Online media as a price monitor: Text analysis using text extraction technique and jaro-winkler similarity algorithm

Vivine Nurcahyawati[a], *Zuriani Mustaffa*[b]

[a] Department of Information System, Faculty Technology and Informatics Universitas Dinamika, Surabaya, Indonesia

[b] Department of Computer, Science Faculty Computing, Universiti Malaysia Pahang, Pahang, Malaysia

ABSTRACT

Online media has become an essential part of everyday life in modern society. Everyone or organization is free to share their opinions and feelings about any topic on it, including information or news about commodity price fluctuations. Commodity price data from the National Strategic Price Information Center (NSPIC) website is not real-time, so it is not sufficient as a basis for monitoring commodity price fluctuations. Meanwhile, the government needs to collect data and information quickly about these price fluctuations, hence immediately strategic decisions and policies can be made to stabilize the prices. This study explores the potential function of online media by extracting the text in it and analyzing text so that it can display the commodity price data sought. The commodities used as search keywords were commodities that had the highest consumption level in 2016 in Indonesia. The texts analyzed were taken from three online media, namely Twitter, Liputan6.com, and Detik.com. It was analyzed using text extraction techniques and the application of the Jaro-Winkler algorithm to find commodity prices in the text collection. Then compare the results of text analysis with commodity prices from the NSPIC website. The experimental data were 99,007 with a data collection time of three months. From only 122 data that match the keywords, it consists of 100 training data and 22 testing data. The results of the text analysis show that the text from the Detik.com website shows the commodity prices closest to the price data from the NSPIC, while Twitter shows the farthest results. The accuracy test with the confusion matrix is 75%. Based on this research, online media texts are a viable source for monitoring commodity price fluctuations.

KEYWORDS

Classification; Online Media; Similarity; Text Extraction; Text Mining; Text-Preprocessing; Visualization

REFERENCES

1. T.-Y. Liu, C. N. Scollon and W. Zhu, "Social Informatics", *7th International Conference SocInfo 2015*, 2015.

2. *Competitive Analysis Marketing Mix and Traffic*, 2020, [online] Available: https://www.alexa.com/siteinfo/kompas.com#section_audience.

3. *BI Belum Lihat Risiko dari Kenaikan Harga Pangan*, January 2019, [online] Available: https://www.cnnindonesia.com/ekonomi/20190118165750–532-361987/bi-belum-lihat-risiko-dari-kenaikan-harga-pangan.

4. *APJII: Jumlah Pengguna Internet di Indonesia Tembus 171 Juta Jiwa*, May 2019, [online] Available: https://tekno.kompas.com/read/2019/05/16/03260037/apjii-jumlah-pengguna-internet-di-indonesia-tembus-171-juta-jiwa.

5. I. E. Agbehadji, H. Yang, S. Fong and R. Millham, "The Comparative Analysis of Smith-Waterman Algorithm with Jaro[] Winkler Algorithm for the Detection of Duplicate Health Related Records", *IEEE*, 2018.