International Journal of Artificial Intelligence & Robotics (IJAIR)
Vol.7, No.1, 2025, pp.10-17
E - ISSN : 2686-6269

Sentiment Analysis of Social Media X Users on the Decline of Marriage Rates in Indonesia

Novi Amelia Kristanti[1], Wildan Mahmud[2], Galuh Wilujeng Saraswati[3], Erba Lutfina[4]
[1,2,3,4]Information Systems Department, Universitas Dian Nuswantoro, Kediri, Indonesia
[1]612202100025@mhs.dinus.ac.id*; [2]wildan.mahmud@dsn.dinus.ac.id, [3]galuhwilujengs@dsn.dinus.ac.id, [4]erba.lutfina@dsn.dinus.ac.id
*corresponding author

ABSTRACT

This study aims to analyze public sentiment regarding Indonesia's declining marriage rates and identify the most accurate algorithm for sentiment analysis. Data were collected from the social media platform X using crawling techniques, resulting in 1,082 tweets that were processed and classified into positive, negative, and neutral sentiments. The findings reveal that most sentiments are positive at 41.31%, negative at 30.59%, and neutral at 28.10%. The classification model evaluation shows that SVM outperforms Naïve Bayes, achieving an accuracy of 74% compared to 53%. This study is limited to data collected from single social media platform X. Future research is encouraged to expand the scope by collecting opinions from various social media platforms and exploring other machine learning or deep learning algorithms. The findings of this study are expected to contribute to policy-making efforts to improve marriage stability and well-being in Indonesia. This study also serves as a reference for academics and practitioners in understanding public opinion patterns on emerging social issues and providing a foundation for future studies on similar topics.

Keywords: Indonesian Marriage; Naïve Bayes; Sentiment Analysis; Support Vector Machine; Social Media X.

Article History

Received : xx-xx-xxxx	Revised : xx-xx-xxxx	Accepted : xx-xx-xxxx

I. INTRODUCTION

Marriage is a fundamental aspect of social structure that plays a crucial role in forming stable families and ensuring generational continuity. In population administration, marriage is a significant event that must be legally recognized and officially recorded to be considered valid under the law [1]. However, based on data from Indonesia's Central Statistics Agency *(Badan Pusat Statistik*, BPS), there has been a declining trend in the number of marriages over the past five years, with a decrease of 19.88%. In 2019, there were 1,968,978 recorded marriages, whereas in 2023, this number dropped to 1,577,255 [2]. This situation has attracted the attention of various stakeholders, including the government, academics, and society, as it may impact social stability and family well-being. The Minister of Population and Family Development emphasized that this trend is one of the key concerns in formulating population policies [4]. Therefore, understanding public opinion is essential in exploring societal perspectives on Indonesia's declining number of marriages.

Within Natural Language Processing (NLP), sentiment analysis involves identifying and labeling opinions based on their emotional tone, whether positive, negative, or neutral. [5]. Previous studies have examined sentiment analysis related to marriage, including an analysis of public sentiment regarding the decline in marriage rates in Indonesia using the TikTok dataset and the K-Nearest Neighbors (KNN) algorithm [6]. The study found that most TikTok users expressed positive opinions, with KNN achieving an accuracy of 85.83%. However, the study had limitations in data coverage, as it focused solely on a single social media platform.

This study introduces novelty in two key aspects. It utilizes X (formerly Twitter) as the data source for analyzing public sentiment regarding the decline in marriage rates. According to We Are X social media, it has 611.3 million global users, with Indonesia ranking fourth with 24.85 million users [7]. X platform provides real-time information more effectively than other platforms [8]. Previous studies have shown that Twitter is an effective platform for sentiment analysis in various sectors, such as education, analyzed sentiment using a Support Vector Machine (SVM), and confirmed Twitter's potential for extracting public opinions [9]. Therefore, social media X is suitable for monitoring public opinion trends on this phenomenon to gain a broader perspective beyond TikTok users. Sentiment analysis in this study uses the SVM algorithm because of its proven accuracy and robustness when dealing with non-normally distributed datasets. [10].

Additionally, this study compares SVM with Naïve Bayes (NB), a simple classification algorithm commonly used in sentiment analysis tasks [11]. In this study, NB is used as a standard for comparison to measure the accuracy of SVM. Therefore, this research seeks to analyze public sentiment regarding the decreasing marriage rates in Indonesia, whether positive, negative, or neutral, and to determine which algorithm yields the highest accuracy in sentiment classification. The results obtained from this study are

DOI: http://dx.doi.org/10.25139/ijair.v7i1.9953

intended to contribute to policy-making efforts to improve marriage stability and well-being in Indonesia. Moreover, the results can serve as a reference for academics and practitioners in understanding public opinion patterns on emerging social issues and provide a foundation for future studies on similar topics.

II. METHOD

This study employs a quantitative approach utilizing machine learning for sentiment analysis. This study utilizes data obtained via crawling from platform X, which is then analyzed using the SVM and NB methods. Figure 1 presents an overview of the research methodology.

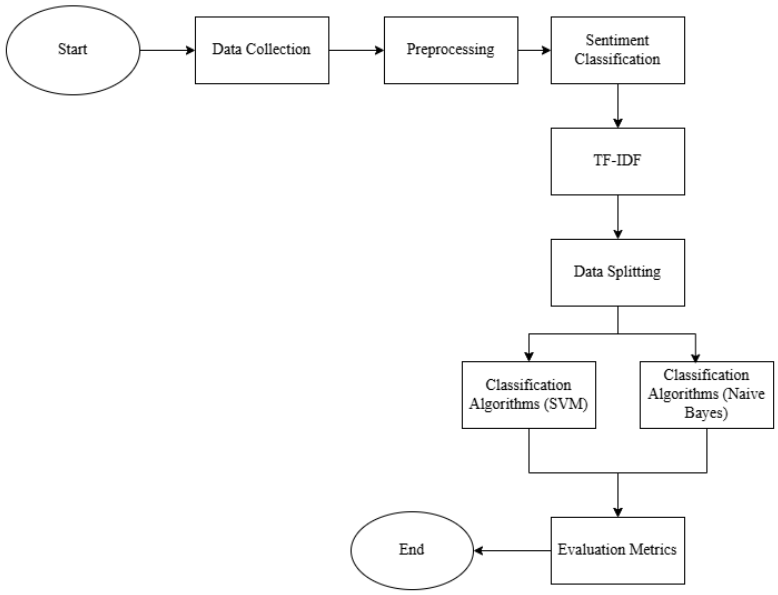

Figure 1. Research Workflow

A. Data Collection

Tweets were gathered using the Tweet Harvest tool, which enables automated data retrieval based on predefined keywords. The collected data was stored in *.csv format to facilitate further processing. Before analysis, the duplicate data was removed to ensure data integrity. All tweets collected were publicly available and obtained following the platform's terms of service. No private or sensitive user information was accessed during data collection. Identifiable information such as usernames and user IDs was removed during preprocessing to maintain user anonymity. Therefore, this study adheres to ethical standards commonly applied in social media-based research.

B. Preprocessing

The preprocessing phase in sentiment analysis is critical for ensuring the data is clean and well-organized, contributing to better model accuracy [12]. This process comprises several stages [13][14]. The first stage, case folding, standardizes the text by converting all characters to lowercase. The second stage, cleaning, removes unnecessary components such as URLs, emojis, and special characters that may disrupt the analysis. Next is normalization, where non-standard terms are replaced with their proper forms according to language rules. Tokenization follows, splitting the text into individual words or phrases. Then, stopword removal filters out commonly used words with minimal analytical value. Finally, stemming simplifies words to their root form to consolidate various word forms.

C. Sentiment Classification

TextBlob generates a polarity score that categorizes sentiment as positive, negative, or neutral [15]. Table I illustrates sample tweets with their corresponding sentiment classifications.

TABLE I
SENTIMENT CLASSIFICATION

Tweet	Classification
takut nikah tambah beban	*Negative*
nikah perempuan kubur mimpi	*Negative*
kampung desa angka nikah nikah kota iya turun angka nikah	*Neutral*
normal tingkat orang melek material banding kurang nikah hubung jangka beda an belakang lekat prinsip nikah ya bagus	*Positive*

D. TF-IDF

TF-IDF transforms text into numerical values by weighting words based on their frequency in a particular document compared to their frequency across the whole dataset. Words that occur frequently in one document but are rare across others receive higher weights, highlighting significant terms while downplaying common ones [16], as shown in Equation (1) [17].

$$W_{t,d} = tf_{t,d} \times idf_t \tag{1}$$

E. Data Splitting

The data was partitioned into training and testing sets at an 80:20 ratio through stratified sampling, ensuring the positive, negative, and neutral sentiment categories remained proportionally represented. This method was implemented to maintain class balance across both sets. A consistent random state (random state=42) was applied to ensure the data split could be reliably reproduced. This strategy helps ensure the models are trained on representative samples and assessed accurately. By applying stratified sampling, the risk of bias caused by class imbalance during training and testing is reduced.

F. Classification Algorithms

This research utilizes machine learning SVM and NB techniques to categorize public sentiment into positive, negative, and neutral classes. SVM is the primary classifier, while NB is a benchmark model for evaluating performance. This comparison aims to determine which algorithm offers the highest accuracy in classifying sentiments related to the decreasing marriage rate in Indonesia.

SVM works by identifying the optimal hyperplane that effectively separates data points from different classes in a high-dimensional space [17]. This study chose a linear kernel for SVM, as previous sentiment analysis studies have shown better performance than other kernels like polynomial, Radial Basis Function (RBF), and sigmoid. Research has demonstrated that the linear kernel can achieve up to 87% accuracy [19]. The model was configured with a regularization parameter of C = 1 and random_state = 42 for consistency and reproducibility. The SVM decision function is mathematically defined in Equation (2) [18], where $h(x)$ denotes the decision function, yi is the label of the i -th data point, αi represents the corresponding alpha value, $K\left(\vec{x_i} \cdot \vec{x_j}\right)$ is the kernel function, and b is the bias term.

$$h(x)=\sum_{i=1}^{m} \alpha_i \cdot y_i \cdot K\left(\vec{x_i} \cdot \vec{x_j}\right)+b \tag{2}$$

Conversely, Naïve Bayes is a probabilistic classification method widely used in text analysis because of its simplicity and speed [20]. The Multinomial Naïve Bayes variant is especially popular for text classification tasks involving discrete data such as word counts or term frequencies. This algorithm uses Bayes' theorem to calculate the probability of a class given a set of observed features, as shown in Equation (3) [21]. Where $P(C|X)$ represents the posterior probability of class *C* given features *X*, $P(X|C)$ is the likelihood $P(C)$ is the prior probability, and $P(X)$ s the marginal likelihood.

$$P(C|X)=\frac{P(X|C) \cdot P(C)}{P(X)} \tag{3}$$

G. Evaluation Metrics

Model evaluation is carried out to assess how well the classification algorithms perform in sentiment analysis. A commonly adopted approach for this is the Confusion Matrix, which summarizes correct and incorrect predictions. It is a standard tool for evaluating classification models by comparing the predicted labels against the actual labels from the test dataset [22]. The evaluation in this research relies on several key metrics [23]. Accuracy reflects the ratio of correctly predicted instances to total instances, as Equation (4) defines. Precision indicates the model's accuracy in predicting a specific class, as shown in Equation (5). Recall measures the model's ability to correctly identify all relevant positive cases, calculated using Equation (6). Finally, the F1-Score provides a balanced metric by computing the harmonic mean of precision and recall, as illustrated in Equation (7) [24].

$$Accuracy= \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision= \frac{TP}{TP + FP} \tag{5}$$

$$Recall= \frac{TP}{TP + FN} \tag{6}$$

$$F1\text{-}Score=2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

DOI: http://dx.doi.org/10.25139/ijair.v7i1.9953

H. Streamlit

To facilitate interactive sentiment analysis, this study develops a web-based application using Streamlit, a Python library that enables simple user interfaces for Machine Learning [25]. Users can input text, and the system will determine whether the sentiment is positive, negative, or neutral based on sentiment lexicon weighting. Figure 2 presents the Sentiment Analysis Application Interface developed in this study, which displays sentiment results based on user-inputted text.

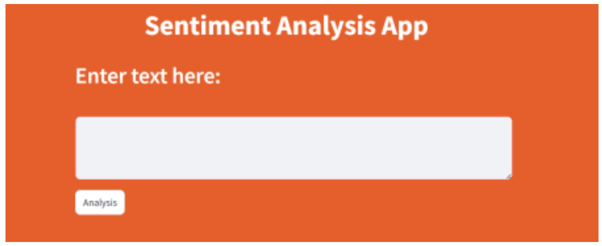

Figure 2. Sentiment Analysis Application Interface

III. RESULT AND DISCUSSION

The data was gathered from the social media platform X through a crawling method, using the keyword "decline in marriage rates in Indonesia," covering the period from January to November 2024. A total of 1,112 tweets were collected and saved in CSV format for subsequent analysis.

After removing duplicates, the number of tweets ready for analysis was reduced to 1,082. Table II presents examples of the collected data. The next step is to pre-process the tweets by converting the text to lowercase to maintain consistency and prevent discrepancies due to the use of capital letters, as shown in the case folding results in Table III. User tags such as '@strategi_bisnis' were then removed because they were irrelevant to sentiment analysis, and the results of this cleaning step can be seen in Table IV. Non-standard words were then normalized to correct Indonesian, such as "*gw*" to "*saya*",' "*uda*" to "*sudah*", and "*ortu*" to "*orang tua*", with the results in Table V. The tweet text was then tokenized by dividing the text into individual words, the results are presented in Table V. Common words (stop words) such as "*saya*," "*diri*," "*saja*," and "*sudah*", were then removed, the results in Table VII. Finally, stemming was applied using the Sastrawi library to reduce words to their basic forms, such as "*mengurus*" to "*urus*" and "*mikirnya*" to "*pikir*", the results in Table VIII.

TABLE II
DATASET

Tweet
@CNNIndonesia ya bagus lah
@CNNIndonesia Bekerja utk memenuhi hidup sendiri saja susah mau berumah tangga ????
@CNNIndonesia Padahal malah bagus.
@Strategi_Bisnis Generasi sekarang udh mulai bodo amat ama pernikahan lebih fokus ke pencapaian lain mau di ghibahin satu RW juga mereka pada cuek
@kekocakandunia Setuju kl nggak punya duit jangan nikah. Bener nanti tuhan kasih rejeki... Tapi ya dicari dulu rejeki tuhan baru nikah... Jangan kebalik ya adik adick!

TABLE III
CASE FOLDING

Tweet	Case Folding
@Strategi_Bisnis Gw salah satu nya diri sendiri aja uda ribet adek masih kuliah belum ortu kalo sakit belum buat hidup sehari2 gt kok mau nikah ngurus anak org + anak kita kalo ntar hamil bisa auto kriting rambut gw kali mikirnya mending single gini dah e	*@strategi_bisnis gw salah satu nya diri sendiri aja uda ribet adek masih kuliah belum ortu kalo sakit belum buat hidup sehari2 gt kok mau nikah ngurus anak org + anak kita kalo ntar hamil bisa auto kriting rambut gw kali mikirnya mending single gini dah e*

TABLE IV
CLEANING

Case Folding	Cleaning
@strategi_bisnis gw salah satu nya diri sendiri aja uda ribet adek masih kuliah belum ortu kalo sakit belum buat hidup sehari2 gt kok mau nikah ngurus anak org + anak kita kalo ntar hamil bisa auto kriting rambut gw kali mikirnya mending single gini dah e	*gw salah satu nya diri sendiri aja uda ribet adek masih kuliah belum ortu kalo sakit belum buat hidup sehari gt kok mau nikah ngurus anak org anak kita kalo ntar hamil bisa auto kriting rambut gw kali mikirnya mending single gini dah*

TABLE V
NORMALIZATION

Cleaning	Normalization
gw salah satu nya diri sendiri aja uda ribet adek masih kuliah belum ortu kalo sakit belum buat hidup sehari gt kok mau	*saya salah satu ya diri sendiri saja sudah ribet adek masih kuliah belum orang tua kalau sakit belum buat hidup sehari begitu kok mau*

DOI: http://dx.doi.org/10.25139/ijair.v7i1.9953

Cleaning	Normalization
nikah ngurus anak org anak kita kalo ntar hamil bisa auto kriting rambut gw kali mikirnya mending single gini dah	*nikah mengurus anak orang anak kita kalau nanti hamil bisa langsung keriting rambut saya kali mikirnya mending tunggal begini sudah*

TABLE VI
TOKENIZATION

Normalization	Tokenization
saya salah satu ya diri sendiri saja sudah ribet adek masih kuliah belum orang tua kalau sakit belum buat hidup sehari begitu kok mau nikah mengurus anak orang anak kita kalau nanti hamil bisa langsung keriting rambut saya kali mikirnya mending tunggal begini sudah	*['saya', 'salah', 'satu', 'ya', 'diri', 'sendiri', 'saja', 'sudah', 'ribet', 'adek', 'masih', 'kuliah', 'belum', 'orang', 'tua', 'kalau', 'sakit', 'belum', 'buat', 'hidup', 'sehari', 'begitu', 'kok', 'mau', 'nikah', 'mengurus', 'anak', 'orang', 'anak', 'kita', 'kalau', 'nanti', 'hamil', 'bisa', 'langsung', 'keriting', 'rambut', 'saya', 'kali', 'mikirnya', 'mending', 'tunggal', 'begini', 'sudah']*

TABLE VII
STOPWORD REMOVAL

Tokenization	Stopword Removal
['saya', 'salah', 'satu', 'ya', 'diri', 'sendiri', 'saja', 'sudah', 'ribet', 'adek', 'masih', 'kuliah', 'belum', 'orang', 'tua', 'kalau', 'sakit', 'belum', 'buat', 'hidup', 'sehari', 'begitu', 'kok', 'mau', 'nikah', 'mengurus', 'anak', 'orang', 'anak', 'kita', 'kalau', 'nanti', 'hamil', 'bisa', 'langsung', 'keriting', 'rambut', 'saya', 'kali', 'mikirnya', 'mending', 'tunggal', 'begini', 'sudah']	*['salah', 'ya', 'ribet', 'adek', 'kuliah', 'orang', 'tua', 'sakit', 'hidup', 'sehari', 'nikah', 'mengurus', 'anak', 'orang', 'anak', 'nanti', 'hamil', 'langsung', 'keriting', 'rambut', 'kali', 'mikirnya', 'mending', 'tunggal']*

TABLE VIII
STEMMING

Stopword Removal	Stemming
['salah', 'ya', 'ribet', 'adek', 'kuliah', 'orang', 'tua', 'sakit', 'hidup', 'sehari', 'nikah', 'mengurus', 'anak', 'orang', 'anak', 'nanti', 'hamil', 'langsung', 'keriting', 'rambut', 'kali', 'mikirnya', 'mending', 'tunggal']	*salah ya ribet adek kuliah orang tua sakit hidup hari nikah urus anak orang anak nanti hamil langsung keriting rambut kali pikir mending tunggal*

The processed tweets were then classified into three sentiment categories based on their polarity scores generated using the TextBlob library: positive, negative, and neutral. The classification results showed a variety of public opinions regarding the decline in marriage rates in Indonesia, with the detailed distribution illustrated in Figure 3.

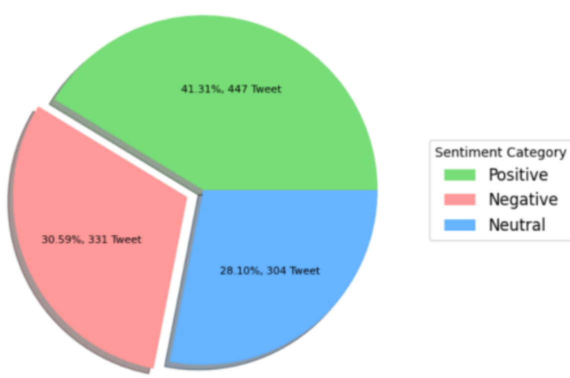

Figure 3. Sentiment Distribution

Most user sentiments on social media X were positive, accounting for 41.31%. Figure 4 displays the results of a word cloud analysis to examine the distribution of words more closely.

Figure 4. Word Cloud

 E - ISSN : 2686-6269

The word cloud analysis identifies key terms such as "*buru-buru",* "*nikah*", "*anak*", "*takut*", and "*salah*" as dominant, evidenced by their relatively larger font sizes compared to other words. The preprocessed textual data was then transformed into numeric features using the TF-IDF technique. TF-IDF assigns weights to words by evaluating their frequency within an individual document concerning their occurrence across the entire corpus. Terms that appear frequently in a single tweet but infrequently across the full dataset receive higher TF-IDF scores, thereby underscoring their significance in representing the expressed sentiment. The resulting data representation following the application of TF-IDF weighting is depicted in Figure 5.

```
(0, 881)      0.7713494042049016
(0, 2359)     0.6364118922779045
(1, 741)      0.35628477536839265
(1, 942)      0.23739949260881052
(1, 1693)     0.635129675813023
(1, 1270)     0.21440197884768405
(1, 2227)     0.40746974853153495
(1, 2249)     0.40746974853153495
(1, 2359)     0.18783379403320907
(2, 1194)     0.3575384218406826
(2, 1000)     0.43533423556394896
(2, 564)      0.48988123721912985
(2, 1429)     0.21070380651414486
(2, 826)      0.6310868875160778
(3, 881)      1.0
(4, 1116)     0.6512137678154827
(4, 1197)     0.3169404807453799
(4, 1472)     0.689542863259747
(5, 53)       0.1553719811872923
(5, 1290)     0.3065531755641855
(5, 447)      0.2774228663096625
(5, 653)      0.3065531755641855
(5, 632)      0.3065531755641855
```

Figure 5 TF-IDF

Following the TF-IDF transformation, the data was split into training and testing subsets, allocating 80% for training and 20% for testing. Table IX displays the breakdown of the training and testing datasets.

TABLE IX
DATA SPLITTING

Sentiment	Training Set	Testing Set
Positive	357	90
Negative	265	66
Neutral	243	61

Following the data splitting process, the next step involved training two machine learning models, SVM and NB, on the training set. Each model was trained separately using the TF-IDF feature vectors. After training, both models were evaluated on the testing set by calculating their accuracy, precision, recall, and F1-score. Figure 6 presents the comparison of accuracy achieved by the two algorithms.

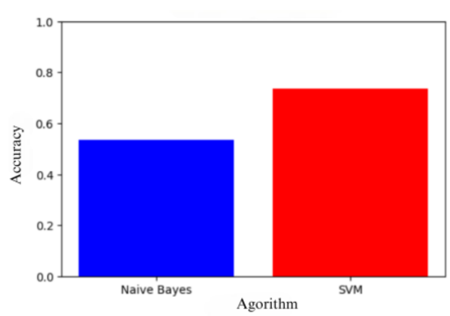

Figure 6. Accuracy Comparison

Evaluation findings demonstrate that the SVM model performed with 74% accuracy, surpassing the NB model's 53%. This disparity suggests that the SVM algorithm is more proficient at identifying patterns within tweet data related to Indonesia's declining marriage rates. Figure 7 compares the SVM and NB models' precision, recall, and F1-score metrics.

DOI: http://dx.doi.org/10.25139/ijair.v7i1.9953

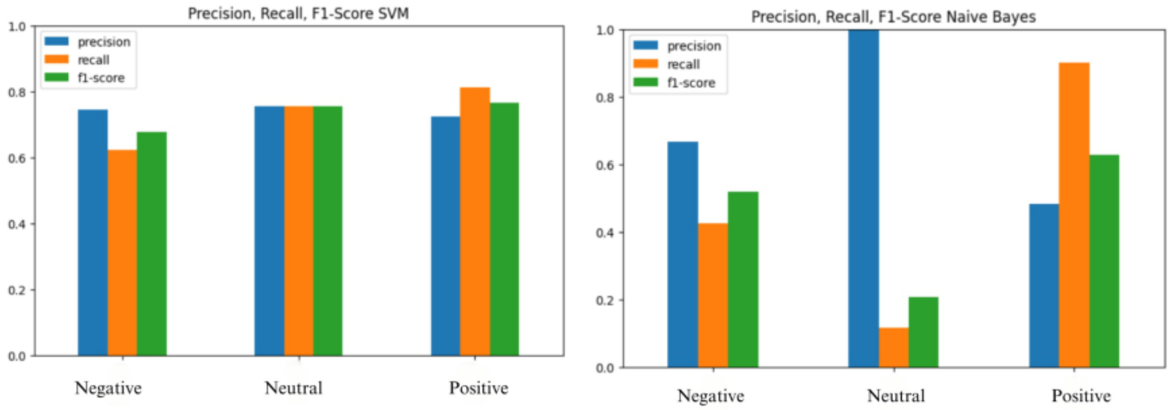

Figure 7. Precision, Recall, and F1-score Comparison

Based on the comparison results, the SVM model consistently outperformed NB across all sentiment categories. For the positive sentiment, SVM achieved a precision of 0.74, a recall of 0.83, and an F1-score of 0.78. In the negative category, SVM recorded a precision of 0.75, a recall of 0.62, and an F1-score of 0.68. Precision and recall reached 0.75 for the neutral class, resulting in an F1-score of 0.75.

In contrast, the NB model showed lower performance metrics. The positive class attained a precision of 0.48, a recall of 0.90, and an F1-score of 0.63. For the negative class, NB achieved a precision of 0.67, a recall of 0.42, and an F1-score of 0.52. Although NB demonstrated a perfect precision of 1.00 for the neutral class, its recall was notably low at 0.11, yielding an F1-score of merely 0.21.

The challenges in accurately classifying neutral sentiment can be attributed to the brevity of social media texts, which restricts contextual information. Additionally, the reduced effectiveness of NB in this study likely stems from its assumption of feature independence, which limits its ability to capture contextual dependencies between words within tweets. Consequently, NB tends to struggle with neutral sentiment classification and is less capable of modeling complex word interactions, leading to lower accuracy and F1-scores when compared to SVM.

After the model evaluation, the sentiment analysis application developed using Streamlit was successfully tested to analyze public opinion regarding Indonesia's declining marriage rates. This application can categorize tweets into three sentiment classes: positive, negative, and neutral, as illustrated in Figure 8.

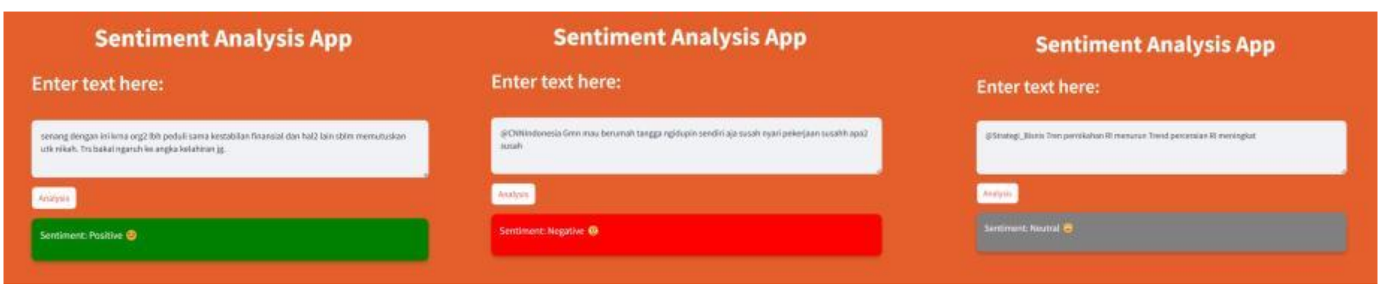

Figure 8. Streamlit Application Display for Positive, Negative, and Neutral Sentiments

IV. CONCLUSION

This study applied sentiment analysis using SVM as the primary method, with NB as a benchmark, to determine the highest classification accuracy in analyzing public sentiment on the decline of marriage rates in Indonesia using data from the X platform. The results showed that SVM achieved the highest accuracy of 74%, outperforming NB, which reached 53%. Sentiment analysis revealed that most public sentiment was positive, followed by negative and neutral sentiments.

Nonetheless, challenges persist in accurately classifying neutral sentiments due to the brevity and limited contextual information in social media posts and the independence assumption of NB, which reduces its ability to model complex word relationships. This study demonstrates that SVM is the most effective algorithm for achieving the highest classification accuracy in sentiment analysis related to social issues discussed on social media. Future research is encouraged to expand data collection from multiple platforms and explore advanced machine learning or deep learning approaches to enhance sentiment analysis performance.

REFERENCES

[1] I. S. Ningtias, "Faktor Yang Mempengaruhi Penurunan Angka Pernikahan Di Indonesia Indira Setia Ningtias Pertumbuhan penduduk yang tidak terkendali dapat menimbulkan berbagai permasalahan pada suatu negara jika tidak diikuti peningkatan kualitas hidup . Maria dalam penelit," vol. 4, no. 16, pp. 87–98, 2022.

[2] BPS, “Statistik Demografi Sosial,” bps.go.id. Accessed: Oct. 14, 2024. [Online]. Available: https://www.bps.go.id/id/statistics-table/3/VkhwVUszTXJPVmQ2ZFRKamNIZG9RMVo2VEdsbVVUMDkjMw==/nikah-dan-cerai-menurut-provinsi--2023.html?year=2023
[3] A. F. Adhani and A. Aripudin, “Perspektif Generasi Z di Platform X Terhadap Penurunan Angka Pernikahan di Indonesia.,” *J-KIs J. Komun. Islam*, vol. 5, no. 1, pp. 185–198, 2024.
[4] A. Ibrahim, “Menteri Wihaji Lapor Wapres Masalah Angka Pernikahan Ri Terus Menurun,” CNN. Accessed: Nov. 01, 2024. [Online]. Available: https://www.cnnindonesia.com/nasional/20241029153245-32-1160848/menteri-wihaji-lapor-wapres-masalah-angka-pernikahan-ri-terus-turun
[5] Wisnalmawati, A. S. Aribowo, and Y. Herawati, "Semi-supervised Learning Models for Sentiment Analysis on Marketplace Dataset," *Int. J. Artif. Intell. Robot.*, vol. 4, no. 2, pp. 78–85, 2022.
[6] A. A. Lailany and S. Lestari, “Analisis Sentimen Publik Terhadap Penurunan Jumlah Pernikahan di Indonesia menggunakan Algoritma K-Nearest,” vol. 5, no. 3, pp. 3043–3053, 2024.
[7] M. A. Rizaty, “Daftar Negara Pengguna X Terbanyak di Dunia per April 2024, Indonesia Keempat,” DataIndonesia.id. Accessed: Oct. 14, 2024. [Online]. Available: https://dataindonesia.id/internet/detail/daftar-negara-pengguna-x-terbanyak-di-dunia-per-april-2024-indonesia-keempat
[8] Kepios, "Global social media Statistics," Datereportal. Accessed: Oct. 19, 2024. [Online]. Available: https://datareportal.com/social-media-users
[9] G. Testiana, D. Erlina, P. Studi, S. Informasi, and S. Vectormachine, “Analisis Sentimen Pada Twitter Terhadap UIN Raden Fatah Menggunakan,” vol. 9, no. 1, pp. 698–708, 2022.
[10] I. S. Aisah, B. Irawan, and T. Suprapti, “Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur’an Digital,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3759–3765, 2024, doi: 10.36040/jati.v7i6.8263.
[11] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*. 2023.
[12] I. T. Julianto, “Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut,” *J. Algoritm.*, vol. 19, no. 1, pp. 449–456, 2022, doi: 10.33364/algoritma/v.19-1.1112.
[13] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm,” *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
[14] I. Verawati and B. S. Audit, “Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u,” *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1411, 2022, doi: 10.30865/mib.v6i3.4132.
[15] R. Azhar, A. Surahman, and C. Juliane, “Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes,” *J. Sains Komput. Inform. (J-SAKTI*, vol. 6, no. 1, pp. 267–281, 2022.
[16] H. Zakiyudin and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta Application of the Cosine Similarity Algorithm and Weighting of the TF-IDF System for New Student Admissions on Private Campuses," vol. 3, no. 1, pp. 19–27, doi: 10.30812/bite.v3i1.1110.
[17] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH,” *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
[18] R. Tineges, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
[19] T. M. P. Aulia, N. Arifin, and R. Mayasari, “Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19,” *SINTECH (Science Inf. Technol. J.*, vol. 4, no. 2, pp. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.
[20] P. Morgan, B. Show, and N. Media, “Analisis Sentimen Pengguna Youtube Terhadap Tayangan # Matanajwamenantiterawan Dengan Metode Naïve Bayes Sentiment Analysis Of User Comments On Youtube Video # Matanajwamenantiterawan With Naive Bayes Classifier Method,” vol. 5, no. 1, pp. 1–6, 2022, doi: 10.33387/jiko.
[21] R. A. D. Yunas, A. Triayudi, and I. D. Sholihati, “Implementasi Sistem Pakar untuk Mendeteksi Virus Covid-19 dengan Perbandingan Metode Naïve Bayes dan Certainty Factor,” vol. 5, no. 3, 2021.
[22] Zakaria, Kusrini, and D. Ariatmanto, "Sentiment Analysis to Measure Public Trust in the Government Due to the Increase in Fuel Prices Using Naïve Bayes and Support Vector Machine," *Int. J. Artif. Intell. Robot.*, vol. 5, no. 2, pp. 54–62, 2023.
[23] E. Suryati and A. A. Aldino, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” vol. 4, no. 1, pp. 96–106, 2023.
[24] A. Syahri, A. Angraini, and F. Muttakin, “Sentimen Analisis Pada Ulasan Aplikasi Ajaib Di Google Play Store Dengan Algoritma Support Vector Machine,” *INOVTEK Polbeng - Seri Inform.*, vol. 9, no. 1, pp. 257–271, 2024, doi: 10.35314/isi.v9i1.4047.
[25] R. S. Irawansyah, L. A. S. Irfan A, and G. W. Wiriasto, “Analisis Sentimen Terhadap Program Merdeka Belajar-Kampus Merdeka (Mbkm) Pada Twitter Menggunakan Algoritma Naive Bayes Classifier,” *J. Teknol. Informasi, Komputer, dan Apl. (JTIKA)*, vol. 5, no. 2, pp. 237–244, 2023, doi: 10.29303/jtika.v5i2.283.