

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang begitu pesat membawa peradaban manusia menuju ke era yang serba digital. Sejalan dengan perkembangan tersebut, tentunya membawa dampak yang positif. Salah satunya adalah perkembangan teknologi internet dengan pengguna yang terus bertambah setiap tahunnya. Berdasarkan hasil survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada periode 2021-2022 mengalami kenaikan dari tahun sebelumnya yang mencapai sekitar 210,03 juta pengguna internet di dalam negeri. Hasil survei ini juga memberikan hasil bahwa penggunaan internet paling banyak di akses adalah ke situs media sosial (89,15%), terbanyak kedua adalah *chatting online* (73,86%), dan terbanyak ketiga adalah penggunaan internet untuk belanja *online* (21,26%). (Survei APJII, 2022).

Kemajuan teknologi saat ini diciptakan untuk membantu manusia mengurangi beban dalam kehidupan sehari-hari dan memberikan informasi secara cepat, akurat dan ringkas. (Nidaul Khasanah et al., 2020). Kebutuhan akan sumber informasi semakin meningkat dan informasi menjadi sesuatu yang sangat penting terutama bagi semua orang. Internet kini banyak digunakan untuk mencari informasi yang ditawarkan dari berbagai *website*, antara lain informasi kebijakan, informasi produk, maupun situs penyedia berbagai produk yang dijual. (Narulita, 2018). Kemudahan dan kepraktisan yang ditawarkan internet tidak hanya terbatas

dalam pencarian informasi, namun juga peranan teknologi informasi berpengaruh dalam dunia ekonomi, khususnya dalam hal transaksi jual dan beli.

Dari sekian banyaknya penyedia layanan *marketplace* di Indonesia salah satunya adalah PT.Tokopedia yang merupakan salah satu perusahaan *e-commerce* terbesar dan memiliki keunikan yang luar biasa dalam menjalankan bisnisnya. Sejak didirikan pada tahun 2009 Tokopedia telah bertransformasi menjadi sebuah *unicorn* yang berpengaruh di Asia Tenggara. Tokopedia termasuk dalam *marketplace* yang paling banyak dikunjungi oleh masyarakat Indonesia (Afandi, 2020).

Pertumbuhan teknologi informasi terutama pada pengguna media sosial yang banyak mengubah cara masyarakat untuk melakukan berkomunikasi antar pengguna, serta untuk menyampaikan berupa opini atau persepsi umum, hingga pengalaman yang telah mereka alami. Bahkan terhadap Tokopedia tentunya banyak opini atau persepsi umum yang disampaikan pengguna terkait layanan *marketplace* tersebut didalam media sosial, melihat jumlah pengguna internet dalam belanja *online* yang termasuk dalam penggunaan internet paling banyak diakses.

Salah satu media sosial yang sering digunakan dalam menyampaikan sebuah opini atau persepsi umum hingga berkomunikasi yaitu *Twitter*. *Twitter* adalah salah satu media sosial yang paling populer bertindak sebagai platform komunikasi di masyarakat, sehingga *twitter* menjadi salah satu situs yang menyediakan kumpulan data opini dari masyarakat di seluruh dunia. (Fikri et al., 2020). Media sosial *Twitter* menjadi salah satu tempat yang sering digunakan oleh pelanggan untuk menyampaikan keluhan serta pendapat terkait layanan dari suatu

perusahaan (Pratiwi et al., 2021). Untuk mengetahui pandangan atau persepsi umum tersebut, maka dibutuhkan suatu metode analisa persepsi umum atau tanggapan yang diberikan pengguna terhadap layanan dari Tokopedia. Hal ini dapat dijadikan sebagai acuan untuk dilakukan analisis sentimen masyarakat terhadap layanan Tokopedia.

Berdasarkan uraian tersebut, maka akan dilakukan analisis sentimen terhadap Tokopedia. Analisis sentimen merupakan cara untuk melakukan proses klasifikasi terhadap dokumen-dokumen teks yang dilandasi dari opini dan pandangan dari masyarakat tertentu bisa dilakukan dengan proses memahami, mengolah hingga mengekstrak data bersifat dokumen teks secara otomatis. (Hendriani & Sianturi, 2021). Analisis sentimen akan melakukan pengelompokan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut (Aliyah et al., 2020).

Berdasarkan hasil penelusuran terdapat beberapa penelitian menggunakan algoritma klasifikasi yang berbeda dalam penelitiannya dengan tujuan membandingkan kinerja antar algoritma. Banyaknya metode algoritma untuk analisis sentimen data membuat beragam penelitian yang menggunakan dua atau lebih untuk membandingkan seberapa akurat algoritma yang dipilih dalam mengklasifikasi data yang ada. Salah satunya adalah penelitian dari (Marutho, 2019). Dengan judul “Perbandingan Metode *Naïve Bayes*, *Knn*, *Decision Tree* Pada Laporan *Water Level* Jakarta” yaitu membandingkan ketiga model algoritma tersebut untuk mengetahui hasil tingkat akurasi yang lebih tinggi dengan

menggunakan model pengujian *cross validation* menggunakan *k-fold* 10 yang dapat disimpulkan untuk penelitian dataset ketinggian air di Jakarta, yang terbaik menggunakan model algoritma *Decision Tree*. (Marutho, 2019). Kemudian penelitian yang dilakukan oleh (Budiyantara et al., 2020) dengan judul “Komparasi Algoritma *Decision Tree*, *Naive Bayes*, Dan *K-Nearest Neighbor* Untuk Memprediksi Mahasiswa Lulus Tepat Waktu”. Bertujuan untuk menentukan 4ulfil akademis yang berpengaruh terhadap masa studi dan membangun model prediksi terbaik dengan teknik *Data Mining* menggunakan klasifikasi algoritma *Decision Tree*, *Naïve Bayes* dan K-NN. Dari hasil proses pengujian dengan *tools RapidMiner* Menggunakan tiga metode yang telah dilakukan. *Decision Tree* (C4.5) memperoleh hasil akurasi tertinggi sebesar 98.04% pada pengujian ke 3. Metode *Naïve Bayes* memperoleh hasil akurasi tertinggi sebesar 96.00% pada pengujian ke 4, dan Metode *K-Nearest Neighbor* (K-NN) memperoleh hasil akurasi tertinggi sebesar 90.00% pada pengujian ke 2. (Budiyantara et al., 2020).

Berdasarkan latar belakang yang telah dipaparkan penelitian ini akan mengambil topik/judul penelitian “**KOMPARASI TINGKAT AKURASI ALGORITMA DECISION TREE DAN NAÏVE BAYES DALAM ANALISIS SENTIMEN TERHADAP TOKOPEDIA**”. Penelitian ini juga bertujuan untuk melakukan komparasi tingkat akurasi tertinggi dari algoritma *Decision Tree* dan algoritma *Naïve Bayes* dalam melakukan klasifikasi data.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja algoritma *Decision Tree* dan algoritma *Naïve Bayes* dalam klasifikasi data sentiment dari media sosial *Twitter*?
2. Bagaimana hasil komparasi algoritma *Decision Tree* dan algoritma *Naïve Bayes* dalam mengklasifikasikan data sentimen tersebut?
3. Bagaimana opini pengguna terhadap layanan Tokopedia berdasarkan *tweet* pada media sosial *Twitter*?

1.3 Batasan Masalah

Batasan masalah yang ditentukan dalam penelitian ini adalah sebagai berikut:

1. Algoritma yang digunakan *Decision Tree* dan *Naïve Bayes*.
2. Sumber data yang di uji dalam penelitian ini diambil dari beberapa *Tweet* yang ada di Twitter dalam kurun waktu 30 September 2022 sampai 01 Desember 2022.
3. Hanya mengambil data *tweet* berdasarkan kata kunci pencarian “Tokopedia” dan *tweet* terhadap akun @Tokopedia.
4. *Software* pembantu yang digunakan untuk *Collecting Data* / Pengumpulan Data pada penelitian ini menggunakan *RapidMiner*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang diperoleh maka tujuan dari diadakan penelitian ini yaitu:

1. Menganalisis kinerja algoritma *Decision Tree* dengan algoritma *Naïve Bayes*.
2. Melakukan Komparasian terhadap algoritma *Decision Tree* dengan *Naïve Bayes*.
3. Mencari tahu opini pengguna terhadap layanan Tokopedia berdasarkan *tweet* pada media sosial *Twitter*.

1.5 Manfaat Penelitian

Manfaat dari penelitian tugas akhir ini adalah sebagai berikut:

1. Manfaat akademik:
 - a. Menambah wawasan mengenai proses menentukan algoritma yang cocok dan baik untuk melakukan data analisis sentimen.
 - b. Dapat menjadi bahan referensi untuk penelitian lebih lanjut atau penelitian dengan topik yang sama.
 - c. Memberikan informasi untuk memudahkan penelitian sejenis dalam memilih algoritma klasifikasi terbaik.
2. Manfaat umum:
 - a. Memberikan informasi tentang persepsi umum dari pengguna tentang layanan *marketplace* Tokopedia.

- b. Dapat digunakan sebagai pertimbangan bagi masyarakat yang akan menggunakan layanan *marketplace* dalam memilih *marketplace* yang terbaik.
- c. Memberikan informasi yang dapat digunakan pihak perusahaan untuk meningkatkan kualitas pelayanan perusahaan.

1.6 Metodologi Penelitian

Metodologi penelitian ini berisikan tentang tahapan-tahapan penelitian Pengolahan data, diantaranya ;

1. *Collecting Data* / Pengumpulan Data

Collecting data atau disebut juga pengumpulan data ini merupakan proses yang dilakukan dengan melakukan proses *crawling* data dari media sosial Twitter dengan menggunakan *software Rapidminer Studio*.

2. *Text Preprocessing*

Tahap ini merupakan tahapan awal untuk mengubah struktur isi dari data agar menjadi format yang sesuai dan diproses oleh algoritma yang diantaranya *Cleansing, Case Folding, Tokenization, Stopword Removal* dan *Stemming*.

3. Pembobotan Kata

Pada pembobotan kata dilakukan proses pemberian nilai atau bobot terhadap setiap kata yang ada pada setiap *tweet* yang telah melewati tahap *preprocessing*.

4. Klasifikasi algoritma *Decision Tree* dan *Naïve Bayes*.

Pada proses ini dibutuhkan bobot setiap kata yang ada pada data training. Nilai bobot setiap kata ini didapat dari proses pembobotan TF-IDF yang akan

digunakan dalam pembuatan model dari masing masing algoritma yang digunakan.

5. Evaluasi

Pengujian kinerja dari algoritma *Decision Tree* dan *Naïve Bayes* dilakukan dengan metode *confusion matrix* yang menghasilkan empat nilai yaitu *True Positive*, *True Negative*, *False Positive* dan *False Negative* selanjutnya dilakukan perhitungan mulai dari *accuracy*, *precision* dan *recall*. Kemudian, menghasilkan informasi yang dapat digunakan untuk penarikan kesimpulan penelitian.

1.7 Sistematika Pembahasan

Sistematika Pembahasan yang digunakan dalam laporan tugas akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menguraikan tentang garis besar keseluruhan penelitian. Terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini memuat dasar-dasar teoritis maupun penjelasan umum yang berhubungan dengan penelitian, diantaranya mengenai *Komparasi Tingkat Akurasi Algoritma Decision Tree* dan

Algoritma *Naïve Bayes* Dalam Analisis Sentimen Terhadap Layanan Tokopedia

BAB III METODOLOGI

Bab ini berisi tentang metodologi yang digunakan dalam pembahasan serta langkah-langkah penyelesaian masalah selama melakukan penelitian. Langkah-langkah tersebut memuat konsep dari metode yang digunakan, analisis kebutuhan, serta uraian-uraian lainnya yang berkaitan dengan penelitian yang akan dibahas.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi uraian mengenai hasil dan proses pencapaian dalam menyelesaikan penelitian ini. Serta pembahasan yang meliputi penjelasan mengenai data yang akan di analisis.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang merupakan rangkuman dari pembahasan masalah pada penelitian dan saran yang perlu diperhatikan berdasarkan keterbatasan-keterbatasan yang ada dalam penelitian serta untuk bahan evaluasi dasar pengembangan sistem atau penelitian selanjutnya.