

SISTEM PREDIKSI BERITA PALSU TENTANG VIRUS COVID-19 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Muhammad Fadhil Muttaqin¹, Tatang Bukhori², Yanto³, Nova Agustina⁴, Muchammad Naseer⁵
Program Studi Teknik Informatika^{1,2,3,4,5}
Sekolah Tinggi Teknologi Bandung^{1,2,3,4,5}
mfadhilmuttaqin22@gmail.com¹, tatangb909@gmail.com², yan95452@gmail.com³, nova@sttbandung.ac.id⁴

Abstrak

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 berita palsu. Berita palsu yang paling mudah beredar adalah berita palsu tentang informasi isu-isu terkini contohnya tentang Covid-19 dan bencana alam. Agar terhindar dari berita *hoax* yang beredar, salah satu caranya yaitu dengan lebih teliti membaca judul berita pada situs yang terpercaya seperti Liputan6, Kompas.com, dan CNN. Dalam penelitian ini penulis mengambil topik berita *hoax* tentang vaksinasi, penyebaran virus *Corona* dan isu-isu yang memang tidak benar tentang pandemi Covid-19, tidak hanya itu untuk memprediksinya penulis juga membutuhkan berita fakta untuk bisa menjadi perbandingan antara berita palsu dengan fakta dari kasus Covid-19. Karena itu penelitian ini akan membuat model klasifikasi berita palsu virus Covid-19 dengan menerapkan algoritma *Support Vector Machine (SVM)*. Hasil penelitian menyatakan bahwa model klasifikasi berita palsu Covid-19 dengan Algoritma *Support Vector Machine (SVM)* memiliki nilai akurasi 78%. Setelah uji coba model *SVM* terhadap *dataset* Covid-19, deploy model dilakukan dan diharapkan untuk masa yang akan datang masyarakat dapat mengakses situs pengecekan berita palsu. Selain mendapatkan prediksi judul berita yang dikategorikan dengan "Fake" dan "True", web juga dapat menampilkan persentase probabilitas dari prediksi yang dilakukan oleh model. Dengan adanya sistem ini, diharapkan dapat memprediksi judul berita tentang virus Covid-19 seakurat mungkin, baik berita palsu maupun berita yang berisikan fakta.

Kata kunci : Covid-19, *Support Vector Machine (SVM)*, Berita Palsu.

Abstract

Along with the development of information technology, the news media has also developed in presenting information in online media. However, it is very unfortunate that in its distribution there are still many hoax news or fake news. The most easily circulated fake news is fake news about information on current issues, for example about Covid-19 and natural disasters. In order to avoid hoax news circulating, one way is to read more carefully the news titles on trusted sites such as Liputan6, Kompas.com, and CNN. In this study the authors took the topic of hoax news about vaccinations, the spread of the corona virus and issues that were not true about the Covid-19 pandemic, not only that to predict it the authors also needed factual news to be able to be a comparison between fake news and facts from the Covid-19 case. Therefore this research will create a classification model for fake news of the Covid-19 virus by applying the *Support Vector Machine (SVM)* algorithm. The results of the study stated that the Covid-19 fake news classification model with the *Support Vector Machine (SVM)* Algorithm had an accuracy value of 78%. After testing the *SVM* model on the Covid-19 dataset, the model was deployed and it is hoped that in the future the public will be able to access fake news checking sites. In addition to getting news headline predictions categorized by "Fake" and "True", the web can also display the probability percentage of the predictions made by the model. With this system, it is hoped that it can predict news headlines about the Covid-19 virus as accurately as possible, both fake news and news containing facts.

Keywords : Covid-19, *Support Vector Machine (SVM)*, Fake News.

I. PENDAHULUAN

Seiring dengan perkembangan teknologi informasi, media informasi seperti berita, artikel dan sosial media juga turut berkembang pesat dalam menyajikan informasi dalam media *online* atau *internet*. Dalam penyebarannya sangat disayangkan karena selalu ada pihak yang kurang bertanggung jawab dalam menyebarkan informasi. Pihak yang kurang bertanggung jawab tersebut seringkali menyebarkan berita palsu atau sering disebut *hoax* demi keuntungan sepihak atau keisengan semata. Dalam menghentikan penyebaran berita palsu atau *hoax* diperlukan kesadaran dari seluruh masyarakat, contohnya bijak dalam menggunakan media sosial, mencermati berita atau isu yang diterima sebelum menyebarkan kembali, dan memahami resiko kerugian orang lain jika berita palsu tersebut tersebar. Dikutip dari [1], Kementerian Komunikasi dan Informatika atau Kominfo menemukan 9.546 *hoax* telah tersebar di berbagai *platform* media sosial di internet. Data itu terangkum dalam kurun tiga tahun mulai Agustus 2018 hingga awal 2022.

Berita palsu atau *hoax* yang paling cepat penyebarannya adalah berita palsu tentang informasi isu-isu terkini, contohnya tentang virus Covid-19 dan bencana alam. Berita tentang Covid-19, pada Agustus tahun 2020 saja sudah tersebar banyak *hoax* yang berasal dari berbagai sumber. Dikutip dari [2], Kementerian Komunikasi dan Informatika (Kominfo) mencatat hingga 8 Agustus 2020 sebanyak 1.028 *hoax* tersebar di berbagai *platform* media sosial terkait disinformasi tentang virus Corona (COVID-19).



Gambar 1. Data Hoaks Covid-19 2020
(Sumber : [2])

Mengurangi penyebaran berita palsu tentang Covid-19, penulis membuat sistem prediksi berita palsu tentang virus Covid-19 ini. Sistem dibuat dengan pengklasifikasian pada judul berita tentang isu-isu virus Covid-19 yang ada. Model yang dibuat akan menggunakan algoritma klasifikasi *Support Vector Machine (SVM)*. Di sistem prediksi berita palsu ini, pengguna atau *user* dapat memasukkan judul berita atau isu yang beredar, lalu pengguna atau *user* akan mendapatkan hasil prediksi dengan label “FAKE” atau “TRUE” lengkap dengan probabilitas *fake* dan *true* dari judul berita atau isu yang dimasukkan.

II. LANDASAN TEORI

1. Hoax

Hoax adalah suatu kata yang digunakan untuk menunjukkan pemberitaan palsu atau usaha untuk menipu atau mengakali pembaca untuk mempercayai sesuatu. Pemberitaan yang tidak berdasarkan kenyataan atau kebenaran (*nonfactual*) untuk maksud tertentu [3]. *Hoax* atau berita bohong adalah salah satu bentuk *Cyber Crime* yang kelihatannya sederhana, mudah dilakukan namun berdampak sangat besar bagi kehidupan sosial masyarakat. Tujuan dari pembuat dan penyebar. *Hoax* adalah menggiring opini masyarakat dan kemudian membentuk persepsi yang salah terhadap suatu informasi yang sebenarnya [4].

Hoax adalah informasi sesat dan berbahaya karena menyesatkan persepsi manusia dengan menyampaikan informasi palsu sebagai kebenaran. *Hoax* sendiri dapat bertujuan untuk mempengaruhi pembaca dengan informasi palsu sehingga pembaca mengambil tindakan sesuai dengan isi *hoax* [5]. Indonesia merupakan negara demokrasi terbesar ketiga di dunia setelah India dan Amerika yang mengalami permasalahan serius soal penyebaran berita palsu (*Fake News/Hoax*). Pemerintah melalui Kepolisian Republik Indonesia telah mengeluarkan ancaman untuk memproses hukum pihak-pihak yang menyebarkan berita palsu [6].

2. Media Sosial

Media sosial adalah kumpulan saluran komunikasi *online* yang didedikasikan untuk input, interaksi, berbagi *content* serta kolaborasi berbasis komunitas tertentu. Aplikasi dan situs web yang didedikasikan untuk beberapa forum sosial, jejaring sosial, *microblogging* adalah beberapa diantaranya [3]. Media sosial sebagai bagian dari inovasi teknologi informasi, memberikan ruang bagi seseorang untuk menyuarakan pikirannya yang sebelumnya mungkin tidak pernah bisa terdengar. Media sosial menjadi ruang ekspresi baru bagi masyarakat dunia dalam beberapa tahun terakhir ini. Kehadiran media sosial sebagai wadah berbincang dan bertukar informasi antara satu dengan yang lain tentu memberikan dampak yang sangat positif. Disisi lain, media sosial juga menjadi arena bagi penyampaian opini, ujaran penuh kebencian (*hate speech*) dan berita-berita palsu (*hoax*) [7].

Berbagai Media Sosial *Online* merupakan sarana atau media bagi seseorang ataupun berbagai pihak dalam menyampaikan aspirasi pikirannya, pendapatnya ataupun sebagai tempat untuk menyampaikan berbagai informasi. Sebenarnya jika media *online* tersebut digunakan untuk hal-hal yang positif maka tidak ada masalah yang perlu dikuatirkan. Sayangnya media sosial *online* sering kali digunakan untuk menyampaikan berbagai hal negatif oleh seseorang ataupun pihak-pihak tertentu untuk berbagai kepentingan, baik kepentingan pribadi ataupun kepentingan pihak lain. Harus diakui bahwa media sosial merupakan tempat yang subur bagi munculnya informasi yang bersifat fitnah, hasutan, *hoax* dan sebagainya [4].

3. Support Vector Machine (SVM)

SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* [8]. *Support Vector Machines (SVM)* adalah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi. SVM memiliki kelebihan yaitu mampu mengidentifikasi *hyperplane* terpisah yang memaksimalkan *margin* antara dua kelas yang berbeda. Namun SVM memiliki kekurangan terhadap masalah

pemilihan parameter atau fitur yang sesuai. Pemilihan fitur sekaligus penyetingan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi [9].

Support Vector Machine (SVM) mengharuskan sebuah teks harus diubah menjadi *vector* sebelum digunakan untuk klasifikasi. *Key idea* dari SVM adalah untuk menemukan permukaan keputusan (*Hyperlane*) yang maksimal dari setiap titik data. Untuk melakukan *training* mesin yang didukung oleh *vector* atau biasa disebut *Support Vector Machine* (SVM) memerlukan solusi *Quadratic Programming* (QP) yang sangat besar. *Quadratic Programming* adalah masalah matematika untuk menemukan *vector* “x” yang meminimalkan fungsi kuadrat, dengan melakukan pembagian kelas menggunakan *hyperplane* maka masing-masing kelas positif, netral dan negatif dapat dibagi berdasarkan area masing-masing sehingga ketika terdapat data baru dapat ditentukan kelasnya berdasarkan area positif, netral maupun negatif [10].

4. Klasifikasi

Pengolahan dan komputasi data pada teknologi memiliki banyak metode. Salah satu metode dalam pengolahan data adalah klasifikasi. Klasifikasi merupakan cara pengelompokan benda berdasarkan ciri – ciri yang dimiliki oleh objek klasifikasi. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya *Naïve Bayes*, *Support Vector Machine*, *Decision Tree*, *Fuzzy* dan *Jaringan Saraf Tiruan* [11].

Klasifikasi merupakan salah satu tugas utama dalam *machine learning* dan *data mining* yang termasuk ke dalam tipe *supervised learning*. Istilah klasifikasi didapat dari tujuan utama teknik ini yaitu untuk memprediksi sebuah kategori dari masukan data [12]. Sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan. Klasifikasi merupakan pengelompokan fitur ke dalam kelas yang sesuai. Vektor fitur pelatihan tersedia dan telah diketahui kelas- kelasnya, kemudian vektor fitur pelatihan tersebut dimanfaatkan untuk merancang pemilah. Pengenalan pola ini disebut terbimbing, *supervised* [11].

5. Machine Learning

Machine learning adalah bidang yang mempelajari pengembangan algoritma komputer untuk mengubah data menjadi aksi yang cerdas atau secara singkat dapat juga diartikan sebagai proses mengubah data menjadi informasi (*Machine Learning in Action*) [12]. *Machine Learning* adalah teknik yang meningkatkan kinerja sistem dengan belajar dari pengalaman melalui metode komputasi. Dalam sistem komputer, pengalaman ada dalam bentuk data, dan tugas utama *machine learning* adalah mengembangkan algoritma pembelajaran yang membangun model dari data. Mengumpamakan algoritma pembelajaran dengan data pengalaman, dapat memperoleh suatu model yang memprediksi observasi baru [13].

Machine Learning terbagi menjadi tiga kategori: *Supervised Learning*, *Unsupervised Learning*, *Reinforcement Learning*. *Machine learning* dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [14]. *Machine learning* memiliki fokus pada pengembangan sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia. Hal ini menjadikan mesin tidak hanya mampu berperilaku mengambil keputusan, namun juga dapat beradaptasi dengan perubahan yang terjadi. *Machine learning* bekerja apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data besar (*big data*) sehingga menemukan pola tertentu [15].

III. ANALISIS DAN PERANCANGAN

Perancangannya, sistem prediksi berita palsu ini menggunakan dataset yang dibuat oleh penulis sendiri. *Dataset* dibuat dengan terdiri dari kolom id, status fakta, judul berita dan *link* sumber berita. Kolom status fakta berfungsi untuk pelabelan judul berita, label berita terdiri dari 2 variabel yaitu *hoax* dan fakta. Kolom judul berita berisi tentang judul berita atau isu-isu tentang virus covid-19, sedangkan kolom *link* sumber berita berisi *link* sumber judul berita Covid-19 itu berasal. Total judul berita yang dimasukkan ke *dataset* berjumlah 505 judul. Berikut contoh data yang ada pada *dataset* yang telah dibuat.

TABEL I
DATASET

ID	Status Fakta	Judul Berita	Link Sumber
1	<i>Hoax</i>	Virus Corona Tersebar Akibat Kebocoran Laboratorium Di Wuhan	Kompas.com
2	<i>Hoax</i>	Berendam Air Panas Uap Panas Dari Peningg Tangan Dapat Membunuh Virus Covid-19	Kompas.com

3	Hoax	Mengkonsumsi Bawang Putih Dapat Mencegah Penularan Covid-19	Kompas.com
...
505	Fakta	Pemerintah Waspada Varian Baru Corona AY.4.2 dari Inggris	https://menpan.go.id/site/berita-terkini/pemerintah-waspada-varian-baru-corona-ay-4-2-dari-inggris

Tahap perancangan, sistem prediksi berita palsu tentang virus covid-19 ini terbagi kedalam beberapa tahapan yaitu *Data Preprocessing, Split Dataset, Modelling, Testing, dan Deployment Model*.

1. *Data Preprocessing*

Pada tahap *data preprocessing*, langkah pertama yang dilakukan adalah dengan memberikan label untuk judul berita Covid-19 yang ada pada *dataset*. Pelabelan atau *indexing data* pada judul berita ini menggunakan *library pandas* dengan memanfaatkan fungsi *loc*. *Loc* untuk seleksi dengan menggunakan label/bilangan bulat [16]. Data judul berita yang memiliki variabel status fakta “Hoax” diseleksi untuk dimasukkan ke data berlabel “fake”, sedangkan data judul berita yang memiliki variabel status fakta “Fakta” diseleksi untuk dimasukkan ke data berlabel “true”.

```
#filter label = 0 adalah fake news
#filter label = 1 adalah true news
fake = data.loc[data['Status Fakta'] == "Hoax"]
true = data.loc[data['Status Fakta'] == "Fakta"]

[ ] display("fake",fake)
display("true",true)
83 84.0 Hoax China Minta Persetujuan Bunuh 20 Ribu Pasien V... https://www.kominfo.go.id/content/detail/24321...
87 88.0 Hoax Pasien di Makasar Positif Virus Corona https://www.kominfo.go.id/content/detail/24814...
89 90.0 Hoax Situs Penanganan Covid-19 Belanda Diluncurkan ... https://www.kominfo.go.id/content/detail/37598...

ID Status Fakta Judul Berita Link Sumber
23 24.0 Fakta PPKM Dicabut, Satgas Covid-19 Tegaskan "Booste... PPKM Dicabut, Satgas Covid-19 Tegaskan "Booste...
24 25.0 Fakta Pemerintah Bakal Terapkan PPKM Lagi Jika Kasus... https://nasional.tempo.co/read/1674049/pemerin...
```

Gambar 2. Pelabelan

Langkah selanjutnya pada *data preprocessing* yaitu *data cleaning* dan normalisasi. Metode yang digunakan dalam *data cleaning* dan normalisasi adalah *tokenization*. *Tokenization* adalah proses memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal yang berdiri sendiri [17]. Tahap pertama dari *data cleaning* dan normalisasi adalah menghitung token setiap kata pada *dataset*. Pada tahap ini kata yang berduplikat dihitung sebagai 1 kata dengan nama variabelnya adalah *unnormalize_clean*.

```
[ ] # token for unnormlize data
unnormlize = []

for index, row in data.iterrows():
    text = row['Judul Berita']
    unnormlize += nltk.word_tokenize(text)

print("UNNORMALIZE TOKEN",len(unnormlize))
#remove Duplicates
unnormlize_clean = list(dict.fromkeys(unnormlize) )
print("UNNORMALIZE REMOVE DUPS",len(unnormlize_clean))

UNNORMALIZE TOKEN 5718
UNNORMALIZE REMOVE DUPS 1618
```

Gambar 3. Unnormalize Token

Pada gambar 3, terdapat 5718 token kata yang belum dinormalisasi. Tahap selanjutnya adalah normalisasi teks dan menghapus kata yang tidak diperlukan pada judul berita Covid-19, dimulai dari menghapus tanda baca, simbol, dll. Setelah teks dari judul berita Covid-19 tersebut dinormalisasi, tahap selanjutnya adalah menghitung token setiap kata pada *dataset* judul berita Covid-19 yang sudah dinormalisasi. Pada tahap ini kata yang berduplikat dihitung 1 kata dengan nama variabelnya adalah *normalize_clean*.

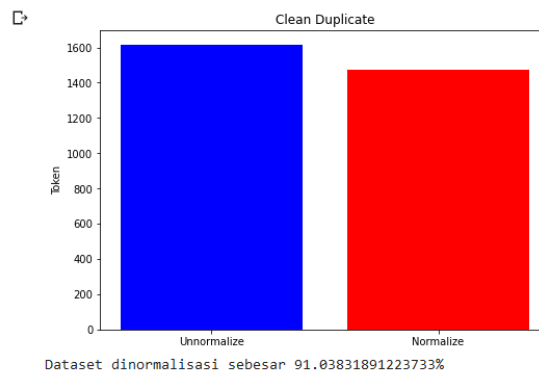
```
[ ] # token for normalize data
normalize = []
fake = []
true = []
for index, row in data.iterrows():
    test= row['Judul Berita']
    test = casefolding(test)
    test = filtering(test)
    test = replaceThreeOrMore(test)
    test = removeDoubleSpaces(test)
    test = convertToSlangword(test)
    test = removeStopword(test)
    test = tokenize(test)

    normalize += test
    if row['Status Fakta'] == "Fakta":
        true += test
    else:
        fake += test
print("NORMALIZE TOKEN",len(normalize))
#remove Duplicates
normalize_clean = list( dict.fromkeys(normalize) )
print("NORMALIZE REMOVE DUPS",len(normalize_clean))
```

NORMALIZE TOKEN 4406
 NORMALIZE REMOVE DUPS 1473

Gambar 4. *Normalize Token*

Pada gambar 4, terdapat 4406 *token* kata yang sudah dinormalisasi maka jika dilihat dari total *token* kata yang sebelum dinormalisasi dan yang sudah dinormalisasi. Normalisasi *dataset* ini memiliki peran penting dalam tahapan *data preprocessing*. Berikut tampilan persentase perbandingan *token unnormalize* dan *normalize*.



Gambar 5. *Persentase Normalisasi*

Tahapan setelah dilakukan *data cleaning* dan normalisasi data, masing-masing kata dari judul berita covid-19 tersebut perlu dilakukan pengkonversian ke *lowercase* agar setiap token kata berada dalam satu versi *string* yaitu huruf kecil. *Convert* string dari setiap kata judul berita covid-19 adalah tahap terakhir dalam *data preprocessing*, data yang didapatkan dari hasil tahap *preprocessing* berupa token kata dari judul berita covid-19 yang ada pada dataset. Berikut merupakan contoh dari hasil tahapan *data preprocessing* yang sudah dilakukan.

```
[ ] dataProcesed['token']

34 [dokter, sebut, vaksin, efektif, cegah, subvar...
324 [studi, total, kasus, covid19, china, sentuh, ...
288 [covid19, china, diprediksi, meledak]
299 [arti, coronavirus, bahasa, latin, serangan, j...
195 [omicron, sembuh, menghirup, bubuk, jahe]
...
104 [indonesia, patut, mewaspada, covid19, xbb15,...
25 [varian, omicron, menyebar, lebih, cepat, geja...
268 [nigeria, hancurkan, 1, juta, vaksin, covid19,...
21 [linhua, qingwen, sejenis, obat, herbal, memba...
456 [menyebarkan, lewat, sinyal, telepon, seluler]
Name: token, Length: 505, dtype: object
```

Gambar 6. *Hasil Tokenization*

2. *Split Dataset*

Tahapan setelah mendapatkan hasil dari *data preprocessing*, selanjutnya diperlukan tahapan *split dataset*. Pada tahap ini *dataset* dibagi menjadi dua bagian data, yaitu *data training* dan *data testing*. Sebanyak 70% *dataset* digunakan untuk *data training*, sedangkan 30% sisanya digunakan untuk *data testing*. Untuk *data training*, data diberi nama dengan *news_train* dan *status_train*. Sedangkan untuk *data testing* diberi nama *news_test* dan *status_test*. *Data training* akan berguna untuk pelatihan model SVM dari sistem prediksi berita palsu Covid-19, sedangkan *data testing* akan berfungsi pada tahap *testing* untuk menguji model SVM dari sistem prediksi berita palsu Covid-19.

3. Modelling

Tahap ini model *Support Vector Machine* dari sistem prediksi berita palsu Covid-19 dibuat. Model dibuat dengan menggabungkan data *news_train* dan data *status_train*.

```
from sklearn import svm #svc : svm.SVC()
from sklearn.pipeline import Pipeline

model = Pipeline([
    ('bow', CountVectorizer(analyzer=process_text)), # strings to token integer counts
    ('tfidf', TfidfTransformer()), # integer counts to weighted TF-IDF scores
    ('classifier', svm.SVC(probability=True)), # train on TF-IDF vectors w/ SVM Support Vector Classifier
])
tmodel = model.fit(news_train,status_train)
```

Gambar 7. Model SVM

Tahap model berhasil dibuat, maka selanjutnya dilakukan tahap *model evaluation*. Pada tahap *model evaluation*, model akan memprediksi judul berita yang ada pada data *news_test*. Berikut hasil dari model *evaluation* dengan hasil prediksi judul berita “true”.

```
#Model Evaluation
predictions = tmodel.predict(news_test)

for index, item in enumerate(news_test):
    if predictions[index] == "true":
        print(item, predictions[index])

kasus covid-19 di korea selatan dan china meningkat, pemerintah kembali bersiap true
seminggu naik 5 persen, eropa satu-satunya wilayah dengan kematian covid-19 masih bertambah true
potret puluhan ribu warga wuhan tahun baruan saat covid-19 melonjak true
```

Gambar 8. Model Evaluation

Tahap *model evaluation* dilakukan, penulis melakukan pengujian akurasi dari model SVM yang dibuat. Hasil dari pengujian menunjukkan model SVM dari sistem prediksi berita palsu Covid-19 memiliki akurasi 78%.

TABEL II
TES AKURASI

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Fake</i>	0.80	0.67	0.73	89
<i>True</i>	0.77	0.87	0.82	113
<i>accuracy</i>	0.78			202
<i>Macro avg</i>	0.79	0.77	0.77	202
<i>Weighted avg</i>	0.78	0.78	0.78	202
<i>Accuracy: 0.78</i>				

Evaluasi model yang dihasilkan digolongkan dalam kelompok berikut[18]:

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Hasil kesimpulan yang didapat, model klasifikasi berita palsu tentang Covid-19 menggunakan algoritma SVM ini memiliki akurasi klasifikasi yang cukup.

4. Testing

Tahap selanjutnya setelah model dibuat adalah pengujian model atau *testing*. Dalam testing ini, penulis menguji model SVM dengan memasukan judul berita atau isu tentang covid-19 yaitu “minum kayu putih dapat menyembuhkan pasien Covid-19”, dan prediksi model SVM menghasilkan label “fake” untuk judul berita tersebut.

```
test = news_test.iloc[0]
test = 'minum kayu putih dapat menyembuhkan pasien covid-19'
pre = tmodel.predict([test])
pre

array(['fake'], dtype=object)
```

Gambar 9. Testing

5. Deployment Model

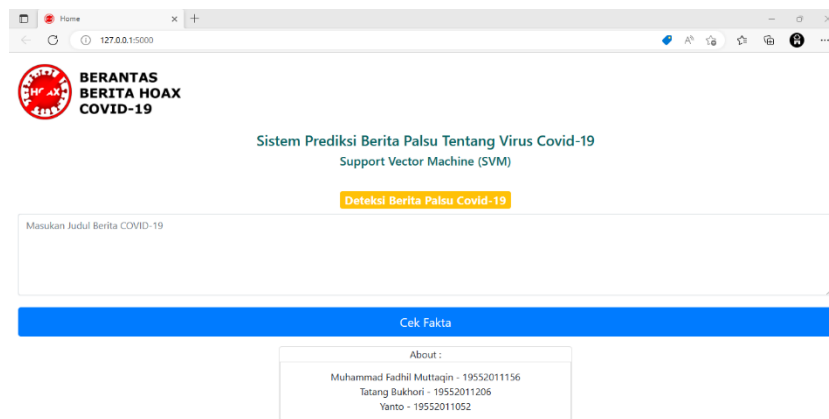
Tahap terakhir dalam perancangan sistem prediksi berita palsu Covid-19 adalah *deployment model SVM* ke basis web. *Deployment model* ini bertujuan untuk memudahkan dalam penggunaan model SVM yang telah dibuat oleh penulis. Langkah pertama, penulis melakukan *deployment model* ke format *.pkl (pickle)*.

```
import joblib
print(joblib.__version__)
import pickle
# Pipeline
joblib.dump(tmodel, 'fakenews.pkl')
```

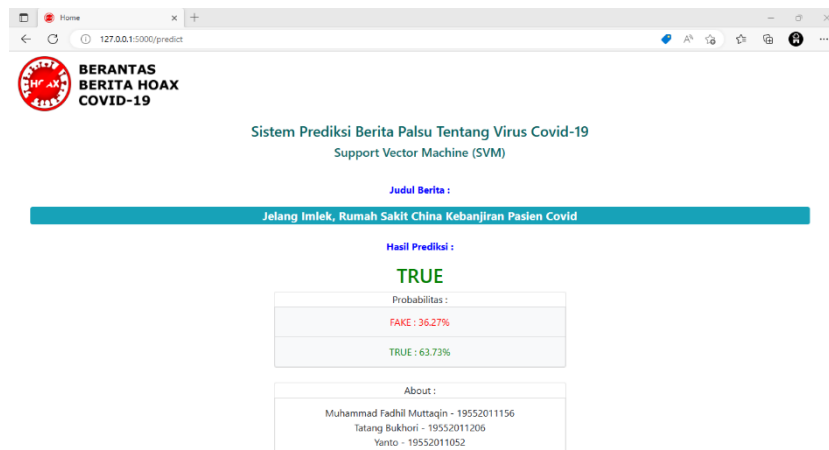
```
1.2.0
['fakenews.pkl']
```

Gambar 10. Deploy Model

Tahap *deploy model* ke format *.pkl* selesai, maka didapatkan model dengan format *.pkl* untuk keperluan *deployment model SVM* tersebut ke basis web. Penulis menggunakan *framework CSS Bootstrap* untuk mendesain tampilan *website* sistem prediksi berita palsu Covid-19, dan menggunakan *framework Python Flask* untuk penerapan model SVM format *.pkl* ke dalam *website*. Berikut merupakan hasil *deployment model SVM* ke basis web dengan contoh hasil prediksi sistemnya.



Gambar 11. Deployment (Home)



Gambar 12. Deployment (Predict)

Gambar 13 merupakan contoh dari hasil prediksi menggunakan model SVM. Hasil prediksi pada judul berita “Jelang Imlek, Rumah Sakit China Kebanjiran Pasien Covid” adalah “TRUE” dengan probabilitas persentase *true* 63,73%.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal, antara lain dalam pembuatan sistem prediksi berita palsu diperlukan model klasifikasi terlebih dahulu. Salah satu metode atau algoritma yang digunakan untuk klasifikasi adalah *Support Vector Machine (SVM)*. Sistem prediksi berita palsu tentang Covid-19 ini dibuat dengan beberapa tahapan atau proses yaitu pembuatan *dataset* judul berita, *data preprocessing*, *split dataset*, *modelling*, *testing*, dan *deployment model*.

Model klasifikasi dengan algoritma *Support Vector Machine (SVM)* ini memiliki nilai akurasi yang cukup. Hasil pengujian mendapat persentase 78% untuk akurasi model klasifikasi SVM ini. Model klasifikasi SVM ini di *deployment* ke *website* untuk pembuatan sistem prediksi berita palsu tentang Covid-19. Model di *deploy* terlebih dahulu ke format *.pkl (pickle)* untuk digunakan dalam pengimplementasiannya.

Menerapkan model klasifikasi SVM ke web, penulis menggunakan *framework Python Flask* sedangkan untuk mendesain tampilan web, penulis menggunakan *framework CSS Bootstrap*. Selain memberi label “FAKE” atau “TRUE” pada judul berita yang dimasukkan, sistem prediksi berita palsu ini dapat juga menampilkan persentase probabilitas dari *true* atau *fake* nya judul berita tersebut.

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 meningkatkan persentase akurasi dari model yang digunakan menggunakan algoritma lain seperti *Naïve Bayes* dan *Decision Tree*. Besar harapan jika menggunakan algoritma lain mampu meningkatkan persentase akurasi dari model yang digunakan.

REFERENSI

- [1] F. C. Rosana, ‘Hingga Awal 2022, Kominfo Temukan 9.546 Hoaks di Internet’, TEMPO.CO, 2022. <https://bisnis.tempo.co/read/1558213/hingga-awal-2022-kominfo-temukan-9-546-hoaks-di-internet> (accessed Jan. 17, 2023).
- [2] B. Mufarida, ‘Kominfo Mencatat Sebanyak 1.028 Hoaks Tersebar terkait COVID-19’, SINDONEWS.com, 2020. <https://nasional.sindonews.com/read/131216/15/kominfo-mencatat-sebanyak-1028-hoaks-tersebar-terkait-covid-19-1597219726> (accessed Jan. 17, 2023).
- [3] C. Juditha, ‘Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya Hoax Communication Interactivity in Social Media and Anticipation’, *J. Pekommas*, vol. 3, no. 1, pp. 31–44, 2018.
- [4] H. Septanto, ‘Pengaruh Hoax dan Ujaran Kebencian Sebuah Cyber Crime dengan Teknologi Sederhana di Kehidupan Sosial Masyarakat’, *J. Sains dan Teknol.*, vol. 5, no. 2, pp. 157–162, 2018.
- [5] A. Afriza and J. Adisantoso, ‘Metode Klasifikasi Rocchio untuk Analisis Hoax’, *J. Ilmu Komput. dan Agri-informatika*, vol. 5, no. 1, pp. 1–10, 2018.
- [6] R. Firmansyah, ‘Web Klarifikasi Berita Untuk Meminimalisir Penyebaran Berita Hoax’, *J. Inform.*, vol. 4, no. 2, pp. 230–235, 2017, doi: <https://doi.org/10.31294/ji.v4i2.2138>.
- [7] R. Pakpahan, ‘Analisis Fenomena Hoax Diberbagai Media Sosial Dan Cara Menanggulangi Hoax’, *Konf. Nas. Ilmu Sos. Teknol.*, vol. 1, no. 1, pp. 479–484, 2017.
- [8] A. S. Nugroho, A. B. Witarto, and D. Handoko, ‘Support Vector Machine’, *Mach. Learn. Methods Appl. to Brain Disord.*, 2003, doi: 10.1016/B978-0-12-815739-8.00006-7.
- [9] D. Ajeng Kristiyanti, ‘Analisis Sentimen Review Produk Kosmetik Menggunakan Algoritma Support Vector Machine Dan Particle Swarm Optimization Sebagai Metode Seleksi Fitur’, in *Seminar Nasional Inovasi dan Tren (SNIT) 2015*, 2015, vol. 1, no. 1, p. A 134-141.
- [10] N. Saputra, T. B. Adji, and A. E. Permasari, ‘Analisis Sentimen Data Presiden Jokowi Dengan Pre Processing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM’, *J. Din. Inform.*, vol. 5, no. 1, 2015.
- [11] A. P. Wibawa, M. Guntur Aji Purnama, M. Fathony Akbar, and F. Andika Dwiyanto, ‘Metode-metode Klasifikasi’, in *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 2018, vol. 3, no. 1, pp. 134–138.
- [12] M. R. Faisal and D. T. Nugrahadi, *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. Banjarbaru: Scripta Cendekia, 2019.
- [13] Z.-H. Zhou, *Machine Learning*. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-15-1967-3.
- [14] A. R. Dikananda, O. Nurdiawan, A. Faqih, A. Surip, and D. B. Saputra, ‘Estimasi Kemampuan Computational Thingking Pemain Game Dota Dua Menggunakan Teknik Machine Learning’, *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1599, Oct. 2022, doi: 10.30865/jurikom.v9i5.5025.
- [15] E. Retnoningsih and R. Pramudita, ‘Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python’, *BINA Insa. ICT J.*, vol. 7, no. 2, p. 156, Dec. 2020, doi: 10.51211/biict.v7i2.1422.
- [16] I. Mutmainnah, ‘Mengenal Pandas Dalam Python’, Medium, 2019. <https://medium.com/@16611092/mengenal-pandas-dalam-python-cc66d0c5ea40> (accessed Jan. 23, 2023).
- [17] L. Robinson, ‘Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval Untuk Pencarian Informasi Pada Kumpulan Dokumen Teknik Elektro di UPT BPI LIPI’, *J. Ilm. Komput. dan Inform. (KOMPUTA)*, 2014.
- [18] F. Gorunescu, ‘Data Mining’, in *Concepts, Models and Techniques*, 1st ed., Berlin: Springer Berlin, Heidelberg, 2011, p. 360.