

# Temu Kembali Informasi untuk Visualisasi Riwayat Berita Online Covid-19 Di Indonesia

Muhammad Naufal Refadi<sup>1</sup>, Diana Purwitasari<sup>2</sup>, Agus Budi Raharjo<sup>3</sup>, Dwi Sunaryono<sup>4</sup>

<sup>1234</sup>Departemen Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas,

Institut Teknologi Sepuluh Nopember (ITS)

E-mail: <sup>1</sup>naufal.refadi@yahoo.co.id, <sup>2</sup>diana@if.its.ac.id, <sup>3</sup>agus.budi@its.ac.id, <sup>4</sup>dwis@if.its.ac.id

Pada 2 Maret 2021, Indonesia mendapatkan kasus Coronavirus Disease of 2019 atau COVID-19 pertama dan sampai sekarang jumlah kasus virus COVID-19 terus meningkat dan memakan banyak korban jiwa. Banyak kejadian atau hal yang menarik yang dapat menggambarkan situasi pandemi di Indonesia. Dari kejadian itu, bisa digambarkan menjadi sebuah visualisasi yang dapat menggambarkan perjalanan COVID-19 di Indonesia dan ditampilkan kedalam sebuah sistem informasi berbentuk website agar pengguna terutama masyarakat awam dapat menganalisa manfaat usaha pemerintah pusat atau daerah terkait kenaikan dan penurunan kasus COVID-19 di Indonesia. Untuk mendapatkan data kejadian COVID-19, data dapat diambil berupa berita dari portal berita online dengan melakukan web scraping dan diolah agar mendapatkan informasi sesuai dengan tema COVID-19 di Indonesia. Selain itu, dengan melakukan pemrosesan teks dan klasifikasi teks menggunakan berbagai macam model untuk mencari dan memvisualisasikan berita dapat memvisualisasikan informasi yang tidak ada pada berita, seperti kategori berita. Dari hasil penelitian yang diperoleh, didapatkan akurasi model klasifikasi teks menggunakan model Random Forest adalah 94,5% dan sistem informasi berita covid-19 dapat berjalan baik dengan menampilkan berita covid-19 di setiap daerah provinsi/kota.

**Kata kunci:** Berita Covid-19, Sistem Informasi, Web Scraping, Pemrosesan Teks, Klasifikasi Teks

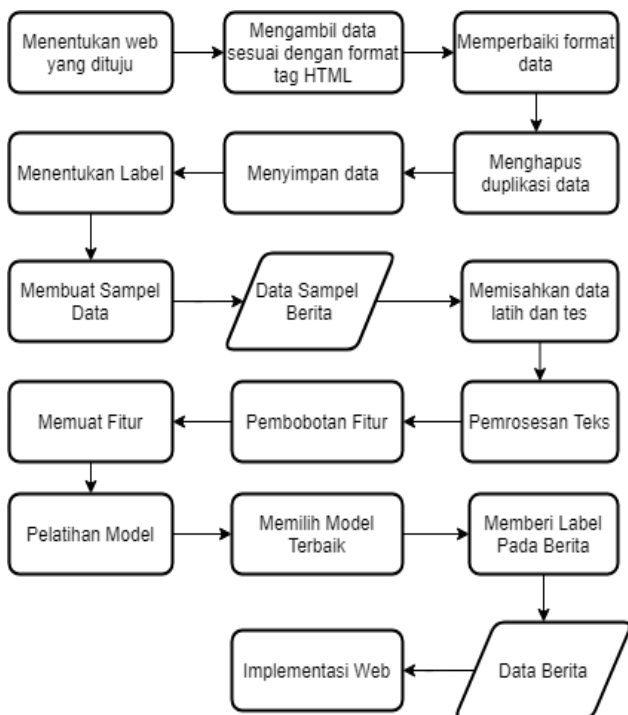
## I. PENDAHULUAN

Pada akhir 2019, dunia dilanda dengan wabah virus mematikan bernama Coronavirus Disease of 2019 atau bisa disebut COVID-19. Tercatat pada 27 Januari 2020, sudah 100 juta kasus infeksi COVID-19 di seluruh dunia dengan diantaranya 2,1 juta jiwa meninggal dunia karena virus tersebut termasuk Indonesia [1]. Di Indonesia sendiri, kasus COVID-19 pertama dan kedua di Indonesia terdeteksi pada tanggal 2 Maret 2020. Tercatat sampai 27 Januari, sudah lebih dari 1 juta kasus di Indonesia dengan diantaranya 28.468 meninggal karena virus tersebut [2]. Banyak upaya yang dilakukan pemerintah untuk mengurangi kenaikan kasus COVID-19 dari Pembatasan Sosial Berskala Besar (PSBB) sampai vaksinasi. Upaya tersebut dapat di gambarkan dalam berbagai hal seperti laporan, data, dan media berita. Media berita selalu memiliki peranan penting dalam menginformasikan berita kepada pembaca untuk memuaskan keingintahuan mereka dalam menghadapi wabah COVID-19 [3]. Informasi tersebut dapat dibaca melalui majalah, koran, dan portal berita online. Portal berita online sendiri merupakan media yang paling sering digunakan oleh pengguna dalam membaca berita karena memiliki kemudahan akses untuk membaca. Selain itu pada portal berita online jumlah data berita yang besar dalam berbagai bidang. Data-data tersebut dapat dianalisa dan diklasifikasikan untuk mendapatkan informasi terutama mengenai COVID-19 di Indonesia [4]. Hasil data tersebut

nantinya dapat divisualisasikan menjadi sebuah informasi baru kepada masyarakat, yang selama ini memiliki kesulitan akses dalam mendapatkan situasi terkini di daerah tinggal masing-masing. Selain itu, visualisasi ini dapat membantu pemerintah, organisasi kesehatan dan masyarakat agar bisa menjadi patokan dalam program mengurangi kenaikan angka kasus COVID-19, atau kasus virus yang baru kemudian harinya.

Dari Informasi diatas, dibutuhkan visualisasi berita mengenai riwayat perjalanan COVID-19 di setiap daerah provinsi di Indonesia dan membuat sebuah sistem informasi berita visualisasi Riwayat Berita Online. Sistem informasi ini merupakan sebuah web yang menampilkan visualisasi riwayat perjalanan COVID-19 di setiap provinsi di Indonesia. Sistem informasi terkait COVID-19 Indonesia pernah dibuat pada penelitian sebelumnya. Sistem Informasi ini menampilkan terkait grafik kasus barus COVID-19, pasien yang meninggal karena COVID-19, dan pasien yang sembuh dari COVID-19 pada setiap daerah [5]. Sistem informasi tersebut dapat dikembangkan dengan menampilkan keterkaitan antara angka kenaikan kasus COVID-19 dengan usaha pemerintah pusat atau daerah dalam menekan angka kasus COVID-19 di Indonesia. Usaha tersebut didapatkan melalui berita dari portal berita online menggunakan teknik web scraping. Berita yang diambil berasal dari portal berita online terkenal yaitu Kompas dan Tribunnews. Pemilihan kedua portal berita tersebut dikarenakan kedua portal berita tersebut memiliki banyak informasi terkait Informasi COVID-19 di berbagai provinsi atau kota di seluruh Indonesia. Selain itu, kedua portal berita tersebut memiliki kemudahan akses mendapatkan artikel berita terdahulu untuk dilakukan web scraping dibandingkan portal berita lainnya. Berita yang didapatkan harus berhubungan dengan COVID-19 dan berada pada sesuai daerah provinsi di Indonesia agar berita yang didapatkan sesuai dengan tema COVID-19 pada provinsi masing-masing.

Selain mendapatkan data berita dari portal berita online, diperlukannya informasi yang tidak ada dalam artikel berita untuk membantu mendapatkan hasil analisa berita yang didapatkan melalui klasifikasi teks pada data berita COVID-19 yang telah didapatkan. Proses klasifikasi teks pernah digunakan pada penelitian sebelumnya. Penelitian tersebut menggunakan klasifikasi teks terkait berita online pada umumnya dengan model Random Forest [6]. Pengujian fitur menggunakan n-gram juga dilakukan pada penelitian tersebut agar fitur yang didapatkan pada klasifikasi berita memiliki akurasi prediksi yang tinggi. Dengan adanya visualisasi ini diharapkan dapat membantu menginformasikan masyarakat dan membantu pemerintah pusat dan daerah dalam menentukan kebijakan dalam mengurangi angka kenaikan kasus COVID-19.



Gambar 1 Diagram Alir Implementasi Temu Kembali Informasi Berita Covid-19

## II. METODOLOGI

Tahapan pertama dalam pembuatan temu kembali informasi Berita Covid-19 dapat adalah persiapan data yang didapatkan dari hasil *web scraping*. Kedua adalah klasifikasi teks yang dilakukan untuk melengkapi data berita yang didapatkan dari proses *web scraping*. Dan yang terakhir adalah pengimplementasian data kedalam sebuah web. Pembahasan latar dasar teori dalam pembuatan Sistem Informasi Berita COVID-19 dapat dilihat sebagai berikut.

### A. Persiapan Data

Tahapan pertama dalam proses *web scraping* adalah menentukan data yang ditampilkan didalam Sistem Informasi. Data yang akan digunakan adalah data COVID-19 dan data berita. Proses pengambilan data COVID-19 dilakukan setiap jam 00.00 WIB sementara pengambilan data berita dilakukan setiap 1 jam. Data COVID-19 menggunakan data KawalCOVID-19 dalam bentuk *spreadsheet*. Proses pengambilan data COVID-19 menggunakan bahasa pemrograman Python dengan pustaka *gsread* dan *ServiceAccountCredentials*. Library *gsread* merupakan pustaka API Python untuk google Sheets sedangkan *ServiceAccountCredentials* merupakan pustaka python untuk mendapatkan otorisasi API dari kredensial Google API. Data yang diambil pada dataset KawalCOVID-19 adalah data linimasa peningkatan jumlah kasus COVID-19, jumlah pasien meninggal karena COVID-19 dan jumlah sembuh dari COVID-19 pada setiap provinsi di Indonesia.

Sedangkan data berita diambil menggunakan metode *Web scraping*. *Web scraping* atau *web extracting* merupakan sebuah teknik untuk mengekstrak data spesifik dari *World Wide Web* (WWW) dan disimpan ke dalam berkas sistem atau database untuk pengambilan keputusan atau analisis selanjutnya [7]. Proses *web scraping* menggunakan bahasa pemrograman Python dan *framework* Scrapy. Scrapy merupakan sebuah *framework* yang cocok untuk melakukan

*web scraping* terutama pada web yang terdiri dari banyak struktur halaman yang sama. Proses *web scraping* dapat dilihat pada gambar 1 yang dimulai dengan menentukan web tujuan sampai menyaitu portal berita online Kompas.com dan Tribunnews.com. Proses *web scraping* dilakukan setiap 1 jam untuk mendapatkan berita terbaru pada portal berita tersebut. Berita yang diambil harus memiliki tag “COVID-19”. Scrapy mengumpulkan elemen HTML, menuju link berita dan mengunduh konten data pada setiap halaman dan artikel berita.

Data yang didapatkan berupa judul berita, sumber berita, link berita, link gambar berita, tanggal, isi konten, tag, provinsi dan kota berita itu dimuat. Data yang didapatkan diperbaiki format teks agar setiap berita memiliki format yang sama. Data yang diambil akan dihapus apabila memiliki data yang sama atau tidak sesuai dengan topik COVID-19 di Indonesia atau daerah provinsi dan kota di Indonesia. Kesesuaian tag berita diperlukan dalam membedakan berita COVID-19 pada setiap daerah provinsi dan kota di Indonesia. Apabila tag berita tidak mengandung kata nama daerah, nama singkatan daerah atau nama kepala daerah tersebut, maka data berita tersebut dihapus kecuali jika tag berita tersebut mengandung kata “Indonesia” atau program pemerintah pusat seperti pembatasan sosial berskala besar (PSBB) dan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM). Data berita yang didapatkan disimpan kedalam database MYSQL. Jumlah data berita yang didapatkan dari hasil *web scraping* sampai sekarang adalah 30000 data.

### B. Klasifikasi Teks

Klasifikasi teks atau kategorisasi teks merupakan proses otomatis yang menempatkan teks dokumen ke dalam suatu kategori berdasarkan isi dari teks tersebut. Proses klasifikasi teks dapat dilihat pada gambar 1 dimulai dari persiapan data, pemrosesan teks, pemilihan fitur/rekayasa fitur, pelatihan model sampai pemberian kategori. Proses klasifikasi teks digunakan pada data berita untuk mendapatkan informasi yang tidak ada data berita yaitu label berita. Proses pemberian label berita diperlukan karena belum adanya label atau kategori berita yang didapatkan dari proses *web scraping*.

TABEL I. JENIS LABEL BERITA COVID-19

Nama Label	Keterangan	Contoh Judul Berita
Informasi	Pengumuman jumlah peningkatan atau penurunan kasus COVID-19	UPDATE 9 April: Tambah 121, Total 42.348 Orang Meninggal Dunia akibat Covid-19 di Indonesia
Donasi	Donasi atau bantuan dalam upaya pencegahan dan pengendalian wabah	12 Ton Peralatan Kesehatan Yang Baru Tiba di Indonesia Bantuan Pemerintah Tiong-kok
Kritik	Mempertanyakan tindakan pemerintah pusat atau daerah atau pihak lain terhadap kebijakan atau tindakan terkait COVID-19	4 Kritik Anggota Komisi IX DPR Agar Pemerintah Fokus Tangani Corona
Hoaks	Klarifikasi Berita Palsu, Menanggapi rumor atau kritikan	12 Karyawan Pertamina ODP Covid-19 Diusir dari

### 1) Menentukan Label dan Sampel Data

Tahap awal dalam proses klasifikasi teks, menentukan kategori label berita yang digunakan untuk mengelompokkan setiap teks data kedalam kategori label tersebut. Penentuan label mengutamakan kategori berita berdasarkan gambaran usaha pemerintah dalam menangani COVID-19 yang digambarkan pada Tabel 1. Setelah menentukan Label, diperlukannya sampel data berita untuk dilakukan pemrosesan teks dan pembuatan fitur. Sampel data berita dibuat dengan memilih secara acak data berita dan memberikan label secara manual berdasarkan label yang telah ditentukan pada Tabel 1. Dari hasil sampel data berita didapatkan 3300 data. Data tersebut dibagi menjadi sampel data latih sebesar 85% dan sampel data tes sebesar 15%. Data latih digunakan untuk mengukur akurasi pada setiap model. Sedangkan data tes digunakan untuk mengukur akurasi pada setiap model dan menguji fitur model dan menguji performa setiap model.

### 2) Pemrosesan Teks

Tahap kedua adalah pemrosesan teks. Tahap ini adalah mengubah teks menjadi bentuk yang lebih mudah dimengerti sehingga tidak ada distorsi yang diperkenalkan ke model dan algoritma pembelajaran mesin dapat bekerja lebih baik. Tahap Pemrosesan Teks terdiri dari penghapusan spasi ganda, mengubah huruf besar menjadi kecil, menghapus tanda baca dan angka, mengubah semua kata menjadi kata dasar atau lematematisasi teks dan yang terakhir adalah menghapus beberapa kata yang tidak penting atau *stopword*. List *stopword* dibuat berdasarkan *Natural Language Toolkit* (NLTK) pada bahasa Indonesia. Kata-kata seperti “Indonesia”, “COVID-19”, korona dan beberapa kata lainnya juga dihapus untuk meningkatkan kualitas fitur dalam proses pelatihan model.

### 3) Pembobotan fitur

Pembobotan fitur adalah sebuah proses mengubah data menjadi fitur untuk model klasifikasi sehingga fitur berkualitas baik dan membantu meningkatkan kinerja model. Proses rekayasa fitur dijalankan menggunakan fitur TF-IDF atau *Term Frequency – Inverse Document Frequency*. TF-IDF adalah bobot dari sejumlah kata dalam dokumen dalam seluruh korpus [8]. Formula dalam proses pembobotan TF-IDF dapat dilihat dibawah ini.

$$TFIDF(t, d) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right) \quad (1)$$

t adalah kata dalam satu dokumen, d adalah dokumen, tf(t) adalah frekuensi kata yang muncul di dokumen d, N adalah jumlah dokumen di dalam korpus dan DF(t) adalah jumlah dokumen di dalam korpus.

### 4) Pelatihan Model

Tahap terakhir adalah pelatihan model, tahap ini melakukan uji performa data latih pada beberapa jenis model. Model digunakan dalam melabelkan atribut label pada dataset berita adalah *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbor*, dan *Multinomial Naïve Bayes*.

*Random Forest* adalah sebuah klasifier yang terdiri dari kumpulan dari klasifier pohon terstruktur. Setelah pohon terbentuk, setiap pohon memberikan vote untuk kelas paling populer [9]. *K-Nearest Neighbor* atau disingkat KNN adalah

sebuah klasifier yang memberi label terhadap objek berdasarkan jarak dari k buah tetangga pada data latih. SVM atau *Support Vector Machine* adalah algoritma pelatihan membangun model yang memprediksi label untuk setiap data masukan dengan data contoh yang sudah diberi label dengan mencari hyperplane terbaik dalam ruang n-dimensi dengan memisahkan jarak antar kelas menggunakan data yang sudah diberi label atau data latih. Dan yang terakhir adalah *Multinomial Naïve Bayes*, yaitu sebuah model klasifier yang memberi nilai berdasarkan perhitungan probabilitas kelas pada data latih [10].

Keempat model tersebut dibandingkan berdasarkan akurasi data latih dan tes paling tinggi. Model yang memiliki akurasi data latih dan tes paling tinggi digunakan sebagai model utama dalam melabelkan data berita. Sebelum melakukan pelabelan berita, diperlukan penentuan ambang batas untuk menentukan atribut label. Apabila data berita memiliki salah satu nama label sesuai Tabel I dengan probabilitas bersyarat diatas 65% maka atribut label akan diberi nilai sesuai nama atribut tersebut. Jika ada dua nama label yang memiliki probabilitas bersyarat diatas 65%, maka akan dipilih nama label yang memiliki probabilitas tertinggi. Data berita yang tidak memiliki nama label yang diatas 65% akan diberi nilai “lain-lain”.

### C. Laravel

Hasil dataset berita ditampilkan pada web dengan framework Laravel, Laravel adalah *framework* aplikasi web untuk pengembangan web dengan menerapkan konsep *Model View Controller*. Laravel bertujuan untuk membuat proses pengembangan mudah bagi pengembang tanpa mengorbankan fungsionalitas aplikasi seperti otentikasi, routing, session dan caching.

TABEL II. CONTOH DATA BERITA COVID-19

Atribut	Deskripsi	Contoh Berita
judul	Judul berita	UPDATE 31 Maret: Bertambah 1 Orang, Total 3 Pasien Positif Covid-19 asal Gresik
portal	Nama Portal Berita	Kompas
link	Link URL berita	<a href="https://regional.kompas.com/read/2020/03/31/1816.....">https://regional.kompas.com/read/2020/03/31/1816.....</a>
image	Link gambar dari berita	<a href="https://asset.kompas.com/crops/hKqbeF438BgQ15.....">https://asset.kompas.com/crops/hKqbeF438BgQ15.....</a>
tanggal	Tanggal Berita dimuat	2020-03-31
konten	Isi dari berita	Satuan Tugas Gugus Pencegahan Penyebaran Covid-19 Kabupaten mengumumkan....
tag	Tag dari berita	Gresik, Virus Corona di Indonesia, Gresik darurat pencegahan Covid-19, gresik positif corona
provinsi	Kategori berita berdasarkan provinsi	jawa timur
Kota	Kategori berita berdasarkan kota	Gresik
Label	Kategori berita berdasarkan label	Informasi *

### III. HASIL DAN DISKUSI

#### A. Persiapan Data

Dapat dilihat pada Tabel II merupakan hasil Atribut Berita yang didapatkan pada proses web scraping dapat dilihat dan tidak ditandai (\*). Sementara Atribut yang ditandai (\*), yaitu label, menandakan atribut tersebut tidak didapatkan pada proses web scraping melainkan dilakukan pada proses klasifikasi teks.

#### B. Hasil dan Uji Coba Klasifikasi Teks

Dari hasil Tabel III, menandakan bahwa *Random Forest* adalah Model yang digunakan pada proses klasifikasi Teks karena memiliki akurasi set latih dan akurasi set tes yang tinggi. Pemberian label pada dataset berita juga menggunakan model tersebut.

Meskipun sudah mendapatkan model klasifikasi teks, diperlukannya pengujian performa untuk menganalisa kesesuaian label dengan judul berita menggunakan masing-masing model. Pengujian fitur dan performa dilakukan dengan menggunakan data tes diambil dari 15% dari sampel data berita.

Pengujian fitur dilakukan agar setiap label mempunyai kata kunci yang mempunyai hubungan dengan nama label masing-masing. Pengujian ini juga dilakukan agar bisa menguji fitur data latih agar membantu dalam proses pelatihan model. Kesesuaian kata kunci menggunakan cara dengan mencari kata *unigram* dan *bigram* yang paling cocok. *Unigram* dan *bigram* adalah bagian dari *n-gram*, yaitu sebuah metode pengolahan dokumen dengan membagi kalimat sebanyak 1 dan 2 kata sesuai urutan teks. Proses mencari *unigram* dan *bigram* menggunakan data tes dan dirubah menjadi sebuah fitur menggunakan metode TFIDF.

Dapat dilihat bahwa setiap label pada Tabel V memiliki korelasi *bigram* yang sama kecuali label donasi sementara *unigram* sama sekali tidak ada label yang memiliki *unigram* yang sama. Ini dapat disebabkan karena kata *bigram* tersebut memiliki sering muncul pada label informasi. Sehingga saat kata tersebut muncul di tabel kritik dan donasi menyebabkan kata tersebut mendapatkan skor yang besar dan menjadi kata *bigram* tertinggi. Meskipun begitu pada Tabel IV, label *random forest* memiliki nilai *precision*, *recall* dan *f1-score* yang sempurna pada masing-masing tabel. Ini dikarenakan

pada label informasi, kritik, dan hoaks memiliki berbagai macam keunikan kata *unigram* sehingga pohon keputusan memutuskan prediksi label sesuai dengan data aktual label. Selain itu, jumlah sampel berita yang sangat banyak membantu dalam proses pelatihan model sehingga pada setiap label memiliki nilai *precision*, *recall*, dan *f1-Score* yang tinggi.

Jika dilakukan analisa pada *unigram* di masing-masing label, label donasi memiliki kata “ikan” yang tidak ada hubungannya dengan donasi. Tetapi setelah diteliti lagi, kata ikan berasal dari kata “berikan” yang telah dilakukan proses lemmatisasi teks. Sedangkan pada label kritik dan hoaks, seharusnya salah satunya memiliki kata “pemerintah” yang merupakan adalah aktor utama dalam pengendalian covid-19. Tetapi setelah diteliti lebih lanjut, kata “pemerintah” telah dilemmatisasi menjadi “perintah” dan kata termasuk kata yang dihapus pada proses penghapusan stopword karena termasuk dalam list stopword NLTK Bahasa Indonesia. Dari contoh kata tersebut dapat disimpulkan bahwa banyak kata yang mempresentasikan sebuah label tetapi dikarekanakan kata tersebut termasuk dalam list stopword NLTK Bahasa Indonesia sehingga kata tersebut dihapus dari fitur.

Berdasarkan prediksi label yang salah pada setiap model pada Tabel IV, label kritik dan hoaks memiliki nilai *precision*, *recall* dan *f1-score* yang rendah. Ini disebabkan karena pada tabel V label hoaks dan kritik memiliki kata “kritik” sebagai lima kata *bigrams* tertinggi pada kedua label tersebut sehingga ada kemungkinan kedua label tersebut memiliki hasil prediksi yang salah. Selain itu kedua label tersebut memiliki penjelasan label yang sangat berlawanan. Label kritik lebih mengarah ke berita tentang kritikan kepada pemerintah pusat sedangkan hoaks mengarah ke berita tentang menanggapi kritikan. Akibatnya, tidak menutup kemungkinan jika terjadi kesalahan prediksi pada kedua label tersebut karena beberapa kata termasuk kata “kritik”.

TABEL IV. PENGUJIAN PEFORMA KLASIFIKASI TEKS

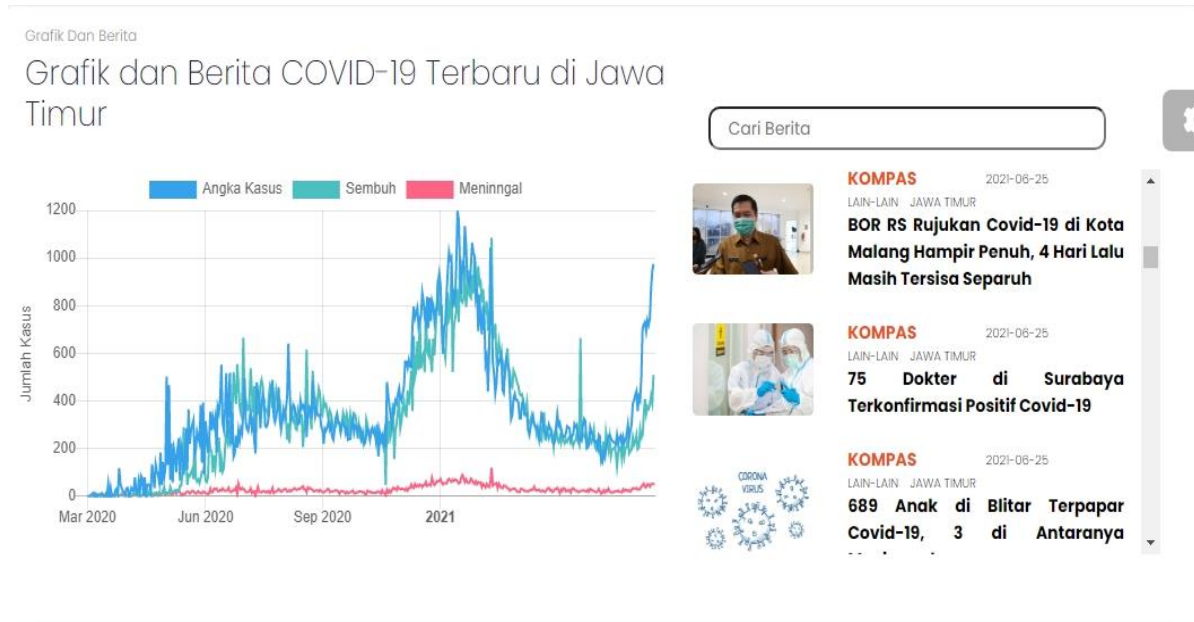
Nama Model	Akurasi Set Latih	Akurasi Set Tes
<i>Random Forest</i>	0.978	0.945
SVM	0.930	0.926
MNB	0.891	0.873
KNN	0.858	0.850

TABEL IV. PENGUJIAN PEFORMA KLASIFIKASI TEKS

Label	<i>Random Forest</i>			<i>Support Vector Machine</i>		
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Informasi	1.0	1.0	1.0	0.99	1.0	0.99
Donasi	0.98	0.94	0.96	0.98	0.92	0.95
Kritik	0.80	0.89	0.84	0.76	0.93	0.83
Hoaks	0.89	0.83	0.86	0.88	0.75	0.81
Rata-Rata	0.92	0.92	0.91	0.92	0.92	0.91
Label	<i>K-Nearest Neighbor</i>			<i>Multinomial Naïve Bayes</i>		
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Informasi	0.91	0.98	0.94	0.91	0.97	0.94
Donasi	0.86	0.74	0.80	0.88	0.74	0.81
Kritik	0.73	0.77	0.75	0.79	0.68	0.74
Hoaks	0.77	0.69	0.73	0.83	0.91	0.87
Rata-Rata	0.82	0.80	0.80	0.85	0.83	0.84

TABEL V. UNIGRAM DAN BIGRAM PALING BANYAK PADA MASING-MASING LABEL

No	Informasi		Donasi		Kritik		Hoaks	
	Unigram	Bigram	Unigram	Bigram	Unigram	Bigram	Unigram	Bigram
1	Update	Positif sembuh	Bantu	Salur bantu	Gagal	Sembuh tinggal	Fakta	Positif sembuh
2	Total	Total positif	Sumbang	Bantu tangan	Bukti	Total Positif	Klaim	Total Positif
3	Sembuh	Sembuh tinggal	Donasi	Bantu warga	Kritik	Positif sembuh	Hoaks	Sembuh Tinggal
4	Positif	Update februari	Ikan	Terima bantu	Singgung	Pemprov dki	Klarifikasi	Wali Kota
5	Bantu	Update januari	Dampak	Warga dampak	Buruk	Pasien Positif	Kritik	Update Februari



Gambar 2 Tampilan Halaman Utama Berita COVID-19

### C. Hasil Implementasi Web

Hasil dataset COVID-19 dan berita COVID-19 yang didapatkan dari *web scraping* dan dilabelkan dari proses klasifikasi teks diimplementasikan kedalam web menggunakan framework Laravel. Hasil implementasi web dapat dilihat pada Gambar 2.

#### 1) Halaman Utama Berita COVID-19

Halaman utama riwayat berita COVID-19 merupakan halaman yang pertama kali muncul sebelum menuju halaman statistik berita atau daftar berita. Halaman ini menampilkan grafik COVID-19 dan berita di seluruh Indonesia dari tanggal 18 Maret 2020 sampai tanggal saat membuka web seperti gambar 2. Tujuan dari halaman ini adalah agar pengguna bisa membandingkan antara data COVID-19 dengan kejadian pada tanggal tersebut melalui berita. Pengguna bisa menyaring data COVID-19 dan berita berdasarkan judul berita, tanggal, tempat. Selain itu data COVID-19 juga bisa disaring berdasarkan kenaikan total kasus atau kenaikan munculnya kasus dalam sehari. Halaman ini juga menampilkan tombol menu yang mengarah ke halaman daftar riwayat berita COVID-19 dan halaman statistik berita COVID-19.

#### 2) Halaman Daftar Riwayat Berita COVID-19

Halaman daftar riwayat berita COVID-19 merupakan halaman yang menampilkan setiap berita COVID-19 pada masing-masing provinsi dan kota di seluruh Indonesia. Setiap

halaman daftar riwayat berita COVID-19 akan menampilkan 10 halaman berita agar mengurangi waktu eksekusi saat membuka halaman tersebut. Berita pada halaman ini dapat disaring berdasarkan judul berita, tanggal, tempat, dan data harian atau total. Tujuan dari halaman ini adalah agar memudahkan pengguna dalam mengakses dan membaca kelengkapan berita pada daerah masing-masing.

#### 3) Halaman Daftar Pencarian Berita COVID-19

Halaman ini digunakan untuk menampilkan daftar riwayat berita COVID-19 berdasarkan hasil pencarian judul berita pada halaman daftar riwayat berita. Setiap halaman daftar riwayat berita COVID-19 akan menampilkan 10 halaman berita agar mengurangi waktu eksekusi saat

#### 4) Halaman Statistik Berita COVID-19

Halaman ini digunakan untuk menampilkan statistik riwayat berita COVID-19 berdasarkan kategori berita. Grafik statistik berupa *bar chart* dan *stacked bar chart* berdasarkan label. Statistik berupa *bar chart* dan *stacked bar chart*. *Bar chart* menampilkan jumlah berita berdasarkan kategori berita. Sementara *stacked bar chart* menampilkan persebaran kategori berita pada setiap bulannya. Tujuan dari halaman ini adalah untuk membantu pengguna dalam menganalisa data berita COVID-19 berdasarkan kategori berita.

TABEL VI. PERTANYAAN SURVEI DAN JAWABAN RESPONDEN

Pertanyaan	Skor (%)			
	1	2	3	4
Apakah berita COVID-19 di provinsi/kota Anda sesuai dengan apa yang ditampilkan di web?	0	0	66	34
Apakah menurut anda halaman statistik berita diperlukan?	0	0	2	98
Apakah memahami maksud dari halaman statistik berita?	0	16	58	26
Bagaimana nilai Anda terkait web sistem informasi berita COVID-19 di Indonesia?	0	8	44	48

#### A. Pengujian Web oleh User

Pengujian ini dilakukan untuk menguji hasil visualisasi data COVID-19 dan berita yang ditampilkan dalam sistem informasi COVID-19. Pengujian dilakukan oleh pengguna web dengan membuat form survei kepada pengguna. Isi survei terdiri dari profil responden yang terdiri dari nama, provinsi dan kota tempat tinggal pengguna dan media kebiasaan pengguna dalam membaca berita mengenai COVID-19.

Didapatkan bahwa rata-rata usia responden berusia 50 tahun dan memiliki usia rata-rata 22,5 tahun. Responden rata-rata berasal dari provinsi Jawa timur sebanyak 32%, 44% dari provinsi di Jawa lainnya dan sisanya berasal dari provinsi di luar Jawa. Dalam hal membaca berita COVID-19, 32 persen responden sering membaca berita COVID-19, 50 persen kadang-kadang membaca berita COVID-19 dan sisanya jarang atau tidak pernah sekali membaca berita COVID-19. Media yang sering digunakan responden untuk membaca berita COVID-19 adalah 90% melalui media sosial, 72% melalui portal berita online dan 16% melalui grup chat seperti WA/Line, 16% melalui koran. Tetapi jika dilihat berdasarkan berita COVID-19 pada daerah responden masing-masing, ada peningkatan minat baca melalui koran sebesar 30% dan grup chat sebesar 25%. Sementara 88% responden membaca berita terkait COVID-19 di daerah tempat tinggal melalui media social dan 60% melalui portal berita online.

Selama pengujian web oleh responden, responden dipersilahkan menjelajahi web sistem informasi COVID-19. Setelah itu melakukan survei pemahaman responden terkait kesesuaian berita pada masing-masing tanggal, visualisasi data berita, serta pendapat dan saran responden mengenai web.

Dapat dilihat dari hasil Tabel VI, data berita yang ditampilkan oleh web sudah sesuai dengan tempat tinggal responden dan responden memahami yang ditampilkan didalam web. Meskipun begitu, diperlukan evaluasi lebih lanjut terkait kesesuaian berita pada setiap daerah provinsi dan kota. Responden juga merasa terbantu dengan adanya Sistem Informasi Berita topik COVID-19 karena informatif dan mudah mengakses berita daerah terutama pada bagian kota. Responden juga memberi saran untuk meningkatkan desain

halaman Sistem Informasi dan informasi berita tidak hanya berasal dari berita online tetapi juga dari sumber lain.

#### IV. KESIMPULAN DAN SARAN

Setelah melakukan terkait penelitian dapat disimpulkan bahwa proses pembuatan Sistem Informasi Riwayat Berita Online COVID-19 berhasil dibangun. Data berita yang didapatkan berasal dari portal berita online dengan menggunakan metode web scraping. Proses web scraping mengambil data dari halaman tersebut dengan mengambil elemen HTML pada halaman web yang dituju. Data berita pada masing-masing daerah provinsi dan kota di Indonesia yang digunakan pada halaman sistem informasi ini sudah sesuai. Penggunaan klasifikasi teks diperlukan untuk mengkategorikan data berita COVID-19 dan melengkapi data berita. Model yang digunakan pada klasifikasi teks adalah Random Forest dengan akurasi set latih sebesar 97,8 dan akurasi set tes sebesar 94,5%.

Saran dalam penelitian ini adalah memperbaiki sampel data berita untuk meningkatkan kualitas model saat proses klasifikasi teks dan data berita COVID-19 bukan berasal dari portal berita online saja, tetapi juga bisa berasal dari informasi dari pemerintah pusat atau daerah sehingga kebijakan dan informasi pada setiap daerah dapat diperoleh dari satu sumber. Dan yang terakhir diperlukannya penambahan visualisasi atau analisa terkait COVID-19 pada Web Sistem Informasi Riwayat COVID-19 di Indonesia selain dari data berita sehingga setiap visualisasi dapat mendukung analisa satu sama lain dan sistem informasi ini dapat membantu terkait penanganan COVID-19.

#### V. REFERENSI

- [1] Worldometer. (n.d.). Retrieved January 27, 2021, from COVID-19 Coronavirus Pandemic: <https://www.worldometers.info/coronavirus/>
- [2] Kawal Covid-19, "Kawal informasi seputar COVID-19 secara tepat dan akurat," <https://kawalcovid19.id/> (accessed Jan. 27, 2021).
- [3] S. Anggarini, "Fenomena Dalam Berita Covid-19," *J. Audience*, vol. 3, no. 2, pp. 224–249, 2020.
- [4] K. Jahanbin and V. Rahmadian, "Using twitter and web news mining to predict COVID-19 outbreak," *Asian Pac. J. Trop. Med.*, vol. 13, no. 8, pp. 378–380, 2020.
- [5] S. Susanto, J. P. Manurung, and F. W. Setyawan, "Information System Design COVID-19 with Prototype Model," *Buana Inf. Technol. Comput. Sci.*, vol. 1, no. 2, pp. 46–50, 2020.
- [6] D. Liparas, Y. Hacoheh-kerner, A. Moumtzidou, S. Vrochidis, and I. Kompatsiaris, "News Articles Classification Using Random Forests and Weighted Multimodal Features News articles classification using Random Forests and weighted multimodal features," no. November, 2014.
- [7] A. V. Saurkar, "An Overview On Web Scraping Techniques And Tools," *Int. J. Futur. Revolut. Comput. Sci. Commun. Eng.*, vol. 4, no. 4, pp. 363–367, 2018.
- [8] C. Yang and X. He, "A Text Feature Selection Algorithm Based on Improved TFIDF," *2008 Chinese Conference on Pattern Recognition*, pp. 1–4, 2008.
- [9] B. Xu, X. Guo, Y. Ye, and J. Chen, "An Improved Random Forest Classifier for Text Categorization.," *J. Comput.*, vol. 7, pp. 2913–2920, 2012.
- [10] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, and O. Rusli, "News Article Text Classification in Indonesian Language," *Procedia Comput. Sci.*, vol. 116, pp. 137–143, 2017.

