



TUGAS AKHIR

IMPLEMENTASI WEB SISTEM INFORMASI UNTUK VISUALISASI RIWAYAT BERITA ONLINE TOPIK COVID-19 DI INDONESIA

Penyusun Tugas Akhir:

Muhammad Naufal Refadi
NRP :05111740000097

Pembimbing Tugas Akhir:

DIANA PURWITASARI, S.Kom., M.Sc.
AGUS BUDI RAHARJO, PhD

Departemen Teknik Informatika
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Latar Belakang

Jumlah kasus COVID-19 yang terus meningkat terutama di Indonesia meskipun sudah ada program pemerintah untuk menekan angka kasus COVID-19 seperti vaksinasi. Sehingga diperlukan visualisasi antara Kasus COVID-19 dan berita COVID-19 di seluruh daerah (provinsi/kota) Indonesia. Visualisasi juga dibantu dengan proses klasifikasi teks untuk memberi informasi yang tidak ada di berita dan membantu dalam menganalisa keterkaitan antara berita dengan kenaikan kasus COVID-19. Diharapkan visualisasi ini dapat membantu menginformasikan masyarakat dan membantu pemerintah pusat dan daerah dalam menentukan kebijakan untuk mengurangi angka kenaikan kasus COVID-19.

Rumusan Masalah

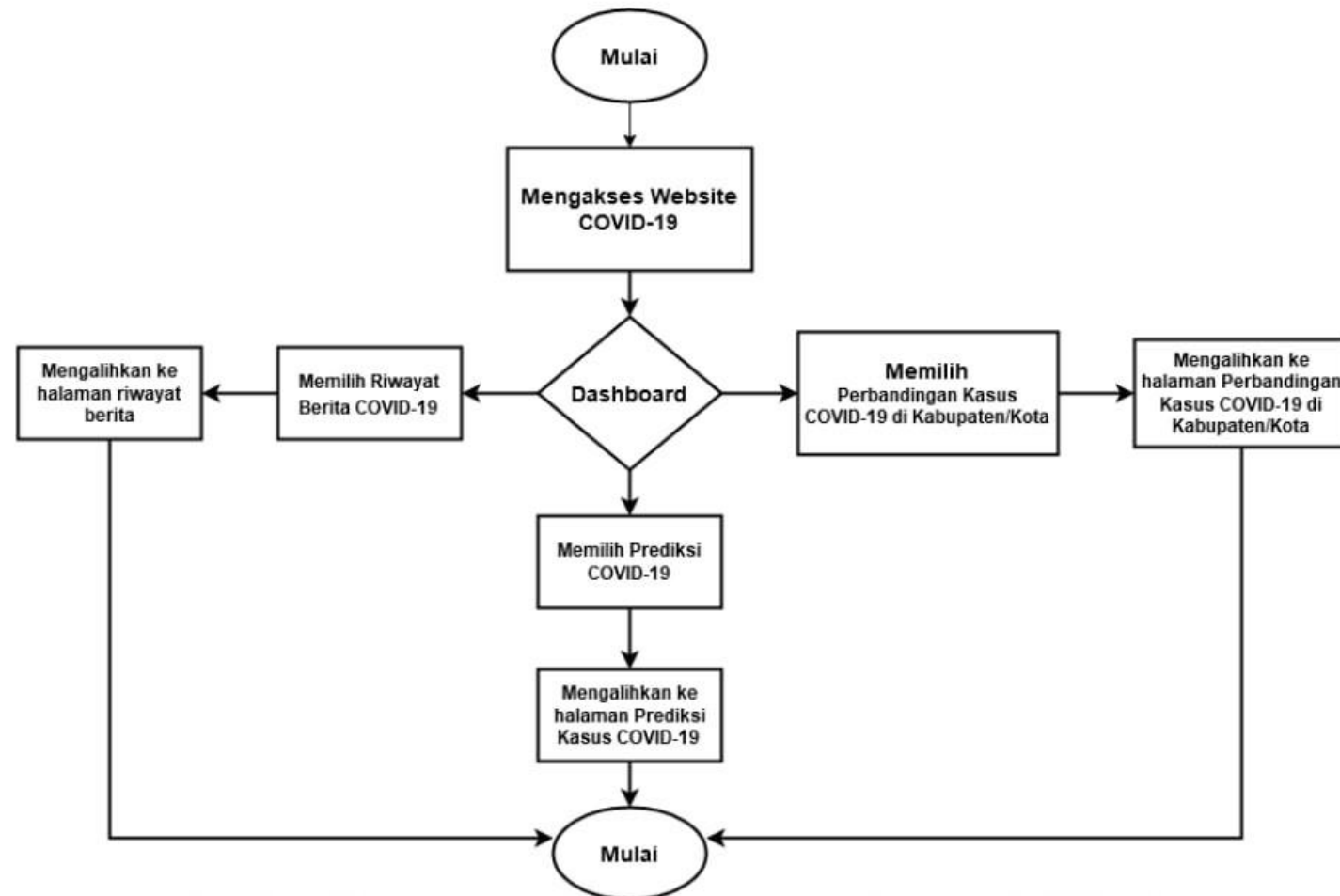
1. Bagaimana melakukan pengumpulan data dari portal berita online?
2. Bagaimana melakukan proses klasifikasi teks dalam mendapatkan informasi baru?
3. Bagaimana menampilkan visualisasi Riwayat Berita Online Topik COVID-19?
4. Bagaimana melakukan implementasi Sistem Informasi untuk Visualisasi Riwayat Berita Online ke dalam sebuah web?

Tujuan

Tujuan dari pembuatan tugas akhir ini adalah untuk mengklasifikasikan berita COVID-19 dan membangun sebuah website sistem informasi yang menampilkan visualisasi Riwayat Berita Online Topik COVID-19 dari awal muncul COVID-19 di Indonesia.

Perancangan Sistem

Desain Umum Sistem



Desain Sistem Informasi Berita

**Pengambilan Data
Covid-19 dan Berita**

1. Pengambilan data melakukan metode web scraping
2. Pengambilan data COVID-19 dilakukan setiap jam 1 malam
3. Pengambilan data Berita terbaru setiap 1 jam

Pemrosesan Teks

Mendapatkan fitur klasifikasi teks dan menampilkan wordcloud

Klasifikasi Teks

Proses ini terdiri dari penentuan label, pemilihan data latih dan tes, pembuatan fitur, pelatihan model, pemilihan model dan pemberian atribut label pada data berita

Visualisasi Berita

Tampilan visualisasi dalam bentuk website

Pengambilan Data

Dataset COVID-19

1. Dataset COVID-19 didapatkan dari data kawalcovid-19
2. Sistem mengupdate data terbaru setiap jam 01.00 WIB
3. Dataset terdiri dari 6 Tabel yaitu Kasus Positif, Meninggal dan Sembuh (Harian dan Total)

Dataset Berita

1. Data Berita diambil dari Tribunnews dan Kompas dengan metode web scraping
2. Sistem melakukan scraping data berita terbaru setiap 1 jam
3. Data yang didapatkan dari hasil web scraping berita adalah judul, portal berita, URL, URL gambar, tanggal, konten, dan tag Berita.
4. Data Berita yang baru berupa data provinsi, kota dan label berita

Pemrosesan Teks

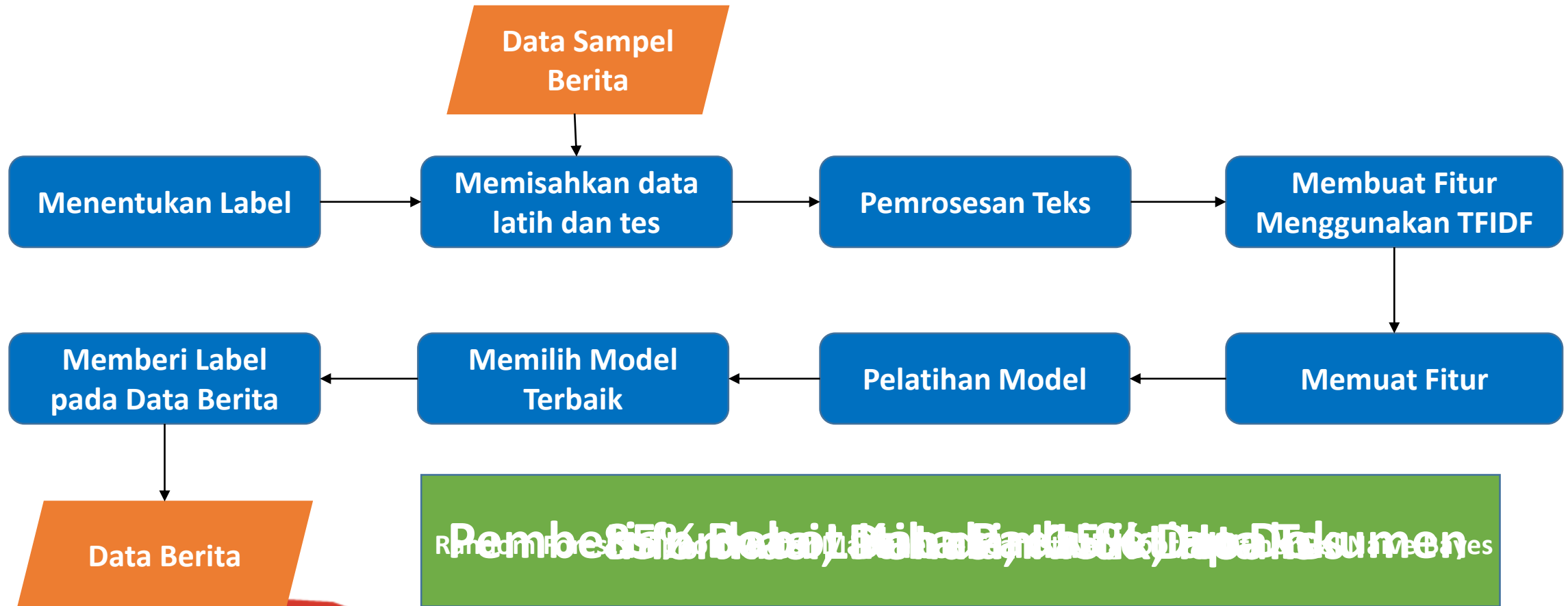
**1.379.662 Kasus Covid-19 di Indonesia, PPKM
Mikro Diklaim Tekan Kasus Harian**



ppkm mikro klaim tekan

1. Menghapus tab, enter dan spasi ganda
2. Menghapus tanda baca dan angka
3. Lemmatisasi/stemming judul berita
4. Menghapus kata yang mengandung stopwords

Klasifikasi Teks



Pemilihan Model/Model Training

Model	Akurasi Set Latih	Akurasi Set Tes
Random Forest	0.963004	0.945032
SVM	0.930493	0.926004
Multinomial Naïve Bayes	0.891256	0.873150
KNN	0.857997	0.849894

- Random Forest adalah model terbaik dalam klasifikasi teks
- Setiap Data Berita akan dilabelkan dengan memprediksikan judul berita
- Apabila judul berita memiliki tingkat kecocokan dibawah 65% pada semua label. Maka Label berita pada data berita tersebut adalah “lain-lain”.

Kata-Kata yang Sering Muncul pada Setiap Label

Informasi

No	Unigrams	Bigrams
1	<u>Update</u>	<u>Positif sembuh</u>
2	<u>Total</u>	<u>Total positif</u>
3	<u>Sembuh</u>	<u>Sembuh tinggal</u>
4	<u>Positif</u>	<u>Update februari</u>
5	Bantu	<u>Update januari</u>
6	Tinggal	<u>Update maret</u>
7	Fakta	<u>Pasien positif</u>
8	April	<u>News update</u>
9	<u>Perintah</u>	<u>Pasien sembuh</u>
10	Februari	<u>Breaking news</u>
11	<u>Jokowi</u>	<u>Update pasien</u>
12	Januari	<u>Update mei</u>
13	<u>Vaksin</u>	<u>Update virus</u>
14	Orang	<u>Update april</u>
15	<u>Klaim</u>	Total pasien

Donasi

No	Unigrams	Bigrams
1	<u>Bantu</u>	Salur bantu
2	Sumbang	Bantu tangan
3	Donasi	Bantu warga
4	Ikatan	Terima bantu
5	Dampak	Warga dampak
6	Salur	Rp miliar
7	Rp	Tenaga medis
8	<u>Update</u>	<u>Positif sembuh</u>
9	Sosial	<u>Total positif</u>
10	Apd	Rp juta
11	Gaji	<u>Sembuh tinggal</u>
12	Alat	Rumah sakit
13	Miliar	<u>Update februari</u>
14	<u>Positif</u>	<u>Update januari</u>
15	Medis	<u>Update maret</u>

Kritik

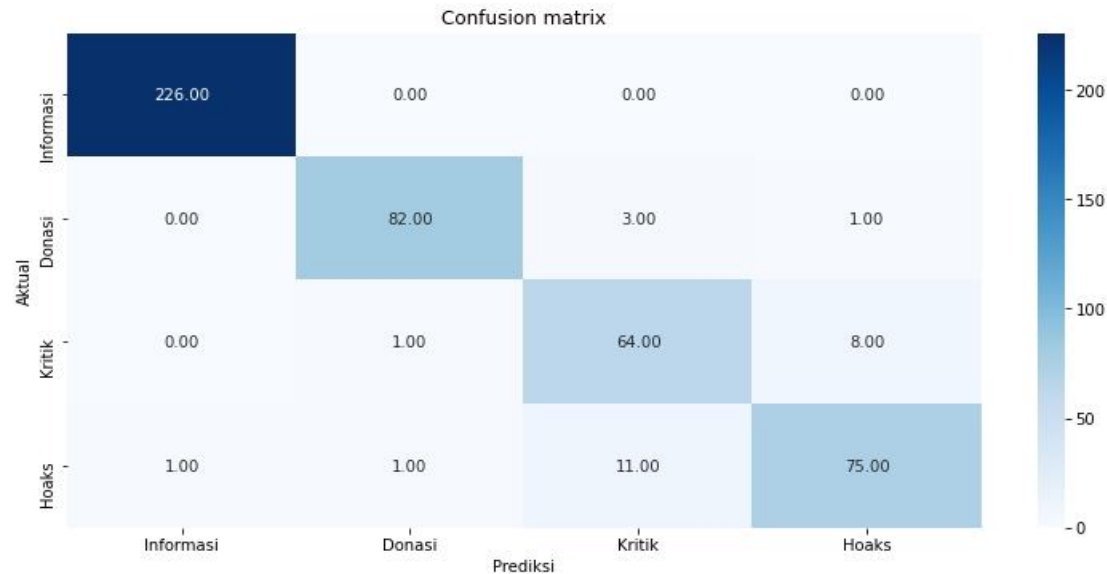
No	Unigrams	Bigrams
1	Gagal	<u>Positif Sembuh</u>
2	Bukti	<u>Total positif</u>
3	Kritik	<u>Sembuh tinggal</u>
4	Singgung	Pemprov dki
5	Buruk	<u>Pasien positif</u>
6	Perintah	<u>Update februari</u>
7	<u>Update</u>	<u>Update januari</u>
8	Ekonomi	<u>Update maret</u>
9	Masyarakat	<u>Pasien sembuh</u>
10	<u>Jokowi</u>	<u>Breaking news</u>
11	Dpr	<u>News update</u>
12	<u>Bantu</u>	Tinggal dunia
13	<u>Sembuh</u>	Anies baswedan
14	Bijak	<u>Update pasien</u>
15	Terap	<u>Update mei</u>

Hoaks

No	Unigrams	Bigrams
1	Fakta	<u>Positif sembuh</u>
2	<u>Klaim</u>	<u>Total positif</u>
3	Hoaks	<u>Sembuh tinggal</u>
4	Klarifikasi	Wali kota
5	<u>Update</u>	<u>Update februari</u>
6	<u>Vaksin</u>	<u>Update januari</u>
7	Video	<u>Update maret</u>
8	Obat	<u>News update</u>
9	<u>Bantu</u>	<u>Pasien sembuh</u>
10	Viral	<u>Breaking news</u>
11	Kendali	<u>Pasien positif</u>
12	<u>Total</u>	<u>Update pasien</u>
13	Awat	<u>Update mei</u>
14	Psbb	<u>Update virus</u>
15	Tular	<u>Update april</u>

Pengujian Model Klasifikasi Teks

Random Forest



Label	Precision	Recall	F1-Score
Informasi	1.0	1.0	1.0
Donasi	0.98	0.94	0.96
Kritik	0.80	0.89	0.84
Hoaks	0.89	0.83	0.86
Rata-Rata	0.92	0.92	0.91

VISUALISASI WEB

Survei Pengguna

Kriteria Responden

1. Responden terdiri dari pelajar, mahasiswa dan pekerja dengan usia rata-rata 19,6 Tahun.
2. 28,6% responden berasal dari Jawa Timur , 50% dari Provinsi di Jawa (non-Jawa Timur), dan 22,8% dari provinsi luar jawa.
3. 32% responden sering membaca berita terkait COVID-19, 50% kadang-kadang dan 18% Jarang/ Tidak pernah membaca berita.
4. 92,9% responden membaca berita terkait COVID-19 melalui media sosial, 64,3% melalui portal berita online dan 17,9% melalui grup chat seperti WA/Line, 3,6% melalui Koran.
5. 85,7% responden membaca berita terkait COVID-19 di daerah tempat tinggal melalui media sosial, 57,1% melalui portal berita online dan 25% melalui grup chat seperti WA/Line, 17,9% melalui Koran.

Skenario Responden

1. Membuka halaman awal
2. Membuka halaman daftar berita
3. Mencari berita COVID-19 di provinsi/kota sesuai tempat tinggal responden
4. Mencari berita berdasarkan judul
5. Mencari halaman statistik berita
6. Responden diizinkan menjelajahi halaman web berita COVID-19 dan melaporkan apabila terjadi bug/masalah.

Pendapat Responden

1. Sangat membantu, terkait penyebaran informasi pesebaran COVID-19 di daerah masing-masing
2. Informatif dan menarik
3. Bagus dan mudah dipahami

Saran dan Kritik Responden

1. Tidak responsive pada tampilan potrait Mobile
2. Halaman Statistik Berita lama sekali dibuka
3. Keterlambatan masuknya data berita

Pertanyaan

No	Pertanyaan	Skor (%)			
		1	2	3	4
1	Apakah berita COVID-19 di provinsi/kota Anda sesuai dengan apa yang ditampilkan di web?	0	0	66	34
2	Apakah menurut anda halaman statistik berita diperlukan?	0	0	2	98
3	Apakah memahami maksud dari halaman statistik berita?	0	16	58	26
4	Bagaimana nilai Anda terkait web sistem informasi berita COVID-19 di Indonesia?	0	8	44	48

Kesimpulan dan Saran

Kesimpulan

- 1) Proses pengambilan data berita dilakukan dengan mengambil elemen HTML pada halaman web yang dituju. Sedangkan proses pengambilan data covid-19 adalah mencari data label berdasarkan lokasi *sheet* dan *cell* pada halaman spreadsheet.
- 2) Model yang digunakan pada klasifikasi teks pada data berita adalah Random Forest dengan akurasi sebesar 94,5%.
- 3) Meskipun memiliki akurasi data latih dan data tes diatas 94,5%. Pesebaran data label berita lebih banyak mengarah ke label kritik.
- 4) Berdasarkan hasil survei, tampilan berita covid-19 sesuai dengan apa yang ditampilkan kedalam website. pengguna merasa terbantu dengan adanya web berita topik COVID-19.

Saran

1. Memperbaiki sampel data berita untuk meningkatkan kualitas model saat proses klasifikasi teks.
2. Menambah jenis label berita untuk membedakan berbagai jenis berita terkait COVID-19.
3. Data berita COVID-19 bukan berasal dari portal berita online saja, tetapi juga bisa berasal dari informasi dari pemerintah pusat/daerah.
4. Menambahkan visualisasi atau analisa terkait COVID-19 pada Web Sistem Informasi COVID-19 sehingga sistem informasi ini dapat membantu terkait penanggulangan COVID-19.

- TERIMA KASIH -