

---

# Klasifikasi Penggunaan Kata *Urgent* Pada Percakapan *Whatsapp* Menggunakan Algoritma *Support Vector Machine*

Desman Frans N. Hulu<sup>1\*</sup>, Jatmika<sup>2</sup>, Yo'el Pieter Sumihar<sup>3</sup>

Program Studi Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel Yogyakarta  
Jln. Ukrim No. KM 11, Kadirojo I, Purwomartani, Kalasan, Sleman, Daerah Istimewa Yogyakarta 55571, Indonesia

E-mail: [desman.f2042@student.ukrimuniversity.ac.id](mailto:desman.f2042@student.ukrimuniversity.ac.id)<sup>1</sup>

---

## Abstrak

Penelitian ini bertujuan mengembangkan pendekatan otomatis untuk mengklasifikasikan pesan "urgent" dalam percakapan grup WhatsApp menggunakan algoritma Support Vector Machine (SVM) dalam bidang Pemrosesan Bahasa Alami (NLP). Data dari grup WhatsApp K-BIMI UKRIM dikumpulkan melalui proses crawling dan melalui tahapan preprocessing, seperti cleaning, case folding, normalization, tokenization, stopwords removal, dan stemming. Dataset dilabeli secara manual untuk menentukan pesan "urgent" dan "not urgent" untuk pelatihan dan pengujian model. Model klasifikasi SVM menunjukkan hasil yang efektif dengan nilai precision 0,99, recall 0,94, dan f-1 score 0,96 untuk sentimen "urgent", serta precision 0,99, recall 1,00, dan f-1 score 0,99 untuk sentimen "not urgent". Temuan ini menunjukkan bahwa SVM efektif dalam memprioritaskan pesan mendesak di grup komunitas WhatsApp.

---

## Abstract

This research aims to develop an automated approach to classify "urgent" messages in WhatsApp group conversations using the Support Vector Machine (SVM) algorithm in the field of Natural Language Processing (NLP). Data from the K-BIMI UKRIM WhatsApp group was collected through a crawling process and underwent preprocessing stages such as cleaning, case folding, normalization, tokenization, stopwords removal, and stemming. The dataset was manually labeled to identify "urgent" and "not urgent" messages for model training and testing. The SVM classification model demonstrated effective results with a precision of 0.99, recall of 0.94, and an f-1 score of 0.96 for "urgent" sentiments, and a precision of 0.99, recall of 1.00, and an f-1 score of 0.99 for "not urgent" sentiments. These findings indicate that SVM is effective in prioritizing urgent messages in WhatsApp community groups.

---

## Info Naskah:

Naskah masuk: 02 Juli 2024

Direvisi: 10 Juli 2024

Diterima: 12 Juli 2024

---

## Keywords:

Whatsapp;

Classification;

Support Vector Machine;

Urgent;

Natural Language Processing;

---

\*Penulis korespondensi:

Desman Frans Natalius Hulu

E-mail: [desman.f2042@student.ukrimuniversity.ac.id](mailto:desman.f2042@student.ukrimuniversity.ac.id)

## 1. Pendahuluan

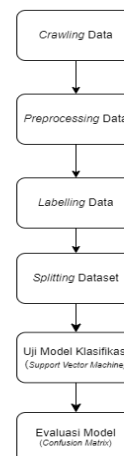
Dalam sebuah grup komunitas Whatsapp, pesan-pesan yang masuk mencakup beragam topik dan tingkat urgensi yang berbeda. Pentingnya mengidentifikasi pesan-pesan yang memiliki tingkat urgensi tinggi sangatlah krusial. Hal ini dapat memastikan bahwa tindakan cepat dapat diambil untuk menanggapi situasi darurat atau untuk menangani masalah yang membutuhkan perhatian segera. Namun, dalam grup yang besar dan aktif, menyaring pesan-pesan penting dari sekian banyaknya bisa menjadi tugas yang kompleks dan membutuhkan banyak waktu. Oleh karena itu, diperlukan suatu pendekatan yang otomatis dan efisien untuk mengklasifikasikan penggunaan kata yang bersifat "*urgent*" pada percakapan di grup chat whatsapp komunitas.

Teknologi Pemrosesan Bahasa Alami (*Natural Language Processing/NLP*) dengan Algoritma *Support Vector Machine* (SVM) memiliki potensi besar untuk menangani masalah diatas. Algoritma dengan *Machine Learning* ini banyak digunakan dalam NLP karena akurasi yang tinggi [1]. Dengan memanfaatkan kemampuan SVM untuk memahami dan menganalisis teks, kita dapat mengembangkan suatu model yang dapat mengidentifikasi pesan-pesan yang memuat kata "*urgent*" atau kata-kata terkait dengan tingkat urgensi. Kata "*urgent*" mengacu pada kata atau pesan yang menyampaikan kepentingan mendesak atau pentingnya suatu situasi atau tindakan yang segera dilakukan. Pesan yang memiliki tingkat urgensi tinggi biasanya mengandung kata-kata atau frasa seperti "segera", "penting", "darurat", atau ekspresi serupa.

Penelitian ini akan mengeksplorasi dan mengembangkan model NLP dengan Algoritma *Support Vector Machine*, yang dapat mengklasifikasikan penggunaan kata "*urgent*" pada chat whatsapp grup Komunitas Bidikmisi UKRIM (K-BIMI UKRIM). Dengan mengintegrasikan algoritma SVM ke dalam aplikasi Whatsapp, dapat meningkatkan responsivitas dan efektivitas dalam menanggapi situasi darurat dan memprioritaskan pesan-pesan yang memerlukan perhatian segera.

## 2. Metode Penelitian

Pada penelitian ini, algoritma yang digunakan adalah *Support Vector Machine* (SVM). Dataset yang digunakan merupakan data yang diperoleh dari chat Whatsapp, yang kemudian akan melalui tahapan *preprocessing*. Setelah itu, akan diuji dengan SVM dan dievaluasi dengan parameter pengujian. Gambar 3.1 dibawah ini merupakan skema daripada penelitian ini.



Gambar 2.1 Skema Penelitian

### 2.1 Crawling Data

Data diambil dari chat log grup WhatsApp K-BIMI UKRIM. Data yang digunakan merupakan data yang hasil dari fitur *export chat Whatsapp*. Dataset (*chat*) tersebut diperoleh dari rentang waktu antara 2020 s/d 2023. Dataset yang diambil sebanyak 10480 baris data, yang telah diubah kedalam format *.csv*.

### 2.2 Wordcloud

*WordCloud* adalah tampilan visual dari data teks di mana kata-kata yang paling sering muncul ditampilkan lebih besar dan unik dalam satu gambar. Biasanya, semakin sering banyak kata muncul dalam teks, maka semakin besar juga ukuran kata tersebut dalam *wordcloud*.

### 2.3 Preprocessing Data

*Preprocessing* data adalah tahapan untuk meningkatkan kualitas, struktur, dan relevansi informasi dari suatu teks. Dalam penelitian ini, tahapan *preprocessing* meliputi *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal*, dan *stemming data*.

- 1) *Cleaning* adalah suatu proses untuk mengidentifikasi dan memperbaiki kumpulan data dan tabel yang rusak atau tidak akurat [2].
- 2) *Case Folding* adalah prose untuk mengubah karakter alfabet yang telah dibersihkan menjadi huruf kecil (*lower case*) [3].
- 3) *Normalization* atau Normalisasi Data adalah proses mengubah sebuah atribut numerik yang telah di *scaling* ke dalam bentuk yang lebih sederhana, misalnya dari rentang 0 sampai 1 [4]. *Normalization* adalah proses untuk memperbaiki kata tidak baku menjadi kata yang baku.
- 4) *Tokenization* adalah proses mengurai urutan kata dalam sebuah kalimat, paragraf, atau halaman menjadi potongan kata tunggal yang berdiri sendiri [5].



| cleaning                                                  | case_folding                                       | normalization                                     |
|-----------------------------------------------------------|----------------------------------------------------|---------------------------------------------------|
| Selamat tahun baru                                        | selamat tahun baru                                 | selamat tahun baru                                |
| Selamat tahun baru                                        | selamat tahun baru                                 | selamat tahun baru                                |
| Happy new year                                            | happy new year                                     | happy new year                                    |
| Selamat tahun baru juga kak                               | selamat tahun baru juga kak                        | selamat tahun baru juga kakak                     |
| SELAMAT TAHUN BARU SEMUANYAAA                             | selamat tahun baru semuanyaaa                      | selamat tahun baru semuanyaaa                     |
| ...                                                       | ...                                                | ...                                               |
| Selamat Ulang Tahun Dan Panjang Umur sehat ...            | selamat ulang tahun dan panjang umur sehat ...     | selamat ulang tahun dan panjang umur sehat sel... |
| Selamat Ulang Tahun Dan sukses selalu panj...             | selamat ulang tahun dan sukses selalu panj...      | selamat ulang tahun dan sukses selalu panjang ... |
| Terima kasih semuaa teman-teman kbimi yayy                | terima kasih semuaa teman-teman kbimi yayy         | terima kasih semua teman kbimi yayy               |
| Happy birthday juga buatmu                                | happy birthday juga buatmu                         | happy birthday juga buatmu                        |
| Perpanjangan Waktu KRS online Mohon bantu sha...          | perpanjangan waktu krs online mohon bantu sha...   | perpanjangan waktu krs online mohon bantu shar... |
| tokenize                                                  | stopword removal                                   | stening_data                                      |
| [selamat, tahun, baru]                                    | [selamat]                                          | selamat                                           |
| [selamat, tahun, baru]                                    | [selamat]                                          | selamat                                           |
| [happy, new, year]                                        | [happy, new, year]                                 | happy new year                                    |
| [selamat, tahun, baru, juga, kakak]                       | [selamat, kakak]                                   | selamat kakak                                     |
| [selamat, tahun, baru, semuanyaaa]                        | [selamat, semuanyaaa]                              | selamat semuanyaaa                                |
| ...                                                       | ...                                                | ...                                               |
| [selamat, ulang, tahun, dan, panjang, umur, se...]        | [selamat, ulang, umur, sehat, selalugbu]           | selamat ulang umur sehat selalugbu                |
| [selamat, ulang, tahun, dan, sukses, umur, p...]          | [selamat, ulang, sukses, umur, tym]                | selamat ulang sukses umur tym                     |
| [terima, kasih, semua, teman, kbimi, yayy]                | [terima, kasih, teman, kbimi, yayy]                | terima kasih teman kbimi yayy                     |
| [happy, birthday, juga, buatmu]                           | [happy, birthday, buatmu]                          | happy birthday buat                               |
| [perpanjangan, waktu, krs, online, mohon, bantu, shar...] | [perpanjangan, krs, online, mohon, bantu, shar...] | panjang krs online mohon bantu share group prodi  |

Gambar 3.2 Hasil *Preprocessing*

### 3.3 Wordcloud

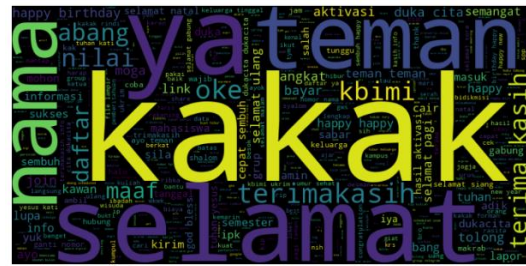
#### 1) Wordcloud sebelum *Preprocessing*

Visualisasi *wordcloud* sebelum *preprocessing* menunjukkan kata-kata yang jumlahnya lebih banyak ditandai dengan ukuran yang lebih besar, seperti "kak," "ya," "dan," "di," "pak," dan "yang." Sehingga kata-kata yang ada dalam data belum dapat diidentifikasi dengan tepat. Hasil daripada visualisasi *wordcloud* sebelum *preprocessing* ditunjukkan pada Gambar 3.3 berikut.

Gambar 3.3 Wordcloud Sebelum *Preprocessing*

#### 2) Wordcloud setelah *preprocessing*

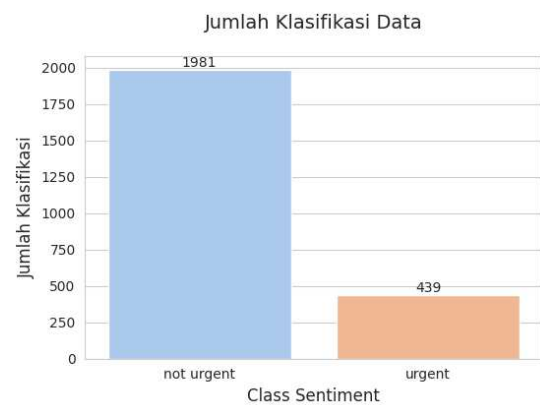
Visualisasi *wordcloud* setelah *preprocessing* menunjukkan bahwa beberapa unsur yang tidak diinginkan telah dihapus dan diperbaiki. Kata kakak, selamat, ya, teman, nama, terimakasih menjadi kata-kata yang frekuensi jumlahnya paling banyak. sementara untuk kata-kata yang bersifat *urgent* memiliki frekuensi jumlah yang cukup, misalnya kata tolong, ayo, cepat, dan lain sebagainya. Hasil *wordcloud* setelah *preprocessing* ditunjukkan pada Gambar 3.4 berikut.

Gambar 3.4 Wordcloud Setelah *Preprocessing*

### 3.4 Implementasi *Support Vector Machine*

#### 1) Pelabelan Data

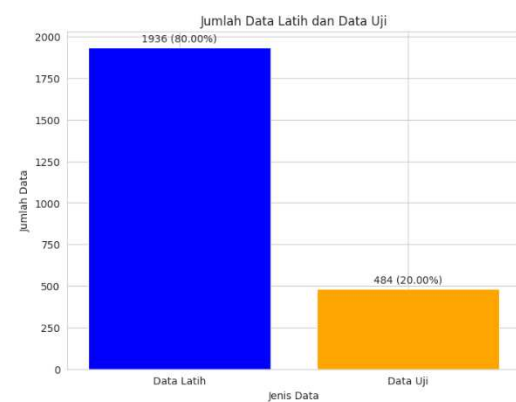
Hasil daripada pelabelan data memperlihatkan bahwa jumlah data kelas *not urgent* adalah sebanyak 1981 data, sedangkan kelas *urgent* sebanyak 439 data. Hal ini menunjukkan bahwa dari total 2420 data, kelas *urgent* memiliki jumlah yang sedikit dibanding kelas *not urgent*. Hasil daripada distribusi data dapat dilihat pada Gambar 3.5 dibawah ini.



Gambar 3.5 Distribusi Hasil Pelabelan Data

#### 2) Data Latih dan Data Uji

Data latih yang dihasilkan sebanyak 1936 data (80%), sedangkan data uji sebanyak 484 data (20%). Gambar 3.6 berikut adalah *barplot* distribusi jumlah data latih dan data uji.

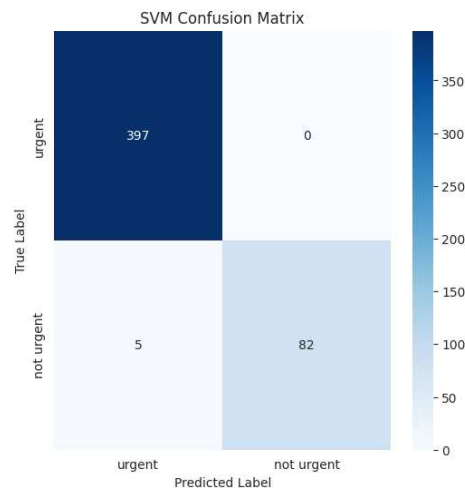


Gambar 3.6 Distribusi Jumlah Data Latih dan Data Uji

#### 3) *Confusion Matrix*

Setelah melalui tahapan pengujian, model akan dievaluasi dengan pengujian SVM sehingga

menghasilkan hasil *confusion matrix* seperti yang terlihat pada Gambar 3.7 dibawah ini.



Gambar 3.7 Hasil Visualisasi *Confusion Matrix* SVM

Deskripsi hasil prediksi untuk setiap kategori *Confusion Matrix* diatas adalah sebagai berikut :

- Kelas *Urgent*

Jumlah data kategori sentimen *urgent* yang diinput ke model sebanyak 397 data. Hasil prediksi menunjukkan bahwa seluruh data (397) diidentifikasi benar sebagai label *urgent*. Sementara itu, tidak ada yang diprediksi sebagai *not urgent*.

- Kelas *Not urgent*

Jumlah data kategori sentimen *not urgent* yang diinput ke model sebanyak 87 data. Hasil prediksi menunjukkan sebanyak 82 data diidentifikasi benar sebagai label *not urgent*. Sementara, sisanya diprediksi sebanyak 5 data sebagai *urgent*.

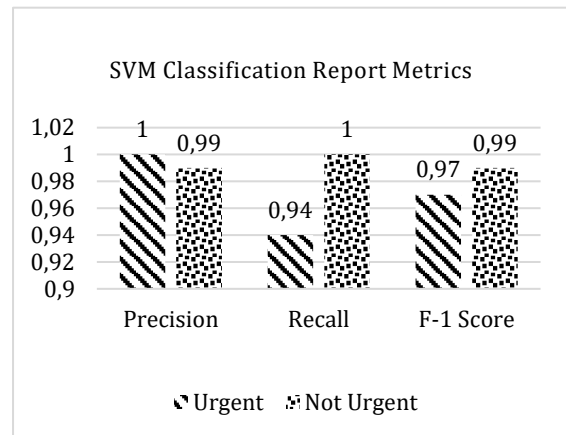
### 3.5 Evaluasi Model

Dari hasil pengujian kinerja model, *precision*, *recall*, dan *f-1 score* memiliki nilai yang besar untuk kedua kelas tersebut. Nilai *precision* untuk kelas "*urgent*" mencapai 100%, *recall* 94%, dan *f-1 score* 97% yang menunjukkan bahwa dari 397 data yang dianggap sebagai "*urgent*", hampir semua data tersebut benar-benar termasuk kelas "*urgent*". Selain itu, model juga menunjukkan kinerja yang sangat bagus dalam melakukan klasifikasi pada kelas "*not urgent*", dengan nilai *precision* mencapai 99%, nilai *recall* mencapai 100%, dan nilai *f-1 score* mencapai 99%. Hal ini juga menunjukkan bahwa dari 87 data pengujian yang sebenarnya termasuk kelas "*not urgent*", hampir hanya 0,01 (1 data) dari semua data tersebut tidak dikenali oleh model. Dengan demikian, model klasifikasi yang diterapkan dalam penelitian ini memiliki performa yang lebih baik untuk mengklasifikasi kelas *not urgent* dibandingkan dengan kelas *urgent*. Distribusi hasil evaluasi kinerja model ditampilkan pada Tabel 3. 1 dibawah ini.

Tabel 3.1 Distribusi Hasil Evaluasi Kinerja Model

| Kelas             | <i>Precision</i> | <i>Recall</i> | <i>F-1 Score</i> |
|-------------------|------------------|---------------|------------------|
| <i>Urgent</i>     | 1,00             | 0,94          | 0,97             |
| <i>Not Urgent</i> | 0,99             | 1,00          | 0,99             |

Hasil diatas juga dapat ditampilkan kedalam *barplot SVM Classification Report Metrics*, yang ditunjukkan pada Gambar 3.8 berikut.



Gambar 3.8 SVM Classification Report Metrics

## 4. Kesimpulan

Berdasarkan hasil dan pembahasan diatas, maka dapat disimpulkan bahwa pelabelan data pada penelitian ini menghasilkan sebanyak 1981 data (82%) bernilai *not urgent*, dan 439 data (18%) bernilai *urgent*. Hasil ini menunjukkan bahwa data dengan nilai sentimen "*urgent*" berbanding jauh dengan hasil data dengan nilai sentimen "*not urgent*". Sehingga, data hanya sedikit memuat kata kata yang bermaksud *urgent*, bersifat "*urgent*", atau bersifat mendesak. Selain itu, hasil pengujian terhadap model menunjukkan bahwa kelas *urgent* pada parameter pengujian *precision* memiliki nilai yang 1,00 dibandingkan dengan kelas *not urgent* yang memiliki nilai 0,99. Selanjutnya, pada parameter pengujian *recall*, kelas *not urgent* memperoleh nilai 1,00 dan lebih besar dibandingkan dengan kelas *urgent* yang memiliki nilai, yakni 0,94. Sementara itu, pada parameter pengujian *f-1 score*, kelas *not urgent* memiliki nilai yang lebih tinggi yakni 0,99 dibandingkan dengan kelas *urgent* 0,97. Secara keseluruhan, model dapat mengklasifikasikan kata *not urgent* dengan konsisten dibandingkan dengan *urgent*.

## Daftar Pustaka

- [1] I. Puspasari and P. H. Rusmin, "Klasifikasi Wazan pada Kata-Kata Al Qur'an Menggunakan Natural Language Processing," *J. Technol. Informatics*, vol. 3, no. 2, pp. 41–48, 2022, doi: 10.37802/joti.v3i2.224.
- [2] A. Riezka, I. Atastina, and K. Maulana, "ANALISIS DAN IMPLEMENTASI DATA-CLEANING DENGAN MENGGUNAKAN

METODE MULTI-PASS NEIGHBORHOOD (MPN) Anandary,” 2011.

[3] V. K. Sitanayah, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.

[4] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, “Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.

[5] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.

[6] I. Saputra and D. Rosiyadi, “Perbandingan Kinerja Algoritma K-Nearest Neighbor, Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp,” *Fakt. Exacta*, vol. 12, no. 2, p. 101, 2019, doi: 10.30998/faktorexacta.v12i2.4181.

[7] A. S. Ritonga and E. S. Purwaningsih, “PENERAPAN METODE SUPPORT VECTOR MACHINE ( SVM ) DALAM KLASIFIKASI KUALITAS PENGELASAN SMAW ( SHIELD METAL ARC WELDING ),” vol. 5, no. 1, pp. 17–25, 2018.

[8] M. T. Anjasmos, Istiadi, and F. Marisa, “Seminar Nasional Hasil Riset Prefix-RTR ANALISIS SENTIMEN APLIKASI GO-JEK MENGGUNAKAN METODE SVM DAN NBC (STUDI KASUS: KOMENTAR PADA PLAY STORE),” *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020.

[9] S. S. Berutu, H. Budiati, J. Jatmika, and F. Gulo, “Data preprocessing approach for machine learning-based sentiment classification,” *J. Infotel*, vol. 15, no. 4, pp. 317–325, 2023, doi: 10.20895/infotel.v15i4.1030.

[10] H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issuue3.year2020.page133-143.

[11] Y. S. HARIYANI, S. HADIYOSO, and T.

S. SIADARI, “Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network,” *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 8, no. 2, p. 443, 2020, doi: 10.26760/elkomika.v8i2.443.