

Eksplorasi Pola Interaksi Pengguna Microblogging dengan Big Data dan Visualisasi Interaktif

**Oppir Hutapea^{*1}, Lamria Sitorus², Desrico Hizkia Siallagan³, Mega Paramita Malau⁴,
Frans Panjaitan⁵, Inggrid Pardede⁶, Roulina Sianipar⁷**

^{1,2,3,4,5,6,7}Teknologi Rekayasa Perangkat Lunak, Institut Teknologi Del, Laguboti, Toba

e-mail: ^{*1}oppir.hutapea@del.ac.id, ²lamriaa04@gmail.com, ³siallagandesrico@gmail.com,

⁴Megamalau912@gmail.com, ⁵franspanjaitan21z@gmail.com, ⁶inggridpardede72@gmail.com,

⁷roulinasianipar18@gmail.com

Abstrak

Perkembangan *platform microblogging* menghasilkan *volume* data yang sangat besar, menimbulkan tantangan dalam memahami pola interaksi pengguna secara efektif. Penelitian ini penting dilakukan untuk mengidentifikasi dinamika sosial dan tren topik melalui analisis big data dan visualisasi interaktif. Tujuan dari penelitian ini adalah mengembangkan metode eksplorasi pola interaksi pengguna dengan memanfaatkan teknik pengolahan data besar dan visualisasi yang interaktif. Metode yang digunakan meliputi pengumpulan data dari *platform microblogging*, analisis menggunakan algoritma *clustering* dan analisis jaringan, serta penyajian data melalui visualisasi dinamis. Hasil penelitian menunjukkan bahwa visualisasi interaktif mampu mempercepat pemahaman pola sosial dan tren informasi dalam data besar, serta memfasilitasi pengambilan keputusan yang lebih akurat. Kesimpulan ini menegaskan pentingnya penerapan big data dan visualisasi dalam memahami dinamika pengguna *microblogging*, serta memberikan kontribusi signifikan terhadap pengembangan sistem analisis sosial berbasis teknologi tersebut.

Kata kunci: *big data, visualisasi interaktif, pola interaksi, microblogging, analisis sosial*

1. PENDAHULUAN

Perkembangan media sosial telah merevolusi cara individu berinteraksi secara global. *Twitter* atau sekarang dikenal dengan *X*, sebagai salah satu *platform microblogging* paling populer, memungkinkan pengguna membagikan informasi singkat secara *real-time* dalam bentuk teks, gambar, maupun tautan [1]. Dari karakteristiknya yang cepat dan ringkas, *Twitter* telah menjadi sarana komunikasi yang responsif dan dinamis. Seiring dengan meningkatnya jumlah pengguna dan *volume tweet*, *platform* ini telah berkembang menjadi sumber data sosial yang melimpah untuk dianalisis, mencerminkan opini publik dan perilaku digital pengguna secara langsung. Di dalam konteks big data, data dari *Twitter* yang berjumlah jutaan per hari memerlukan pendekatan pemrosesan data berskala besar. *Hadoop* dan teknologi komputasi terdistribusi menjadi solusi penting untuk mengelola dan menganalisis data yang masif ini [2], [3]. Berbagai studi sebelumnya lebih banyak berfokus pada analisis isi (*content analysis*) dan jaringan sosial (*network analysis*) dari tweet. Namun, aspek temporal seperti pola waktu aktivitas pengguna, distribusi sentimen per jam, serta variasi aktivitas berdasarkan hari, masih jarang dieksplorasi secara mendalam menggunakan pendekatan big data dan visualisasi interaktif.

Pola interaksi menggunakan teknologi big data telah banyak dimanfaatkan, diantaranya pola interaksi untuk menganalisis perilaku belajar siswa [4], melihat tren pasar dalam industri retail [5], menganalisis data E-commerce dalam skala besar untuk meningkatkan pengalaman pelanggan, mengoptimalkan operasi, dan memanfaatkan peluang bisnis [6], menganalisis data

akademik [7] dan menganalisis data lainnya [8]–[10]. Penerapannya yang telah banyak diberbagai jenis data membuat penelitian ini menggunakan teknologi yang sama untuk mengeksplorasi dan memvisualisasikan dinamika aktivitas pengguna *Twitter* dari berbagai perspektif, tidak hanya dari segi waktu tetapi juga konten dan emosi. Analisis yang dilakukan mencakup visualisasi frekuensi kata menggunakan *word cloud*, analisis jumlah sentimen per tanggal, distribusi harian tiap jenis sentimen, distribusi sentimen berdasarkan jam, serta distribusi sentimen berdasarkan maskapai. Penelitian diharapkan dapat memberikan wawasan komprehensif tentang ritme digital masyarakat dan opini publik secara kontekstual melalui teknik visualisasi data yang informatif dan interaktif.

2. METODE PENELITIAN

Bab ini menjelaskan pendekatan dan tahapan metodologis yang digunakan dalam penelitian untuk mengeksplorasi pola interaksi pengguna pada *platform mikroblogging Twitter*. Metodologi yang diterapkan meliputi proses pengumpulan data, *preprocessing* menggunakan ekosistem *Hadoop*, hingga analisis dan visualisasi data berbasis *Python*. Setiap tahapan dirancang secara sistematis agar hasil analisis dapat merepresentasikan pola waktu interaksi pengguna secara akurat dan efisien.

2.1 Twitter

Penelitian ini berfokus pada pemrosesan big data, yaitu kumpulan data berukuran sangat besar, dengan kecepatan tinggi dan variasi yang kompleks, sehingga tidak dapat ditangani dengan pendekatan konvensional. Big Data memiliki lima karakteristik utama yang dikenal sebagai 5V: *Volume*, *Velocity*, *Variety*, *Veracity*, dan *Value* [11]. Data yang diperoleh dari *Twitter* berjumlah ribuan hingga jutaan entri pengguna secara *real-time* merupakan representasi nyata dari karakteristik Big Data, yang mencerminkan dinamika sosial dalam skala luas. *Twitter* sebagai objek penelitian merupakan *platform mikroblogging*, yaitu media sosial yang memungkinkan pengguna membagikan informasi singkat seperti teks, gambar, atau tautan secara cepat. Sejak diluncurkan pada tahun 2006, *Twitter* telah menjadi platform utama dalam penyebaran opini publik dan interaksi sosial. *Twitter* menyimpan berbagai jenis interaksi, seperti mention, retweet, dan hashtag, yang dapat dianalisis secara kronologis untuk mengungkap pola komunikasi pengguna [12]–[14].

Dalam pengelolaan dan analisis data *Twitter* yang berskala besar, penelitian ini memanfaatkan *Hadoop* sebagai kerangka kerja untuk proses *preprocessing*. *Hadoop* memungkinkan pemrosesan paralel dan terdistribusi terhadap kumpulan data besar, sehingga mampu meningkatkan efisiensi dalam menyaring, membersihkan, dan mengelompokkan data mentah sebelum dianalisis lebih lanjut [15]. Tahapan selanjutnya melibatkan proses data mining, yaitu teknik untuk menggali informasi atau pola tersembunyi dari data besar [16]–[18]. Dalam penelitian ini, data mining digunakan untuk mengeksplorasi isi tweet dan menemukan struktur topik yang sering muncul. Salah satu metode utama yang digunakan adalah *Latent Dirichlet Allocation* (LDA). LDA merupakan teknik pemodelan topik yang memandang setiap dokumen sebagai campuran dari beberapa topik tersembunyi [19].

Penerapan LDA dalam penelitian ini bertujuan untuk mengidentifikasi tema-tema utama yang dibahas pengguna *Twitter* dalam periode waktu tertentu, sehingga memungkinkan pemetaan topik terhadap dimensi temporal. Setelah topik utama diidentifikasi, informasi divisualisasikan menggunakan pendekatan visualisasi interaktif. Visualisasi interaktif mempermudah pengguna dalam mengidentifikasi pola dari kumpulan data kompleks [20]. Visualisasi dalam penelitian ini tidak hanya menyajikan hasil analisis dalam bentuk grafis, tetapi juga memungkinkan eksplorasi lebih lanjut berdasarkan waktu, topik, dan intensitas interaksi.

Salah satu jenis visualisasi yang digunakan adalah Word Cloud, yaitu representasi visual dari kumpulan kata, di mana ukuran kata mencerminkan frekuensi atau tingkat kepentingannya dalam teks [21], [22]. Semakin sering suatu kata muncul, semakin besar tampilannya dalam Word Cloud. Tata letak dan pewarnaan kata dirancang untuk memperjelas kategori, frekuensi, serta memberikan nilai estetika. Word Cloud digunakan dalam penelitian ini untuk menampilkan kata-kata dominan dari hasil topik modeling LDA, serta memberikan gambaran umum topik pembicaraan pengguna dalam rentang waktu tertentu. Representasi visual ini membantu memperkuat pemahaman terhadap hasil analisis teks secara intuitif dan informatif. Dengan kombinasi konsep dan pendekatan tersebut, metodologi yang diterapkan dalam penelitian ini memungkinkan identifikasi dan penyajian pola waktu interaksi pengguna Twitter secara komprehensif, dari tahap data mentah hingga bentuk visualisasi yang mudah dipahami.

2.2 Arsitektur Sistem

Penelitian ini menggunakan arsitektur sistem berbasis Hadoop untuk menangani pemrosesan data berskala besar, serta memanfaatkan Python sebagai alat utama dalam proses analitik dan visualisasi. Adapun tahapan dalam arsitektur sistem ini meliputi [3]:

- Pengambilan data tweet dari dataset berformat CSV.
- Penyimpanan data ke dalam Hadoop Distributed File System (HDFS).
- Pemrosesan data menggunakan MapReduce untuk mengekstraksi informasi waktu dan isi tweet.
- Pemodelan topik menggunakan algoritma Latent Dirichlet Allocation (LDA).
- Visualisasi hasil analisis menggunakan pustaka Python seperti matplotlib, pyLDAvis, dan wordcloud.

Struktur ini mendukung efisiensi pengelolaan data skala besar serta memberikan fleksibilitas dalam eksplorasi hasil analisis.

2.3 Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle, berjudul "*Twitter US Airline Sentiment*", yang berisi sebanyak 14.640 entri tweet dari pengguna Twitter di Amerika Serikat. Dataset ini dikumpulkan oleh CrowdFlower dan bertujuan untuk mengklasifikasikan opini publik terhadap layanan beberapa maskapai penerbangan besar di AS, seperti United, American, Delta, Southwest, dan US Airways. Dataset ini terdiri dari 15 atribut, di antaranya:

- tweet_id: ID unik dari setiap tweet.
- airline_sentiment: Kategori sentimen dari tweet (positif, netral, atau negatif).
- airline_sentiment_confidence: Nilai keyakinan terhadap klasifikasi sentimen.
- negativereason: Alasan spesifik untuk sentimen negatif (jika tersedia).
- airline: Nama maskapai yang dibahas dalam tweet.
- text: Isi teks tweet.
- tweet_created: Waktu dan tanggal tweet dikirim.
- tweet_location dan user_timezone: Informasi lokasi dan zona waktu pengguna

Dataset ini sangat relevan untuk penelitian berbasis Big Data karena mencakup aspek volume dan variasi yang tinggi. Selain mengandung teks mentah, data ini juga memungkinkan analisis sentimen, waktu, dan korelasi topik. Dalam penelitian ini, dataset digunakan untuk melakukan analisis temporal terhadap pola interaksi pengguna dan mengidentifikasi topik dominan dengan metode LDA. Selanjutnya, hasil dari topik modeling divisualisasikan dalam bentuk Word Cloud untuk memudahkan pemahaman isi pembicaraan pengguna dalam berbagai rentang waktu.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksplorasi pola interaksi pengguna pada platform *microblogging* dengan mengambil twitter sebagai contoh. Kemudian data di *preprocessing* menggunakan big data *framework* yaitu X lalu dilakukan analisis dengan data mining untuk mendapatkan insight menggunakan python. Analisis dipusatkan pada distribusi sentimen terhadap maskapai penerbangan yang terekam dalam tweet pengguna. Data yang telah diproses divisualisasikan secara interaktif untuk mengidentifikasi tren, persepsi, dan kecenderungan opini publik terhadap masing-masing maskapai. Pembahasan dilakukan untuk menafsirkan hasil visualisasi dan memberikan insight terhadap pola komunikasi digital di media sosial. Setiap sub bagian berikut akan membahas tahapan analisis secara rinci berdasarkan pendekatan Big Data dan visualisasi interaktif yang telah diterapkan.

3.1 Preprocessing Data

Preprocessing dataset sentimen Twitter tentang maskapai di Amerika dengan Hadoop dilakukan untuk membersihkan dan menyiapkan data teks yang tidak terstruktur agar siap dianalisis. Dataset umumnya berformat CSV dan berisi kolom seperti tweet ID, nama pengguna, maskapai, isi tweet, dan label sentimen (positif, netral, atau negatif). Karena tweet mengandung banyak noise seperti URL, mention, hashtag, dan simbol khusus, tahap pembersihan dilakukan menggunakan MapReduce. Pada proses Map, setiap baris dibaca dan teks tweet dibersihkan dengan menghapus karakter tidak penting, mengubah ke huruf kecil, dan menghilangkan stopwords atau baris kosong. Selanjutnya, proses Reduce mengumpulkan dan menyusun ulang tweet yang telah dibersihkan untuk membentuk dataset yang konsisten dan layak pakai. Hasil akhir dari preprocessing ini dapat digunakan untuk analisis sentimen, pelatihan model klasifikasi, atau pemantauan reputasi maskapai, dan Hadoop memungkinkan semua proses ini dilakukan secara efisien dalam skala besar dengan pemrosesan paralel.

3.2 Topik Modeling setiap Sentimen dengan metode LDA

Setelah dilakukan preprocessing terhadap data, langkah pertama yang dilakukan adalah mengidentifikasi topik-topik yang dibahas dalam data tersebut. Untuk tujuan ini, digunakan metode Latent Dirichlet Allocation (LDA) guna melakukan topic modeling berdasarkan masing-masing kategori sentimen. LDA membantu mengungkap topik-topik utama dari setiap sentimen dan mengelompokkannya sesuai dengan kemunculannya dalam teks seperti terlihat pada Gambar 1.

```
Topik 1: ['jetblue', 'lost', 'help', 'just', 'bag', 'customer', 'americanair', 'service', 'usairways', 'united']
Topik 2: ['hours', 'help', 'flightled', 'jetblue', 'southwestair', 'hold', 'usairways', 'americanair', 'cancelled', 'flight']
Topik 3: ['jetblue', 'gate', 'plane', 'time', 'southwestair', 'delayed', 'americanair', 'usairways', 'united', 'flight']
```

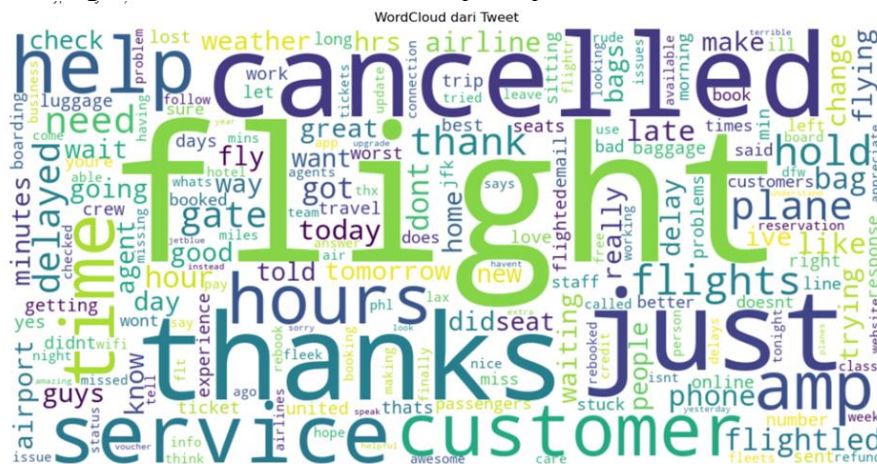
Gambar 1. Hasil LDA

Berdasarkan hasil analisis dan visualisasi yang diperoleh, ditemukan beberapa insight penting. Pada sentimen negatif, topik-topik yang dominan mencakup keluhan terkait bagasi yang hilang, penundaan dan pembatalan penerbangan, serta pengalaman buruk di bandara, seperti keterlambatan dan permasalahan pada gate keberangkatan.

3.3 Visualisasi Frekuensi Kata Menggunakan Word Cloud

Word Cloud digunakan untuk memvisualisasikan kata-kata utama dari masing-masing topik yang dihasilkan oleh LDA. Setelah LDA mengelompokkan dokumen ke dalam beberapa topik berdasarkan distribusi kata, Word Cloud dapat dibuat untuk setiap topik secara terpisah agar

kita bisa melihat kata-kata apa saja yang paling dominan dalam satu topik tertentu atau seberapa penting atau sering kata tersebut muncul dalam topik seperti Gambar 2.

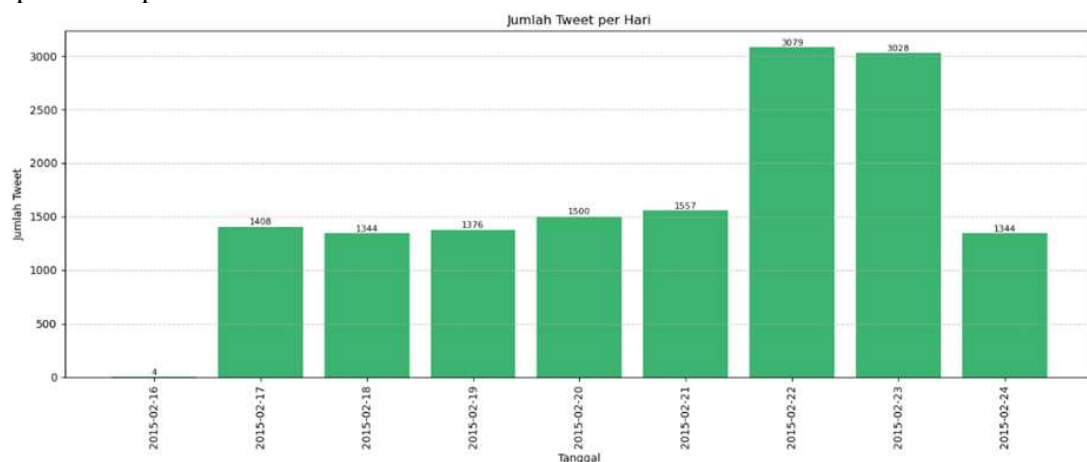


Gambar 2. Hasil penggunaan Word Cloud

Dari hasil visualisasi menggunakan Word Cloud, kita dapat mengidentifikasi pola-pola berupa keluhan maupun pujian dari pelanggan. Kata-kata seperti *flight*, *cancelled*, *delay*, *help*, *lost*, dan *waiting* menunjukkan adanya keluhan dari customer. Sementara itu, istilah seperti *service* dan *airport* cenderung merefleksikan sentimen yang netral. Di sisi lain, kata-kata seperti *thanks*, *thank*, *really*, dan *great* mencerminkan aspek-aspek yang diapresiasi dan menjadi bentuk pujian.

3.4 Analisis Jumlah Sentimen per Tanggal

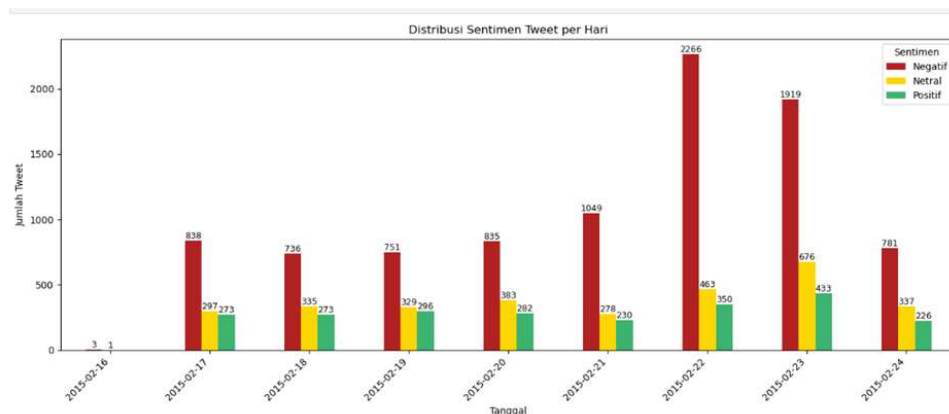
Perolehan insight mengenai tanggal-tanggal di mana pengguna lebih aktif menuliskan tweet di Twitter, dibuatlah sebuah visualisasi data. Dari hasil visualisasi tersebut terlihat adanya lonjakan jumlah tweet terkait maskapai yang signifikan selama dua hari berturut-turut, yaitu pada tanggal 22 dan 23 Mei 2015. Jumlah tweet pada periode tersebut tercatat hampir tiga kali lebih banyak dibandingkan dengan hari-hari biasa. Kenaikan ini kemungkinan dipengaruhi oleh kejadian atau layanan dari maskapai tertentu. Oleh karena itu, diperlukan analisis lebih lanjut terhadap isi dan sentimen tweet yang muncul pada tanggal tersebut. Visualisasi hasil analisis dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Sentimen per Tanggal

3.5 Distribusi Harian Tiap Jenis Sentimen

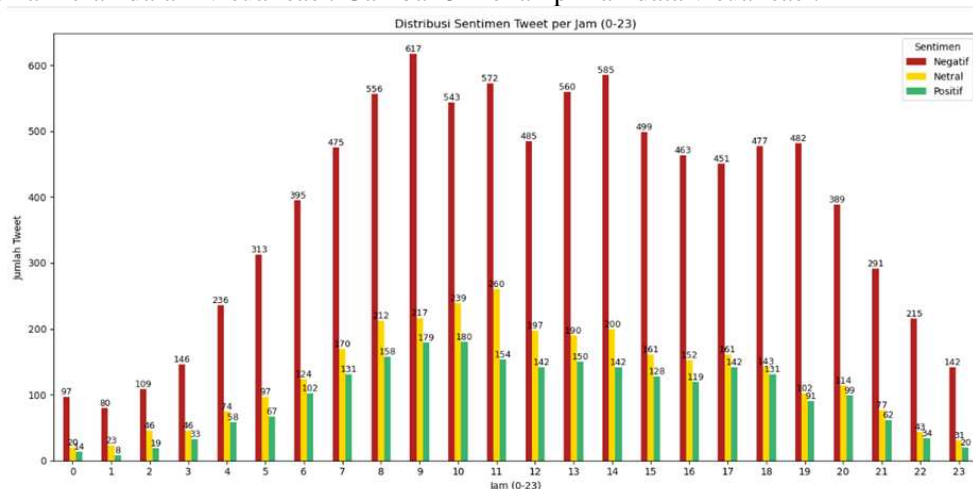
Selanjutnya dilakukan analisis lanjutan terhadap isi dan sentimen tweet yang muncul pada masing-masing tanggal, guna melihat persebaran jenis sentimen, apakah negatif, netral, atau positif. Hasil visualisasi disajikan dalam tiga warna: merah untuk sentimen negatif, kuning untuk sentimen netral, dan hijau untuk sentimen positif. Dari visualisasi tersebut terlihat bahwa lebih dari 50% tweet setiap harinya didominasi oleh sentimen negatif dari para pengguna. Selain itu, terdapat beberapa tanggal dengan lonjakan aktivitas tweet yang diduga disebabkan oleh buruknya pelayanan maskapai, seperti keterlambatan (delay) atau pembatalan penerbangan (cancel). Insight yang diperoleh dari temuan ini adalah bahwa pelanggan maskapai cenderung lebih sering menyampaikan kekecewaan dibandingkan kepuasan terhadap layanan yang mereka terima. Hasil visualisasi dapat dilihat pada Gambar 4.



Gambar 4. Visualisasi Distribusi Tipe Sentimen Setiap Harinya

3.6 Distribusi Sentimen Berdasarkan Jam

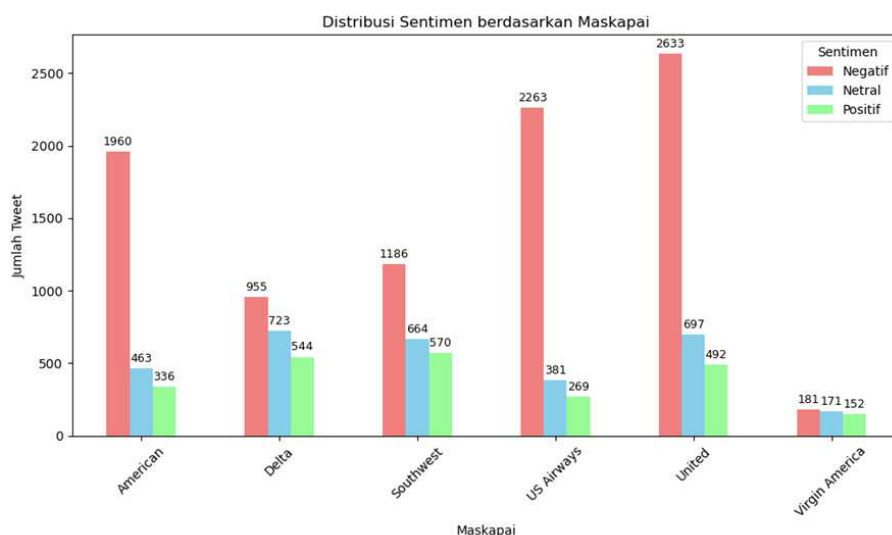
Selain mengetahui hari dan jenis sentimen yang paling sering dituliskan oleh pengguna Twitter, penting juga untuk memahami waktu atau jam tertentu ketika pengguna paling aktif menulis tweet. Oleh karena itu, dibuatlah sebuah visualisasi interaktif guna menganalisis pola aktivitas pengguna berdasarkan jam. Hasil visualisasi menunjukkan bahwa aktivitas pengguna Twitter cenderung meningkat antara pukul 08.00 hingga 14.00. Pada rentang waktu tersebut, mayoritas tweet yang diposting memiliki sentimen negatif, yang ditampilkan melalui batang berwarna merah dalam visualisasi. Gambar 5 menampilkan data visualisasi.



Gambar 5. Visualisasi Distribusi Sentimen Berdasarkan Jam

3.7 Distribusi Sentimen Berdasarkan Maskapai

Insight penting lainnya yang perlu diperoleh adalah informasi terkait masing-masing maskapai, mulai dari maskapai yang paling banyak disebut hingga yang paling jarang. Selain itu, penting juga untuk mengetahui maskapai mana yang paling banyak menerima sentimen negatif maupun positif. Berdasarkan tujuan tersebut, dibuatlah sebuah visualisasi interaktif. Dari hasil visualisasi, terlihat bahwa terdapat enam maskapai yang menerima komentar dari pengguna Twitter. Maskapai United muncul sebagai yang paling banyak disebut, menunjukkan tingkat aktivitas tertinggi berdasarkan jumlah tweet yang membahas maskapai tersebut. Sebaliknya, Virgin America menjadi maskapai yang paling jarang disebut, dilihat dari jumlah tweet yang jauh lebih sedikit. Namun, perlu dicermati bahwa sebagian besar maskapai cenderung menerima sentimen negatif secara dominan dibandingkan sentimen netral maupun positif. Kondisi ini kemungkinan disebabkan oleh faktor yang serupa, seperti keterlambatan (*delay*) atau pembatalan penerbangan (*cancel*) yang mungkin terjadi akibat perubahan cuaca atau gangguan operasional lainnya. Visualisasi dapat dilihat pada Gambar 6.



Gambar 6. Visualisasi Distribusi Sentimen Berdasarkan Maskapai

4. KESIMPULAN

Penelitian ini membuktikan bahwa pendekatan berbasis big data dan visualisasi interaktif mampu memberikan pemahaman dan sejumlah insight yang mendalam terhadap pola interaksi pengguna di *platform microblogging Twitter*. Insight tersebut mencakup informasi mengenai waktu-waktu ketika pengguna paling aktif menggunakan Twitter, serta faktor-faktor yang menyebabkan sebuah maskapai mendapatkan sentimen negatif dari pengguna dan lain-lain. Penggunaan model *Latent Dirichlet Allocation (LDA)*, telah berhasil mengidentifikasi lima topik utama percakapan publik yang mencerminkan dinamika sosial digital. Analisis temporal menunjukkan bahwa interaksi pengguna memiliki pola waktu tertentu, dengan puncak aktivitas terjadi pada pagi dan siang hari. Hal ini sangat berguna bagi pelaku industri komunikasi digital dan pemasaran untuk menentukan waktu yang tepat dalam menyampaikan pesan atau kampanye. Visualisasi interaktif yang dibangun juga terbukti efektif dalam menyederhanakan kompleksitas data besar menjadi bentuk representasi yang mudah dipahami dan ditelusuri oleh pengguna. Kelebihan dari pendekatan ini adalah kemampuannya dalam mengolah data berukuran besar secara efisien, serta menyajikan hasil dalam bentuk visual yang informatif. Sistem ini juga fleksibel untuk dikembangkan lebih lanjut dengan berbagai jenis data sosial lainnya.

Penelitian ini juga menggarisbawahi pentingnya pengolahan big data secara tepat, mulai dari proses pembersihan (*cleaning*), transformasi, hingga analisis. Data yang telah dibersihkan

dengan baik tidak hanya siap digunakan, tetapi juga meningkatkan keakuratan dalam proses analisis selanjutnya. Visualisasi yang dihasilkan dari data bersih mampu memberikan gambaran yang lebih jelas, relevan, dan dapat diandalkan untuk pengambilan keputusan. Sebaliknya, data yang masih kotor atau tidak konsisten berisiko menghasilkan visualisasi yang menyesatkan dan tidak mencerminkan kondisi sebenarnya. Oleh karena itu, proses pembersihan data menjadi tahap krusial dalam setiap analisis berbasis data besar, terutama saat digunakan untuk mengidentifikasi pola, tren, dan insight yang bernilai strategis.

5. SARAN

Implementasi Hadoop dalam penelitian ini menunjukkan adanya tantangan teknis, terutama bagi pengguna yang belum memiliki pemahaman mendalam terkait infrastruktur big data. Oleh karena itu, pengembangan ke depan dapat diarahkan pada pemanfaatan layanan cloud seperti Amazon EMR atau Google Dataproc yang mendukung framework seperti Hadoop, Spark, dan Hive secara lebih praktis dan terkelola. Selain itu, untuk meningkatkan akurasi analisis dan eksplorasi data yang lebih dalam, disarankan integrasi teknologi pembelajaran mesin lanjutan seperti *deep learning*. Model seperti LSTM atau Transformer dapat memberikan pemahaman yang lebih baik terhadap konteks bahasa dalam tweet, terutama pada teks informal yang kompleks. Dengan pendekatan ini, sistem dapat berkembang menjadi lebih cerdas dan adaptif dalam menangkap pola serta tren yang tersembunyi dalam data besar.

UCAPAN TERIMA KASIH

Terima kasih kepada semua pihak yang telah memberikan dukungan dalam penyusunan artikel ini, khususnya kepada rekan penulis yang telah berkontribusi besar dalam penelitian dan penulisan. Terima kasih juga saya sampaikan kepada teman-teman serta keluarga yang telah memberikan dukungan dan motivasi selama proses pengerjaan penelitian ini.

DAFTAR PUSTAKA

- [1] F. W. Umbara, "User Generated Content di Media Sosial Sebagai Strategi Promosi Bisnis," *J. Manaj. Strateg. dan Apl. Bisnis*, vol. 4, no. 2, pp. 572–581, 2021.
- [2] P. A. Aziz, S. B. N. Ilahi, S. Moka, and A. M. Sajiah, "Penerapan Hadoop untuk Analisis Sentimen Berbasis Big Data pada Ulasan Aplikasi Transportasi Online," *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 51–60, 2025.
- [3] I. B. K. Manuaba *et al.*, *Big Data: Teori dan Penerapan Teknologi Big Data dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2025.
- [4] J. Hutahaean and S. Amelia, "Penggunaan Teknologi Big Data Untuk Menganalisis Perilaku Belajar Siswa," *J. Teknol. Pendidik.*, vol. 3, no. 1, pp. 152–160, 2024.
- [5] P. Halawa, B. Bangun, and V. Sihombing, "Penggunaan Big Data untuk Prediksi Tren Pasar dalam Industri Retail," *J. Sist. Informasi, Tek. Inform. dan Teknol. Pendidik.*, vol. 4, no. 1, pp. 32–36, 2024.
- [6] S. R. Kamal, "Analisis big data pada aplikasi e-commerce dengan pendekatan human dan komputer interaction," *J. Literasi Inform.*, vol. 2, no. 4, 2023.
- [7] D. Hartama, "Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data," *Jurasik*

- (*Jurnal Ris. Sist. Inf. Dan Tek. Inform.*, vol. 3, pp. 46–55, 2018.
- [8] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, “Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook,” *J. Sist. Inf. dan Ilmu Komput.*, vol. 4, no. 2, pp. 23–27, 2021.
 - [9] A. Ahmad, “Analisis Visualisasi Data Bencana Alam Di Indragiri Hilir Menggunakan Tableau Public,” *TEKNOFILE J. Sist. Inf.*, vol. 2, no. 8, pp. 647–652, 2024.
 - [10] R. Al Ghivary, M. Mawar, N. Wulandari, and N. Srikandi, “Peran Visualisasi Data Untuk Menunjang Analisa Data Kependudukan Di Indonesia,” *PENTAHHELIX*, vol. 1, no. 1, pp. 57–62, 2023.
 - [11] R. T. Aldisa, P. Maulana, and M. A. Abdullah, “Penerapan Big Data Analytic Terhadap Strategi Pemasaran Job Portal di Indonesia dengan Karakteristik Big Data 5V,” *J. Sist. Komput. dan Inform. Hal*, vol. 267, p. 272, 2022.
 - [12] P. Raut, R. Rathod, R. Tidke, R. Pande, N. Rathod, and N. Kulkarni, “Sentiment Analysis of Twitter,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 12, pp. 621–627, 2022.
 - [13] D. Suparto and A. Habibullah, “Efektivitas penggunaan sosial media Twitter dalam penyebaran informasi dalam pelayanan publik (Studi kasus Kabupaten Pematang),” *Indones. Gov. J. Kaji. Polit.*, vol. 4, no. 2, 2021.
 - [14] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter,” *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 02, pp. 71–76, 2020.
 - [15] H. M. Putra, T. Akbar, A. Ahmadi, and M. I. Darmawan, “Analisa Performa Klastering Data Besar pada Hadoop,” *Infotek J. Inform. dan Teknol.*, vol. 4, no. 2, pp. 174–183, 2021.
 - [16] N. Hidayati and W. M. Baihaqi, “PENERAPAN DATA MINING DALAM MENGANALISIS POLA PEMBELIAN MENGGUNAKAN ALGORITMA APRIORI,” *J. Informatics Interact. Technol.*, vol. 1, no. 2, pp. 73–79, 2024.
 - [17] A. B. Trisnawan, “PEMANFAATAN DATA MINING DALAM MENENTUKAN POLA PEMILIHAN MATA KULIAH MAHASISWA,” *J. Adv. Res. Informatics*, vol. 3, no. 2, pp. 145–151, 2025.
 - [18] N. Puspitasari, U. Hairah, V. Z. Kamila, H. Hamdani, A. P. A. Masa, and R. P. Santi, “Classification of Crude Palm Oil Quality Using K-Nearest Neighbor (K-NN),” in *2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2024, pp. 675–681.
 - [19] F. Gurcan, O. Ozyurt, and N. E. Cagitay, “Investigation of emerging trends in the e-learning field using latent dirichlet allocation,” *Int. Rev. Res. Open Distrib. Learn.*, vol. 22, no. 2, pp. 1–18, 2021.
 - [20] R. Y. Rusdianto, “Pemanfaatan Visualisasi Data dalam Meningkatkan Pengambilan Keputusan Bisnis,” *J. Informasi, Sains dan Teknol.*, vol. 7, no. 2, pp. 200–208, 2024.
 - [21] P. Agusia, M. U. A. Manurung, V. Calista, and V. C. Mawardi, “Pemanfaatan Word Cloud

Pada Analisis Sentimen Dalam Menggali Persepsi Publik,” in *Seminar Nasional Penelitian (SEMNAS CORISINDO 2024)*, 2024, pp. 25–30.

- [22] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, “Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek dengan Support Vector Machine,” *J. Serina Sains, Tek. dan Kedokt.*, vol. 2, no. 1, pp. 61–70, 2024.