

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah mendorong peningkatan dalam bidang pemrosesan bahasa alami (*Natural Language Processing*), terutama dalam analisis sentimen berbasis teks. Salah satu teknik yang krusial dalam ekstraksi fitur dari data teks adalah *N-gram*, yaitu metode statistika yang digunakan untuk membentuk rangkaian kata sebanyak  $n$  berdasarkan urutan kemunculannya dalam kalimat atau dokumen (Dhinora & Mailoa, 2025). *N-gram* dapat berupa *unigram* (satu kata), *bigram* (dua kata), *trigram* (tiga kata) dan seterusnya. Dengan pendekatan ini, relasi antar kata dalam sebuah konteks menjadi lebih terlihat sehingga mampu menangkap makna yang lebih kompleks dibandingkan representasi kata tunggal. Dalam analisis sentimen, teknik *N-gram* memegang peranan penting karena dapat mendeteksi pola-pola kalimat yang mengindikasikan sentimen tertentu, baik positif maupun negatif, secara lebih presisi. Dengan penerapan *N-gram*, informasi kontekstual antar kata dapat lebih mudah dipetakan sehingga mendukung proses klasifikasi, khususnya dalam analisis sentimen.

Analisis sentimen dilakukan terhadap ulasan aplikasi ojek daring yang tersedia di *Google Play Store*. Aplikasi yang menjadi objek pada penelitian ini adalah ACI, yaitu aplikasi layanan transportasi daring asli karya anak bangsa Indonesia. ACI merupakan singkatan dari "Aku Cinta Indonesia", yang tidak hanya berfungsi sebagai aplikasi ojek online, tetapi juga mengusung semangat nasionalisme dan pemberdayaan ekonomi lokal. Secara singkat, ACI dapat didefinisikan sebagai aplikasi transportasi daring berbasis lokal yang menyediakan layanan ojek, pengantaran makanan, kurir, dan layanan lainnya. Seiring dengan meningkatnya penggunaan layanan transportasi berbasis aplikasi, ulasan dan opini pengguna menjadi indikator penting dalam menilai kualitas layanan yang diberikan (Sahabuddin dkk., 2025). Oleh karena itu, untuk memperkuat pemahaman terhadap definisi tersebut, dilakukan analisis teks menggunakan metode *N-gram* untuk mengidentifikasi pola kata yang dominan dalam ulasan aplikasi ACI.

Dalam penelitian ini, pendekatan yang dipilih adalah menggunakan metode *Support Vector Machine* (SVM). *Support Vector Machine* (SVM) ditemukan oleh

Vladimir N. Vapnik dan Alexey Ya. Chervonenkis pada tahun 1963, sebagai bagian dari pengembangan teori *statistical learning*. Pada awalnya, SVM digunakan untuk menyelesaikan masalah klasifikasi linier sederhana. Namun, seiring berkembangnya teknologi dan kebutuhan akan metode klasifikasi yang lebih kompleks, SVM mengalami berbagai penyempurnaan, terutama melalui pengenalan konsep kernel pada tahun 1992 yang memungkinkan algoritma ini digunakan untuk klasifikasi data non-linier (Iqbal dkk., 2024). SVM banyak digunakan karena efektif dalam menyelesaikan masalah klasifikasi, regresi, dan prediksi pada data linier maupun nonlinier dalam jumlah besar (Mukhtar dkk., 2022).

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin yang berfungsi untuk melakukan klasifikasi maupun regresi dengan cara mencari sebuah *hyperplane* atau garis pemisah terbaik yang mampu memisahkan data dari dua kelas secara optimal (Pratama dkk., 2025). Algoritma ini memaksimalkan margin antara dua kelas data, sehingga meningkatkan kemampuan generalisasi model terhadap data baru. Titik data yang paling berpengaruh dalam pembentukan *hyperplane* disebut sebagai *support vectors*. SVM sangat efektif dalam menangani data berdimensi tinggi dan sering digunakan dalam analisis sentimen karena kemampuannya dalam mengolah data teks (Br Sinulingga & Sitorus, 2024), terutama ketika dikombinasikan dengan teknik representasi fitur seperti *N-gram*.

Beberapa penelitian terkait telah mengkaji penggunaan metode klasifikasi dalam analisis sentimen dengan penerapan *N-gram*. Salah satu penelitian berjudul "Analisis Sentimen pada Twitter Mengenai Pasca Bencana Menggunakan Metode *Naïve Bayes* dengan Fitur *N-Gram*" melakukan empat kali pengujian dengan persentase data pelatihan yang berbeda-beda. Hasil pengujian menunjukkan bahwa jumlah data latih memiliki pengaruh terhadap peningkatan akurasi. Pada ekstraksi fitur *unigram*, diperoleh akurasi sebesar 76,67%, 84,44%, 90,00%, dan 93,33%. Sementara itu, akurasi pada *bigram* tercatat sebesar 64,17%, 68,89%, 75,00%, dan 86,67%. Dari keseluruhan pengujian, *unigram* menghasilkan akurasi tertinggi sebesar 93,33%, lebih tinggi dibandingkan *bigram* yang mencapai akurasi tertinggi sebesar 86,67%. Hasil ini menunjukkan bahwa fitur *unigram* cenderung memberikan hasil yang lebih baik dalam analisis sentimen dibandingkan dengan

*bigram*. Penelitian lain berjudul "Analisis Sentimen *Review Skincare Skintific* dengan Algoritma *Support Vector Machine (SVM)*" membahas penerapan algoritma SVM dengan kernel linear untuk mengklasifikasikan sentimen positif dan negatif pada ulasan produk *skincare* berbahasa Indonesia. Sebanyak 958 ulasan dari hasil penelitian menunjukkan bahwa model SVM mampu mencapai akurasi sebesar 94%, dengan nilai *precision*, *recall*, dan *f1-score* yang masing-masing melebihi 0,9. Dari hasil klasifikasi, ditemukan sebanyak 862 ulasan dengan sentimen positif dan 96 ulasan dengan sentimen negatif. Dengan demikian dalam penelitian ini menggunakan metode Klasifikasi *Support Vector Machine (SVM)* dengan penerapan *N-gram* untuk analisis sentimen ulasan pada aplikasi ACI di *Google Play Store*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat diambil beberapa rumusan masalah sebagai berikut:

1. Bagaimana proses penerapan fitur *N-Gram (unigram, bigram dan trigram)* dalam ekstraksi data teks ulasan pengguna aplikasi ACI?
2. Bagaimana kinerja algoritma *Support Vector Machine (SVM)* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi ACI berdasarkan fitur *N-Gram*?
3. Apakah kombinasi fitur *unigram, bigram, dan trigram* memberikan hasil klasifikasi yang lebih baik dibandingkan dengan penggunaan *unigram* saja?

## 1.3 Tujuan Penelitian

Berikut adalah beberapa tujuan penelitian yang akan dilakukan:

1. Menerapkan teknik *N-Gram (unigram, bigram dan trigram)* untuk mengekstraksi fitur dari data ulasan pengguna aplikasi ACI.
2. Mengembangkan model klasifikasi sentimen berbasis *Support Vector Machine (SVM)* dengan fitur hasil ekstraksi *N-Gram*.

3. Mengevaluasi dan membandingkan performa model klasifikasi menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

#### 1.4 Batasan Penelitian

Adapun batasan masalah yang ditetapkan adalah sebagai berikut:

1. Data yang dianalisis terbatas pada 1.419 ulasan pengguna aplikasi ACI yang tersedia di platform *Google Play Store*.
2. Tahapan *tokenizing* dilakukan dengan menerapkan fitur *unigram*, *unigram + bigram* dan *unigram + trigram*.
3. Pelabelan data dilakukan dengan pendekatan berbasis leksikon (*Lexicon-Based*).
4. Algoritma *Support Vector Machine* (SVM) digunakan dalam tahap pemodelan data pada penelitian ini.
5. Pengolahan data pada penelitian ini dilakukan menggunakan bahasa pemrograman *python* melalui platform *Google Colaboratory* (*Google Colab*).

#### 1.5 Manfaat Penelitian

Beberapa manfaat yang dapat diambil dari penelitian ini adalah:

1. Memberikan kontribusi terhadap pengembangan metode analisis sentimen, khususnya dalam pendekatan algoritma *Support Vector Machine* (SVM) dengan penerapan *n-gram*.
2. Menjadi referensi bagi pengembang aplikasi layanan ojek *online*, khususnya ACI, dalam memahami persepsi dan kepuasan pengguna melalui analisis ulasan di *platform* digital.
3. Menyediakan dasar bagi penelitian lanjutan di bidang *text mining* dan *Natural Language Processing*, khususnya yang berkaitan dengan klasifikasi opini berbasis ulasan pengguna berbahasa Indonesia.