ANALISIS SENTIMEN ANCAMAN DISRUPSI PEKERJAAN AKIBAT KECERDASAN BUATAN DI MEDIA SOSIAL YOUTUBE DENGAN ALGORITMA SVM

Putri Maharani, Salsabila Alfath Annisaa', Dinda Aulika Elfarin Aritonang, Fathoni, Ali Ibrahim
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang – Prabumulih No.KM. 32, Indralaya Indah, Kec.
Indaralaya, Kabupaten Ogan Ilir, Sumatera Selatan, Indonesia
ptrm4hrnii@gmail.com

ABSTRAK

Meningkatnya penggunaan kecerdasan buatan berpotensi menimbulkan disrupsi pekerjaan dan memicu kekhawatiran masyarakat terhadap masa depan dunia kerja. Opini publik mengenai isu tersebut banyak muncul di media sosial, khususnya YouTube, dalam bentuk komentar tidak terstruktur sehingga membutuhkan analisis sentimen secara sistematis. Penelitian ini bertujuan mengkaji sentimen publik terhadap disrupsi pekerjaan akibat kecerdasan buatan menggunakan algoritma Support Vector Machine (SVM). Data penelitian berupa komentar YouTube berbahasa Indonesia yang dikumpulkan dari Kaggle. Tahapan penelitian meliputi praproses teks, ekstraksi fitur menggunakan TF-IDF, serta klasifikasi sentimen dengan SVM menggunakan kernel Linear dan Radial Basis Function (RBF). Hasil penelitian menunjukkan bahwa kedua model memperoleh akurasi sebesar 75%. Kernel Linear menghasilkan klasifikasi yang lebih optimal pada kelas netral dan positif, sedangkan kernel RBF masih memiliki keterbatasan dalam mengidentifikasi kelas negatif. Penelitian ini menunjukkan bahwa penyaringan data berbasis konteks pekerjaan serta optimasi kernel SVM dapat membantu memetakan sentimen publik secara lebih relevan terhadap isu disrupsi pekerjaan akibat kecerdasan buatan.

Kata kunci : *Analisis Sentimen, Kecerdasan Buatan, Disrupsi Pekerjaan, YouTube, Support Vector Machine.*

1. PENDAHULUAN

Kehadiran Kecerdasan Buatan (AI) di dunia kerja telah memicu kekhawatiran yang meluas di masyarakat, di mana fenomena ini dikenal secara luas dengan istilah *AI Anxiety* [1], [2]. Fenomena ini menyebabkan banyaknya opini publik tersaji di platform YouTube dalam bentuk teks yang tidak terstruktur dan tidak dievaluasi secara sistematis [3]. Masalah nyata yang muncul adalah tingginya volume komentar dengan beragam sudut pandang yang menghambat proses pemetaan persepsi secara cepat, serta adanya opini yang bercampur aduk mengenai apakah AI akan menggantikan atau sekadar membantu pekerja [4]. Hal ini menyulitkan sistem dalam menemukan pola sentimen publik yang bermakna dari sekumpulan data yang sangat besar.

Dampak dari permasalahan tersebut sangat krusial, terutama pada ketidakjelasan informasi mengenai tingkat kecemasan publik yang sebenarnya. Banyaknya opini yang tidak terfilter mengakibatkan terjadinya bias informasi, di mana masyarakat justru mengalami kesulitan dalam menyaring isu ancaman pekerjaan yang relevan [5]. Secara teknis, analisis sentimen tanpa metode ekstraksi yang tepat menyebabkan ketidakakuratan dalam mengklasifikasikan polaritas opini, sebagaimana terlihat pada penggunaan dataset umum yang belum mampu menangkap konteks spesifik mengenai isu ketenagakerjaan [6].

Penelitian sebelumnya telah menggunakan beragam pendekatan untuk menangani persoalan ini, salah satunya melalui penerapan algoritma *machine learning* dalam analisis sentimen opini publik terhadap AI. [7]. Namun, metode tersebut memiliki keterbatasan karena sering kali menggunakan dataset terbuka secara langsung, sehingga memuat topik yang terlalu luas dan tidak relevan dengan isu karier. Penggunaan data media sosial juga menunjukkan bahwa sentimen publik tidak selalu fokus pada isu pekerjaan, melainkan bercampur dengan opini lain terkait fungsi dan performa AI [6]. Selain itu, penggunaan algoritma klasifikasi telah banyak dikembangkan, namun sering kali belum optimal dalam menemukan tingkat akurasi terbaik karena tidak melalui tahapan eksplorasi konfigurasi parameter secara mendalam [8], [9]. Terdapat *research gap* di mana optimasi algoritma Support Vector Machine (SVM) serta penyaringan data secara khusus masih belum banyak diterapkan pada komentar YouTube untuk mengukur kecemasan kehilangan pekerjaan secara spesifik.

Sebagai solusi, penelitian ini menerapkan algoritma Support Vector Machine (SVM) dengan teknik penyaringan data yang difokuskan pada konteks pekerjaan untuk melakukan klasifikasi sentimen secara otomatis. Pendekatan ini mencakup penyaringan komentar menggunakan kata kunci yang berkaitan dengan isu pekerjaan, seperti *job*, *work*, dan *career*, serta evaluasi terhadap dua kernel SVM, yakni Linear dan Radial Basis Function (RBF), untuk memperoleh model klasifikasi yang paling sesuai [8], [10]. Tujuan penelitian ini adalah menghasilkan sistem klasifikasi sentimen yang efisien dan akurat dalam memetakan respons publik terhadap ancaman disrupsi pekerjaan akibat kecerdasan buatan. Kebaruan penelitian ini terletak pada penerapan penyaringan data berbasis konteks pekerjaan pada komentar YouTube yang awalnya bersifat umum, serta optimasi

model SVM melalui perbandingan kernel untuk memperoleh hasil klasifikasi yang lebih relevan dengan isu yang dikaji.

2. TINJAUAN PUSTAKA

Penggunaan kecerdasan buatan (AI) yang semakin masif di berbagai sektor menjadi elemen krusial dalam proses transformasi dunia kerja [1]. Teknologi ini tidak hanya dimanfaatkan untuk meningkatkan efisiensi, tetapi juga mendorong perubahan signifikan pada pelaksanaan berbagai jenis pekerjaan rutin. Fenomena otomatisasi tersebut berpotensi mengurangi kebutuhan tenaga manusia pada bidang tertentu, sehingga memicu kekhawatiran terkait job displacement di masyarakat [2]. Perkembangan teknologi ini rupanya berdampak pada kondisi psikologis individu melalui munculnya bentuk kecemasan baru terhadap ketidakpastian karier [3]. Fenomena tersebut dikenal sebagai *AI anxiety* atau *algorithmic anxiety*, yang mencerminkan respons emosional negatif terhadap potensi kehilangan pekerjaan akibat otomatisasi [4]. Selain itu, persepsi terhadap risiko ini sangat dipengaruhi oleh faktor sosial dan demografis, sehingga respons setiap individu terhadap AI cenderung bersifat heterogen [5].

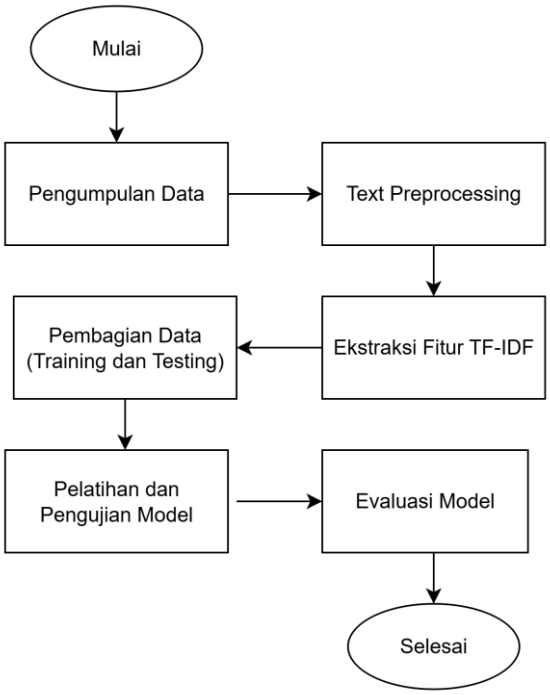
Dinamika respons publik dapat diamati secara mendalam melalui analisis sentimen terhadap opini masyarakat di ruang digital [6], [7]. Pendekatan ini efektif untuk mengidentifikasi kecenderungan sikap masyarakat, baik yang mendukung maupun yang merasa terancam oleh perkembangan teknologi. Dalam konteks penggantian peran kerja oleh AI secara langsung, sentimen negatif sering kali muncul lebih dominan sebagai bentuk reaksi terhadap ancaman otomatisasi tersebut [8]. Platform media sosial saat ini menjadi instrumen utama dalam menggambarkan opini publik secara luas dan transparan. YouTube memfasilitasi pengguna untuk menyampaikan pandangan secara terbuka melalui kolom komentar, sehingga mampu mencerminkan pola opini terhadap isu tertentu [7]. Karakteristik komentar yang spontan dan ekspresif menjadikan data tersebut sangat relevan untuk mengkaji sentimen publik terkait isu disrupsi pekerjaan secara lebih mendalam.

Guna mengolah data teks bervolume besar tersebut, algoritma Support Vector Machine (SVM) mempunyai kapabilitas yang sangat baik dalam mengolah data berdimensi tinggi [9]. Berbagai penelitian membuktikan bahwa SVM mampu menghasilkan performa klasifikasi yang tinggi pada data komentar YouTube, bahkan sering kali mengungguli metode klasifikasi lainnya [10]. Akurasi model ini juga dapat dioptimalkan melalui tahapan preprocessing, seperti normalisasi bahasa informal, yang terbukti meningkatkan kualitas klasifikasi secara signifikan [11]. Penelitian ini dilakukan untuk mengkaji sentimen terhadap ancaman disrupsi pekerjaan akibat kecerdasan buatan dengan memanfaatkan data komentar YouTube dan klasifikasi berbasis SVM. Kebaruan penelitian ini terletak pada dua aspek utama. Pertama, penyaringan data dilakukan secara kontekstual agar analisis tidak menggunakan seluruh komentar secara umum, melainkan difokuskan pada komentar yang berkaitan dengan isu pekerjaan dan potensi disrupsi kerja akibat AI. Kedua, penelitian ini membandingkan penggunaan kernel Linear dan Radial Basis Function (RBF) pada algoritma SVM untuk mendapatkan model klasifikasi yang paling selaras dengan karakteristik data. Melalui pendekatan tersebut, penelitian ini diharapkan mampu menghasilkan analisis yang lebih terarah, relevan, dan tepat dengan konteks ancaman pekerjaan akibat kecerdasan buatan.

3. METODE PENELITIAN

Pada metodologi penelitian dijelaskan tahapan diterapkan dalam proses analisis tanggapan di YouTube yang berkaitan dengan risiko disrupsi pekerjaan akibat kecerdasan buatan. Penelitian ini memanfaatkan algoritma Support Vector Machine (SVM) sebagai metode pengklasifikasian untuk menentukan polaritas sentimen. Adapun tahapan penelitian mencakup pengumpulan data, pra pemrosesan teks, ekstraksi fitur dengan TF-IDF, pemisahan data menjadi set pelatihan dan set pengujian, pelatihan serta pengujian model SVM, dan evaluasi performa model. Setiap tahapan dirancang secara terstruktur agar dapat menghasilkan model klasifikasi sentimen yang akurat dan relevan dengan tujuan penelitian.

Alur proses penelitian dapat terlihat pada gambar 1.



Gambar 1. Diagram Alur penelitian

3.1. Pengumpulan Data

Dalam studi ini menggunakan informasi yang dikumpulkan dari komentar pengguna YouTube yang berkaitan dengan kecerdasan buatan dan dampaknya terhadap dunia kerja. Data diperoleh dari dataset komentar YouTube yang telah tersedia, dengan jumlah sebanyak 1.237 data, serta diolah menggunakan Python dan YouTube Data API.

Dataset tersebut terdiri atas komentar berbahasa Indonesia yang memuat opini masyarakat mengenai ancaman kecerdasan buatan terhadap pekerjaan. Data telah disusun dalam bentuk terstruktur dan dilabeli kedalam tiga kelompok, yaitu positif, negatif, dan netral, agar dapat langsung digunakan dalam pendekatan *supervised learning* pada tahap klasifikasi.

3.2. Text Preprocessing

Tahap *text preprocessing* dilakukan untuk membersihkan serta menyiapkan data komentar YouTube sebelum proses ekstraksi fitur dilakukan. Tahapan ini penting karena mutu data akan mempengaruhi kinerja model dalam menjalankan klasifikasi sentimen. Proses *preprocessing* yang diterapkan mencakup beberapa tahapan berikut:

a. *Cleansing* : Penghapusan unsur yang tidak penting seperti tanda baca, angka, URL, simbol, emoji, serta karakter khusus lainnya.
b. *Case Folding* : Mengubah seluruh huruf pada teks menjadi huruf kecil (*lowercase*) untuk menghindari variasi penulisan kata.
c. *Normalization* : Proses penyesuaian kata tidak baku atau istilah tertentu menjadi bentuk baku sesuai kaidah bahasa Indonesia.
d. *Tokenizing* : Tahapan memisahkan teks menjadi unit-unit kata (*token*) agar setiap kata dapat dianalisis secara terpisah.
e. *Stopword Removal* : Menghilangkan istilah-istilah yang tidak memiliki arti penting dalam proses penglasifikasian, seperti kata hubung.
f. *Stemming* : Tahapan memodifikasi kata berimbuhan menjadi format dasar guna mengurangi variasi kata.

3.3. Ekstraksi Fitur TF-IDF

Setelah melewati fase preprocessing, proses ekstraksi ciri dilakukan dengan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini diterapkan agar bisa mengonversi data berbentuk teks menjadi berbentuk angka hingga dapat diolah oleh algoritma machine learning.

TF-IDF memberi pembobotan di tiap-tiap kata dengan mempertimbangkan frekuensi kemunculannya dalam suatu dokumen serta tingkat keunikannya di seluruh dataset. Istilah yang sering muncul dalam suatu dokumen, tetapi jarang terdapat dalam dokumen lain akan mendapatkan nilai lebih besar. Hasil dari langkah ini berupa representasi matriks fitur numerik yang selanjutnya akan dipakai sebagai masukan pada model klasifikasi SVM. Implementasi ekstraksi fitur dilakukan menggunakan TfidfVectorizer yang ada dalam library Scikit-learn.

3.4. Pemisahan Data

Data yang sudah melewati tahap preprocessing dan ekstraksi fitur kemudian dibagi menjadi dua kelompok, yaitu data latih sebesar 80% dan data uji sebesar 20%. Data latih dimanfaatkan untuk melatih model agar dapat mengidentifikasi pola dalam data, sementara data uji memiliki tujuan agar menilai kemampuan model dalam mengklasifikasikan data yang belum pernah diproses sebelumnya.

3.5. Pelatihan dan Pengujian Model

Proses pelatihan model dilakukan dengan menggunakan algoritma Support Vector Machine (SVM). Model SVM diterapkan karena mempunyai kapabilitas yang baik dalam mengolah data teks berdimensi tinggi dan mampu menghasilkan klasifikasi yang konsisten.

Model dilatih memakai data *training* untuk mempelajari pola sentimen dari komentar YouTube. Setelah tahap pelatihan selesai, model kemudian diuji menggunakan data uji guna memperkirakan label sentimen. Selanjutnya, hasil estimasi tersebut dibandingkan dengan label yang sebenarnya agar menilai seberapa baik tingkat performa model dalam melakukan klasifikasi sentimen.

3.6. Evaluasi Model

Tahap penilaian model dilakukan dengan memakai *confusion matrix* sebagai alat ukur agar bisa menilai kinerja model dalam mengelompokkan sentimen komentar YouTube dalam tiga kategori, yaitu positif, negatif, dan netral. Berdasarkan hasil *confusion matrix*, dilakukan perhitungan berbagai metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy berfungsi untuk mengevaluasi tingkat ketepatan model secara menyeluruh, *precision* berfungsi untuk menilai seberapa tepat prediksi di tiap kelas, sedangkan *recall* merepresentasikan kapasitas model dalam mengidentifikasi seluruh data pada setiap kategori. *F1-score* adalah metrik yang menunjukkan keseimbangan antara *precision* dan *recall*. Metrik tersebut menjadi acuan dalam mengevaluasi performa model Support Vector Machine (SVM) dalam menjalankan pengklasifikasian sentimen dengan optimal.

4. HASIL DAN PEMBAHASAN

Studi ini menggunakan data sebanyak 1.237 data komentar Youtube yang diambil dari website dataset Kaggle.

4.1. Pengumpulan Data

Langkah awal dari studi ini adalah pengumpulan dan pemilihan data yang akan digunakan pada proses analisis sentimen. Data yang digunakan berasal dari dataset Sentiment Analysis ChatGPT YouTube

Comments pada situs Kaggle. Dataset tersebut diperoleh melalui teknik web scraping pada video YouTube berjudul "ChatGPT dan Masa Depan Pekerjaan Kita" dan terdiri atas 1.237 komentar. Data yang diambil berisi atribut komentar (comment) sebagai data teks dan sentimen (sentimen) sebagai label kelas, seperti yang terlihat pada Tabel 1.

Tabel 1. Pengumpulan Data

No	Comment	Sentimen
1	Luar biasa pak dampak chatGPT ini bagi mahasiswa, saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya. Proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatGPT ini	Positif
2	Tapi ingat bahwa chat gpt, berdampak buruk terutama mampu membuat puisi cinta dan juga mampu membuat mahasiswa lulus tes pekerjaan hanya pakai Ai terlebih jika makin banyak yang salah gunakan chat gpt ini sehebat apapun manusia Bakalan terbongkar kalau dia pakar palsu	Negatif

4.2. *Data Preprocessing*

Dalam *Pre-Processing*, terdapat beberapa tahapan yang perlu dilakukan, yaitu *;*

4.3. *Cleansing*

Sebelum dilakukan cleansing, teks masih mengandung tanda titik dan pemisah kalimat, seperti pada kalimat "Luar biasa pak dampak chatGPT ini" Setelah proses cleansing, tanda baca dihilangkan sehingga teks menjadi rangkaian kalimat bersih tanpa simbol tambahan.

Tabel 2. *Cleansing*

Sebelum *Cleaning*	Sesudah *Cleaning*
Luar biasa pak dampak chatGPT ini bagi mahasiswa, saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya. Proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatGPT ini	Luar biasa pak dampak chatGPT ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya Proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatGPT ini

4.4. *Case Folding*

Di fase ini dilakukan proses *case folding*, yaitu proses normalisasi tulisan dengan mengganti seluruh huruf jadi huruf kecil.

Tabel 3. *Case Folding*

Sebelum Case Folding	Sesudah Case Folding
Luar biasa pak dampak chatGPT ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya Proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatGPT ini	luar biasa pak dampak chatgpt ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatgpt ini

4.5. *Normalization*

Normalization dilakukan dengan menyelaraskan kata yang tidak tepat menjadi bentuk yang tepat dan mengikuti aturan bahasa Indonesia.

Tabel 4. *Normalization*

Sebelum Normalization	Sesudah Normalization
luar biasa pak dampak chatgpt ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya proses pendidikan konvensional bakal tidak relevan lagi setelah adanya chatgpt ini	luar biasa pak dampak chatgpt ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya proses pendidikan konvensional akan tidak relevan lagi setelah adanya chatgpt ini

4.6. *Tokenize*

Tokenize melakukan pemecahan teks menjadi unit-unit kata tunggal, seperti "luar" "biasa" "pak" "dampak" "chatGPT" "ini", dan seterusnya.

Tabel 5. *Tokenize*

Sebelum tokenize	Sesudah tokenize
luar biasa pak dampak chatgpt ini bagi mahasiswa saya rasa dunia pendidikan akan dipaksa berkembang ke arah yang tidak pernah kita pikirkan sebelumnya proses pendidikan konvensional akan tidak relevan lagi setelah adanya chatgpt ini	['luar', 'biasa', 'pak', 'dampak', 'chatgpt', 'ini', 'bagi', 'mahasiswa', 'saya', 'rasa', 'dunia', 'pendidikan', 'akan', 'dipaksa', 'berkembang', 'ke', 'arah', 'yang', 'tidak', 'pernah', 'kita', 'pikirkan', 'sebelumnya', 'proses', 'pendidikan', 'konvensional', 'akan', 'tidak', 'relevan', 'lagi', 'setelah', 'adanya', 'chatgpt', 'ini']

4.7. *Stopward Removal*

Pada tahap ini, kata-kata seperti "dan", "di", "ke", dan "pada" dihapus karena tidak memiliki peran penting dalam proses penentuan sentimen. Setelah tahap ini, tersisa kata-kata yang lebih substantif dan relevan terhadap isi komentar.

Tabel 6. Hasil *stopword removal*

Sebelum stopword removal	Setelah stopword removal
['luar', 'biasa', 'pak', 'dampak', 'chatgpt', 'ini', 'bagi', 'mahasiswa', 'saya', 'rasa', 'dunia', 'pendidikan', 'akan', 'dipaksa', 'berkembang', 'ke', 'arah', 'yang', 'tidak', 'pernah', 'kita', 'pikirkan', 'sebelumnya', 'proses', 'pendidikan', 'konvensional', 'akan', 'tidak', 'relevan', 'lagi', 'setelah', 'adanya', 'chatgpt', 'ini']	['dampak', 'chatgpt', 'mahasiswa', 'dunia', 'pendidikan', 'dipaksa', 'berkembang', 'arah', 'pikirkan', 'proses', 'pendidikan', 'konvensional', 'relevan', 'chatgpt']

4.8. *Stemming*

Stemming diimplementasikan dengan mereduksi kata berimbuhan ke bentuk dasarnya, seperti "pekerjaan" menjadi "kerja", "menggantikan" menjadi "ganti", dan "berkembang" menjadi "kembang". Tujuan dari langkah ini adalah mempermudah variasi bentuk kata pada komentar YouTube sehingga representasi fitur menjadi lebih ringkas dan dapat membantu model Support Vector Machine dalam melakukan klasifikasi sentimen secara lebih efektif.

Tabel 7. Hasil *stemming*

Sebelum stopword removal	Setelah stopword removal
['dampak', 'chatgpt', 'mahasiswa', 'dunia', 'pendidikan', 'dipaksa', 'berkembang', 'arah', 'pikirkan', 'proses', 'pendidikan', 'konvensional', 'relevan', 'chatgpt']	['dampak', 'chatgpt', 'mahasiswa', 'dunia', 'didik', 'paksa', 'kembang', 'arah', 'pikir', 'proses', 'didik', 'konvensional', 'relevan', 'chatgpt']

4.9. Ekstraksi Fitur TF-IDF

Pada tahap ini dilakukan ekstraksi fitur dengan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk mengubah data teks komentar menjadi bentuk numerik. Komentar YouTube yang telah melalui tahap preprocessing kemudian ditransformasikan menggunakan TfidfVectorizer, sehingga setiap komentar dapat direpresentasikan dalam bentuk vektor berdasarkan bobot kata yang muncul. Dalam penelitian ini, jumlah fitur terbatas, yaitu sampai 5000 kata untuk mendapatkan representasi teks yang lebih terarah dan efisien.

Hasil representasi TF-IDF menunjukkan bahwa data komentar memiliki keragaman kosakata yang cukup tinggi. Pemberian bobot pada setiap kata membantu model dalam mengenali istilah-istilah yang lebih relevan terhadap isi komentar. Dengan demikian, hasil ekstraksi fitur dapat menunjang proses pengklasifikasian sentimen menggunakan algoritma Support Vector Machine (SVM) secara lebih optimal.

4.10. Pemisahan Data

Setelah tahapan pra pemrosesan dan ekstraksi fitur TF-IDF, dataset dipisahkan menjadi dua subset, yaitu data latih dan data uji dengan perbandingan 80:20. Pemisahan data ini dilakukan agar model bisa mengenali pola sentimen dari data latih, kemudian diuji menggunakan data yang belum pernah dipakai sebelumnya.

Berdasarkan hasil penyaringan data, jumlah dataset yang digunakan dalam penelitian ini berjumlah 371 komentar. Dari total tersebut, sebanyak 296 data dimanfaatkan sebagai data latih, sedangkan 75 data digunakan sebagai data uji. Pemisahan ini bertujuan untuk menilai kemampuan algoritma Support Vector Machine (SVM) dalam mengelompokkan sentimen komentar YouTube.

4.11. Pelatihan & Pengujian Model

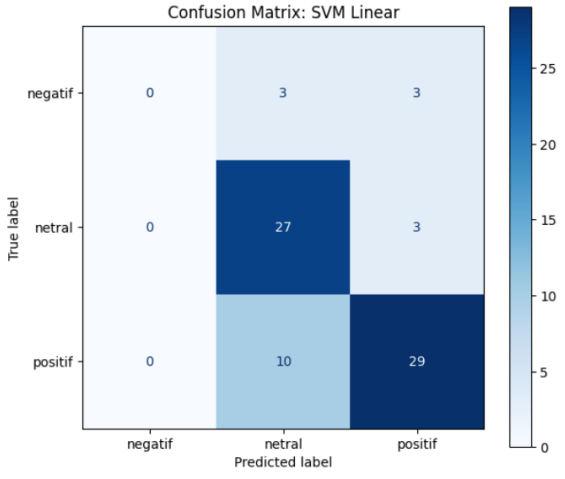
Setelah tahap *preprocessing* dan ekstraksi fitur selesai, model Support Vector Machine (SVM) dilatih menggunakan data latih dan dievaluasi dengan data uji. Pada penelitian ini diterapkan dua jenis kernel, yakni Linear dan Radial Basis Function (RBF). Evaluasi model dilakukan dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang ditampilkan dengan bentuk *classification report*.

Hasil pengujian memperlihatkan bahwa kedua model memperoleh tingkat akurasi sebesar 75% pada data uji. Performa model pada kelas netral dan positif tergolong baik, sedangkan pada kelas negatif model belum mampu mengklasifikasikan data dengan optimal, yang terlihat dari nilai evaluasi yang sangat rendah. Hasil *classification report* dilihat pada Gambar 2.

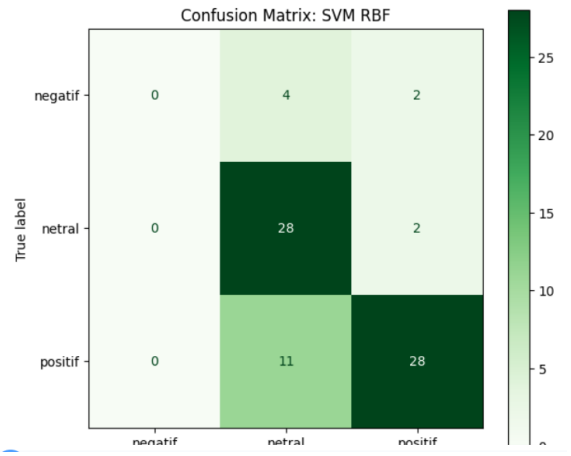
```
=== LAPORAN KLASIFIKASI: SVM LINEAR ===
              precision    recall  f1-score   support

     negatif       0.00      0.00      0.00         6
      netral       0.68      0.90      0.77        30
     positif       0.83      0.74      0.78        39

    accuracy                           0.75        75
   macro avg       0.50      0.55      0.52        75
weighted avg       0.70      0.75      0.72        75


=== LAPORAN KLASIFIKASI: SVM RBF ===
              precision    recall  f1-score   support

     negatif       0.00      0.00      0.00         6
      netral       0.65      0.93      0.77        30
     positif       0.88      0.72      0.79        39

    accuracy                           0.75        75
   macro avg       0.51      0.55      0.52        75
weighted avg       0.72      0.75      0.72        75
```

Gambar 2. Hasil *Classification Report*

Berdasarkan *confusion matrix*, model dapat mengelompokkan sebagian besar data dalam kategori netral dan positif secara tepat. Namun, seluruh data pada kelas negatif belum mampu diprediksi dengan benar dan cenderung diklasifikasikan ke kelas lain. Hal ini menunjukkan bahwa kapasitas model dalam mendeteksi sentimen negatif masih memiliki batasan. Visualisasi *confusion matrix* disajikan pada Gambar 3 dan Gambar 4.



Gambar 3. *Confusion Matrix*: SVM Linear



Gambar 4. *Confusion Matrix*: SVM RBF

4.12. Evaluasi Model

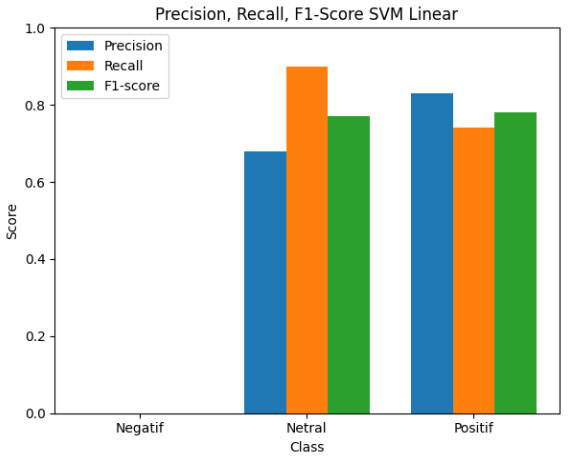
Tahap evaluasi model dibuat agar bisa mengukur seberapa efektif model dalam pengklasifikasian sentimen terkait ancaman disrupsi pekerjaan akibat kecerdasan buatan pada media sosial YouTube ke dalam kelas negatif, netral, dan positif. Adapun metrik evaluasi yang digunakan sebagai acuan dalam penelitian ini meliputi:

a. *Precision*
Presisi digunakan untuk mengukur ketepatan model dalam mengklasifikasikan suatu kelas sentimen. Nilai yang tinggi menunjukkan bahwa hasil prediksi model sebagian besar tepat. Pada SVM Linear, *precision* untuk kelas positif sebanyak 0,83 dan kelas netral sebanyak 0,68, yang memperlihatkan kinerja cukup baik pada kedua kelas tersebut. Namun, kelas negatif memperoleh nilai 0,00, yang menandakan model belum mampu memprediksi kelas ini dengan baik. Pada SVM RBF, *precision* untuk kelas positif meningkat menjadi 0,88, sedangkan kelas netral sedikit menurun menjadi 0,65. Hasil ini menunjukkan adanya peningkatan performa pada kelas positif, tetapi kelas negatif tetap memperoleh nilai 0,00.

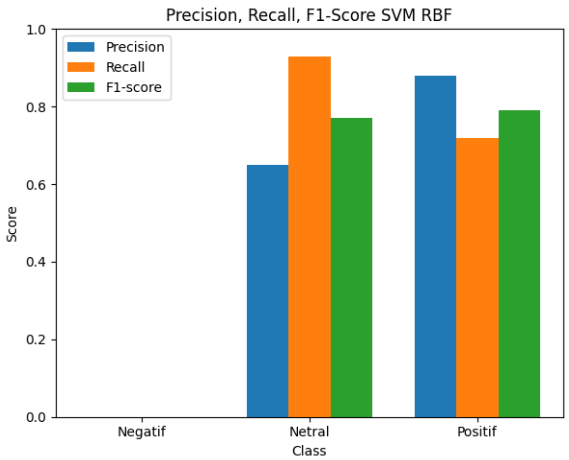
b. *Recall*
Recall memperlihatkan kapabilitas model dalam mengenali semua data yang masuk ke dalam suatu kelas. Semakin tinggi nilainya, semakin banyak data yang dapat diidentifikasi secara tepat. Pada SVM Linear, *recall* kelas netral sebanyak 0,90 dan kelas positif sebanyak 0,74, yang mengindikasikan kemampuan identifikasi yang cukup baik. Namun, kelas negatif memperoleh nilai 0,00, yang berarti data negatif belum berhasil dikenali. Pada SVM RBF, *recall* kelas netral meningkat menjadi 0,93, sementara kelas positif sedikit menurun menjadi 0,72. Meskipun demikian, kelas negatif tetap memperoleh nilai 0,00.

c. *F1-score*
F1-score adalah perpaduan antara presisi dan *recall* yang dimanfaatkan untuk melihat keseimbangan performa model. Pada SVM Linear, *F1-score* untuk kelas positif sebanyak 0,78 dan kelas netral sebanyak 0,77, yang menunjukkan keseimbangan kinerja yang cukup baik. Namun, kelas negatif memperoleh nilai 0,00. Pada SVM RBF, *F1-score* kelas positif meningkat menjadi 0,79 dan kelas netral tetap sebanyak 0,77. Hal ini mengindikasikan bahwa model RBF sedikit lebih unggul, meskipun kelas negatif masih belum mampu diklasifikasikan.

Model SVM memperlihatkan performa yang cukup baik dalam mengklasifikasikan sentimen terkait ancaman disrupsi pekerjaan akibat kecerdasan buatan, baik dengan kernel Linear maupun RBF. Hal ini terlihat dari nilai *precision*, *recall*, dan *F1-score* yang relatif tinggi pada kelas netral dan positif, meskipun masih rendah pada kelas negatif. Model SVM dengan kernel RBF memperlihatkan kinerja yang sedikit lebih unggul dibandingkan SVM Linear, terutama pada peningkatan nilai *precision* dan *F1-score* pada kelas positif serta *recall* pada kelas netral. Kedua model juga memperlihatkan kemampuan generalisasi yang cukup baik terhadap data pengujian. Visualisasi hasil evaluasi disajikan pada Gambar 5 dan Gambar 6.



Gambar 5. *Precision*, *Recall*, *F1-Score* SVM Linear



Gambar 6. *Precision*, *Recall*, *F1-Score* SVM RBF

5. KESIMPULAN DAN SARAN

Penelitian ini memperlihatkan bahwa penerapan algoritma Support Vector Machine (SVM) dengan representasi fitur Term Frequency–Inverse Document Frequency (TF-IDF) cukup efektif dalam mengklasifikasikan sentimen terkait ancaman disrupsi pekerjaan akibat kecerdasan buatan pada komentar YouTube. Berdasarkan hasil pengujian pada dataset yang digunakan, model SVM dengan kernel Linear dan RBF sama-sama memperoleh tingkat akurasi sebanyak 75%, dengan nilai *precision*, *recall*, dan *F1-score* yang cukup tinggi pada kelas netral dan positif. Hasil tersebut mengindikasikan bahwa kombinasi TF-IDF dan SVM bisa mengelompokkan sentimen dengan cukup memadai, serta model SVM dengan kernel RBF mempunyai kinerja yang sedikit lebih unggul dibandingkan SVM Linear.

Penelitian ini juga menyatakan bahwa metode penyaringan data berbasis konteks pekerjaan mampu meningkatkan relevansi dataset, sehingga proses klasifikasi dapat lebih fokus pada sentimen yang berkaitan dengan isu ancaman disrupsi pekerjaan akibat kecerdasan buatan. Dengan demikian, kebaruan penelitian ini terletak pada penerapan penyaringan data yang lebih khusus terhadap konteks pekerjaan, serta pengujian dua kernel SVM, yaitu Linear dan RBF, untuk memperoleh model klasifikasi yang paling sesuai dengan karakteristik data komentar YouTube.

Namun, model masih mempunyai keterbatasan dalam mengelompokkan kelas negatif yang terlihat dari nilai evaluasi yang sangat rendah. Oleh sebab itu, penelitian berikutnya disarankan menggunakan dataset yang lebih luas dan seimbang agar model dapat mempelajari distribusi data secara lebih optimal. Diharapkan hasil dari penelitian ini dapat dijadikan rujukan dalam pengembangan analisis sentimen serta memberikan gambaran mengenai pandangan masyarakat terhadap dampak kecerdasan buatan di dunia kerja.

DAFTAR PUSTAKA

[1] A. Faliza, dkk., "Job-Related Anxiety in the Age of Artificial Intelligence: A Systematic Review of Workplace Dynamics," *Formosa Journal of Multidisciplinary Research*, vol. 4, no. 8, hlm. 3965–3976, 2025.

[2] N. Badhurunnisa dan R. Dass, "Artificial Intelligence Anxiety and Perceived Future Employability Among Aviation Students," *DergiPark*, vol. 12, no. 1, hlm. 1-15, 2025.

[3] X. Sha, dkk., "When digital-AI transformation sparks adaptation: job crafting and AI knowledge in job insecurity contexts," *Frontiers in Psychology*, vol. 16, 2025.

[4] Kaya, dkk., "Will AI Take My Job? Evolving Perceptions of Automation and Labor Risk in Latin America," *AAAI Publications*, vol. 38, hlm. 1-10, 2024.

[5] A. Moghayedi, dkk., "Algorithmic anxiety: AI, work, and the evolving psychological contract in digital discourse," *PMC*, vol. 15, hlm. 100-110, 2024.

[6] A. N. Muhammad, S. Bukhori, dan P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes - Support Vector Machine (NBSVM) Classifier," 2019 *International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 2019

[7] M. Y. Khan, dkk., "Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments," *Sinkron*, vol. 9, no. 2, 2025.

[8] A. S. Iedwan, dkk., "Comparative Performance of SVM and Multinomial Naïve Bayes in Sentiment Analysis of the Film 'Dirty Vote'," *Scientific Journal of Informatics*, vol. 11, no. 3, 2024.

[9] A. N. Muhammad, dkk., "Optimizing Sentiment Analysis in Educational YouTube Videos: A Comparative Study of RoBERTa and Multinomial Naive Bayes," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 22, no. 2, hlm. 83–90, 2024.

[10] Syafia, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," *Jurnal Informatika*, vol. 9, no. 1, 2024.

[11] M. Garvey dan M. Maskal, "Sentiment Analysis of the News Media on Artificial Intelligence Does Not Support Claims of Negative Bias Against Artificial Intelligence," *Semantic Scholar*, vol. 10, hlm. 1-10, 2024.

[12] D. A. Constantin, dkk., "Anxiety in the Age of AI: Constructing a Tool to Assess Public Perceptions," *BRAIN*, vol. 15, no. 1, 2025.

[13] A. R. Putra, dkk., "Ensuring Reliability: Adaptation and Validation of the AI Anxiety Scale (AIAS) in Indonesia," *JOPP UK Institute*, vol. 3, no. 1, 2025.

[14] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Pencemaran Nama Baik di Media Sosial Twitter," KOMPUTA: *Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, 2023.

[15] D. Ghatge, dkk., "YouTube Comment Analyzer Using Sentimental Analysis," *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2024.

[16] A. Azrul, A. I. Purnamasari, dan I. Ali, "Analisis Sentimen Pengguna Twitter terhadap Perkembangan Artificial Intelligence dengan Penerapan Algoritma Long Short-Term Memory (LSTM)," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, 2024.

[17] M. Gerlich, "Public Anxieties About AI: Implications for Corporate Strategy and Societal Impact," *Administrative Sciences*, vol. 14, no. 11, 2024.

[18] M. D. Rachim dan L. O. M. Yamin, "Analisis

Sentimen Publik terhadap Penggunaan Teknologi AI dalam Berita Politik dan Implikasinya terhadap Pertumbuhan Ekonomi," *Jurnal Online Program Studi Pendidikan Ekonomi*, 2024.

[19] G. Y. Ardyaputra, I. B. N. Pascima, dan P. H. Suputra, "Analisis Sentimen Penggunaan CekatAI dalam Menggantikan Customer Service Menggunakan Logistic Regression dan TF-IDF," *Journal of Computer System and Informatics*, vol. 5, 2026.