BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada Januari 2025 Indonesia memiliki 143 juta pengguna media sosial, setara dengan 50,2% populasi nasional (SIMON, 2025). Dari jumlah tersebut, 25,2 juta di antaranya aktif di X atau yang dulu dikenal Twitter (Ribka, 2024). Platform media sosial X juga dikenal karena memberi kebebasan berpendapat yang ditawarkannya, sehingga berbagai isu sosial sering diperbincangkan di sana (Mahilla, 2023). Salah satu topik yang paling banyak dibahas baru-baru ini adalah peluncuran lembaga Danantara, lembaga baru milik negara yang berfungsi sebagai badan pengelola kekayaan dan investasi negara Indonesia (Nur, 2025). Diluncurkan pada 24 Februari 2025, Danantara bertugas mengelola seluruh aset BUMN dan investasi strategis untuk mendukung pertumbuhan ekonomi nasional jangka panjang (Melan, 2025). Peluncuran lembaga ini menjadi topik pembicaraan di berbagai kalangan. Oleh karena itu, memahami opini publik dapat menjadi landasan penting dalam menilai keberhasilan suatu proyek, Data media sosial yang umumnya tidak terstruktur dan bervolume besar dapat menyulitkan proses analisis manual. Dibutuhkan metode klasifikasi otomatis untuk mengidentifikasi kecenderungan opini masyarakat terkait isu tersebut.

Analisis sentimen merupakan sebuah gabungan dari beberapa bidang ilmu seperti *Machine Learning*, *Natural Language Processing* (*NLP*), umumnya analisis sentimen sering digunakan untuk mengetahui kecenderungan opini publik terkait suatu isu, masalah, atau peristiwa tertentu (Purnamasari dkk. 2023). Dalam praktiknya terdapat beberapa metode algoritma yang sering digunakan dalam menganalisis sentimen seperti *Support Vector Machine* (*SVM*), *Naive Bayes*, *Decision Tree*, dan *K-nearest neigbhor* (*KNN*).

Support Vector Machine adalah sebuah algoritma *machine learning* yang bekerja dengan konsep *supervised learning* agar model dapat mempelajari pola-pola dalam data tersebut. pola-pola dalam data tersebut, SVM dapat melakukan tugas klasifikasi dan regresi. Dalam bidang analisis sentimen algoritma SVM dinilai sangat efektif untuk tugas klasifikasi teks (Purnamasari dkk. 2023).

Penelitian yang dilakukan Hendrastuty dkk. (2021) dengan menggunakan *Support Vector Machine* untuk mengklasifikasi opini publik pada program kartu kerja, hasilnya SVM mencapai akurasi 98,67% dan F1-score mencapai 98% ini artinya bahwa SVM dapat melakukan klasifikasi opini dari teks yang berasal dari sosial media *Twitter* dengan optimal. Penelitian menggunakan SVM untuk melakukan klasifikasi opini publik menggunakan data yang berasal dari sosial media juga pernah dilakukan Tukino dan Fifi (2024) dengan topik layanan Gojek di Kota Batam dengan hasil akurasi mencapai 88,5%. Hasil kedua penelitian ini menunjukkan bahwa SVM dapat melakukan klasifikasi data teks yang berasal dari sosial media dengan efektif. Penelitian yang dilakukan (Damayanti dkk. 2024) dengan menggunakan algoritma *support vector machine* untuk melakukan klasifikasi opini publik terkait kebijakan BPJS yang diterapkan pemerintah, hasil evaluasi menunjukkan bahwa model mencapai akurasi 94,28%.

Berdasarkan beberapa penelitian terdahulu juga menyatakan jika algoritma *support vector machine* lebih unggul dari beberapa algoritma machine learning lainnya, seperti yang dilakukan Septiana dan Alita (2024) dengan melakukan perbandingan antara SVM dan *Random Forest* untuk melakukan klasifikasi opini publik terkait hasil perhitungan cepat pemilu 2024, hasilnya SVM mencapai akurasi 80% dan *Random Forest* yang hanya mencapai akurasi 78%. Senada dengan temuan tersebut, dalam penelitian Pamungkas dan Kharisudin (2021) yang membandingkan SVM dengan *Naïve Bayes* dan KNN di mana hasil yang diperoleh menunjukkan bahwa SVM memiliki performa yang lebih baik dari kedua algoritma tersebut

Berdasarkan hasil yang ditunjukkan penelitian terdahulu, algoritma *Support Vector Machine* dinilai sangat tepat untuk digunakan dalam penelitian ini, di mana SVM digunakan untuk melakukan klasifikasi opini publik terkait lembaga investasi negara Danantara.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, berikut ini merupakan permasalahan dalam penelitian ini:

1. Bagaimana efektivitas algoritma *Support Vector Machine* dalam melakukan klasifikasi opini publik berbasis teks terhadap peluncuran lembaga Danantara.
2. Bagaimana mayoritas sentimen publik terkait peluncuran lembaga Danantara.

1.3 Ruang Lingkup

Penelitian ini akan dilakukan dengan ruang lingkup sebagai berikut:

1. Objek Penelitian: Opini publik di platform X (twitter) terkait peluncuran Lembaga Danantara
2. Pengumpulan Data: Data dikumpulkan dengan *tools tweet-harvest* pada platform sosial media X (twitter). Pengumpulan ini dilakukan pada rentang waktu 24 Februari 2025 (Peluncuran Danantara) sampai dengan 14 Maret 2025 (Pengumpulan data dilakukan).
3. Variabel Penelitian: Opini publik yang terbagi menjadi dua yaitu positif dan negatif.
4. Metode Klasifikasi: Metode klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM).
5. Model SVM: Menggunakan model SVM dengan *kernel linear* yang telah dilatih menggunakan data berlabel (positif atau negatif) yang diperoleh dengan metode *lexicon-based*.
6. Pengujian Dan Evaluasi Kinerja: Model yang telah dilatih kemudian diuji dengan algoritma *Support Vector Machine*. Setelah model diuji, untuk mengetahui performanya dievaluasi dengan *confusion matrix* untuk menghitung beberapa kriteria penilaian dalam model klasifikasi, kriterianya sebagai berikut ini:
    a. Akurasi
    b. Presisi
    c. Recall
    d. F1-Score

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan sebagai berikut:

1. Menganalisis dan mengevaluasi performa algoritma *Support Vector Machine* dalam melakukan klasifikasi opini publik berdasarkan teks terkait peluncuran lembaga Danantara.
2. Mengidentifikasi mayoritas sentimen publik terkait peluncuran lembaga Danantara

1.5 Manfaat Penelitian

Manfaat yang dihasilkan dari penelitian ini adalah mengetahui kecenderungan opini publik terhadap peluncuran Danantara yang dapat dijadikan landasan bagi lembaga/instansi terkait untuk merumuskan strategi komunikasi yang tepat.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat secara terstruktur dan sistematis untuk mempermudah pembaca dalam memahami alur penelitian yang dilakukan, berikut ini merupakan sistematika penulisan dari penelitian ini:

BAB 1 PENDAHULUAN

Bab ini menjelaskan alasan yang mendasari sehingga penelitian ini dilakukan seperti perumusan masalah yang akan dijawab, ruang lingkup penelitian, tujuan yang ingin dicapai, manfaat yang diharapkan dari penelitian, dan sistematika penulisan secara keseluruhan.

BAB 2 TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini berisi tinjauan literatur terhadap penelitian terdahulu yang memiliki topik atau tema serupa dengan penelitian ini. Dan juga bab ini akan berisi penjelasan konsep-konsep dasar dari metode yang diterapkan dalam penelitian ini seperti Data Mining, Analisis Sentimen, *Natural Language Processing*, Ekstraksi fitur (TF-IDF), dan pelabelan (lexicon).

BAB 3 METODE PENELITIAN

Bab ini berisi metode penelitian yang dilakukan seperti metode pengumpulan data dan prosedur penelitian yang meliputi pra-pemrosesan teks, pelabelan, pembagian

data, ekstraksi fitur, penanganan data tidak seimbang dengan SMOTE, hingga pelatihan dan pengujian model.

BAB 4 IMPLEMENTASI DAN PEMBAHASAN

Bab ini berisi implementasi teknis yang dilakukan dalam penelitian ini seperti pengumpulan data, pembersihan data, pra-pemrosesan teks, pemodelan dan evaluasi model. Dan juga bab ini berisi pembahasan hasil implementasi yang dilakukan.

BAB 5 PENUTUP

Bab ini berisi kesimpulan dari penelitian yang menjawab permasalahan dan tujuan penelitian ini dilakukan. Dan juga bab ini berisi saran-saran yang relevan bagi penelitian dengan topik serupa pada masa yang akan datang.

BAB 6 DAFTAR PUSTAKA

Bab ini berisi seluruh sumber-sumber referensi yang dijadikan sebagai acuan dalam melakukan penelitian ini.