**SENTIMENT ANALYSIS PEMILIHAN CALON PRESIDEN 2024**

**MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE**

*[****SENTIMENT ANALYSIS 2024 PRESIDENTIAL CANDIDATE ELECTION USING SUPPORT VECTOR MACHINE ALGORITHM****]*

**Michael Alfonso, s32190039@student.ubm.ac.id1), Dionisia Bhisetya Rarasati, S.Kom., M.T.I, l1783@lecturer.ubm.ac.id 2)**

1) Program Studi Informatika, Universitas Bunda Mulia

2) Program Studi Informatika, Universitas Bunda Mulia

Disetujui 15 Agustus 2023

***ABSTRACT***

*The election of presidential candidates is held every 5 years with various candidates running for office. On social media Twitter there are often arguments about political matters which are the topic of discussion regarding the election of presidential candidates. Thus, this research focuses on the sentiment analysis process of Twitter users responses to the election of presidential candidates followed by validation by looking for correlations between electability survey results and tweets using Pearson Correlation. Method used to search for the best model using 10-Fold Cross Validation, then data weighting using TF-IDF and validating the best model using Confusion Matrix to get the closest f1-score to the average f1-score across all folds. The algorithm used to create the model is the Support vector machine algorithm with a Gaussian RBF (Radial Basis Function) kernel. The results of analysis were compared with the news portal electability survey containing the 3 candidates using Pearson Correlation. Correlation for candidates Anies Baswedan and Prabowo Subianto, the higher the positive sentiment, the higher the electability survey data, while for candidate Ganjar Pranowo, the lower the positive sentiment, the higher the electability survey data.*

***Keywords:***  *NLP, Pearson Correlation, Sentiment analysis, SVM, TF-IDF*

**ABSTRAK**

Pemilihan calon presiden dilaksanakan setiap 5 tahun dengan berbagai kandidat yang mencalonkan diri, pada media sosial Twitter sering ternjadi argumen seputar hal-hal politik yang menjadi topik diskusi tentang pemilihan calon presiden ini. Sehingga, penelitian ini berfokus pada proses Analisa sentimen terhadap respon pengguna twitter terhadap pemilihan calon presiden dilanjutkan dengan melakukan validasi dengan mencari korelasi antara hasil survei elektabilitas dan tweet dengan menggunakan Korelasi Pearson. Pencarian model terbaik menggunakan 10-Fold Cross Validation, lalu pembobotan menggunakan TF-IDF serta melakukan validasi dari model terbaik menggunakan Confusion Matrix untuk mendapat f1-score terdekat dari rata-rata f1-score keseluruah *fold*. Algoritma yang digunakan untuk membuat model adalah algoritma Support vector machine dengan kernel Gaussian RBF (Radial Basis Function). Hasil analisa dibandingkan dengan hasil survey elektabilitas portal berita yang memuat 3 calon tersebut dengan menggunakan Korelasi Pearson. Pada Korelasi Peaason, pada calon Anies Baswedan dan Prabowo subianto, semakin tinggi sentiment positif, maka semakin tinggi juga data survei elektabilitas, sedangkan untuk calon Ganjar Pranowo, semakin rendah sentiment positif, maka semakin tinggi data survei elektabilitas.

**Kata Kunci:** NLP, Pearson Correlation, Sentiment analysis, SVM, TF-IDF

**PENDAHULUAN**

Pemilu terjadi setiap 5 tahun sekali dengan jumlah pemilih yang semakin bertambah setiap periodenya, penambahan pemilih tetap sejak tahun 2004 sampai 2019 terhitung sebanyak 28,90%. Media sosial seperti Facebook, Twitter, LinkedIn, YouTube sudah menjadi bagian dari kehidupan masyarakat kita, hal ini dibuktikan dengan cara berinteraksi kebanyakan dapat dilakukan dengan sosial media, sampai menjadi identitas diri ke masyarakat umum [1].

Pada sosial media Twitter, terjadi penyebaran informasi, hiburan, bahkan opini politik, salah satu informasi yang tersebar adalah terjadinya pemilu pada tahun 2024 yang disebut sebagai pesta demokrasi. Cara penyampaian opini melalui social media Twitter adalah dengan menggunakan tanda pagar tertentu pada tweet yang dipost di Twitter. Twitter digunakan karena kemudahan dalam membalas *tweet* dengan *like*, komentar bahkan *reply*, sehingga topik tersebut dapat dengan mudah tersebar [3].

Kerap kali sosial media didominasi oleh banyak masyarakat yang tidak memiliki pendapat yang kurang didengar lalu dituangkan aspirasi tersebut dalam sosial media, sosial media dapat dikategorikan sebagai kumpulan informasi dan dijadikan pengetahuan politik oleh masyarakat Indonesia [2] khususnya Twitter tentang hak pilih mereka terhadap pemilihan calon presiden 2024. Sebab itu, diperlukannya analisa lebih lanjut untuk mengatahui kesetujuan dan ketidaksetujuan pengguna Twitter terhadap pemilihan calon presiden 2024.

Sentiment analysis atau dapat disebut Opinion Mining merupakan studi yang mempelajari opini, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas [1], [3]. Disebutkan juga oleh sumber lain bahwa analisis sentimen adalah sebuah cara yang dilakukan untuk mendapatkan Opini Publik [4].

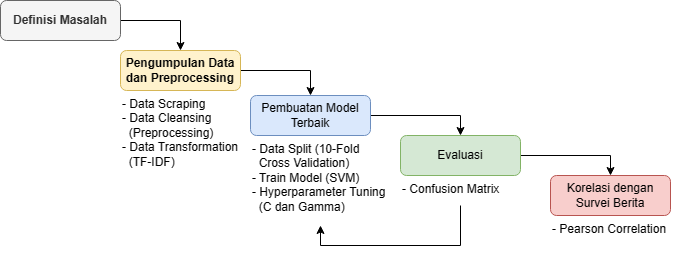
Penelitian ini menggunakan data yang diperoleh dari tweet pengguna Twitter yang berkaitan dengan topik pemilihan umum 2024. Penelitian ini menggunakan metode klasfikasi *supervised* yaitu Support vector machine (SVM). Klasifikasi dilakukan dengan membuat garis pemisah (hyperlane) antara kelas positif dan negatif [1] dengan bantuan garis pemisah yang dibentuk melalui proses pemilihan data yang disebut *support vector* sehingga membentuk garis pemisah terbaik dari sebuah dataset [5].

Sehingga hasil dari penelitian ini adalah sentiment pengguna Twitter terhadap pemilihan calon presiden pada tahun 2024 dengan hasil korelasi antara setiap calon presiden dengan survei elektabilitas portal berita.

Dengan adanya system ini, diharapkan kita sebagai masyarakat yang kurang aktif dalam sosial media, dapat mengetahui ringkasan sentiment pengguna Twitter terhadap pemilihan calon presiden 2024.

**METODOLOGI PENELITIAN**

Penelitian ini terdiri dari 6 proses secara garis besar yaitu pengumpulan data, preprocessing, pembobotan Term Frequency-Inverse Document Frequency, train Support Vector Machine menggunakan 10-Fold Cross Validation, Memilih fold terbaik untuk dijadikaan model, dan pengujian seperti pada Gambar 1.

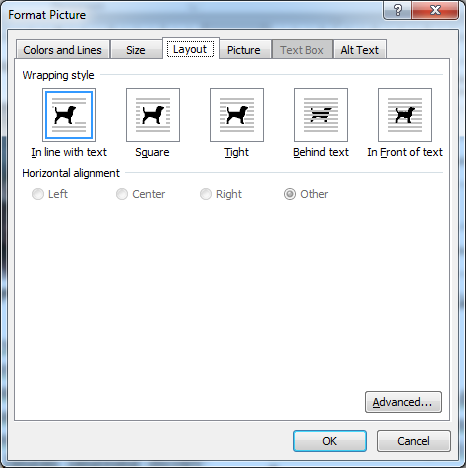


***Gambar******1.******Flowchart Penelitian***

(Times New Roman 11 poin, spasi single, dua kolom) Menjelaskan langkah-langkah sistematis yang dilakukan dalam kegiatan penelitian. Yang dibahas dalam metode misalnya metodologi penelitian, subjek penelitian, pengembangan instrumen, teknik analisis data / uji statistik yang dilakukan. Metode penelitian menjelaskan tentang tahap penelitian secara rinci dan jelas sehingga dapat dirujuk oleh peneliti lain (repeatable and reproduceable).

**Gambar**

Gambar diformat alignment center, dengan format layout/wrap “Inline with Text” seperti yang diperagakan pada Gambar 1. Jangan menggunakan Layout lainnya karena akan beresiko gambar hilang ketika dilakukan editing.



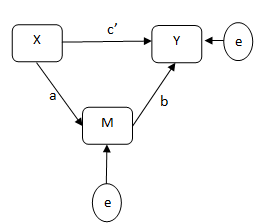
**Gambar 1. *Wraping Inline With Text***

Harus ada kalimat yang mengacu pada gambar tersebut dengan menyebutkan Gambar+nomor label gambar. Tidak boleh menggunakan kalimat “gambar di atas”, ”gambar di bawah ini”, atau “gambar berikut” dsb. Agar jika posisi gambar berubah ketika editing, editor tidak perlu menyesuaikan ulang.

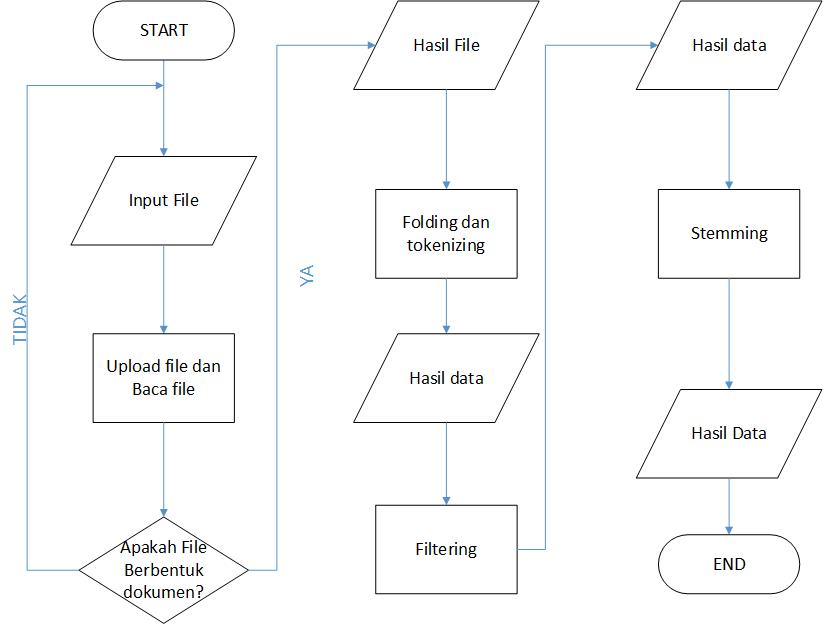
Label dan caption gambar diformat dengan alignment tengah dan di bold. Sertakan referensi sumber gambar!

Gambar yang dibuat dengan menggunakan visio seperti pada contoh Gambar 2, ataupun perangkat pengolah grafik lainnya termasuk yang menggunakan shape dari Microsoft word seperti pada contoh Gambar 3, Jangan langsung dicopy paste dalam bentuk objek. Adanya objek visio ataupun shape ms word pada artikel akan mengakibatkan hasil cetak kurang jelas dan menyulitkan editing, ubahlah menjadi image dahulu dengan copy-paste ke paint atau perangkat gambar sejenisnya, atau bisa juga dengan printscreen atau dengan perangkat capture.

Gambar tidak boleh melampaui batas kiri dan kanan margin, karena ketika pada proses editing gambar tersebut berpindah tempat akan beresiko menutupi kolom sebelahnya.



**Gambar 2. Contoh Gambar dari Shape MS.Word**

****

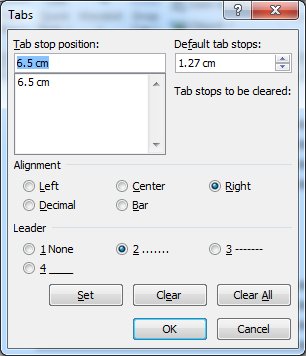
**Gambar 3. Contoh Gambar Hasil Visio**

**Penulisan Rumus**

Rumus menggunakan Microsoft equation. Rumus tidak boleh disajikan dalam bentuk gambar. Rumus harus diacu dan dijelaskan oleh kalimat dalam artikel. Contoh penulisan rumus seperti pada rumus (1).

 (1)

Jika memungkinkan, rumus disarankan menggunakan leader titik. Gunakan *format tab-stop* untuk memformat *tab-stop* *right allignment* dengan *leader* titik. Untuk menampilkan *tab-stop* *format* tersebut dapat dilakukan dengan cara menaruh tab-stop pada posisi yang dikehendaki pada *ruler* yang terdapat di atas *editor*, kemudian *double click* pada *tab-stop* tersebut. Atau dapat juga menggunakan format equation yang disediakan oleh ms word.

****

**Gambar 4. Tab-Stop Format**

**Tabel**

Tabel bikin dalam *text box*, *textbox* diformat “*inline with text*” Seperti pada contoh Tabel 1. Harus ada kalimat yang mengacu pada tabel tersebut dengan menyebutkan Tabel+nomor label tabel. Tidak boleh menggunakan kalimat “tabel di atas”, “tabel di bawah ini”, atau “tabel berikut” dsb. Agar jika posisi tabel berubah ketika editing, editor tidak perlu menyesuaikan ulang.

Tabel tidak menggunakan garis pemisah horisontal kecuali pada *header* table. Tabel rata kiri, label dan keterangan “Tabel” di bold.

Jumlah gambar dan tabel pada artikel sebaiknya tidak lebih dari 6 gambar berikut tabelnya.

**Tabel 1. Contoh Tabel**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Header  1 | Header  2 | Header  3 | Header  4 | Header  5 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Penulisan Daftar Pustaka**

Artikel ilmiah harus memuat daftar referensi. Daftar pustaka dibentuk secara otomatis menggunakan mendeley, hanya artikel yang dicitasi yang dimunculkan pada daftar pustaka.

Daftar referensi disusun konsisten dengan mengikuti model sistem Harvard atau IEEE. Kutipan dari referensi ditulis dengan pemberian kurung siku berisi nomor referensi dari daftar pustaka pada bagian akhir kutipan. Contoh kutipan pada naskah adalah seperti pada paragraph berikut ini:

Suatu citra watermark dijadikan citra *grayscale*, yang hanya mempunyai satu nilai intensitas.[6] Jika nilai intensitas pada piksel-piksel dari citra watermark yang telah diubah menjadi citra *grayscale* tersebut digunakan untuk menaikkan atau menurunkan nilai intensitas channel RGB suatu citra.[6][7] Posisi dan variasi intensitas masing-masing piksel yang sama dengan citra watermark akan menghasilkan variasi gelap dan terang pada piksel-piksel citra yang akan diwatermark dengan posisi-posisi yang sama dengan posisi pada citra watermark.[8] Piksel yang dianggap tepi merupakan piksel yang berada pada tepian objek.[9][10]

Gunakan Mendeley untuk menyusun citasi dan daftar Pustaka!

# HASIL DAN PEMBAHASAN

(Times New Roman 11 poin, spasi single, dua kolom) Menjelaskan apakah pelaksanaan kegiatan penelitian yang dilakukan mampu mencapai tujuan dan manfaat yang diharapkan. Hasil dan pembahasan didukung dengan data dan menggunakan uji dan indikator yang umum digunakan di bidang keilmuan masing-masing. Pada dalam pembahasan dikemukakan keterkaitan antara hasil penelitian dengan teori, tujuan penelitian dan pembandingan dengan penelitian lain yang telah dipublikasikan. Pembahasan juga menjelaskan implikasi/kontribusi temuan bagi ilmu pengetahuan. Hasil dan pembahasan diungkapkan secara jelas dan lugas menggunakan kalimat sederhana. Hasil dapat ditampilkan dalam grafik ataupun tabel.

**SIMPULAN**

(Times New Roman 11 poin, spasi single, dua kolom). Simpulan adalah jawaban atas hipotesis yang diajukan (atau judul). Simpulan harus didasari fakta yang ditemukan, serta implikasinya perlu dikemukakan untuk memperjelas manfaat yang dihasilkan. Simpulan diharapkan bersifat kritis, cermat, logis, dan jujur berdasarkan fakta yang diperoleh.

**DAFTAR PUSTAKA**

Berikut ini contoh penulisan daftar pustaka:

[1] B. W. Sari and F. F. Haranto, “IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET,” *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, Sep. 2019, doi: 10.33480/pilar.v15i2.699.

[2] C. N. Zempi, A. Kuswanti, and S. Maryam, “ANALISIS PERAN MEDIA SOSIAL DALAM PEMBENTUKAN PENGETAHUAN POLITIK MASYARAKAT,” *EKSPRESI DAN PERSEPSI : JURNAL ILMU KOMUNIKASI*, vol. 6, no. 1, pp. 116–123, Jan. 2023, doi: 10.33822/jep.v6i1.5286.

[3] Fatihah Rahmadayana and Yuliant Sibaroni, “Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.

[4] I. A. Mastan and Y. Toni, “ANALISIS SENTIMEN TERHADAP TEMPAT KULINER AYAM GEDEBUK DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” *JBASE - Journal of Business and Audit Information Systems*, vol. 3, no. 1, Mar. 2020, doi: 10.30813/jbase.v3i1.2062.

[5] F. Feiters Tampinongkol, C. Herdian, H. Basri, and L. Halim, “Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi Identifikasi Penyakit Daun Tomat Menggunakan Gray Level Co-occurrence Matrix (GLCM) dan Support Vector Machine (SVM),” *Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 8, no. 1, 2023, [Online]. Available: https://www.kaggle.com/

[6] T. M. S. Mulyana, “PENGGUNAAN NILAI SKALA KEABUAN DARI CITRA WATERMARK SEBAGAI SEBAGAI CETAK BIRU DARI VISIBLE WATERMARKING Teady,” *Seminar Nasional Informatika*, vol. 2013, no. semnasIF, pp. 23–30, 2013.

[7] “Morphological Edge Detection Algorithm Based on Multi-Structure Elements of Different Directions,” vol. 1, no. 1, pp. 37–43, 2011.

[8] J.-H. Park, G. Baasantseren, N. Kim, G. Park, J.-M. Kang, and B. Lee, “View image generation in perspective and orthographic projection geometry based on integral imaging,” *Optics Express*, vol. 16, no. 12, p. 8800, 2008, doi: 10.1364/oe.16.008800.

[9] D. Chudasama, T. Patel, S. Joshi, and G. I. Prajapati, “Image Segmentation using Morphological Operations,” *International Journal of Computer Applications*, vol. 117, no. 18, pp. 16–19, 2015.

[10] T. M. S. Mulyana, “EFEK HIGH PASS FILTERING DENGAN KOEFESIEN NOL PADA CITRA BINER,” *Jurnal Muara Sains, Teknologi, Kedokteran dan Ilmu Kesehatan*, vol. 1, no. 1, May 2017, doi: 10.24912/jmstkik.v1i1.394.