

---

## **Analisis Sentimen Mengenai Penerimaan Calon Pegawai Negeri Sipil Tahun 2024 Menggunakan Support Vector Machine**

**Khairunissabina<sup>1\*</sup>, Rakhmat Kurniawan<sup>2</sup>**

Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

\*e-mail *Corresponding Author*: khairunissabina0701202158@uinsu.ac.id

### **Abstract**

*The 2024 Civil Servants Candidates (CPNS) recruitment has attracted public attention with various opinions reflecting societal sentiment. This study applies machine learning and text mining techniques to analyze sentiment based on comments collected from various sources. The process begins with text preprocessing, including tokenization, stopword removal, and text transformation using TF-IDF. The Support Vector Machine (SVM) model is used to classify sentiment into positive and negative categories. The data is divided into training and testing sets to evaluate the model's performance. Experimental results show that the model achieves an accuracy of 82.5%, indicating a fairly good performance in recognizing public opinion patterns. This analysis provides insights into aspects that are appreciated or criticized by the public regarding the 2024 CPNS recruitment. These findings can be utilized by policymakers to develop more responsive strategies that address public needs and expectations in the CPNS selection process.*

**Keyword:** *Machine Learning; Text Mining; Civil Servants Candidates; Support Vector Machine*

### **Abstrak**

Penerimaan Calon Pegawai Negeri Sipil (CPNS) 2024 menarik perhatian publik dengan berbagai opini yang mencerminkan sentimen masyarakat. Penelitian ini menerapkan machine learning dan teknik text mining untuk menganalisis sentimen berdasarkan komentar yang dikumpulkan dari berbagai sumber. Proses dimulai dengan pra-pemrosesan teks, termasuk tokenisasi, penghapusan *stopwords*, serta transformasi teks menggunakan TF-IDF. Model *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan sentimen menjadi positif dan negatif. Data dibagi menjadi data pelatihan dan pengujian dengan perbandingan 80% data training dan 20% data uji guna mengevaluasi kinerja model. Hasil eksperimen menunjukkan bahwa model mencapai akurasi 82,5%, menandakan kinerja yang cukup baik dalam mengenali pola opini masyarakat. Analisis ini memberikan wawasan mengenai aspek yang diapresiasi maupun dikritik oleh publik terkait penerimaan CPNS 2024. Temuan ini dapat dimanfaatkan oleh pemangku kebijakan untuk merancang strategi yang lebih responsif terhadap kebutuhan dan harapan masyarakat dalam proses seleksi CPNS.

**Kata kunci:** *Machine Learning; Text Mining; Calon Pegawai Negeri Sipil; Support Vector Machine*

### **1. Pendahuluan**

Pendaftaran Calon Pegawai Negeri Sipil (CPNS) 2024 selalu menyita perhatian sebagian masyarakat, terutama bagi mereka yang memiliki rencana untuk mendaftar. Seperti tahun-tahun sebelumnya, peminat CPNS pada tahun ini kemungkinan besar akan meningkat drastis. Sebelumnya, pemerintah telah mengumumkan akan membuka seleksi CASN dengan 2,3 juta formasi pada 2024. Formasi yang dibuka terdiri 690 ribu Calon Pegawai Negeri Sipil dan 1,6 juta pegawai pemerintah dengan perjanjian kerja (PPPK).

Antusiasme masyarakat yang terlihat pada media sosial terkait penerimaan CPNS 2024 ini cukup tinggi, yang pastinya juga menjadi salah satu bahan evaluasi dan pertimbangan pihak pemerintah dan penyelenggara lainnya untuk dapat menyesuaikan dengan kapasitas dan kesiapan dalam melaksanakan penerimaan CPNS 2024 [1]. Hal inilah yang menjadi acuan

peneliti untuk melakukan penelitian ini, dengan demikian proses penerimaan CPNS 2024 akan dapat dilaksanakan dengan baik karena adanya dukungan dari pihak penyelenggara dan masyarakat.

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi [2]. Salah satu keunggulan utama SVM adalah kemampuannya menangani data non-linier melalui kernel trick, yang mentransformasikan data ke dimensi lebih tinggi agar dapat dipisahkan secara linier [3]. Banyak penelitian sebelumnya yang menggunakan metode *Support Vector Machine* (SVM) untuk klasifikasi sentimen, seperti yang dilakukan oleh Nofandi dengan akurasi 69% menggunakan 550 data [4], dan Santoso Teguh dengan akurasi 94% menggunakan 300 data. Kedua penelitian ini menggunakan fitur Kernel Trick, yang membantu memetakan data non-linear ke ruang dimensi lebih tinggi untuk mempermudah klasifikasi. Meskipun menggunakan jumlah data dan fitur yang hampir sama, hasil akurasi kedua penelitian tersebut berbeda signifikan [5].

Penelitian ini dilakukan guna mengidentifikasi antusiasme masyarakat terhadap pendaftaran CPNS 2024, yang dimana data dari hasil penelitian ini akan menunjukkan data yang berpengaruh dan diperlukan dalam proses evaluasi untuk meningkatkan sistem penerimaan CPNS di masa mendatang.

## 2. Tinjauan Pustaka

Hasil penelitian yang dilakukan oleh Fajar dan Sugiyono, dengan judul penelitian berupa Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada *Twitter* Dengan Metode *Naive Bayes* Dan *Support Vector Machine*, guna mencari tahu pendapat rakyat terkait proses perekrutan PPPK yang dicuitkan pada *Twitter*. Algoritma yang diterapkan pada penelitian ini adalah algoritma *Naive Bayes* dan *Support Vector Machine* (SVM), dengan tujuan sebagai variabel dalam melakukan analisis penelitian ini. Hasil akhir yang diimplementasikan pada metode SVM (*Support Vector Machine*) menunjukkan akurasi sebesar 94,80%. Sedangkan pengujian dengan metode *Naive bayes* menunjukkan nilai akurasi sebesar 96,14% [6].

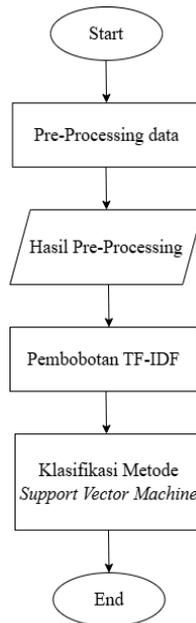
Sedangkan penelitian yang dilakukan oleh Rahma, dkk, dengan menerapkan metode yang berbeda, dengan judul Analisis Sentimen terhadap Karyawan Dirumahan pada Media Sosial *Twitter* menggunakan Fitur N-Gram dan Pembobotan *Augmented TF – IDF Probability* dengan *K-Nearest Neighbour*. Dengan menerapkan jarak *Manhattan Distance* pada pembobotan *Augmented TF – IDF Probability* yang berakurasi senilai 0,6, *presicion* senilai 0,415, *recall* senilai 0,404, dan *f-measure* senilai 0,406 untuk *term unigram*. Dan hasil evaluasi terbaik dihasil dengan menggunakan jarak *Cosine Distance* dengan nilai *accuracy* 0,776, *presicion* senilai 0,59 *recall* senilai 0,408, dan *f-measure* senilai 0,437 untuk *term bigram* [7].

Begitu pula dengan penelitian yang dilakukan oleh Susi dengan judul Analisis Sentimen Cuitan *Twitter* Pada Proses Seleksi CPNS Menggunakan *Algoritma Ensembles Vote*, yang dimana metode yang diterapkan dalam penelitian ini jelas berbeda. Dengan melakukan analisis sentimen cuitan *twitter* pada proses seleksi CPNS seleksi CPNS 2021, penggunaan algoritma *k-Nearest Neighbor* (k-NN) akurasi 61.90% dengan AUC 0.912 dibandingkan algoritma *Naive Bayes* akurasi 100.00% dengan AUC 1.000, dan untuk algoritma C4.5 akurasi 85.71% dengan AUC 0.906 [8].

Berbeda dengan penelitian-penelitian tersebut, penelitian yang dilakukan saat ini menggunakan metode *Support Vector Machine* (SVM) dalam analisis sentimen, yang bertujuan untuk mengklasifikasikan opini masyarakat terkait penerimaan CPNS 2024. Penggunaan SVM dalam penelitian ini memungkinkan pemisahan antara sentimen positif dan negatif dengan cara menemukan hyperplane yang optimal, sehingga dapat memberikan wawasan yang lebih jelas tentang pandangan masyarakat. SVM dipilih karena kemampuannya dalam menangani data yang kompleks dan non-linear, serta efektif dalam mengklasifikasikan data berbasis teks, seperti komentar atau tweet yang berhubungan dengan topik CPNS 2024.

## 3. Metodologi

Dalam penelitian ini, *flowchart* yang digunakan menggambarkan alur proses analisis sentimen dengan menggunakan metode SVM.



Gambar 1 Flowchart Metode SVM

### 3.1 Preprocessing Data

Data dikumpulkan melalui *scrapping Twitter* menggunakan kata kunci #CPNS2024 untuk rentang waktu 1 Januari 2024 hingga sekarang. Sebanyak 800 *tweets* berhasil dikumpulkan dan diekspor ke format .xlsx dan terdiri dari 80% *data training* dan 20% *data testing*. Setiap *tweet* diberi label positif atau negatif secara manual untuk menyusun dataset berlabel. Proses pelabelan ini dilakukan dengan mempertimbangkan konteks dan makna dalam setiap *tweet* untuk memastikan akurasi data. Dataset berlabel ini akan digunakan untuk melatih model klasifikasi dalam analisis sentimen, yang nantinya dapat mengidentifikasi sentimen secara otomatis pada data baru dengan tingkat akurasi yang optimal. Data ini juga akan dianalisis lebih lanjut untuk mengidentifikasi tren opini publik terhadap CPNS 2024.

Tahap *preprocessing* data meliputi *cleaning*, *case folding*, *tokenizing*, *filtering*, *stopword removal*, dan *stemming* untuk menghasilkan data yang bersih dan terstruktur [9]. Contoh menghitung manual menggunakan 5 data latih dan 2 data uji disertakan untuk mempermudah pemahaman proses klasifikasi. Data ini mencakup teks, label *pril nt*, dan bobot fitur. Tahap ini bertujuan untuk menggambarkan *pril nt*-langkah detail dari *preprocessing* hingga prediksi *pril nt*, yang kemudian akan dibandingkan dengan output sistem untuk memvalidasi akurasi metode SVM.

Tabel 1 Sampel Data Latih

Data Latih	
sentimen	komentar
Negatif	@_pegipegi blm kaak pril tuh keknya masih ngurusin cpns 2024 itu sk cpns nya baru pada keluar paling cepet agustus atau sept
Positif	Penting! 5 Hal yang Wajib Diperhatikan saat Pengisian DRH CPNS 2024 Jangan Sampai Salah <a href="https://t.co/rssFPRyZns">https://t.co/rssFPRyZns</a>
Positif	berikut Minaru sampaikan Pengumuman Hasil Seleksi Calon Pegawai Negeri Sipil (CPNS) Kementerian ATR/BPN Tahun 2024. <a href="https://t.co/wdWYqri8IG">https://t.co/wdWYqri8IG</a>
Positif	Seleksi CPNS 2025 – Mau dibantu belajar soal CPNS 2025 dengan lebih mudah? Klik lini di bio yaa! Kuota terbatas #cpns2025 <a href="https://t.co/22G2IX2Fla">https://t.co/22G2IX2Fla</a>
Positif	Kelulusan bagi calon pegawai negeri sipil (CPNS) formasi 2024 di lingkungan Pemkab Mojokerto final. <a href="https://t.co/xpWDir2HvG">https://t.co/xpWDir2HvG</a> #CPNS #NIP #mojokerto

Tabel 2 Sampel Data Uji

Data Uji
@kempnrb Bgmn untuk yang masuk data P3K yg udah mendaftar CPNS 2024? Mereka sdh tidak bisa ikutan P3K Gel 2? 20 Link Instansi Pusat yang Sudah Umumkan Hasil Akhir CPNS 2024 Salah Satunya Ada KemenPAN RB – Ayo Bandung <a href="https://t.co/tPkSPJ8hD9">https://t.co/tPkSPJ8hD9</a> #link #CPNS

1) *Cleaning Data*

Pada tahap *cleaning data*, elemen-elemen tidak relevan seperti tanda baca, angka, URL, emotikon, pril, dan karakter non-alfabet dihapus untuk memastikan data pril pada teks utama yang relevan [10]. Proses ini bertujuan agar analisis selanjutnya, seperti *tokenizing* dan pembobotan, dapat dilakukan lebih efektif dan menghasilkan *output* yang lebih akurat [11].

Tabel 3 *Cleaning*

Komentar	<i>Cleaning Data</i>
	<b>Data Latih</b>
@_pegipegi blm kaak pril tuh keknya masih ngurusin cpns 2024 itu sk cpns nya baru pada keluar paling cepet agustus atau sept Penting! 5 Hal yang Wajib Diperhatikan saat Pengisian DRH CPNS 2024 Jangan Sampai Salah <a href="https://t.co/rssFPRyZns">https://t.co/rssFPRyZns</a> ...	Blm Kaak April Tuh Keknya Masih Ngurusin Cpns Itu Sk Cpns Nya Baru Pada Keluar Paling Cepet Agustus Atau Sept  Penting Hal Yang Wajib Diperhatikan Saat Pengisian Drh Cpns Jangan Sampai Salah  ...
Kelulusan bagi calon pegawai negeri sipil (CPNS) formasi 2024 di lingkungan Pemkab Mojokerto final. <a href="https://t.co/xpWDir2HvG">https://t.co/xpWDir2HvG</a> #CPNS #NIP #mojokerto	Kelulusan Bagi Calon Pegawai Negeri Sipil Cpns Formasi Di Lingkungan Pemkab Mojokerto Final
	<b>Data Uji</b>
@kempnrb Bgmn untuk yang masuk data P3K yg udah mendaftar CPNS 2024? Mereka sdh tidak bisa ikutan P3K Gel 2? 20 Link Instansi Pusat yang Sudah Umumkan Hasil Akhir CPNS 2024 Salah Satunya Ada KemenPAN RB - Ayo Bandung <a href="https://t.co/tPkSPJ8hD9">https://t.co/tPkSPJ8hD9</a> #link #CPNS	Bgmn Untuk Yang Masuk Data PK Yg Udah Mendaftar Cpns Mereka Sdh Tidak Bisa Ikutan PK Gel  Link Instansi Pusat Yang Sudah Umumkan Hasil Akhir Cpns Salah Satunya Ada Kemenpan Rb Ayo Bandung

2) *Casefolding*

Tahapan ini merupakan *case folding*, yaitu proses untuk mengubah seluruh teks menjadi huruf kecil agar lebih konsisten [12].

Tabel 4 *Casefolding*

<i>Cleaning Data</i>	<i>Casefolding</i>
	<b>Data Latih</b>
Blm Kaak April Tuh Keknya Masih Ngurusin Cpns Itu Sk Cpns Nya Baru Pada Keluar Paling Cepet Agustus Atau Sept Penting Hal Yang Wajib Diperhatikan Saat Pengisian Drh Cpns Jangan Sampai Salah ...	blm kaak april tuh keknya masih ngurusin cpns itu sk cpns nya baru pada keluar paling cepet agustus atau sept penting hal yang wajib diperhatikan saat pengisian drh cpns jangan sampai salah ...
Kelulusan Bagi Calon Pegawai Negeri Sipil Cpns Formasi Di Lingkungan Pemkab Mojokerto Final	kelulusan bagi calon pegawai negeri sipil cpns formasi di lingkungan pemkab mojokerto final
	<b>Data Uji</b>

<b>Cleaning Data</b>	<b>Casefolding</b>
Bgmn Untuk Yang Masuk Data PK Yg Udah Mendaftar Cpns Mereka Sdh Tidak Bisa Ikutan PK Gel	bgmn untuk yang masuk data pk yg udah mendaftar cpns mereka sdh tidak bisa ikutan pk gel
Link Instansi Pusat Yang Sudah Umumkan Hasil Akhir Cpns Salah Satunya Ada Kemenpan Rb Ayo Bandung	link instansi pusat yang sudah umumkan hasil akhir cpns salah satunya ada kemenpan rb ayo bandung

### 3) Tokenize

Tahapan ini merupakan *tokenizing*, yaitu proses memecah teks menjadi unit-unit terkecil, yang disebut token, seperti kata atau frasa [13].

Tabel 5 Tokenize

<b>Casefolding</b>	<b>Tokenizing</b>
<b>Data Latih</b>	
blm kaak april tuh keknya masih ngurusin cpns itu sk cpns nya baru pada keluar paling cepet agustus atau sept	"blm", "kaak", "april", "tuh", "keknya", "masih", "ngurusin", "cpns", "itu", "sk", "cpns", "nya", "baru", "pada", "keluar", "paling", "cepat", "agustus", "atau", "sept"
penting hal yang wajib diperhatikan saat pengisian drh cpns jangan sampai salah	"penting", "hal", "yang", "wajib", "diperhatikan", "saat", "pengisian", "drh", "cpns", "jangan", "sampai", "salah"
...	...
kelulusan bagi calon pegawai negeri sipil cpns formasi di lingkungan pemkab mojokerto final	"kelulusan", "bagi", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "di", "lingkungan", "pemb", "mojokerto", "final"
<b>Data Uji</b>	
Bgmn untuk yang masuk data PK yg udah mendaftar cpns Mereka sdh tidak bisa ikutan PK Gel	"bgmn", "untuk", "yang", "masuk", "data", "pk", "yg", "udah", "mendaftar", "cpns", "mereka", "sdh", "tidak", "bisa", "ikutan", "pk", "gel"
Link instansi pusat yang sudah umumkan hasil akhir cpns salah satunya ada kemenpan rb ayo bandung	"link", "instansi", "pusat", "yang", "sudah", "umumkan", "hasil", "akhir", "cpns", "salah", "satunya", "ada", "kemenpan", "rb", "ayo", "bandung"

### 4) Filtering

Tahap ini merupakan *filtering*, yaitu proses untuk menghilangkan elemen atau kata-kata yang tidak relevan atau kurang signifikan dalam teks [14].

Tabel 6 Filtering

<b>Tokenizing</b>	<b>Filtering</b>
<b>Data Latih</b>	
"blm", "kaak", "april", "tuh", "keknya", "masih", "ngurusin", "cpns", "itu", "sk", "cpns", "nya", "baru", "pada", "keluar", "paling", "cepat", "agustus", "atau", "sept"	"kaak", "april", "ngurusin", "cpns", "sk", "cpns", "keluar", "agustus", "sept"
"penting", "hal", "yang", "wajib", "diperhatikan", "saat", "pengisian", "drh", "cpns", "jangan", "sampai", "salah"	"penting", "wajib", "diperhatikan", "pengisian", "drh", "cpns"
...	...
"kelulusan", "bagi", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "di", "lingkungan", "pemb", "mojokerto", "final"	"kelulusan", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "lingkungan", "pemb", "mojokerto", "final"

<i>Tokenizing</i>	<i>Filtering</i>
<b>Data Uji</b>	
"bgmn", "untuk", "yang", "masuk", "data", "pk", "yg", "udah", "mendaftar", "cpns", "mereka", "sdh", "tidak", "bisa", "ikutan", "pk", "gel" "link", "instansi", "pusat", "yang", "sudah", "umumkan", "hasil", "akhir", "cpns", "salah", "satunya", "ada", "kemenpan", "rb", "ayo", "bandung"	masuk, "data", "pk", "mendaftar", "cpns", "ikutan", "pk", "gel" "link", "instansi", "pusat", "umumkan", "hasil", "akhir", "cpns", "kemenpan", "rb", "bandung"

5) *Stopword*

Tahap ini merupakan *stopword removal*, yaitu proses menghapus kata-kata umum yang sering muncul dalam teks namun tidak memiliki makna atau kontribusi signifikan dalam analisis sentimen. Kata-kata seperti "dan", "atau", "di", "ke", "yang", dan "ini" [15].

Tabel 7 *Stopword*

<i>Filtering</i>	<i>Stopword</i>
<b>Data Latih</b>	
"kaak", "april", "ngurusin", "cpns", "sk", "cpns", "keluar", "agustus", "sept" "penting", "wajib", "diperhatikan", "pengisian", "drh", "cpns" "seleksi", "cpns", "dibantu", "belajar", "soal", "cpns", "mudah", "klik", "lini", "bio", "kuota", "terbatas"	"ngurusin", "cpns", "cpns", "keluar" "penting", "wajib", "pengisian", "cpns" "seleksi", "cpns", "soal", "cpns", "mudah", "kuota", "terbatas"
...	
"kelulusan", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "lingkungan", "pemkab", "mojokerto", "final"	"kelulusan", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "final"
<b>Data Uji</b>	
masuk, "data", "pk", "mendaftar", "cpns", "ikutan", "pk", "gel" "link", "instansi", "pusat", "umumkan", "hasil", "akhir", "cpns", "kemenpan", "rb", "bandung"	"masuk", "data", "mendaftar", "cpns", "ikutan", "instansi", "pusat", "umumkan", "hasil", "akhir", "cpns",

6) *Stemming*

Pada tahap ini merupakan *stemming*, yaitu proses untuk mengubah kata-kata ke bentuk dasarnya [16].

Tabel 8 *Stemming*

<i>Stopword</i>	<i>Stemming</i>
<b>Data Latih</b>	
"ngurusin", "cpns", "cpns", "keluar" "penting", "wajib", "pengisian", "cpns" "seleksi", "calon", "pegawai", "negeri", "sipil", "cpns", "kementerian", "atr", "bpn"	"urus", "cpns", "cpns", "keluar" "penting", "wajib", "isi", "cpns" "seleksi", "calon", "pegawai", "negeri", "sipil", "cpns", "kementerian", "atr", "bpn"
...	
"kelulusan", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "final"	"lulus", "calon", "pegawai", "negeri", "sipil", "cpns", "formasi", "final"
<b>Data Uji</b>	
"masuk", "data", "mendaftar", "cpns", "ikutan", "instansi", "pusat", "umumkan", "hasil", "akhir", "cpns",	"masuk", "data", "daftar", "cpns", "ikut", "instansi", "pusat", "umum", "hasil", "akhir", "cpns",

## 4. Hasil dan Pembahasan

### 4. 1. Nilai TF-IDF

Berikut adalah nilai TF-IDF yang diperoleh setelah tahap preprocessing data. Nilai TF-IDF ini menunjukkan bobot penting dari setiap kata dalam dokumen, dengan mempertimbangkan frekuensi kata dalam dokumen serta distribusi kata di seluruh koleksi dokumen.

Tabel 9 Nilai TF-IDF

TERM	TF-IDF				
	D1	D2	D3	D4	D5
urus	1,477	0	0	0	0
cpns	6,125	6,125	6,125	6,125	6,125
formasi	0	0	0	0	1,477
...	...	...	...	...	...
final	0	0	0	0	1,477

### 4. 2. Kernel Linier

Adapun hasil dari perhitungan kernel linear dari data sampel yang dimiliki sebagaimana yang disajikan pada tabel berikut.

Tabel 10 Kernel Linier

No	1	2	3	...	19	20	21
1	0,00757	0,03141	0,00757	...	0	0	0
2	0,03141	0,65131	0,03141	...	0,03141	0,03141	0,03141
3	0,00757	0,03141	0,00757	...	0	0	0
...	...	...	...	...	...	...	...
20	0	0,03141	0	...	0,00757	0,00757	0,00757
21	0	0,03141	0	...	0,00757	0,00757	0,00757

### 4. 3. Menghitung Nilai *Matriks Hessian*

Berikut adalah hasil perhitungan *matriks Hessian* yang diperoleh selama proses pelatihan model.

Tabel 11 Nilai *Matriks Hessian*

	urus	cpns	keluar	...	lulus	formasi	final
urus	0,25757	0,21859	0,25757	...	0,25	0,25	0,25
cpns	0,21859	0,90131	0,21859	...	0,28141	0,28141	0,28141
keluar	0,25757	0,21859	0,25757	...	0,25	0,25	0,25
penting	0,25	0,28141	0,25	...	0,25	0,25	0,25
...	...	...	...	...	...	...	...
final	0,25	0,28141	0,25	...	0,25757	0,25757	0,25757

### 4. 4. Penentuan *Support Vector*

Menentukan *support vector* untuk setiap dokumen. Proses ini dilakukan berdasarkan nilai citerbaru yang telah diperoleh.

Tabel 12 Penentuan *Support Vector*

No	D1	D2	D3	D4	D5	$\alpha$	Kelas
1	0,087	0	0	0	0	0,23993435	-1
2	0,361	0,361	0,361	0,361	0,361	0,223021859	1
3	0,087	0	0	0	0	0,23993435	-1
...	...	...	...	...	...	...	...
20	0	0	0	0	0,087	0,238580144	1
21	0	0	0	0	0,087	0,238580144	1

Langkah selanjutnya adalah menghitung fungsi kernel untuk masing-masing kelas menggunakan nilai  $\alpha$  tertinggi. Nilai  $K(x_i, x^+)$  dihitung dengan  $\alpha_i$  terbesar dari kelas positif (0,239163741), dan  $K(x_i, x^-)$  dihitung dengan  $\alpha_i$  terbesar dari kelas negatif (0,23993435). Nilai-nilai ini diambil dari elemen *matriks Hessian* pada kolom yang sesuai untuk masing-masing kelas. Perhitungan kernel ini menjadi komponen penting dalam proses klasifikasi menggunakan SVM untuk menentukan posisi data terhadap hyperplane pemisah dan mengambil keputusan akhir.

$$\begin{aligned}
 K(x_i, x^+) &= \sum \alpha_i y_i D_i \\
 &= (0.23993435 \times -1 \times 0.25) + (0.223021859 \times 1 \times 0.281412323) + (0.23993435 \times -1 \times 0.25) + (0.239163741 \times 1 \times 0.257574949) + (0.239163741 \times 1 \times 0.257574949) + (0.239163741 \times 1 \times 0.257574949) + (0.235989159 \times 1 \times 0.25) + (0.235367369 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) \\
 &= 1,018662157
 \end{aligned}$$

$$\begin{aligned}
 K(x_i, x^-) &= \sum \alpha_i y_i D_i \\
 &= (0.23993435 \times -1 \times 0.257574949) + (0.223021859 \times 1 \times 0.218587677) + (0.23993435 \times -1 \times 0.257574949) + (0.239163741 \times 1 \times 0.25) + (0.239163741 \times 1 \times 0.25) + (0.239163741 \times 1 \times 0.25) + (0.235989159 \times 1 \times 0.25) + (0.235367369 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238434333 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238934951 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) + (0.238580144 \times 1 \times 0.25) \\
 &= 0,995580948
 \end{aligned}$$

Nilai  $K(x_i, x^+)$  dan  $K(x_i, x^-)$  sudah dihitung, langkah selanjutnya adalah menghitung nilai bias dengan menggunakan persamaan dibawah ini.

$$b = \frac{1}{2} [\sum_{i=1} \alpha_i y_i K(x_i x^+) + \sum_{i=1} \alpha_i y_i K(x_i x^-)]$$

Dari persamaan diatas, maka diperoleh nilai b sebagai berikut.:

$$\begin{aligned}
 b &= -\frac{1}{2} [w \cdot x^+ + w \cdot x^-] \\
 &= -\frac{1}{2} [1,018662157 + 0,995580948] \\
 &= -\frac{1}{2} (2,01424) \\
 &= -1,00712
 \end{aligned}$$

#### 4. 5. Pengujian SVM

Setelah menghitung nilai bias, langkah selanjutnya adalah menguji model dengan menghitung nilai TF-IDF untuk setiap data uji. TF-IDF digunakan untuk mengukur pentingnya kata dalam dokumen relatif terhadap kumpulan dokumen, sehingga membantu menentukan bobot kata yang signifikan dalam pengklasifikasian data.

Tabel 13 Nilai TF-IDF Data Uji

No.	Term	TF	DF	IDF	TF-IDF	Norm
1	masuk	1	0	1,778	1,778	0,3167
2	data	1	0	1,778	1,778	0,3167
3	daftar	1	0	1,778	1,778	0,3167
4	cpns	2	7	0,875	1,750	0,3117
...	...	...	...	...	...	...
10	akhir	1	0	1,778	1,778	0,3167

Tahap berikutnya adalah menghitung kernel untuk setiap data uji dengan data latih yang telah dihitung sebelumnya. Perhitungan kernel ini bertujuan untuk mengukur hubungan atau kesesuaian antara data uji dan data latih berdasarkan fitur-fitur yang diwakili dalam ruang dimensi tertentu. Hasil dari perhitungan kernel ini akan digunakan sebagai dasar untuk menentukan kelas dari data uji. Berikut adalah hasil perhitungan kernel antara data latih dan data uji yang telah dianalisis.

Tabel 14 Kernel Linier Data Uji

No	masuk	data	daftar	cpns	...	umum	hasil	akhir
1	0,0276	0,0276	0,0276	0,0271	...	0,0276	0,0276	0,0276
2	0,5716	0,5716	0,5716	0,5625	...	0,5716	0,5716	0,5716
3	0,0276	0,0276	0,0276	0,0271	...	0,0276	0,0276	0,0276
...	...	...	...	...	...	...	...	...
20	0,0276	0,0276	0,0276	0,0271	...	0,0276	0,0276	0,0276
21	0,0276	0,0276	0,0276	0,0271	...	0,0276	0,0276	0,0276

Setelah nilai kernel diperoleh, langkah berikutnya adalah menghitung bobot untuk data uji. Bobot dihitung dengan menggabungkan nilai kernel yang telah diperoleh sebelumnya dan nilai alpha yang terkait dengan masing-masing data. Perhitungan ini akan memberikan bobot yang digunakan dalam proses klasifikasi, yang pada akhirnya akan membantu menentukan posisi data terhadap hyperplane pemisah dan mengklasifikasikan data ke dalam kelas yang sesuai.

Tabel 15 Nilai Bobot Data Uji

No	masuk	data	daftar	cpns	...	umum	hasil	akhir
1	-0,0066	-0,0066	-0,0066	-0,0065	...	-0,0066	-0,0066	-0,0066
2	0,1275	0,1275	0,1275	0,1255	...	0,1275	0,1275	0,1275
3	-0,0066	-0,0066	-0,0066	-0,0065	...	-0,0066	-0,0066	-0,0066
...	...	...	...	...	...	...	...	...
20	0,0066	0,0066	0,0066	0,0065	...	0,0066	0,0066	0,0066
21	0,0066	0,0066	0,0066	0,0065	...	0,0066	0,0066	0,0066

Setelah bobot dihitung, nilai  $f(x)$  digunakan untuk mengklasifikasikan data. Jika hasil positif, data termasuk kelas positif (mendukung CPNS 2024), dan jika negatif, data masuk kelas negatif (tidak mendukung). Proses ini memastikan klasifikasi sentimen yang akurat terkait CPNS 2024.

#### Data uji 1

["masuk", "data", "daftar", "cpns", "ikut"]

$$\begin{aligned}
 f(x) &= w \cdot x + b \\
 &= \sum \alpha_i y_i K(x_i, x_j) + b \\
 &= (0,3142 + 0,3142 + 0,3142 + 0,3092 + 0,3142) + (-1,00712) \\
 &= 2,573
 \end{aligned}$$

$$\begin{aligned}
 \text{Fungsi Klasifikasi} &= \text{sign}(2,573) \\
 &= 1
 \end{aligned}$$

#### Data uji 2

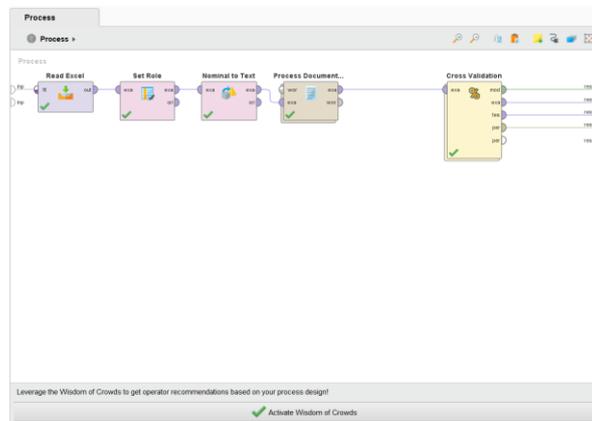
["instansi", "pusat", "umum", "hasil", "akhir", "cpns"]

$$\begin{aligned}
 f(x) &= w \cdot x + b \\
 &= \sum \alpha_i y_i K(x_i, x_j) + b \\
 &= (0,3142 + 0,3142 + 0,3142 + 0,3142 + 0,3142 + 0,3092) + (-1,00712) \\
 &= 2,88717
 \end{aligned}$$

$$\begin{aligned}
 \text{Fungsi Klasifikasi} &= \text{sign}(2,88717) \\
 &= 1
 \end{aligned}$$

#### 4. 6. Penerapan Algoritma SVM pada *RapidMiner*

Gambar berikut merupakan proses pengujian algoritma SVM menggunakan *RapidMiner* Studio melibatkan beberapa tahapan utama, dimulai dengan memasukkan data latih dan data uji, diikuti oleh langkah preprocessing untuk membersihkan dan menormalkan data. Selanjutnya, model SVM dibangun untuk memisahkan data ke dalam kelas positif dan negatif, kemudian data uji diterapkan pada model untuk melakukan klasifikasi.



Gambar 2 Proses Pengujian SVM pada *RapidMiner*

#### 4. 7. Akurasi Model

Gambar di bawah ini menunjukkan nilai akurasi model yang diuji menggunakan *RapidMiner*, yaitu 82,5%. Nilai ini menggambarkan bahwa model dapat mengklasifikasikan data uji dengan benar sebanyak 82,5% dari total data yang diuji. Hasil ini mencerminkan kinerja model dalam mengklasifikasikan data ke dalam kategori yang tepat. Dengan akurasi sebesar 82,5%, model dapat dianggap cukup efektif, namun masih ada ruang untuk perbaikan, seperti dengan optimasi parameter atau menggunakan teknik lain untuk meningkatkan hasil klasifikasi lebih lanjut.

Table View  Plot View

accuracy: 82.50% +/- 1.67% (micro average: 82.50%)

	true Negatif	true Positif	class precision
pred. Negatif	26	4	86.67%
pred. Positif	136	634	82.34%
class recall	16.05%	99.37%	

Gambar 3 Akurasi Model

#### 4. 8. Pembahasan

Penelitian analisis sentimen dengan menerapkan metode *Support Vector Machine* terhadap data X dengan kata kunci #CPNS2024 menunjukkan hasil akurasi sebesar 82,5%. Nilai akurasi ini tergolong baik dibandingkan dengan penelitian sebelumnya yang dilakukan oleh Susi dengan judul Analisis Sentimen Cuitan *Twitter* Pada Proses Seleksi CPNS Menggunakan *Algoritma Ensembles Vote* serta penggunaan algoritma *k-Nearest Neighbor* (k-NN) yang nilai akurasinya hanya 61.90% [8].

Namun berbeda penelitian yang dilakukan oleh Fajar dan Sugiyono, dengan judul penelitian berupa Analisis Sentimen Masyarakat Terhadap Perekrutan PPPK Pada *Twitter* Dengan Metode *Naive Bayes* Dan *Support Vector Machine*, yang menunjukkan nilai akurasi sebesar 94,80 % [6]. Penelitian tersebut hanya menggunakan 482 data, namun berhasil mendapatkan nilai akurasi yang lebih tinggi. Hal ini menunjukkan bahwa jumlah data yang digunakan tidak memberikan pengaruh dalam nilai akurasi dari hasil penelitian.

## 5. Simpulan

Analisis sentimen terhadap penerimaan CPNS 2024 bertujuan untuk menggali pandangan masyarakat, baik positif maupun negatif. Proses dimulai dengan pengumpulan data komentar, yang kemudian diproses melalui tokenisasi, penghapusan *stopwords*, penyaringan, dan transformasi teks menggunakan TF-IDF. Algoritma *Support Vector Machine* (SVM) digunakan untuk klasifikasi sentimen. Model diuji dengan data pelatihan dan pengujian, menghasilkan akurasi 82,5%. Hasil ini memberikan wawasan tentang penerimaan CPNS 2024, membantu pembuat kebijakan merancang langkah-langkah strategis yang lebih baik.

## Daftar Referensi

- [1] M. Nurkarifin, T. I. Hermanto, and M. G. Resmi, "Analisis Sentimen Terhadap Aplikasi Lowongan Kerja Menggunakan Lexicon Based Features Dan Support Vector Machine," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, vol. 13, no. 1, p. 505, May 2024, doi: 10.35889/jutisi.v13i1.1867.
- [2] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 8, no. 1, p. 147-151, Feb. 2021, doi: 10.25126/jtiik.0813944.
- [3] E. Hartati, "Penggunaan Klasifikasi Sayur Segar dan Sayur Busuk Menggunakan Algoritma Support Vector Machine," *J. Tek. Inform. Dan Sist. Inf.*, vol. Vol 07, No 03, pp. 678–687, Desember 2020.
- [4] A. Nofandi, N. Y. Setiawan, and D. W. Brata, "Analisis Sentimen Ulasan Pelanggan dengan Metode Support Vector Machine (SVM) untuk Peningkatan Kualitas Layanan pada Restoran Warung Wareg," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 7, no. 1, pp. 458–466, 2023.
- [5] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM," *J. Data Min. Dan Sist. Inf.*, vol. 2, no. 1, pp. 31-41, Feb. 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [6] F. N. Hidayat and S. Sugiyono, "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains Dan Teknol.*, vol. 5, no. 2, pp. 665–672, Dec. 2023, doi: 10.55338/saintek.v5i2.1359.
- [7] R. Chairunnisa, "Analisis Sentimen terhadap Karyawan Dirumahkan pada Media Sosial Twitter menggunakan Fitur N-Gram dan Pembobotan Augmented TF – IDF Probability dengan K-Nearest Neighbour," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, vol. 6, no. 4, pp. 1960–1965, 2022.
- [8] Susi, "Analisis Sentimen Cuitan Twitter Pada Proses Seleksi CPNS Menggunakan Algoritma Ensembles Vote," *IKRAM: Jurnal Ilmu Komputer Al Muslim*, vol. 1, no. 1, pp. 42–46, 2022.
- [9] A. H. Lubis, L. P. A. Lubis, and Sriani, "Sentiment analysis on twitter about the death penalty using the support vector machine method," *TEKNOSAINS J. Sains Teknol. Dan Inform.*, vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373.
- [10] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains Teknol. Dan Inform.*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.
- [11] M. A. P. Ginting and S. Sriani, "Developing a Web-Based Application for Palm Seedling Eligibility Using C5.0 Algorithm and CART Algorithm," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 12, no. 1, pp. 97–108, Mar. 2024, doi: 10.33558/piksel.v12i1.8810.
- [12] M. Furqan, S. Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.Com*, vol. 21, no. 1, pp. 51–60, Feb. 2022, doi: 10.33633/tc.v21i1.5446.
- [13] R. W. Pulungan, S. Sriani, and A. Armansyah, "Implementation of Naïve Bayes Method Diagnosing Diseases Nile Tilapia," *J. Comput. Netw. Archit. High Perform. Comput.*, vol. 6, no. 2, pp. 817–828, May 2024, doi: 10.47709/cnahpc.v6i2.3834.
- [14] M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 162–166, Nov. 2020, doi: 10.31539/intecom.v3i2.1654.

- 
- [15] A. Handayani and I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res. JOSH*, vol. 5, no. 1, pp. 53–63, Oct. 2023, doi: 10.47065/josh.v5i1.4379.
- [16] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjee.v5i1.16830.