ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Sentiment Analysis on UNHCR Accepting Illegal Immigrants into Indonesia Using Naïve Bayes and SVM Algorithms on Instagram

Modesty Tumio Luciana Sianipar[1], Lucky Lhaura Van FC*[2]

[1,2] Informatics Engineering, Faculty of Computer Science, Universitas Lancang Kuning, Pekanbaru, Indonesia.

* Lucky Lhaura Van FC
Email: lucky@unilak.ac.id

Received 05/12/2024, Revised 29/12/2024, Accepted 29/12/2024, Published 31/12/2024

Abstract

Technology encompasses everything that can be used to convey messages, process data, or solve problems. Human resource development is a process of improving the quality, skills, and competencies of individuals to adapt to environmental changes and face future challenges. This has significantly increased the use of social media, which, in turn, has led to a surge in the number of comments or opinions from users. A concrete example is the issue of illegal immigrants entering Indonesia, which has resulted in UNHCR receiving both positive and negative comments from Instagram users. Sentiment analysis is a suitable technique for analyzing user opinions and sentiments on specific topics, including issues occurring in Indonesia. In this study, the Naïve Bayes and SVM methods were selected due to their effectiveness in text classification, especially for large text datasets, with satisfactory accuracy levels.

Keywords: Sentiment Analysis, Naïve Bayes, SVM, UNHCR, Illegal Immigrants, Reviews

Introduction

The use of technology by humans as a tool to complete tasks has become a necessity in daily life. However, this technological advancement must be accompanied by improvements in the quality of human resources. As technology users, humans must be able to utilize current technologies effectively while keeping up with future technological developments.

Instagram is a social media platform that serves as a fast and extensive information medium. Currently, Instagram is one of the most popular social media applications, used by nearly everyone. Many news organizations, journalists, and public figures use Instagram as a tool to disseminate news and provide the latest updates to the public. Instagram enables the dissemination of information both globally and locally. Users can stay updated on global developments while also accessing information about local communities.

Information on Instagram can be accessed by anyone with an account, making the platform open to various groups and offering high accessibility to information. With these characteristics, Instagram has become a dynamic medium for information and significantly impacts the rapid and efficient dissemination of information.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

UNHCR (United Nations High Commissioner for Refugees) is a UN agency that handles refugee affairs, serving as a primary instrument for countries facing refugee issues. As a result, there are many pro and con arguments simultaneously expressed by the Indonesian public.

Sentiment analysis is conducted to identify sentiments or emotions in a text and classify the characteristics of the text in documents or sentences. This process helps determine whether the sentiment expressed is positive, negative, or neutral.

Materials and Methods

This research implements Naïve Bayes and SVM to analyze sentiment with the following research model:

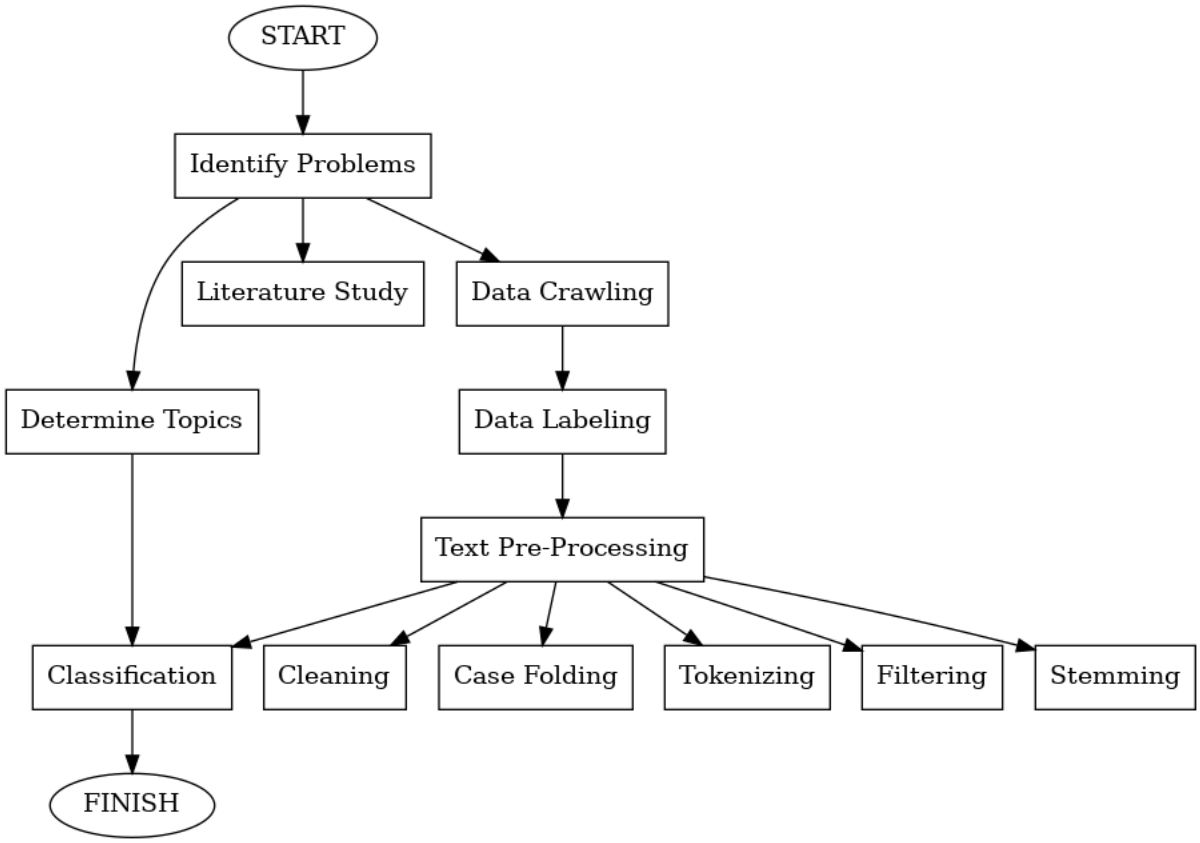

Fig 1. Research Method

Preparation Stage

In the preparation stage, the author performs several steps to prepare for the research, including formulating the problem statement, determining the background, defining the research title, and designing the research plan.


ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Planning Stage

In the planning stage, the author conducts preparations for literature review and data collection.

Final Stage

In the final stage, the researcher compares the sentiment analysis results using Naïve Bayes with those obtained using SVM.

Data Collection Techniques

Data Crawling

Data collection involves gathering reviews or comments from netizens about the research topic to support the study using data crawling techniques in Google Colab.

Pre-processing

Cleaning the data or reducing attributes that do not significantly contribute by performing the following steps:

- Cleaning
- Case folding
- Tokenizing
- Filtering
- Stemming

Naïve Bayes Algorithm

This step involves testing or predicting based on the collected data. The process involves gathering features related to the tested results and calculating how often each feature appears in different data groups. This produces a feature model that helps classify data into two categories: positive and negative.

Support Vector Machine

Similarly, this step involves testing or researching based on the collected data. Like Naïve Bayes classification, the process first involves gathering relevant features and calculating how often each feature appears in different data groups. This results in a feature model that helps classify data into two categories: positive and negative.

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Results and Discussion

Results

Crawling Data

From the code shown above, researchers obtained a dataset of 1,000 comments on UNHCR's Instagram posts about illegal immigrants.

Pre-processing

Cleaning

The cleaning process removes unnecessary characters, numbers, punctuation, emojis, and other irrelevant attributes from the dataset.

Table 3. Cleaning Results

Original Comment	After Cleaning
"Liat noh berita PEMERINTAH DENGAN TEGAS MENOLAK ROHINGYA..."	"Liat noh berita PEMERINTAH DENGAN TEGAS MENOLAK ROHINGYA..."
"Ingat sila ke 5 'keadilan sosial bagi seluruh rakyat indonesia'..."	"Ingat sila ke keadilan sosial bagi seluruh rakyat indonesia..."

Case Folding

This step converts all text in the dataset to lowercase.

Table 2. Case Folding Results

Cleaned Data	After Case Folding
"Liat noh berita PEMERINTAH DENGAN TEGAS MENOLAK ROHINGYA..."	"liat noh berita pemerintah dengan tegas menolak rohingya..."

Tokenizing

Splits sentences into individual tokens (words).

Table 3. Tokenizing Results

Case Folded Data	After Tokenizing
"liat noh berita pemerintah dengan tegas menolak rohingya..."	['liat', 'noh', 'berita', 'pemerintah', ...]

ComniTech: Journal of Computational Intelligence and Informatics
https://dx.doi.org/000000000
P-ISSN: 3063-315X E-ISSN: 3063-0630

Filtering

Removes unnecessary words that are not relevant for analysis.

Table 4. Filtering Results

Tokenized Data	After Filtering
['liat', 'noh', 'berita', 'pemerintah', ...]	['liat', 'noh', 'berita', 'pemerintah', 'tegas', 'menolak', ...]

Stemming

Converts words to their root forms.

Table 5. Stemming Results

Filtered Data	After Stemming
['liat', 'noh', 'berita', 'pemerintah', ...]	['liat', 'noh', 'berita', 'perintah', 'tegas', 'tolak', ...]

Data Labeling

After pre-processing, the dataset is labeled to identify positive and negative sentiments based on Indonesian positive and negative word dictionaries.

Table 6. Labeling Results

Positive Comments	Negative Comments
"Terima kasih UNHCR terima rohingya..."	"UNHCR beban negara UNHCR ga asuuuuuuw..."

Classification Report

Naïve Bayes Accuracy

Using 80% of the dataset for training and 20% for testing, Naïve Bayes achieved an accuracy of 89%, with the following metrics: Negative Precision: 99%, Positive Precision: 25%, Negative Recall: 89%, Positive Recall: 88%, F1 Score: Negative 94%, Positive 39%.

SVM Accuracy

Using the same dataset split, SVM achieved an accuracy of 98%, with the following metrics: Negative Precision: 98%, Positive Precision: 100%, Negative Recall: 100%, Positive Recall: 50%, F1 Score: Negative 99%, Positive 67%.

Discussion

This study on sentiment analysis of UNHCR's acceptance of illegal immigrants into Indonesia used data crawling techniques to collect 1,000 Instagram comments. The pre-processing involved cleaning, case folding, tokenizing, filtering, and stemming to prepare the data for analysis.

After labeling the dataset, the researchers found 946 negative comments and 41 positive comments. Using Naïve Bayes and SVM algorithms, the sentiment analysis showed that SVM achieved higher accuracy (98%) compared to Naïve Bayes (89%).

The findings indicate that the majority of comments about UNHCR's acceptance of illegal immigrants in Indonesia are negative, reflecting significant opposition among Instagram users.

Conclusion

This study successfully conducted sentiment analysis on UNHCR's acceptance of illegal immigrants into Indonesia using Naïve Bayes and SVM algorithms. A dataset of 1,000 comments from UNHCR's Instagram posts was collected through data crawling and processed using cleaning, case folding, tokenizing, filtering, and stemming techniques. The pre-processed data was labeled to categorize sentiments as positive or negative.

The results revealed that 946 comments expressed negative sentiment, while 41 comments were positive. Sentiment analysis using the Naïve Bayes algorithm achieved an accuracy of 89%, while the SVM algorithm achieved a higher accuracy of 98%. These findings indicate that SVM performed better in analyzing sentiment in this dataset.

The study concluded that the majority of Instagram users expressed negative opinions regarding UNHCR's decision to accept illegal immigrants into Indonesia. This highlights the public's critical stance on this issue, which could serve as valuable feedback for policymakers and organizations in handling refugee-related matters. Future studies may expand the scope by including larger datasets or exploring additional sentiment analysis methods for more comprehensive insights.

References

Afifah, N. F., & Voutama, A. (2023). Analisis Sentimen Isu Perselingkuhan pada Postingan Autobase Twitter @tanyarlfes Menggunakan Metode Naïve Bayes. Bianglala Informatika, 11(1), 32–37. https://doi.org/10.31294/bi.v11i1.16396

Alfarizi, S., & Fitriani, E. (2023). Analisis Sentimen Kendaraan Listrik Menggunakan Algoritma Naive Bayes dengan Seleksi Fitur Information Gain dan Particle Swarm Optimization. Indonesian Journal on Software Engineering (IJSE, 9)(1), 19–27.

Alita, D., & Shodiqin, R. A. (2023). Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine. Journal of Artificial Intelligence and Technology Information (JAITI, 1)(1), 1–12. https://doi.org/10.58602/jaiti.v1i1.20

Apri Wenando, F. (2023). Analisis Opini Publik Terhadap Undang-Undang KUHP Tahun 2022 Menggunakan Algoritma Naïve Bayes Classifier. Jurnal Fasilkom, 13(02), 334–339. https://doi.org/10.37859/jf.v13i02.5670

Arvyantomo, M., & Ratama, N. (2023). Analisis Sentimen Masyarakat Indonesia terhadap Invasi Russia di Ukraina menggunakan Metode Naive Bayes pada Media Sosial Facebook. LOGIC: Jurnal Ilmu Komputer, 1(4), 705–717.

Azura, R., & Munir, F. (2023). Peran United Nations High Commissioner for Refugees (UNHCR) dalam proses resettlement pengungsi asing di Indonesia (Studi Kasus Pengungsi Afghanistan di Kota Pekanbaru). Journal of Diplomacy and International Studies, 6(1), 51–64.

Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. Journal of Dinda: Data Science, Information Technology, and Data Analytics, 1(1), 42–51. https://doi.org/10.20895/dinda.v1i1.180

Febriyani, E., & Februariyanti, H. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes di Twitter. Jurnal TEKNO KOMPAK, 17(2), 25–38.

Gilbert, Syariful Alam, & Sulistyo, M. I. (2023). Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina Pada Google Playstore Menggunakan Metode Naïve Bayes. STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer, 2(3), 100–108. https://doi.org/10.55123/storage.v2i3.2333

Hardi, N., Alkahfi, Y., Handayani, P., Gata, W., & Firdaus, M. R. (2021). Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier. Sistemasi, 10(1), 131. https://doi.org/10.32520/stmsi.v10i1.1118

Hasibuan, F., Priatna, W., & Lestari, T. S. (2022). Analisis Sentimen Terhadap Kementrian Perdagangan Pada Sosial Media Twitter Menggunakan Metode Naïve Bayes. Techno.Com, 21(4), 741–752. https://doi.org/10.33633/tc.v21i4.6565

Ilmu, J., An, K., & Volume, N. (2024). Analisis Sentimen pada Twitter tentang Isu Perilaku Antisosial dengan Algoritma Naïve Bayes. Retika Nur Fadila, Andri Triyono, Dhika Malita P.A, 4(13–20).

Kaka, D. L., Pati, G. K., & Rato, K. W. (2023). Analisis Sentimen Komentar SIAKAD Menggunakan Metode Naive Bayes Classifier. Jurnal Kridatama Sains dan Teknologi, 5(02), 266–277. https://doi.org/10.53863/kst.v5i02.933

Kausar, A., Irawan, A., Wahyuddin, & Fernando, I. (2023). Implementasi Algoritma Naïve Bayes Classifier untuk Penilaian Kinerja Dosen. PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 10(2), 117–127. https://doi.org/10.30656/prosisko.v10i2.6922

Ketut Mediana Ayu Candrayani, I. M. A. D. S., & Wiranatha, A. A. K. A. C. (2023). Analisis Sentimen Pembelajaran Daring Era Pandemi COVID-19 Menggunakan Naive Bayes dan SVM. Tematik, 10(1), 47–53. https://doi.org/10.38204/tematik.v10i1.1274

Kusuma, A., & Nugroho, A. (2021). Analisa Sentimen pada Twitter terhadap Kenaikan Tarif Dasar Listrik dengan Metode Naïve Bayes. Jurnal Ilmiah Teknologi Informasi Asia, 15(2), 137. https://doi.org/10.32815/jitika.v15i2.557

Parintan Manurung, & Guntoro. (2024). Analisis Sentimen Layanan Aplikasi Kesehatan Pada Ulasan Play Store: Systematic Literature Review. SEMASTER: Seminar Nasional Teknologi Informasi & Ilmu Komputer, 3(1), 318-333. Retrieved from https://pustaka-psm.unilak.ac.id/index.php/Semaster/article/view/23938

Costaner, L., Guntoro, G., & Lisnawita, L. (2023). Analysis of Students to Follow Computer Extracurricular Activities using Fuzzy Logic Method. Sistemasi: Jurnal Sistem Informasi, 12(1), 244-256.