

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan internet di Indonesia juga turut berkembang seiring dengan perkembangan internet di dunia. Pengguna internet di Indonesia terus bertambah dari tahun ke tahun. Berdasarkan survei dari Asosiasi Penyelenggara Jasa Internet Indonesia atau APJII hingga kuartal II tahun 2020, jumlah pengguna internet di Indonesia terdapat 196,7 juta orang atau 73,7% dari total populasi Indonesia sebanyak 266,9 juta berdasarkan *data* BPS. Jumlah pengguna internet di Indonesia meningkat sekitar 8,9% atau sekitar 25,5 juta pengguna dari 171 juta di tahun 2019 [1].

Internet yang berkembang dengan pesat memudahkan masyarakat dalam mendapatkan informasi serta kebutuhan – kebutuhan lainnya [2]. Masyarakat dapat dengan mudah mendapatkan produk maupun jasa secara *online* melalui aplikasi *e-commerce*. Berdasarkan *data* dari *GlobalWebIndex*, Indonesia merupakan negara yang memiliki tingkat adopsi *e-commerce* tertinggi di dunia pada 2019. Industri *e-commerce* mampu membuka peluang bisnis baru dan menghasilkan dampak bagi industri pendukung seperti logistik, infrastruktur TI, dan *operator e-commerce* [3].



Gambar 1. 1. Data Pengguna E-Commerce di Indonesia

(Sumber : <https://datareportal.com/reports/digital-2020-indonesia>)

Platform manajemen media sosial *HootSuite* dan agensi *marketing* sosial *We Are Social* dengan judul “*Global Digital Reports 2020*” menunjukkan pengguna *E-Commerce* di Indonesia. Pada *data* tersebut menunjukkan bahwa penduduk Indonesia yang melakukan belanja *online* berjumlah 168,3 juta penduduk. Pengguna *E-Commerce* menghabiskan uang senilai US\$18,76 miliar atau setara dengan Rp 262,810 triliun. Rata-rata pengguna *E-Commerce* di Indonesia melakukan belanja *online* senilai \$11 atau setara dengan Rp 1,555,011.

Toko Online	Pengunjung Web Bulanan	Ranking AppStore	Ranking PlayStore	Twitter	Instagram	Facebook	Jumlah Karyawan
1 Shopee	96,532,300	#1	#1	486,100	5,965,200	18,870,500	7,000
2 Tokopedia	84,997,100	#2	#4	611,900	2,120,700	6,385,100	4,300
3 Bukalapak	31,409,200	#4	#5	193,800	1,123,600	2,501,900	2,300
4 Lazada	22,674,700	#3	#3	398,300	2,327,200	30,072,000	3,500
5 Blibli	18,695,000	#5	#6	504,500	1,334,500	8,568,100	1,900
6 JD ID	4,785,800	#8	#7	32,100	492,000	781,300	1,100
7 Orami	3,071,900	#27	n/a	6,000	n/a	352,800	183
8 Bhinneka	2,803,800	#20	#20	6,900	41,800	1,051,400	605
9 Sociolla	1,986,700	#7	#2	3,200	907,500	10,900	392
10 Zalora	1,828,500	#6	#8	900	619,100	7,843,300	562

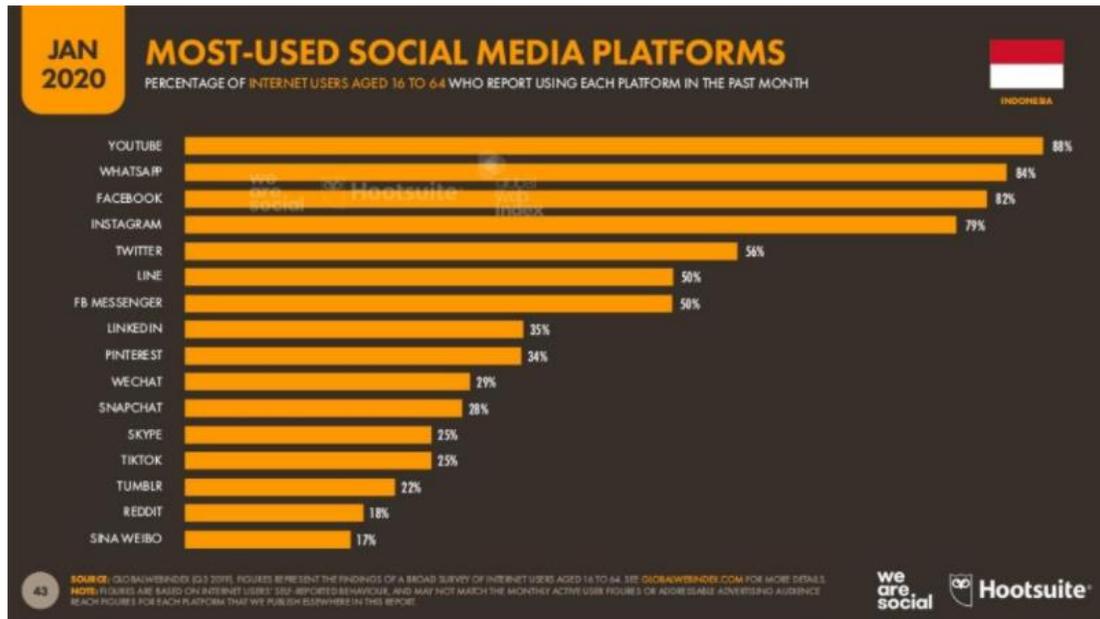
Gambar 1. 2. E-Commerce Yang Sering Dikunjungi Masyarakat Indonesia

(Sumber : <https://iprice.co.id/insights/mapofecommerce/>)

Berdasarkan *data* riset dari *iPrice* tentang “Peta *E-Commerce* Indonesia” pada periode Q3 2020, *iPrice* mengumpulkan *data* berdasarkan rata-rata pengunjung *website* setiap kuartal, peringkat aplikasi, pengikut pada setiap media sosial, dan jumlah karyawan di *E-Commerce*. *Data* yang akan digunakan pada penelitian ini merupakan 5 *E-Commerce* yang sering dikunjungi oleh masyarakat Indonesia pada tahun 2020.

Internet yang sudah berkembang dengan pesat di Indonesia menyebabkan penggunaan media sosial juga terus meningkat. Berdasarkan riset dari *platform* manajemen media sosial *HootSuite* dan agensi *marketing* sosial *We Are Social* dengan judul “*Global Digital Reports 2020*” menunjukkan pengguna media sosial di

Indonesia sudah mencapai 160 juta. Riset tersebut juga memberitahu *platform* media sosial yang sering digunakan oleh masyarakat Indonesia.



Gambar 1. 3. Platform Media Sosial Terbanyak Digunakan di Indonesia

(Sumber : <https://datareportal.com/reports/digital-2020-indonesia>)

Data Governance adalah suatu organisasi dan implementasi dari kebijakan, prosedur, struktur, peran, dan tanggung jawab yang menguraikan dan menegakkan aturan pihak yang terlibat, hak keputusan, dan akuntabilitas untuk pengelolaan aset informasi yang efektif. *Data Governance* memberikan petunjuk dan panduan tentang bagaimana cara menggunakan dan merawat aset *data*. *Data Governance* juga merupakan pengawas untuk memastikan pengguna dan pengelola *data* menggunakan material yang berkualitas dan mengelola dengan cara yang benar [4].



Gambar 1. 4. Komentar di Media Sosial Instagram E-Commerce

Instagram merupakan salah satu media sosial yang penggunaannya begitu pesat dan memiliki *UGC* (*User Generated Content*). *UGC* merupakan *data* pada media sosial dalam berbagai bentuk seperti *video* dan foto yang diunggah [5]. Pada *UGC* yang diunggah terdapat komentar yang dapat diberikan oleh pengguna *Instagram*. Komentar yang diberikan dapat berupa komentar yang positif ataupun komentar negatif. Komentar negatif dapat diakibatkan oleh ketidakpuasan pengguna terhadap berbagai hal dan mampu mengakibatkan rusaknya citra dari suatu perusahaan apabila tidak ditindaklanjuti.

Sentimen analisis digunakan untuk memperoleh tanggapan masyarakat terkait dengan *E-Commerce* [6]. Dengan memperoleh tanggapan masyarakat, maka penelitian ini dapat digunakan untuk mengevaluasi atau *me-review E Commerce* apakah sudah menjadi media yang baik dalam melakukan jual-beli berdasarkan tanggapan masyarakat Indonesia. Evaluasi atau *review* dimulai dari awal Maret 2020 karena pada

Maret 2020 wabah *COVID-19* mulai menyerang sehingga setiap orang sulit dalam melakukan jual beli [7].

Sentimen analisis ini akan menggunakan algoritma *SVM* dan *Naïve Bayes*. Algoritma tersebut digunakan untuk mengklasifikasikan *data* [6]. *Support Vector Machine* merupakan teknik untuk melakukan prediksi yang baik dalam melakukan klasifikasi dan regresi [8]. *Naïve Bayes* merupakan teknik untuk melakukan prediksi probabilitas keanggotaan suatu kelas pada teorema *bayes* dan memiliki kemampuan dalam klasifikasi [9]. *SVM* dan *Naïve Bayes* merupakan metode klasifikasi yang sudah teruji dan mendapatkan nilai akurasi yang tinggi, maka *SVM* dan *Naïve Bayes* dinilai cocok untuk dibandingkan dalam mencari nilai akurasi [10].

Beberapa penelitian telah dilakukan yang menjadi dasar dari terjadinya penelitian ini yaitu penelitian dengan judul Algoritma Klasifikasi *Naïve Bayes* untuk Analisa Sentimen Aplikasi *Shopee* oleh Siti Masripah dan Lila Dini, penelitian dengan judul Apakah Youtuber Indonesia Kena *Bully* Netizen? oleh Joviano Siahaan; Wella; dan Ririn I. Desanti, penelitian dengan judul *Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine* oleh Anang Anggono Lufti; Adhistya Erna Permanasari; dan Silmi Fauziati, dan penelitian dengan judul *Sentiment Analysis on E-commerce Application by using Opinion Mining* oleh Nitu Kumari dan Dr. Shailendra Narayan Singth. Penelitian ini mengadopsi beberapa metode dan cara yang dilakukan pada penelitian terdahulu.

1.2. Rumusan Masalah

Berdasarkan latar belakang pada paragraf sebelumnya maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana hasil analisis sentimen dari pengguna media sosial *Instagram* terhadap *E-Commerce* yaitu *Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli* di Indonesia menggunakan algoritma *Support Vector Machine* dan *Naïve Bayes*?
2. Bagaimana perbandingan kinerja dari *Support Vector Machine* dan *Naïve Bayes* berdasarkan *accuracy* pada analisis sentiment ?
3. Bagaimana urutan tertinggi pada *E-Commerce* berdasarkan hasil klasifikasi ?

1.3. Batasan Masalah

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, diperlukan batasan masalah agar penelitian tidak keluar dari inti. Maka dari itu batasan masalah pada penelitian ini adalah :

1. Penelitian ini terbatas hanya untuk beberapa kelas yaitu klasifikasi sentimen positif dan negatif.
2. Media sosial yang digunakan untuk mendapatkan data adalah media sosial *Instagram*.
3. *Data* sentimen yang akan diprediksi didapatkan dari komentar *Instagram* pada 5 akun *E-Commerce* (*Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli*) berdasarkan jumlah pengunjung tertinggi dari *website iprice insights*.
4. *Data* yang digunakan akan diambil pada bulan Maret 2020 hingga Maret 2021.

5. *Data* yang digunakan sebanyak 10.000 komentar pada *Instagram E Commerce* yang diambil menggunakan *Instagram Scraper* dengan aplikasi *PyCharm*.
6. *Post* yang akan digunakan tidak akan termasuk *post* yang mengandung unsur *giveaway*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, diperlukan tujuan agar penelitian ini semakin jelas. Maka, tujuan dari dilakukannya penelitian ini adalah :

1. Mengetahui hasil analisis sentimen dari pengguna media sosial *Instagram* terhadap *E-Commerce* yaitu *Shopee, Tokopedia, Bukalapak, Lazada, dan Blibli* di Indonesia menggunakan algoritma *Support Vector Machine* dan *Naïve Bayes*.
2. Mengetahui perbandingan *accuracy* dan *AUC* dari algoritma *Support Vector Machine* dan *Naïve Bayes* dalam melakukan klasifikasi dan prediksi opini pengguna media sosial *Instagram* terhadap *E-Commerce* di Indonesia.
3. Mengetahui urutan tertinggi dari *E-Commerce* pada hasil klasifikasi.

1.5. Manfaat Penelitian

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, diperlukan manfaat agar penelitian ini dapat digunakan. Maka, manfaat dari dilakukannya penelitian ini adalah:

1. Menambahkan pengetahuan terkait dengan penggunaan algoritma *Support*

Vector Machine dan *Naïve Bayes* untuk melakukan klasifikasi dan prediksi dari opini pengguna media sosial *Instagram* terhadap *E-Commerce* di Indonesia.

2. Menambah pengetahuan terkait dalam melakukan penambangan *data* dan pengelolaan *data* untuk klasifikasi dan prediksi opini.