



Penerapan *Text Mining* Untuk Klasifikasi Judul Berita *Hoax* Vaksinasi COVID-19 Menggunakan Algoritma *Support Vector Machine*

Aditya Wijaya¹, Chaerur Rozikin², Betha Nurina Sari³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang

Abstract

Received: 12 Agustus 2022

Revised: 18 Agustus 2022

Accepted: 24 Agustus 2022

Abstract: Covid-19 has become a worldwide pandemic and eventually spread to Indonesia. The government is taking preventive measures by COVID-19 vaccination program. People looking for this vaccination information program through online media. However, hoax news about the covid-19 vaccination is spreading more and more because of the lack of ability of the public to distinguish between valid news and hoaxes. This makes hoax news increase its spread quickly. So this study will utilize text mining in classifying hoax news texts for the Covid-19 vaccine using 4 kernels of Support Vector Machine. The data collected from 2 sources, hoax news title data are collected from Kominfo and non-hoax news title data are collected from Kompas. The research methodology used is Knowledge Discovery in Databases. The data will go through several preprocessing stages before testing the model. The results show that the best test results for the model obtained by the linear kernel in the 80:20 data split scenario which after going through an evaluation using a confusion matrix produces an accuracy value of 89,43%, precision 94,93%, recall 87,20%, and f1-score 90,90%.

Keywords: *Text Mining, Classification, Hoax, Support Vector Machine*

(*) Corresponding Author: aditya.wijaya18083@student.unsika.ac.id

How to Cite: Wijaya, A., Rozikin, C., & Sari, B. (2022). Application of Text Mining for Classification of COVID-19 Vaccination Hoax News Using Support Vector Machine Algorithm. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 11-20. <https://doi.org/10.5281/zenodo.7058890>.

PENDAHULUAN

Coronavirus atau COVID-19 merupakan sebuah virus yang dapat mengakibatkan pneumonia dan sindrom pernapasan akut kepada orangnya yang telah terjangkit oleh virus ini. Wabah penyakit ini pertama kali muncul pada akhir tahun 2019 hingga menjadi pandemik dan mewabah ke Indonesia pada tahun 2020. Maka dalam upaya penekanan dan mengurangi resiko terjangkit virus COVID-19 ini, pemerintah Indonesia melakukan program vaksinasi COVID-19 secara serentak pada 34 provinsi di seluruh wilayah Indonesia (Kemkes, 2021).

Di zaman yang sudah digital sekarang, mendapatkan informasi secara daring melalui platform digital menjadi sangat mudah. Dengan dimulainya program vaksinasi tersebut, masyarakat mulai mencari-cari informasi mengenai vaksinasi secara daring. Namun pelaksanaan program vaksinasi COVID-19 ini kurang berjalan mulus seperti yang diduga, karena maraknya penyebaran berita *hoax* seputar program dan vaksin COVID-19 ini di sosial media (Fasihullisan dkk., 2022). Berita *hoax* adalah berita yang berisi informasi yang belum pasti kebenarannya dan bukan berisi fakta yang sebenarnya (Juditha, 2018).



Menurut data yang dirilis per tanggal 30 maret 2022 oleh Kementerian Komunikasi dan Informatika di situs resminya, terdapat total 2.769 isu berita dan konten *hoax* tentang vaksin COVID-19 yang tersebar di lima platform digital berbeda. Penyebaran berita *hoax* ini terjadi semakin cepat karena masih banyak pembaca berita yang menerima berita *hoax* tersebut langsung mempercayai dan menyebarkannya kembali kepada orang lain tanpa mengetahui kebenaran informasi dari berita *hoax* tersebut. Menurut data hasil survei yang dilakukan oleh Katadata Insight Center (2020) kepada responden yang pernah menyebarkan *hoax* tentang penyebab alasan mereka menyebarkan *hoax*, hasilnya didapatkan sebanyak 68,4% menjawab bahwa mereka hanya meneruskan berita yang tersebar tanpa mencari tahu kebenarannya terlebih dahulu. Jika penyebaran berita *hoax* tentang program dan vaksin COVID-19 ini terus menyebar, maka dapat membuat keresahan dan perpecahan diantara masyarakat dan mengurangi kepercayaan masyarakat kepada pemerintah dan media (Priastuty dkk., 2020).

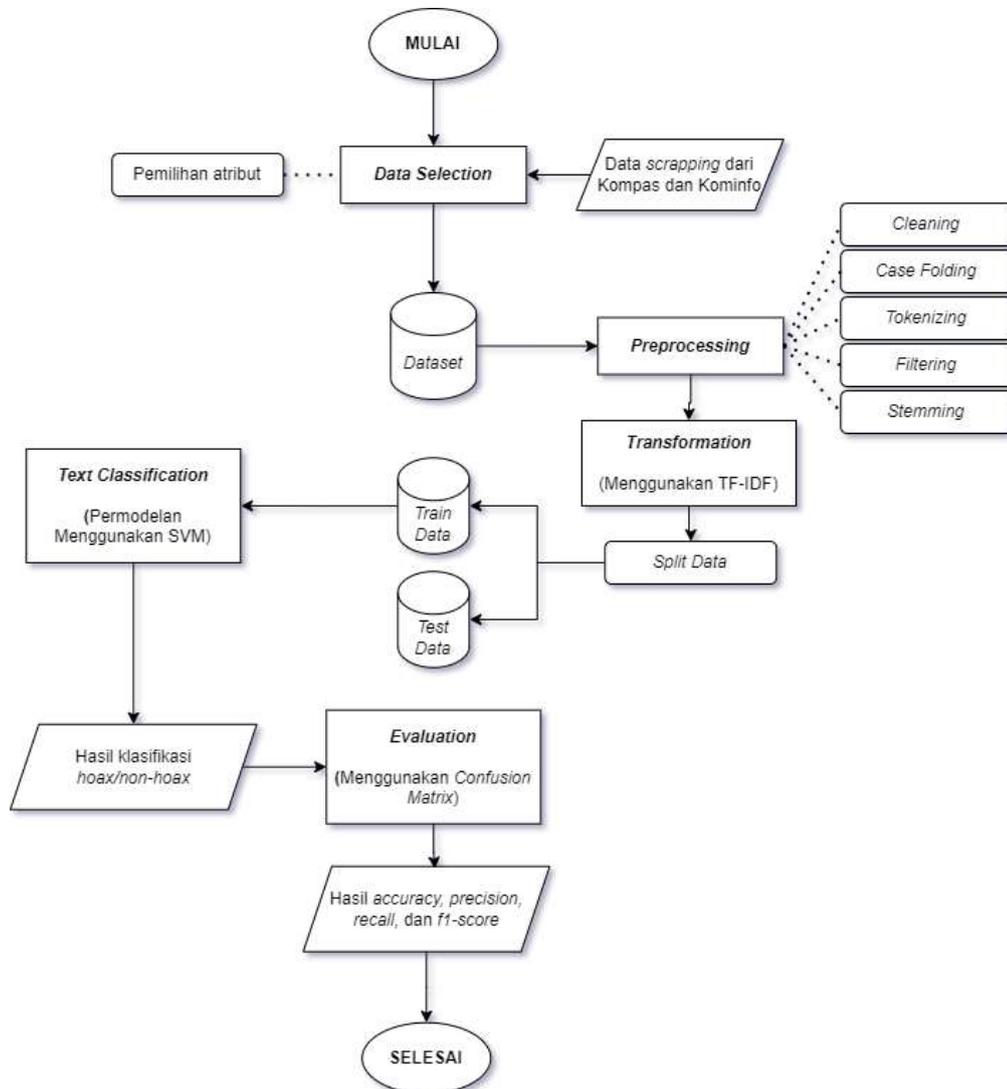
Salah satu cara mengetahui apakah sebuah berita mengandung informasi *hoax* atau tidak yaitu melakukan *text classification* dengan memanfaatkan teknik *text mining*. *Text mining* bisa digunakan dalam melakukan analisis sentimen maupun melakukan klasifikasi pada sebuah teks. Metode yang dapat dipakai untuk melakukan *text mining* salah satu diantaranya adalah *Support Vector Machine*. Pada SVM terdapat fungsi *kernel trick* yang mencari pemisah (*hyperplane*) dengan cara mentransformasikan dataset ke ruang vektor/kernel yang berdimensi lebih tinggi yang kemudian proses klasifikasi dilakukan pada *feature space* tersebut sehingga membuat SVM lebih baik dari metode klasifikasi lain. Beberapa fungsi kernel pada SVM yaitu kernel *Linear*, kernel *Radial Basis Function* (RBF), kernel *Polynomial*, dan kernel *Sigmoid* (Setiyono & Pardede, 2019). Hasil akurasi yang didapatkan SVM akan berbeda dan dipengaruhi oleh kernel yang dipilih (Mutawalli et al, 2019).

Pada penelitian yang membandingkan fungsi kernel *Linear* dan RBF dalam mengklasifikasi teks judul berita *hoax* tentang COVID-19, mendapatkan hasil bahwa kernel *Linear* menjadi kernel terbaik dengan tingkat akurasi sebesar 92,90% (Ropikoh et al, 2021). Pada penelitian lainnya yang dilakukan oleh Rahman et al (2021), yang membandingkan 3 fungsi kernel SVM yaitu kernel *Linear*, *Radial Basis Function* (RBF), dan *Sigmoid* untuk mengklasifikasi teks pada media sosial Twitter, hasil penelitian menunjukkan bahwa kernel RBF memiliki performa terbaik dengan mendapatkan tingkat akurasi 93%.

Berdasarkan uraian di atas, karena belum ada kepastian mengenai kernel SVM yang terbaik dalam melakukan *text classification*, maka pada penelitian ini akan melakukan penerapan dan perbandingan performa empat kernel SVM, yaitu antara kernel *Linear*, kernel *Radial Basis Function* (RBF), kernel *Polynomial*, dan kernel *Sigmoid* dalam mengklasifikasi judul berita *hoax* tentang vaksin COVID-19 yang bertujuan untuk mengetahui performa kernel SVM terbaik dalam melakukan klasifikasi teks.

METODOLOGI PENELITIAN

Metodologi yang diterapkan pada penelitian ini adalah KDD (*Knowledge Discovery in Databases*). Prosesnya terdiri dari lima tahapan, yaitu *data selection*, *preprocessing*, *transformation*, *text classification*, dan *evaluation*. Algoritma yang digunakan untuk klasifikasi adalah algoritma *Support Vector Machine*.



Gambar 1. Rancangan Penelitian

1. *Data Selection*

Pada penelitian ini data diambil dari dua sumber, dimana data kategori berita *hoax* diambil dari rekapan data yang dirilis oleh Kementerian Komunikasi dan Informatika sedangkan data dengan kategori berita *non-hoax* diambil dari situs berita *online* Kompas menggunakan teknik *web scraping*. Pada tahap *data selection* ini akan melakukan pemilihan atribut tabel yang hanya diperlukan saat penelitian.

2. *Preprocessing*

Tahap *preprocessing* adalah tahap melakukan pembersihan *dataset* sebelum dilakukan proses klasifikasi teks di tahap selanjutnya. Tahap *preprocessing* terdiri

dari beberapa proses seperti *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

3. Transformation

Pada tahap ini data dilakukan pembobotan setiap kata pada dokumen menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Lalu dilakukan *split data* latih dan uji hingga menjadi 3 skenario berbeda. Skenario 1 pembagian data dengan rasio 60:40, skenario 2 dengan rasio 70:30, dan skenario 3 dengan rasio 80:20.

4. Text Classification

Tahap ini akan melakukan pengujian model klasifikasi teks. Algoritma yang digunakan adalah *Support Vector Machine* yang dilakukan dengan 4 kernel, yaitu kernel *Linear*, kernel *Polynomial*, kernel *RBF*, dan kernel *Sigmoid*.

5. Evaluation

Model yang sudah dibuat akan dilakukan evaluasi dengan menggunakan *Confusion Matrix* yang akan menghasilkan nilai *accuracy*, *precision*, *recall* dan *f1-score*.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Penelitian ini akan menerapkan text mining dalam melakukan klasifikasi teks pada judul berita hoax vaksinasi covid-19 menggunakan algoritma *Support vector Machine*. Berikut tahap/proses dan hasil penelitian yang telah didapatkan.

1. Data Selection

Data Judul berita yang berkategori *hoax* dikumpulkan dari hasil rekapan kumpulan berita *hoax* vaksin COVID-19 yang dipublikasikan oleh Kementerian Komunikasi dan Informatika. Sedangkan data judul berita yang berkategori bukan *hoax* diambil dari situs berita Kompas dengan kata kunci ‘vaksin covid-19’ menggunakan teknik *web scraping* dengan bantuan *Google Colaboratory* dan *Python 3*. Data yang diambil adalah data judul berita tanggal 1 Maret 2021 sampai dengan 1 Maret 2022. Data yang dikumpulkan memiliki tiga atribut, yaitu tanggal, judul berita, dan label. Data diurutkan berdasarkan waktu dan tanggal saat berita tersebut diterbitkan. Data yang terkumpul berjumlah 3735 baris dataset yang terdiri dari 3382 data berita bukan *hoax* dan 353 data berita *hoax*. Atribut tabel yang dipilih adalah kolom ‘Judul’ dan kolom ‘Label’.

	A	B	C
	Tanggal	Judul	Label
1	01/03/2021	Kepala Peneliti Pfizer Sebut Vaksin Covid-19 untuk Sterilisasi Wanita	HOAX
2	01/03/2021	Dinkes Tangsel Belum Atur Jadwal Vaksinasi Covid-19 untuk Pedagang Pasar	NON-HOAX
3	01/03/2021	Tunggu Arahan Pusat, Pemkot Tangsel Belum Mulai Vaksinasi Covid-19 untuk Lansia	NON-HOAX
4	01/03/2021	Dapat 33.400 Dosis, Depok Mulai Vaksinasi Covid-19 Tahap Dua	NON-HOAX
5	01/03/2021	Vaksinasi Covid-19 Lansia Digelar di Sekolah Tiap Kecamatan di Jakpus	NON-HOAX
6	01/03/2021	KPU Siapkan Data Pemilih Pemilu 2019 dan Pilkada 2020 untuk Vaksinasi Covid-19	NON-HOAX
7	01/03/2021	Besok, KPU dan Kemenkes Tebak MUI Akas Data Pemilih untuk Vaksinasi Covid-19	NON-HOAX
8	01/03/2021	Vaksinasi Covid-19 di Lingkungan DPR Serjalan Tertib dan Sesuai Prokes	NON-HOAX
9	01/03/2021	UPDATE 1 Maret: 1.002.218 Nakes Sudah Divaksinasi Dosis Kedua, 1.720.529 Dosis Pertama	NON-HOAX
10	01/03/2021	Ini Alasan Pemerintah Gunakan Istilah Vaksinasi Gotong Royong	NON-HOAX
11	02/03/2021	Guru Besar UI: Keamanan Vaksin Covid-19 Harus Dievaluasi	NON-HOAX
12	01/03/2021	Disuntik Vaksin Covid-19, Pedagang Pasar di Kota Tangerang: Terima Kasih, Pak Jokowi	NON-HOAX
13	01/03/2021	7.450 Lansia di Banjarmasin Jalani Vaksinasi Covid-19	NON-HOAX
14	01/03/2021	2.500 Anggota Polda Bali Terima Vaksin Covid-19, Kapolda: Tetap Patuhi Protokol Kesehatan	NON-HOAX
15	01/03/2021	H-1 Vaksinasi, Mayoritas Pedagang Pasar di Tangsel Belum Dapat Sosialisasi	NON-HOAX
16	01/03/2021	Dinkes Kota Tangerang Tak Berikan Sanksi Administratif kepada Pedagang Penolak Vaksin Covid-19	NON-HOAX
17	01/03/2021	Soal Vaksinasi Mandiri, IDI Minta Pemerintah Perhatikan Pekerja yang Bukan Karyawan	NON-HOAX
18	01/03/2021	5 Hal yang Perlu Diketahui tentang Vaksinasi Drive Thru	NON-HOAX
19	01/03/2021	Dinkes Kota Bekasi Pertimbangkan Vaksinasi Covid-19 Tahap Dua Dilakukan Masual	NON-HOAX
20	01/03/2021	Tertarik 50.700 Pasisi Vaksinasi Covid-19, Kota Bekasi Berencana Lakukan Vaksinasi Tahap 2	NON-HOAX

Gambar 2. Contoh Dataset

2. *Preprocessing*

a. *Cleaning*

Komposisi data yang masih tidak seimbang antara jumlah judul berita *hoax* dengan *non-hoax*, sehingga akan dilakukan *undersampling data* agar komposisinya menjadi seimbang dan jumlah masing-masing data menjadi sama. Setelah melalui proses *undersampling*, jumlah data judul berita *non-hoax* berubah menjadi 353 data dan menjadi seimbang jumlahnya dengan data judul berita *hoax* dengan total jumlah data menjadi 706 baris data. Setelah melakukan *undersampling* pada data, maka dilakukan penghapusan atribut karakter seperti tanda baca, angka, dan emoji yang dianggap tidak diperlukan saat proses klasifikasi.

	Tanggal	Judul	Label
0	2021-03-01	Kepala Peneliti Pfizer Sebut Vaksin Covid-19 u...	HOAX
1	2021-03-01	Dinkes Tangsel Belum Atur Jadwal Vaksinasi Cov...	NON-HOAX
2	2021-03-01	Tunggu Arahan Pusat, Pemkot Tangsel Belum Mula...	NON-HOAX
3	2021-03-01	Dapat 33.400 Dosis, Depok Mulai Vaksinasi Covi...	NON-HOAX
4	2021-03-01	Vaksinasi Covid-19 Lansia Digelar di Sekolah T...	NON-HOAX

Gambar 3. Sebelum Proses *Cleaning*

	Tanggal	Judul	Label
0	2021-03-01	Kepala Peneliti Pfizer Sebut Vaksin Covid untu...	HOAX
1	2021-03-03	Surat Elektronik Mengatas Namakan IPD Balitban...	HOAX
2	2021-03-03	Pesan Berantai Vaksinasi Lansia di Kantor BPPS...	HOAX
3	2021-03-04	Covid Diciptakan Agar Manusia Divaksinasi	HOAX
4	2021-03-04	Potasium dalam Vaksin Covid Berbahaya bagi Org...	HOAX

Gambar 4. Setelah Proses *Cleaning*

b. *Case Folding*

Proses *case folding* ini merupakan proses dalam melakukan perubahan semua karakter alfabet yang semula berupa huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*).

	Tanggal	Judul	Label
0	2021-03-01	kepala peneliti pfizer sebut vaksin covid untu...	HOAX
1	2021-03-03	surat elektronik mengatas namakan ipd balitban...	HOAX
2	2021-03-03	pesan berantai vaksinasi lansia di kantor bpps...	HOAX
3	2021-03-04	covid diciptakan agar manusia divaksinasi	HOAX
4	2021-03-04	potasium dalam vaksin covid berbahaya bagi org...	HOAX

Gambar 5. Hasil Proses *Case Folding*

c. *Tokenizing*

Proses selanjutnya adalah *tokenizing*. Pada proses ini, data yang telah diubah menjadi *lowercase* akan dipecah menjadi kepingan/pecahan kata. Proses *tokenizing* ini menggunakan *library* NLTK pada *Python* 3.

	Tanggal	Judul	Label
0	2021-03-01	[kepala, peneliti, pfizer, sebut, vaksin, covi...	HOAX
1	2021-03-03	[surat, elektronik, mengatas, namakan, ipd, ba...	HOAX
2	2021-03-03	[pesan, berantai, vaksinasi, lansia, di, kanto...	HOAX
3	2021-03-04	[covid, diciptakan, agar, manusia, divaksinasi]	HOAX
4	2021-03-04	[potasium, dalam, vaksin, covid, berbahaya, ba...	HOAX

Gambar 6. Hasil Proses *Tokenizing*

d. Filtering

Pada proses *filtering* ini akan dilakukan penghapusan kata-kata yang tidak mempunyai makna dan tidak diperlukan untuk tahap klasifikasi nantinya. Kata-kata yang akan dihilangkan contohnya seperti kata ‘di’, ‘yang’, ‘dan’, ‘ke’ dan sebagainya. Proses ini akan menggunakan metode *stopword removal* dari *library* NLTK pada *Python 3*.

	Tanggal	Judul	Label
0	2021-03-01	[kepala, peneliti, pfizer, vaksin, covid, ster...	HOAX
1	2021-03-03	[surat, elektronik, mengatas, namakan, ipd, ba...	HOAX
2	2021-03-03	[pesan, berantai, vaksinasi, lansia, kantor, b...	HOAX
3	2021-03-04	[covid, diciptakan, manusia, divaksinasi]	HOAX
4	2021-03-04	[potasium, vaksin, covid, berbahaya, organ, tu...	HOAX

Gambar 7. Hasil Proses *Filtering*

e. Stemming

Proses terakhir dari tahap *preprocessing* ini adalah proses *stemming*. Pada proses ini, teks yang mempunyai imbuhan akan diubah menjadi kata dasar sesuai dengan Kamus Besar Bahasa Indonesia. *Stemming* ini akan dilakukan dengan menggunakan *library* Sastrawi pada *Python 3*.

	Tanggal	Judul	Label
0	2021-03-01	[kepala, teliti, pfizer, vaksin, covid, steril...	HOAX
1	2021-03-03	[surat, elektronik, atas, nama, ipd, balitbang...	HOAX
2	2021-03-03	[pesan, beranta, vaksinasi, lansia, kantor, bp...	HOAX
3	2021-03-04	[covid, cipta, manusia, vaksinasi]	HOAX
4	2021-03-04	[potasium, vaksin, covid, bahaya, organ, tubuh...	HOAX

Gambar 8. Hasil Proses *Stemming*

3. Transformation

Pada tahap *transformation* ini, dilakukan pembobotan dan ekstraksi fitur dengan TF-IDF. Hasil dari proses ini adalah terbentuknya matriks 706x1386. Pada matriks tersebut terdapat 706 dokumen dari *dataset* dan 1386 *term* yang memiliki nilai bobot kata atau nilai TF-IDF.

Gambar 9. Hasil *Transformation* dengan TF-IDF

Setelah melakukan transformasi, maka akan dilakukan *split data*. Pada penelitian ini terdapat tiga skenario *split data* yang dilakukan, yaitu 60% data *train* dan 40% data *test*, 70% data *train* dan 30% data *test*, serta 80% data *train* dan 20% data *test*. Data juga diberi parameter *random state=42*.

Tabel 1. Hasil *Split Data*

Rasio <i>Split Data</i>	Jumlah Data <i>Train</i>	Jumlah Data <i>Test</i>	Total
60:40	423	283	706
70:30	494	212	706
80:20	564	142	706

4. *Text Classification*

Pada tahap *text classification* ini, data yang sudah siap setelah melalui tahap *preprocessing* dan *transformation* akan dilakukan proses implementasi klasifikasi menggunakan algoritma *Support Vector Machine* dengan 4 kernel, yaitu kernel *Linear*, kernel *Polynomial*, kernel RBF, dan kernel *Sigmoid*. Proses ini dilakukan dengan tiga skenario pembagian data.

5. *Evaluation*

Tahap terakhir yang dilakukan adalah tahap *evaluation*. Di tahap ini dilakukan evaluasi terhadap performa dari implementasi dan pengujian klasifikasi menggunakan algoritma *Support Vector Machine*. Evaluasi model akan menggunakan *Confusion Matrix*. Hasil tahap evaluasi ini akan menentukan kernel SVM yang paling unggul dalam implementasi dan pengujian model klasifikasi yang sudah dilakukan.

Tabel 2. *Confusion Matrix* Hasil Pengujian

Rasio <i>Split Data</i> (Skenario)	Kernel	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
60:40 (Skenario 1)	<i>Linear</i>	85,51%	86,79%	87,34%	87,06%
	<i>Polynomial</i>	78,09%	98,00%	62,02%	75,96%

	RBF	86,21%	88,38%	86,70%	87,54%
	<i>Sigmoid</i>	86,21%	87,42%	87,97%	87,69%
	<i>Linear</i>	85,37%	87,28%	86,55%	86,92%
	<i>Polynomial</i>	83,49%	92,00%	77,31%	84,01%
70:30 (Skenario 2)	RBF	86,32%	88,13%	87,39%	87,76%
	<i>Sigmoid</i>	86,32%	88,13%	87,39%	87,76%
	<i>Linear</i>	89,43%	94,93%	87,20%	90,90%
	<i>Polynomial</i>	80,98%	94,03%	73,25%	82,35%
80:20 (Skenario 3)	RBF	87,32%	92,50%	86,04%	89,15%
	<i>Sigmoid</i>	88,73%	94,87%	86,04%	90,24%

Berdasarkan pada tabel di atas, perbandingan evaluasi pengujian memberikan hasil bahwa setelah melakukan pengujian 4 kernel SVM, didapatkan bahwa hasil performa kernel terbaik diperoleh kernel *Linear* pada skenario 3 dengan *split data* 80:20 yang menghasilkan nilai *accuracy* sebesar 89,43%, *precision* 94,93%, *recall* 87,20%, dan *f1-score* 90,90%.

PEMBAHASAN

Penelitian ini melakukan implementasi algoritma *Support Vector Machine* untuk mengklasifikasi berita tentang vaksinasi COVID-19 dari situs berita daring. Metodologi penelitian yang digunakan pada penelitian ini adalah metode *Knowledge Discovery in Databases (KDD)* yang terdiri dari beberapa tahapan, yaitu *data selection*, *preprocessing*, *transformation*, *text classification*, dan *evaluation*.

Data dikumpulkan dari website Kompas dan Kementerian Komunikasi dan Informatika. Data yang digunakan berjumlah 3735 yang terdiri dari 3382 judul berita *non-hoax* dan 353 judul berita *hoax*. Data dilakukan *undersampling* agar jumlahnya menjadi seimbang. Data diolah melalui tahap *preprocessing* untuk membersihkan data dan dilakukan *transformation* menggunakan TF-IDF. Setelah itu dilakukan *split data* hingga menjadi tiga skenario pembagian data. Selanjutnya melakukan tahap *text classification* menggunakan algoritma *Support Vector Machine* dengan empat buah kernel yang berbeda. Pengujian dilakukan dengan tiga skenario pembagian data, yaitu skenario 1 dengan rasio 60:40, skenario dengan rasio 70:30, dan skenario 3 dengan rasio 80:20. Terakhir melakukan *evaluation* dengan menggunakan *Confusion Matrix* hingga mendapatkan nilai akurasi, presisi, *recall*, dan *f1-score* untuk mengetahui perbandingan performa antar kernel SVM yang dihasilkan saat tahap *text classification*.

Hasil dari pengujian klasifikasi menghasilkan bahwa kernel *Linear* pada skenario 3 (80:20) memiliki performa paling baik dalam melakukan klasifikasi judul berita hoax dengan mendapatkan nilai *accuracy* sebesar 89,43%, *precision* 94,93%, *recall* 87,20%, dan *f1-score* 90,90%. dimana sebanyak 75 data judul berita *hoax* diklasifikasi dengan benar dan sebanyak 52 data judul berita *non-hoax* juga diklasifikasikan dengan benar. Sedangkan hasil performa kernel terendah diperoleh kernel *Polynomial* pada skenario 1 (60:40) yang menghasilkan nilai *accuracy* sebesar 78,09%, *precision* 98,00%, *recall* 62,02%, dan *f1-score* 75,96%.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan, yaitu:

1. Proses melakukan klasifikasi teks untuk berita hoax vaksinasi COVID-19 menggunakan penerapan algoritma *Support Vector Machine* dengan keempat kernelnya, yaitu kernel *Linear*, *Polynomial*, RBF, dan *Sigmoid* dilakukan melalui beberapa tahapan. Tahap awal adalah melakukan pengumpulan dan pemilihan data judul berita *hoax* dan *non-hoax* dari situs berita daring terpercaya. Setelah itu data melalui tahap *preprocessing* dan *transformation* menggunakan TF-IDF. Setelahnya adalah melakukan klasifikasi teks dengan pemodelan dan pengujian yang dilakukan dengan 3 skenario pembagian data hingga mendapatkan model terbaik dari keempat kernel *Support Vector Machine* yang diuji.
2. Pengujian performa masing-masing keempat kernel *Support Vector Machine* untuk melakukan klasifikasi teks berita *hoax* vaksinasi COVID-19 dilakukan dengan evaluasi menggunakan *Confusion Matrix* dan mendapatkan nilai akurasi, presisi, *recall*, dan *f1-score* dari masing-masing kernel dan skenario yang dilakukan.
3. Hasil evaluasi perbandingan performa antar keempat kernel SVM mendapatkan hasil bahwa kernel *Linear* pada skenario pembagian data ketiga memiliki performa yang paling baik dengan mendapatkan nilai *accuracy* sebesar 89,43%, *precision* 94,93%, *recall* 87,20%, dan *f1-score* 90,90%. Sedangkan hasil performa terendah diperoleh kernel *Polynomial* pada skenario pembagian data pertama yang hanya mendapatkan nilai *accuracy* sebesar 78,09%, *precision* 98,00%, *recall* 62,02%, dan *f1-score* 75,96%.

REFERENCES

- Direktorat Jenderal Pencegahan dan Pengendalian Penyakit. (2021). *Program Vaksinasi COVID-19 Mulai Dilakukan, Presiden Orang Pertama Penerima Suntikan Vaksin COVID-19*. Jakarta.
- Euis Saraswati, Yuyun Umaidah, & Apriade Voutama. (2021). Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19. *Generation Journal*, 5(2), 109–118.
- Fashihullisan, M., Martini, M., & Iriyanti, S. (2022). Peredaran Berita Bohong Program Vaksinasi COVID 19 di Media Sosial Facebook. *Jurnal Baksooka*.
- Juditha, C. (2018). Interaksi Komunikasi Hoax di Media Sosial Serta Antisipasinya. *Journal Pekommas*, 3(1), 31–34.

- Juditha, C. (2020). Perilaku Masyarakat Terkait Penyebaran Hoaks Covid-19. *Journal Pekommas*, 5(2), 105.
- Muhammad, F., Mulyana Maghfur, N., & Voutama, A. (2022). Sentiment Analysis Dataset on COVID-19 Variant News: Kumpulan Data Analisis Sentimen pada Berita Varian COVID-19. *SYSTEMATICS*, 4(1), 382–391.
- Mutawalli, L., Zaen, M. T. A., & Bagye, W. (2019). Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto). *Jurnal Informatika Dan Rekayasa Elektronik*, 2(2), 43.
- Nana Mulyana Maghfur, Fakhri Muhammad, & Apriade Voutama. (2021). Analysis of the Relationship between Public Sentiment on Social Media and Indonesian Covid-19 Dynamics: Analisis Hubungan Sentimen Publik di Media Sosial dengan Dinamika Covid-19 Indonesia. *SYSTEMATICS*, 3(3), 336–345.
- Oryza Habibie Rahman, Gunawan Abdillah, & Agus Komarudin. (2021). Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 17–23.
- Priastuty, C. W., Rahmanto, A. N., Maret, U. S., Surakarta, K., Maret, U. S., Surakarta, K., ... Surakarta, K. (2020). Hoaks tentang Vaksin Covid-19 di Tengah Media Sosial. *Prosiding Seminar Nasional Unimus*, 3(1), 391–399.
- Pusparisa, Y. (2020). *Ragam Sebab Masyarakat Sebar Hoaks*. Databoks.
- Ropikoh, I. A., Abdulhakim, R., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19. *Journal of Applied Informatics and Computing*, 5(1), 64–73.
- Setiyono, A., & Pardede, H. F. (2019). Klasifikasi SMS Spam Menggunakan Support Vector Machine. *Jurnal Pilar Nusa Mandiri*, 15(2), 275–280.
- Utomo, W. H., & Prayoga, K. J. (2021). Hoax Classification Corona Virus (COVID-19) News in Indonesian using the Support Vector Machine (SVM) Method. *Journal of Computer Science*, 17(8), 692–708.
- Yoga Pratama, A., Umaidah, Y., Voutama, A., Informatika, T., Ilmu Komputer, F., Singaperbangsa Karawang Ds Paseurjaya, U., ... Barat, J. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, pp. 897–910).