

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial seperti Twitter, Facebook, dan Instagram mengalami lonjakan pengguna dalam beberapa tahun terakhir, memungkinkan penyebaran opini dan emosi secara luas terhadap berbagai isu global maupun lokal. Twitter, dengan lebih dari 330 juta pengguna aktif dan sifat komunikasinya yang real-time, menjadi sumber utama data digital yang dimanfaatkan dalam banyak studi karena volume dan kecepatan arus informasi yang tinggi (Krishna & Prashanth, 2023). Namun, tingginya jumlah dan keberagaman data menjadikan analisis manual tidak efisien, sehingga dibutuhkan pendekatan otomatis yang mampu mengekstraksi, mengolah, dan mengklasifikasikan opini publik secara sistematis. Dalam konteks ini, metode berbasis machine learning menjadi solusi yang relevan untuk analisis data besar dari media sosial (Mishra, 2023).

Salah satu pendekatan utama dalam mengolah opini publik adalah analisis sentimen, proses untuk mengidentifikasi sikap atau emosi dalam teks, apakah bersifat positif, negatif, atau netral (Alsaeedi & Khan, 2020). Teknik ini memungkinkan sistem untuk memahami persepsi publik secara otomatis dan telah menjadi fokus penting dalam penelitian data science dan kecerdasan buatan. Jumlah publikasi ilmiah terkait sentiment analysis tercatat meningkat secara konsisten sejak 2008 (Rodríguez-Ibáñez et al., 2023).

Berbagai penelitian sebelumnya membuktikan efektivitas metode machine learning dalam klasifikasi sentimen. Gusti, (2023) menunjukkan bahwa Logistic Regression memberikan akurasi tertinggi dalam klasifikasi bencana alam di Twitter (92.4%), diikuti oleh SVM (91.82%) dan Naïve Bayes (91.85%). Dahlian & Sitanggang, (2023) dalam analisis sentimen migrasi televisi digital menemukan bahwa SVM unggul dengan akurasi 94%. Studi lainnya oleh Hadi & Sugiarto, (2025) menunjukkan efektivitas SVM dalam menganalisis opini publik terhadap pembangunan IKN dengan akurasi 80%. Penelitian Mustopo & Afiyati, (2025) 3 pada isu Food Estate menunjukkan Logistic Regression unggul (74%), disusul SVM (73%) dan Naïve Bayes (72%). Sahat et al., (2023) dan Assiroj et al., (2023)

juga menunjukkan dominasi Logistic Regression dalam konteks institusi pendidikan dan layanan imigrasi.

Meskipun hasil-hasil tersebut membuktikan efektivitas ketiga algoritma, sebagian besar penelitian masih terbatas pada isu lokal seperti kebijakan pemerintah atau layanan publik. Dataset yang digunakan pun cenderung kecil dan bersifat statis dalam satu rentang waktu, sehingga kurang merepresentasikan dinamika topik yang cepat berubah. Selain itu, pengujian model sering kali tidak memperhitungkan kestabilan performa terhadap variasi data.

Oleh karena itu, penelitian ini bertujuan untuk melakukan evaluasi komparatif terhadap algoritma Naïve Bayes, SVM, dan Logistic Regression dalam menganalisis sentimen publik terhadap Bitcoin menggunakan data Twitter berbahasa Inggris tahun 2025 dalam jumlah besar. Penelitian ini diharapkan memberikan kontribusi dalam pemilihan model klasifikasi paling efektif untuk menangani data besar bertopik global dengan opini yang kompleks.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, penelitian ini difokuskan untuk menjawab pertanyaan sebagai berikut :

1. Bagaimana performa algoritma Naïve Bayes, Support Vector Machine, dan Logistic Regression dalam mengklasifikasikan Sentimen Tweet tentang Bitcoin di tahun 2025?
2. Algoritma manakah yang menghasilkan akurasi dan efektivitas terbaik berdasarkan metrik evaluasi klasifikasi seperti akurasi, presisi, recall dan F1-Score?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk :

1. Menganalisis dan membandingkan performa algoritma Naïve Bayes, Support Vector Machine, dan Logistic Regression dalam klasifikasi sentimen tweet tentang Bitcoin.
2. Menentukan algoritma terbaik berdasarkan hasil evaluasi klasifikasi menggunakan metrik akurasi, presisi, recall dan F1-Score

1.4 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut :

1. Teoritis

Dapat memberikan pengetahuan sebagai landasan dan referensi untuk penelitian selanjutnya yang berhubungan dengan Bitcoin serta menjadi bahan kajian lebih lanjut.

2. Praktis

a. Bagi Masyarakat

Memberikan informasi kepada masyarakat mengenai opini publik terhadap cryptocurrency, khususnya Bitcoin, melalui hasil analisis sentimen dari media sosial Twitter.

b. Bagi Penulis

Menambah wawasan dan pengalaman dalam penerapan data mining machine learning, khususnya dalam melakukan analisis sentimen menggunakan algoritma Naïve Bayes, Support Vector Machine dan Logistic Regression

