

Penerapan Algoritma *Support Vector Machine (SVM)* untuk Klasifikasi Berita *Hoax* Covid-19

Isnin Apriyatin Ropikoh^{1*}, Rijal Abdulhakim², Ultach Enri^{3*}, Nina Sulistiyowati^{4*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

isnin.apriyatin17029@student.unsika.ac.id¹, rijal.abulhakim16182@student.unsika.ac.id², ultach@staff.unsika.ac.id³, nina.sulistio@unsika.ac.id⁴

Article Info

Article history:

Received 01-07-2021

Revised 16-07-2021

Accepted 19-07-2021

Keyword:

Covid-19,
Hoax,
KDD,
Klasifikasi,
SVM.

ABSTRACT

Seiring dengan perkembangan teknologi informasi, media berita juga turut berkembang menyajikan informasi dalam media *online*. Namun, sangat disayangkan pada penyebarannya masih banyak ditemukan berita *hoax* atau tidak benar. Berita *hoax* yang paling mudah beredar adalah *hoax* tentang kesehatan. Di Indonesia sendiri semenjak diberitakan masuknya virus Covid-19, berita *hoax* tentang hal itu terus meningkat berdasarkan data yang dirilis oleh Kominfo periode Januari-Agustus 2020. Agar terhindar dari berita *hoax* ialah dengan lebih teliti membaca judul berita pada situs yang terpercaya seperti Kompas. Karena itu penelitian ini akan mengembangkan dan menganalisis model klasifikasi berita *hoax* Covid-19 dengan menerapkan algoritma *Support Vector Machine (SVM)* dengan metodologi *Knowledge Discovery in Databases (KDD)*. Studi kasus penelitian ini dibagi dalam 2 kategori yaitu berita *hoax* yang didapat dari situs *Trunbackhoax & Hoax buster* sedangkan berita bukan *hoax* diambil dari situs berita Kompas. Hasil penelitian menyatakan bahwa Algoritma *Support Vector Machine (SVM)* dengan *kernel linear* memiliki hasil prediksi yang bagus pada skenario 3 (80:20) karena model sanggup dalam mengklasifikasikan berita *hoax* dan bukan *hoax* Covid-19. Akurasi yang didapat pada skenario 3 juga memiliki nilai akurasi tertinggi sebesar 97,06%. Sedangkan pada *kernel RBF* memiliki akurasi terendah pada skenario 4 (90:10) yaitu 90.46% dan model kurang bagus dalam mengklasifikasikan berita *hoax* maupun bukan *hoax* Covid-19.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

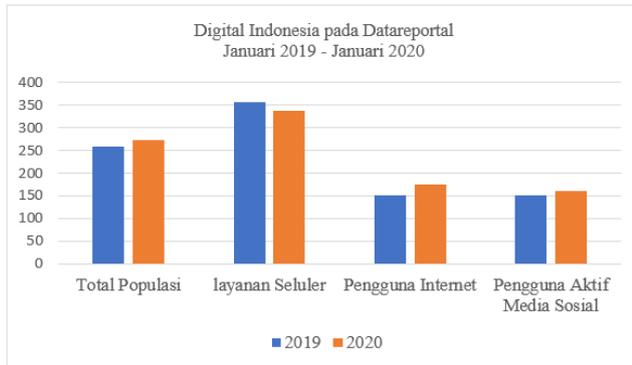
I. PENDAHULUAN

Berdasarkan Kamus Besar Bahasa Indonesia (KBBI) 2021, dijelaskan bahwa informasi tidak benar atau palsu yang didapat tanpa sumber yang jelas merupakan pengertian dari *hoax* [1]. Sedangkan dalam *Oxford English Dictionary* [2], *hoax* didefinisikan sebagai suatu tindakan atau perbuatan yang dibuat dengan tujuan untuk membuat orang lain mempercayai sesuatu yang tidak benar. Masyarakat memiliki penjelasan tersendiri mengenai *hoax*. Berdasarkan survey yang dilakukan oleh “Masyarakat Telematika Indonesia (MASTEL)” pada 2019 dan diikuti sebanyak 941 responden mendapatkan hasil bahwa 88% memilih penjelasan berita *hoax* sebagai berita tidak benar dan sengaja dibuat, 49% menjawab berita penghasut, 15% menyebutkan *hoax*

merupakan berita perkiraan atau yang berkaitan dengan sains, 14% menyebutkan sebagai berita yang membuat pemerintah merasa terpojokan, 31% menganggap berita *hoax* untuk menjelekan orang lain dan 1% lainnya menjawab tidak tahu mengenai definisi *hoax* [3].

Berdasarkan data dari situs datareportal, pengguna internet dari januari tahun 2019 [4] sampai januari 2020 [5] mengalami peningkatan sekitar 25 juta pengguna. Menurut Jamalul izza, peningkatan terjadi disebabkan oleh pemerataan infrastruktur internet dengan adanya Palapa Ring [6]. Hal ini merupakan salah satu faktor yang mempengaruhi penyebaran informasi menjadi tidak terkendali, sehingga berbagai informasi dapat diakses oleh masyarakat melalui internet tanpa validasi, dimana informasi itu dapat berupa

berita palsu atau *hoax*. Peningkatan jumlah pengguna internet dapat dilihat pada gambar 1 berikut.



Gambar 1. Digital Indonesia dari Datareportal

Kata berita sendiri menurut Kamus Besar Bahasa Indonesia (KBBI), 2021 merupakan suatu informasi yang menunjukkan tempat atau keterangan waktu dan biasanya terjadi di masyarakat [7]. Seperti telah disinggung sebelumnya, penyebaran berita dapat berupa informasi palsu atau kebohongan yang akan sulit dikendalikan apalagi dengan kondisi yang belum dapat dipastikan. Sekretaris Dewan Kehormatan Persatuan Wartawan Indonesia (PWI), Wina Armanda Sukardi menyampaikan bahwa berita *hoax* tentang kesehatan adalah yang paling banyak beredar menurut survey yang dilakukan sejak Februari 2016 sampai Februari 2017. Dengan kurang lebih seribu sampel, *hoax* kesehatan mendapat nilai tertinggi yaitu sebanyak 27%, posisi kedua merupakan berita *hoax* politik sebanyak 22%, berita *hoax* hiburan sebanyak 15% dan sisanya merupakan berita *hoax* mengenai bisnis dan lainnya [8]. Dapat diketahui bahwa berita *hoax* kesehatan adalah *hoax* yang populer jika dibandingkan dengan berita lainnya. Berdasarkan data yang dirilis oleh Kominfo melalui *website* resminya telah mencatat dari bulan Januari – Agustus 2020 bahwa sebanyak 1028 hoaks tersebar terkait berita Covid-19. Dalam hal ini, berita hoaks covid merupakan salah satu berita hoaks kesehatan yang paling banyak penyebarannya seperti ditamikan pada table berikut.

TABEL I
DATA HOAX TERKAIT COVID-19

Bulan	Berita Hoax
Januari	40
Februari	100
Maret	265
April	219
Mei	172
Juni	102
Juli	108
Agustus (8 Agustus)	22
Total	1.028

(Sumber: https://kominfo.go.id/content/detail/28536/kominfo-mencatat-sebanyak-1028-hoaks-tersebar-terkait-covid-19/0/sorotan_media)

Di masa pandemi covid-19 ini, banyak masyarakat mencari informasi mengenai virus seperti awal mula

terbentuknya virus, gejala, cara pengobatan dan berbagai informasi lainnya. Namun sejak diberitakan masuknya virus Covid-19 ke Indonesia pada 2 Maret 2020 oleh Presiden RI Joko Widodo yang menjangkit 2 warga negara Indonesia [9], membuat penyebaran informasi *hoax* mengenai virus tersebut semakin marak dan tidak terkendali seperti dapat dilihat pada gambar 1.

Pada tahun 2020 dilakukan survey oleh Juditha [10] dengan 625 responden dengan kategori pengguna internet yang menetap di Indonesia mendapatkan hasil bahwa masyarakat Indonesia mendapatkan informasi mengenai Covid-19 dari media *online* (situs berita) sebanyak 89,9% dimana merupakan yang terbanyak, lalu dari media sosial sebanyak 83,70%, dari televisi 78,20%, melalui pesan instan sebanyak 69,60% dan 65% dari *website* resmi pemerintah, sedangkan 60% lainnya dari media lain. Sehingga berdasarkan survey tersebut, media *online* merupakan sumber informasi yang paling ditunggu.

Salah satu cara untuk tidak terjerumus pada berita *hoax* adalah dengan membaca berita atau informasi dari media terpercaya. Kompas salah satu situs berita *online* yang mendapat dua kali penghargaan oleh Suoerbrands untuk kategori berita *online* terpercaya [11]. Selain itu kompas secara resmi tersertifikasi dalam jaringan internasional penguji fakta (IFCN – *International Fact-Checking*) [12].

Berita mengenai *hoax* dapat diakses melalui situs *Turnbackhoax & Hoax-buster*. *Turnbackhoax* merupakan situs pengaduan berbagai jenis *hoax*. Situs ini dibuat oleh KOMINFO & MASTEL [13]. Sedangkan *hoax buster* merupakan bagian dari *website* www.covid19.co.id yang secara resmi telah diluncurkan oleh gugus tugas penanganan covid-19 dibuat dengan tujuan sebagai acuan apakah informasi tersebut merupakan *hoax* berita covid-19 di Indonesia [14].

Menurut Monarizqa et al [15], *text mining* merupakan proses penambangan pengetahuan dan informasi data berupa teks atau dokumen. Hal ini bertujuan untuk mendapatkan informasi dari sekumpulan format teks yang tidak terstruktur [16]. Salah satu bagian dari *text mining* adalah klasifikasi, dimana klasifikasi ialah proses prediksi pada kelas objek yang belum diketahui kelas dan karakteristik tipe data klasifikasinya [17]. Proses klasifikasi dapat menggunakan algoritma *Decision Tree*, *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Neural Network (NN)*, *Random Forest (RF)* dan *Support Vektor Machine (SVM)*.

Algoritma *Support Vector Machine (SVM)* merupakan metode yang paling populer digunakan untuk klasifikasi berita *hoax* menurut survey yang dilakukan pada 2016-2018 [16]. Penelitian tersebut menggunakan fungsi *kernel SVM* [18] yaitu *kernel Linear* dengan metode *cross validation fold 10* mendapatkan nilai akurasi sebesar 95,84%. Perhitungan akurasi SVM lebih unggul dengan nilai akurasi pada proses *k-fold validation* sebesar 88,76% dibandingkan dengan Algoritma KNN dengan akurasi 88,1%. Kemudian perbandingan SVM dengan *Multinomial Naive Bayes*

(MNB)[19] mendapatkan hasil SVM lebih unggul dengan akurasi sebesar 82%, presisi 84% dan recall 82%. Sedangkan MNB memperoleh akurasi sebesar 62%, presisi 68% dan recall 62%.

Penelitian yang membandingkan empat fungsi *linear* untuk mengklasifikasikan *helpdesk* dengan metode SVM menghasilkan akurasi paling tinggi pada *kernel RBF* atau *gaussian* dengan nilai 81%, *kernel linear* dan *polynomial* mendapatkan nilai setara dengan 78% serta *kernel sigmoid* mendapatkan nilai terendah dengan 51% [20]. Kemudian pada penelitian lain yang serupa [21] mendapatkan hasil akurasi tertinggi pada RBF 81,44%, *kernel linear* sebesar 81,26% dan *polynomial* 67,12%. Namun pada nilai presisi, *recall* dan *f1-score kernel linear* justru mendapatkan nilai tertinggi dengan 90%, 89% dan 89% secara berurutan.

Berdasarkan pada penelitian sebelumnya, maka penelitian ini akan menggunakan metode klasifikasi dengan menetapkan algoritma *Support Vector Machine (SVM)* dengan dua fungsi *kernel linear* dan *kernel RBF*. Serta dataset yang digunakan diambil dari situs berita *online Kompascom* untuk berita bukan *hoax* dan situs *Turnbackhoax & hoaxbuster* untuk berita *hoax* pada periode Maret 2020 sampai Desember 2020.

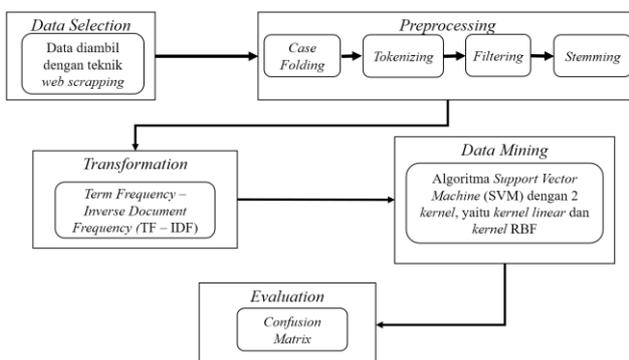
II. METODE PENELITIAN

A. Metodologi Penelitian

Penelitian menggunakan Algoritma *Support Vector Machine (SVM)* untuk klasifikasi berita *hoax* Covid-19. Dengan metodologi *Knowledge Discovery in Database (KDD)*. Metode KDD ini dapat berguna menemukan suatu teknik dari *database* yang ada. Terdapat 5 proses *text mining* dalam tahapan KDD yaitu, *data selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation*.

B. Rancangan Penelitian

Penelitian ini akan menggunakan bahasa pemrograman *Python 3*, dengan rancangan penelitian seperti berikut.



Gambar 2. Rancangan Penelitian.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Penelitian ini menggunakan proses *text mining*, hasil penelitian yang sudah dilakukan yaitu bagaimana menerapkan algoritma *Support Vector Machine (SVM)* dengan dua fungsi *kernel* yaitu *kernel linear* dan *kernel RBF* untuk menghasilkan berita *hoax* dan bukan *hoax* dengan menggunakan data berita Covid-19 yang diambil dari situs *online*.

1) *Data Selection*: Data diambil menggunakan teknik *web scrapping* dengan *Python 3*. Data awal memiliki 3 atribut yaitu Judul, Tanggal dan Sumber. Kemudian data dikonversikan berdasarkan tanggal agar data konsisten. Data yang digunakan adalah data berita Covid-19 pada tanggal 1 Maret 2020 sampai 31 Desember 2020 yang diperoleh dari 3 situs berita *online* dengan kategori berita *hoax* dan berita bukan *hoax*. Data yang akan diproses berjumlah 8.172 *dataset* yang berisikan 837 berita *hoax* dan 7.335 berita bukan *hoax*. Contoh data judul berita dapat dilihat pada tabel berikut.

TABEL II
CONTOH DATASET

No.	Judul	Sumber	Label
1	[SALAH] Orang yang Bergolongan Darah O Lebih Rentan Terinfeksi Virus Corona	Turnbackhoax	0
2	[SALAH] Area Kantor Nokia di Menara Mulia Dibersihkan Setelah Ada Karyawan Meninggal	Turnbackhoax	0
3	[SALAH] "Jokowi hari ini tolak disuntik vaksin Covid-19 duluan"	Hoaxbuster	0
4	[SALAH] Video Perawat Pingsan Setelah Mendapat Vaksin Covid-19, Potensi Genosida Massal	Hoaxbuster	0
5	Update Virus Corona 2 Maret: 89.212 Orang di 68 Negara Terinfeksi	Kompas	1
6	Bertambah 34 Negara dalam Sepekan, Ini Daftar 64 Negara Terinfeksi Virus Corona	Kompas	1
7	Update Terbaru Virus Corona: Positif di 64 Negara, 3.006 Orang Meninggal, 88.227 Terinfeksi	Kompas	1
8	Hadapi Corona, Wali Kota Jakbar Imbau Stop Panic Buying	Kompas	1

Pelabelan dilakukan secara manual dengan ketentuan label (0) untuk berita *hoax* dengan alasan mengandung kata atau isi berita yang tidak sesuai dengan fakta sebenarnya berasal dari situs *turnbackhoax.id*, sedangkan berita bukan *hoax* diberi label (1) yang berasal dari situs berita *online Kompascom* dan *Detikcom* dimana merupakan berita dengan isi berita mengandung fakta sebenarnya yang terjadi pada masyarakat.

2) *Pre-processing*: pada bagian ini terdapat beberapa tahap yang harus dilakukan yaitu tahap *case folding*, *tokenizing*, *filtering* dan *stemming*.

a) Pada tahap *case folding* dilakukan penyeragaman kata, dimana huruf atau kata yang semula kapital (*uppercase*)

diubah menjadi huruf kecil (*lowercase*). Berikut contoh tahap *case folding*:

```
0 "Corona itu adanya di China bukan di sini. Di...
1 "menghilangkan virus adalah mengkonsumsi lebi...
2 Kunyah Daun Sirih Untuk Mengatasi Covid-19
3 "Minum Kopi 3 x Sehari untuk menangkal Covid-19"
4 Pelarangan Masker Scuba Politik Perusahaan
...
7748 The Next Indonesia Big Stars, Ajang Pencarian ...
7749 BIN Sumbang Alat Medis dan Obat ke Gugus Tugas...
7750 Pandemi Covid-19, Perusahaan Negara Ini Tetap ...
7751 Menjawab Pertanyaan Anak, Kapan Pandemi Covid-...
7752 Perawat RSUP Kariadi Positif Covid-19 Meningga...
```

Gambar 3. Sebelum Proses *Case Folding*

```
0 corona itu adanya di china bukan di sini di ...
1 menghilangkan virus adalah mengkonsumsi lebih...
2 kunyah daun sirih untuk mengatasi covid
3 minum kopi x sehari untuk menangkal covid
4 pelarangan masker scuba politik perusahaan
...
7748 the next indonesia big stars ajang pencarian ...
7749 bin sumbang alat medis dan obat ke gugus tugas...
7750 pandemi covid perusahaan negara ini tetap r...
7751 menjawab pertanyaan anak kapan pandemi covid ...
7752 perawat rsup kariadi positif covid meninggal...
```

Gambar 4. Setelah Proses *Case Folding*

b) *Tahap Tokenizing* merupakan proses selanjutnya setelah dilakukan *Case Folding*, dimana proses ini merupakan kegiatan dengan tujuan untuk memecah sebuah kalimat ke dalam kepingan-kepingan kata agar dapat diketahui kata mana yang merupakan kata benda, kata sifat, kata kerja, kata sambung dan tanda baca pada sebuah kalimat, kemudian kata yang tidak dibutuhkan dapat dihilangkan.

```
0 corona itu adanya di china bukan di sini di ...
1 menghilangkan virus adalah mengkonsumsi lebih...
2 kunyah daun sirih untuk mengatasi covid
3 minum kopi x sehari untuk menangkal covid
4 pelarangan masker scuba politik perusahaan
...
7748 the next indonesia big stars ajang pencarian ...
7749 bin sumbang alat medis dan obat ke gugus tugas...
7750 pandemi covid perusahaan negara ini tetap r...
7751 menjawab pertanyaan anak kapan pandemi covid ...
7752 perawat rsup kariadi positif covid meninggal...
```

Gambar 5. Sebelum Proses *Tokenizing*

```
0 [corona, itu, adanya, di, china, bukan, di, si...
1 [menghilangkan, virus, adalah, mengkonsumsi, l...
2 [kunyah, daun, sirih, untuk, mengatasi, covid]
3 [minum, kopi, x, sehari, untuk, menangkal, covid]
4 [pelarangan, masker, scuba, politik, perusahaan]
...
7748 [the, next, indonesia, big, stars, ajang, penc...
7749 [bin, sumbang, alat, medis, dan, obat, ke, gug...
7750 [pandemi, covid, perusahaan, negara, ini, teta...
7751 [menjawab, pertanyaan, anak, kapan, pandemi, c...
7752 [perawat, rsup, kariadi, positif, covid, menin...
```

Gambar 6. Setelah Proses *Tokenizing*

c) *Filtering* merupakan tahap pengambilan kata penting dari hasil *tokenizing*. Penghilangan kata sambung atau kata tidak penting dilakukan menggunakan teknik

removes stopword. *Stopword* adalah kata tidak deskriptif yang dapat dibuang ke dalam *bag of word*.

```
0 [corona, itu, adanya, di, china, bukan, di, si...
1 [menghilangkan, virus, adalah, mengkonsumsi, l...
2 [kunyah, daun, sirih, untuk, mengatasi, covid]
3 [minum, kopi, x, sehari, untuk, menangkal, covid]
4 [pelarangan, masker, scuba, politik, perusahaan]
...
7748 [the, next, indonesia, big, stars, ajang, penc...
7749 [bin, sumbang, alat, medis, dan, obat, ke, gug...
7750 [pandemi, covid, perusahaan, negara, ini, teta...
7751 [menjawab, pertanyaan, anak, kapan, pandemi, c...
7752 [perawat, rsup, kariadi, positif, covid, menin...
```

Gambar 7. Sebelum Proses *Filtering*

```
0 [corona, adanya, china, bukan, sini, sini, cum...
1 [menghilangkan, virus, mengkonsumsi, lebih, ba...
2 [kunyah, daun, sirih, mengatasi, covid]
3 [minum, kopi, x, sehari, menangkal, covid]
4 [pelarangan, masker, politik, perusahaan]
...
7748 [the, indonesia, big, stars, ajang, pencarian,...
7749 [bin, sumbang, alat, medis, obat, gugus, tugas...
7750 [pandemi, covid, perusahaan, negara, tetap, ra...
7751 [menjawab, pertanyaan, anak, kapan, pandemi, c...
7752 [perawat, rsup, kariadi, positif, covid, menin...
```

Gambar 8. Setelah Proses *Filtering*

d) *Stemming* merupakan tahap terakhir dari *pre-processing*. Tahap ini diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, dapat pula melakukan penghapusan kata berimbuhan dan merubahnya menjadi kata dasar. Proses ini menggunakan *Library Phyton Sastrawi* yang merupakan *library* sederhana yang dapat mengolah kata berimbuhan bahasa Indonesia menjadi bentuk kata dasarnya.

```
0 [corona, adanya, china, bukan, sini, sini, cum...
1 [menghilangkan, virus, mengkonsumsi, lebih, ba...
2 [kunyah, daun, sirih, mengatasi, covid]
3 [minum, kopi, x, sehari, menangkal, covid]
4 [pelarangan, masker, politik, perusahaan]
...
7748 [the, indonesia, big, stars, ajang, pencarian,...
7749 [bin, sumbang, alat, medis, obat, gugus, tugas...
7750 [pandemi, covid, perusahaan, negara, tetap, ra...
7751 [menjawab, pertanyaan, anak, kapan, pandemi, c...
7752 [perawat, rsup, kariadi, positif, covid, menin...
```

Gambar 9. Sebelum Proses *Stemming*

```
0 [corona, ada, china, bukan, sini, sini, cuma, ...
1 [hilang, virus, konsumsi, lebih, banyak, makan...
2 [kunyah, daun, sirih, atas, covid]
3 [minum, kopi, x, hari, tangkal, covid]
4 [larang, masker, politik, usaha]
...
7748 [the, indonesia, big, stars, ajang, cari, baka...
7749 [bin, sumbang, alat, medis, obat, gugus, tugas...
7750 [pandemi, covid, usaha, negara, tetap, rampung...
7751 [jawab, tanya, anak, kapan, pandemi, covid, ak...
7752 [awat, rsup, kariadi, positif, covid, tinggal,...
```

Gambar 10. Setelah Proses *Stemming*

3) *Transformation*: pada tahap ini dilakukan pembobotan kata dengan mencari nilai *Tern Frequency* (TF), namun sebelumnya dilakukan penghitungan jumlah kata pada judul berita seperti pada gambar di bawah ini:

```
[{'ada': 1,
'adakan': 1,
'bukan': 1,
'china': 1,
'corona': 1,
'cuma': 1,
'sini': 2},
{'alkali': 1,
'asam': 1,
'atas': 1,
'banyak': 1,
'hilang': 1,
'konsumsi': 1,
'lebih': 1,
'makan': 1,
'minum': 1,
'tingkat': 1,
'virus': 2},
{'atas': 1, 'covid': 1, 'daun': 1, 'kunyah': 1, 'sirih': 1}]
```

Gambar 11. Jumlah Kata pada Judul Berita

```
[{'ada': 0.125,
'adakan': 0.125,
'bukan': 0.125,
'china': 0.125,
'corona': 0.125,
'cuma': 0.125,
'sini': 0.25},
{'alkali': 0.08333333333333333,
'asam': 0.08333333333333333,
'atas': 0.08333333333333333,
'banyak': 0.08333333333333333,
'hilang': 0.08333333333333333,
'konsumsi': 0.08333333333333333,
'lebih': 0.08333333333333333,
'makan': 0.08333333333333333,
'minum': 0.08333333333333333,
'tingkat': 0.08333333333333333,
'virus': 0.16666666666666666},
{'atas': 0.2, 'covid': 0.2, 'daun': 0.2, 'kunyah': 0.2, 'sirih': 0.2}]
```

Gambar 12. Hasil Term Frequency (TF)

Kemudian setelah nilai TF dan IDF ditemukan dilakukan pencarian nilai TF-IDF dengan mengkalikan hasil keduanya.

	aa	aamir	aaron	abai	abang	abdel	abduh	abdul	abigail	abk	abon	absen	abu	ac	acara	accor	aceh	achmad	aci	
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 14. Hasil TF-IDF

Perhitungan nilai TF secara manual dilakukan dengan persamaan (1).

$$tf_i = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)} \tag{1}$$

Dilanjutkan dengan mencari nilai IDF dengan persamaan (2).

$$idf_i = \log \frac{|D|}{|\{d:t_i \in d\}|} \tag{2}$$

Setelah kedua nilai didapatkan, dilakukan perhitungan manual nilai TF-IDF dengan persamaan (3).

$$(tf - idf)_{ij} = tf_i * idf_i \tag{3}$$

1) *Data Mining*: pada tahap ini dilakukan pemodelan menggunakan algoritma *Support Vector Machine (SVM)*

Setelah mendapatkan nilai TF selanjutnya dilakukan pencarian nilai IDF seperti pada gambar di bawah ini.

```
{'a': 6.653250051224913,
'aa': 7.857222855550849,
'aamir': 8.955835144218959,
'aaron': 8.955835144218959,
'abai': 6.470928494430959,
'abang': 7.569540783099068,
'abdel': 8.955835144218959,
'abduh': 8.955835144218959,
'abdul': 8.955835144218959,
'abigail': 8.955835144218959,
'abk': 6.247784943116749,
'abon': 8.955835144218959,
'absen': 8.955835144218959,
'abu': 8.955835144218959,
'ac': 6.557939871420588,
'acara': 5.864792690860643,
'accor': 8.955835144218959,
'aceh': 6.011396165052519,
'achmad': 7.1640756749909045,
'aci': 8.955835144218959,
'ad': 8.262687963659014,
'ada': 5.911312706495536,
'adakan': 8.262687963659014,
'adam': 7.3463972317848585,
'adaptasi': 8.955835144218959,
'adat': 7.857222855550849,
```

Gambar 13. Hasil Inverse Document Frequency (IDF)

dengan 2 fungsi *kernel* yaitu *linear* dan *RBF*. Pemodelan ini dilakukan pembagian data *training* dan data *testing* menjadi 4 skenario yaitu (60%:40%), (70%:30%), (80%:20%) dan (90%:10%). Pembagian ini dilakukan untuk mendapatkan nilai performa. Hasil pengolahan klasifikasi akan menghasilkan nilai tingkat akurasi, *precision*, *recall*, *f-measure* dan *error rate*. Pembagian data *training* dan data *testing* dapat dilihat pada tabel berikut.

TABEL III
PEMBAGIAN DATA TRAINING DAN DATA TESTING

Skenario 1	Data training (60%)	Data testing (40%)
Skenario 2	Data training (70%)	Data testing (30%)
Skenario 3	Data training (80%)	Data testing (20%)
Skenario 4	Data training (90%)	Data testing (10%)

Kemudian dilakukan pengujian terhadap tiap skenario dengan menggunakan 2 fungsi *kernel* yaitu *kernel linear* dan *kernel RBF*.

Skenario 1 (60:40): hasil dari pengujian SVM pada pemodelan klasifikasi berita *hoax* dan berita bukan *hoax* dengan skenario 1 adalah sebagai berikut.

• Pengujian dengan *Kernel Linear*

Kernel linear

	predicted:hoax	predicted:bukan hoax		
actual:hoax	187	156		
actual:bukan hoax	93	2833		
	precision	recall	f1-score	support
0	0.66786	0.54519	0.60032	343
1	0.94781	0.96822	0.95790	2926
accuracy			0.92383	3269

Gambar 15. Hasil Pengujian SVM dengan *Kernel Linear* Skenario 1

TABEL IV
HASIL PREDIKSI SVM DENGAN KERNEL LINEAR SKENARIO 1

	Nilai
True Positive	187
True Negative	2833
False Positif	156
False Negative	93

• Pengujian dengan *Kernel RBF*

Kernel rbf

	predicted:hoax	predicted:bukan hoax		
actual:hoax	73	270		
actual:bukan hoax	3	2923		
	precision	recall	f1-score	support
0	0.96053	0.21283	0.34845	343
1	0.91544	0.99897	0.95538	2926
accuracy			0.91649	3269

Gambar 16. Hasil Pengujian SVM dengan *Kernel RBF* Skenario 1

TABEL V
HASIL PREDIKSI SVM DENGAN KERNEL RBF SKENARIO 1

	Nilai
True Positive	73
True Negative	2923
False Positif	270
False Negative	3

a) *Skenario 2 (70:30)*: hasil dari pengujian SVM pada pemodelan klasifikasi berita *hoax* dan berita bukan *hoax* dengan skenario 2 adalah sebagai berikut:

• Pengujian dengan *Kernel Linear*

Kernel linear

	predicted:hoax	predicted:bukan hoax		
actual:hoax	149	107		
actual:bukan hoax	77	2119		
	precision	recall	f1-score	support
0	0.65929	0.58203	0.61826	256
1	0.95193	0.96494	0.95839	2196
accuracy			0.92496	2452

Gambar 17. Hasil Pengujian SVM dengan *Kernel Linear* Skenario 2

TABEL VI
HASIL PREDIKSI SVM DENGAN KERNEL LINEAR SKENARIO 2

	Nilai
True Positive	149
True Negative	2119
False Positive	107
False Negative	77

• Pengujian dengan *Kernel RBF*

Kernel rbf

	predicted:hoax	predicted:bukan hoax		
actual:hoax	62	194		
actual:bukan hoax	3	2193		
	precision	recall	f1-score	support
0	0.95385	0.24219	0.38629	256
1	0.91873	0.99863	0.95702	2196
accuracy			0.91966	2452

Gambar 18. Hasil Pengujian SVM dengan *Kernel RBF* Skenario 2

TABEL VII
HASIL PREDIKSI SVM DENGAN KERNEL RBF SKENARIO 2

	Nilai
True Positive	62
True Negative	2193
False Positive	194
False Negative	3

b) *Skenario 3 (80:20)*: hasil dari pengujian SVM pada pemodelan klasifikasi berita *hoax* dan berita bukan *hoax* dengan skenario 3 adalah sebagai berikut:

• Pengujian dengan *Kernel Linear*

Kernel linear

	predicted:hoax	predicted:bukan hoax		
actual:hoax	111	61		
actual:bukan hoax	55	1408		
	precision	recall	f1-score	support
0	0.66867	0.64535	0.65680	172
1	0.95848	0.96241	0.96044	1463
accuracy			0.92905	1635

Gambar 19. Hasil Pengujian SVM dengan *Kernel Linear* Skenario 3

TABEL VIII
HASIL PREDIKSI SVM DENGAN KERNEL LINEAR SKENARIO 3

	Nilai
True Positive	111
True Negative	1408
False Positive	61
False Negative	55

• Pengujian dengan *Kernel RBF*

Kernel rbf

	predicted:hoax	predicted:bukan hoax		
actual:hoax	50	122		
actual:bukan hoax	3	1460		
	precision	recall	f1-score	support
0	0.94340	0.29070	0.44444	172
1	0.92288	0.99795	0.95895	1463
accuracy			0.92355	1635

Gambar 20. Hasil Pengujian SVM dengan *Kernel RBF* Skenario 3

TABEL IX
HASIL PREDIKSI SVM DENGAN KERNEL RBF SKENARIO 3

	Nilai
True Positive	50
True Negative	1460
False Positive	122
False Negative	3

c) *Skenario 4 (90:10)*: hasil dari pengujian SVM pada pemodelan klasifikasi berita *hoax* dan berita bukan *hoax* dengan skenario 1 adalah sebagai berikut:

• Pengujian dengan *Kernel Linear*

Kernel linear

	predicted:hoax	predicted:bukan hoax		
actual:hoax	70	33		
actual:bukan hoax	29	686		
	precision	recall	f1-score	support
0	0.70707	0.67961	0.69307	103
1	0.95410	0.95944	0.95676	715
accuracy			0.92421	818

Gambar 21. Hasil Pengujian SVM dengan *Kernel Linear* Skenario 4

TABEL X
HASIL PREDIKSI SVM DENGAN KERNEL LINEAR SKENARIO 4

	Nilai
True Positive	70
True Negative	686
False Positive	33
False Negative	29

• Pengujian dengan *Kernel RBF*

Kernel rbf

	predicted:hoax	predicted:bukan hoax		
actual:hoax	28	75		
actual:bukan hoax	3	712		
	precision	recall	f1-score	support
0	0.90323	0.27184	0.41791	103
1	0.90470	0.99580	0.94807	715
accuracy			0.90465	818

Gambar 22. Hasil Pengujian SVM dengan *Kernel RBF* skenario 4

TABEL XI
HASIL PREDIKSI SVM DENGAN KERNEL RBF SKENARIO 4

	Nilai
True Positive	28
True Negative	712
False Positive	75
False Negative	3

5. *Evaluation*: Hasil yang diperoleh dari seluruh pengujian dengan model SVM akan dibandingkan untuk mengetahui model terbaik dalam proses klasifikasi. Hasil perbandingan evaluasi menggunakan 4 skenario dapat dilihat pada tabel berikut ini:

TABEL XII
PERBANDINGAN EVALUASI MODEL

Skenario	Akurasi (%)		Precision (%)				
	Kernel Linear	Kernel RBF	Kernel Linear		Kernel RBF		
			0	1	0	1	
60:40	92.38	91.65	66.78	94.78	96.05	91.54	
70:30	92.49	91.96	65.92	95.19	95.38	91.87	
80:20	92.90	92.35	66.86	95.84	94.34	92.28	
90:10	92.42	90.46	70.70	95.41	90.32	90.47	
Recall (%)		F-score (%)					
Kernel Linear	Kernel RBF	Kernel Linear	Kernel RBF	Kernel Linear	Kernel RBF	Kernel Linear	Kernel RBF
0	1	0	1	0	1	0	1
54.51	96.82	21.28	99.89	60.03	95.79	34.84	95.53
58.20	96.49	24.21	99.86	61.82	95.83	38.62	95.70
64.53	96.24	29.07	99.79	65.68	96.04	44.44	95.89
67.96	95.94	27.18	99.58	69.30	95.67	41.79	94.48

Berdasarkan tabel di atas, dapat dilihat hasil akurasi menggunakan *Support Vector Machine (SVM)* dengan 2 fungsi *kernel* yaitu *kernel linear* mendapatkan nilai akurasi terbesar pada skenario 1 (60:40) sebesar 97,06%. Sedangkan hasil akurasi terendah juga didapat *kernel linear* pada skenario 90:10 sebesar 95,87%. Namun hasil dari nilai

akurasi saja tidak cukup untuk menentukan model terbaik pada penelitian ini. Dapat dilihat juga hasil prediksi dari masing-masing kernel SVM dengan 4 skenario pada **tabel 12**.

Skenario 1 (60:40) dengan *kernel linear* mendapatkan hasil yang baik karena model dapat meng-klasifikasikan data berita *hoax* dan berita bukan *hoax* dengan baik, ditunjukkan pada tabel di bawah.

TABEL XIII
HASIL PREDIKSI SVM DENGAN 2 FUNGSI KERNEL 4 SKENARIO

Skenario	Hasil		
Skenario 60:40	Kernel Linear		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	187	156
	Bukan Hoax	93	2833
	Kernel RBF		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	73	270
Bukan Hoax	3	2923	
Skenario 70:30	Kernel Linear		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	149	107
	Bukan Hoax	77	2119
	Kernel RBF		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	62	194
Bukan Hoax	3	2193	
Skenario 80:20	Kernel Linear		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	111	61
	Bukan Hoax	55	1408
	Kernel RBF		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	50	61
Bukan Hoax	3	1408	
Skenario 90:10	Kernel Linear		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	70	33
	Bukan Hoax	29	686
	Kernel RBF		
		Prediksi Hoax	Prediksi Bukan Hoax
	Hoax	28	75
Bukan Hoax	3	712	

B. Pembahasan

Penelitian ini mengimplementasikan algoritma Support Vector Machine (SVM) dalam mengklasifikasikan *hoax* dan bukan *hoax* berita Covid-19 yang diambil dari situs berita online. Metodologi yang digunakan adalah proses text mining yang terdiri dari beberapa tahapan yaitu data selection, preprocessing, transformation, data mining dan evaluation.

Pada data *selection* dilakukan pengambilan data dengan proses *web scrapping* dan pemberian *labeling*, selanjutnya ke

proses *preprocessing* dengan beberapa tahapan diantaranya *case folding* (mengubah keseluruhan kata dalam judul menjadi *lowercase* atau huruf kecil), *tokenizing*, *filtering* (penghapusan kata sambung) dan *stemming* (penghapusan kata imbuhan). Pada tahap *transformation* dilakukan pembobotan kata dengan menggunakan salah satu *feature selection* yaitu TF-IDF.

Pada pemodelan klasifikasi SVM dengan *kernel linear* dan RBF didapatkan hasil pengujian terbaik oleh *kernel linear* pada skenario 3 (80:20) karena model dapat mengklasifikasikan dengan baik data berita *hoax* dan berita bukan *hoax* Covid-19, yaitu dengan hasil data *hoax* yang diprediksi *hoax* ada 111, data *hoax* yang diprediksi bukan *hoax* ada 61. Data bukan *hoax* diprediksi *hoax* ada 55 dan data bukan *hoax* diprediksi bukan *hoax* ada 1408. Pada skenario (80:20) juga dapat dilihat bahwa *kernel linear* memiliki nilai akurasi tertinggi yaitu sebesar 92,90%.

Sedangkan untuk nilai terendah didapat oleh *kernel RBF* pada skenario (90:10) yaitu 90,46% dan model yang didapat juga kurang baik dalam mengklasifikasikan berita *hoax* maupun bukan *hoax* yaitu dengan hasil data *hoax* yang diprediksi sebagai *hoax* ada 28 dan data *hoax* yang diprediksi bukan *hoax* ada 75. Data bukan *hoax* yang diprediksi *hoax* ada 3 dan data bukan *hoax* yang diprediksi bukan *hoax* ada 712.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, antara lain penerapan algoritma Support Vector Machine dengan kernel linear dan RBF untuk klasifikasi berita *hoax* Covid-19 dilakukan dengan menggunakan beberapa proses seperti *case folding*, *tokenizing*, *filtering*, *stemming*, pembobotan kata (TF-IDF) dan melakukan pemodelan dengan 4 skenario pengujian secara acak untuk mendapatkan model terbaik dari kedua kernel algoritma *Support Vector Machine*.

Algoritma *Support Vector Machine* (SVM) dengan *kernel linear* memiliki hasil prediksi yang bagus pada skenario 3 (80:20) karena model sanggup dalam mengklasifikasikan berita *hoax* dan bukan *hoax* covid-19. Hasil yang didapat yaitu ada 111 data *hoax* yang diprediksi *hoax*, ada 61 data *hoax* yang diprediksi bukan *hoax*. Sedangkan data bukan *hoax* yang diprediksi *hoax* ada 55 dan data bukan *hoax* diprediksi bukan *hoax* ada 1408. Selain itu akurasi yang didapat pada skenario 80:20 juga memiliki nilai akurasi tertinggi yaitu sebesar 92,90%.

Sedangkan untuk nilai akurasi terendah didapat oleh *kernel RBF* pada skenario (90:10) yaitu 90,46% dan model yang didapat juga kurang baik dalam mengklasifikasikan berita *hoax* maupun bukan *hoax* yaitu dengan hasil data *hoax* yang diprediksi *hoax* ada 28 & data *hoax* yang diprediksi bukan *hoax* ada 75. Data bukan *hoax* yang diprediksi *hoax* ada 3 & data bukan *hoax* yang diprediksi bukan *hoax* ada 712.

B. Saran

Penelitian ini masih belum sempurna dan membutuhkan pengembangan lebih lanjut. Oleh karena itu, peneliti menyarankan hal yang dapat menjadi saran untuk mengembangkan penelitian selanjutnya yaitu diharapkan dapat menangani imbalance data untuk proses klasifikasi berita *hoax* dan bukan *hoax* covid-19 serta mampu menerapkan hasil pemodelan klasifikasi pada sistem

DAFTAR PUSTAKA

- [1] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, "Arti Kata Hoax," Kamus Besar Bahasa Indonesia, 2021. <https://kbbi.kemdikbud.go.id/entri/hoaks>.
- [2] Oxford University, "Oxford Learners Dictionaries," Oxford University Press, 2021. https://www.oxfordlearnersdictionaries.com/definition/english/hoax_1?q=hoax.
- [3] MASTEL, "Hasil Survey Wabah Hoax Nasional 2019," MASTEL, 2019. <https://mastel.id/hasil-survey-wabah-hoax-nasional-2019/> (accessed Jan. 24, 2021).
- [4] Simon Kemp, "DIGITAL 2019: INDONESIA," Data Reportal, 2019. <https://datareportal.com/reports/digital-2019-indonesia>.
- [5] Simon Kemp, "Digital 2020: Indonesia," Data Reportal, 2020. <https://datareportal.com/reports/digital-2019-indonesia>.
- [6] A. F. I. Tachta Citra Elfira, "APJII: Pandemi COVID-19 Buat Pengguna Internet di Indonesia Meningkat Hampir 200 Juta," *voi.id*, 2020. <https://voi.id/teknologi/19331/apjii-pandemi-covid-19-buat-pengguna-internet-di-indonesia-meningkat-hampir-200-juta>.
- [7] Kementerian Pendidikan dan Kebudayaan Republik Indonesia, "Arti kata Berita," Kamus Besar Bahasa Indonesia, 2021. <https://kbbi.kemdikbud.go.id/entri/berita>.
- [8] A. M. Hasan, "Info Hoax Soal Kesehatan Paling Banyak Beredar di Masyarakat," *tirto.id*, 2017. <https://tirto.id/info-hoax-soal-kesehatan-paling-banyak-beredar-di-masyarakat-cnQZ>.
- [9] Tim Detikcom, "Kapan Sebenarnya Corona Pertama Kali Masuk RI?," *Detikcom*, 2020. <https://news.detik.com/berita/d-4991485/kapan-sebenarnya-corona-pertama-kali-masuk-ri?single=1>.
- [10] C. Juditha, "People Behavior Related To The Spread Of Covid-19's Hoax," *J. Pekommas*, vol. 5, no. 2, p. 105, 2020, doi: 10.30818/jpkm.2020.2050201.
- [11] F. N. Ulya and B. P. Jatmiko, "Kompas.com Kembali Jadi Pemenang Kategori Media Online Tepercaya," *Kompas.com*, 2019. <https://money.kompas.com/read/2019/08/01/124215826/kompascom-kembali-jadi-pemenang-kategori-media-online-tepercaya?page=all>.
- [12] H. Margianto, "Kompas.com Tersertifikasi dalam Jaringan Internasional Penguji Informasi," *Kompas.com*, 2018. <https://nasional.kompas.com/read/2018/10/17/08133311/kompascom-tersertifikasi-dalam-jaringan-internasional-penguji-informasi>.
- [13] R. P. Aji, "Pelatihan Identifikasi dan Pelaporan Berita Hoax melalui portal 'turnbackhoax.id' kepada Masyarakat Desa Kedungwringin," vol. 1, no. 2, pp. 120–127.
- [14] F. Prayoga, "Gugus Tugas Covid-19 Luncurkan Situs Resmi, Berisi Edukasi hingga Hoax Buster," *Okezone*, 2020. <https://nasional.okezone.com/read/2020/03/19/337/2186044/gugus-tugas-covid-19-luncurkan-situs-resmi-berisi-edukasi-hingga-hoax-buster>.
- [15] H. Hermanto, A. Mustopa, and A. Y. Kuntoro, "Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 211–220, 2020, doi: 10.33480/jitk.v5i2.1181.
- [16] P. Assiroj, Meyliana, A. N. Hidayanto, H. Prabowo, and H. L. H. S. Warnars, "Hoax News Detection on Social Media: A Survey," *1st 2018 Indones. Assoc. Pattern Recognit. Int. Conf. Ina. 2018 - Proc.*, pp. 186–191, 2019, doi: 10.1109/INAPR.2018.8627053.
- [17] H. Mustofa and A. A. Mahfudh, "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes," *Walisono J. Inf. Technol.*, vol. 1, no. 1, p. 1, 2019, doi: 10.21580/wjit.2019.1.1.3915.
- [18] D. Maulina and R. Sagara, "Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency-Inverse Document Frequency," *J. Mantik Penusa*, vol. 2, no. 1, pp. 35–40, 2018.
- [19] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [20] A. T. Akhir, "1, 2, 3."
- [21] Y. Wulandari, Wiranto, and Wiharto, "Klasifikasi Keluhan Pelanggan Berbasis Twitter Menggunakan Metode Support Vector Machine (Svm) (Studi Kasus Pos Indonesia)," *Pus. Dokumentasi dan Inf. Ilm. UNS*, p. 1, 2017, [Online]. Available: <https://eprints.uns.ac.id/43593/>.