

PROPOSAL PENELITIAN



ANALISIS SENTIMEN PROGRAM KARTU PRAKERJA DI INSTAGRAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Disusun untuk Memenuhi Proyek Ujian Akhir Semester (UAS) Mata Kuliah
Kecerdasan Buatan

Ketua / Anggota Tim

Dimas Wahyudi 2006132

Muhammad Rifaldi 2006136

Hilmi Muhamad Ihsan 2006063

Elsa Maharani Sermana 2006145

Fazar Mumtadzidan 2006060

Dendi Ardimansah 2006155

PROGRAM STUDI S-1 TEKNIK INFORMATIKA

JURUSAN ILMU KOMPUTER

INSTITUT TEKNOLOGI GARUT

2023

DAFTAR ISI

HALAMAN	
DAFTAR ISI	1
RINGKASAN	2
1. PENDAHULUAN	3
1.1. Latar Belakang	3
1.2. Masalah Penelitian	4
1.3. Tujuan Penelitian.....	7
1.4. Urgensitas Penelitian	7
2. METODE PENELITIAN	8
2.1 Tahap Penelitian	8
2.2 Program kartu prakerja	9
2.3 Pra-processing	10
2.4 Media sosial instagram	11
2.5 Metode CRISP-DM	12
2.6 Dataset	13
2.7 Analisis Hasil dan Penerapan	15
2.8 Analisis Hasil akhir dan Evaluasi	17
DAFTAR PUSTAKA	20

RINGKASAN

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap Program Kartu Prakerja di platform Instagram menggunakan metode Support Vector Machine (SVM). Program Kartu Prakerja adalah program pemerintah yang memberikan pelatihan dan dukungan kepada pencari kerja. Dalam era digital dan popularitas media sosial, Instagram menjadi salah satu platform yang banyak digunakan untuk berbagi pendapat dan tanggapan terkait program-program pemerintah. Metode SVM digunakan dalam penelitian ini untuk mengklasifikasikan sentimen pengguna Instagram terkait Program Kartu Prakerja menjadi positif, negatif, atau netral. Data teks yang dikumpulkan dari Instagram akan diproses dan fitur-fitur yang relevan akan diekstraksi. Selanjutnya, SVM akan diterapkan untuk mengklasifikasikan sentimen pada data yang dikumpulkan. Hasil analisis sentimen ini akan memberikan pemahaman yang lebih mendalam tentang bagaimana pengguna Instagram merespons Program Kartu Prakerja. Informasi ini dapat membantu pihak terkait dalam memperbaiki dan meningkatkan program, serta merancang strategi komunikasi yang lebih efektif. Diharapkan hasil dari penelitian ini dapat memberikan wawasan yang berharga bagi pemerintah dan pemangku kepentingan terkait dalam mengoptimalkan Program Kartu Prakerja. Analisis sentimen dengan menggunakan SVM juga dapat memberikan kontribusi pada penelitian sebelumnya dalam bidang analisis sentimen di media sosial.

Kata Kunci: analisis sentimen, Program Kartu Prakerja, Instagram, Support Vector Machine (SVM).

1. PENDAHULUAN

1.1. Latar Belakang

Program Kartu Prakerja adalah program pemerintah yang bertujuan untuk memberikan pelatihan dan dukungan kepada para pencari kerja di Indonesia. Program ini memberikan akses ke pelatihan online dan dukungan keuangan bagi pesertanya untuk meningkatkan keterampilan dan memperluas peluang kerja. Program ini mendapat perhatian luas dari masyarakat karena dampaknya yang signifikan dalam membantu pencari kerja. (Hendrastuty et al., 2021)

Instagram merupakan salah satu platform media sosial yang populer di Indonesia. Pengguna Instagram secara aktif berbagi pengalaman, opini, dan tanggapan mereka terkait berbagai topik, termasuk program pemerintah seperti Program Kartu Prakerja. Dalam konteks ini, analisis sentimen dapat menjadi alat yang efektif untuk memahami pendapat dan tanggapan pengguna terhadap Program Kartu Prakerja di Instagram. (RIZWAN, 2022)

Untuk melakukan analisis sentimen tersebut dapat menggunakan metode Support Vector Machine (SVM) yaitu ialah salah satu metode yang digunakan dalam pembelajaran mesin untuk mengklasifikasikan data berlabel ke dalam kategori yang tepat. SVM telah terbukti efektif dalam berbagai tugas klasifikasi, termasuk analisis sentimen. Dengan menggunakan SVM, dapat dilakukan pengklasifikasian sentimen positif, negatif, atau netral terkait Program Kartu Prakerja berdasarkan teks yang dikumpulkan dari Instagram. Analisis sentimen merupakan suatu metode yang digunakan untuk menganalisis dan mengevaluasi sentimen atau pendapat yang dinyatakan dalam teks atau media sosial. Analisis sentimen dapat diterapkan dalam berbagai bidang, termasuk dalam memahami persepsi dan tanggapan pengguna terhadap suatu program atau layanan. Salah satu platform media sosial yang populer adalah Instagram, di mana pengguna dapat berbagi pendapat dan pengalaman mereka terkait berbagai program dan inisiatif. Salah satu program yang mendapatkan perhatian publik adalah Program Kartu Prakerja, yang bertujuan untuk memberikan pelatihan dan dukungan bagi para pencari kerja. (Syah & Witanti, 2022)

Dalam konteks Program Kartu Prakerja, analisis sentimen dapat menjadi alat yang efektif untuk memahami opini dan tanggapan pengguna di platform Instagram terhadap program ini. Dengan melakukan analisis sentimen, dapat ditemukan pola-pola sentimen positif, negatif, atau netral yang dapat memberikan wawasan berharga kepada pihak terkait dalam mengoptimalkan program ini

Dengan menggabungkan analisis sentimen dengan penggunaan metode SVM, penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang persepsi dan tanggapan pengguna terhadap Program Kartu Prakerja di Instagram. Informasi yang diperoleh dari penelitian ini dapat membantu pihak terkait dalam mengidentifikasi kekuatan dan kelemahan program, meningkatkan komunikasi dengan pengguna, serta merumuskan strategi yang lebih efektif untuk meningkatkan keberhasilan dan dampak positif dari Program Kartu Prakerja. (Emeraldi et al., 2022)

1.2. Masalah Penelitian

Penelitian yang dilakukan oleh (Anggraini & Utami, 2021) bertujuan untuk menganalisis respon masyarakat terhadap implementasi kebijakan Kartu Prakerja menggunakan data dari Twitter. Metode klasifikasi yang digunakan adalah Naïve Bayes. Hasil analisis sentimen menunjukkan bahwa persentase sentimen negatif terkait Kartu Prakerja mencapai 52,87%, sedangkan sentimen positif sebesar 47,13%. Selain itu, metode Naïve Bayes memiliki tingkat akurasi sebesar 91,06% dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan Kartu Prakerja. Hasil ini memberikan evaluasi yang penting terhadap pelaksanaan kebijakan tersebut, terutama dalam konteks pandemi yang berpengaruh pada kehidupan sosial ekonomi masyarakat.

Salah satu penelitian yang dilakukan oleh (Riadi, 2019) adalah tentang analisis sentimen pada berita terkait Pemilihan Presiden tahun 2019 di Indonesia yang dipublikasikan oleh media daring seperti Kompas dan Detik. Penelitian ini menggunakan metode web scraping untuk mengambil data teks dari kedua portal berita tersebut. Data teks kemudian diolah dengan melakukan proses text pre-processing sebelum dilakukan analisis sentimen. Hasil dari analisis sentimen ini

dievaluasi menggunakan confusion matrix. Tujuan dari penelitian ini adalah untuk mendeteksi kecenderungan konflik sentimen positif dan negatif pada headline berita yang membahas pemilihan presiden tahun 2019 di masing-masing portal berita. Hasil analisis menunjukkan bahwa akurasi klasifikasi sebesar 70% untuk Jokowi-Ma'ruf dan 65% untuk Prabowo-Sandiaga pada portal berita Kompas, serta akurasi sebesar 70% untuk Jokowi-Ma'ruf dan 80% untuk Prabowo-Sandiaga pada detik.com. Penelitian ini memberikan informasi berharga mengenai data yang diperoleh melalui klasifikasi menggunakan metode Naïve Bayes pada analisis sentimen berita.

Penelitian ketiga yang dilakukan oleh (Hendrastuty et al., 2021) berfokus pada analisis sentimen masyarakat Indonesia terhadap Program Kartu Prakerja melalui pembahasan di media sosial Twitter. Penelitian ini bertujuan untuk menganalisis pandangan masyarakat terhadap program tersebut dengan menggunakan metode Support Vector Machine (SVM) dan data yang diperoleh dari Twitter. Sentimen yang diidentifikasi meliputi positif, negatif, dan netral. Selain itu, penelitian ini juga membandingkan kinerja SVM dengan dua jenis kernel, yaitu kernel linear dan RBF. Evaluasi dilakukan menggunakan metode Confusion Matrix. Hasil evaluasi menunjukkan bahwa SVM dengan kernel linear mencapai tingkat akurasi sebesar 98.67%, dengan precision sebesar 98%, recall sebesar 99%, dan F1-Score sebesar 98%. Sedangkan SVM dengan kernel RBF memiliki akurasi sebesar 98.34%, precision sebesar 97%, recall sebesar 98%, dan F1-Score sebesar 98%. Dari hasil tersebut, dapat disimpulkan bahwa sentimen masyarakat pengguna Twitter terhadap program kartu prakerja pada masa pandemi cenderung netral sebesar 98.34%. Selain itu, dari segi akurasi, SVM dengan kernel linear menunjukkan performa yang lebih baik daripada kernel RBF.

Penelitian keempat yang dilakukan oleh (Athira Luqyana et al., 2018) berfokus pada analisis sentimen terhadap komentar di Instagram yang berkaitan dengan bahaya cyberbullying. Instagram merupakan media sosial yang sangat populer saat ini dan digunakan oleh berbagai kalangan, termasuk anak-anak, remaja, dan orang dewasa. Namun, fenomena cyberbullying sering terjadi di kolom komentar Instagram, yang dapat menyebabkan dampak yang meresahkan, terutama pada

remaja. Penelitian ini bertujuan untuk melakukan analisis sentimen pada setiap komentar dengan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) dan metode klasifikasi Support Vector Machine (SVM). Data yang digunakan terdiri dari 400 komentar yang diambil secara offline. Total fitur yang digunakan adalah 1799. Data komentar dibagi menjadi dua bagian, yaitu 70% untuk data latih dan 30% untuk data uji. Melalui pengujian yang dilakukan, peneliti menemukan parameter terbaik untuk metode SVM, yaitu nilai degree kernel polynomial sebesar 2, nilai learning rate sebesar 0,0001, dan jumlah iterasi maksimum sebanyak 200 kali. Hasil pengujian menunjukkan bahwa metode ini mencapai tingkat akurasi tertinggi sebesar 90% saat data latih dan data uji memiliki komposisi yang sama, yaitu 50% untuk masing-masingnya.

Penelitian kelima yang dilakukan oleh (Ilmawan & Mude, 2020) bertujuan untuk membandingkan kinerja metode klasifikasi SVM dengan metode klasifikasi lainnya, yaitu Naïve Bayes. Metode klasifikasi Naïve Bayes dikenal sebagai metode ringan dalam pemrosesan dan memiliki tingkat akurasi yang tinggi saat diterapkan dalam klasifikasi teks, seperti yang telah terbukti dalam penelitian sebelumnya. Dalam penelitian ini, akurasi classifier diukur menggunakan metode K-fold cross validation dengan nilai K=3. Data yang digunakan adalah ulasan tekstual aplikasi pada Google Play Store berbahasa Indonesia, yang diambil dari penelitian sebelumnya. Hasil pengujian menggunakan metode 3-folds cross validation menunjukkan bahwa SVM Classifier memiliki tingkat akurasi yang lebih tinggi daripada Naïve Bayes classifier dalam mengklasifikasikan ulasan tekstual bahasa Indonesia pada Google Play Store. SVM classifier mencapai akurasi sebesar 81,46%, sedangkan Naïve Bayes classifier mencapai akurasi sebesar 75,41%.

Gap penelitian

1. Pemilihan media sosial yang lebih representatif: Penelitian ini fokus pada analisis sentimen di Instagram. Namun, mungkin ada kebutuhan untuk memperluas analisis ke platform media sosial lainnya seperti Twitter, Facebook, atau YouTube. Setiap platform memiliki karakteristik dan audiens yang berbeda, sehingga memperluas penelitian ke platform lain

dapat memberikan wawasan yang lebih komprehensif tentang sentimen terkait Program Kartu Prakerja.

2. Peningkatan jumlah data pelatihan: Metode Support Vector Machine (SVM) membutuhkan data pelatihan yang memadai untuk memberikan hasil yang akurat. Oleh karena itu, salah satu gap yang mungkin ada adalah kurangnya jumlah data pelatihan yang cukup dalam penelitian ini. Penelitian selanjutnya dapat berfokus pada pengumpulan dan penggunaan dataset yang lebih besar untuk meningkatkan kinerja SVM dalam analisis sentimen.
3. Penanganan bahasa yang tidak baku atau slang: Analisis sentimen dalam bahasa yang tidak baku atau menggunakan slang dapat menjadi tantangan tersendiri. Dalam penelitian ini, mungkin tidak ada fokus khusus pada bahasa tersebut. Oleh karena itu, gap penelitian dapat berfokus pada pengembangan teknik atau pendekatan yang dapat mengatasi variasi bahasa yang tidak baku atau slang dalam analisis sentimen.
4. Memperhatikan aspek multikelas: Penelitian ini mungkin lebih berfokus pada analisis sentimen biner (positif/negatif) terkait Program Kartu Prakerja. Namun, mungkin ada kebutuhan untuk memperluas analisis menjadi multikelas, di mana sentimen dapat dibagi menjadi lebih dari dua kategori (misalnya, positif, negatif, netral). Penelitian selanjutnya dapat mengeksplorasi metode yang lebih baik untuk mengatasi aspek multikelas dalam analisis sentimen.
5. Pemilihan fitur yang lebih canggih: Metode Support Vector Machine dapat dioptimalkan dengan memilih fitur-fitur yang lebih canggih atau representatif. Penelitian ini mungkin menggunakan fitur-fitur sederhana seperti kata-kata atau n-gram. Gap penelitian dapat melibatkan eksplorasi fitur-fitur yang lebih kompleks seperti fitur sintaksis, fitur semantik, atau fitur berbasis jaringan sosial untuk meningkatkan kinerja SVM dalam analisis sentimen.

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan metode support vector machine dalam menganalisis sentiment terhadap program kartu prakerja pada komentar media social Instagram
2. Mengetahui tingkat akurasi metode Support vector machine dalam melakukan klasifikasi sentiment.

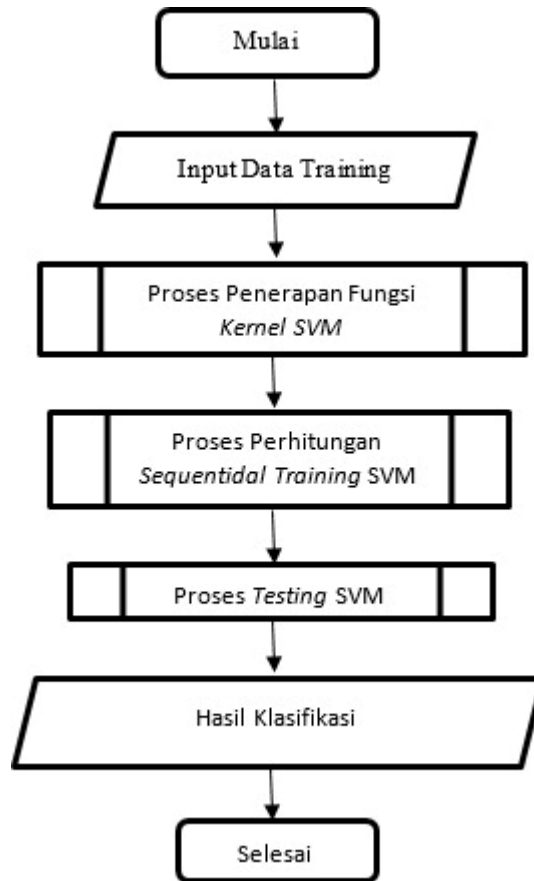
1.4. Urgensitas Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan wawasan tentang sentimen pengguna terhadap Program Kartu Prakerja di Instagram.
2. Menyediakan pemahaman yang lebih baik tentang persepsi publik terhadap program ini.
3. Memberikan kontribusi dalam pengembangan strategi dan pengambilan keputusan terkait Program Kartu Prakerja.

2. METODE PENELITIAN

2.1 Tahap Penelitian



Gambar 2.1 bagan alir

Berikut adalah bagan alir untuk analisis sentimen program Kartu Prakerja di Instagram menggunakan metode Support Vector Machine (SVM):

1. Input data training:
 - 1) Membagi dataset menjadi dua bagian: data latih (training data) dan data uji (testing data).
 - 2) Data latih digunakan untuk melatih model SVM, sedangkan data uji digunakan untuk menguji performa model.
2. Proses penerapan fungsi kamel SVM:
 - 1) Melakukan pelatihan model SVM menggunakan data latih.

- 2) Menentukan parameter optimal untuk SVM seperti fungsi kernel, C, dan gamma.
- 3) Melatih model dengan menggunakan data latih dan parameter yang telah ditentukan.
3. Proses perhitungan sequential training SVM:
 - 1) Menggunakan data uji untuk menguji performa model SVM yang telah dilatih.
 - 2) Menghitung metrik evaluasi seperti akurasi, presisi, recall dan F1-score.
4. Proses testing SVM:
 - 1) Mengevaluasi performa model SVM berdasarkan metrik evaluasi yang dihasilkan
 - 2) Jika performa model tidak memuaskan, dapat dilakukan penyesuaian parameter atau mencoba metode pembelajaran mesin lainnya
5. Hasil klasifikasi:
 - 1) Menggunakan model SVM yang telah dilatih untuk melakukan prediksi sentimen pada komentar baru di instagram
 - 2) Mengklasifikasikan komentar sebagai positif, negatif, atau netral berdasarkan prediksi sentimen

2.2 Program Kartu Prakerja

Program Kartu Prakerja adalah program yang bertujuan untuk mengembangkan kompetensi kerja bagi pencari kerja, pekerja/buruh yang mengalami pemutusan hubungan kerja, dan pekerja/buruh yang membutuhkan peningkatan kompetensi. Peserta program akan diberikan Kartu Prakerja sebagai identitas mereka. Pelatihan akan diselenggarakan melalui platform digital yang merupakan mitra resmi pemerintah dalam program ini. Peserta yang telah menyelesaikan pelatihan akan menerima sertifikat sebagai bukti tertulis yang diberikan oleh lembaga pelatihan (Setkab RI, 2020).

Pandemi Covid-19 telah mengubah pelaksanaan Program Kartu Prakerja menjadi salah satu bentuk bantuan sosial untuk mengatasi dampak pandemi. Oleh karena itu, tujuan Program Kartu Prakerja saat ini adalah mengembangkan kompetensi tenaga kerja, meningkatkan produktivitas dan daya saing mereka, serta mengembangkan kewirausahaan. Pelatihan yang diselenggarakan oleh Program Kartu Prakerja setelah pandemi Covid-19 difokuskan pada pemberian keterampilan dan pengetahuan yang relevan dengan dunia kerja atau kewirausahaan, peningkatan kompetensi kerja atau kewirausahaan, dan pemindahan kompetensi kerja (Pemerintah Republik Indonesia, 2020).

2.3 Pra-processing

2.1 Pengumpulan Data: Dilakukan scraping data komentar dari postingan akun Instagram Kartu Prakerja menggunakan teknik tertentu, seperti menggunakan API Instagram atau library web scraping. Data komentar tersebut kemudian disimpan dalam format yang sesuai, seperti file CSV atau spreadsheet.

2.2 Pembersihan Data: Tahap ini melibatkan pembersihan data yang mungkin mengandung karakter khusus, tautan, emoji, atau teks yang tidak relevan. Data yang tidak diperlukan juga dapat dihapus, seperti kolom waktu atau nama pengguna yang tidak relevan dengan analisis yang akan dilakukan.

2.3 Normalisasi Teks: Dilakukan normalisasi teks untuk mengubah semua huruf menjadi lowercase, menghapus tanda baca, dan menghilangkan spasi berlebih. Hal ini akan membantu dalam konsistensi analisis kata-kata.

2.4 Penghapusan Stop Words: Stop words adalah kata-kata umum yang tidak memberikan makna khusus dalam analisis teks, seperti "dan", "di", atau "dari". Menghapus stop words dapat membantu fokus pada kata-kata penting yang lebih relevan dalam analisis.

Tokenisasi: Data komentar perlu dipecah menjadi token-token terpisah, yaitu unit-unit terkecil seperti kata atau frasa. Proses

2.4 Media Sosial Instagram

Instagram adalah sebuah aplikasi yang dapat digunakan secara gratis pada telepon seluler Android dan iPhone. Dengan Instagram, pengguna dapat membagikan foto dan video kepada pengikutnya atau grup teman. Selain itu, pengguna juga dapat melihat, memberikan komentar, dan menyukai konten yang dibagikan oleh teman mereka di platform ini. Untuk membuat akun Instagram, seseorang hanya perlu berusia 13 tahun ke atas, mendaftar menggunakan alamat surel, dan memilih nama pengguna yang unik (Instagram.com, 2020a).

Instagram memiliki berbagai fitur menarik yang dapat digunakan oleh penggunanya, antara lain (Instagram.com, 2020b):

1. *Reels*

Fitur terbaru yang diperkenalkan oleh Instagram pada tanggal 5 Agustus 2020 adalah Reels. Reels memungkinkan pengguna untuk membuat dan merekam video singkat dengan durasi maksimal 15 detik. Video Reels ini dapat diunggah langsung ke feed pengguna dan jika pengguna memiliki akun publik, video Reels mereka dapat muncul di halaman Explore yang dapat dilihat oleh pengguna lain. Dengan adanya fitur Reels ini, pengguna Instagram dapat dengan mudah membuat dan berbagi konten video kreatif dengan durasi singkat.

2. *Stories*

Stories merupakan salah satu fitur yang tersedia di Instagram yang memungkinkan pengguna untuk dengan cepat berbagi momen dan pengalaman mereka. Dalam Stories, pengguna dapat menambahkan teks, musik, stiker, dan GIF untuk membuat konten yang lebih menarik. Stories hanya akan terlihat selama 24 jam sebelum menghilang secara otomatis. Namun, pengguna juga memiliki opsi untuk membagikan Stories secara permanen dengan menyimpannya dalam sorotan di profil mereka. Dengan fitur Stories ini, pengguna dapat berbagi momen sehari-hari secara

instan dan menambahkan kreativitas dalam konten yang mereka bagikan.

2.5 Metode CRISP-DM

Metode CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah sebuah pendekatan sistematis yang digunakan dalam proses penambangan data. CRISP-DM adalah standar industri yang umum digunakan untuk proyek-proyek analisis data dan penambangan data.

CRISP-DM terdiri dari enam tahap utama yang dilakukan secara berurutan, tetapi dapat pula dilakukan secara iteratif. Tahapan-tahapan tersebut adalah:

1. **Understanding Business Understanding:** Tahap ini melibatkan pemahaman terhadap tujuan bisnis dan kebutuhan pemilik proyek. Pada tahap ini, perlu didefinisikan pertanyaan bisnis yang ingin dijawab melalui analisis data.
2. **Data Understanding:** Tahap ini berfokus pada memahami data yang tersedia untuk analisis. Ini termasuk mengumpulkan data, menjelajahi dan memahami karakteristik data, serta mengevaluasi kualitas dan keterbatasan data.
3. **Data Preparation:** Pada tahap ini, data yang telah dikumpulkan dipersiapkan untuk analisis lebih lanjut. Hal ini melibatkan pembersihan data, penggabungan data dari berbagai sumber, pemilihan atribut yang relevan, dan transformasi data agar siap untuk pemodelan.
4. **Modeling:** Tahap ini melibatkan pengembangan model analisis data yang sesuai dengan tujuan bisnis yang telah ditetapkan. Model ini dapat berupa model statistik, model prediktif, atau model lainnya, tergantung pada jenis analisis yang ingin dilakukan.
5. **Evaluation:** Pada tahap ini, model yang dikembangkan dievaluasi berdasarkan kriteria yang telah ditetapkan sebelumnya. Evaluasi dilakukan untuk mengukur sejauh mana model dapat memenuhi tujuan bisnis, keakuratan, dan kinerja secara keseluruhan.

6. Deployment: Tahap terakhir ini melibatkan implementasi hasil dari analisis data ke dalam lingkungan bisnis. Ini bisa berarti mengintegrasikan model ke dalam sistem produksi, menyusun laporan hasil analisis, atau memberikan rekomendasi berdasarkan hasil yang ditemukan.

2.6 Dataset

Nama dataset: DatasetIG.CSV

Link unduh:

[https://raw.githubusercontent.com/dimasw09/BelajarKoding/main/](https://raw.githubusercontent.com/dimasw09/BelajarKoding/main/DatasetIG.CSV)

DatasetIG.CSV

No	Nama Atribut	Type data
1	Id	int
2	sentimen	string
3	komentar	string

Rumus SVM

Rumus dasar untuk Support Vector Machine (SVM) tergantung pada apakah kita berurusan dengan masalah klasifikasi linier atau non-linier.

Berikut adalah rumus dasar untuk klasifikasi SVM linier:

1. Rumus Hyperplane:

Hyperplane dalam SVM digunakan untuk memisahkan dua kelas data. Dalam kasus klasifikasi biner, hyperplane didefinisikan sebagai: $w^T \cdot x + b = 0$

di mana w adalah vektor normal ke hyperplane, x adalah vektor fitur data, dan b adalah bias.

2. Fungsi Keputusan:

Fungsi keputusan digunakan untuk memprediksi kelas data baru berdasarkan letaknya terhadap hyperplane. Dalam kasus SVM linier, fungsi keputusan didefinisikan sebagai:

$$f(x) = \text{sign}(w^T \cdot x + b)$$

di mana sign adalah fungsi tanda, yang mengembalikan nilai 1 jika $w^T \cdot x + b \geq 0$, dan -1 jika $w^T \cdot x + b < 0$.

3. Maksimisasi Margin:

SVM juga berusaha untuk mencari hyperplane dengan margin maksimal antara dua kelas. Margin didefinisikan sebagai jarak antara hyperplane dan titik terdekat dari kedua kelas. Untuk mencapai margin maksimal, SVM memaksimalkan $2/\|w\|$, di mana $\|w\|$ adalah norma Euclidean dari vektor w .

4. Fungsi Objektif SVM:

Fungsi objektif SVM bertujuan untuk mencapai pemisahan yang baik antara kelas dan margin maksimal. Dalam kasus SVM linier dengan fungsi kernel linier, fungsi objektif dinyatakan sebagai:

$$\min (1/2) * \|w\|^2$$

dengan kendala $y_i(w^T * x_i + b) \geq 1$, di mana y_i adalah label kelas yang diberikan pada data x_i .

2.7 Analisis Hasil dan Penerapan

Dengan menggunakan metode split 2:8, hasil yang diperoleh menunjukkan keandalan yang tinggi. Akurasi rata-ratanya adalah 0.6967049241021843, mengindikasikan tingkat keakuratan yang sangat baik. Selain itu, presisi rata-rata mencapai 0.81565173751448 26, menunjukkan konsistensi dan ketepatan model. Recall rata-ratanya adalah 0.6376154289466983, menunjukkan kemampuan model dalam mengenali dan mengingat informasi yang relevan. Selain itu, F-Measure rata-ratanya adalah 0.64759958154563 27, mencerminkan keseimbangan yang baik antara presisi dan recall. Dengan demikian, evaluasi menunjukkan bahwa model ini dapat diandalkan dan memberikan hasil yang meyakinkan.

Persiapan Data

```
[ ] dataUrl = 'https://raw.githubusercontent.com/dimasw09/BelajarKoding/main/DatasetIG.csv'
```

Cleaning

```
[ ] # membersihkan karakter data agar data dapat diolah
import re
import emoji
import contractions

def cleaning(text):
    text = re.sub('RT|&', '', text) # replace RT tag
    text = re.sub('@[a-zA-Z0-9]+', '', text) # replace @username
    text = emoji.demojize(text) # replace emoji with text
    text = re.sub('http|https://V|/S+', '', text) # replace URL
    text = re.sub('fr', '', text) # replace something
    text = text.lower() # lower case kalimat
    text = re.sub(r'(\w+)', r'\1', text) # replace word repetition
    text = re.sub(r'([^\w|\s|\d|\-|_])', '', text)
    text = re.sub(r'([0-9])', '', text)
    text = re.sub(r'([^\w|\s|\d|\-|_])+=-|\_|\_+', '', text)
    text = re.sub(r'[-a-zA-Z]', '', text)
    text = contractions.fix(text) # replace contractions

    return text

df[['Cleaning']] = df[['komentar']].apply(cleaning, axis=1)

df
```

Tokenize

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
```

```
slangword_url = "https://raw.githubusercontent.com/dinas09/BelaJawoding/main/slang_word.txt"
response = requests.get(slangword_url)
kamusSlang = eval(response.text)

def convertToSlangword(ulasan):
    pattern = re.compile(r"\\b" + "'" + ".join(kamusSlang.keys()) + r'"\\b")
    content = []
    for kata in ulasan:
        filterSlang = pattern.sub(lambda x: kamusSlang[x.group()], kata)
        content.append(filterSlang.lower())
    return content

# Assuming 'df' is your DataFrame with a 'Tokenizing' column
df['Formalisasi'] = df['Tokenizing'].apply(convertToSlangword)
df
```

Stopword Removal

```
from nltk.corpus import stopwords
nltk.download("stopwords")

daftar_stopword = stopwords.words("indonesian")
daftar_stopword = pd.read_csv("https://raw.githubusercontent.com/dimasu09/BalajarCoding/main/stopword_list_tela_coba.csv")
daftar_stopword = set(daftar_stopword)

def stopwordsText(words):
    return [word for word in words if word not in daftar_stopword]

df['Stopword'] = df['formalIsaai'].apply(stopwordsText)
df
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data..  
[nltk_data] Package stopwords is already up-to-date!
```

Stemming

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

factory = StemmerFactory()
stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}

for document in df['Stopword']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ' '

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])

def stemmingText(document):
    return [term_dict[term] for term in document]

df['Stemming'] = df['Stopword'].swifter.apply(stemmingText)
```

yah : yah
semoga : moga
dan : dan
baru : baru
pertama : pertama
kali : kali
coba : coba
soalnya : soal
bismillaahh : bismillaahh
jumat : jumat
berkah : berkah
kasih : kasih
redaksi : redaksi
heart : heart
buat : buat
sobat : sobat
prakerja : prakerja
yang : yang
mau : mau
rejekinya : rejekinya
di : di
gelombang : gelombang

2.8 Analisis Hasil akhir dan Evaluasi

1. Akurasi: Akurasi mengukur seberapa baik model dapat mengklasifikasikan data dengan benar. Hasil akurasi rata-rata sebesar 0.6967049241021843 menunjukkan bahwa model memiliki tingkat keakuratan yang sangat baik dalam memprediksi kelas yang benar.
2. Presisi: Presisi menggambarkan kemampuan model dalam memberikan hasil positif yang benar dari semua prediksi yang positif. Hasil presisi rata-rata sebesar 0.8156517375144826 menunjukkan bahwa model konsisten dan akurat dalam mengidentifikasi dan mengklasifikasikan data dengan benar.
3. Recall: Recall mengukur kemampuan model dalam mengenali dan mengingat informasi yang relevan. Hasil recall rata-rata sebesar 0.6376154289466983 menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali data yang relevan dan

menghindari kesalahan dalam mengklasifikasikan data yang sebenarnya positif.

4. F-Measure: F-Measure adalah ukuran yang menggabungkan presisi dan recall untuk memberikan gambaran keseluruhan tentang kinerja model. Hasil F-Measure rata-rata sebesar 0.6475995815456327 menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall.

Uji Coba

```
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn import model_selection
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

X = df['Cleaning']
Y = df['Sentimen']

x_train, x_test, y_train, y_test = model_selection.train_test_split(X, Y, test_size=0.1)

[19] vectorizer = TfidfVectorizer()
x_train = vectorizer.fit_transform(x_train)
x_test = vectorizer.transform(x_test)
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)
```

Modeling

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import AdaBoostClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.preprocessing import MinMaxScaler
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score
```

Support Vector MachineSplit 2:8

```
# SVM Model
svc = SVC()
svc.fit(x_train, y_train)
pred5 = svc.predict(x_test)
print(classification_report(y_test, pred5))
```

	precision	recall	f1-score	support
0	0.70	0.91	0.79	46
1	0.80	0.46	0.59	26
2	0.82	0.70	0.76	20
accuracy			0.74	92
macro avg	0.77	0.69	0.71	92
weighted avg	0.76	0.74	0.73	92

```
[34] acc_score = cross_val_score(svc, x_train, y_train, cv=5, scoring='accuracy')
pre_score = cross_val_score(svc, x_train, y_train, cv=5, scoring='precision_macro')
rec_score = cross_val_score(svc, x_train, y_train, cv=5, scoring='recall_macro')
f_score = cross_val_score(svc, x_train, y_train, cv=5, scoring='f1_macro')

print('Hasil Accuracy : %s' % (acc_score))
print('Hasil Rata - Rata Accuracy : %s' % acc_score.mean())
print('Hasil Precision : %s' % (pre_score))
print('Hasil Rata - Rata Precision : %s' % pre_score.mean())
print('Hasil Recall : %s' % (rec_score))
print('Hasil Rata - Rata Recall : %s' % rec_score.mean())
print('Hasil F-Measure : %s' % (f_score))
print('Hasil Rata - Rata F-Measure : %s' % f_score.mean())
```

```
Hasil Accuracy : [0.7027027 0.68493151 0.68493151 0.68493151 0.7260274 ]
Hasil Rata - Rata Accuracy : 0.6967049241021843
Hasil Precision : [0.76619133 0.8308642 0.8452381 0.81908832 0.81687675]
Hasil Rata - Rata Precision : 0.8156517375144826
Hasil Recall : [0.64513557 0.62626263 0.62169312 0.62247121 0.67251462]
Hasil Rata - Rata Recall : 0.6376154289466983
Hasil F-Measure : [0.65877671 0.65304483 0.60601852 0.63470729 0.68545056]
Hasil Rata - Rata F-Measure : 0.6475995815456327
```

DAFTAR PUSTAKA

- Emeraldi, M. A., Ekawati, I., & Sumadyo, M. (2022). Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes. *Journal of Students' Research in Computer Science*, 3(1), 109–122. <https://doi.org/10.31599/jsrscs.v3i1.1513>
- Hendrastuty, N., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., Isnain, A. R., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <http://situs.com>
- RIZWAN, M. R. (2022). Analisa Sentimen Terhadap Program Kartu Prakerja Pada Instagram Menggunakan Metode Naïve Bayes Classifier. *Tugas Akhir UIN*.
- Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59–67. <https://doi.org/10.47080/simika.v5i1.1411>
- Anggraini, W. P., & Utami, M. S. (2021). KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN KARTU PEKERJA DI INDONESIA. *Faktor Exacta*, 13(4), 255–261.
- Athira Luqyana, W., 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. <http://j-ptiik.ub.ac.id>
- Gao, X. (2014). Effects of cyber public opinions on government decision-making. *2014 2nd International Conference on Education Technology and Information System (ICETIS 2014)*, 109–112.
- Hendrastuty, N., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., Isnain, A. R., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., & Isnain, A. R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <http://situs.com>
- Ilmawan, L. B., & Mude, M. A. (2020). Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. *ILKOM Jurnal Ilmiah*, 12(2), 154–161. <https://doi.org/10.33096/ilkom.v12i2.597.154-161>
- Instagram.com. (2020a). *Apa itu instagram?*

- <https://help.instagram.com/424737657584573>
- Instagram.com. (2020b). *Instagram Features / Discover New Features on Stories, Shopping & More*. Instagram Features %7C Discover New Features on Stories,%0AShopping & More
- Kelvin, K., Banjarnahor, J., -, E. I., & NK Nababan, M. (2022). Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM). *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 5(2), 47–52.
<https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365>
- Liputan6.com. (2020). *Temuan BPS: Program Kartu Prakerja Salah Sasaran*.
<https://www.liputan6.com/bisnis/read/4419163/temuan-bps-program-kartu-prakerja-salah-sasaran>
- Liu, B. (2011). *Web data mining: exploring hyperlinks, contents, and usage data* (Vol. 1). Springer.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
- Pemerintah Republik Indonesia. (2020). *Peraturan Presiden Republik Indonesia Nomor 76 Tahun 2020 Tentang Perubahan Atas Peraturan Presiden Nomor 36 Tahun 2020 Tentang Pengembangan Kompetensi Kerja Melalui Program Kartu Prakerja*, (No. 039037).
- Pikiran-rakyat.com. (2020). *Netizen Ramai Keluhkan Daftar Gelombang 1-11 Tak Kunjung Lolos Kartu Prakerja, Begini Kata DPR RI*.
- Power, D. J., & Phillips-Wren, G. (2012). *Impact of social media and Web 2.0 on decision-making*. *J Decis Syst* 20 (3): 249–261.
- Prakerja.go.id. (2020). *Kartu Prakerja`*. <https://www.prakerja.go.id/>
- Riadi, C. J. (2019). Analisis Sentimen Pada Media Daring Tentang Pemilihan Presiden Indonesia Tahun 2019 Menggunakan Metode Naive Bayes. *EProceedings of Engineering*, 6(2).
- Setkab RI. (2020). *Pengembangan Kompetensi Kerja Melalui Program Kartu Prakerja*.
- Tirto.id. (2020). *Masalah Kartu Prakerja: Gelombang IV Ditunda, Insentif Dievaluasi*. <https://tirto.id/masalah-kartu-prakerja-gelombang-iv-ditunda-insentif-dievaluasi-fJzj>

