

ANALISIS DISTRIBUSI PROPORSI PENGGUNA PLATFORM SOSIAL MEDIA BERDASARKAN PERILAKU DAN PREFERENSI MENGGUNAKAN DATA MINING

Sihar Haryanto Pandiangan¹, Rafli Rafiqri², Jadiaman Parhusip³

¹Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

²Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

³Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

siharharyantopandiangan4@gmail.com, RafliRafiqri578@gmail.com,

parhusip.jadiaman@it.upr.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis distribusi dan proporsi pengguna platform sosial media berdasarkan perilaku dan preferensi mereka menggunakan teknik data mining. Dataset yang digunakan mencakup atribut seperti sentimen, aktivitas (likes dan retweets), waktu penggunaan, dan distribusi geografis. Metodologi ini mencakup preprocessing data, analisis distribusi, dan visualisasi menggunakan Python. Hasilnya menunjukkan preferensi platform yang bervariasi berdasarkan aktivitas pengguna, waktu, dan lokasi geografis, dengan Twitter dan Instagram sebagai platform dominan. Penelitian ini memberikan wawasan strategis bagi pengelola platform dan pengiklan dalam memahami perilaku pengguna.

Kata Kunci: Analisis Data, Preferensi Pengguna, Perilaku Pengguna, Platform Sosial Media.

I. PENDAHULUAN LATAR BELAKANG

Dalam beberapa tahun terakhir, platform sosial media seperti Facebook, Instagram, Twitter, dan TikTok telah mengalami perkembangan pesat dan menjadi bagian penting dalam kehidupan digital masyarakat. Layanan sosial media ini memfasilitasi interaksi antar pengguna serta menyediakan berbagai jenis konten yang dapat diakses oleh penggunanya secara fleksibel. Menurut laporan dari Statista, jumlah pengguna aktif di platform sosial media terus mengalami peningkatan setiap tahunnya, dengan penetrasi pasar yang semakin meluas. Hal ini mencerminkan peran sosial media dalam kehidupan sosial dan hiburan digital masyarakat [1].

Platform sosial media tidak hanya menyediakan konten yang beragam, tetapi juga mengumpulkan data dari aktivitas pengguna yang memungkinkan pemahaman lebih dalam terhadap perilaku dan preferensi mereka. Data yang diperoleh, seperti frekuensi penggunaan, jenis konten yang diakses, serta fitur-fitur yang digunakan, sangat penting untuk memahami pola pengguna. Data ini dapat dimanfaatkan untuk mengembangkan algoritma rekomendasi yang lebih personal dan relevan dengan kebutuhan pengguna. [2]

Namun, meskipun platform sosial media mengumpulkan sejumlah besar data, penelitian yang membahas distribusi proporsi pengguna berdasarkan berbagai aspek perilaku dan preferensi masih sangat terbatas. Sebagian besar penelitian yang ada lebih fokus pada analisis preferensi konten saja, tanpa memperhatikan faktor-faktor lain seperti perangkat yang digunakan atau pola interaksi dalam jangka waktu tertentu. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan menganalisis distribusi proporsi pengguna platform sosial media berdasarkan durasi penggunaan, kategori konten, dan perangkat yang digunakan, menggunakan teknik data mining untuk mengidentifikasi pola perilaku yang lebih kompleks dan tersembunyi dalam data besar [2].

PERTANYAAN PENELITIAN

1. Bagaimana distribusi pengguna pada berbagai platform sosial media?
2. Apa saja pola aktivitas pengguna berdasarkan perilaku (likes dan retweets)?
3. Bagaimana distribusi pengguna berdasarkan waktu?

2. METODE PENELITIAN

2.1 DATASET

Dataset adalah kumpulan data yang terstruktur dan terorganisir dalam format yang memudahkan analisis. Dataset biasanya disusun dalam bentuk tabel, di mana setiap baris mewakili unit data (seperti individu, transaksi, atau kejadian) dan setiap kolom menggambarkan atribut atau karakteristik data tersebut (seperti umur, jenis kelamin, lokasi, dan lain-lain). Dataset digunakan dalam berbagai bidang untuk menganalisis pola, membuat prediksi, atau mendapatkan wawasan terkait fenomena yang sedang diteliti [3].

Atribut yang terdapat dalam dataset yang digunakan meliputi:

1. Platform: Platform media sosial tempat pengguna melakukan interaksi (misalnya, Twitter, Instagram, Facebook).
2. Sentimen: Sentimen yang tercatat dari setiap interaksi pengguna (positif, netral, atau negatif).
3. Aktivitas: Jumlah interaksi pengguna dalam bentuk likes dan retweets.
4. Waktu: Waktu yang tercatat untuk setiap aktivitas yang dilakukan oleh pengguna (termasuk tahun, bulan, jam, menit) [4].

2.2 PYTHON

Python adalah bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido van Rossum pada akhir 1980-an dan dirilis pertama kali pada 1991. Bahasa ini dirancang dengan fokus utama pada kemudahan pembacaan dan penulisan kode. Python mendukung berbagai [4]paradigma pemrograman seperti berorientasi objek, fungsional, dan imperatif, sehingga memberikan fleksibilitas kepada pengembang. Salah satu keunggulannya adalah sintaks yang sederhana dan intuitif, yang memungkinkan pengembang untuk lebih fokus pada logika pemrograman tanpa harus bergantung pada detail teknis yang rumit [5].

Dalam penelitian ini, Python dipilih sebagai alat untuk melakukan analisis distribusi dan proporsi pengguna platform sosial media karena kemampuan bahasa ini dalam mengolah data besar dan melakukan analisis statistik secara efisien [5]. Python memungkinkan pengolahan dataset yang besar dengan mudah, serta memiliki berbagai pustaka yang dapat membantu dalam tahap preprocessing data, seperti pembersihan data dan transformasi format waktu. Pustaka seperti pandas digunakan untuk manipulasi data, sementara matplotlib dan seaborn digunakan untuk visualisasi yang memudahkan interpretasi hasil analisis [6]. Penggunaan Python dalam penelitian ini memungkinkan peneliti untuk mengidentifikasi pola-pola tersembunyi dalam perilaku pengguna, seperti pola aktivitas berdasarkan waktu dan platform yang digunakan, yang sangat penting untuk pengambilan keputusan berbasis data dalam konteks platform sosial media.

2.2.1 PREPROCESSING

Data preprocessing adalah teknik pengolahan data yang mencakup langkah-langkah yang perlu kita ikuti untuk mengubah atau mengkodekan data sehingga dapat dengan mudah diuraikan oleh mesin [7].

Preprocessing data berfungsi untuk mengubah data menjadi format yang lebih mudah dan efektif. Proses ini menjadi langkah awal yang penting dalam upaya data mining. Selain itu, teknik data preprocessing juga telah diadaptasi untuk melatih model machine learning. Pembersihan Data: Langkah pertama adalah membersihkan dataset dengan menghapus data duplikat dan menangani nilai kosong (missing values). Data yang hilang dapat disebabkan oleh ketidaksesuaian dalam pengumpulan data, sehingga perlu diisi atau dihapus sesuai dengan konteks analisis [7].

1. Transformasi Waktu: Format waktu dalam dataset awal tidak selalu dalam bentuk yang dapat dianalisis secara langsung. Oleh karena itu, waktu diubah menjadi atribut yang lebih relevan,

seperti tahun, bulan, dan jam. Hal ini mempermudah analisis berdasarkan waktu aktivitas pengguna.

2. Kategorisasi Aktivitas: Aktivitas pengguna (likes dan retweets) dikategorikan berdasarkan frekuensi aktivitas. Pengguna yang memiliki jumlah aktivitas tinggi dapat digolongkan ke dalam kategori "aktif", sementara pengguna dengan aktivitas rendah bisa digolongkan sebagai "pasif". Hal ini mempermudah analisis perilaku berdasarkan tingkat aktivitas.
3. Pemilihan Subset Data: Mengingat dataset yang digunakan sangat besar, dengan lebih dari ribuan entri, untuk tujuan penelitian ini diambil subset data yang terdiri dari sekitar ± 700 entri secara acak. Pemilihan ini bertujuan untuk menjaga keberagaman data, sekaligus membuat proses analisis lebih efisien tanpa kehilangan informasi yang penting. [4]

2.3 ANALISIS DATA

Analisis data adalah proses untuk mengubah data mentah menjadi informasi yang bermakna. Hal ini dilakukan dengan menggunakan metode kuantitatif atau kualitatif untuk memahami pola, hubungan, atau makna dalam data. Analisis ini mencakup pengorganisasian, deskripsi, dan interpretasi data untuk menjawab pertanyaan penelitian atau menguji hipotesis [8]

Metode yang digunakan:

1. *Analisis Distribusi Platform*: Menghitung jumlah pengguna di setiap platform.
2. *Analisis Aktivitas Berdasarkan Waktu*: Menghitung aktivitas pengguna per jam.
3. *Analisis Geografis*: Mengelompokkan pengguna berdasarkan negara.
4. *Visualisasi*: Menggunakan grafik batang dan stacked bar chart untuk distribusi data [6].

3. HASIL DAN PEMBAHASAN

3.1 Analisis Distribusi Platform

Tabel 3.1 Program distribusi sampel berdasarkan platform

```
import pandas as pd
import matplotlib.pyplot as plt

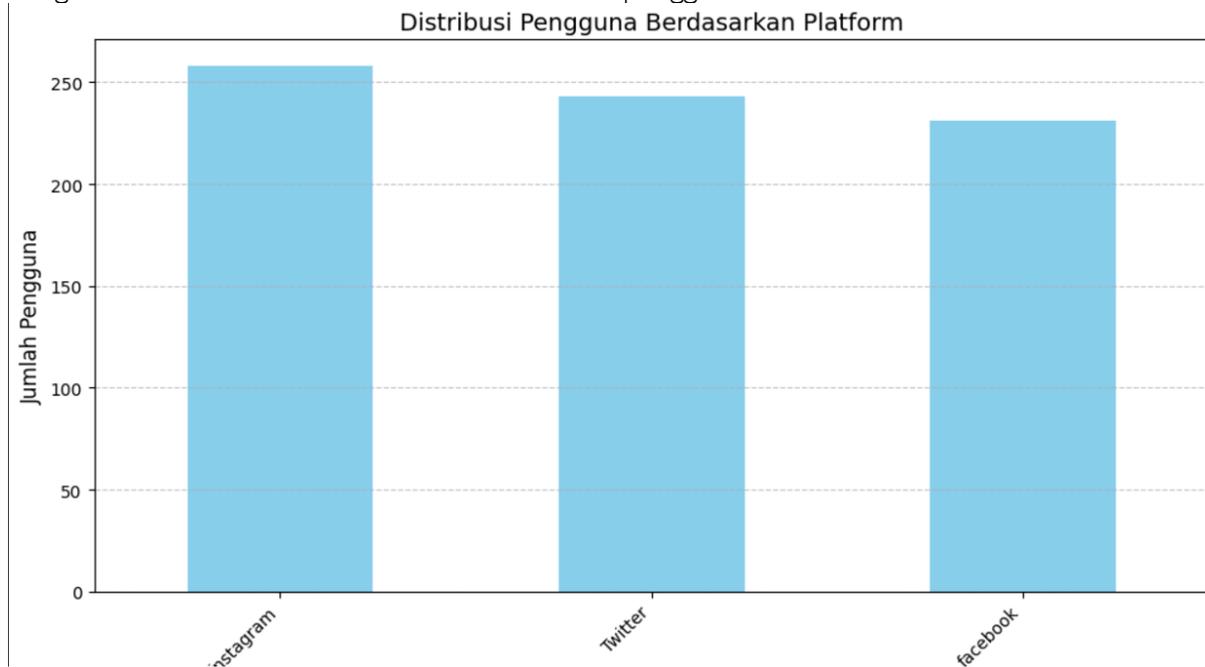
# Membaca dataset
df = pd.read_csv('dataset.csv')

# Menghitung distribusi pengguna berdasarkan platform
platform_distribution = df['platform'].value_counts()

# Menampilkan distribusi dalam bentuk grafik batang
plt.figure(figsize=(10,6))
platform_distribution.plot(kind='bar', color='skyblue')
plt.title('Distribusi Pengguna Berdasarkan Platform')
plt.xlabel('Platform')
plt.ylabel('Jumlah Pengguna')
plt.xticks(rotation=45)
plt.show() Hasil dan Pembahasan
```

Pada tahap ini, dilakukan analisis distribusi pengguna berdasarkan platform sosial media yang digunakan. Data yang digunakan berasal dari kolom platform dalam dataset yang mencatat aktivitas pengguna pada berbagai platform seperti Instagram, Twitter, dan Facebook. Dengan menggunakan

teknik pemrosesan data, dilakukan perhitungan frekuensi kemunculan setiap platform untuk mengetahui jumlah pengguna yang berinteraksi di masing-masing platform tersebut. Hasil dari analisis ini menunjukkan bahwa Instagram dan Twitter adalah platform yang paling dominan, mencerminkan tren global di mana kedua platform tersebut memiliki basis pengguna yang lebih besar dibandingkan dengan Facebook dalam konteks aktivitas interaksi pengguna.



Gambar 1. Visualisasi distribusi pengguna berdasarkan platform

Distribusi pengguna berdasarkan platform menunjukkan bahwa *Instagram* (258 pengguna) dan *Twitter* (243 pengguna) menjadi platform utama, diikuti oleh *Facebook* (231 pengguna). Hal ini mencerminkan popularitas platform tersebut dalam menjangkau audiens yang lebih luas.

3.2 Analisis Aktivitas Berdasarkan Waktu

Tabel 3.2 Program distribusi sampel berdasarkan jam

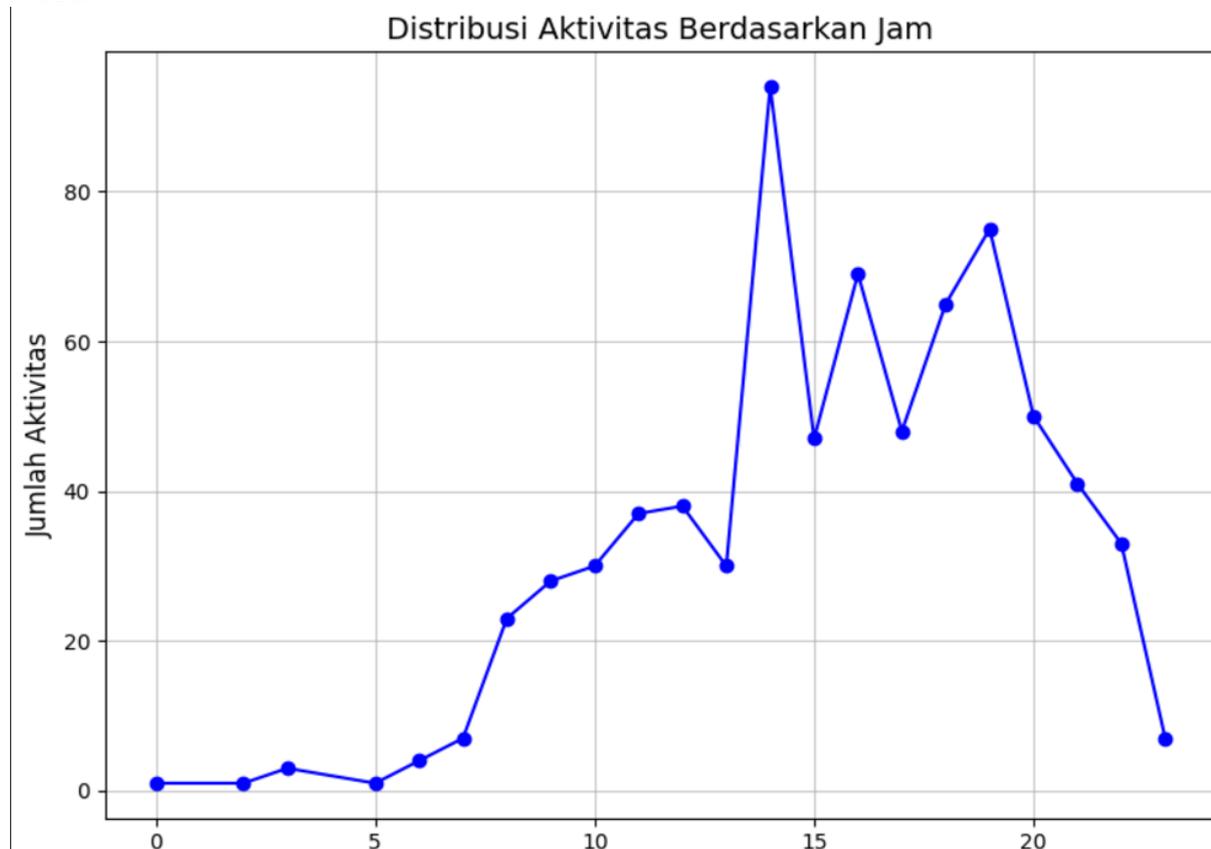
```
# Mengonversi kolom waktu ke format datetime
df['waktu'] = pd.to_datetime(df['waktu'])

# Menambahkan kolom jam untuk analisis berdasarkan waktu
df['jam'] = df['waktu'].dt.hour

# Menghitung distribusi aktivitas berdasarkan jam
activity_by_hour = df.groupby('jam').size()

# Menampilkan distribusi aktivitas dalam bentuk grafik batang
plt.figure(figsize=(10,6))
activity_by_hour.plot(kind='bar', color='salmon')
plt.title('Distribusi Aktivitas Pengguna Berdasarkan Waktu')
plt.xlabel('Jam')
plt.ylabel('Jumlah Aktivitas')
plt.xticks(rotation=0)
plt.show()
```

Dalam analisis ini, dilakukan pemrosesan waktu pada dataset dengan cara mengonversi kolom waktu menjadi format yang dapat dianalisis, seperti datetime untuk memperoleh informasi yang lebih relevan, seperti tahun, bulan, dan jam [6]. Selanjutnya, dilakukan pengelompokan data berdasarkan jam untuk menganalisis kapan pengguna paling aktif. Hasil analisis menunjukkan adanya puncak aktivitas yang signifikan pada jam-jam tertentu, seperti pukul 14.00, 16.00, dan 19.00. Pola ini mengindikasikan bahwa waktu-waktu tersebut mungkin bertepatan dengan waktu istirahat atau setelah jam kerja, ketika pengguna lebih cenderung untuk melakukan interaksi di platform sosial media.



Gambar 2. Visualisasi distribusi pengguna berdasarkan waktu

Distribusi aktivitas pengguna menunjukkan puncak aktivitas terjadi pada pukul 14.00 (94 aktivitas), diikuti pukul 19.00 (75 aktivitas) dan 16.00 (69 aktivitas). Pola ini menggambarkan waktu istirahat atau setelah jam kerja sebagai waktu paling aktif bagi pengguna.

3.3 Analisis Aktivitas (Likes dan Retweets)

Tabel 3.2 Program distribusi sampel berdasarkan perilaku

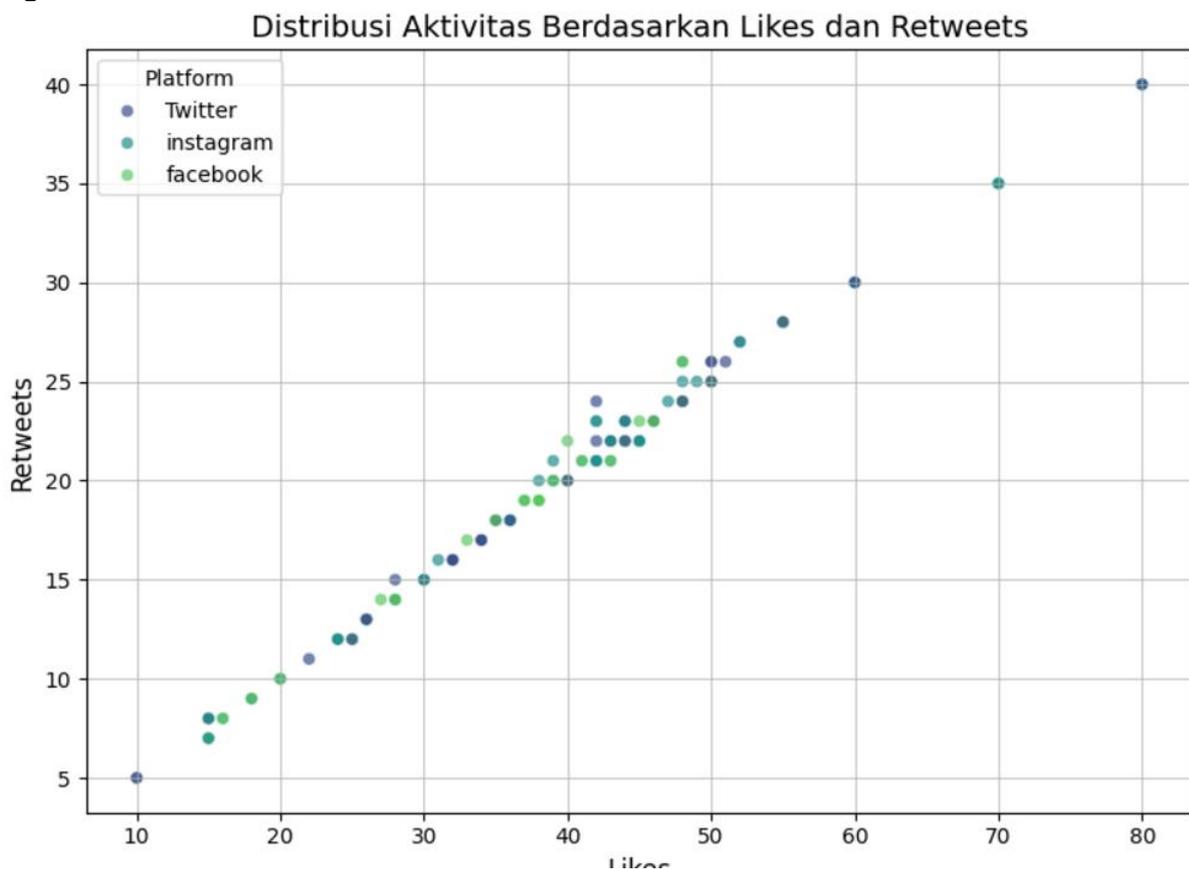
```
# Kategorisasi aktivitas pengguna (likes dan retweets)
df['aktivitas'] = df['likes'] + df['retweets']
df['kategori_aktivitas'] = df['aktivitas'].apply(lambda x: 'Aktif' if x > 100 else 'Pasif')

# Menghitung distribusi aktivitas berdasarkan kategori
activity_distribution = df['kategori_aktivitas'].value_counts()

# Menampilkan distribusi aktivitas dalam bentuk grafik batang
plt.figure(figsize=(8,5))
activity_distribution.plot(kind='bar', color='lightgreen')
```

```
plt.title('Distribusi Aktivitas Pengguna (Likes dan Retweets)')
plt.xlabel('Kategori Aktivitas')
plt.ylabel('Jumlah Pengguna')
plt.xticks(rotation=0)
plt.show()
```

Pada bagian ini, analisis fokus pada jumlah aktivitas pengguna yang diukur melalui jumlah likes dan retweets. Aktivitas pengguna dikelompokkan menjadi dua kategori: "aktif" dan "pasif", berdasarkan jumlah interaksi yang dilakukan. Pengguna dengan lebih dari 100 likes dan retweets digolongkan sebagai pengguna "aktif", sedangkan yang memiliki jumlah lebih rendah digolongkan sebagai "pasif". Hasil dari analisis ini menunjukkan bahwa platform Instagram dan Twitter memiliki proporsi pengguna yang lebih banyak dalam kategori "aktif", yang mengindikasikan bahwa kedua platform ini cenderung lebih efektif dalam memfasilitasi interaksi dan keterlibatan pengguna, dibandingkan dengan Facebook yang menunjukkan keterlibatan yang lebih rendah. Analisis ini memberikan gambaran tentang keterlibatan pengguna dengan platform sosial media berdasarkan tingkat interaksi mereka.



Gambar 3. Visualisasi distribusi pengguna berdasarkan perilaku

Pengguna yang memiliki lebih dari 100 likes sebagian besar aktif di Instagram dan Twitter. Preferensi ini menunjukkan bahwa platform tersebut lebih efektif untuk meningkatkan engagement dibandingkan Facebook.

4. PENUTUP

Kesimpulan

Penelitian ini berhasil mengidentifikasi perilaku dan preferensi pengguna di platform media sosial, dengan Instagram dan Twitter menjadi platform yang paling dominan dalam hal jumlah

pengguna. Waktu dan aktivitas menunjukkan pola tertentu dalam tingkat keterlibatan pengguna, dengan jam-jam tertentu yang memiliki interaksi lebih tinggi. Temuan ini memberikan wawasan penting tentang bagaimana pengguna berinteraksi dengan konten di media sosial.

Implikasi

Implikasi dari temuan ini adalah bahwa pengelola platform sosial media dan pengiklan dapat menggunakan wawasan ini untuk lebih menyesuaikan strategi konten dan kampanye mereka berdasarkan pola aktivitas dan preferensi geografis pengguna. Platform dapat meningkatkan interaksi dengan audiens dengan memanfaatkan jam-jam tertentu yang menunjukkan puncak aktivitas, serta menyesuaikan konten berdasarkan lokasi geografis pengguna.

Penelitian ini memberikan kontribusi terhadap pemahaman lebih dalam mengenai distribusi dan perilaku pengguna sosial media, serta menunjukkan pentingnya analisis data besar dalam merancang strategi digital yang lebih efektif.

Keterbatasan Penelitian

Dataset terbatas pada beberapa negara dan platform. Penelitian mendatang dapat mencakup platform lain seperti TikTok atau LinkedIn, serta menganalisis data demografis seperti usia dan gender.

5. DAFTAR PUSTAKA

- [1]. Statista. (2023). *Global social media usage*.
- [2]. Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- [3]. Reviewtekno(2023). Apa Itu Dataset. Reviewtekno. <https://www.reviewtekno.com/>
- [4]. Python.org. (2024). Python programming language. Python.org. <https://www.python.org/>
- [5]. Famili, A., Shen, W. M., Weber, R., & Simoudis, E. (1997). Data preprocessing and intelligent data analysis. *Intelligent data analysis*, 1(1), 3-23.
- [6]. Taherdoost, H. (2022). Different types of data analysis; data analysis methods and techniques in research projects. *International Journal of Academic Research in Management*, 9(1), 1-9.
- [7]. García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72, pp. 59-139). Cham, Switzerland: Springer International Publishing.
- [8]. Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, 3(1), 12-27.
- [9]. Quaranta, L., Calefato, F., & Lanubile, F. (2021, May). KGTorrent: A dataset of python jupyter notebooks from kaggle. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)* (pp. 550-554). IEEE.
- [10]. (2023). Social Media Sentiments Analysis Dataset [Dataset]. <https://www.kaggle.com/datasets/kashishparmar02/social-media-sentiments-analysis-dataset?select=sentimentdataset.csv>