

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu masalah utama yang dihadapi di wilayah metropolitan Jakarta (WMJ) adalah kemacetan lalu lintas. Dampak dari permasalahan ini menimbulkan kerugian ekonomi yang diperkirakan mencapai 50,2 triliun rupiah setiap tahunnya, dan angka tersebut diproyeksikan akan meningkat menjadi 65 triliun rupiah pada tahun 2020. Oleh karena itu, diperlukan langkah-langkah untuk mengurangi kemacetan di Jakarta. Salah satu solusi yang dapat diterapkan adalah dengan mendorong penggunaan transportasi umum sebagai alternatif bagi para komuter, yang dapat meningkatkan mobilitas harian dan mempercepat waktu perjalanan [1].

Pemerintah mengembangkan sektor transportasi untuk menyediakan layanan yang efisien, aman, terjangkau, nyaman, serta mendukung kemajuan sosial dan ekonomi di berbagai daerah. Terdapat berbagai jenis moda transportasi yang tersedia bagi masyarakat, namun kereta api masih menjadi pilihan utama saat ini. Hal ini dikarenakan kereta api memiliki beragam keunggulan, seperti hemat energi, biaya operasional yang rendah, dan kapasitas angkut yang besar [2].

Salah satu moda transportasi kereta api yang digunakan oleh masyarakat di Indonesia untuk menghindari kemacetan dan kepenatan di jalan raya adalah *Commuter Line* (Kereta Rel Listrik). Namun, kondisi fisik serta kenyamanan KRL sering kali belum memenuhi keinginan pelanggan. Sehubungan dengan hal tersebut, guna menyelesaikan permasalahan ini pemerintah kemudian menyediakan alternatif transportasi umum lainnya, yaitu MRT. *Mass Rapid Transit* (MRT) adalah sistem transportasi massal perkotaan berbasis rel cepat yang pembangunannya dimulai pada tahun 2013. Jalur pertama MRT Jakarta resmi beroperasi pada 24 Maret 2019 [3].

Seiring dengan kemajuan teknologi dan informasi PT. MRT Jakarta juga meluncurkan Aplikasi resmi bernama “MyMRTJ” pada April 2019. Aplikasi ini dirancang untuk menyediakan informasi mengenai jadwal, rute perjalanan, stasiun, tarif, pembelian tiket *online*, pelacakan perjalanan, serta informasi terbaru terkait kondisi operasional MRT Jakarta. Aplikasi MyMRTJ dapat diunduh di perangkat

dengan sistem operasi *android* dan *IOS* melalui *Google Playstore* maupun *Appstore*. Hingga kini, Aplikasi tersebut telah diunduh sebanyak 500 lebih pengguna dengan memperoleh *rating* sebanyak 4.5 dari 5.0 dan telah di review 52 ribu lebih pengguna di *Google Play Store*[4].

Setiap aplikasi pasti memiliki kekurangan dan kelebihan yang dapat dilihat dari ulasan para penggunanya. Jumlah ulasan yang ada di *Google Play Store* sangat banyak dan tidak terorganisir dengan baik, sehingga menjadi sulit untuk dipahami. Ulasan tersebut mengandung berbagai opini, baik yang positif (seperti kepuasan, pujian, dan saran) maupun yang negatif (seperti keluhan, kekecewaan, umpatan, dan sarkasme) yang menggambarkan kepuasan pengguna terhadap aplikasi MyMRTJ [5]. Oleh karena itu, analisis sentimen diperlukan untuk mengidentifikasi kecenderungan ulasan pengguna, apakah lebih condong ke sentimen positif, negatif, atau netral [6].

Analisis sentimen merupakan studi yang fokus pada pengidentifikasian pendapat, perasaan, dan opini yang diungkapkan melalui tulisan. Tujuan dari analisis ini adalah untuk menentukan arah kecenderungan opini seseorang terkait suatu kejadian atau isu, apakah opini tersebut cenderung positif, negatif, atau netral. Namun, banyak penelitian mengenai analisis sentimen yang hanya mengklasifikasikan sentimen dalam dua kategori yaitu positif dan negatif dan mengabaikan kategori netral karena dianggap tidak penting atau tidak relevan. Metode yang digunakan adalah *text mining*, sebuah teknik yang dirancang untuk mengekstrak informasi yang berguna dari data teks yang tidak terorganisir [7].

Analisis sentimen saat ini telah digunakan di berbagai sektor termasuk produk konsumen, layanan, kesehatan, keuangan, serta acara sosial dan politik. Pengaruh dan manfaatnya di berbagai sektor tersebut mendorong banyak penelitian tentang analisis sentimen. Hal ini digunakan sebagai landasan untuk mengembangkan produk dan layanan yang berkualitas dalam dunia usaha [8].

Berbagai algoritma dapat diterapkan untuk mengklasifikasikan data berbentuk teks, seperti *K-Nearest Neighbor* (K-NN), *Naïve Bayes*, *Maximum Entropy*, *Deep Learning*, dan *Support Vector Machine* yang umumnya digunakan dalam analisis sentimen. Selanjutnya, proses klasifikasi dilakukan dengan memanfaatkan pendekatan yang berbasis kamus atau *Knowledge Base Approach*, yaitu *Lexicon-*

Based dan *Point Mutual Information*. Dalam penelitian ini, peneliti menggunakan metode *K-Nearest Neighbor* (K-NN) yang dipadukan dengan pendekatan *Lexicon-Based* digunakan untuk mengklasifikasikan ulasan aplikasi MyMRTJ ke dalam tiga kategori sentimen: positif, negatif, dan netral.

Peneliti memilih metode *K-Nearest Neighbor* (K-NN) dengan pendekatan *Lexicon-Based* karena setelah melakukan tinjauan pustaka terhadap beberapa metode serupa yang sudah ada sebelumnya. Adapun beberapa referensi yang mendukung penelitian ini antara lain: Penelitian yang membahas Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN. Melalui tahap pemodelan dengan pembagian data 60:40 antara data latih dan data uji, hasil yang diperoleh menunjukkan nilai akurasi, presisi, dan *recall* untuk aplikasi Bibit masing-masing sebesar 85,14%, 91,91%, dan 76,44%. Sedangkan untuk aplikasi Bareksa, nilai-nilai tersebut berturut-turut adalah 81,70%, 87,15%, dan 75,73% [9].

Penelitian selanjutnya membahas tentang Penggunaan Metode Lexicon Untuk Analisis Sentimen pada Ulasan Aplikasi KAI Access di *Google Play Store*. Penelitian ini bertujuan untuk mengklasifikasikan opini pengguna ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, dengan menggunakan metode *lexicon*. Hasil penelitian menunjukkan bahwa mayoritas ulasan pengguna aplikasi KAI Access cenderung memiliki sentimen negatif dengan total 24.700 ulasan atau 60,01% dari keseluruhan data yang dianalisis [10].

Penelitian berikutnya berkaitan dengan Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode *Lexicon Based* dan *Naïve Bayes Classifier*. Proses awal dalam analisis sentimen melibatkan pra-pemrosesan data yang mencakup langkah-langkah seperti *filtering*, *case folding*, koreksi kata tidak baku, penanganan kata bernegasi, penghapusan *stopword*, dan penghilangan spasi yang berlebih. Hasil pengujian menunjukkan bahwa analisis sentimen dengan algoritma *Naive Bayes Classifier* pada data tweet memperoleh akurasi sistem sebesar 71%, dengan tingkat kesalahan sistem sebesar 29%, yang ditunjukkan melalui *confusion matrix* [7].

Penelitian tentang Analisis Sentimen Pengguna Gopay Menggunakan Metode *Lexicon Based* Dan *Support Vector Machine*. Sentimen dianalisis melalui media

sosial Twitter, dengan langkah-langkah penelitian yang mencakup *preprocessing* dan pemberian label opini dalam kategori positif dan negatif menggunakan metode *Lexicon Based*. Data yang digunakan terdiri dari 1210 opini terkait ulasan Go-Pay yang diperoleh dari Twitter. Proses pelabelan menggunakan metode *Lexicon Based* menghasilkan 923 opini dalam kategori positif dan 287 opini dalam kategori negatif. Di sisi lain, klasifikasi menggunakan metode SVM dengan kernel linear mencapai akurasi 89,17%, sedangkan dengan kernel *Polynomial* menghasilkan akurasi 84,38% [11].

Penelitian mengenai Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode *Support Vector Machine* dan *K Nearest Neighbor*. Tujuan penelitian ini adalah melakukan klasifikasi pada ulasan Aplikasi Gojek menggunakan metode *Support Vector Machine* (SVM) dan *K Nearest Neighbor* (KNN). Diperoleh hasil akurasi menggunakan Metode KNN dengan nilai $K=22$ memperoleh nilai akurasi, presisi, dan recall berturut-turut sebesar 82,14%, 82,28%, dan 95,43%, sedangkan metode SVM dengan kernel linear dan parameter $C=1$ memperoleh nilai akurasi, presisi, dan recall berturut-turut sebesar 87,98%, 88,55%, dan 95,43% [12].

1.2 Perumusan Masalah

1.2.1 Permasalahan

Berdasarkan penjelasan latar belakang di atas, masalah yang dapat diidentifikasi dalam penelitian ini berkaitan dengan banyaknya ulasan pada aplikasi MyMRTJ di *Google Play Store* yang masih tidak terstruktur, sehingga sulit dipahami. Selain itu, ulasan tersebut mengandung berbagai opini, baik yang positif (seperti kepuasan, pujian, dan saran) maupun yang negatif (seperti keluhan, kekecewaan, umpatan, dan sarkasme) dari pengguna aplikasi MyMRTJ di *Google Play*. Dengan demikian, diperlukan analisis sentimen untuk mengidentifikasi kecenderungan dalam ulasan pengguna, apakah lebih condong ke sentimen positif, negatif, atau netral.

1.2.2 Metode Usulan

Dalam menganalisis sentimen, peneliti mengusulkan 2 metode yaitu metode *K-Nearest Neighbor* dengan pendekatan *Lexicon-Based* untuk memberi bobot positif,

negarif, dan netral pada hasil dari masing-masing metode yang memungkinkan pengambilan keputusan yang lebih akurat.

1.2.3 Pertanyaan Penelitian

1. Bagaimana akurasi yang dihasilkan dari penggabungan metode *K-Nearest Neighbor* dengan pendekatan *Lexicon Based* dalam menganalisis sentimen ulasan pengguna aplikasi MyMRTJ di *Google Play Store*?
2. Bagaimana dampak penerapan *SMOTE* terhadap kinerja model *K-Nearest Neighbor* dalam klasifikasi sentimen pada ulasan aplikasi?

1.3 Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis dan mengevaluasi tingkat akurasi dan efisiensi diperoleh dari penerapan metode *K-Nearest Neighbor* (K-NN) dengan pendekatan *Lexicon-Based* dalam mengklasifikasikan sentimen pada ulasan pengguna aplikasi MyMRTJ. Selain itu, melihat pengaruh penggunaan *SMOTE* terhadap ketidakseimbangan data dalam ulasan tersebut.

1.3.2 Manfaat Penelitian

Penelitian ini dapat memberikan pemahaman bagi pengembang untuk mengidentifikasi pola sentimen dalam merespon umpan balik pengguna dengan lebih cepat dan efektif, guna meningkatkan kualitas layanan aplikasi berdasarkan pengalaman dan kebutuhan pengguna. Selain itu juga diharapkan dapat memberi wawasan kepada mahasiswa mengenai kombinasi metode *K-Nearest Neighbor* (K-NN) dan pendekatan *Lexicon-Based* dengan penerapan teknik *SMOTE* dapat digunakan dalam analisis sentimen untuk memahami lebih dalam kepuasan dan ketidakpuasan pengguna terhadap layanan aplikasi, khususnya pada konteks aplikasi transportasi umum di Indonesia. Sehingga dapat diambil keputusan-keputusan yang lebih baik untuk perbaikan layanannya.

1.4 Batasan Masalah

- 3.6.1 Data yang digunakan dalam penelitian diperoleh dari komentar pengguna aplikasi MyMRTJ yang ada di *Google Play Store*.
- 3.6.2 Data ulasan yang dianalisis dibatasi dari 3 tahun terakhir dengan total data sebanyak 1210 ulasan.

- 3.6.3 Data ulasan aplikasi My MRTJ diperoleh menggunakan teknik *web scrapping* dengan memanfaatkan *library google_play_scraper* pada *Python* melalui situs *Google Play Store*.
- 3.6.4 Hasil dari analisis sentimen berupa sentiment positif, negatif, dan netral.
- 3.6.5 Penelitian ini menggunakan kombinasi algoritma *K-Nearest Neighbor* dengan pendekatan berbasis *Lexicon* sebagai metode yang diterapkan.
- 3.6.6 Bahasa yang digunakan dalam pendekatan *Lexicon-Based* adalah Bahasa Indonesia.

1.5 Metodologi Penulisan Proposal

1. BAB I PENDAHULUAN

Pada bab Pendahuluan berisikan penjelasan mengenai apa yang melatarbelakangi masalah pada tugas akhir ini. Melalui rumusan masalah akan menjadi tujuan memperoleh solusi, tujuan dan manfaat untuk menyelesaikan permasalahan pada tugas akhir ini.

2. BAB II: TINJAUAN PUSTAKA

Pada bab ini membahas berbagai teori terkait metode *K-Nearest Neighbor* (K-NN) dan *Lexicon-Based* yang telah diterapkan dalam penelitian-penelitian terdahulu.

3. BAB III: METODE PENELITIAN

Pada bab ini membahas tentang langkah-langkah/alur untuk melakukan analisis sentimen ulasan aplikasi MyMRTJ menggunakan metode *K-Nearest Neighbor* (KNN) dan pendekatan *Lexicon-Based*.

4. BAB IV: HASIL DAN PEMBAHASAN

Pada bab ini berisi penjelasan lebih detail hasil klasifikasi metode *K-Nearest Neighbor* (KNN) dan pendekatan *Lexicon-Based* serta penerapan teknik *oversampling* SMOTE.

5. BAB V: KESIMPULAN DAN SARAN

Pada bab ini disajikan kesimpulan dan saran yang diperoleh dari hasil keseluruhan penelitian ini.