

Analysis Of Public Sentiment Regarding The 2024 Pre-Employment Training System Update On Social Media Using The SVM Method

Ferdiansyah Ijian Sagama¹, Dr. Heny Pratiwi², Kusno harianto³

¹Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika Dan Komputer, Widya Cipta Dharma
¹Jalan M Yamin No 14 Kelurahan Samarinda Ulu Kota/Kabupaten Kec. Samarinda Ulu Kota Samarinda Prov. Kalimantan Timur Kode Pos 75123
Email: ferdiansya1805@gmail.com

ABSTRACT

This research aims to analyze public sentiment regarding the 2024 pre-employment training system update on Instagram social media using the SVM method. Social media, especially Instagram, has become an important platform for monitoring people's reactions to various issues, including politics. In this research, Instagram comment data was collected and analyzed to identify positive, negative and neutral sentiments. The analysis process includes Data Collection, Pre-Processing, Labeling, TF-IDF Weighting, Modeling with Support Vector Machine (SVM), Confusion Matrix and Visualization. The results of research conducted at the social media showed that the SVM method was able to classify sentiment with an accuracy rate of 86%, precision of 85%, and recall of 85%. It is hoped that this research can contribute to the development of more sophisticated and effective sentiment analysis methods and become a basis for deeper research into understanding the dynamics of public opinion in the increasingly complex context of modern politics.

Keywords: *Sentiment Analysis, Politics, Social Media, Instagram, Support Vector Machine (SVM).*

ANALISIS SENTIMEN MASYARAKAT TERKAIT PEMBARUAN SISTEM PELATIHAN PRAKERJA TAHUN 2024 PADA MEDIA SOSIAL INSTAGRAM DENGAN MENGGUNAKAN METODE SVM

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terkait pembaruan sistem pelatihan prakerja tahun 2024 pada media sosial instagram menggunakan metode SVM. Media sosial, khususnya Instagram, menjadi platform penting untuk mengamati reaksi masyarakat terhadap berbagai isu, termasuk politik. Dalam penelitian ini, data komentar Instagram dikumpulkan dan dianalisis untuk mengidentifikasi sentimen positif, negatif, dan netral. Proses analisis meliputi Pengumpulan Data, *Pre-Processing*, Pelabelan, Pembobotan TF-IDF, Pemodelan Dengan *Support Vector Machine (SVM)*, *Confusion Matrix* dan Visualisasi. Hasil penelitian, yang dilakukan pada media sosial menunjukkan bahwa metode SVM mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 86%, presisi sebesar 86%, dan recall sebesar 85%. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen yang lebih canggih dan efektif serta menjadi landasan bagi penelitian lebih lanjut dalam memahami dinamika opini publik dalam konteks politik modern yang semakin kompleks.

Kata Kunci: Analisis Sentimen, Politik, Media Sosial, Instagram, Support Vector Machine (SVM)

1. PENDAHULUAN

Program prakerja merupakan salah satu upaya strategis pemerintah Indonesia yang disahkan pada awal tahun 2020 sebagai upaya dalam membantu masyarakat untuk memperoleh keterampilan yang dibutuhkan dalam dunia pekerjaan. Program ini menggabungkan pelatihan daring dengan pemberian insentif berupa uang tunai setelah peserta menyelesaikan pelatihan. Sistem pelatihan yang diterapkan pada tahun itu cukup sukses karena memberikan akses yang mudah bagi banyak

orang untuk mengikuti pelatihan dari berbagai penyedia pelatihan *online* yang terdaftar.

Meskipun tujuan dari program prakerja adalah untuk meningkatkan keterampilan tenaga kerja di Indonesia, namun pada kenyataannya pada tahun 2024 setelah dimulainya gelombang baru program ini menghadapi banyak tantangan yang menyebabkan penurunan efektivitas pelatihan yang diberikan. Salah satu masalah utama yang muncul adalah kesulitan dalam mengakses pelatihan secara *online*. Hal ini disebabkan oleh

penerapan sistem verifikasi wajah yang baru pada tahapan proses pembelajaran. Sistem verifikasi wajah ini menambah langkah tambahan yang harus dilalui peserta sebelum bisa melanjutkan ke tahap berikutnya. Bagi sebagian besar peserta terutama yang kurang familiar dengan teknologi atau yang memiliki kendala akses internet, hal ini menjadi hambatan yang cukup besar. Akibatnya banyak peserta yang merasa frustrasi dan enggan untuk melanjutkan pelatihan pada gelombang sebelumnya.

Merujuk pada fakta diatas, peneliti ingin mencari tau berapa jumlah hasil sentimen yang beredar dikalangan masyarakat dengan melakukan penelitian di DPD Partai Golkar untuk menganalisis sentimen masyarakat terkait pembaruan sistem pelatihan prakerja tahun 2024 pada media sosial instagram dengan menggunakan metode *support vector machine* (SVM).

Oleh karena itu, penelitian ini diharapkan mampu memberikan kontribusi yang signifikan dalam memahami dinamika persepsi publik terhadap sentimen yang beredar dikalangan masyarakat terkait pembaruan sistem pelatihan prakerja dan dapat menjadi referensi bagi mahasiswa terkait bagaimana cara merancang sebuah analisis sentimen dengan menggunakan penerepan metode *support vector machine* (SVM) dalam melakukan analisis untuk kedepanya.

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

1. Rumusan masalah yang akan diangkat pada penelitian ini adalah “Bagaimana cara menganalisis sentimen yang beredar dikalangan masyarakat terkait pembaruan sistem pelatihan prakerja tahun 2024 pada media sosial instagram dengan menggunakan metode *support vector machine* (SVM?)”.
2. Agar tidak meluasnya permasalahan ini, maka dibuatlah batasan masalah sebagai berikut:
 1. Data diambil dari media sosial instagram Prakerja.go.id.
 2. Diluar komentar bahasa Indonesia tidak dipergunakan.
 3. Data yang dikumpulkan antara 1000-5000 data.
 4. Data yang dibutuhkan untuk menentukan sentimen hanya diambil dari data komentar tahun 2024.
3. Dengan menggunakan metode *support vector machine* (SVM), penelitian ini akan dapat menyediakan informasi terkait berapa jumlah hasil sentimen yang beredar dikalangan masyarakat pada media sosial instagram.

3. METODE PENELITIAN

3.1 Tempat Dan Waktu Penelitian

Tempat penelitian akan dilakukan di DPD Partai Golkar Kalimantan Timur, Jl. Mulawarman No.95, Karang Mumus, Kec. Samarinda Kota, Kota Samarinda, Kalimantan Timur 75113. Penelitian ini dilakukan

selama kurang lebih satu bulan yang dimulai pada bulan Desember sampai Januari 2025.

3.2 Tahapan Penelitian

3.2.1 Observasi

Observasi dilakukan sebagai studi pendahuluan, peneliti untuk mengamati secara langsung postingan prakerja.go.id dengan melihat pendapat atau komentar yang ada pada media sosial Instagram. Observasi dilakukan peneliti pada bulan November 2024. Beberapa hal yang diperoleh peneliti dari observasi yang telah dilakukan adalah memperoleh informasi bahwa terdapat tanggapan atau komentar yang beragam dari postingan prakerja terkait pengembangan sistem pelatihan pada instagram.

3.2.2 Studi Pustaka

Metode pengumpulan studi pustaka ini dilakukan untuk mencari, mempelajari, dan menggunakan berbagai literatur seperti, buku, jurnal, *paper*, *e-book*, atau literatur lain yang berhubungan dengan tema penelitian ini. Informasi yang terkumpul akan dijadikan data pendukung dan pembanding untuk penelitian ini.

3.2.3 Crawling

Pada tahapan ini dilakukan pengumpulan dan pencarian dataset instagram Bahasa Indonesia. Data yang dikumpulkan adalah data komentar masyarakat dari instagram yang merupakan pesan ataupun opini dari masyarakat terkait pembaruan sistem pelatihan prakerja 2024.

3.3 Pre-Processing

Pre-processing merupakan tahapan awal yang akan dilalui dalam memproses data mentah hingga menjadi data yang lebih terstruktur. Tahapan *pre-processing* yang dilakukan meliputi:

1. *Cleansing*, merupakan tahapan yang bertujuan menghilangkan karakter atau *symbol link url* (<http://situs.com>), *username* atau *mention* (*@username*), *hastage*(#), *retweet*, dan *emoticon*.
2. *Case Folding*, adalah proses mengubah semua karakter dalam sebuah dokumen dari yang awalnya huruf kapital besar menjadi huruf kapital kecil.
3. *Tokenization*, dilakukan untuk memisahkan teks menjadi bagian-bagian yang lebih terstruktur sehingga mempermudah proses analisis selanjutnya, seperti penghapusan karakter khusus, menghitung frekuensi kata, atau membangun representasi vektor dari teks.
4. *Stopword*, berfungsi untuk membuang kata-kata tidak bermakna yang tidak mempengaruhi klasifikasi sentimen, seperti kata penghubung dan kata depan berdasarkan kamus *stopword*.
5. *Stemming*, tahap ini bertujuan untuk menghilangkan kata imbuhan awalan, akhiran, sisipan atau pun kombinasi imbuhan. Pada penelitian ini akan menggunakan *library* sastrawi untuk melakukan

proses *stemming*. Sastrawi merupakan *library* pada Bahasa pemrograman *python* yang dibangun dengan algoritma NA. Algoritma tersebut dibuat berdasarkan aturan Bahasa Indonesia.

3.4 Pelabelan

Pelabelan bertujuan untuk memberikan kategori atau tanda pada dataset agar atribut tertentu dapat diidentifikasi. Dengan menggunakan tahapan pelabelan maka akan memudahkan penelitian ini dalam menentukan kalimat sentimen yang terdiri dari sentimen negatif, positif ataupun netral. Pelabelan akan dibantu dengan kamus *lexicon*, berupa kamus kata positif dan negatif.

3.5 Pembobotan TF-IDF

Setelah melalui tahapan pelabelan, berikutnya adalah tahap pembobotan kata. Pembobotan bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. TF-IDF (*Term Frequency-Inverse Document Frequency*) memberi nilai penting pada kata dalam satu dokumen dengan mempertimbangkan seberapa sering kata tersebut muncul didokumen itu dan seberapa jarang kata itu muncul diseluruh kumpulan dokumen. Rumus untuk menghitung bobot (W) masing masing dokumen terhadap kata kunci dengan rumus yaitu:

$$TF(t, d) = \frac{\text{Number of times } t \text{ appear in document } d}{\text{Total number of terms in document } d}$$

$$IDF(t, d) = \log \left(\frac{\text{Total number of documents } D}{\text{Number of document the term in it}} \right)$$

$$TFIDF(t, d, D) = TF(t, d) * IDF(t, d)$$

Keterangan:

TF = *Term Frekuensi*

DF = *Document Frekuensi*

IDF = *Inverse Document Frekuensi*

n = Banyaknya Data

3.6 Confusion Matrix

Confusion Matrix berisi informasi asli dan yang diprediksi. Dari matriks tersebut dapat diketahui *accuracy*, *precision* dan *recall*. Perhitungan *Confusion Matrix* dengan 3 kelas yaitu positif, netral dan negatif. *Confusion Matrix* berisi informasi asli dan yang diprediksi, dari matriks tersebut dapat diketahui *accuracy*, *precision* dan *recall*. Akurasi merupakan perbandingan antara prediksi benar pada bagian positif dan negatif dengan keseluruhan data. Dalam konteks pembelajaran mesin, akurasi sering digunakan untuk menilai kinerja model klasifikasi akurasi. Gambar 1 merupakan perhitungan *Confusion Matrix* dengan 3 kelas yaitu positif, netral dan negatif.

Confusion Matrix		Predicted Class		
		Positive	Neutral	Negative
True Class	Positive	True Positive (TP)	False Neutral (FN)	False Negative (FN)
	Neutral	False Positive (FP)	True Neutral (TN)	False Negative (FN)
	Negative	False Positive (FP)	False Neutral (FN)	True Negative (TN)

Gambar 1 Confusion Matrix

3.7 Visualisasi

Tahap ini merupakan tahap memvisualisasikan kata-kata yang sering muncul pada data komentar ke dalam bentuk *word cloud*. Hal ini bertujuan untuk mengetahui kata-kata apa saja yang paling banyak yang terdapat pada data komentar dan berapa frekuensi kata tersebut muncul dalam sebuah komentar. Visualisasi juga digunakan untuk menampilkan jumlah emosi pada data komentar dalam bentuk barplot yang terdiri atas beberapa parameter emosi.

4. PEMBAHASAN

4.1 Sejarah Tempat Penelitian

Partai Golkar didirikan pada tahun 1964 yang berfungsi sebagai alat politik untuk mendukung pemerintahan orde baru, setelah era reformasi partai ini terus berkembang menjadi salah satu kekuatan politik utama di Indonesia. Gedung DPD Partai Golkar di Samarinda, yang terletak di Jalan Mulawarman, memiliki sejarah yang erat kaitannya dengan perkembangan Partai Golkar di Kalimantan Timur, khususnya di kota Samarinda. Sebagai salah satu partai politik terbesar di Indonesia, Golkar memainkan peran penting dalam kehidupan politik nasional dan daerah, termasuk dalam membentuk struktur organisasi dan infrastruktur partai di tingkat lokal.

4.2 Pengumpulan Data



Gambar 2 Prose Pengumpulan Data

Tahap pertama yang akan dilakukan yaitu melakukan proses *crawling* data menggunakan *export comment* dari *google chrome* yang akan secara otomatis melakukan pengumpulan data dengan cara menyalin link postingan ya ada pada unggahan media sosial *prakerja.go.id*. Setelah link berhasil disalin kemudian dimasukan kedalam *export comment* lalu klik *export*, maka data

yang terkumpul akan disimpan dalam format yang terstruktur untuk memudahkan analisis lebih lanjut. Selain itu, alat ini memastikan bahwa setiap komentar dan metadata yang relevan diambil dengan akurasi tinggi, sehingga memberikan gambaran menyeluruh tentang interaksi pengguna.

4.3 Pre-Processing

Setelah data sudah berhasil dimasukkan kedalam *google collab*, maka tahap selanjutnya yaitu melakukan tahapan *cleansing* atau pembersihan data pada teks kolom komentar.

1. Cleansing



Gambar 3 Tahapan *Cleansing*

Tahapan *cleansing* bertujuan untuk menghilangkan atau menghapus bagian yang tidak diperlukan seperti *URL*, *hashtag*, *mention*, *emoticon*, dan karakter spesial lainnya. Proses pembersihan ini sangat penting untuk memastikan data yang digunakan dalam analisis lebih akurat dan relevan. Setelah data dibersihkan.

2. Case Folding



Gambar 4 Tahapan *Case Folding*

Tahapan *case folding* bertujuan untuk merubah semua huruf kapital yang ada pada kolom komentar dari yang awalnya huruf kapital besar menjadi huruf kapital kecil. Contoh dari tahapan sebelum dan sesudah melakukan *case folding* yaitu pada kalimat "Jangan mau ikut gelombang 71 parah loh", menjadi "jangan mau ikut gelombang 71 parah loh".

3. Tokenization

Setelah berhasil merubah huruf kapital pada komentar, langkah selanjutnya yaitu melakukan *Tokenization*. Tahapan *tokenization* atau yang sering

disebut token merupakan tahapan yang bertujuan untuk membagi satu kalimat pada komentar menjadi beberapa kata.



Gambar 5 Tahapan *Tokenization*

4. Stopword



Gambar 6 Tahapan *Stopword*

Tujuan dari tahapan *stopword* yaitu untuk menghilangkan sebagian kata yang tidak dianggap penting pada kolom komentar, seperti "jangan, mau, ikut, dan lain-lain". Contoh kalimat yang dihilangkan pada kolom yaitu "jangan, mau, ikut, gelombang, 71, parah, loh, main, curang" menjadi "gelombang, 71, parah, loh, main, curang".

5. Stemming



Gambar 7 Tahapan *Stemming*

Tahapan sebelum dan sesudah melakukan tahapan *stemming* yaitu pada kalimat "hubungannya" setelah dilakukan *stemming* menjadi "hubung". Kalimat sebelumnya dihilangkan karena dianggap tidak terlalu memiliki arti penting dalam proses *text mining* dan

- Husni, Nurhikmahwati (2022) Sentimen Analisis Pada Komentar Instagram Selebgram Dan Influencer Terkait Pariwisata di Yogyakarta Menggunakan Metode *Lexicon Based Dan Svm*. <http://eprints.upnyk.ac.id/31018/2/Cover.pdf>
- Handoko, I. T., & Suyanto, S. (2019). Klasifikasi gender dan usia berdasarkan suara pembicara menggunakan hidden markov model. *Indonesia Journal on Computing (Indo-JC)*, 4(3), 99-106. <http://eprints.upnyk.ac.id/30283/4/Daftar%20Pustaka%20Skripsi%20-%20Muhammad%20Iqbal%20-123160079.pdf>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis sentimen cyberbullying pada komentar instagram dengan metode klasifikasi support vector machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704-4713. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3051>
- Mujiati, H. (2013). Analisis dan perancangan sistem informasi stok obat pada apotek arjowinangun. *Indones. J. Comput. Sci.-Speed FTI UNSA*, 9330(2), 1-6 (5). <https://ejournal.bsi.ac.id/ejournal/index.php/Bianglala/article/viewFile/536/428>
- Maulana, G., & Danar Dana, R. (2024). Prediksi Hasil Produksi Jagung Di Jawa Barat Dengan Metode Algoritma Regresi Linear Menggunakan Google Collab. *Jati (Jurnal Mahasiswa Teknik Informatika) Technology (J-HEST)*, 6(2), 151-161. <https://ejournal.itn.ac.id/index.php/jati/article/download/8816/5119/>
- Muharram, F. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kinerja Walikota Medan Menggunakan Metode Naive Bayes Classifier (Doctoral dissertation, Universitas Negeri Medan). <https://www.semanticscholar.org/paper/Analisis-Sentimen-Pengguna-Twitter-Terhadap-Kinerja-Twitter/ca852a0e575db27b64c1174bb6ee151c4ded376f>
- Noventa, C., Soraya, I., & Muntazah, A. (2023). Pemanfaatan media sosial Instagram buddyku sebagai sarana informasi terkini. *Jurnal Ilmu Komunikasi Dan Media Sosial (JKOMDIS)*, 3(3), 626-635. <https://jurnal.minartis.com/index.php/jkomdis/article/view/1124>
- Putra, E. A. (2021). Sentiment Analysis Using Bidirectional Encoder Representations from Transformers (BERT). *Diakses dari* <https://medium.com/@eza.a.putra/i-plementasi-bert-untuk-analisis-sentimen-terhadap-ulasan-aplikasi-flip-berbahasa-indonesia-557d691e0440>. <https://medium.com/@eza.a.putra/implementasi-bert-untuk-analisis-sentimen-terhadap-ulasan-aplikasi-flip-berbahasa-indonesia-557d691e0440>
- Parhusip, Martin (2023) Analisis Sentimen Masyarakat terhadap Tragedi Kanjuruhan dengan Metode Support Vector Machine. Undergraduate Thesis thesis, Institut Telkom Telkom. <https://repository.itelkom-pwt.ac.id/10170/>
- Rahmawati, S. (2023). Implementasi Algoritma BERT untuk Analisis Sentimen Ulasan Pengguna Aplikasi PeduliLindungi (Doctoral dissertation, UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA). <https://eprints.utdi.ac.id/9863/>
- Sinaga, H. H., & Agustian, S. (2022). Perbandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter. Perbandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter, 8(03), 107-114. <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/2170>
- Sonny, S. (2021). Pengembangan Sistem Presensi Karyawan Dengan Teknologi GPS Berbasis Web Pada PT BPR Dana Makmur Batam (Doctoral dissertation, Prodi Teknik Informatika). <https://ejournal.upbatam.ac.id/index.php/comasiejournal/article/view/3247>
- Samsudiney, (2019). "Penjelasan Sederhana tentang apa itu SVM? <https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02>. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/11045/4891/77145>
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *INSYST: Journal of Intelligent System and Computation*, 1(1), 43-49. <https://jurnal.stts.edu/index.php/INSYST/article/view/36>
- Wahyu Surya Gemilang, Purwantoro, Carudin (2023) Analisis Sentimen Pengguna Instagram Pada Calon Presiden 2024 Menggunakan Algoritma *Support Vector Machine*. <https://ejournal.itn.ac.id/index.php/jati/article/view/7256>

