

PEMANFAATAN *CROWDSOURCED* DATA TWITTER (X) UNTUK PEMETAAN SENTIMEN PENGGUNA TERHADAP DESTINASI WISATA DI DAERAH ISTIMEWA YOGYAKARTA TAHUN 2023

Nisa Amelia
21/483034/SV/20055

INTISARI

Pariwisata menjadi salah satu sektor yang mendukung pertumbuhan ekonomi di Daerah Istimewa Yogyakarta. Era *big data* yang semakin berkembang terus didukung oleh ketersediaan informasi dari media sosial, salah satunya X atau Twitter. Penelitian ini bertujuan (1) melakukan pengambilan dan storing data dari Twitter (X) terkait wisata di Daerah Istimewa Yogyakarta, (2) menganalisis sebaran dan pola sentimen terhadap destinasi wisata secara spasial di Daerah Istimewa Yogyakarta selama 2023, (3) menganalisis sentimen terhadap destinasi wisata secara temporal di Daerah Istimewa Yogyakarta berdasarkan sentimen pengguna Twitter (X) selama 2023, dan (4) membuat dashboard untuk visualisasi data hasil analisis.

Penambahan data dilakukan dengan pustaka tweet-harvest dan *pre-processing* secara otomatis menggunakan bahasa pemrograman Python. Selain itu, *geocoding* melalui persamaan dengan data Dinas Pariwisata berdasarkan destinasi wisata yang tercantum di dalam *posts* dilakukan untuk memberikan koordinat geografis. Analisis sentimen dilakukan dengan VADER atau *Valance Aware Dictionary and Sentiment Reasoner* berbasis leksikon, mempertimbangkan jumlah data dan kompleksivitas kata. Visualisasi spasial melalui *heatmap* dipilih untuk merepresentasikan pola dan sebaran melalui gradasi warna tanpa memperhatikan aspek statistik karena data sosial media yang rentan terjadi ketidaknormalan distribusi.

Hasil menunjukkan bahwa dari total 25.068 unggahan media sosial yang berhasil ditambah, sebanyak 2.685 data memenuhi kriteria untuk dianalisis lebih lanjut setelah melalui tahap pra-pemrosesan. Sebaran unggahan secara spasial cenderung padat di wilayah perkotaan dan destinasi wisata alam populer seperti kawasan pantai mencerminkan tingginya aktivitas digital di lokasi dengan fasilitas wisata yang baik. Analisis sentimen menghasilkan akurasi klasifikasi sebesar 80%, dengan distribusi sentimen netral dan positif yang mendominasi, khususnya di area wisata utama. Sentimen netral umumnya berupa informasi atau promosi, sedangkan sentimen positif menggambarkan kepuasan pengunjung. Walaupun hasil korelasi Spearman yang rendah, adanya keselarasan jumlah antara sentimen netral yang menunjukkan berita atau promosi wisata dan sentimen positif yang menunjukkan kepuasan terhadap destinasi wisata, mengindikasikan bahwa informasi mampu membentuk persepsi di kalangan masyarakat secara digital.

Kata kunci : *data mining*, X atau Twitter, pariwisata, analisis sentimen

MAPPING PUBLIC SENTIMENT ON TOURISM DESTINATIONS IN THE SPECIAL REGION OF YOGYAKARTA IN 2023 USING CROWDSOURCED TWITTER (NOW X) DATA

Nisa Amelia
21/483034/SV/20055

ABSTRACT

Tourism is one of the sectors that support economic growth in the Special Region of Yogyakarta. The growing era of big data continues to be supported by the availability of information from social media, one of which is X or Twitter. This research aims to (1) retrieve and store data from Twitter (X) related to tourism in the Special Region of Yogyakarta, (2) analyze the distribution and pattern of sentiment towards tourist destinations spatially in the Special Region of Yogyakarta during 2023, (3) analyze sentiment towards tourist destinations temporally in the Special Region of Yogyakarta based on the sentiment of Twitter (X) users during 2023, and (4) create a dashboard for visualizing the data analysis results.

Data mining is done with the tweet-harvest library and pre-processing automatically using the Python programming language. In addition, geocoding through equation with the Tourism Office data based on the tourist destinations listed in the posts was performed to provide geographic coordinates. Sentiment analysis was performed with a lexicon-based VADER or Valance Aware Dictionary and Sentiment Reasoner, considering the amount of data and word complexity. Spatial visualization through heatmaps was chosen to represent patterns and distribution through color gradations without considering statistical aspects because social media data is prone to distribution abnormalities.

The results show that from a total of 25,068 successfully mined social media posts, 2,685 data met the criteria for further analysis after going through a pre-processing stage. The spatial distribution of posts tends to be dense in urban areas and popular natural tourist destinations such as beach areas, reflecting the high digital activity in locations with good tourist facilities. Sentiment analysis resulted in a classification accuracy of 80%, with neutral and positive sentiment distributions dominating, especially in major tourist areas. Neutral sentiments are generally informational or promotional, while positive sentiments reflect visitor satisfaction. Despite the low Spearman correlation results, the alignment between neutral sentiments indicating news or promotion of tourism and positive sentiments indicating satisfaction with tourist destinations, indicates that information is able to shape perceptions among the public digitally.

Keywords : *data mining, X or Twitter, tourism, sentiment analysis*