

BAB 3

METODE PENELITIAN

Penelitian ini merupakan penelitian *topic modeling* menggunakan algoritma *Latent Dirichlet Allocation* pada data umpan balik pelanggan yang diterima oleh JuraganCOD dengan mengimplementasikan metode *life cycle of text mining*. Metode *life cycle of text mining* merupakan metode yang dapat dipakai dalam proses pemrosesan teks secara keseluruhan yang digunakan juga oleh peneliti (Shaikh et al., 2019). Adapun 8 tahap dalam metode *life cycle of text mining* tersebut meliputi beberapa langkah penting, seperti mendefinisikan masalah (*problem definition*), memilih pendekatan *text data mining* yang tepat (*select text data mining approach*), mengumpulkan data (*data collection*), melakukan standarisasi teks (*text standardization*), pemrosesan data awal (*text pre-processing*), ekstraksi fitur (*feature extraction*), tahap analisis (*analysis*), dan interpretasi pola atau pengetahuan yang ditemukan (*discovery*).

3.1 OBJEK PENELITIAN

Objek penelitian dalam skripsi ini adalah perusahaan JuraganCOD, yang berfokus pada manajemen media dan platform untuk mengatur pengiriman barang melalui berbagai kurir, baik yang menggunakan sistem *cash on delivery* (COD) maupun non-COD (JuraganCOD, 2022). Layanan JuraganCOD mencakup proses pendaftaran yang membutuhkan data diri, data bank, dan foto KTP; pembuatan pesanan pengiriman melalui *dashboard* pilihan jenis penjemputan, yaitu *pick up* tanpa minimal paket atau *drop off* ke ekspedisi; pelacakan perjalanan paket; serta monitoring transaksi pembayaran yang mencakup status sukses, *cashback*, *refund*, dan lainnya. Platform JuraganCOD memudahkan pengguna dalam mengelola pengiriman dan memastikan transparansi dalam setiap transaksi.

3.2 BAHAN DAN ALAT PENELITIAN

Bahan utama yang digunakan dalam penelitian ini adalah data umpan balik atau ulasan sebanyak 2.010 yang diberikan oleh pengguna layanan JuraganCOD.

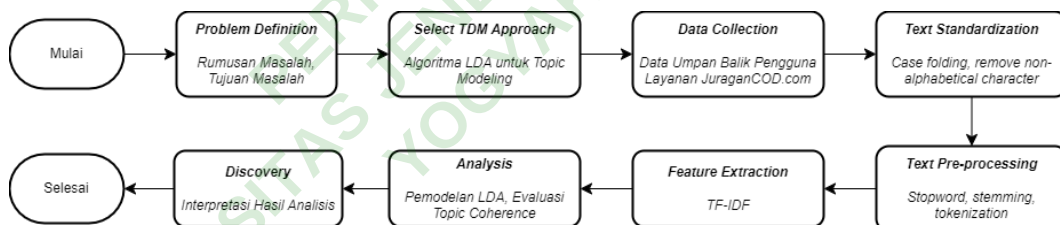
Data ini diperoleh dari tim CRM yang mencatat keluhan pelanggan melalui telepon dan dimasukkan ke dalam database umpan balik. Selain itu, data juga diperoleh dari umpan balik yang diterima melalui *website* JuraganCOD, yaitu JuraganCOD.com, selama tahun 2023.

Perangkat yang diperlukan untuk penelitian ini termasuk komputer dengan spesifikasi yang memadai untuk menjalankan sistem operasi dan perangkat lunak pengembangan yang dibutuhkan, serta memiliki akses internet. Komponen perangkat lunak yang digunakan dalam pengembangan aplikasi mencakup:

1. Sistem Operasi: Windows 11
2. Google Colab
3. Bahasa Pemrograman Python 3.12.0

3.3 JALAN PENELITIAN

Penelitian ini akan mengadopsi metode *life cycle of text mining* yang digunakan oleh peneliti (Shaikh et al., 2019). Metode penelitian terdiri dari beberapa tahap yang dapat dilihat pada **Gambar 3.1** berikut:



Gambar 3. 1 Alur Jalan Penelitian

1. TAHAP PROBLEM DEFINITION

Pada tahap *problem definition*, penulis mengidentifikasi masalah berdasarkan latar belakang penelitian. Setelah itu, dirumuskan suatu masalah yang bertujuan untuk mengidentifikasi topik-topik umum yang sering dibicarakan oleh pelanggan terkait dengan JuraganCOD. Selain itu, ditentukan bagaimana algoritma LDA dapat diterapkan untuk menganalisis topik-topik tersebut. Tujuannya adalah untuk memberikan analisis yang dapat mendukung upaya peningkatan kepuasan dan loyalitas pelanggan terhadap JuraganCOD.

2. TAHAP SELECTING TEXT DATA MINING APPROACH

Pada tahap *selecting TDM approach* atau pendekatan TDM, penulis melakukan studi literatur dan pemutusan penggunaan *topic modeling* menggunakan LDA telah dijelaskan juga pada bagian latar belakang dan tinjauan pustaka dimana penelitian (Churchill & Singh, 2022) menyoroti 7 faktor yang harus dipertimbangkan dalam memilih metode, namun secara keseluruhan, LDA masih dianggap sebagai pilihan yang tepat dan dapat diandalkan untuk pemodelan topik secara umum. Selain itu, terdapat 24 makalah yang diterbitkan antara tahun 2020 dan 2022 yang juga menegaskan bahwa LDA tetap menjadi salah satu teknik terbaik dalam pemodelan topik. Hal ini menunjukkan bahwa metode LDA masih relevan dalam konteks pemodelan topik saat ini (Sharif et al., 2023).

3. TAHAP DATA COLLECTION

Pada tahap *data collection* (pengumpulan data), penulis mengirimkan surat permohonan izin penelitian kepada perusahaan untuk mendapatkan akses data. Langkah ini diperlukan karena data umpan balik hanya dapat diakses oleh pihak perusahaan. Data umpan balik terdiri dari 2.010 ulasan berbahasa Indonesia yang dikumpulkan selama tahun 2023 dari dua sumber yang berbeda. Pertama, 1.810 ulasan berasal dari keluhan yang disampaikan pelanggan langsung kepada tim CRM melalui telepon. Kedua, 200 ulasan tambahan diperoleh dari database umpan balik pengguna di *website* JuraganCOD.com.

4. TAHAP TEXT STANDARDIZATION

Pada tahap *text standardization*, penulis menyesuaikan format data agar sesuai dengan aturan yang telah ditetapkan. Penulis menggabungkan dua sumber data yang memiliki format dan jumlah tabel yang berbeda, serta memilih kolom-kolom tertentu untuk digunakan dalam analisis. Selain itu, penulis mengonversi format file dari Excel ke CSV. Penulis juga menyamakan format data dengan melakukan *case folding* dan menghapus karakter non-alfabetikal.

5. TAHAP TEXT PRE-PROCESSING

Pada tahap *text pre-processing*, penulis menghapus *stopword* atau kata-kata umum yang tidak memiliki makna signifikan. Selain itu, dilakukan *stemming*, yaitu

proses mengubah kata berimbuhan menjadi kata dasarnya. Penulis juga melakukan *tokenizing*, yang mengubah kalimat menjadi serangkaian kata terpisah.

6. TAHAP *FEATURE EXTRACTION*

Pada tahap *feature extraction*, dilakukan pengambilan fitur-fitur yang representatif dan relevan dari data mentah untuk digunakan dalam model. Fitur-fitur ini bisa berupa kata, frasa, atau karakteristik lainnya yang diekstrak dari dokumen. Dalam konteks ini, penulis menggunakan metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk membentuk model topik yang lebih akurat dan informatif.

7. TAHAP *ANALYSIS*

Pada tahap ini, dilakukan *topic modeling* menggunakan algoritma *Latent Dirichlet Allocation* (LDA) dengan parameter K topik, yang bervariasi dari 1 hingga 15, dan jumlah iterasi model yang bervariasi dari 1 hingga 100 kali. Pemilihan jumlah topik dan iterasi dilakukan secara eksploratif untuk menemukan konfigurasi yang menghasilkan skor koherensi terbaik. Setelah model topik dibentuk, evaluasi dilakukan menggunakan metode *topic coherence* untuk memastikan kualitas dan relevansi model yang dihasilkan. *Topic coherence* digunakan untuk menentukan jumlah topik terbaik berdasarkan nilai *topic coherence* yang dihasilkan, di mana skor koherensi yang tinggi menandakan kualitas model yang baik.

8. TAHAP *DISCOVERY*

Pada tahap *Discovery*, dilakukan interpretasi dari pola atau pengetahuan (*knowledge*) yang ditemukan dari hasil analisis. Penulis juga melakukan visualisasi data untuk memperoleh pemahaman yang lebih mendalam dan untuk mendukung analisis yang lebih baik. Visualisasi ini membantu dalam mengidentifikasi hubungan dan tren yang mungkin tidak terlihat dari data mentah saja.