yang negatif atau opini yang netral [6]. Sehingga dengan

adanya analisis sentimen ini dapat membantu perusahaan untuk melihat bagaimana respon atau penilaian dari pengguna atas barang atau layanan mereka, maka dari itu perusahaan dapat menciptakan rencana-rencana yang aktual untuk masa yang akan datang [5]. Selain digunakan pada teks yang berisi opini pribadi terhadap produk atau layanan dari sebuah perusahaan, analisis sentimen juga dapat digunakan untuk menilai sebuah teks yang berasal dari sebuah topik atau peristiwa yang sedang terjadi, lokasi atau suatu tempat, dan opini pribadi untuk seseorang ataupun untuk organisasi.

Algoritma *Machine Learning* dapat diimplementasikan untuk membantu dalam menghasilkan akurasi untuk klasifikasi sebuah data pada analisis sentimen, beberapa diantaranya yaitu *Naïve Bayes*, *K-Nearest Neighbour* (KNN), *Random Forest*, dan *Support Vector Machine* (SVM), dan lain-lain [7]. Akurasi dari klasifikasi data tersebut menjadi salah satu hal yang paling dicari dalam sebuah data mining. Berdasarkan penelitian sebelumnya, *K-Nearest Neighbour* (KNN) adalah sebuah algoritma *supervised learning* yang menggunakan prinsip kerja dalam mempraktikkan klasifikasi pada objek yang bersumber dari data pembelajaran memiliki jarak paling dekat dengan objek [8]. Algoritma KNN ini menggunakan prinsip kerja yang mendapatkan jarak paling dekat dengan data yang diuji dengan K tetangga (*neighbour*) terdekatnya pada data latih. Algoritma KNN ini ditemukan oleh Fix dan Hodges pada tahun 1951. Pada penelitian lain yang menggunakan algoritma KNN, data yang digunakan pada penelitian ini yaitu mengenai sentimen masyarakat terhadap larangan mudik 2021 pada twitter dengan jumlah data sebanyak 4.799 data dengan nilai $k = 3$ menghasilkan nilai akurasi sebesar 86.67% [9].

Algoritma *Support Vector Machine* (SVM) adalah salah satu algoritma pembelajaran mesin yang diciptakan oleh Vladimir Vapnik bersama rekannya yaitu Bernhard Boser dan Isabelle Guyon pada tahun 1992 yang memiliki dasar dari teori *Structural Risk Minimization* [10]. Algoritma SVM ini adalah salah satu algoritma baru yang mampu digunakan untuk pengklasifikasian data, baik linear maupun non-linear. Algoritma ini adalah salah satu algoritma *supervised learning* yang berarti algoritma ini dapat mengetahui suatu data dari model label tertentu yang ada pada dataset yang digunakan. SVM ini banyak diimplementasikan dalam penelitian

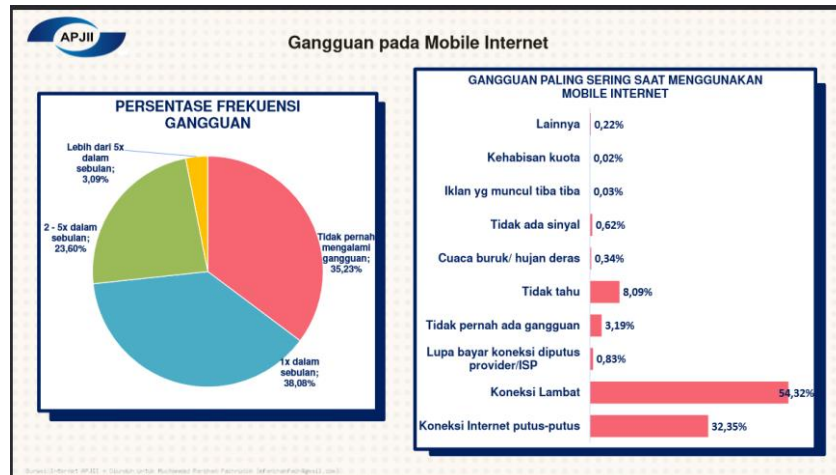
dalam studi klasifikasi teks, salah satu kelebihan dari SVM ini yaitu mampu menghasilkan nilai yang lebih baik pada sebuah teks yang berisi opini [6]. Pada penelitian sebelumnya, menggunakan data yang berasal dari twitter mengenai sentimen masyarakat Indonesia terkait vaksin covid-19 pada media sosial twitter, dengan jumlah data yang digunakan sebanyak 5000 data, dengan menggunakan kernel *linear*, *RBF*, *sigmoid*, dan *polynomial*, dari empat kernel tersebut menghasilkan nilai akurasi pada kernel *linear* sebesar 88.3%, pada kernel *RBF* menghasilkan nilai akurasi sebesar 88.8%, pada kernel *sigmlid* menghasilkan nilai akurasi sebesar 87%, pada kernel *polynomial* menghasilkan nilai akurasi sebesar 85.5% [11].

Salah satu penelitian yang menggunakan algoritma *K-Nearest Neighbour* dan *Support Vector Machine* pada data yang berasal dari media sosial twitter adalah [12], data yang digunakan pada penelitian ini yaitu terkait sentimen terhadap persepsi masyarakat Indonesia terkait Pandemi Covid 19 untuk digunakan sebagai analisis sentimen dengan menggunakan tiga algoritma *Support Vector Machine* dengan menggunakan kernel *linear*, *Naïve Bayes*, dan *K-Nearest Neighbour* dengan menggunakan teknik evaluasi *K-Fold Cross Validation* yang menghasilkan nilai akurasi rata-rata pada algoritma *Support Vector Machine* dengan nilai 90.1% sedangkan rata-rata nilai akurasi pada algoritma *K-Nearest Neighbour* yaitu 62.10%. Pada penelitian yang lain yang juga menggunakan data yang berasal dari media sosial twitter yaitu mengenai vaksin *Sinovac*, pada penelitian ini menggunakan algoritma SVM dan KNN dengan menggunakan 2105 data yang menggunakan bahasa inggris, kernel yang digunakan pada algoritma SVM yaitu kernel *Linear*, kernel *Polynomial*, dan kernel *RBF*, sedangkan algoritma KNN menggunakan jumlah n sebanyak 3, 5, dan 7. Hasil pengujian dari masing masing algoritma tersebut dievaluasi dengan memakai *K-Fold Cross Validation* dengan jumlah *Fold* sebesar 10 *fold*. Didapatkan nilai akurasi untuk algoritma SVM untuk kernel *Linear* dengan nilai 0.70, untuk kernel *Polynomial* dengan nilai 0.57, dan untuk kernel *RBF* dengan nilai 0.66, sedangkan pada algoritma KNN pada jumlah n 3 menghasilkan nilai akurasi sebesar 0.55, pada jumlah n 5 menghasilkan nilai akurasi sebesar 0.55, dan pada jumlah n 7 menghasilkan nilai akurasi 0.56.

Sehingga dapat disimpulkan bahwa algoritma SVM menghasilkan nilai akurasi terbaik sebesar 0.7 sedangkan *KNN* menghasilkan nilai akurasi sebesar 0.56 [13].

Peneliti menambahkan metode *ensemble* untuk menggabungkan kedua algoritma tersebut. Metode *ensemble* merupakan sebuah metode pembelajaran mesin yang menggabungkan dari beberapa model untuk menghasilkan sebuah solusi untuk memecahkan masalah tertentu dan akan memberikan hasil yang lebih baik daripada hanya menggunakan satu model saja [14]. Metode *ensemble* merupakan sebuah metode yang mampu untuk menggabungkan beberapa *single classifier* sehingga terbentuklah suatu *classifier* baru yang mampu untuk meningkatkan nilai akurasi yang telah diciptakan [15].

Studi kasus yang digunakan oleh peneliti yaitu opini publik terhadap kualitas layanan jaringan internet provider Telkomsel. Peneliti menggunakan studi kasus tersebut dikarenakan Telkomsel merupakan operator seluler yang memiliki persentase pengguna terbanyak yaitu sebanyak 41.94% untuk mengakses internet melalui smartphone pada tahun 2021-2022, persentase tersebut didapatkan dari survei yang dilaksanakan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) [16]. Selain itu, terdapat gangguan yang paling sering digunakan pada saat menggunakan *mobile* internet yaitu koneksi yang lambat dan koneksi internet yang putus putus. Kedua gangguan tersebut memiliki persentase 54.32% dan 32.35%, akan tetapi persentase ini dihitung tidak hanya pada operator seluler Telkomsel saja melainkan dengan operator seluler yang lain seperti XL Axiata, Indosat Ooredoo, 3 (Three), dan Smartfren. Persentase tersebut diperoleh dari survei yang dilaksanakan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) dan dapat dilihat pada Gambar 1.2.



Gambar 1.2 Gangguan pada Mobile Internet

Pada penelitian ini difokuskan terhadap pengguna telkomsel yang menggunakan layanan paket internet telkomsel secara keseluruhan. Oleh karena itu peneliti ingin membuktikan apakah kualitas layanan jaringan internet yang diberikan oleh operator seluler Telkomsel cukup baik sehingga Telkomsel menjadi provider operator seluler dengan persentase pengguna paling banyak.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Di Indonesia terdapat beberapa macam perusahaan provider yang memberikan jasa layanan internet salah satu yang menyediakan dan yang banyak digunakan yaitu PT Telekomunikasi Seluler. Perusahaan ini memberikan banyak layanan jasa jaringan telekomunikasi salah satunya yaitu memberikan layanan jasa jaringan internet melalui operator seluler Telkomsel. Dikarenakan banyak yang menggunakan operator seluler Telkomsel maka umumnya pengguna dari operator seluler tersebut akan menyampaikan opini mereka terhadap kualitas layanan jaringan internet yang diberikan. Terdapat persentase gangguan yang paling sering dialami saat menggunakan mobile internet yaitu koneksi yang lambat dan koneksi internet yang putus-putus. Kedua gangguan tersebut memiliki persentase sebesar 54.32% dan 32.35%, persentase tersebut didapatkan dari hasil survei yang dilaksanakan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII). Selain itu juga peneliti juga merasakan hal yang sama terkait koneksi yang lambat, oleh karena itu peneliti menggunakan operator seluler sebagai topik pada penelitian ini.

1.2.2 Metode Usulan

Pada penelitian ini mengusulkan penggunaan metode algoritma *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) dengan kernel yang digunakan yaitu kernel *Radial Basis Function*, dan *ensemble learning* dari KNN dan SVM yang berbasis *majority voting* dan berbasis *average*.

1.2.3 Pertanyaan Penelitian

Berdasarkan permasalahan diatas maka dapat dirumuskan beberapa pertanyaan pada penelitian ini yaitu:

1. Bagaimana hasil akurasi yang didapatkan pada analisis sentimen dengan mengimplementasikan algoritma *K-Nearest Neighbor*, *Support Vector Machine*, dan *K-Nearest Neighbor-Support Vector Machine* yang di *Ensemble* berbasis *majority voting* dan berbasis *average* dengan menggunakan data *tweets* opini publik mengenai kualitas layanan jaringan internet Telkomsel?
2. Bagaimana hasil sentimen dari implementasi algoritma *K-Nearest Neighbor* dan *Support Vector Machine* dalam melakukan klasifikasi data *tweets* opini publik mengenai kualitas layanan jaringan internet Telkomsel?

1.3 Tujuan Penelitian dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Mengetahui hasil akurasi pada penerapan algoritma *K-Nearest Neighbor*, *Support Vector Machine*, dan *K-Nearest Neighbor-Support Vector Machine* yang di *Ensemble* berbasis *majority voting* dan berbasis *average* dengan menggunakan data *tweets* opini publik mengenai kualitas layanan jaringan internet Telkomsel.
2. Untuk mengetahui hasil sentimen dari implementasi algoritma KNN dan SVM dengan variasi kernel *Radial Basis Function* dalam melakukan klasifikasi data *tweets* opini publik mengenai kualitas layanan jaringan internet Telkomsel.

1.3.2 Manfaat Penelitian

Berdasarkan tujuan yang telah dijelaskan maka adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Manfaat bagi pihak perusahaan

Menjadi bahan evaluasi untuk pihak pengelola layanan internet PT Telekomunikasi Seluler untuk terus meningkatkan kualitas layanan internet yang diberikan.

2. Manfaat bagi pembaca

Adapun beberapa manfaat yang dapat diambil oleh pembaca adalah sebagai berikut:

- a. Sebagai bahan studi penelitian dan tambahan wawasan pada topik analisis sentimen
- b. Sebagai referensi tambahan untuk mengetahui nilai akurasi yang diberikan oleh algoritma *SVM*, *KNN*, dan yang telah di *Ensemble*.

3. Manfaat bagi penulis

Manfaat dilakukan penelitian ini bagi penulis yaitu tambahan wawasan terkait analisis sentimen menggunakan algoritma *SVM*, *KNN*, serta yang telah di *Ensemble* dan pengimplementasian ilmu pengetahuan yang telah diperoleh pada saat belajar di tingkat universitas.

1.4 Batasan Masalah

Adapun batasan-batasan masalah pada penelitian ini adalah:

1. Studi kasus penelitian adalah opini publik terhadap kualitas layanan jaringan internet pada provider Telkomsel.
2. Data *tweets* yang digunakan yaitu data *tweets* pada tanggal 7 Juli 2020 hingga 31 Desember 2022.
3. Data *tweets* yang digunakan hanya *tweets* yang menggunakan bahasa Indonesia.
4. Penelitian ini menggunakan 2 keyword yaitu jaringan telkomsel dan sinyal telkomsel.
5. Bahasa pemrograman yang digunakan untuk mengimplementasikan metode yang digunakan yaitu Python

6. *Library* Python yang digunakan yaitu *snsrape*, *nlTK*, *scikit-learn*, *sastrawi*, *pandas*, dan *numpy*.
7. Hasil teks sentimen akan dibagi menjadi tiga kelas yaitu positif, negatif, netral.
8. Pada operator seluler atau mobile internet Telkomsel.
9. Jaringan internet telkomsel dengan seluruh layanan paket internet yang tersedia.

1.5 Sistematika Penulisan

Penulisan proposal skripsi ini disusun secara runtut dan dikelompokkan menjadi beberapa BAB, seperti berikut ini:

BAB I PENDAHULUAN

Pada bab pendahuluan ini berisi tentang latar belakang yang akan dilakukan pada penelitian, rumusan masalah penelitian, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan laporan.

BAB II TINJAUAN PUSTAKA

Pada bab tinjauan pustaka atau dapat disebut juga dengan landasan teori, berisi tentang analisis sentimen, twitter, metode yang digunakan pada skripsi ini, dan penelitian yang relevan dengan topik yang diambil.

BAB III METODE PENELITIAN

Pada bab metode penelitian, berisi tentang tahapan-tahapan yang akan dilaksanakan dalam penelitian ini, cara mengumpulkan dan mengolah data yang digunakan dalam penelitian, serta penerapan metode pengolahan data dalam penelitian ini.

BAB IV PEMBAHASAN

Pada bab pembahasan ini, menjelaskan lebih detail mengenai pengimplementasian dan hasil klasifikasi sentimen yang dibentuk oleh model algoritma K-Nearest Neighbour (KNN), Support Vector Machine (SVM) dengan variasi kernel *Radial Basis Function* (RBF), dan *Ensemble* KNN-SVM dengan menggunakan model *Majority Vote* atau voting mayoritas dan *average* atau rata-rata.

BAB V PENUTUP

Pada bab penutup ini, berisi kesimpulan yang didapatkan dari penelitian yang telah dan saran untuk pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA