

1. Pendahuluan Latar Belakang

Perkembangan teknologi dilakukan dengan tujuan meningkatkan taraf hidup manusia. Salah satu perkembangan yang berpengaruh besar terhadap kehidupan manusia yaitu internet. Hingga tahun 2018, pengguna internet di Indonesia telah mencapai sekitar 143 juta [1]. Keberadaan internet membuat pemasaran dapat dilakukan tanpa harus interaksi langsung antara penjual dan pembeli. Pemasaran menggunakan internet disebut dengan *e-commerce*. Terdapat banyak pengunjung *e-commerce* di Indonesia, bahkan Tokopedia yang merupakan salah satu contoh *e-commerce* memiliki ratusan juta pengunjung pada 2018 [2]. Perusahaan *e-commerce* perlu memiliki pemahaman terhadap sentimen pelanggan untuk memberikan pelayanan yang lebih baik. Dengan adanya media sosial yang merupakan pemanfaatan dari internet, sebagian besar pengguna media sosial dapat menyalurkan sentimen atau opini mereka di media sosial. Dapat disimpulkan bahwa media sosial merupakan pilihan tepat dalam mendapatkan data sentimen pelanggan.

Twitter merupakan salah satu media sosial dengan pengguna yang sangat banyak di Indonesia [3]. Banyaknya pengguna Twitter di Indonesia, tentunya opini pelanggan dapat ditemukan dalam jumlah yang besar. Dari opini tersebut, diharapkan *e-commerce* di Indonesia dapat berkembang ke arah yang lebih baik dengan terus mengkaji kepuasan pelanggan melalui sentimen pelanggan yang tersurat di media sosial. Opini dapat bersifat positif, negatif, ataupun netral. Dengan banyaknya jumlah opini yang dapat ditemukan dan terus bertambah, diperlukan sistem yang dapat mengklasifikasikan opini-opini tersebut.

Terdapat beberapa metode yang dapat digunakan untuk klasifikasi data sentimen, salah satunya Naïve Bayes. Naïve Bayes merupakan metode pembelajaran mesin yang memiliki model dalam bentuk probabilitas atau peluang [4]. Ketika ada tweet baru yang ingin diklasifikasikan, maka nilai probabilitas dari setiap fitur untuk masing-masing kelas dikalikan. Nilai probabilitas yang paling besar diantara kelas positif, negatif, dan netral, dijadikan kelas tweet baru tersebut. Selain itu, klasifikasi data menggunakan Naïve Bayes *Classifier*, karena metode tersebut memiliki akurasi yang tinggi dalam berbagai situasi di dunia nyata [5].

Penelitian mengenai analisis sentimen pada media sosial Instagram menggunakan metode klasifikasi Naïve Bayes, pernah dilakukan oleh Antonius Rachmat Chrismanto [6], menggunakan 25.000 data. Penelitian tersebut menghasilkan akurasi sebesar 77,25%. Berdasarkan penelitian tersebut, penelitian ini menggunakan data sebanyak 50.000 *tweet*.

Topik dan Batasan

Berdasarkan latar belakang, masalah yang dapat dirumuskan pada penelitian ini adalah bagaimana performansi metode naïve bayes dengan pembobotan TF-IDF pada tweet sentimen pengguna terhadap suatu *e-commerce*. Studi kasus yang dilakukan terbatas pada *e-commerce* Indonesia karena mesin yang dibuat digunakan untuk memprediksi teks berbahasa Indonesia. Data yang digunakan sebanyak 50.000 tweet dengan isi tweet yang merujuk pada *e-commerce* di Indonesia. Setelah data tersebut melalui klasifikasi Naïve Bayes, dilakukan pengukuran performansi menggunakan *Precision*, *Recall*, dan Akurasi. Penulis menggunakan Google Colab untuk menjalankan program dikarenakan banyaknya data sehingga membutuhkan sumber daya yang cukup besar, namun untuk tahap hingga preprocessing, penulis menggunakan PyCharm.

Tujuan

Melalui Twitter, pengguna dapat menyampaikan opini mereka mengenai layanan toko *online* yang mereka gunakan. Opini tersebut dijadikan data, untuk selanjutnya diklasifikasikan dan ditemukan pola yang terkandung di dalam data tersebut, Naïve bayes merupakan salah satu metode yang dapat digunakan untuk mendapatkan pola. Tujuan dari penelitian ini adalah untuk mengetahui apakah metode Naïve Bayes efektif terhadap data sebanyak 50.000 data yang berupa teks.

Organisasi Tulisan

Organisasi penulisan ini dimulai dengan latar belakang, topik, batasan dan tujuan penelitian yang termasuk dalam pendahuluan. Selanjutnya ada studi terkait yang berisi tentang penelitian-penelitian yang berhubungan dengan analisis sentimen dan metode yang digunakan. Ketiga, sistem yang dibangun, gambaran umum sistem beserta penjelasan setiap tahapnya. Keempat, dituliskan hasil implementasi dan analisis berdasarkan penelitian yang dilakukan. Kelima adalah poin kesimpulan serta saran.