

# 1. Pendahuluan

## 1.1 Latar Belakang

Pertumbuhan volume transaksi pada toko online semakin lama semakin meningkat. Saat ini, semua orang yang terhubung dengan internet sudah dapat membeli barang apa pun yang ia mau, kapan pun dan dimana pun. Pada tahun 2020 diperkirakan penjualan pada toko online secara global yaitu sejumlah USD 4 Triliun, dan setidaknya mengalami peningkatan penjualan 18% setiap tahunnya [1]. Di Indonesia sendiri, setidaknya mengalami peningkatan penjualan 19.8% per tahun dan pada tahun 2020 akan mencapai angka USD 12.38 Miliar [2]. Hal ini mengindikasikan bahwa tren transaksi elektronik semakin diminati oleh masyarakat.

Ulasan suatu produk yang tersebar di berbagai media juga tidak kalah penting untuk dicermati. Penelitian yang dilakukan pada awal 2013 menunjukkan bahwa ulasan produk yang tersebar di sosial media telah mempengaruhi 90% keputusan seseorang terhadap pembelian produk tersebut [3]. Bagi pihak produsen tentu saja hal ini bisa menjadi acuan apakah produk yang mereka pasarkan dapat diterima dengan baik oleh masyarakat dengan cara membaca dan menilai suatu ulasan dengan nilai positif atau negatif. Akan tetapi, untuk mengkategorikan ulasan produk tersebut ke dalam kategori komentar positif dan negatif yang jumlahnya ribuan tentu akan memakan waktu yang lama untuk mengetahui opini masyarakat mengenai produk tersebut.

Berdasarkan permasalahan tersebut, salah satu solusi yang dapat ditawarkan adalah menggunakan *sentimen classification*. *Sentimen classification* merupakan salah satu bagian dari klasifikasi teks yang dikerjakan oleh komputer dengan tujuan mengklasifikasikan teks berdasarkan polaritasnya seperti positif atau negatif, suka atau tidak suka [4].

Jika ditinjau dari data yang digunakan, maka peneliti dapati bahwa data ulasan produk dalam bentuk teks mengandung ketidakpastian, contohnya “*the camera is very easy to use*” memiliki sentimen positif, sedangkan teks “*the included lens cap is very loose on the camera*” memiliki sentimen negatif. Secara intuisi manusia hal tersebut mudah untuk dibedakan, namun bagi komputer hal semacam itu yang tidak mudah karena memiliki kesamaan kata “*camera*”, “*the*”, dan “*very*” di dalamnya namun memiliki sentimen yang berbeda. Atas dasar itu peneliti menggunakan model probabilistik untuk menyelesaikan masalah tersebut karena model probabilistik adalah salah satu teknik yang dapat digunakan untuk permasalahan *uncertainty reasoning*, dan metode yang peneliti pilih adalah *Multinomial Naive Bayes*.

*Multinomial Naive Bayes* mengasumsikan fitur pada data bersifat *conditionally independent given the class* sehingga setiap kelas memiliki fitur yang merepresentasikan kelasnya masing-masing [5]. Sehingga metode ini dapat dijadikan *classifier* yang tepat dalam permasalahan klasifikasi sentimen data ulasan produk.

## 1.2 Perumusan Masalah

Rumusan masalah dari tugas akhir ini adalah sebagai berikut:

1. Bagaimana cara membangun *Multinomial Naïve Bayes* untuk mengklasifikasikan data ulasan produk toko *online*?
2. Bagaimana performa *Multinomial Naïve Bayes* dalam mengklasifikasikan sentimen pada data ulasan produk toko *online*?

### 1.3 Tujuan

Tujuan dari tugas akhir ini adalah sebagai berikut:

1. Dapat membangun *classifier* yang mampu mengklasifikasikan sentimen pada ulasan produk secara otomatis.
2. Menghasilkan performa klasifikasi yang baik dari *Multinomial Naïve Bayes*.

### 1.4 Batasan Masalah

Batasan masalah dari tugas akhir ini adalah sebagai berikut:

1. Data ulasan produk berasal dari situs *amazon.com* dan menggunakan bahasa inggris [6].
2. Orientasi sentimen yang diklasifikasikan berupa positif dan negatif.
3. Kalimat ulasan yang tidak memiliki orientasi sentimen tidak akan dimasukkan sebagai data *training* maupun data *testing*, namun akan dihilangkan saat proses data *cleaning*.

### 1.5 Metodologi Penyelesaian Masalah

Metodologi yang digunakan dalam menyelesaikan Tugas Akhir ini adalah sebagai berikut.

1. Identifikasi Masalah  
Pada tahap ini peneliti mencari permasalahan dengan mengunjungi website yang memiliki ulasan-ulasan produk untuk dijadikan studi kasus yang dapat dijadikan bahan untuk membuat tugas akhir.
2. Studi literatur dan pengumpulan data  
Pada tahap ini peneliti melakukan kajian terhadap jurnal dan paper terkait untuk memberikan solusi yang tepat saat pengerjaan tugas akhir. Untuk pengumpulan data, peneliti menggunakan dataset mengenai ulasan produk situs *amazon.com* [6].
3. Perancangan Sistem  
Pada tahap ini peneliti melakukan perancangan sistem yang direpresentasikan dalam bentuk *flowchart*, merancang skema pengujian, membuat desain sistem dengan mendefinisikan terlebih dahulu fungsi dan struktur data.
4. Penulisan Buku Laporan Tahap 1  
Pada tahap ini peneliti menuliskan skenario yang akan diimplementasikan pada sistem. Hal ini dilakukan akan penulis memiliki alur kerja yang jelas sehingga memudahkan peneliti pada proses-proses selanjutnya.
5. Implementasi

Pada tahap ini peneliti melakukan pembuatan sistem yang sudah dirancang pada tahap sebelumnya. Implementasi dilakukan menggunakan bahasa pemrograman *Python*, serta menggunakan *compiler* PyCharm.

#### 6. Pengujian

Pada tahap ini peneliti menguji sistem yang sudah dibuat dengan data yang dimiliki. Peneliti melakukan pengujian menggunakan *K-fold cross-validation* menggunakan *10-fold* yaitu 90% dataset digunakan untuk proses *training* dan 10% sisanya digunakan untuk *testing*. Seluruh hasil pengujian ditampilkan dalam *F1-score*.

#### 7. Analisis

Pada tahap ini peneliti menganalisis hasil yang diperoleh dari proses sebelumnya. Analisis yang dilakukan yaitu menghubungkan hasil yang diterima setelah pembangunan sistem serta pengujiannya dengan tujuan yang akan dicapai dari penelitian ini.

#### 8. Penulisan Buku Laporan Tahap 2

Pada tahap ini peneliti mendokumentasikan hasil yang telah dikerjakan dalam bentuk buku tugas akhir sebagai persyaratan sidang dan untuk pengembangan lebih lanjut oleh pihak lain.

### 1.6 Sistematika Penulisan

Untuk mempermudah pembaca dalam memahami laporan penelitian ini, penulis mengelompokkan laporan menjadi beberapa subbab dengan sistematika penulisan sebagai berikut.

#### 1. Pendahuluan

Bab ini menjelaskan tentang informasi umum yaitu latar belakang penelitian, perumusan masalah, batasan masalah, tujuan penelitian, metodologi penyelesaian masalah, dan sistematika penulisan.

#### 2. Landasan Teori

Bab ini berisikan teori yang diperlukan dalam penelitian berupa definisi dan persamaan matematika. Teori tersebut diambil dari beberapa literatur terkait di antaranya kutipan buku, *conference proceeding*, dan artikel.

#### 3. Perancangan Sistem

Bab ini berisi tentang gambaran umum sistem dan tahapan tiap prosesnya. Selain itu, bab ini juga menjelaskan kebutuhan perangkat lunak dan keras.

#### 4. Pengujian dan Analisis

Dalam bab ini peneliti melakukan analisis performa yang dihasilkan dari tiap skema pengujian untuk membangun model classifier yang terbaik dalam kasus ini.

#### 5. Kesimpulan dan Saran

Bab ini berisi kesimpulan dan saran yang berkaitan dengan hasil analisis peneliti dan pada bab sebelumnya.