

1. Pendahuluan

Sentimen adalah pandangan yang dapat dilihat dari seseorang tentang suatu subjek tertentu [1]. Memungkinkan untuk memperoleh pendapat atau sudut pandang publik dengan menganalisis sentimen. Sentimen dapat diambil dari berbagai sumber, termasuk komentar media sosial [2]. Perusahaan dapat menggunakan sentimen untuk menentukan apakah keputusan, kebijakan, atau kinerja mereka pantas disebut "baik" [3]. Twitter adalah platform media sosial yang memungkinkan pengguna untuk menyuarakan pendapat mereka tentang berbagai topik. Menurut [1], Indonesia menempati peringkat keenam dalam hal jumlah pengguna Twitter per negara [4]. Hal ini menunjukkan bahwa pengguna Twitter Indonesia memainkan peran yang signifikan dalam hal keterlibatan. Twitter di Indonesia sering digunakan untuk berkomunikasi tentang informasi dan menyatakan pendapat tentang masalah terkait perusahaan dan saham.

Memprediksi pergerakan harga saham sulit dilakukan karena volatilitas pasar saham [2]. Dalam iklim ekonomi saat ini, berinvestasi di pasar saham adalah praktik umum. Karena pengaruh penawaran dan permintaan, nilai saham fluktuatif, dan harga saham bisa mengalami kenaikan maupun penurunan. Jika permintaan tinggi, harga akan naik, sedangkan jika pasokan tinggi, harga akan turun. Dasar-dasar makroekonomi, fluktuasi nilai tukar Rupiah terhadap mata uang asing, kebijakan pemerintah, faktor-faktor kepanikan, dan faktor-faktor manipulasi pasar adalah contoh faktor eksternal. Dasar-dasar perusahaan, tindakan korporat, dan kinerja adalah contoh faktor internal. Terdapat korelasi antara sentimen Twitter dan harga saham suatu perusahaan, menurut penelitian oleh [5] dan [6].

Menurut *Efficient Market Hypothesis* (EMH), semua informasi pasar yang tersedia dapat mempengaruhi harga saham [7]. Menurut penelitian oleh [5], terdapat korelasi antara sentimen dari tweet tentang sebuah perusahaan dan harga sahamnya. Beberapa peneliti telah melakukan penelitian tentang analisis sentimen, termasuk Musthofa Galih Pradana, yang menyelidiki dampak sentimen media sosial pada harga saham perusahaan. Studi ini menggunakan data dari bagian komentar Facebook. Data yang terkumpul diklasifikasikan, dan analisis regresi digunakan untuk menyelidiki hubungan antara sentimen Facebook dan harga saham. Studi ini menemukan bahwa hanya sentimen optimis yang memiliki korelasi dengan harga saham [8]. Studi lain oleh [9] menggunakan metode CNN untuk analisis sentimen. Dataset yang digunakan berasal dari Twitter, khususnya evaluasi produk. Studi ini menemukan bahwa metode CNN untuk klasifikasi teks mendapatkan *F-measure* sebesar 91,6% dalam skema skip-gram dan *F-measure* sebesar 86,6% dalam skema CBOW. Dalam investigasi terpisah, [10] menerapkan LSTM pada dataset sebanyak 50.000 contoh. Ulasan-ulasan tersebut dikategorikan sebagai positif atau negatif, dan model yang dihasilkan memperoleh tingkat akurasi sebesar 86,75%.

Dalam penelitian ini, analisis korelasi harga saham dengan menggunakan metode Spearman dan analisis sentimen dengan metode CNN-LSTM dibandingkan. Kelas-kelas dibagi menjadi tiga kategori: positif, negatif, dan netral. Selain itu, hasil terbaik pada dataset ulasan pariwisata Tiongkok dicapai dengan model CNN-LSTM dua saluran yang terintegrasi dengan leksikon. Eksperimen yang luas menunjukkan bahwa informasi dari leksikon sentimen dan model dua saluran paralel secara signifikan meningkatkan akurasi analisis sentimen [11]. Frekuensi-dokumen terbalik (dikenal juga sebagai TF-IDF) adalah metode terkenal untuk menentukan signifikansi suatu kata dalam dokumen. Dalam membandingkan TF-IDF dan N-Gram, kita dapat menyimpulkan bahwa TF-IDF adalah pilihan fitur yang lebih unggul jika algoritma pembelajaran mesin akan digunakan untuk klasifikasi teks. Ketika menggabungkan word2vec sebagai ekspansi fitur, akurasi meningkat secara signifikan karena algoritma belajar dari statistik jumlah kali setiap pasangan kata muncul [12]. Menyetel ulang *hyperparameter* model meningkatkan kinerja model pada set validasi. Dalam konteks pembelajaran mesin, *hyperparameter* adalah parameter yang nilainya ditentukan sebelum memulai proses pembelajaran. Sebaliknya, nilai-nilai parameter model diperoleh melalui pelatihan data. Parameter model adalah bobot dan koefisien yang algoritma hasilkan dari data [13]. Metode ini mengurangi sensitivitas terhadap *noise* atau *outlier*, yang dapat diukur dalam peringkat data daripada nilai aktual [14].