SENTIMENT CLASSIFICATION ON E-COMMERCE USER REVIEWS WITH NATURAL LANGUANGE PROCESSING (NLP) AND SUPPORT VECTOR MACHINE (SVM) METHODS

Jimmy Iqbal Wiranata Siregar[1], Andysah Putera Utama Siahaan[2], Muhammad Iqbal[3], Darmeli Nasution[4], Rian Farta Wijaya[5]
jimmywiranatasiregar@gmail.com[1], andiesiahaan@gmail.com[2], darmelinasution@gmail.com[4], rianfartawijaya@dosen.pancabudi.ac.id[5]
[1,2,3,4,5] Universitas Pembangunan Panca Budi, Indonesia

Article Info

ABSTRACT

This research aims to build a classification model that can categorize e-commerce user reviews into positive, negative, or neutral sentiments. By using NLP techniques to process the review text and SVM as a classification algorithm, it is expected that this model can provide high accuracy in determining user sentiment. Common words that do not contribute to sentiment analysis, such as "and," "which," "for", are removed, and SVM is applied after the review data is transformed into vectors using the TF-IDF method. The SVM model will be trained using training data that has been labeled with sentiment.

Corresponding Author:

Jimmy Iqbal Wiranata Siregar,
Magister Teknologi Informasi,
Universitas Pembangunan Panca Budi Medan, Indonesia
Alamat : Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing, 20122, Kota Medan, Provinsi Sumatera Utara, Indonesia.
Email : jimmywiranatasiregar@gmail.com

1. INTRODUCTION

Technological developments have changed the movement of customer behavior from purchasing goods directly to purchasing goods online or through e-commerce. This change affects the buyer's decision to buy a product, this is greatly influenced by consumer views on price, product, promotion, place (marketing mix) that has been implemented by the company. The more e-commerce that is developed, many users provide reviews of the e-commerce applications they have and from the analysis that has been done, most users who install e-commerce applications will provide reviews of the applications used. Reviews or ratings are important for new users who want to install e-commerce applications and determine which e-commerce applications will be used for transactions.

Rating is a part of the review that uses star symbol rather than text form in expressing the opinion of the user, rating can mean as an assessment of the user on the preference of an application to their

experience refers to the psychological and emotional state they go through when interacting with the application.

2.   RESEARCH METHOD

Data collection in this study uses web scraping techniques from Shopee e-commerce reviews. The research diagram illustrates the process from the initial level to the final level of research completion. In the early stages, researchers collect research materials from various library sources, including books, journals (both print and online), proceedings, articles, and other sources of information that are in line with the field of science. The proposed research diagram can be seen in Figure 3.1 below :
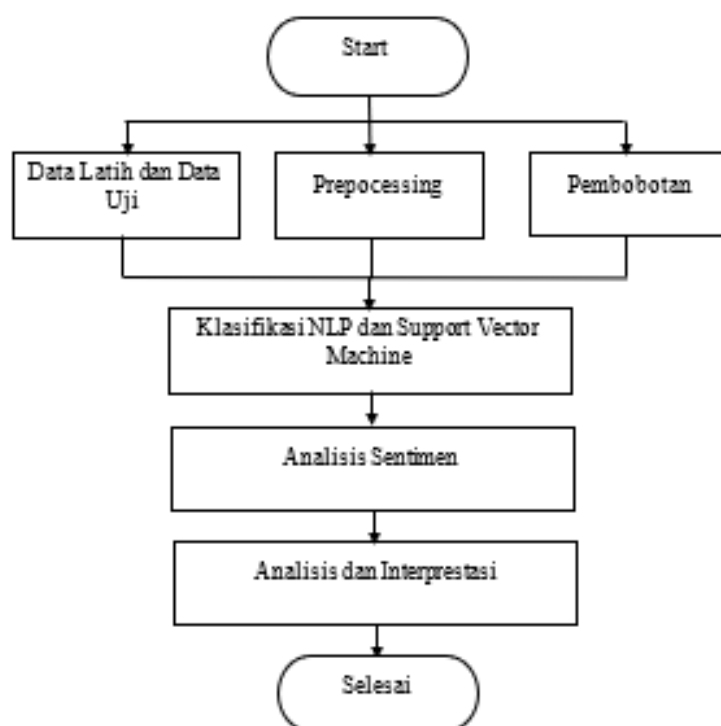

Gambar 1. research framework

2.1. Sentiment Analysis

Sentiment analysis is a technique for extracting information about opinions, sentiments and emotions from text, which can be documents, product reviews, tweets and so on. Sentiment and opinion are usually used interchangeably, they have a fundamental difference. Sentiment is a feeling or emotion expressed by someone towards something, such as positive, negative and neutral. Opinion is a view expressed by someone about something, which can be positive or negative.

2.2. Preprocessing using Neural Languange Processing (NLP) method
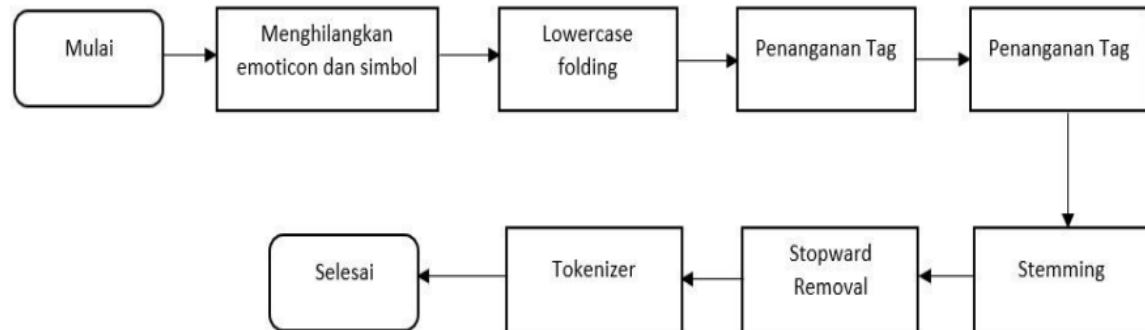
Gambar 2. Pre-processing stage using NLP approach

2.3. Support Vector Machine

Support Vector Machine (SVM) is one of the algorithms used to establish optimal boundaries between data classes, SVM is used to find the best hyperplane by optimizing the distance between data groups. hyperline equations are :

$f(x) = \boldsymbol{w} \cdot \boldsymbol{x} + b$, with :

$\boldsymbol{w}$ = weight parameter,

$\boldsymbol{x}$ = vektor input,

b = bias.

3.   RESULTS AND DISCUSSIONS

From research with a total of 5000 taken from shopee application reviews, consisting of five variables namely username, score, at, content, label. The compound score calculation process is carried out to display labels with positive, negative, neutral reviews, The compound score value with a limitation of < -0.05 shows the results of negative reviews with a value of 2.26% by 113 responses. For the calculation of compound score> 0.05 shows the results of positive reviews with a value of 15.02% by 751 responses, and for neutral reviews shows the results of 82.72% by 4136 responses.
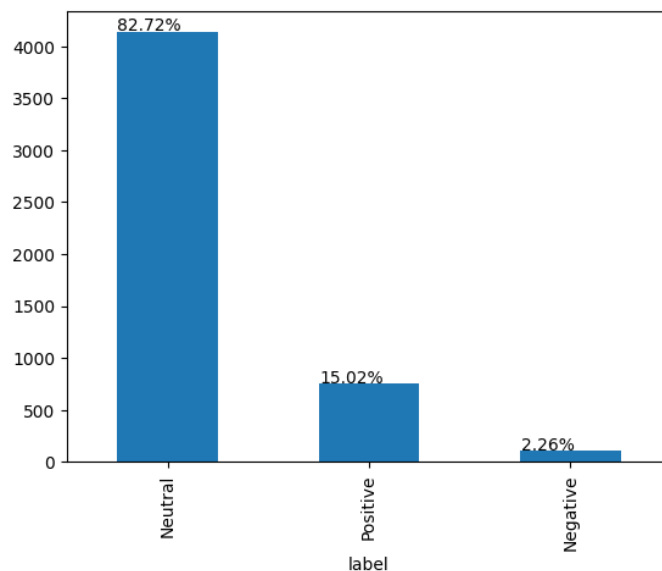
Fig. 3 Review label

For score assessment based on the results of label data processing shows the highest data with a score of 5 and neutral reviews of 3040 responses by 61%, for the least review results with a score of 2 and 3 on negative reviews
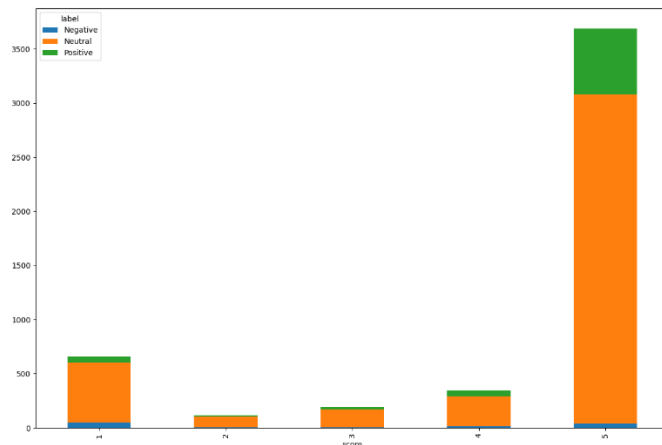

Figure 4. graph of review label and score

In the process of testing from 5000 data sets against the Support Vector Machine (SVM) model, randomly obtained accuracy results of 91.6%. with the amount of data classification from 1250 data sets showing the results of positive reviews of 15.84%, neutral reviews of 82.08% and negative reviews of 2.08%.

4.   CONCLUSION

The use of Neural Languange Processing (NLP) helps the review normalization process. From the classification results using the SVM model with a total of 1250 test data resulting in 26 negative reviews, 1026 neutral reviews, and 198 positive reviews. With an accuracy value of 91.6%

REFERENCES

[1]   I. Setiawan, S. Nirwan, and F. Maisya Amelia, "Rancangan Bangun Aplikasi Marketplace Bagi Usaha Mikro Kecil, Dan Menengah Berbasis Web (SUB MODUL : PEMBELIAN)," 2018.R. Arulmozhiyal and K. Baskaran, "Implementation of a Fuzzy PI Controller for Speed Control of Induction Motors Using FPGA," *Journal of Power Electronics,* vol. 10, pp. 65-71, 2010.

[2]   A. Alwendi, "Penerapan E-Commerce Dalam Meningkatkan Daya Saing Usaha," Manajemen Bisnis, vol. 17, no. 3, pp. 317–325, 2020, [Online]. Available: http://journal.undiknas.ac.id/index.php/magister-manajemen/

[3]   Y. S. Mulyani, "E-Commerce Solusi Pemasaran UMKM Dalam Mengembangkan Industri Pariwisata Di Tengah Pademi Covid-19 (Studi Kasus UMKM di Kota Tasikmalaya)," Khasanah Ilmu - Jurnal Pariwisata Dan Budaya, vol. 12, no. 2, pp. 131–141, 2021, doi: 10.31294/khi.v12i2.11293.

[4]   E. Y. Nasution, P. Hariani, L. S. Hasibuan, and W. Pradita, "Perkembangan Transaksi Bisnis E-Commerce terhadap Pertumbuhan Ekonomi di Indonesia," Jesya, vol. 3, no. 2, pp. 506–519, 2020, doi: 10.36778/jesya.v3i2.227.

[5]   J. Sintadi, "Analisis Pengaruh Brand Image, Harga, Dan Review Produk Terhadap Keputusan Pembelian Smartphone Samsung Pada Mahasiswa Undip Semarang," Diponegoro Journal of Management, vol. 8, no. 2, pp. 33–43, 2019, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/djom

[6]   N. Annisa, M. A. Nasution, and A. Firah, "Pengaruh Atribut Produk Dan Informasi Produk Terhadap Keputusan Pembelian Pada Pt . Trimitra Swadaya Medan the Effect of Product Attributes and Product Information on Purchase Decisions At Pt . Trimitra," no. 1, pp. 67–76, 2023.

[7]   Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," Journal of Information System Management (JOISM), vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.

[8]   N. Munasatya and S. Novianto, "Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron," Techno.Com, vol. 19, no. 3, pp. 237–244, 2020, doi: 10.33633/tc.v19i3.3630.

[9]   R. Yusuf, T. A. Saputri, and A. A. Wicaksono, "Penerapan Natural Language Processing Berbasis Virtual Assistant Pada Bagian Administrasi Akademik Stmik Dharma Wacana," International Research on Big-Data and Computer Technology: I-Robot, vol. 5, no. 1, pp. 33–47, 2022, doi: 10.53514/ir.v5i1.228.

[10] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," Jurnal Teknik Informatika, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.

[11] A. Wijaya, "Analisis Sentimen Kepuasan Pelanggan pada Penyedia Layanan Pengantaran Makanan," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 3, pp. 1699–1708, 2022, doi: 10.35957/jatisi.v9i3.2196.

[12] Fadhilah Dwi, Ananda, and Y. Pristyanto, "Analisis Sentimen Pengguna Twitter Terhadap Layanan Internet Provider Menggunakan Algoritma Support Vector Machine," MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput., vol. 20, no. 2, pp. 407–416, 2021, doi: 10.30812/matrik.v20i2.1130.

[13] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," J. Gaussian, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.

[14] M. Z. Farhan, "Analisis Sentimen Layanan Shopeefood Pada Twitter Dengan Metode K-Nearest Neighbor, Support Vector Machine, dan Decision Tree," J. Ilm. Inform., vol. 7, no. 2, pp. 95–106, 2023, doi: 10.35316/jimi.v7i2.95-106.

[15] K. A. Rokhman, Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," J. Inf. Syst. Manag., vol. 2, no. 2, 2021.

[16] D. L. Rianti, Y. Umaidah, and A. Voutama, "Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine," STRING (Satuan Tulisan Ris. dan Inov. Teknol., vol. 6, no. 1, p. 98, 2021, doi: 10.30998/string.v6i1.9993.