

BAB I.

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital di sektor keuangan mengubah cara pelaku pasar mengakses dan merespons informasi. Selain laporan keuangan dan indikator teknikal, percakapan publik di media sosial, portal berita, dan komunitas investor kini memengaruhi persepsi dan perilaku transaksi. Studi menunjukkan sentimen publik, perhatian, dan intensitas pencarian daring dapat berkorelasi dengan volatilitas harga (Audrino dkk., 2020; Costola dkk., 2023). Namun, data historis harga dan volume perdagangan tetap menjadi prediktor utama karena mencerminkan respons pasar terhadap informasi dan kondisi ekonomi (Broby, 2022; Dakalbab dkk., 2024). Oleh karena itu, peran sentimen sebagai variabel pelengkap perlu diuji secara kontekstual, khususnya di pasar negara berkembang seperti Indonesia.

Dalam konteks domestik, pertumbuhan bank digital menambah kompleksitas pasar. Saham bank digital cenderung volatil, memiliki basis pengguna daring, dan terekspos percakapan media sosial. PT Bank Jago Tbk (ARTO) sering menjadi topik di *platform X*, terkait layanan, integrasi *fintech*, dan isu operasional. Intensitas percakapan ini berpotensi memengaruhi persepsi investor dan keputusan jangka pendek. Namun, data sentimen bersifat tidak terstruktur, mengandung *noise* (*spam*, promosi, *bot*), ironi, dan bahasa tidak baku, sehingga sinyal sentimen lemah tanpa teknik *Natural Language Processing* (NLP) yang memadai (Wu dkk., 2022; Chaudhari dan Mahajan, 2025). Selain itu, perbedaan frekuensi dan sinkronisasi waktu antara percakapan dan jadwal perdagangan menimbulkan asinkronisasi temporal yang dapat menurunkan korelasi (Audrino dkk., 2020; Costola dkk., 2023).

Secara teoretis, peran sentimen terhadap harga saham dijelaskan melalui dua perspektif. *Efficient Market Hypothesis* (EMH) menyatakan harga mencerminkan seluruh informasi yang tersedia, sehingga sentimen publik diperkirakan tidak berpengaruh signifikan, khususnya dalam jangka pendek (Han,

2025). Sebaliknya, *behavioral finance* menyoroiti bias psikologis seperti *herding*, *overreaction*, dan *underreaction* yang membuat harga menyimpang dari nilai fundamental, sehingga sentimen publik dapat berperan, terutama di pasar yang belum efisien (Fazlija dan Harder, 2022; Awan dkk., 2021). Perbedaan ini menunjukkan bahwa pengaruh sentimen bersifat empiris dan kontekstual, sehingga perlu diuji pada objek dan periode spesifik.

Perkembangan *big data* dan *machine learning* menyediakan kerangka komputasional untuk menguji hipotesis secara sistematis. Algoritma seperti *Linear Regression*, *Random Forest*, dan *Neural Network* mampu mengolah data multivariabel dan sebagian hubungan *nonlinier*, sedangkan pendekatan NLP memungkinkan ekstraksi representasi kuantitatif dari teks (Huang dkk., 2022; Jain dan Kulkarni, 2023). Integrasi data historis (harga, volume, indikator teknikal) dan data sentimen dari media sosial dapat digunakan untuk menilai korelasi antara persepsi publik dan fluktuasi harga, serta mengevaluasi kontribusi sentimen terhadap akurasi estimasi berbasis data historis (Shahi dkk., 2020; Lin dan Wei, 2024).

Studi terdahulu menunjukkan hasil beragam terkait integrasi sentimen dan data historis. Wu dkk. (2022) mengusulkan model *Long Short-Term Memory* (LSTM) dengan *Convolutional Neural Network* (CNN) untuk sentimen dan melaporkan penurunan *Mean Absolute Error* (MAE) dibanding pendekatan tunggal. Polireddi (2024) menyoroiti potensi *Artificial Intelligence* (AI) di perbankan, meski bukti pada saham individual masih terbatas. Costola dkk. (2023) menggunakan FinBERT untuk berita COVID-19 dan menemukan korelasi positif dengan *return* indeks saham yang mencakup 500 perusahaan besar di Amerika Serikat. Hu dkk. (2023) dan Bhandari dkk. (2022) membandingkan algoritma klasik dan *boosting*, menunjukkan *Light Gradient Boosting Machine* (LightGBM) unggul untuk jangka waktu investasi panjang, sedangkan *Random Forest* efektif di jangka waktu investasi pendek. Lanbouri dan Achhab (2020) menegaskan LSTM efektif untuk prediksi intraday, sementara Koratamaddi dkk. (2021) memadukan *deep reinforcement learning* dengan sentimen untuk portofolio. Kajian sistematis

oleh Dakalbab dkk. (2024) menekankan dominansi *deep learning* dan metrik *Root Mean Squared Error* (RMSE), sedangkan Broby (2022) menegaskan data historis tetap menjadi fondasi utama. Fazlija dan Harder (2022) menunjukkan sentimen berita dengan *Random Forest* dapat mengungguli *random walk*, meski berbasis berita, bukan media sosial umum.

Secara umum, celah yang tampak ialah terbatasnya studi yang menguji saham bank digital Indonesia (bukan pasar maju atau indeks lebar), mengintegrasikan sentimen media sosial umum (X/Twitter) dengan data historis untuk saham individual, dan menangani tantangan temporal (sinkronisasi hari bursa), *noise* bahasa Indonesia (ironi/sarkasme), serta validasi ketat (analisis residual & *cross-validation*) untuk menilai kontribusi sentimen terhadap akurasi estimasi.

Kontribusi utama penelitian ini adalah memberikan bukti empiris berbasis data lokal mengenai keterbatasan integrasi sentimen media sosial terhadap akurasi estimasi harga saham bank digital Indonesia, sekaligus menawarkan kerangka metodologis yang transparan dan dapat direplikasi. Penelitian ini menekankan pentingnya penyelarasan temporal, validasi silang, dan analisis residual sebagai praktik evaluasi yang ketat, serta menggarisbawahi bahwa sentimen sebaiknya diposisikan sebagai indikator pelengkap, bukan variabel utama, dalam sistem pendukung keputusan investasi. Dengan demikian, penelitian ini tidak hanya mengisi celah literatur terkait pasar Indonesia, tetapi juga memberikan dasar konseptual bagi pengembangan sistem informasi keuangan yang lebih adaptif terhadap dinamika data tekstual.

1.2 Identifikasi masalah

Agar analisis korelasi valid dan terukur, sejumlah persoalan utama diidentifikasi berdasarkan fenomena empiris dan temuan studi terdahulu, yaitu:

1. Kualitas & representativitas sentimen: percakapan di X mengandung *noise* (*spam*, promosi, *bot*), ironi/sarkasme, serta bahasa tidak baku yang mengaburkan

sinyal sentimen sehingga korelasi terhadap fluktuasi harga menjadi lemah atau tidak stabil (Wu dkk., 2022; Chaudhari dan Mahajan, 2025).

2. Asinkronisasi temporal: puncak percakapan sentimen tidak selalu bertepatan dengan titik perubahan harga; tanpa rekonsiliasi waktu dan agregasi yang tepat, keterkaitan dapat terekam terlalu rendah atau bias (Audrino dkk., 2020; Costola dkk., 2023).
3. Dominannya informasi historis: pada banyak emiten, harga atau volume historis menjadi prediktor utama sehingga kontribusi sentimen sebagai variabel komplementer sering tidak signifikan atau inkonsisten, terutama untuk jangka waktu investasi sangat pendek (Broby, 2022; Dakalbab dkk., 2024).
4. Keterbatasan sumber sentimen: bergantung pada satu platform (X) dapat menyebabkan bias audiens sehingga hasil korelasi tidak mewakili lanskap opini publik yang lebih luas. Penelitian ini juga menggunakan satu platform (X) karena keterbatasan akses data, sehingga potensi bias diakui sebagai batasan (Awan dkk., 2021; Wu dkk., 2022).
5. Tantangan metodologis: rekayasa fitur sentimen konversi teks menjadi skor, pemilihan jendela agregasi, dan metrik kinerja (MAE, RMSE, R^2) harus selaras agar penilaian pengaruh sentimen terhadap estimasi berbasis historis tidak menyesatkan (Lin dan Wei, 2024; Shahi dkk., 2020).

Identifikasi masalah tersebut menjadi dasar dalam merumuskan pertanyaan penelitian yang lebih spesifik, sebagaimana dijabarkan pada subbab berikutnya.

1.3 Perumusan masalah

Merujuk identifikasi di atas, pertanyaan penelitian diarahkan untuk memberikan fokus analitis yang jelas, yaitu:

1. Apakah penambahan sentimen memperbaiki akurasi estimasi berbasis historis yang diukur dengan MAE, RMSE, dan R^2 ?
2. Sejauh mana integrasi sentimen media sosial X dengan data historis ARTO berkorelasi dengan fluktuasi harga saham?
3. Faktor kualitas sentimen apa saja yang paling memengaruhi keterdeteksian korelasi?

4. Bagaimana implikasi temuan terhadap teori *Efficient Market Hypothesis* (EMH) dan *behavioral finance* serta penerapannya dalam sistem informasi keuangan?
5. Bagaimana rekomendasi metodologis dan praktis yang dapat disusun berdasarkan hasil evaluasi integrasi sentimen untuk pengembangan sistem informasi keuangan yang transparan dan adaptif?

Rumusan masalah ini selanjutnya dijabarkan menjadi maksud dan tujuan penelitian yang difokuskan pada korelasi integrasi antara data sentimen pasar dan data historis saham.

1.4 Maksud dan Tujuan Penelitian

Sebagai tindak lanjut dari perumusan masalah, penelitian ini memiliki maksud menguji korelasi integrasi data historis dan sentimen media sosial X terhadap fluktuasi harga saham ARTO dengan menggunakan algoritma *Linear Regression*, *Random Forest*, dan *Neural Network*. Tujuan berikut dirancang untuk menjawab rumusan masalah secara menyeluruh:

1. Mengevaluasi pengaruh integrasi sentimen media sosial terhadap akurasi estimasi harga saham berbasis data historis menggunakan metrik MAE, RMSE, dan R^2 .
2. Membandingkan performa model berbasis data historis dengan model berbasis integrasi sentimen menggunakan algoritma *Linear Regression*, *Random Forest*, dan *Neural Network*.
3. Menganalisis faktor-faktor yang memengaruhi kontribusi sentimen terhadap hasil korelasi integrasi data.
4. Menginterpretasikan temuan empiris dalam kerangka teori *Efficient Market Hypothesis* (EMH) dan *behavioral finance* untuk menilai relevansi sentimen terhadap efisiensi pasar.
5. Merumuskan rekomendasi metodologis dan praktis bagi pengembangan sistem informasi keuangan yang transparan dan adaptif terhadap integrasi data tekstual.

Tujuan tersebut tidak hanya diharapkan mampu memberikan solusi atas permasalahan yang ada, tetapi juga berkontribusi pada pengembangan pengetahuan baru dalam literatur. Oleh karena itu, pada bagian selanjutnya akan dijelaskan kebaruan yang ditawarkan oleh penelitian ini.

1.5 Kebaruan

Kebaruan penelitian bersifat kontekstual, metodologis, dan konseptual:

1. Kontekstual: fokus pada emiten bank digital Indonesia (ARTO) untuk menguji korelasi integrasi sentimen historis terhadap fluktuasi harga; studi spesifik semacam ini relatif jarang di konteks lokal.
2. Metodologis: algoritma yang telah ada digunakan sebagai instrumen analitis untuk membandingkan estimasi berbasis historis dibandingkan dengan berbasis integrasi sentimen, dengan penekanan pada keselarasan waktu, normalisasi fitur, dan analisis *residual*.
3. Konseptual: penyusunan kerangka analisis yang mengaitkan peran sentimen terhadap fluktuasi harga dalam kerangka *Efficient Market Hypothesis* (EMH) dan *behavioral finance*, sehingga temuan empiris memiliki arah interpretasi jelas terhadap efisiensi lemah atau setengah kuat.

1.6 Manfaat Penelitian

Mengacu pada kebaruan yang telah diuraikan, manfaat penelitian ini dapat dilihat dari dua perspektif utama, yaitu manfaat akademis dan manfaat praktis. Berikut ini penjabaran manfaat yang diharapkan dari penelitian ini.

Dari sisi akademis, penelitian ini memperjelas arah korelasi antara integrasi sentimen pasar dan data historis terhadap fluktuasi harga saham ARTO, sekaligus menyediakan kerangka interpretatif yang menghubungkan perspektif *Efficient Market Hypothesis* (EMH) dengan pendekatan perilaku. Selain itu, penelitian ini menawarkan kerangka evaluatif berbasis metrik seperti MAE, RMSE, dan R^2 , termasuk analisis *residual*, untuk menilai kekuatan keterkaitan dalam lingkungan data yang memiliki *noise*. Penelitian ini juga memperluas kajian ilmiah mengenai

peran variabel tekstual khususnya sentimen media sosial dalam model estimasi yang selama ini didominasi oleh variabel numerik.

Dari sisi manfaat praktis, penelitian ini memberikan panduan bagi penggunaan data sentimen sebagai variabel pelengkap dalam sistem pendukung keputusan investasi, bukan sebagai determinan tunggal. Temuan penelitian ini juga menawarkan rekomendasi terkait proses pengolahan data dan kebijakan komunikasi korporat dalam mengelola sentimen publik yang berpotensi memengaruhi persepsi terhadap emiten. Selain itu, penelitian ini memberikan informasi mengenai batasan implementasi integrasi sentimen historis sehingga ekspektasi pelaku pasar tetap realistis.

Bagi pengembang sistem informasi keuangan, penelitian ini dapat dijadikan acuan dalam merancang sistem yang mampu mengolah data dari berbagai sumber, mempertimbangkan peran sentimen secara proporsional, serta menampilkan indikator kualitas estimasi secara transparan. Untuk memastikan bahwa seluruh manfaat tersebut dapat dicapai secara efektif, penelitian ini dibatasi oleh ruang lingkup tertentu yang dijelaskan pada subbab berikut.

1.7 Pembatasan masalah

Untuk menjaga fokus dan relevansi penelitian, ditetapkan beberapa pembatasan agar penelitian dapat dilaksanakan secara terarah dan menghasilkan temuan yang terukur. Pembatasan ini bertujuan memperjelas ruang lingkup analisis dalam mengkaji korelasi antara sentimen pasar dan fluktuasi harga saham bank digital di industri fintech. Adapun batasan penelitian yang digunakan adalah sebagai berikut:

1.7.1. Sumber Data Sentimen dan Harga Saham

Penelitian ini akan mengumpulkan data sentimen dan harga saham dari sumber tertentu saja yaitu PT Bank Jago Tbk melalui akun resmi di media sosial X (Twitter) pada url <https://x.com/jadijago> dan yahoo *finance* pada url <https://finance.yahoo.com/quote/ARTO.JK/>. Meskipun terdapat berbagai platform media sosial yang banyak digunakan di Indonesia, penelitian ini memilih platform X (Twitter) sebagai sumber data sentimen karena menyediakan data publik terbuka

dan didominasi oleh konten berbasis teks opini yang relevan dengan diskursus pasar keuangan. Platform seperti WhatsApp bersifat tertutup, sedangkan TikTok dan Instagram lebih berfokus pada konten visual, sehingga kurang sesuai untuk analisis sentimen berbasis teks dalam penelitian ini.

1.7.2. Periode Pengambilan Data

Data yang akan digunakan dalam penelitian ini dibatasi pada periode waktu tertentu yaitu Juli 2019 – Juli 2024. Hal ini bertujuan untuk memastikan bahwa model yang dikembangkan relevan dengan kondisi pasar terbaru dan untuk menghindari bias temporal. Data historis yang terlalu lama mungkin tidak mencerminkan sentimen pasar dan kondisi yang sekarang.

1.7.3. Ruang Lingkup Perusahaan Bank Digital

Fokus penelitian ini adalah pada saham perusahaan Bank digital dalam industri *finTech* tertentu yaitu PT Bank Jago Tbk. Tidak semua perusahaan *fintech* akan dianalisis, melainkan hanya perusahaan yang memiliki data sentimen dan harga saham yang cukup untuk dianalisis. Oleh karena itu, hasil penelitian mungkin tidak dapat digeneralisasikan ke semua perusahaan *fintech* atau sektor industri lainnya.

1.7.4. Jenis Algoritma Machine Learning

Penelitian ini akan menggunakan algoritma *Machine learning* tertentu untuk analisis dan prediksi. Meskipun beberapa algoritma mungkin akan dibandingkan, tidak semua metode atau teknik *Machine learning* akan diuji. Fokus akan berada pada algoritma yang paling relevan dan efektif untuk tujuan penelitian ini.

1.7.5. Indikator Teknis dan Fundamental

Indikator teknis dan fundamental yang digunakan dalam model ini akan dibatasi pada beberapa indikator yang dianggap paling signifikan. Contohnya, indikator teknis seperti harga penutupan dan volume perdagangan serta indikator fundamental seperti laporan keuangan. Indikator lain yang mungkin juga mempengaruhi harga saham tetapi tidak termasuk dalam penelitian ini akan menjadi batasan.

Dengan batasan-batasan yang telah dirumuskan di atas, fokus penelitian ini dapat dikembangkan secara terarah dan mendalam. Untuk mendukung pemahaman yang sistematis terhadap keseluruhan proses penelitian, bagian selanjutnya akan menjelaskan sistematika penulisan disertasi ini.

1.8 Sistematika penulisan

Untuk memberikan gambaran menyeluruh mengenai struktur dan alur logika penyusunan, sistematika penulisan disertasi ini disusun mengacu pada Buku Panduan Penulisan Proposal dan Disertasi DSI Undip Edisi 4 tahun 2024. Secara umum, disertasi ini terdiri dari tiga bagian utama, yaitu:

1. Bagian depan Disertasi

- a. Halaman sampul
- b. Halaman judul
- c. Halaman pengesahan
- d. Kata pengantar
- e. Abstrak (Bahasa Indonesia)
- f. *Abstract* (Bahasa Inggris)
- g. Daftar isi
- h. Daftar tabel
- i. Daftar gambar
- j. Daftar Istilah
- k. Daftar lampiran

2. Bagian isi Disertasi

Bab 1. Pendahuluan

- a. Latar belakang masalah
- b. Identifikasi masalah
- c. Perumusan masalah
- d. Maksud dan tujuan penelitian
- e. Kebaruan (*novelty*)
- f. Pembatasan masalah
- g. Sistematika penulisan

Bab 2. Kajian Pustaka

- a. Tinjauan pustaka
- b. Keaslian (Orisinalitas) penelitian
- c. Landasan teori

Bab 3. Metodologi penelitian

- a. Waktu dan tempat penelitian
- b. Alat dan Bahan penelitian
- c. Prosedur penelitian

Bab 4. Hasil Penelitian dan Pembahasan

- a. Hasil Penelitian
- b. Pembahasan

Bab 5. Penutup

- a. Kesimpulan
 - b. Saran dan rekomendasi untuk penelitian selanjutnya
3. Bagian Akhir Disertasi
 - a. Daftar pustaka
 - b. Lampiran
 4. Ringkasan Eksekutif

Sistematika penulisan ini memberikan gambaran menyeluruh mengenai struktur disertasi yang disusun secara sistematis dan logis. Setelah memberikan latar belakang dan ruang lingkup penelitian dalam Bab 1, pembahasan selanjutnya akan difokuskan pada Landasan Teori yang memuat tinjauan pustaka, keaslian penelitian, dan dasar-dasar teoretis yang menjadi fondasi utama dalam pengembangan model penelitian ini.