

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi berdampak pada perubahan di berbagai bidang, seperti sosial, ekonomi, politik, dan budaya, serta berdampak pada perubahan gaya hidup, termasuk pola konsumsi serta cara berjualan dan berbelanja masyarakat. Di era ini, masyarakat memanfaatkan teknologi informasi dan komunikasi untuk membeli dan/atau menjual barang dan/atau jasa melalui internet. Fenomena ini dikenal dengan perdagangan elektronik atau *E-Commerce*. Fenomena *E-Commerce* menyediakan pilihan cara berbelanja bagi masyarakat dengan tidak perlu datang langsung ke toko.

E-Commerce merupakan suatu konsep baru yang biasa digambarkan sebagai proses jual beli barang atau jasa pada *World Wide Web* Internet atau proses jual beli atau pertukaran produk, jasa, dan informasi melalui jaringan informasi termasuk internet. *E-Commerce* merupakan kegiatan bisnis yang dijalankan secara elektronik melalui suatu jaringan internet atau kegiatan jual beli barang atau jasa melalui jalur komunikasi digital. (Nugroho, 2006). Indonesia merupakan pasar *E-Commerce* terbesar di Asia Tenggara. Menurut data Wearesocial dan Hootsuite, sekitar 90% pengguna internet di Indonesia pernah berbelanja *online*. Pada tahun 2019, nilai kapitalisasi pasar *E-Commerce* di Indonesia mencapai *USD* 21 miliar atau sekitar Rp 294 triliun. Berdasarkan laporan McKinsey, industri *E-Commerce* di Indonesia diprediksi

akan mencapai nilai *USD* 40 miliar pada tahun 2022.

Banyaknya *E-Commerce* yang ada di Indonesia maka pasti semakin banyak kejahatan *online* yang terjadi, sehingga penilaian masyarakat terhadap suatu *E-Commerce* dapat dijadikan analisa terhadap pasar *online E-Commerce* dapat membantu masyarakat lain supaya lebih berhati – hati dalam melakukan transaksi *online*. Penulis merasa perlu untuk melakukan penelitian ini yaitu dengan membuat sebuah sistem penganalisa opini atau biasa disebut analisis sentimen masyarakat sehingga bisa mengetahui dan membantu memberikan informasi mengenai analisa sentimen *E-Commerce* masyarakat. Adapun *E-Commerce* yang menjadi objek penelitian ini adalah Tokopedia dan Shopee. Dilansir dari artikel yang ditulis oleh media *online* iNews.id , kedua *e-commerce* tersebut termasuk tiga situs *E-Commerce* terbaik di Indonesia yang paling sering dikunjungi konsumen dengan data kunjungan peringkat pertama yaitu Tokopedia sebesar 1,2 miliar kali, Shopee sebesar 837,1 juta kali.

Review masyarakat saat ini dapat diperoleh di mana saja salah satunya di aplikasi yaitu *Google Play Store* . Berdasarkan survei yang dilakukan pada sebuah situs 1 Februari 2018 menjelaskan bahwa 69% masyarakat Indonesia mengakses internet dengan perangkat mobile. Bersama negara-negara berkembang lain, seperti Brasil, India, Turki, dan Meksiko, pengguna Android di Indonesia menjadi penyumbang pemasukan terbesar di Android Play Store. Tentu saja, informasi yang terkandung dalam *review* ini sangat berharga sebagai alat penentu kebijakan dan ini bisa dilakukan dengan *Text Mining*. *Teks Mining* adalah salah satu Teknik yang dapat digunakan untuk melakukan

klasifikasi dokumen dimana *Text Mining* merupakan variasi dari *Data Mining* yang berusaha menemukan pola menarik dari sekumpulan data tekstual yang berjumlah besar (Kurniawan, Effendi, & Sitompul, 2012). Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*). Analisis Sentimen atau *Opinion Mining* adalah studi komputasional dari opini – opini orang, sentimen dan emosi melalui entitas atau atribut yang dimiliki yang diekspresikan dalam bentuk teks (Aditya, Hani'ah, Fitrawan, Arifin, & Purwitasari, 2016). Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral (Pang & Lee, 2008).

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes* . Metode pengklasifikasian dg menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes , yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes . Ciri utama dr *Naïve Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang

terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dr " master " tabel keputusan.

Naive Bayes Classifier bekerja sangat baik dibanding dengan model *classifier* lainnya. Hal ini dibuktikan oleh Xhemali , Hinde Stone dalam jurnalnya "*Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*" mengatakan bahwa "*Naive Bayes Classifier* memiliki tingkat akurasi yg lebih baik dibanding model *classifier* lainnya". Keuntungan penggunaan adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yg kecil unt menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian. Karena yg diasumsikan sebagai *variable independent*, maka hanya *varians* dr suatu *variable* dalam sebuah kelas yg dibutuhkan unt menentukan klasifikasi, bukan keseluruhan dr matriks kovarians.

Seleksi Fitur *Information Gain* adalah *Information Gain* dalam *machine learning* digunakan untuk mengukur seberapa relevan / berpengaruh sebuah *feature* terhadap hasil pengukuran. Penggunaan teknik ini dapat mereduksi dimensi *feature* dengan cara mengukur reduksi *Entropy* sebelum dan sesudah pemisahan. *Information Gain (IG)* dikenal juga dengan sebutan *Mutual Information (MI)* dalam kasus untuk mengetahui *dependency* antara dua *variable* (x,y). Keuntungan menggunakan metode ini yaitu penyederhanaan model, membuatnya lebih mudah untuk diinterpretasi oleh *researcher/user*, mempercepat proses *training*, Menghindari *curse of dimensionality* (beragam fenomena yang timbul saat menganalisa atau

mengolah data dengan dimensi tinggi. Dalam *Machine Learning* ini terjadi ketika *high-dimensional feature space* memiliki *n samples* yang terbatas), serta *Enhance generalization* dengan cara menurunkan *overfitting*.

Berdasarkan penelitian (Chandani, Wahono, & Purwanto, 2015) yang meneliti perbandingan akurasi antara *information gain*, *chi square*, *forward selection*, *backward selection* didapatkan *information gain* merupakan yang terbaik. Penelitian mengenai klasifikasi sentiment juga telah dilakukan oleh Syakuro (2017). Pada jurnalnya, Syakuro melakukan klasifikasi sentimen terhadap *review E-Commerce* dengan menggunakan berbagai teknik pembelajaran mesin. Teknik pembelajaran mesin yang digunakan yaitu *Naïve Bayes*. Pada penelitian itu juga digunakan beberapa pendekatan untuk melakukan ekstraksi fitur, yaitu *Information Gain*. Hasil dari eksperimen yang dilakukan dipenelitian ini menemukan bahwa *Naïve Bayes* menjadi metode terbaik ketika dikombinasikan dengan Seleksi Fitur *Information Gain*. dengan akurasi 88.8%.

Penelitian lainnya tentang Klasifikasi Prediksi Model juga telah dilakukan oleh Ilham Susilo Bakti, Ivandari (2019). Pada jurnalnya, mereka melakukan klasifikasi model terhadap penyakit diabetes dengan metode *Naïve Bayes* dan Seleksi Fitur *Information Gain* serta Algoritma *Adaptive Boosting*. Hasil pengujian yang telah dilakukan menunjukkan bahwa akurasi *Naïve Bayes* dengan menggunakan metode boosting *AdaBoost* dan *Information Gain*, meningkatkan akurasi mencapai 79.10%. Peningkatan sebesar 4.2%. Maka

dari itu, terbukti bahwa *Naïve Bayes* dengan menambahkan *Information Gain* dan *AdaBoost*, meningkatkan akurasi pada prediksi penyakit diabetes.

Analisis sentimen juga telah dilakukan oleh Arif Bijaksana Putra Negara, Hafiz Muhandi , Indira Melinda Putri (2020). Pada jurnalnya, mereka melakukan klasifikasi sentiment terhadap ulasan maskapai penerbangan dengan metode *Naïve Bayes* dan Seleksi Fitur *Information Gain*. Hasil pengujian yang telah dilakukan menunjukkan bahwa nilai rata-rata akurasi, precision, recall setelah penambahan *Information Gain* menunjukkan hasil yang lebih baik sebesar 0,865 jika dibandingkan sebelum penambahan *information gain* yakni sebesar 0,81. Sehingga pada penelitian ini penulis akan melakukan klasifikasi sentiment masyarakat pengguna *E-Commerce* dengan metode *Naïve Bayes Classifier* dengan Seleksi Fitur *Information Gain*.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang akan diteliti adalah :

1. Bagaimana hasil pengklasifikasian sentimen terhadap review *E-Commerce* pada *Google Play Store* menggunakan metode *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain* ?
2. Bagaimana tingkat akurasi metode *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain* dalam pengklasifikasian sentimen terhadap review *E-Commerce* pada *Google Play Store* ?

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah ingin menjawab permasalahan dalam rumusan masalah. Adapun tujuan yang ingin dicapai oleh peneliti dalam penelitian ini adalah:

1. Hasil pengklasifikasian dari metode *Naïve Bayes Classifier* dengan Seleksi Fitur *Information Gain* dalam analisis sentimen terhadap review *E-Commerce* pada *Google Play Store*
2. Tingkat akurasi metode *Naive Bayes Classifier* dengan Seleksi Fitur *Information Gain* dalam analisis sentimen terhadap review *E-Commerce* pada *Google Play Store*

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi positif baik secara langsung maupun tidak langsung, antara lain:

1. Manfaat Teoritis

Membantu perkembangan ilmu pengetahuan mengenai metode *text mining* sehingga dapat digunakan sebagai bahan bacaan dan referensi bagi pembaca dalam melakukan penelitian yang berhubungan dengan *text mining* terutama pada metode Analisis Sentimen Terhadap *Review E-Commerce* Pada *Google Play Store* Menggunakan Metode *Naïve Bayes* dengan Seleksi Fitur *Information Gain*.

2. Manfaat Praktis

- a. Bagi Pembaca, pembaca dapat menambah dan memperkaya pengetahuan tentang implementasi statistika dalam dunia perdagangan online serta sebagai sarana informasi bagi pembaca dan pelengkap referensi bagi pihak yang memerlukan.
- b. Bagi Peneliti, peneliti mampu menerapkan metode yang sesuai dalam materi yang telah dipelajari serta peneliti mempunyai pengetahuan dan wawasan mengenai *text mining* menggunakan Analisis Sentimen Terhadap *Review E-Commerce* Pada *Google Play Store* Menggunakan Metode *Naïve Bayes* dengan Seleksi Fitur *Information Gain* .
- c. Bagi Universitas Muhammadiyah Semarang, dapat digunakan sebagai bahan referensi dalam pengaplikasian ilmu statistika khususnya untuk pengolahan data teks dengan menggunakan metode *text mining* kedalam dunia perdagangan.
- d. Bagi Penyedia Layanan *E-Commerce*, Manfaat pengklasifikasian data *tweet* Tokopedia dan Shopee dapat memudahkan pihak perusahaan dalam mengetahui persepsi pengguna jasa dalam bentuk opini positif dan negatif, sehingga dapat dijadikan sebagai acuan dalam upaya menjaga kualitas dan memperbaiki kekurangan serta evaluasi kearah yang lebih baik.

1.5 Batasan Penelitian

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang digunakan adalah data *review* dari aplikasi *Google Play Store* berbahasa Indonesia yang diambil menggunakan software Web Harvy periode 18 Maret- 7 April 2021.
2. *E-Commerce* yang akan dianalisis yaitu Tokopedia dan Shopee
3. Metode yang digunakan yaitu *Naïve Bayes Classifier* (NBC) dengan Seleksi Fitur *Information Gain* (IG).
4. Evaluasi klasifikasi menggunakan tingkat akurasi .

