

**DETEKSI HOAX PADA BERITA KESEHATAN BERBAHASA INDONESIA
MENGUNAKAN ALGORITMA *MULTINOMIAL NAÏVE BAYES***

SKRIPSI

**Oleh :
SALMA AZMI ANGGITA PUTRI
NIM. 19650075**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**DETEKSI HOAX PADA BERITA KESEHATAN BERBAHASA
INDONESIA MENGGUNAKAN ALGORITMA
MULTINOMIAL NAÏVE BAYES**

SKRIPSI

Diajukan Kepada :
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
SALMA AZMI ANGGITA PUTRI
NIM. 19650075

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

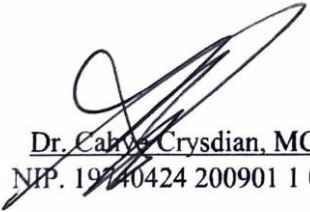
**DETEKSI HOAX PADA BERITA KESEHATAN BERBAHASA
INDONESIA MENGGUNAKAN ALGORITMA
MULTINOMIAL NAÏVE BAYES**

SKRIPSI

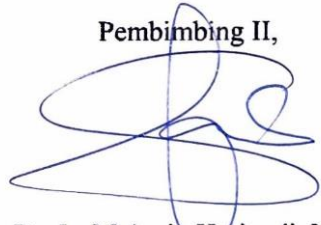
Oleh:
SALMA AZMI ANGGITA PUTRI
NIM. 19650075

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 8 Juli 2024

Pembimbing I,


Dr. Cahyo Crysdian, MCS
NIP. 19740424 200901 1 008

Pembimbing II,


Dr. Ir. M Amin Hariyadi, M.T
NIP. 19670018 200501 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ir. Fachrudin Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

DETEKSI HOAX PADA BERITA KESEHATAN BERBAHASA INDONESIA MENGGUNAKAN ALGORITMA MULTINOMIAL NAÏVE BAYES

SKRIPSI

Oleh:

SALMA AZMI ANGGITA PUTRI
NIM. 19650075

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 15 Oktober 2024

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. Ririen Kusumawati, S.Si., M.Kom</u> NIP. 19720309 200501 2 002
Anggota Penguji I	: <u>Dr. Irwan Budi Santoso, M.Kom</u> NIP. 19770103 201101 1 004
Anggota Penguji II	: <u>Dr. Cahyo Crysdian, MCS</u> NIP. 19740424 200901 1 008
Anggota Penguji III	: <u>Dr. Ir. M. Amin Hariyadi, M.T</u> NIP. 19670118 200501 1 001

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



()
Dr. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Salma Azmi Anggita Putri

NIM : 19650075

Fakultas / Prodi : Sains dan Teknologi / Teknik Informatika

Judul Skripsi : Deteksi Hoax pada Berita Kesehatan Berbahasa Indonesia
Menggunakan Algoritma Multinomial Naïve Bayes

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 15 Oktober 2024

Yang membuat pernyataan,



Salma Azmi Anggita Putri

NIM.19650075

MOTTO

“Happiness constitutes the paramount significance”

“Tidak ada kata terlambat untuk memulai sesuatu. Lakukan jika hal baik itu bermanfaat dan bisa membuatmu bahagia.”

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin. Dengan rasa syukur yang sebesar-besarnya saya menyampaikan terima kasih kepada orang tua saya bapak Yusbanu dan ibu Aries Ernawati, serta adik saya Faya Najwatus Silma atas doa, dukungan, semangat, dan kepercayaan sehingga saya dapat menyelesaikan skripsi dengan baik. Tolong agar tetap sehat dan selalu bahagia kedepannya. Terimakasih juga kepada seluruh keluarga besar saya yang juga telah banyak memberikan dukungan dan semangat untuk saya agar segera lulus. Semoga doa dan harapan yang telah diberikan oleh keluarga saya dapat menjadikan saya menjadi pribadi yang lebih baik, bertanggung jawab, dan sukses di masa mendatang.

Terima kasih kepada teman-teman yang sudah membantu saya dalam bertukar pikiran dan memberikan semangat dan dukungan yang berharga agar saya dapat menyelesaikan skripsi saya, semoga tetap terus berhubungan baik sampai dengan di masa mendatang. Tidak lupa juga saya mengucapkan terima kasih kepada diri saya sendiri yang ternyata mampu untuk menyelesaikan skripsi ini. Terima kasih untuk tidak menyerah dalam mengerjakan skripsi yang terdapat beberapa kendala dalam proses pengerjaannya. Pesan dari saya, tidak ada kata terlambat untuk memulai. Teruslah berproses sampai bertemu dengan dirimu yang sesungguhnya. Jalani hidup yang bahagia, kerjakan seluruh aktivitas dengan senang, karena dengan begitu aktivitas yang berat sekalipun akan menjadi ringan. Karya ini saya persembahkan dengan harapan dapat bermanfaat bagi pihak yang sedang mengembangkan topik penelitian ini. Terima kasih atas dukungan dan doa yang terus menerus diberikan.

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Alhamdulillahirabbil'alamin, tidak ada ungkapan yang paling pantas untuk diungkapkan selain rasa syukur dan terima kasih kepada Allah Subhanahu Wa Ta'ala atas karunia dan rahmat-Nya sehingga saya dapat menyelesaikan skripsi ini dengan baik. Shalawat dan salam semoga tercurahkan kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam yang telah memimpin kita semua dari kegelapan menuju jalan yang terang benderang.

Kerendahan hati dan penuh rasa syukur kebahagiaan ini, saya mempersembahkan skripsi yang telah saya susun ini untuk memenuhi salah satu persyaratan sarjana komputer (S.KOM) di Universitas Islam Negeri Maulana Malik Ibrahim Malang program studi Teknik Informatika Fakultas Sains dan Teknologi.

Berbagai kendala yang ditemui selama proses penyusunan skripsi ini dapat saya lewati berkat kegigihan dan kesabaran, sehingga skripsi ini dapat terselesaikan dengan baik. Serta peran dukungan dan arahan yang tak ternilai harganya yang telah diberikan kepada saya dari berbagai pihak juga membangun saya menjadi lebih giat dalam berproses penyusunan skripsi ini.

Dalam perjalanan penyusunan skripsi ini, saya senantiasa merasakan karunia Allah Subhanahu Wa Ta'ala. Segala pencapaian keberhasilan dan kemajuan yang saya raih tidak lepas dari limpahan rahmat-Nya yang begitu besar dan serta petunjuk-Nya yang selalu ada di saat saya membutuhkan. Saya merasa terhormat dan bersyukur atas kesempatan ini. Oleh sebab itu, saya ingin mengucapkan terimakasih kepada :

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT,IPU., selaku ketua program studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Cahyo Crysdian, MCS., selaku dosen pembimbing I dan Dr. Ir. Mokhammad Amin Hariyadi, M.T., selaku dosen pembimbing II yang telah membimbing penulis untuk mengembangkan pemikiran dalam penyusunan skripsi ini hingga selesai.
5. Dr. Ririen Kusumawati, S.Si., M.Kom., selaku dosen penguji I dan Dr. Irwan Budi Santoso, M.Kom., selaku dosen penguji II yang telah menguji, menasehati, serta memberikan saran untuk menjadikan penyusunan skripsi ini lebih baik.
6. Seluruh dosen dan segenap staff program studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan segala ilmu dan wawasan semasa kuliah.
7. Ibu, Ayah, dan Adik saya yang tidak bosan dalam berdoa yang tak putus-putusnya serta memberikan semangat dan dukungan agar saya bisa menyelesaikan skripsi.
8. Teman-teman program studi Teknik Informatika Angkatan 2019 “ALIEN” yang telah menghabiskan waktu, memotivasi, dan berjuang bersama penulis semasa kuliah.

9. Na Jaemin dan Lee Jeno, serta member grup NCT Dream yang musiknya selalu menemani saya dalam mengerjakan skripsi dan memberikan saya semangat melalui lagu-lagunya.
10. Teman-teman yang tidak bisa saya sebutkan namanya satu persatu yang telah menemani dan mendukung penulis saat proses pengerjaan skripsi.
11. Seluruh pihak yang tidak dapat saya sebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari masih banyak kekurangan dalam pembuatan skripsi ini karena keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi kemajuan penelitian selanjutnya. Akhir kata, semoga skripsi ini dapat bermanfaat bagi kita semua.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 15 Oktober 2024

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN TULISAN	iii
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
ABSTRAK	xiv
ABSTRACT	xv
مستخلص البحث.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
1.6 Sistematika Penulisan	6
BAB II STUDI PUSTAKA.....	7
2.1 Klasifikasi Berita Hoaks.....	7
2.2 Naïve Bayes.....	9
BAB III DESAIN PENELITIAN.....	12
3.1 Alur Penelitian.....	12
3.2 Pengumpulan Data.....	13
3.3 Desain Sistem	17
3.4 <i>Data Preprocessing</i>	18
3.4.1 <i>Cleaning</i>	20
3.4.2 <i>Tokenizing</i>	21
3.4.3 <i>Normalization</i>	23
3.4.4 <i>Stopword</i>	24
3.4.5 <i>Stemming</i>	26

3.5. <i>Term Frequenxy-Inverse Document Frequency (TF-IDF)</i>	28
3.6 <i>Naïve Bayes</i>	33
3.6 Desain Tampilan Aplikasi	37
BAB IV UJI COBA DAN PEMBAHASAN	38
4.1 Langkah-Langkah Uji Coba	38
4.1.1 Pembobotan Kata.....	38
4.1.2 Naïve Bayes Classifier.....	49
4.2 Implementasi Tampilan Antarmuka	51
4.2.1 Beranda.....	51
4.2.2 Hasil Cek Fakta Berita.....	52
4.3 Pengujian	52
4.3.1 Data Uji.....	52
4.3.2 Hasil Uji Coba	53
4.3.3 Pengujian Akurasi.....	57
BAB V PENUTUP	63
5.1 Kesimpulan.....	63
5.2 Saran	63
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 3. 1 Contoh Data Berita Fakta dari website detikcom	14
Tabel 3. 2 Contoh data berita hoax dari website turnbackhoax	16
Tabel 3. 3 Hasil Cleaning	21
Tabel 3. 4 Hasil Tokenizing	22
Tabel 3. 5 Tabel Data Sampel	29
Tabel 3. 6 Perhitungan TF	31
Tabel 3. 7 Perhitungan IDF dan TF-IDF	33
Tabel 4. 1 Hasil TF dan IDF Setiap data Kategori Fakta	39
Tabel 4. 2 Hasil TF dan IDF Setiap data Kategori Hoax	44
Tabel 4. 3 Rata-Rata TF-IDF setiap kategori	48
Tabel 4. 4 Hasil Klasifikasi Berita	53
Tabel 4. 5 Tabel Hasil Uji Berita Kesehatan di dalam Program	54
Tabel 4. 6 Konsep Confusion Matrix	57

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	12
Gambar 3.2 Desain Sistem Deteksi Berita Hoax	18
Gambar 3. 3 Diagram alur Preprocessing (Dewantara, 2022)	19
Gambar 3. 4 Diagram alur Case Folding (Dewantara, 2022)	21
Gambar 3. 5 Diagram Alur Tokenizing (Dewantara, 2022)	23
Gambar 3. 6 Diagram alur Stopword Removal (Dewantara, 2022).....	26
Gambar 3. 7 Diagram alur Stemming (Dewantara, 2022)	28
Gambar 3. 8 Diagram alur TF-IDF	33
Gambar 3. 9 Diagram Alur Metode Multinomial Naive Bayes (Dewantara, 2022)	37
Gambar 4. 1 Wordcloud.....	49
Gambar 4. 2 Halaman Beranda	51
Gambar 4. 3 Tampilan Halaman Hasil.....	52
Gambar 4. 4 Confusion Matrix	59
Gambar 4. 5 Hasil Uji Prediksi pada Program Web	67
Gambar 4. 6 Hasil Uji Prediksi pada Program Web	67
Gambar 4. 7 Hasil Uji Prediksi pada Program Web	68
Gambar 4. 8 Hasil Uji Prediksi pada Program Web	68
Gambar 4. 9 Hasil Uji Prediksi pada Program Web	68
Gambar 4. 10 Hasil Uji Prediksi pada Program Web	69
Gambar 4. 11 Hasil Uji Prediksi pada Program Web	69
Gambar 4. 12 Hasil Uji Prediksi pada Program Web	69
Gambar 4. 13 Hasil Uji Prediksi pada Program Web	70
Gambar 4. 14 Hasil Uji Prediksi pada Program Web	70

ABSTRAK

Putri, Salma Azmi Anggita. 2024. **Deteksi Hoax Pada Berita Kesehatan Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdian, MCS. (II) Dr. Ir. Mokhammad Amin Hariyadi, M.T.

Kata kunci: *Berita Kesehatan, Naïve Bayes, Multinomial Naïve Bayes, Deteksi Hoax*

Salah satu manfaat dari sosial media adalah menyebarkan informasi. Sosial media digunakan untuk wadah dari persebaran berita. Salah satunya adalah berita kesehatan. Namun, realitanya berita mengenai kesehatan yang tersebar melalui sosial media tidak seluruhnya adalah merupakan fakta. Saat ini berita hoaks semakin banyak bermunculan. Untuk meminimalisasi penyebaran berita hoaks terutama pada berita kesehatan, maka dilakukan penelitian deteksi berita hoaks mengenai kesehatan berbahasa Indonesia berdasarkan kata-kata yang menyusun berita tersebut. Deteksi hoaks ini akan diimplementasikan ke dalam aplikasi yang berbasis web. Metode yang digunakan dalam proses deteksi hoaks adalah metode klasifikasi Multinomial Naïve Bayes. Multinomial Naïve Bayes, digunakan untuk pengklasifikasian dokumen atau teks yang mana fungsi tersebut tepat untuk diterapkan pada penelitian ini yang menggunakan data berbentuk teks. Pada metode Multinomial Naïve Bayes, frekuensi kemunculan kata-kata dalam dokumen dihitung dengan tujuan untuk proses klasifikasi. Frekuensi kemunculan kata dihitung dengan menghitung bobot kata dengan TF-IDF. Data yang digunakan dalam penelitian ini, diperoleh dengan melakukan web scraping pada situs detikcom untuk kategori berita fakta dan turnbackhoax untuk memperoleh data berita hoaks. Hasil pengujian dari proses identifikasi menggunakan metode confusion matrix dan didapatkan dengan nilai akurasi sebesar 84.3% dan precision, recall, f-measure masing-masing bernilai 100%, 68.6%, dan 81.39%. Selain itu, pada tahap uji coba yang dilakukan di aplikasi berbasis web, dari 12 berita yang terdiri dari 6 berita fakta dan 6 berita hoaks, terdapat kesalahan deteksi karena teks berita yang terlalu singkat. Sehingga, mesin sulit mendeteksi secara tepat. Oleh karena itu, diperlukan penelitian lanjutan untuk menyempurnakan mesin deteksi ini, contohnya dengan menggunakan metode yang lain atau fitur yang lain.

ABSTRACT

Putri, Salma Azmi Anggita. 2024. **Hoax Detection on Indonesian Health News Using Multinomial Naïve Bayes Algorithm**. Thesis. Informatics Engineering Department Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Dr. Cahyo Crysdian, MCS. (II) Dr. Ir. Mokhammad Amin Hariyadi, M.T.

One benefit of social media is a media for distributing news. In reality, not all health news on social media is factual. Nowadays, hoax news often occurs. To minimize the spread of hoax news, especially health news, the researcher studied Indonesian hoax health news based on their words. The detection is implemented into a web-based application. The researcher employed the Multinomial Naïve Bayes classification method in the process of hoax detection to classify documents or texts. The word frequency in the document is counted for classification by calculating the word weight using the TF-IDF. The research data were obtained using web scraping on the detikcom website for fact news and turnback hoax category to get hoax news data. The test result of the identification process using the confusion matrix method gets an accuracy, precision, recall, and f-measure of 84.3%, 100%, 68.6%, and 81.39%, respectively. In addition, during the try-out process, a web-based application consisting of six factual news and six hoaxes experiences detection failure due to too short news text. The machine is hard to detect correctly. Therefore, it needs further study to improve the performance of the detection machine, for example by using other methods or features.

Keywords: *Health News, Naïve Bayes, Multinomial Naïve Bayes, Hoax Detection*

مستخلص البحث

ف فوترى، سلمى عزمى أنجيتا. 2024. كشف الخدعة في الأخبار الصحية باللغة الإندونيسية باستخدام خوارزمية ساذج بايز متعدد الحدود. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. جهيو كريسديان، الماجستير. المشرف الثاني: د. محمد أمين هريدي، الماجستير.

الكلمات الرئيسية: أخبار صحية، ساذج بايز، ساذج بايز متعدد الحدود، كشف خدعة

من فوائد وسائل التواصل الاجتماعي هي نشر المعلومات. استخدمت وسائل التواصل الاجتماعي كحاوية لنشر الأخبار. منها الأخبار الصحية. ومع ذلك، فإن الحقيقة هي أن الأخبار الصحية المنتشرة عبر وسائل التواصل الاجتماعي ليست حقيقة كاملة. حالياً، تظهر أخبار الخدعة بشكل متزايد. لتقليل انتشار الأخبار الخادعة، خاصة في الأخبار الصحية، تم إجراء بحث حول كشف الأخبار الصحية الخادعة باللغة الإندونيسية بناء على الكلمات التي تشكل الأخبار. سيتم تنفيذ كشف الخدعة هذا في تطبيق قائم على الويب. الطريقة المستخدمة في عملية الكشف عن الخدعة هي طريقة تصنيف ساذج بايز متعدد الحدود. استخدمت هذه الطريقة لتصنيف الوثائق أو النصوص حيث تكون الوظيفة مناسبة للتطبيق على هذا البحث الذي يستخدم البيانات في شكل نص. في هذه الطريقة، تم حساب تكرار حدوث الكلمات في المستند لغرض عملية التصنيف. تم حساب تكرار حدوث الكلمة عن طريق حساب وزن الكلمة باستخدام TF-IDF. تم الحصول على البيانات المستخدمة في هذا البحث عن طريق تجريف الويب على موقع detikcom لفئة الأخبار الواقعية وخدعة الارتداد للحصول على بيانات إخبارية خادعة. استخدمت نتائج اختبار عملية تحديد الهوية طريقة مصفوفة الارتباك وتم الحصول عليها بقيمة دقة 84.3٪ وقيمة الدقة والاستدعاء وقياس f بنسبة 100٪ و 68.6٪ و 81.39٪ على التوالي. بالإضافة إلى ذلك، في المرحلة التجريبية التي أجريت في التطبيق المستند إلى الويب، من بين 12 خبراً تتكون من 6 أخبار واقعية و 6 أخبار خادعة، كان هناك خطأ في الكشف بسبب نص إخباري قصير جداً. لذلك، يصعب على الجهاز الكشف بالطريقة الصحيحة. لذلك، هناك حاجة إلى مزيد من البحث لتحسين محرك الكشف هذا، على سبيل المثال باستخدام طرق أخرى أو ميزات أخرى.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Di era industri 4.0 media sosial menjadi salah satu faktor yang paling berpengaruh. Di media sosial, orang bisa mengenal banyak relasi bisnis, menjalin silaturahmi dengan orang yang jauh, menggali berbagai informasi, dan lain sebagainya. Segala sesuatu dapat didapatkan dengan mudah di era sekarang ini. Kemajuan teknologi dan informasi serta semakin canggihnya peralatan elektronik yang diproduksi seperti laptop, gawai, komputer, dan perangkat-perangkat canggih lainnya membuat era sekarang ini seperti memunculkan “dunia dalam genggaman”. Ungkapan “Dunia dalam Genggaman” sejajar dengan apa yang diutarakan oleh Thomas L. Friedman (2007) sebagai *the world is flat* yang memiliki arti, dunia ini semakin rata, masing-masing orang dapat mengakses apa saja dari sumber mana saja. Selain kemajuan teknologi, ada banyak juga bidang yang saat ini banyak berkembang dengan pesat salah satunya, bidang medis atau kesehatan.

Salah satu manfaat dari sosial media adalah menyebarkan informasi. Sosial media digunakan untuk wadah dari persebaran berita. Salah satunya adalah berita kesehatan. Namun, realitanya berita mengenai kesehatan yang tersebar melalui sosial media tidak seluruhnya adalah merupakan fakta. Saat ini berita hoaks semakin banyak bermunculan. Berita hoaks ini memuat informasi palsu yang kebenarannya tidak bisa dipertanggungjawabkan dan memiliki tujuan untuk meyakinkan warganet agar percaya dan terkecoh/tertipu akan adanya

berita hoaks tersebut. Menurut Kamus Besar Bahasa Indonesia atau yang biasa disingkat dengan sebutan KBBI, hoaks dapat diartikan sebagai berita palsu atau bohong. Adanya hoaks memiliki maksud untuk menyampaikan pemahaman atau pengakuan yang salah atau tidak sesuai dengan fakta.

Hoaks sekarang sedang menjadi perhatian serius pemerintah, salah satunya adalah adanya rumor mengenai varian baru virus corona penyebab Covid-19 yang bernama NeoCov. Doni003 (2022) Benar adanya virus NeoCov, namun NeoCov berasal dari jenis virus Corona yang berbeda yang terkait dengan *Middle East Respiratory Syndrom Coronavirus (MERS-CoV)*, yaitu penyakit yang menyerang saluran pernapasan dan virus ini baru ditemukan di kelelawar dadn belum ada penularan manusia, jadi virus ini bukan varian baru dari Covid. Penyebaran hoaks banyak muncul melalui media sosial seperti Whatsapp, X, Facebook, dan lain sebagainya. *Hit and run*, merupakan istilah atau ungkapan untuk orang yang membuka akun, kemudian melemparkan isu, dan dilanjutkan hapus akun atau tutup akun, kemudian di ulang-ulang seperti itu secara terus-menerus. Hal ini membuat pelaku penyebaran hoaks sulit untuk dianalisis.

Ketua Dewan Pers, (Prasetyo, 2017) menyebutkan bahwa tanda-tanda hoaks dapat disebutkan sebagai berikut.

- a. Menimbulkan akibat berupa permusuhan, kebencian, serta kecemasan.
- b. Sumber berita yang ambigu. Hoaks di media sosial biasanya diterbitkan oleh media yang tidak memiliki sertifikasi, tidak memiliki keseimbangan, dan condong untuk menyudutkan pihak tertentu.

- c. Memiliki muatan fanatisme yang mengatasnamakan judul, ideologi, dan pengantar provokatif, memberikan penghukuman serta menyembunyikan fakta dan data.

Berdasarkan data dari Kementerian Komunikasi dan Informatika Republik Indonesia, saat tahun 2017 disebutkan bahwa terdapat kurang lebih 800.000 situs di Indonesia yang telah ditandai sebagai penyebar informasi atau berita palsu. Data lain pada triwulan pertama tahun 2023 yang diadakan oleh Kementerian Komunikasi dan Informatika, ditemukan sebanyak 425 kasus isu hoaks. Hal ini meningkat dari triwulan pertama tahun 2022 yang mencapai 393 isu hoaks. Jumlah keseluruhan sejak bulan Agustus 2018 hingga 31 Maret 2023 terdapat 11.357 isu hoaks. Berdasarkan hasil penelitian masyarakat telematika juga disebutkan bahwa sebanyak 44,3% responden mengatakan menerima berita bohong setiap harinya.

Oleh karena itu, diperlukan adanya mesin deteksi hoaks pada berita kesehatan untuk mengatasi penyebaran berita hoaks kesehatan. Membuat mesin deteksi hoax dapat dilakukan dengan berbagai macam metode, salah satu metode yang dapat digunakan adalah klasifikasi menggunakan *Multinomial Naïve Bayes*. Metode *Multinomial Naïve Bayes* adalah salah satu metode dari tiga jenis metode *naïve bayes*. Putro *et al.* (2020) *Naïve Bayes Classification* memiliki pengertian pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class.

Pada Al-Qur'an surat Al-Baqarah ayat 31 memaparkan mengenai klasifikasi makhluk hidup sebagaimana pada penelitian ini juga dilakukan klasifikasi yang

berupa berita fakta dan berita hoaks. Berikut merupakan bunyi dari Q.S. Al-Baqarah yang terdapat pada ayat 31.

وَعَلَّمَ آدَمَ الْأَسْمَاءَ كُلَّهَا ثُمَّ عَرَضَهُمْ عَلَى الْمَلَائِكَةِ فَقَالَ أَنْبِئُونِي بِأَسْمَاءِ هَؤُلَاءِ إِنْ كُنْتُمْ صَادِقِينَ

“Dan Dia ajarkan kepada Adam nama-nama (benda) semuanya, kemudian Dia perlihatkan kepada para malaikat seraya berfirman, “Sebutkan kepada-Ku nama semua (benda) ini, jika kamu yang benar!”(Q.S Al Baqarah:31).

Tafsir jalalain dari ayat tersebut adalah (Dan diajarkan-Nya kepada Adam nama-nama) maksud dari nama-nama adalah berbagai nama benda (kesemuanya) dengan jalan memasukkan ke dalam kalbunya pengetahuan tentang benda-benda itu (kemudian dikemukakan-Nya mereka) maksudnya benda-benda tadi yang ternyata bukan saja benda-benda mati, tetapi juga makhluk-makhluk berakal, (kepada para malaikat, lalu Allah berfirman) untuk memojokkan mereka, ("Beritahukanlah kepada-Ku) sebutkanlah (nama-nama mereka) yakni nama-nama benda itu (jika kamu memang benar.") bahwa tidak ada yang lebih tahu daripada kamu di antara makhluk-makhluk yang Kuciptakan atau bahwa kamulah yang lebih berhak untuk menjadi khalifah. Sebagai 'jawab syarat' ditunjukkan oleh kalimat sebelumnya.

Multinomial Naïve Bayes, digunakan untuk pengklasifikasian dokumen atau teks yang mana fungsi tersebut tepat untuk diterapkan pada penelitian ini yang menggunakan data berbentuk teks. Pada metode *Multinomial Naïve Bayes*, frekuensi kemunculan kata-kata dalam dokumen dihitung dengan tujuan untuk proses klasifikasi.

1.2 Pernyataan Masalah

Seberapa besar nilai akurasi, presisi, *recall*, dan *f-measure* yang didapatkan dari penerapan metode *Multinomial Naïve Bayes* dalam klasifikasi berita hoaks kesehatan berbahasa Indonesia?

1.3 Tujuan Penelitian

Menghitung keakuratan, presisi, *recall*, dan *f-measure* klasifikasi dari metode naïve bayes pada berita hoaks mengenai kesehatan.

1.4 Manfaat Penelitian

1. Melalui penelitian ini diharapkan masyarakat dapat meninjau kembali isu-isu atau berita mengenai kesehatan yang mudah menyebar agar informasi menyesatkan seperti berita hoaks kesehatan tidak dijadikan patokan.
2. Dengan adanya penelitian ini, diharapkan para pebisnis terutama yang bergerak di bidang kesehatan tidak melakukan kesalahan dalam menilai pasar, mengalami kerugian dan terkena dampak karena mempercayai berita palsu yang tersebar.

1.5 Batasan Masalah

1. Data yang digunakan adalah data berupa berita fakta dan hoaks berbahasa Indonesia dengan topik kesehatan.
2. Data berita fakta yang digunakan berasal dari situs detikcom dan data berita hoaks yang digunakan berasal dari situs turnbackhoax.

1.6 Sistematika Penulisan

Untuk memberikan gambaran umum penelitian ini maka disusun sistematika peneltian sebagai berikut :

- BAB I : Pendahuluan, bab ini menunjukkan latar belakang penelitian, pernyataan masalah, tujuan penelitian, dan manfaat penelitian.
- BAB II : Bab ini membahas studi pustaka mengenai landasan teori dan penelitian terdahulu yang terkait dengan penelitian yang sedang diteliti.
- BAB III : Bab ini berisikan desain penelitian yang membahas alur penelitian, pengumpulan data, dan desain sistem.
- BAB IV : Bab ini menjelaskan mengenai hasil dan pembahasan dari penelitian yang dilakukan.
- BAB V : Bab ini berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya.

BAB II

STUDI PUSTAKA

2.1 Klasifikasi Berita Hoaks

Sebelumnya terdapat beberapa penelitian yang berkaitan dengan klasifikasi berita hoax, beberapa penelitian tersebut yaitu :

Mustofa & Mahfudh (2019) melakukan penelitian mengenai klasifikasi berita hoax dengan menggunakan metode *Naïve Bayes*. Informasi data berita didapatkan secara manual dari situs turnbackhoax.com yang terbit dari bulan November 2018 sampai Februari 2019. Data yang diambil berupa judul, tanggal berita terbit, dan label. Jumlah data yang digunakan dalam penelitian ini adalah 300 data dengan rincian data 297 berita dengan kategori hoax dan 63 berita berkategori fakta. Data yang telah dikumpulkan diolah menggunakan *Tool Weka*. Dokumen yang sudah dilakukan proses klasifikasi dengan metode *naïve bayes*, diuji dengan menggunakan metode *10-Fold Cross Validation*. Penelitian ini menghasilkan nilai akurasi terbaik pada nilai *fold* 6 dengan keakuratan sebesar 85.28 % yang mana terklasifikasi dokumen yang relevan sebanyak 307 dan yang tidak relevan sebanyak 53 atau error rate sebesar 14.72%. Sedangkan nilai rata-rata berdasarkan berita hoax dan berita benar nilai precision 0,896 dan recall 0.853.

Soleman (2021) melakukan penelitian tentang pemanfaatan metode naïve bayes atau tepatnya *gaussian naïve bayes* untuk mendeteksi berita hoaks yang berbahasa Indonesia. Dataset yang digunakan pada penelitian ini berupa 250 artikel dalam bahasa Indonesia yang terdiri dari 10 topik dengan 25 berita pada setiap topiknya. Seleksi fitur pada penelitian ini dilakukan dengan cara menggabungkan algoritma

TF-IDF dengan algoritma *SGD (Stochastic Gradient Descent)*. Cara mengetahui status website yang ada pada dataset aktif atau tidak, penulis mengunjungi satu per satu data set yang didapatkan dari penelitian sebelumnya. Kemudian, artikel akan di-*import* ke dalam *prototype* dengan menggunakan libraries *pandas* di *python*. Hasil akurasi dari penelitian ini menurun jika dibandingkan dari penelitian sebelumnya. Hal ini disebabkan karena penurunan bobot klasifikasi dari website. setiap website penyebar hoax hanya mempublikasi satu topik dan satu berita dan kemudian domain website tersebut dibiarkan expired atau mati. Hasil akurasi terbaik dari percobaan dihasilkan dari data train 80%, data test 20% dan data train 70%, data test 30% dengan nilai akurasi 72%.

Yodi Prayoga *et al.* (2021) melakukan penelitian terkait deteksi hoaks berita online berbahasa inggris. Tujuan dari penelitian tersebut adalah untuk mengetahui performa dari penggunaan algoritma *Bernoulli Naïve Bayes* dengan ekstraksi fitur *TF-IDF* dalam mendeteksi berita hoaks. Metode yang digunakan pada penelitian tersebut adalah *naïve bayes* berjenis *bernoulli naïve bayes* dengan ekstraksi fitur *TF-IDF*. Data yang digunakan pada penelitian bersumber dari kaggle yang dikumpulkan dengan teknik crawling pada berita online yang menggunakan bahasa inggris. Data terdiri dari 23.489 ribu data berita hoaks dan 21.418 ribu data fakta. Pada penelitian ini, model prediksi yang dibangun dengan 8800 data berita, mampu menghasilkan nilai akurasi sebesar 98,5% dari jumlah data uji sebanyak 2.200 data berita, dimana akurasi dari prediksi model untuk label 1 (Fake) sebesar 97,8%, dan akurasi untuk label 0 (True) sebesar 99,1%. Nilai precision menunjukkan berapa

persen berita yang benar Fake dari keseluruhan berita yang diprediksi Fake, nilai precision yang didapatkan sebesar 99,1%.

2.2 Naïve Bayes

Sebelumnya telah terdapat beberapa studi terkait yang menerapkan metode Naïve Bayes, beberapa penelitian tersebut yaitu :

Putro *et al.* (2020) melakukan penelitian menggunakan metode Naïve bayes untuk klasifikasi pelanggan di Sukoharjo. Tujuan dari penelitian dilakukan adalah untuk membuat aplikasi yang dapat membantu mengklasifikasikan pelanggan yang berpotensi dan tidak berpotensi, serta digunakan sebagai dasar pertimbangan keputusan salah satunya sebagai strategi pemasaran. Data yang dikumpulkan melalui metode observasi terhadap prosedur penjualan ayam dan metode wawancara kepada pemilik UD. Samodro untuk mendapatkan data yang akan diteliti untuk menyesuaikan proses data yang akan diolah menggunakan metode Naïve Bayes. Jumlah data latih dan data uji yang digunakan pada penelitian ini masing-masing berjumlah 75 dan 25. Berdasarkan master pelanggan yang dijadikan data latih, yang telah berhasil mengklasifikasikan 23 data dari 25 data yang diuji, didapatkan hasil nilai precision 100%, nilai recall 91%, dan nilai accuracy 92%.

Naomi Chatrina Siregar *et al.* (2020) juga melakukan penelitian dengan metode naïve bayes untuk klasifikasi komentar warga sekolah mengenai pelaksanaan pembelajaran jarak jauh. Tujuan dari penelitian ini adalah menghasilkan suatu sistem berbasis web yang dapat mengklasifikasikan komentar ke dalam 3 kelompok yaitu, positif, negatif, atau netral. Data yang dikumpulkan dari penelitian ini berjumlah 170 data yang terdiri atas 120 data training dan 50 data uji yang

didapatkan dari komentar warga sekolah SDN 060969. Alur dari penelitian dimulai dari melakukan *pre-processing text* yang diawali dengan membaca teks yang akan dilatih lalu mengubah dalam bentuk kata tunggal, kemudian bentuk kata yang sudah diubah akan diproses untuk mendapatkan kata yang penting dengan cara menghilangkan kata yang tidak terdapat pada kamus data yang tersimpan dalam database dan setelah kata didapatkan selanjutnya adalah menghilangkan tanda baca. Kemudian, setelah proses tersebut, dimulailah proses penghitungan pembobotan TF-IDF dan terakhir, metode *naïve bayes* diimplementasikan. Dari penelitian ini didapatkan hasil akurasi dengan nilai 68%.

Rahman *et al.* (2017) melakukan penelitian mengenai klasifikasi berita online menggunakan metode *naïve bayes*. Tujuan dari penelitian adalah melakukan analisis berita online dengan *naïve bayes* berjenis multinomial. Metode *Document Frequency Thresholding* dan pembobotan dengan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* juga digunakan pada penelitiannya demi mendapatkan perbandingan hasil yang terbaik. Dari penelitian ini di dapatkan hasil akurasi akhir sebesar 94,29%. Kesimpulan yang didapatkan pada penelitian ini adalah penggunaan fitur seleksi dengan metode DF-Thresholding dapat mengurangi tingginya dimensi data. Hal ini ditunjukkan dengan pengurangan fitur dari jumlah awal 22.215 menjadi 16.766 dengan hasil akurasi 93,33%. Dengan menggunakan TF-IDF pada multinomial *naïve bayes* dapat meningkatkan akurasi. Nilai rata-rata akurasi yang awalnya hanya 86,28% menjadi 86,62%. Sedangkan dengan menggunakan *DF-Threshold* pada multinomial *naïve bayes* dengan TF-IDF mendapatkan hasil akurasi yang lebih rendah daripada tanpa *DF-Threshold* yakni,

sebesar 92,38 namun jumlah fitur yang digunakan lebih sedikit dalam proses klasifikasi.

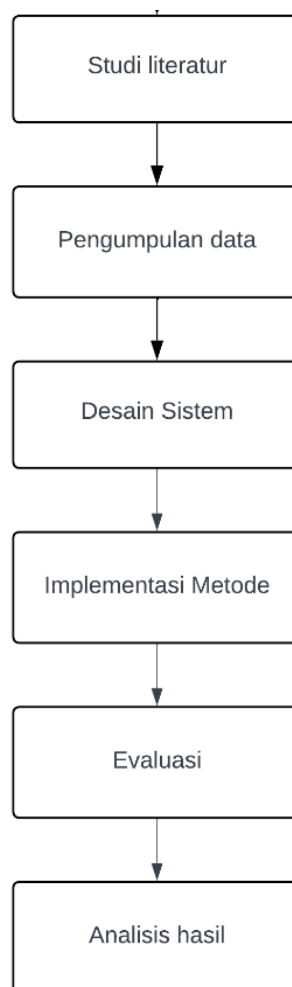
Setianingrum *et al.* (2018) juga melakukan penelitian menggunakan naïve bayes dengan jenis multinomial naïve bayes. Penelitian yang dilakukan mengenai sistem klasifikasi surat keluar studi kasus DISKOMINFO Kabupaten Tangerang. Sistem pengelolaan teks yang digunakan adalah metode *text mining*. Sistem klasifikasi pada penelitian ini didukung dengan confix-stripping stemmer untuk menemukan kata dasar dan TF-IDF untuk pembobotan kata. Penelitian ini juga menggunakan pendekatan *Rapid Application Development* (RAD) sebagai metode pengembangan sistem. Sistem klasifikasi surat yang dibentuk berbasis dekstop dengan bahasa pemrograman Java dan basis data mySQL dengan pembatasan data input yaitu berupa surat keluar dengan format word berekstensi (.docx), (.doc), dan (.txt). Sistem klasifikasi surat keluar didukung dengan penerapan Confix-Stripping Stemmer pada proses stemming dan TF-IDF untuk pembobotan setiap kata yang terkandung pada dokumen (term weighting). Dari penelitian ini didapatkan hasil akurasi, presisi, *recall*, dan *f-measure* masing-masing secara berurutan sebesar 89,58%, 79,17%, 78,72%, dan 77,05%.

BAB III

DESAIN PENELITIAN

3.1 Alur Penelitian

Alur penelitian memberikan gambaran tentang tahapan dari penelitian secara umum. Alur penelitian terkait deteksi hoax pada berita kesehatan berbahasa Indonesia dapat dilihat pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

Pada Gambar 3.1 memberikan penjelasan mengenai tahapan dalam penelitian untuk deteksi berita hoax kesehatan berbahasa Indonesia dengan metode *naïve*

bayes. Tahap pertama adalah melakukan studi literatur dengan cara mencari referensi dari berbagai sumber untuk bahan penelitian terkait. Berikutnya adalah melakukan pengumpulan data yang ada pada website detikcom dan turnbackhoax. Tahap ketiga adalah membuat desain sistem, yaitu penjelasan mengenai proses pengolahan data. Setelah membuat desain sistem, data yang sudah dikumpulkan, diimplementasikan ke dalam metode multinomial *naïve bayes*. Tahap kelima adalah melakukan evaluasi dengan cara melakukan pengujian pada sistem untuk mengetahui apakah sistem sudah berjalan sesuai dengan semestinya. Tahap terakhir adalah melakukan analisis hasil dengan cara menghitung nilai akurasi, presisi, recall, dan *f-measure* dengan menggunakan *confussion matrix*.

3.2 Pengumpulan Data

Data yang digunakan adalah data berita kesehatan berbahasa Indonesia yang diambil dari situs detikcom dan turnbackhoax. Kedua situs ini dipilih karena detikcom sudah terkenal dengan kredibilitasnya yang mengedepankan fakta, sedangkan turnbackhoax merupakan situs tempat berita hoax diklarifikasi berdasarkan sumber terpercaya seperti jurnal penelitian atau website terpercaya. Jumlah data yang diambil adalah sebanyak 504 data berita dengan 252 berita fakta dan 252 berita hoax. Dari 504 data berita, diambil data training sebanyak 201 fakta dan 200 hoaks serta data testing sebanyak 51 fakta dan 52 hoax. Tabel 3.1 menunjukkan contoh data berita kesehatan fakta yang berhasil dikumpulkan dari website detikcom dan Tabel 3.2 menunjukkan contoh data hoax yang berhasil dikumpulkan dari website turnbackhoax. Kategori yang diambil adalah dengan tag

kesehatan tubuh. Berikut merupakan implementasi proses *web scrapping* ke dalam *source code*.

```
def scrape_detik(hal):
    global hades
    a = 1
    for page in range(1,hal):
        url =
f'https://www.detik.com/search/searchnews?query=kesehatan+tubuh&sortby
=time&page={page}'
        ge = req.get(url,hades).text
        sop = bs(ge,'lxml')
        li = sop.find('div',class_='list media_rows list-berita')
        lin = li.find_all('article')
        for x in lin:
            link = x.find('a')['href']
            date =
x.find('a').find('span',class_='date').text.replace('WIB','').replace(
'detikNews','').split(',')[1]
            headline = x.find('a').find('h2').text
            ge_ = req.get(link,hades).text
            sop_ = bs(ge_,'lxml')
            content = sop_.find_all('div',class_='detail__body-text
itp_bodycontent')
            for x in content:
                x = x.find_all('p')
                y = [y.text for y in x ]
                content_ = ''.join(y).replace('\n',
'').replace('ADVERTISEMENT','').replace('SCROLL TO RESUME CONTENT','')
                print(f'done[{a}] > {headline[0:10]}')
                a += 1
            with open('kesehatan.csv','a') as file:
                wr = csv.writer(file, delimiter=',')
                wr.writerow([headline,date,link,content_])
```

Tabel 3. 1 Contoh Data Berita Fakta dari website detikcom

Title	Time stamp	link	Article	label
9 Makanan Rendah Kalori yang Cocok Dikonsumsi Saat Diet	23 Mei 2023 07:00	https://food.detik.com/info-sehat/d-6732871/9-makanan-rendah-kalori-yang-cocok-dikonsumsi-saat-diet	Makanan rendah kalori cocok dikonsumsi untuk kamu yang sedang menjalani program diet. Beberapa diantaranya berasal dari buah-buahan dan sayur-sayuran. Meskipun rendah kalori, jangan lupa untuk mengecek nutrisinya. Apakah makanan tersebut dapat memenuhi kebutuhan gizi dalam tubuh...	fakta

Lanjutan Tabel Contoh Data Berita Fakta

Title	Time stamp	link	Article	label
Batasi Konsumsinya! 5 Sayuran Ini Punya Efek Samping Jika Dimakan Berlebihan	23 Mei 2023 06:00	https://food.detik.com/info-sehat/d-6732824/batasi-konsumsi-5-sayuran-ini-punya-efek-samping-jika-dimakan-berlebihan	Sayuran merupakan makanan terbaik untuk kesehatan. Namun, ada beberapa sayuran ini sebaiknya jangan dikonsumsi berlebihan karena punya efek samping. Sayuran adalah makanan yang kaya akan nutrisi. Mulai dari vitamin, mineral, hingga serat. Mengonsumsi sayur bahkan selalu dianjurkan oleh ahli gizi karena banyak manfaatnya. Mulai dari memperlancar sistem pencernaan, menghindari dari berbagai penyakit tertentu, hingga mempermudah proses penurunan berat badan.....	fakta
Cek Gula Darah: Cara, Penyebab, dan Mengapa Harus Melakukannya?	23 Mei 2023 05:15	https://health.detik.com/berita-detikhealth/d-6732795/cek-gula-darah-cara-penyebab-dan-mengapa-harus-melakukannya	Pada penderita diabetes, cek gula darah dilakukan untuk memantau apakah kadar gula darah tetap terkendali atau tidak. Namun, setiap orang juga dianjurkan untuk melakukan cek gula darah untuk mengetahui kondisi glukosa pada tubuh. Pasalnya, jika kadar glukosa rendah, kita dapat kehilangan kemampuan untuk berpikir dan berfungsi secara normal...	fakta
10 Manfaat Pisang untuk Tubuh yang Jarang Diketahui Orang	23 Mei 2023 05:00	https://food.detik.com/info-sehat/d-6732861/10-manfaat-pisang-untuk-tubuh-yang-jarang-diketahui-orang	Pisang merupakan salah satu jenis buah yang mengandung banyak zat bergizi. Maka dari itu, disarankan untuk mengonsumsi pisang setiap hari demi menjaga kesehatan tubuh. Manfaat pisang yang paling umum diketahui masyarakat adalah dapat melancarkan sistem pencernaan...	fakta

Tabel 3. 2 Contoh data berita hoax dari website turnbackhoax

Title	Time stamp	link	Article	label
[SALAH] “JIKA ANDA DIGIGIT ULAR COBRA JANGAN PANIK, TUMBUKAN KUNYIT, SELEDRI DAN MINYAK KAYU PUTIH OLESKAN KEBAGIAN YANG DIGIGIT DAN DIBUNGKUS SELURUH TUBUH DENGAN KAIN KAFAN”	Januari 22, 2023	https://turnbackhoax.id/2023/01/22/salah-jika-anda-digigit-ular-cobra-jangan-panik-tumbukan-kunyit-seledri-dan-minyak-kayu-putih-oleskan-kebagian-yang-digigit-dan-dibungkus-seluruh-tubuh-dengan-kain-kafan/	GIGITAN ULAR DAPAT DIATASI DENGAN TUMBUKAN KUNYIT, SELEDRI DAN MINYAK KAYU PUTIH DAN DIBUNGKUS DENGAN KAIN KAFAN	Hoax
[SALAH] Vaksin covid-19 fungsinya untuk mengontrol manusia bukan untuk kesehatan	Juli 19, 2022	https://turnbackhoax.id/2022/07/19/salah-vaksin-covid-19-fungsinya-untuk-mengontrol-manusia-bukan-untuk-kesehatan/	Pendeta Lukas Sutrisno. Pemasangan microchip pada tubuh manusia, microchip yang dimasukkan kedalam tubuh manusia yang disuntikan melalui Vaksin covid-19 fungsinya untuk mengontrol manusia bukan untuk kesehatan.	Hoax
[SALAH] Butiran Es Hasil Fenomena Hujan Es Dapat Meningkatkan Imunitas Tubuh	Februari 28, 2022	https://turnbackhoax.id/2022/02/28/salah-butiran-es-hasil-fenomena-hujan-es-dapat-meningkatkan-imunitas-tubuh/	Fenomena Hujan ES! Dulu ketika masih kanak2 terjadi hujan es,ibu membiarkan kami behujan2n dihalaman sambil sesekali kami makan esnya dan sampe sekarang beliau masih msh mengingatkan beliau tidak melarang, Ternyata banyak hikmahnya! 1.imun tubuh 2.bersyukur 3... ‘Allahumma shayyiban nafi’an.’...	hoax

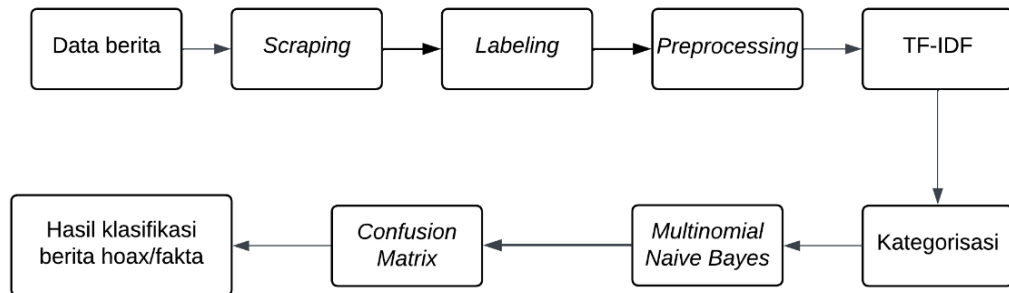
Lanjutan Tabel Contoh Data Berita Hoax

Title	Time stamp	link	Article	label
[SALAH] Garam dan Air Kelapa dapat Hilangkan Vaksin di Dalam Tubuh	Februari 13, 2022	https://turnbackhoax.id/2022/02/13/salah-garam-dan-air-kelapa-dapat-hilangkan-vaksin-di-dalam-tubuh/	Jika anda ragu dan ingin menghilangkan vaksin yg sdh masuk ke tubuh ALHAMDULILLAH. Sekarang ada penangkal/ menetralisir vaksin itu. SILAHKAN DICoba RAMUAN INI YAKIN LAH,,SYARAT. 1.BASMALLAH,. 2.BACA SHOLAWAT NABI, 3.BACA ALFATIHAH INI RAMUAN NYA* “GARAM LAUT” DAN AIR KELAPA MUDA tapi bukan Garam Dapur. *Biasa disebut Garam Krosok (karungan). *- 1/2 Sendok makan Garam* – SeGelas Air Kelapa Muda.kemudian keduanya diaduk. Tunggu sampai mengendap, baru diminum. *2 X Sehari (pagi + malam), selama seminggu *Ditambah “Bekam” lebih bagus lagi *IN SYAA ALLAH... VAKSIN yang bersarang itu lenyap,	hoax

3.3 Desain Sistem

Pada desain sistem berisikan gambaran kerja sistem yang akan dirancang, desain sistem yang akan dirancang berisikan informasi mengenai proses pengolahan data berita hoaks kesehatan berbahasa Indonesia menggunakan metode multinomial *Naïve Bayes*. Dari pengolahan data ini, kemudian diimplementasikan ke dalam kode program dengan menggunakan bahasa pemrograman *Python* dan menghasilkan output *user interface* dalam bentuk *website*.

Berikut Gambar 3.2 yang menjelaskan desain sistem dari deteksi berita hoax pada kesehatan.



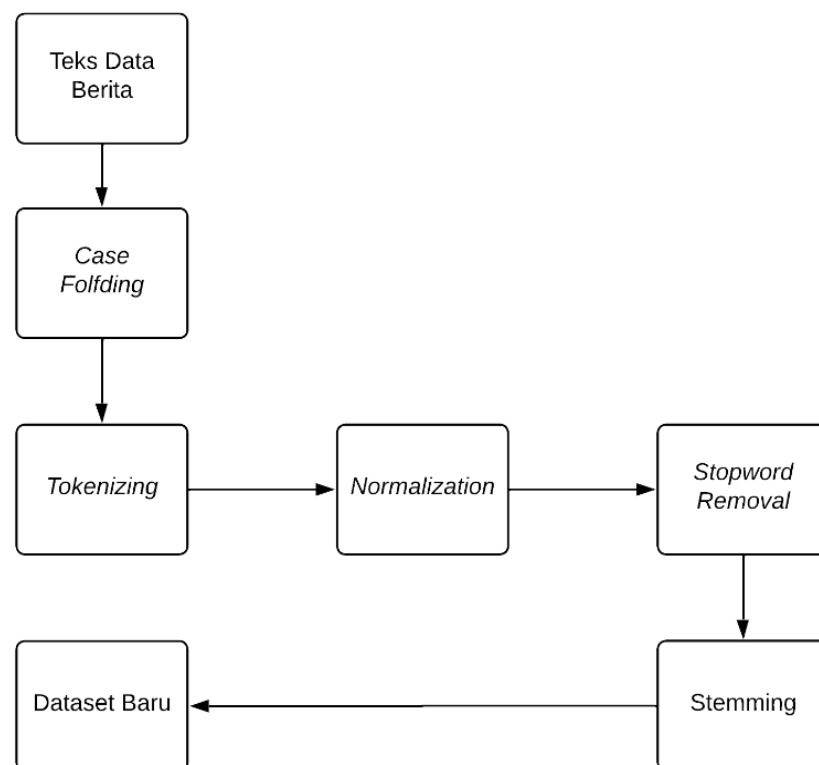
Gambar 3.2 Desain Sistem Deteksi Berita Hoax

Dari Gambar 3.2 dapat dijelaskan terdapat tahap *scraping* yaitu hasil dari klasifikasi berita selanjutnya dilakukan pengujian. Data yang telah dikumpulkan dengan cara *scraping*, kemudian diberi label. Label tersebut berupa hoax atau fakta. Tahap selanjutnya, setelah seluruh data terlabeli, dilakukan proses *preprocessing* data. Tahap berikutnya, dilakukan split data. Data dibagi menjadi data training dan data testing. Setelah itu, proses pembobotan menggunakan TF-IDF dilakukan. Kemudian, dilakukan skenario pengujian dengan cara melakukan pengujian terhadap data yang sudah mengimplementasikan metode *naïve bayes* yang bertujuan untuk menghitung nilai akurasi, presisi, *recall*, dan *f-measure*. Cara untuk menghitung nilai tersebut adalah dengan menggunakan *Confusion Matrix*.

3.4 Data Preprocessing

Data preprocessing merupakan langkah penting dalam melatih model, oleh karena itu artikel berita yang ada dalam *dataset* telah diproses sebelumnya. *Preprocessing* merupakan proses awal yang akan mengubah data masukan menjadi data dengan format yang sesuai dan siap untuk diproses. Berikut merupakan

beberapa contoh proses yang dilakukan dalam *preprocessing*, yakni meliputi berbagai proses yang diperlukan antara lain : penggabungan, pengubahan bentuk, pengubahan data sebagai cara untuk membersihkan, mengintegrasikan, mereduksi, dan mendiskritisasi. Lebih lanjut proses yang ada dalam *preprocessing* dapat terdiri dari salah satu kegiatan proses ataupun gabungan dari beberapa proses di atas (Munirul et al., 2020). Proses yang ada tergantung dari tujuan yang akan dicapai dalam *preprocessing*. Performansi klasifikasi akan meningkat tergantung pada pemilihan proses yang tepat dan sesuai dalam tahapan *preprocessing* data. Gambar 3.3 merupakan diagram alur dari proses *preprocessing*.



Gambar 3. 3 Diagram alur *Preprocessing* (Dewantara, 2022)

Berikut merupakan langkah-langkah yang dilakukan dalam melakukan *preprocessing* data:

3.4.1 *Cleaning*

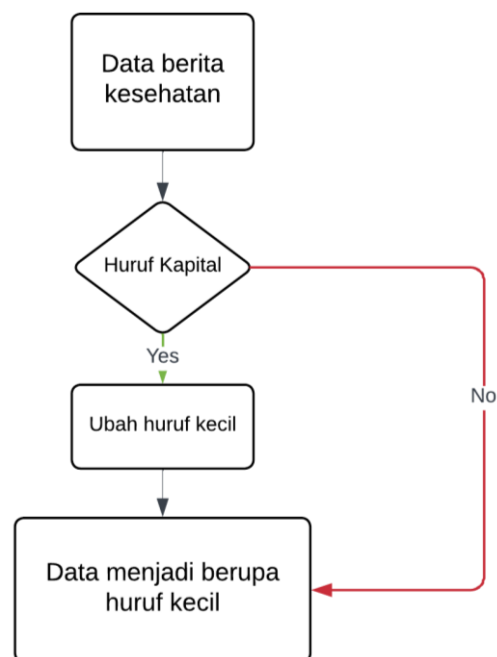
Pada tahap *cleaning* terdapat tahap *case folding* dan menghapus angka, simbol, serta emoji. *Case folding* dibutuhkan untuk menyamakan bentuk pada kata, contohnya seperti yang dilakukan pada data narasi yang digunakan pada penelitian ini yaitu mengubah semua kata menjadi huruf kecil atau *lowercase*. Tujuan dari *case folding* ini adalah untuk mengurangi redudansi data yang nantinya digunakan dalam proses pengklasifikasian, jadi proses perhitungan menjadi optimal. Berita yang ditulis terdapat banyak angka, emoji, dan tanda baca namun ketiganya tidak penting untuk diolah, sehingga perlu dihapus untuk mempermudah dalam pengolahan data.

Berikut merupakan *source code* dari *cleaning* dan diagram alur dari *cleaning*. Tabel 3.3 menunjukkan hasil dari proses *cleaning*.

```
def clean_text(text):
    text=re.sub('@[^s]+', ' ', text)
    text=re.sub(r"(?:\@|http?:\/\/|https?:\/|www)\S+", "", text)
    # text=text.translate(str.maketrans(' ',' ',string.punctuation))
    text=re.sub('<.*?>', ' ', text)
    text=re.sub('[^a-zA-Z]', ' ', text)
    rext=re.sub("\n", " ",text)
    text=text.lower()
    text=re.sub(r"\b[a-zA-Z]\b", " ", text)
    text=' '.join(text.split())
    return text
```

Tabel 3. 3 Hasil *Cleaning*

article	hoax	clean
Seminggu ber-uap: Menurut dokter, Covid -19...	1	seminggu ber uap menurut dokter covid...
Ada makhluk hidup di dalam vaksin. Itu kematian...	1	ada makhluk hidup di dalam vaksin itu kematian...
Meminum air dalam jumlah banyak dapat...	1	meminum air dalam jumlah banyak dapat ...
Tahukah Kamu! Meminum terlalu banyak air dapat...	1	tahukah kamu meminum terlalu banyak air dapat...
Ternyata Biaya utk Cuci Ginjal kurang dari : ...	1	ternyata biaya utk cuci ginjal kurang dari...

Gambar 3. 4 Diagram alur *Case Folding* (Dewantara, 2022)

3.4.2 Tokenizing

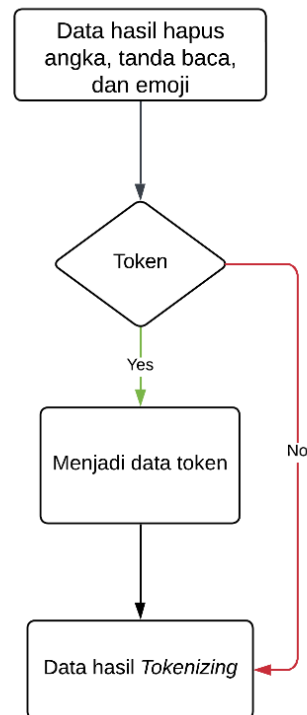
Tokenizing bertujuan untuk membagi kalimat menjadi beberapa bagian seperti kata-kata, frasa atau elemen yang memiliki makna lain. Selain itu

tokenizing merupakan proses membagi atau memotong sebuah kalimat menjadi kata-kata dengan menggunakan spasi, sebagai contoh: "Hoaks merupakan berita palsu yang tidak sesuai dengan fakta", akan dipecah menjadi: "Hoaks", "merupakan", "berita", "palsu", "yang", "tidak", "sesuai", "dengan", "fakta". Teks yang telah dibersihkan dan diseragamkan hurufnya berupa *lowercase*, diolah kembali untuk mendapatkan token per kata. Berikut merupakan *source code* dari *tokenizing* dan hasil dari *tokenizing* yang ditunjukkan pada Tabel 3.4. Sedangkan untuk diagram alur *tokenizing* terdapat pada Gambar 3.5.

```
from nltk.tokenize import RegexpTokenizer
regexp = RegexpTokenizer(r'\w+|$\d+|\S+')
df['token'] = df['clean'].apply(regexp.tokenize)
```

Tabel 3. 4 Hasil *Tokenizing*

article	hoax	token
Seminggu ber-uap: Menurut dokter, Covid -19...	1	['seminggu', 'ber', 'uap', 'menurut', 'dokter', 'covid', ...]
Ada makhluk hidup di dalam vaksin. Itu kematian...	1	['ada', 'makhluk', 'hidup', 'di', 'dalam', 'vaksin', 'itu', 'kematian', ...]
Meminum air dalam jumlah banyak dapat...	1	['meminum', 'air', 'dalam', 'jumlah', 'banyak', 'dapat', ...]
Tahukah Kamu! Meminum terlalu banyak air dapat...	1	['tahukah', 'kamu', 'meminum', 'terlalu', 'banyak', 'air', 'dapat', ...]
Ternyata Biaya utk Cuci Ginjal kurang dari : ...	1	['ternyata', 'biaya', 'utk', 'cuci', 'ginjal', 'kurang', 'dari', ...]



Gambar 3. 5 Diagram Alur *Tokenizing* (Dewantara, 2022)

3.4.3 *Normalization*

Tahap *normalization* merupakan tahap dimana seluruh kata yang memiliki kesalahan dalam hal ejaan, noise, maupun kata yang semula tidak baku menjadi baku. Kata yang mengandung noise memiliki pengertian kata yang memiliki unsur-unsur bahasa daerah atau bahasa gaul yang tidak sesuai dengan Kamus Besar Bahasa Indonesia atau yang biasa disingkat dengan sebutan KBBI.


```
def replace_with_normalization(token_list):
    new_tokens = []
    for token in token_list:
        if token in kn['WORD'].values:
            # Dapatkan indeks di mana token cocok
            index = kn[kn['WORD'] == token].index[0]
            # Ambil nilai normalisasi dari kolom 'di'
            normalization = kn.loc[index, 'NORMALIZATION']
            new_tokens.append(normalization)
        else:
            new_tokens.append(token)
    return new_tokens
df['normalisasi'] =
df['token'].apply(replace_with_normalization)
```

3.4.4 *Stopword*

Stopword merupakan kata-kata umum yang sering muncul yang tidak bermakna atau tidak mempunyai pengaruh yang signifikan dalam kalimat, sehingga penghapusan *stopword* diperlukan dimana penulis melakukan proses. Kata-kata yang tidak unik ataupun sering muncul tetapi tidak signifikan dan tidak memiliki pengaruh akan dihapus dalam tahap *Stopwords removal*. Preposisi, konjungsi, kata keterangan, dan juga kata substitusi merupakan kata yang dimaksudkan, termasuk: "yang", "ke", "di", "sebuah", "pada", "oleh", "ini", "dari", dsb. Berikut merupakan *source code* dari tahap *stopword removal* beserta beberapa list kata yang akan di hapus.

```

from Sastrawi.StopWordRemover.StopWordRemoverFactory
import StopWordRemoverFactory

factory = StopWordRemoverFactory()

stopwords = factory.get_stop_words()

```

```

['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut',
'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika',
'sehingga', 'kembali', 'dan', 'tidak', 'ini', 'karena',
'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah',
'belum', 'kami', 'sekitar', 'bagi', 'serta', 'di', 'dari',
'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah',
'itu', 'dalam', 'bisa', 'bahwa', 'atau', 'hanya', 'kita',
'dengan', 'akan',...]

```

```

def stopwords_(tokens):

    cleaned_tokens = []

    for token in tokens:

        if token not in stopwords:

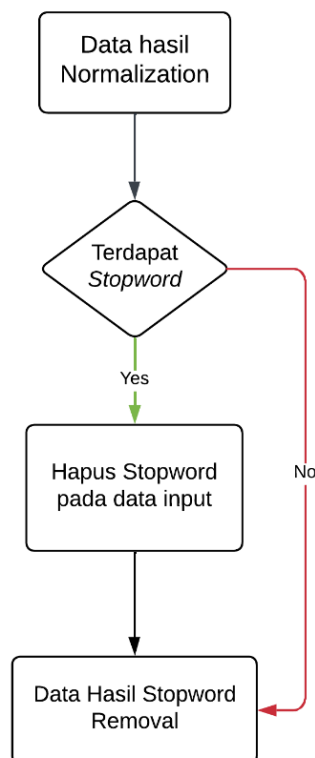
            cleaned_tokens.append(token)

    return cleaned_tokens

df['stopwords'] = df['normalisasi'].apply(stopwords_)

```

Gambar 3.6 menunjukkan diagram alur dari tahap *stopword removal* yang digunakan dalam penelitian.



Gambar 3. 6 Diagram alur *Stopword Removal* (Dewantara, 2022)

3.4.5 *Stemming*

Stemming merupakan suatu teknik untuk mentransformasi kata-kata dalam sebuah dokumen teks menjadi bentuk kata dasar. Setiap bahasa memiliki proses *stemming* yang berbeda-beda. Karena setiap bahasa mempunyai aturan kebahasaan yang berbeda-beda terutama dalam penggunaan kata yang memiliki imbuhan. Contohnya, Bahasa Perancis memiliki perbedaan aturan penggunaan tata bahasa dengan Bahasa Arab. Pada Bahasa Indonesia kompleksitas ada pada variasi imbuhan. Hal tersebut menjadi penting dalam pembentukan kata dasarnya. Berikut merupakan *code* untuk *stemming*.

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

factory = StemmerFactory()
Stemmer = factory.create_stemmer()

def stemmed_wrapper(term):
    return Stemmer.stem(term)

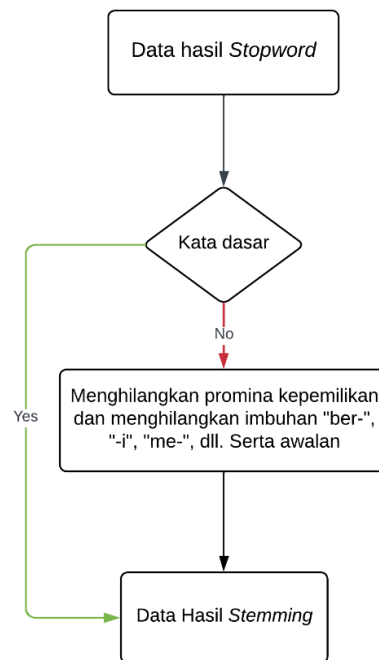
term_dict={}

for doc in df['stopwords']:
    for term in doc:
        if term not in term_dict :
            term_dict[term] = ' '
    print(len(term_dict))

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term, ":", term_dict[term])
    print(len(term_dict))

```

Diagram alur *stemming* yang ditunjukkan pada Gambar 3.7. Pada diagram tersebut menunjukkan proses *stemming* yang dilakukan pada penelitian.



Gambar 3. 7 Diagram alur *Stemming* (Dewantara, 2022)

3.5. *Term Frequenxy-Inverse Document Frequency (TF-IDF)*

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode statistic yang digunakan untuk mengukur seberapa penting suatu kata terhadap suatu dokumen tertentu dari sekumpulan dokumen atau corpus. *Term Frequency* memiliki pengertian frekuensi dari kemunculan sebuah term yang ada di dalam dokumen. Term frequency ini berpengaruh pada besar bobot. Semakin besar jumlah munculnya suatu term, maka semakin besar pula bobot yang ada dan akan memberikan nilai kesesuaian yang semakin besar. Sedangkan *Inverse Document Frequency* memiliki pengertian sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. IDF menunjukkan hubungan ketersediaan sebuah term dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung term yang dimaksud, maka

nilai IDF semakin besar. Tujuan dari IDF adalah untuk menilai sejauh mana sebuah kata kunci dapat membedakan atau menjadi relevan dalam suatu dokumen dibandingkan dengan dokumen-dokumen lain dalam koleksi. Berikut merupakan persamaan untuk perhitungan bobot menggunakan TF, IDF, dan TF-IDF.

$$TF = \frac{\text{jumlah kemunculan kata dalam dokumen}}{\text{jumlah kata pada dokumen}} \quad (3.1)$$

$$IDF = \log \frac{|D|}{|DF|} \quad (3.2)$$

$$TF-IDF = TF \times IDF \quad (3.3)$$

Dengan keterangan sebagai berikut.

TF-IDF = Nilai bobot kata pada sebuah dokumen

TF = Frekuensi pengulangan kata dalam dokumen

IDF = Frekuensi data kata

D = Total dokumen

DF = Kuantitas dokumen yang mengandung kata tersebut

Pada penelitian ini, pembobotan kata dilakukan dengan pendekatan TF-IDF.

Dilakukan pembobotan kata yang ada pada data hoaks dan fakta. Selanjutnya, sebuah array dibuat menggunakan hasil pembobotan. Pada studi ini, perhitungan manual akan menerapkan 5 sampel data yang disajikan pada Tabel 3.5.

Tabel 3. 5 Tabel Data Sampel

Data Berita	Artikel	Pelabelan Manual
DB1	Putri (2023) Konsumsi sayur dan buah masyarakat Indonesia dinilai masih rendah. Diketahui berdasarkan data Riset Kesehatan Dasar (Riskesdas) 2018 hanya 4,5% warga RI yang cukup makan sayur dan buah setiap hari. Meski terdengar sepele, asupan serat ternyata memiliki peranan penting bagi kesehatan tubuh. Bahkan, WHO menganjurkan untuk mengonsumsi serat sebanyak minimal 25 gram per hari. Kekurangan asupan serat dapat menurunkan imunitas tubuh hingga meningkatkan risiko penyakit kardiovaskular. Hal ini karena serat merupakan tipe karbohidrat yang bermanfaat untuk pencernaan...	Fakta

Data Berita	Artikel	Pelabelan Manual
DB2	Kharisma (2023) Menjaga kesehatan otot, tulang dan jantung perlu dilakukan agar tubuh senantiasa terhindar dari risiko penyakit. Selain dengan menjalankan pola gaya hidup dan makanan sehat, menjaga kesehatan otot, tulang dan jantung juga bisa dengan melakukan olahraga kardio. Dilansir dari Mayo Clinic, World Health Organization (WHO) menganjurkan orang dewasa untuk melakukan olahraga kardio atau aerobik setidaknya 150 menit seminggu. Pasalnya, olahraga ini memberikan banyak manfaat bagi kesehatan, khususnya untuk otot, tulang dan jantung. Penasaran apa saja olahraga kardio yang baik untuk menjaga kesehatan otot, tulang dan jantung? Dilansir dari berbagai sumber, berikut ulasannya... Selain itu, berenang juga dapat mengoptimalkan kapasitas paru-paru, meningkatkan kekuatan otot, serta meningkatkan kepadatan tulang. Meski senam yoga bukan merupakan olahraga spesifik kardio, jika dikombinasikan dengan berbagai gerakan atletik olahraga ini bisa membawa banyak manfaat bagi kardiovaskular... Agar manfaat olahraga kardio makin lebih optimal, Anda bisa mengonsumsi asupan nutrisi tambahan seperti Entrasol Platinum. Entrasol Platinum mengandung nutrisi tinggi kalsium dan ekstrak buah zaitun yang bersinergi menjaga kekuatan otot, tulang dan jantung. Minum Entrasol Gold 2x sehari agar otot, tulang, dan jantung selalu sehat.	Fakta
DB3	VAKSIN ITU ARTINYA ZAT YG SENGAJA DIBUAT UNTUK MEMBUAT KEKEBALAN TUBUH TERHADAP PENYAKIT TERTENTU. JADI, VAKSIN COVID19 YG DISUNTIKKAN KE DALAM TUBUH ANDA TUJUANNYA AGAR TUBUH ANDA KEBAL TERHADAP VIRUS COVID19. JIKA ANDA SUDAH DISUNTIK VAKSIN COVID19, MAKA ANDA TIDAK PERLU LAGI PAKEK MASKER, CUCI TANGAN DAN SEBAGAINYA KARENA ANDA SUDAH KEBAL TERHADAP VIRUS COVID19. TAPI, JIKA ANDA SUDAH DI SUNTIK VAKSIN COVID19, TAPI MASIH SAJA DISURUH PAKEK MASKER, DISURUH CUCI TANGAN, DUDUK BERJAUHAN DAN SEBAGAINYA, BERARTI YG DISUNTIKKAN...	Hoaks
DB4	Para ayah dan bunda hati2 memberi minuman instan untuk buah hati anda karena minuman tersebut sudah terbukti mengandung zat berbahaya siklamat yang menyebabkan penyakit lupus yang dapat merusak antibodi. Efeknya 1. dapat mengalami gangguan otak 2. penyebab kanker. Netralkan dan keluarkan zat2 berbahaya yang terlanjur masuk ke dalam tubuh dengan minum air Milagros.. #Milagrosmiracleinside	Hoaks
DB5	Tahukah Kamu! Meminum terlalu banyak air dapat membuat kerusakan pada kulitmu. Meskipun tetap terhidrasi penting bagi kesehatan tubuh secara keseluruhan, jumlah air yang banyak dapat membuat tubuhmu kehilangan kandungan mineral yang penting.	Hoaks

Terdapat 4 tahap yang dilakukan dalam pembobotan kata dengan metode tf-idf. Berikut adalah tahapan-tahapan tersebut:

1) *Term Frequency*

Perhitungan jumlah huruf di dalam sebuah kata merupakan proses dalam *term frequency*. Perhitungan pada *term frequency* dilakukan sama seperti pada persamaan 3.1. Nilai TF ini didapatkan dengan cara membagi jumlah kata “x” yang muncul dalam dokumen berita “db” dengan total kata yang ada dalam dokumen berita “db”. Contoh perhitungan manual dari TF terdapat pada Tabel 3.6 di bawah ini.

Tabel 3. 6 Perhitungan TF

Dokumen Berita	Jumlah Kata	Kata	Frekuensi	TF
DB1	309	minum	3	3/309
DB2	301		1	1/301
DB3	55		0	0/55
DB4	39		3	3/39
DB5	25		1	1/25

2) *Document Frequency*

Pada Tabel 3.6 dipaparkan proses dari menghitung TF (*Term Frequency*). Berikutnya adalah proses dihitungnya banyak dokumen yang ada pada term *ke-t* disebut dengan *document frequency*. Berdasarkan Tabel 3.6 banyak *document frequency* adalah 4. Karena dari 5 data, 4 data diantaranya mengandung kata “minum”.

3) *Inverse Document Frequency*

Inverse Document Frequency merupakan proses penentuan nilai berbanding terbalik atau dapat juga disebut dengan *inverse* dalam *document frequency* dengan menggunakan persamaan 3.2, berikut merupakan contoh perhitungan *IDF* untuk kata atau *term* “minum” pada DB1.

$$\begin{aligned} IDF &= \log \frac{|D|}{|DF|} \\ &= \log \frac{306}{4} \\ &= 1,8836 \end{aligned}$$

4) *TF-IDF*

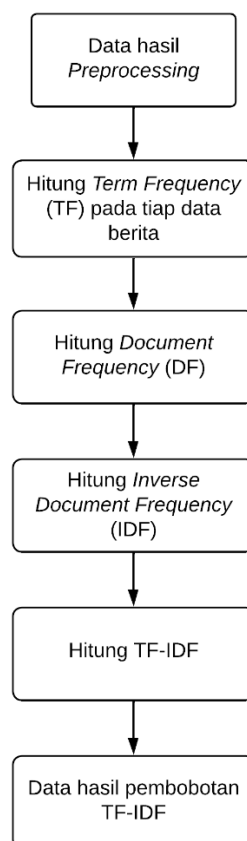
TF-IDF adalah proses untuk memperoleh nilai terhadap dokumen dengan menggunakan persamaan 3.3. Berdasarkan hasil TF dan IDF dari DB1 pada kata atau *term* “minum” maka, hasil TF-IDF adalah sebagai berikut.

$$\begin{aligned} TF-IDF &= TF \times IDF \\ &= 0.0097 \times 1,8836 \\ &= 0.01827092 \end{aligned}$$

Perhitungan IDF dan TF-IDF untuk seluruh data sampel terhadap kata atau *term* “minum” terdapat pada Tabel 3.7. Sedangkan untuk diagram alur dari proses TF-IDF dapat dilihat pada Gambar 3.8.

Tabel 3. 7 Perhitungan IDF dan TF-IDF

Dokumen Berita	kata	IDF	TF-IDF
DB1	minum	1,8836	0.01827092
DB2		1.8750	0.0061875
DB3		1.1383	0
DB4		0.9542	0.07337798
DB5		0.7781	0.031124



Gambar 3. 8 Diagram alur TF-IDF

3.6 Naïve Bayes

Naive Bayes Classifier merupakan algoritma yang digunakan dalam proses klasifikasi yang banyak digunakan dalam pengklasifikasian dokumen, khususnya

pada kasus klasifikasi teks. Metode *Naive Bayes Classifier* adalah salah satu metode yang digunakan untuk pengklasifikasian dengan dasar teorema *Bayes* yang mengasumsikan data bersifat independen dan memiliki arti bahwa data tidak saling terkait satu sama lain. Sifat independen dari teorema *bayes* membuat algoritma *Naive Bayes Classifier* cukup membutuhkan jumlah data *training* yang kecil pada proses klasifikasi. Pada algoritma *naïve bayes*, terdapat tiga tipe, yakni, *Gaussian*, *Multinomial*, dan *Bernoulli*.

Multinomial Naïve Bayes merupakan algoritma yang dapat membantu dalam hal mengetahui kategori suatu dokumen, apakah termasuk dokumen penting, dokumen yang dapat dialihkan, atau dokumen sampah yang berbahaya. Selain itu, bisa juga digunakan untuk mengelompokkan dokumen berdasarkan tema tertentu seperti olahraga, kesehatan, gaya hidup, sosial-politik, dan lain sebagainya.

Fitur yang digunakan oleh classifier adalah frekuensi kata yang ada dalam dokumen. Misalnya, jika suatu dokumen terus menerus menampilkan kata “sepak bola”, “liga”, “pertandingan”, maka bisa dimasukkan dalam kategori olahraga.

Karena data yang digunakan pada penelitian ini adalah data yang berbentuk dokumen teks dan berfokus pada banyaknya frekuensi kata yang muncul pada dokumen, maka metode *Multinomial Naïve bayes* digunakan pada penelitian ini sebagai fokusnya. Persamaan 3.4 merupakan persamaan probabilitas posterior kelas C_k jika mengacu pada data W

$$p(C_k | W) = \frac{p(C_k) \times p(W | C_k)}{p(W)} \quad (3.4)$$

Perhitungan posterior $p(C_k)$ dapat dilihat pada Persamaan 3.5, dengan N adalah jumlah seluruh data.

$$p(C_k) = \frac{\sum C_k}{N} \quad (3.5)$$

Sedangkan untuk menghitung $p(W)$ dapat dilihat pada Persamaan 3.6.

$$p(W) = \frac{\sum W}{N} \quad (3.6)$$

Model *Multinomial Naïve Bayes* (MNB) setiap kata yang muncul dalam dokumen akan dihitung jumlah frekuensi, yang kemudian dihitung menggunakan probabilitas sebelumnya dan dikalikan dengan hasil kemungkinan keseluruhan. Probabilitas prior adalah kemungkinan munculnya suatu kelas di semua dokumen, sedangkan *likelihood* adalah kemungkinan munculnya suatu nilai pada fitur yang terdapat di kelas tertentu (Rahman et al., 2017).

Rumus perhitungan *likelihood* terdapat pada persamaan 3.7 sebagai berikut.

$$p(W | C_k) = \prod_{i=1}^n p(w_i | C_k) \quad (3.7)$$

Dari persamaan yang terdapat pada Persamaan (3.4) sampai Persamaan (3.7), didapatkan Persamaan 3.8. Persamaan 3.8 digunakan untuk menghindari nilai 0.

$$p(C_k | W) = \log p(C_k) + \sum_{i=1}^n w_i \cdot \log p(w_i | C_k) \quad (3.8)$$

Untuk perhitungan $p(w_i | C_k)$ didapatkan dari Persamaan 3.9 sebagai berikut.

$$p(\mathbf{w}_i | C_k) = \frac{\sum w_i}{\sum C_k} \quad (3.9)$$

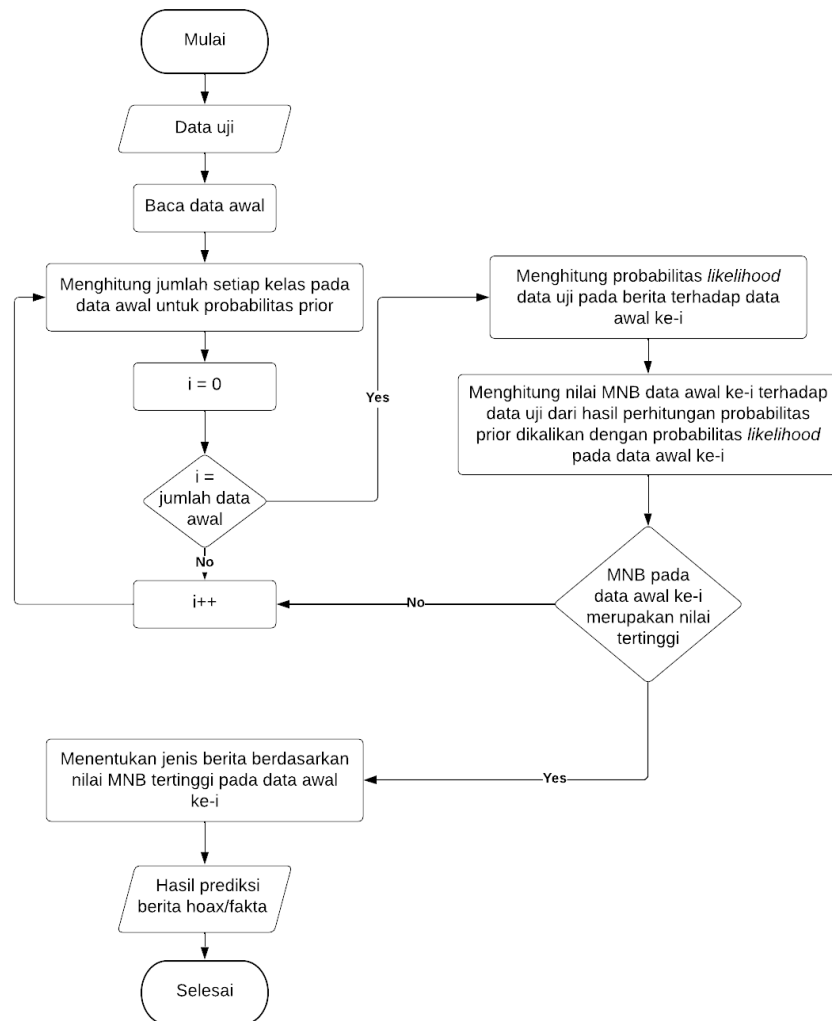
Kemudian dari persamaan (3.8), akan didapatkan kelas yang bernama C_s dari $\mathbf{S}$ yang mana $p(C_k | \mathbf{W})$ $k=1..m$ merupakan nilai maksimum yang dihitung dengan menggunakan Persamaan 3.10.

$$C_s = \operatorname{argmax}_{k \in (1..m)} |p(C_k | \mathbf{W})| \quad (3.10)$$

Terdapat persamaan 3.11 yang merupakan bagian dari Multinomial Naïve Bayes yang mengasumsikan bahwa setiap kata dalam dokumen muncul berdasarkan distribusi multinomial. Pada persamaan 3.11 digunakan m yang merupakan jumlah kata pada dokumen, dimana $\mathbf{w}_i$ adalah frekuensi kemunculan setiap kata. Sedangkan $m!$ merupakan jumlah seluruh kata m dalam dokumen yang berfungsi menghitung seluruh kemungkinan susunan kata dalam dokumen.

$$p(\mathbf{W} | C_k) = \frac{m!}{w_1! w_2! \dots w_n!} \prod_{i=1}^n p(\mathbf{w}_i | C_k)^{w_i} \quad (3.11)$$

Alur metode *multinomial naïve bayes* ditunjukkan pada Gambar 3.9 sebagai berikut.



Gambar 3. 9 Diagram Alur Metode *Multinomial Naive Bayes* (Dewantara, 2022)

3.6 Desain Tampilan Aplikasi

Desain tampilan sistem terdiri dari input teks berita dan *button* proses dengan tampilan “CEK FAKTA” untuk mendeteksi teks yang dimasukkan berupa hoax atau fakta. Kemudian, pada bagian hasil akan ditampilkan teks yang telah diinputkan beserta hasil deteksi yang berupa hoax atau fakta dengan menyertakan persentase probabilitas fakta dan hoaks.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Langkah-Langkah Uji Coba

Pada penelitian ini terdapat beberapa langkah uji coba yang dilakukan. Berikut merupakan langkah-langkah uji coba dari penelitian yang dilakukan.

4.1.1 Pembobotan Kata

Pada penelitian, metode *tf-idf* digunakan untuk menghitung pembobotan kata. Data kata-kata fakta dan hoaks dihitung dengan pembobotan. Pembobotan kata dilakukan dengan 3 langkah. Langkah pertama yaitu menghitung TF pada setiap dokumen di setiap kategori. Setelah menghitung nilai TF, lanjut menghitung IDF. Berikut merupakan sedikit implementasi pembobotan kata fakta ke dalam *source code* menggunakan bahasa *python*.

```
# hitung TF tiap kata
for term in TF_dict:
    TF_dict[term] = TF_dict[term] / len(doc)
return TF_dict

# Definisikan fungsi untuk menghitung IDF
def calc_IDF(docs):
    import math
    N = len(docs)
    IDF_dict = {}
    all_terms = set([term for doc in docs for term in doc])
    for term in all_terms:
        containing_docs = sum([1 for doc in docs if term in
doc])
        IDF_dict[term] = math.log(N / (1 + containing_docs))
    return IDF_dict
```

Tabel 4.1 menunjukkan beberapa hasil dari nilai TF dan IDF pada dokumen yang berkategori fakta, sedangkan Tabel 4.2 untuk dokumen berkategori hoax.

Tabel 4. 1 Hasil TF dan IDF Setiap data Kategori Fakta

No	artikel	TF	IDF
0	oralit rupa obat biasa guna atas ganggu orang alami dehidrasi akibat diare muntah meski oralit diperjualbelikan bebas guna sembarang beri efek samping guna obat milik zat garam gula lalu dua kandung sebut oralit ganti garam mineral hilang tubuh alami dehidrasi akut...	{'oralit': 0.05902777777777776, 'rupa': 0.00347222222222222, 'obat': 0.010416666666666666, 'biasa': 0.00347222222222222, 'guna': 0.01388888888888888, 'atas': 0.00347222222222222, 'ganggu': 0.00347222222222222, 'orang': 0.02430555555555556, 'alami': 0.02777777777777776, 'dehidrasi': 0.01736111111111112,...}	{'oralit': 0.2888141958106869, 'rupa': 0.0038495230018111494, 'obat': 0.01326005912305091, 'biasa': 0.0039025350578208314, 'guna': 0.017312950260343497, 'atas': 0.004883670376518785, 'ganggu': 0.005859025881610459, 'orang': 0.01171713011583345, 'alami': 0.027806443342479045, 'dehidrasi': 0.05289974944342954,...}
1	bentuk ibadah puasa manfaat sehat tubuh karena hindar beberapa biasa buruk ibadah puasa lancar ajar agama islam ibadah puasa tuju lebih takwa allah swt bentuk syukur ibadah tentu beri banyak berkah pahala lipat ganda puasa sehat tubuh mulai bakar lemak tubuh tingkat hormon tumbuh manusia hingga tingkat fungsi otak biasa buruk puasa dampak balik biasa buruk sebut kait menu makan pilih...	{'bentuk': 0.010416666666666666, 'ibadah': 0.02083333333333332, 'puasa': 0.05729166666666664, 'manfaat': 0.00520833333333333, 'sehat': 0.010416666666666666, 'tubuh': 0.02083333333333332, 'karena': 0.010416666666666666, 'hindar': 0.010416666666666666, 'beberapa': 0.00520833333333333, 'biasa': 0.03645833333333336, 'buruk': 0.03645833333333336, 'lancar': 0.00520833333333333,...}	{'bentuk': 0.018305812942820028, 'ibadah': 0.05861272326583409, 'puasa': 0.13795521716234682, 'manfaat': 0.006771440962910616, 'sehat': 0.0033858110312036433, 'tubuh': 0.00620275850636005, 'karena': 0.029306361632917045, 'hindar': 0.01689600723584334, 'beberapa': 0.004675595553652186, 'biasa': 0.04097661810711874, 'buruk': 0.07168119788859287, 'lancar': 0.019212913823510085,...}

Lanjutan Tabel Hasil TF dan IDF Fakta

No	artikel	TF	IDF
2	banyak orang awal hari minum kopi khawatir dampak sistem cerna benar kopi lebih baik minum sarap jelas beri manfaat tambah energi biasa kopi minum pagi hari waktu sarap sedikit minum kopi sarap banyak orang bingung kait waktu tepat seruput kopi gemar kopi sekalipun sering alami ganggu cerna...	{'banyak': 0.004866180048661801, 'orang': 0.024330900243309004, 'awal': 0.0024330900243309003, 'hari': 0.014598540145985401, 'minum': 0.058394160583941604, 'kopi': 0.08029197080291971, 'khawatir': 0.0024330900243309003, 'dampak': 0.0024330900243309003, 'sistem': 0.0024330900243309003, 'cerna': 0.009732360097323601,...}	{'banyak': 0.002843273847434858, 'orang': 0.01172934818686143, 'awal': 0.004656756942097962, 'hari': 0.010944781489756781, 'minum': 0.06745592059308052, 'kopi': 0.2106996780170554, 'khawatir': 0.007056988062651256, 'dampak': 0.004872215329790043, 'sistem': 0.0043839654624855395, 'cerna': 0.019670779888621513,...}
3	lama bulan ramadan umat islam harus laku ibadah puasa tahan lapar haus nafsu fakta puasa dar ibadah tambah pahala lihat sisi sehat puasa milik manfaat baik tubuh sebut spesialis sakit dr herry kongko sppd kemd finasim lalu instagram pribadi setiyantoherry kutip atas izin sangkut Kamis ketua dokter alumni sman jakarta dapat manfaat puasa tubuh manfaat sebut benar dokter kait turun berat badan dokter halim spko rs royal...	{'lama': 0.0058823529411764705, 'bulan': 0.011764705882352941, 'ramadan': 0.0058823529411764705, 'umat': 0.0058823529411764705, 'islam': 0.0058823529411764705, 'harus': 0.0058823529411764705, 'laku': 0.011764705882352941, 'ibadah': 0.011764705882352941, 'puasa': 0.052941176470588235, 'tahan': 0.011764705882352941, 'lapar': 0.011764705882352941, 'haus': 0.0058823529411764705, 'nafsu': 0.0058823529411764705, 'fakta': 0.0058823529411764705,...}	{'lama': 0.005244694819316374, 'bulan': 0.02251678944943838, 'ramadan': 0.018241722289481278, 'umat': 0.0165494748044708, 'islam': 0.017061306433821564, 'harus': 0.01972004245583955, 'laku': 0.00900844556935036, 'ibadah': 0.0330989496089416, 'puasa': 0.12747947339921675, 'tahan': 0.018221332830796064, 'lapar': 0.031285412199209155, 'haus': 0.01931420203532807, 'nafsu': 0.01931420203532807,...}

Lanjutan Tabel Hasil TF dan IDF Fakta

No	artikel	TF	IDF
4	negara dua afrika lama lapor wabah infeksi virus marburg virus marburg sangat tular milik angka mati tinggi organisasi sehat dunia badan sehat dunia ada virus marburg potensi jadi endemi jadi tanda urgensi upaya lanjut kembang vaksin obat virus marburg muncul khawatir pihak ahli sehat masyarakat ubah iklim...	{'negara': 0.009771986970684038, 'dua': 0.003257328990228013, 'afrika': 0.006514657980456026, 'lama': 0.003257328990228013, 'lapor': 0.013029315960912053, 'wabah': 0.013029315960912053, 'infeksi': 0.016286644951140065, 'virus': 0.04560260586319218, 'marburg': 0.05211726384364821, 'sangat': 0.006514657980456026,...}	{'negara': 0.020719187650163753, 'dua': 0.007341351559689296, 'afrika': 0.030001108703505485, 'lama': 0.002904228401575842, 'lapor': 0.03419112335011737, 'wabah': 0.0436795728663547, 'infeksi': 0.031449862158983106, 'virus': 0.07633768752444105, 'marburg': 0.2761338041197674, 'sangat': 0.006565849676870655,...}
5	hormon testosteron pegang peran penting jaga fertilitas subur fungsi seksual sehat tulang otot pria kadar hormon umum alami turun iring tambah umur faktor kondisi medis gaya hidup pola makan turut sebab turun ketidakseimbangan hormon faktor stres berat badan lebih obesitas jadi sebab turun produksi hormon testosteron konsultasi dokter jumlah cara alami laku atas masalah salah satu baik pola makan...	{'hormon': 0.04666666666666667, 'testosteron': 0.03333333333333333, 'pegang': 0.01333333333333334, 'peran': 0.01333333333333334, 'penting': 0.01333333333333334, 'jaga': 0.00666666666666667, 'fertilitas': 0.01333333333333334, 'subur': 0.01333333333333334, 'fungsi': 0.00666666666666667, 'seksual': 0.00666666666666667, 'sehat': 0.02, 'tulang': 0.00666666666666667, 'otot': 0.00666666666666667, 'pria': 0.01333333333333334, 'kadar': 0.01333333333333334,...}	{'hormon': 0.10745397100638882, 'testosteron': 0.13040076684760488, 'pegang': 0.058427021795651755, 'peran': 0.031038705346377815, 'penting': 0.017213122417540878, 'jaga': 0.009445023790697327, 'fertilitas': 0.058427021795651755, 'subur': 0.04918505938818582, 'fungsi': 0.011431989520612845, 'seksual': 0.017494458114143387, 'sehat': 0.006500757179910995, 'tulang': 0.01587031089642505, 'otot': 0.01518594986935783, 'pria': 0.03005059905099485, 'kadar': 0.022319685780955622,...}

Lanjutan Tabel Hasil TF dan IDF Fakta

No	artikel	TF	IDF
6	bulan ramadan tiba umat muslim seluruh dunia mulai jalan ibadah puasa lama bulan penuh jadi wajib beberapa kelompok orang tentu beri ringan laku puasa ramadan salah satu ibu hamil sisi medis ibu hamil boleh laku ibadah puasa beberapa syarat penuh ibu hamil laku ikut beberapa antara kondisi hamil orang beda beda ibu hamil sangat saran...	{'bulan': 0.00823045267489712, 'ramadan': 0.00823045267489712, 'tiba': 0.00411522633744856, 'umat': 0.00411522633744856, 'muslim': 0.00411522633744856, 'seluruh': 0.00411522633744856, 'dunia': 0.00411522633744856, 'mulai': 0.00411522633744856, 'jalan': 0.00823045267489712, 'ibadah': 0.0205761316872428, 'puasa': 0.053497942386831275, 'lama': 0.00411522633744856, 'penuh': 0.00823045267489712, 'jadi': 0.01646090534979424, 'wajib': 0.00411522633744856, 'beberapa': 0.012345679012345678, 'kelompok': 0.00411522633744856, 'orang': 0.00823045267489712,...}	{'bulan': 0.015752486446109156, 'ramadan': 0.025523397442072573, 'tiba': 0.012539199868072188, 'umat': 0.011577821879670932, 'muslim': 0.012996918531077227, 'seluruh': 0.007000021012179386, 'dunia': 0.0068887919077023525, 'mulai': 0.005663967202743714, 'jalan': 0.014345426379083318, 'ibadah': 0.057889109398354656, 'puasa': 0.12882013544228121, 'lama': 0.003669128062896229,...}
7	rasa nyeri serang bagi tubuh tentu mungkin sering anggap sepele meski kadang sebab aktivitas ganggu beberapa orang abai percaya kondisi sebut reda sendiri detikers hati hati tanda sakit saraf jepit lengkap ikut detiksumut hadir gejala saraf jepit sebab hingga cara tangan lansir laman resmi menteri sehat ri saraf jepit sebut radikulopati pinched nerve kondisi jadi jaring tubuh jaring otot...	{'rasa': 0.015527950310559006, 'nyeri': 0.012422360248447204, 'serang': 0.006211180124223602, 'bagi': 0.009316770186335404, 'tubuh': 0.024844720496894408, 'tentu': 0.003105590062111801, 'mungkin': 0.006211180124223602, 'sering': 0.003105590062111801, 'anggap': 0.003105590062111801, 'sepele': 0.003105590062111801, 'meski': 0.003105590062111801, 'kadang': 0.003105590062111801,...}	{'rasa': 0.015230267903908792, 'nyeri': 0.03923293419940082, 'serang': 0.011783353943390567, 'bagi': 0.009915008030597304, 'tubuh': 0.007397078467211985, 'tentu': 0.005368896490890268, 'mungkin': 0.007478091952335006, 'sering': 0.004702260039222905, 'anggap': 0.008044307967223064, 'sepele': 0.012563833534325058, 'meski': 0.005502971558794644,...}

Lanjutan Tabel Hasil TF dan IDF Fakta

No	artikel	TF	IDF
8	malaysia taiwan jegal indomie akibat temu etilen oksida zat karsinogenik sifat kanker risiko picu reaksi buruk tubuh kait masalah pt indofood cbp sukses makmur tbk rupa anak usaha pt indofood sukses makmur beri tanggap atas lapor temu zat picu kanker sebut indomie malaysia taiwan tegas bagaimana sampai badan pom ri produk mi instan indomie aman konsumsi ungkap direktur icbp taufik wiraatmadja konfirmasi jumat aman konsumsi...	{'malaysia': 0.017064846416382253, 'taiwan': 0.013651877133105802, 'jegal': 0.0034129692832764505, 'indomie': 0.020477815699658702, 'akibat': 0.0034129692832764505, 'temu': 0.020477815699658702, 'etilen': 0.023890784982935155, 'oksida': 0.023890784982935155, 'zat': 0.006825938566552901, 'karsinogenik': 0.0034129692832764505, 'sifat': 0.0034129692832764505, 'kanker': 0.006825938566552901, 'risiko': 0.006825938566552901, 'picu': 0.006825938566552901,...}	{'malaysia': 0.05983887196791778, 'taiwan': 0.0573338577184973, 'jegal': 0.018082994425078623, 'indomie': 0.09430382633422714, 'akibat': 0.005408687030183375, 'temu': 0.02588553027431348, 'etilen': 0.11689408125965567, 'oksida': 0.10469005611848864, 'zat': 0.013926422037723922, 'karsinogenik': 0.015717304389037855, 'sifat': 0.007774377066224144, 'kanker': 0.011897401399717563, 'risiko': 0.011336049193321849, 'picu': 0.014616497920804937,...}
9	beberapa waktu lalu gempar produk indomie rasa ayam spesial tarik taiwan produk mi instan asal indonesia sebut tarik otoritas sehat tempat temu residu etilen oksida eto dalam etilen oksida rupa agen sterilisasi pestisida kerap guna bahan alat alat tahan proses panas etilen oksida sebut kait zat...	{'beberapa': 0.004201680672268907, 'waktu': 0.012605042016806723, 'lalu': 0.004201680672268907, 'gempar': 0.004201680672268907, 'produk': 0.03361344537815126, 'indomie': 0.01680672268907563, 'rasa': 0.004201680672268907, 'ayam': 0.004201680672268907, 'spesial': 0.004201680672268907, 'tarik': 0.008403361344537815, 'taiwan': 0.012605042016806723, 'mi': 0.012605042016806723,...}	{'beberapa': 0.0037719090180723517, 'waktu': 0.015822681916556555, 'lalu': 0.004754840935908865, 'gempar': 0.019349454562975175, 'produk': 0.07824883700767515, 'indomie': 0.0773978182519007, 'rasa': 0.00412113131517532, 'ayam': 0.014085744611313962, 'spesial': 0.017645819654957675, 'tarik': 0.022346722999435107,...}

Tabel 4. 2 Hasil TF dan IDF Setiap data Kategori Hoax

No	artikel	TF	IDF
0	maaf bc penting beritahukan anak suami isteri semua teman mu perhati mulai jangan makan makan kaleng an utama buah an masuk merek taiwan keluar thailand negara nya kira ratus orang sakit aids kerja pabrik kaleng masuk darah racun kaleng hari masalah sebut tahu bagi tim sehat thailand kaleng an nya sebut sita contoh lecy rambutan lengkeng mangga puding dan lain lain terima cepat kirim teman...	{'maaf': 0.014285714285714285, 'bc': 0.014285714285714285, 'penting': 0.014285714285714285, 'beritahukan': 0.014285714285714285, 'anak': 0.014285714285714285, 'suami': 0.014285714285714285, 'isteri': 0.014285714285714285, 'semua': 0.02857142857142857, 'teman': 0.02857142857142857, 'mu': 0.02857142857142857, 'perhati': 0.014285714285714285, 'mulai': 0.014285714285714285, 'jangan': 0.014285714285714285, 'makan': 0.02857142857142857, 'kaleng': 0.07142857142857142,...}	{'maaf': 0.04598394035526001, 'bc': 0.06989788940628389, 'penting': 0.01844263116165094, 'beritahukan': 0.06989788940628389, 'anak': 0.021793663263619974, 'suami': 0.04598394035526001, 'isteri': 0.05779363425789527, 'semua': 0.03767530607560216, 'teman': 0.05420342813959661, 'mu': 0.1199915736537122, 'perhati': 0.02353798750363283, 'mulai': 0.01966205757523889, 'jangan': 0.020092815263391572, 'makan': 0.01994742063952827, 'kaleng': 0.22991970177630003,...}
1	sebar arti selamat ribu nyawa mi goreng ujung maut himbauan seluruh masyarakat diposting muslim bagi himbauan semua duduk mohon bantu sebar hati hati edar mi instan palsu ancam sehat selamat anak informasi lewat cara cepat sebar lewat social media informasi dapat satu antara account ...	{'sebar': 0.008379888268156424, 'arti': 0.002793296089385475, 'selamat': 0.0111731843575419, 'ribu': 0.002793296089385475, 'nyawa': 0.002793296089385475, 'mi': 0.0446927374301676, 'goreng': 0.008379888268156424, 'ujung': 0.002793296089385475, 'maut': 0.002793296089385475, 'himbau': 0.008379888268156424, 'seluruh': 0.002793296089385475, 'masyarakat': 0.00558659217877095,...}	{'sebar': 0.017767571532402996, 'arti': 0.00520762615101366, 'selamat': 0.026598286418589467, 'ribu': 0.007635106170632682, 'nyawa': 0.00836796724456422, 'mi': 0.14982825553040105, 'goreng': 0.02430521307611452, 'ujung': 0.008662270360926864, 'maut': 0.012863603871475117, 'himbau': 0.03859081161442535, 'seluruh': 0.00475141091050165, 'masyarakat': 0.010787829813193088,...}

Lanjutan Tabel Hasil TF dan IDF Hoax

No	artikel	TF	IDF
2	sumber siar pesan sama whatsapp narasi institut cegah kanker umum ranking sayur anti kanker jengkol ubi jalar masak ubi jalar mentah asparagus broccoli kubis cabbage kembang kol seledri terong paprica wortel golden cauliflower kmbg kol capsella shepherd purse kol kohlrabi mustard brassica juncea tomat trik utama semua kentang...	{'sumber': 0.002452483139178418, 'siar': 0.0006131207847946045, 'pesan sama': 0.0006131207847946045, 'whatsapp': 0.0006131207847946045, 'narasi': 0.0006131207847946045, 'institut': 0.001226241569589209, 'cegah': 0.002452483139178418, 'kanker': 0.007357449417535254, 'umum': 0.0018393623543838135, 'ranking': 0.001226241569589209, 'sayur': 0.001226241569589209, 'anti': 0.001226241569589209, 'jengkol': 0.003678724708767627, ...}	{'sumber': 0.004456351384947504, 'siar': 0.0028235255585457334, 'pesan sama': 0.0028235255585457334, 'whatsapp': 0.00268671160924211, 'narasi': 0.0022032920136784862, 'institut': 0.004797085230445305, 'cegah': 0.003916653322956575, 'kanker': 0.012823808498285393, 'umum': 0.0029374899922174308, 'ranking': 0.005999818833157415, 'sayur': 0.002574671059469473, 'anti': 0.0029527230026387143, 'jengkol': 0.017999456499472245, ...}
3	institut cegah kanker umum ranking sayur anti kanker jengkol ubi jalar masak ubi jalar mentah asparagus broccoli kubis cabbage kembang kol seledri terong paprica wortel golden cauliflower kmbg kol capsella shepherd...	{'institut': 0.00273224043715847, 'cegah': 0.00546448087431694, 'kanker': 0.01366120218579235, 'umum': 0.00273224043715847, 'ranking': 0.00273224043715847, 'sayur': 0.00273224043715847, 'anti': 0.00273224043715847, 'jengkol': 0.00819672131147541, 'ubi': 0.01092896174863388, 'jalar': 0.01092896174863388, 'masak': 0.01639344262295082, 'mentah': 0.00273224043715847, 'asparagus': 0.00273224043715847, ...}	{'institut': 0.010688587446524988, 'cegah': 0.008726860067953788, 'kanker': 0.023811056080035833, 'umum': 0.004363430033976894, 'ranking': 0.013368448793551565, 'sayur': 0.005736732920757801, 'anti': 0.0065790863624368085, 'jengkol': 0.040105346380654695, 'ubi': 0.044213709268335184, 'jalar': 0.03744825343875897, 'masak': 0.03947451817462085, 'mentah': 0.009159582561455527, 'asparagus': 0.013368448793551565, ...}

Lanjutan Tabel Hasil TF dan IDF Hoax

No	artikel	TF	IDF
4	laku telusur lebih lanjut akun memposting memang sering buat kirim kait bahaya guna alat alat masak jarang beberapa kali laku promosi produk suplemen sehat brand amerika serikat telaah pesan singkat sebut si buat isu bohong giring...	{'laku': 0.011695906432748537, 'telusur': 0.003898635477582846, 'lebih': 0.011695906432748537, 'lanjut': 0.003898635477582846, 'akun': 0.003898635477582846, 'memposting': 0.003898635477582846, 'memang': 0.009746588693957114, 'sering': 0.003898635477582846,...}	{'laku': 0.008955764601108546, 'telusur': 0.017083924501652557, 'lebih': 0.005685766212580341, 'lanjut': 0.006178344248903426, 'akun': 0.012549223488764914, 'memposting': 0.02065620805671749, 'memang': 0.01916289333696718, 'sering': 0.005903032096022516,...}
5	ayah ibu hati beri minum instan buah hati minum sebut bukti kandung zat bahaya siklamat sebab sakit lupus dpat rusak antibodi efek alami ganggu otak sebab kanker netral keluar zat bahaya lanjur masuk dalam tubuh...	{'ayah': 0.02564102564102564, 'ibu': 0.02564102564102564, 'hati': 0.05128205128205128, 'beri': 0.02564102564102564, 'minum': 0.07692307692307693, 'instan': 0.02564102564102564, 'buah': 0.02564102564102564, 'sebut': 0.02564102564102564, 'bukti': 0.02564102564102564, 'kandung': 0.02564102564102564, 'zat': 0.05128205128205128, 'bahaya': 0.05128205128205128,...}	{'ayah': 0.0945866526695881, 'ibu': 0.047393072326062796, 'hati': 0.09728820435312212, 'beri': 0.02862223651043155, 'minum': 0.08886020308896185, 'instan': 0.08419011143604543, 'buah': 0.030871097546818875, 'sebut': 0.0152128678356469, 'bukti': 0.0508751631250131, 'kandung': 0.026918516012786605, 'zat': 0.10462670915520793, 'bahaya': 0.07109201851896875,...}
6	tahun laku teliti akhir harvard university tambah temu betapa bahaya air mineral kandung flourida dapat air mineral aqua punya danone aqua kini saham beli danone kandung zat flourida temu kini harvard university air mineral fluorida kurang iq anak anak buah tanya besar...	{'tahun': 0.005606728073688426, 'laku': 0.002803364036844213, 'teliti': 0.0092110532639167, 'akhir': 0.0012014417300760913, 'harvard': 0.0032038446135362435, 'university': 0.0024028834601521826, 'tambah': 0.0012014417300760913, 'temu': 0.007208650380456548, 'betapa': 0.0008009611533840609, 'bahaya': 0.004004805766920304, 'air': 0.016019223067681217,...}	{'tahun': 0.006844581064073865, 'laku': 0.0021465859486437585, 'teliti': 0.015419086092169982, 'akhir': 0.0017527413543709162, 'harvard': 0.012961327666968904, 'university': 0.007588909578818821, 'tambah': 0.0015955451978324968, 'temu': 0.00911228718907551, 'betapa': 0.0035098331074680665,...}

Lanjutan Tabel Hasil TF dan IDF Hoax

No	artikel	TF	IDF
7	wa teman dokter bahaya sakit leukimia adik ipar umur tahun baru tinggal malam sebab leukimia almarhum masa hidup teliti gelar master in botanical kampus usm kaji jenis tumbuh rekan satu tim research beliau tinggal tahun lalu idap sakit leukimia almarhum tahan hingga malam memang hasil teliti...	{'wa': 0.010101010101010102, 'teman': 0.010101010101010102, 'dokter': 0.010101010101010102, 'bahaya': 0.020202020202020204, 'sakit': 0.020202020202020204, 'leukimia': 0.06060606060606061, 'adik': 0.010101010101010102, 'ipar': 0.010101010101010102, 'umur': 0.010101010101010102, 'tahun': 0.020202020202020204, 'baru': 0.010101010101010102, 'tinggal': 0.020202020202020204,...}	{'wa': 0.03629867953848093, 'teman': 0.01916282813016042, 'dokter': 0.014104491888620124, 'bahaya': 0.028005946689290722, 'sakit': 0.014104236815828073, 'leukimia': 0.2791012233932177, 'adik': 0.031901527303553195, 'ipar': 0.05351835723785896, 'umur': 0.02583310447093774, 'tahun': 0.024662220659440754, 'baru': 0.014627977422605841, 'tinggal': 0.04008346185607093,...}
8	maaf bc penting beritahukan anak suami isteri semua teman mu perhati mulai jangan makan makan kaleng an utama buah an masuk merek taiwan keluar thailand negara nya kira ratus orang sakit aids kerja pabrik kaleng masuk darah racun kaleng...	{'maaf': 0.0072992700729927005, 'bc': 0.0072992700729927005, 'penting': 0.0072992700729927005, 'beritahukan': 0.0072992700729927005, 'anak': 0.0072992700729927005, 'suami': 0.0072992700729927005, 'isteri': 0.0072992700729927005, 'semua': 0.014598540145985401, 'teman': 0.014598540145985401,...}	{'maaf': 0.023495443977140148, 'bc': 0.03571425006160491, 'penting': 0.009423242199383692, 'beritahukan': 0.03571425006160491, 'anak': 0.01113544838287152, 'suami': 0.023495443977140148, 'isteri': 0.029529594146369846, 'semua': 0.019250156388993807, 'teman': 0.027695182261107756,...}
9	tahun laku teliti akhir harvard university tambah temu betapa bahaya air mineral kandung flourida dapat air mineral aqua punya danone aqua kini saham beli danone kandung zat flourida temu kini harvard university air mineral...	{'tahun': 0.012987012987012988, 'laku': 0.008658008658008658, 'teliti': 0.008658008658008658, 'akhir': 0.004329004329004329, 'harvard': 0.012987012987012988, 'university': 0.008658008658008658, 'tambah': 0.004329004329004329, 'temu': 0.021645021645021644,...}	{'tahun': 0.015854284709640485, 'laku': 0.006629591977443988, 'teliti': 0.014493302455165988, 'akhir': 0.006315433134003142, 'harvard': 0.05253966750717753, 'university': 0.02734416626018845, 'tambah': 0.00574902793504725,...}

Setelah dilakukan perhitungan TF dan IDF masing-masing kata pada setiap dokumen di setiap kategori seperti tersebut di atas, langkah berikutnya adalah melakukan perhitungan rata-rata nilai TF-IDF dari masing-masing kata pada masing-masing kategori hoax maupun fakta. Berikut merupakan tabel hasil rata-rata TF-IDF untuk masing-masing kata pada setiap kategori.

Tabel 4. 3 Rata-Rata TF-IDF setiap kategori

kata	fakta	hoax	rata-rata
ajar	0.003911	0.001564	0.002738
akhir	0.006177	0.007399	0.006788
akibat	0.003621	0.004058	0.00384
aktif	0.001192	0.000755	0.000974
aktivitas	0.00452	0.000438	0.002479
aku	0.003494	0.002892	0.003193
akun	0.000958	0.004716	0.002837
al	0.001247	0.000762	0.001005
alam	0.001926	0.004904	0.003415
alami	0.019146	0.007538	0.013342
alas	0.001569	0.000498	0.001033
alat	0.001641	0.003455	0.002548
alir	0.004013	0.002208	0.00311
alir darah	0.001644	0.000533	0.001088
alkohol	0.001791	0.000237	0.001014
allah	0.00023	0.005385	0.002807
alpukat	0.000626	0.000997	0.000812
...	...	...	...
wanita	0.007627	0.00724	0.007433
warna	0.002332	0.001941	0.002137
waspada	0.00245	0.002108	0.002279
wilayah	0.001155	0.002193	0.001674
yakin	0.000906	0.002599	0.001752
zaitun	0.001669	0.000269	0.000969
zat	0.006218	0.006512	0.006365

Dari pembobotan kata yang sudah dilakukan, selain didapatkan hasil TF-IDF untuk seluruh term pada dokumen, maka didapatkan *wordcloud* yang tercantum pada Gambar 4.1.

4.1.2 Naïve Bayes Classifier

a. Menghitung probabilitas prior. Pada program, langkah awal yang dilakukan adalah menghitung nilai probabilitas prior pada dokumen berita kesehatan hoaks dan fakta.

b. Menghitung probabilitas kemunculan kata dalam masing-masing kategori dengan menggunakan probabilitas *Likelihood*. Nilai

probabilitas yang telah dihitung kemudian disimpan ke dalam *dictionary* *hoax* dan *dictionary not hoax*.

```
def training(dataset, test_data, data_dictionary,
dict_hoax_words, dict_not_hoax_words, vocabulary, alpha =
1):
    prob_hoax_words = {}
    prob_not_hoax_words = {}
    hoax_words = {}
    not_hoax_words = {}
    num_of_word_in_hoax = len(data_dictionary['1'])
    num_of_word_in_not_hoax = len(data_dictionary['0'])
    for word in hoax_words:
        prob_hoax_words[word] = (hoax_words[word] +
alpha)/(num_of_word_in_hoax + vocabulary)
    for word in not_hoax_words:
        prob_not_hoax_words[word] = (not_hoax_words[word] +
alpha)/(num_of_word_in_not_hoax + vocabulary)
```

- c. Menghitung probabilitas *Posterior* untuk Kelas Hoax dan Non-Hoax dengan menggunakan probabilitas *Likelihood* yang telah dihitung sebelumnya dalam fungsi training, serta probabilitas Prior untuk Kelas Hoax dan Non-Hoax.

```
def predict(test_data, dict_prob_hoax, dict_prob_not_hoax, PHoax,
PnotHoax, sumdata):

    probability_hoax = []
    probability_not_hoax = []
    prediction = []
    hoax_meter = 1
    not_hoax_meter = 1
    for value in probability_hoax:
        hoax_meter *= value
        hoax_meter = hoax_meter * (PHoax/sumdata)
    for value in probability_not_hoax:
        not_hoax_meter *= value
        not_hoax_meter = not_hoax_meter * (PnotHoax/sumdata)
```

4.2 Implementasi Tampilan Antarmuka

4.2.1 Beranda



Gambar 4. 2 Halaman Beranda

Laman awal atau beranda ini adalah laman yang muncul pertama kali saat membuka *website*. Pada laman ini, terdapat input teks data berita, dan *button* cek fakta berita.

4.2.2 Hasil Cek Fakta Berita



Gambar 4. 3 Tampilan Halaman Hasil

Halaman hasil adalah halaman yang menampilkan hasil prediksi deteksi berita kesehatan yang telah diinputkan pada halaman awal. Pada halaman hasil terdapat narasi berita yang telah diinput pada halaman awal, hasil deteksi, dan persentase dari hasil deteksi.

4.3 Pengujian

4.3.1 Data Uji

Pada tahap pengujian, data berita kesehatan hasil scraping dari web detikcom dan turnbackhoax dengan berita kesehatan tahun 2023. Setelah proses scraping, teks berita diklasifikasikan menjadi berita kesehatan hoaks dan berita kesehatan fakta.

Tabel 4. 4 Hasil Klasifikasi Berita

title	timestamp	link	article	hoax
9 Makanan Rendah Kalori yang Cocok Dikonsumsi Saat Diet	23 Mei 2023 07:00	https://food.detik.com/info-sehat/d-6732871/9-makanan-rendah-kalori-yang-cocok-dikonsumsi-saat-diet	Makanan rendah kalori cocok dikonsumsi untuk kamu yang sedang menjalani program diet. Beberapa diantaranya berasal dari buah-buahan dan sayur-sayuran. Meskipun rendah kalori, jangan lupa untuk mengecek nutrisinya....	0
Cek Gula Darah: Cara, Penyebab, dan Mengapa Harus Melakukannya ?	23 Mei 2023 05:15	https://health.detik.com/berita-detikhealth/d-6732795/cek-gula-darah-cara-penyebab-dan-mengapa-harus-melakukannya	Pada penderita diabetes, cek gula darah dilakukan untuk memantau apakah kadar gula darah tetap terkendali atau tidak. Namun, setiap orang juga dianjurkan untuk melakukan cek gula darah untuk mengetahui kondisi glukosa pada tubuh...	0
10 Manfaat Pisang untuk Tubuh yang Jarang Diketahui Orang	23 Mei 2023 05:00	https://food.detik.com/info-sehat/d-6732861/10-manfaat-pisang-untuk-tubuh-yang-jarang-diketahui-orang	Pisang merupakan salah satu jenis buah yang mengandung banyak zat bergizi. Maka dari itu, disarankan untuk mengonsumsi pisang setiap hari demi menjaga kesehatan tubuh.	0

4.3.2 Hasil Uji Coba

Tahap deteksi hoax pada penelitian ini mengenakan metode *multinomial naïve bayes* yang diimplementasikan ke dalam sebuah program berbasis web yang diprogram dengan *python* sebagai bahasa pemrograman. Tabel 4.5 merupakan tabel hasil uji berita kesehatan yang dilakukan di dalam program. Dari 10 berita yang terdiri dari 4 berita fakta dan 6 berita hoaks, maka mendapatkan perolehan seperti yang tercantum pada Tabel 4.5. Hasil prediksi dapat kurang akurat akibat teks berita yang dimasukkan untuk dites bukan merupakan teks berita. Melainkan hanya sebuah kalimat saja, sedangkan program digunakan untuk klasifikasi bentuk teks. Gambar hasil prediksi

program ditampilkan pada bagian Lampiran Gambar 4.5 sampai Gambar 4.14. Pada Tabel 4.5 pada kolom hasil prediksi terdapat probabilitas fake dan probabilitas true. Hal ini dimaksudkan untuk mendefinisikan berita tersebut adalah berita fakta atau hoax.

Pada penelitian mengenai deteksi berita hoaks ini terdapat ketentuan hasil prediksi untuk probabilitas fake dan probabilitas true. Qubra & Saputra (2024) apabila probabilitas fake lebih tinggi daripada probabilitas true, maka berita akan didefinisikan sebagai berita hoax. Sebaliknya, apabila hasil probabilitas true lebih tinggi daripada probabilitas fake, maka berita akan didefinisikan sebuah berita fakta. Serota *et al.* (2024) mengungkapkan ketika probabilitas memiliki hasil yang sama atau ambigu maka keputusan terbaik adalah mengklasifikasikan ke dalam berita hoax agar dapat meminimalisasi risiko kesalahan penyebaran berita palsu, karena kesalahan mengidentifikasi berita hoaks sebagai benar lebih berisiko daripada sebaliknya. Hal ini dikenal sebagai pendekatan *decision threshold*.

Tabel 4. 5 Tabel Hasil Uji Berita Kesehatan di dalam Program

Narasi berita	Kelas aktual	Hasil prediksi
jakarta lebaran identik ragam hidangan lezat meski kandungan gula lemak garam tinggi ahli gizi tegas tetap nikmat makan enak hari raya rusak diet jalan benar satu buat diet jadi lebih baik rusak jadi gin pola pikir nya diet itu bukan arti makan diet itu arti justru atur pola makan ungkap spesialis gizi klinik dr ana suganda spgk kes hubung detikcom lebaran wajar nikmat makan khas ketupat opor kue kering hidangan lain penting tetap mindful arti perlu sepenuh sadar makan konsumsi tubuh tetep mindful bukan arti cheating kalau cheating kayak mindsetnya bebas makan apa berapa banyak banyak banyak serah mungkin makan...	Fakta	Probabilitas Fake: 17.56% Probabilitas True: 82.44% Kesimpulan berita: FAKTA

Lanjutan Tabel Hasil Uji Berita Kesehatan di dalam Program

Narasi berita	Kelas aktual	Hasil prediksi
jakarta marasmus jenis malnutrisi parah tanda kurang makronutrien khusus protein siapa derita marasmus baik anak maupun orang dewasa ada paling sering serang anak anak utama negara kembang lansir laman cleveland clinic kondisi kurang nutrisi penting sebab masalah sehat lebih serius derita mampu tahan fungsi tubuh normal tubuh sulit tumbuh sel baru lawan sakit ganggu sehat timbul akibat marasmus marasmus umum jadi jumlah negara masyarakat milik cukup makan negara maju kondisi jadi akibat lain makan anoreksia nervosa sebab marasmus kondisi marasmus sebab jumlah umum miskin milik makan anoreksia infeksi sebab diare kronis aids makan terlalu sedikit konsumsi nutrisi...	Fakta	Probabilitas Fake: 7.41% Probabilitas True: 92.59% Kesimpulan berita: FAKTA
Covid-19 merupakan penyakit yang tidak menular	Hoax	Probabilitas Fake: 81.31% Probabilitas True: 18.69% Kesimpulan berita: HOAX
Nyamuk aedes aegypti merupakan nyamuk penyebab DBD	Fakta	Probabilitas Fake: 52.55% Probabilitas True: 47.45% Kesimpulan Berita: HOAX
ingat medis jepang sila bagi keluarga teman sangat penting selamat hidup kelompok dokter jepang konfirmasi air panas efektif atas masalah sehat tentu migrain tekan darah tinggi tekan darah rendah nyeri sendi detak jantung tiba tiba tingkat kurang epilepsi kadar kolesterol tingkat batuk sulit tubuh rasa sakit bodoh pakai pertusis vena jadi putih sakit kait rahim rahim adang nafsu makan buruk semua sakit hubung mata telinga tenggorok sakit kepala cara guna air panas bangun pagi minum gelas air panas perut kosong mungkin minum gelas lebih iring waktu sana catat jangan makan apa menit minum air	Hoax	Probabilitas Fake: 60.78% Probabilitas True: 39.22% Kesimpulan Berita: HOAX

Lanjutan Tabel Hasil Uji Berita Kesehatan di dalam Program

Narasi berita	Kelas aktual	Hasil prediksi
baca habis lihat bawah ijin pak ibu kakak abang adik assalammu alaikum wr wb breaking news terus informasi dr ibu dubes kbri kalau beritahu adik suami isteri semua teman perhati mulai jangan makan makan kaleng utama buah an khusus produksi thailand negara kira orang idap aids kerja pabrik kaleng masuk darah kaleng masalah sebut tahu depkes thailand kaleng an sebut banyak sita lebih banyak lanjut ekspor contoh lecy rambut an lengkung mangga puding dan lain lain terima cepat kirim saudara teman semua konsumsi kaleng apa selamat semua informasi dr ibu dubes kbri rita toisuta arifson tri sehat ri simak berita moga manfaat mohon bantu iya warning sebar luas...	Hoax	Probabilitas Fake: 95.38% Probabilitas True: 4.62% Kesimpulan Berita: HOAX
penting malam jam pagi hingga pagi pasti off ponsel laptop dan lain lain jauhkn dr badan tv singapore mengumumkn berita sebut beritahu keluarga sahabat malam jam pagi hingga pagi bumi hadap radiasi paling tinggi pancar cahaya cosmic lintas dekat bumi off ponsel dan lain lain jauhkn dr badan bagai ikut akn sebab dapat efek radiasi bahaya lihat google nasa berita bbc bagi pesan orang penting keluarga teman sahabat anak istri selamat nyawa banyak orang buat moga manfaat amiin	Hoax	Probabilitas Fake: 65.19% Probabilitas True: 34.81% Kesimpulan berita: HOAX
imbau kena masker sebar disentri amoeba temu akun facebook ikut narasi unggah salah satu akun selasa teman mau ingetin seperti mulai pakai masker main banyak banget orang kena virus serang perut nama amuba rs penuh pasti hampir semua sakit sama jadi perut lilit sakitt sekali orang mau kontraksi terus lemasss badan sakit ujung mencret pipis keluar mencret efek kalium elektrolit nya turun terus terus serang fungsi jantung waspada jaga sehat sahabat unggah serta foto menteri koordinator bidang maritim investasi menko marves luhut binsar pandjaitan serta teks ikut menteri luhut wajib guna masker warga jabodetabek	Hoax	Probabilitas Fake: 75.92% Probabilitas True: 24.08% Kesimpulan Berita: HOAX
sekitar kalian terdapat tower hias daun palem toweri frekuensi timbul radiasi ygtdk baik tubuh manusia makin pendek gelombang makin tinggi frekuensi makin tinggi frekuensi makin bahaya radiasi	Hoax	Probabilitas Fake: 51.71% Probabilitas True: 48.29% Kesimpulan berita: HOAX
air rebus serai teh serai masuk minum herbal populer konsumsi air serai tidur beri efek relaksasi reda cemas efek serai sebut bantu orang milik masalah tidur insomnia sekaligus tingkat kualitas tidur tahu lebih lanjut aneka potensi manfaat minum air serai tidur simak ulasan ikut apa manfaat minum air serai tidur sari healthline tua saude ikut manfaat minum air serai tidur atas susah tidur serai kandung senyawa sirtral tindak obat tenang atas insomnia susah tidur tingkat kualitas tidur relaksasi serai bantu kendur otot otot tubuh milik efek tenang buat tubuh lebih relaks reda cemas efek relaksasi tenang...	Fakta	Probabilitas Fake: 42.85% Probabilitas True: 57.15% Kesimpulan berita: FAKTA

4.3.3 Pengujian Akurasi

Pada langkah uji akurasi, dilakukan pengujian akan data yang telah mengimplementasikan metode *naïve bayes* yang bertujuan untuk mencari nilai akurasi, presisi, *recall*, dan *f-measure*. Cara untuk menghitung nilai tersebut adalah dengan mengenakan *Confusion Matrix*. Maulana Stiawan & Hidayat (2019) *Confusion Matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Konsep dari *confusion matrix* tertuang dalam Tabel 4.6.

Tabel 4. 6 Konsep *Confusion Matrix*

<i>Actual Prediction</i>	<i>Prediction Class</i>	
	<i>Negative</i>	<i>Positive</i>
<i>Class Negative</i>	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
<i>Class Positive</i>	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

Berdasarkan konsep *confusion matrix* yang tertuang pada Tabel 4.6, maka perhitungan nilai *accuracy*, presisi, *recall*, beserta *f-measure* dapat dilakukan seperti pada Persamaan 4.1 hingga dengan 4.4 berikut ini.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4.3)$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (4.4)$$

Pada Tabel 4.4 terdapat *True Negative* (TN), *False Negative* (FN), *True Positive* (TP), dan *False Positive* (FP). TP merupakan kondisi dimana nilai prediksi benar sesuai dengan nilai aktual benar. Sedangkan TN merupakan kondisi saat nilai prediksi salah sesuai dengan nilai aktual salah. FP memiliki pengertian kondisi disaat nilai prediksi benar, tetapi nilai aktual salah. Sedangkan FN, merupakan kondisi dimana nilai prediksi salah tetapi nilai aktual benar (Saputra, 2022).

Dari Persamaan 4.1 sampai 4.4 didapatkan hasil perhitungan dari akurasi, presisi, *recall*, serta *f-measure*. Sedangkan untuk hasil *confusion matrix* didapatkan dari Gambar 4.3. Berikut merupakan perhitungan dari *accuracy*, *precision*, *recall*, serta *f-measure*.

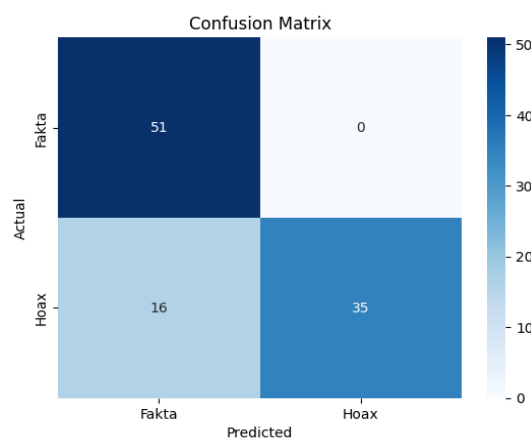
$$\begin{aligned} Accuracy &= \frac{35 + 51}{35 + 0 + 16 + 51} \times 100\% \\ &= 84.313725\% \end{aligned}$$

$$\begin{aligned} Precision &= \frac{35}{35 + 0} \times 100\% \\ &= 100\% \end{aligned}$$

$$\begin{aligned} Recall &= \frac{35}{35 + 16} \times 100\% \\ &= 68,6274\% \end{aligned}$$

$$\begin{aligned} F - Measure &= \frac{2 \times 100 \times 68,62}{100 + 68,62} \times 100\% \\ &= 81.39\% \end{aligned}$$

Dari hasil perhitungan di atas, didapatkan nilai *f-measure* sebesar 81,39%. Hal ini menunjukkan bahwa presisi dan *recall* yang ada memiliki nilai yang seimbang. Tingginya *f-measure* menunjukkan bahwa model multinomial naïve bayes memiliki performa yang baik dalam menangani keseimbangan antara *false positive* dan *false negative*. Presisi yang tinggi menunjukkan bahwa model sangat akurat dalam mengidentifikasi *false positive*.



Gambar 4. 4 Confusion Matrix

Pada Gambar 4.4 menunjukkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dari pengujian masing-masing adalah 35, 51, 0, dan 16. Terdapat beberapa penyebab FP mendapatkan nilai 0, salah satunya adalah pengambilan sampel data uji yang mungkin tidak memadai atau tidak representatif, sehingga tidak ada contoh yang sesuai untuk menghasilkan *false positive*.

Berdasarkan penelitian yang sudah dilakukan, dalam hal menyebarkan informasi, sebaiknya dilakukan berdasarkan fakta. Bukan menyebarkan fitnah atau

kebohongan yang dapat merugikan orang lain. Hal ini sesuai dengan ayat dalam Al-Qur'an, Q.S. An-Nur pada ayat 11 yang berbunyi:

إِنَّ الَّذِينَ جَاءُوا بِالْإِفْكِ عُصْبَةٌ مِّنْكُمْ لَا تَحْسَبُوهُ شَرًّا لَّكُم بَلْ هُوَ خَيْرٌ لَّكُم لِّكُلِّ امْرِئٍ مِّنْهُمْ مَا اكْتَسَبَ مِنَ الْإِثْمِ وَالَّذِي تَوَلَّى كِبْرَهُ مِنْهُمْ لَهُ عَذَابٌ عَظِيمٌ

“Sesungguhnya orang-orang yang membawa berita bohong itu adalah kelompok di antara kamu (juga). Janganlah kamu mengira bahwa peristiwa itu buruk bagimu, sebaliknya itu baik bagimu. Setiap orang dari mereka akan mendapat balasan dari dosa yang diperbuatnya. Adapun orang yang mengambil peran besar di antara mereka, dia mendapat azab yang sangat berat”(QS. An-Nur:11).

Asbabun nuzul ayat ini adalah mengenai kasus berita bohong yang menyebar mengenai Sayyidah Aisyah yang dituduh berselingkuh dengan Sofwan bin Mu'athal. Berita ini disebarkan oleh kaum munafik yang ada di Madinah dengan tujuan merusak citra rumah tangga Rasulullah ﷺ. Ayat ini memberikan pesan kepada manusia bahwa setiap orang akan mendapatkan balasan dari dosa yang telah dikerjakan. Terkhusus dalam hal ini adalah menyebarkan berita hoaks. Menyebarkan berita hoaks dapat juga menimbulkan fitnah dan dalam islam, fitnah merupakan tindakan yang sangat kejam.

Allah memerintahkan umat-Nya untuk selalu selektif dan memilah terlebih dahulu informasi yang diterima dalam Q.S Al-Hujurat ayat 6.

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحَبُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ ﴿٦﴾

“Wahai orang-orang yang beriman, jika seorang fasik datang kepadamu membawa berita penting, maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena ketidaktahuan(-mu) yang berakibat kamu menyesali perbuatanmu itu”(QS. Al-Hujurat:6).

Selain itu, pada penelitian ini dibahas mengenai akurasi atau ketepatan. Allah Subhanahu Wa Ta'ala berfirman dalam Q.S Al-Baqarah ayat 202 mengenai ketepatan dan kecepatan Allah SWT dalam seluruh amal perbuatan manusia.

أُولَٰئِكَ هُم مِّنْ نَّصِيبِ مِّمَّا كَسَبُوا ۖ وَاللَّهُ سَرِيعُ الْحِسَابِ ﴿٢٠٢﴾

“Sesungguhnya agama di sisi Allah ialah Islam. Tidaklah berselisih orang-orang yang telah diberi Kitab kecuali setelah mereka memperoleh ilmu, karena kedengkian di antara mereka. Barang siapa ingkar terhadap ayat-ayat Allah, maka sungguh, Allah sangat cepat perhitungan-Nya” (QS.Al-Baqarah:202).

Pada ayat tersebut dijelaskan secara akurat mengenai kecepatan dan ketetapan Allah dalam menghitung amal perbuatan. Menurut tafsir jalalain, (Mereka itulah orang-orang yang mendapat bagian), maksudnya pahala (dari), artinya disebabkan (apa yang mereka usahakan), yakni amal mereka dari haji dan doa (dan Allah sangat cepat perhitungan-Nya).

Menurut keterangan sebuah hadits, Allah melakukan hisab atau perhitungan untuk semua makhluk dalam tempo yang tidak lebih dari setengah hari waktu dunia. Jika memiliki amal yang baik, maka Allah SWT akan mencatat hal tersebut sebagai kebaikan. Begitu juga sebaliknya. Jika memiliki amal yang jelek, maka Allah SWT akan mencatat hal tersebut sebagai suatu keburukan. Hal ini juga berlaku pada perhitungan niat. Niat baik akan dicatat amal baik, dan niat jelek tidak dicatat sebagai suatu keburukan kecuali pada saat melakukannya. Kebaikan dicatat sebagai amal kebaikan kalau dilakukan dengan niat baik dan ikhlas, kejelekan dicatat sebagai keburukan pada saat melakukannya mau dengan niat baik dan ikhlas. Allah

sangat detail dan akurat dalam memperhitungkan amal perbuatan. Maha besar Allah atas segala firman-Nya.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan yang dapat diperoleh dari dasar uji coba deteksi dan implementasi sistem, hasil akurasi dari deteksi berita hoaks dan fakta menggunakan metode *Multinomial Naïve Bayes* adalah sebesar 84,31%. Sedangkan untuk presisi, *recall*, dan *f-measure* masing-masing mendapatkan nilai 100%, 68,62%, 81,39%. Hal ini membuktikan bahwa dalam kasus ini, metode *Multinomial Naïve Bayes* menghasilkan nilai akurasi yang cukup tinggi. Hasil keakuratan yang dihasilkan oleh program deteksi hoaks berbasis web ini dapat berkurang tergantung pada data teks yang dimasukkan. Apabila data teks yang dimasukkan hanya terdiri dari satu kalimat saja, maka mesin deteksi tidak dapat mendeteksi dengan optimal.

5.2 Saran

Penelitian ini belumlah sempurna. Masih banyak yang bisa dikembangkan dan dipelajari lebih dalam lagi dari penelitian ini. Contohnya dengan pembaharuan data atau menggunakan data yang lain untuk diuji menggunakan metode yang sama. Ekstarksi fitur selain TF-IDF juga dapat digunakan. Ekstraksi fitur tersebut contohnya adalah *word embeddings*. Selain dengan menggunakan metode yang sama, untuk penelitian selanjutnya juga dapat dilakukan dengan mengaplikasikan metode *naïve bayes* yang lain seperti *gaussian* dan *bernoulli*.

DAFTAR PUSTAKA

- Dewantara, F. (2022). *IDENTIFIKASI JENIS PENYAKIT BERDASARKAN GEJALA MENGGUNAKAN METODE MULTINOMIAL NAIVE BAYES*. <http://etheses.uin-malang.ac.id/42900/7/17650063.pdf>
- Doni003. (2022). *Awas Disinformasi, Varian Baru Virus Corona*. KOMINFO. https://www.kominfo.go.id/content/detail/39774/awas-disinformasi-varian-baru-virus-corona/0/berita_satker
- Maulana Stiawan, M., & Hidayat, R. (2019). Pengembangan Sistem Identifikasi Fakta Dan Tidak Fakta Berita di Media Informasi Berbahasa Indonesia. *Seminar Nasional Rekayasa Dan Teknologi*, 27(November), 34–39. <https://jurnal.tau.ac.id/index.php/snartek/article/view/95/63>
- Munirul, Ula, Alvanof, M. M., & Triandi, R. (2020). Analisa Dan Deteksi Konten Hoax Pada Media Berita. *Jurnal Teknologi Terapan & Sains 4.0 Universitas Malikussaleh*, 1, 2.
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1. <https://doi.org/10.21580/wjit.2019.1.1.3915>
- Naomi Chatrina Siregar, Riki Ruli A. Siregar, & M. Yoga Distra Sudirman. (2020). Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ). *Jurnal Teknologia*, 3(1), 102–110.
- Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomsin)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.500>
- Qubra, R., & Saputra, R. A. (2024). Classification of Hoax News Using the Naïve Bayes Method. *International Journal Software Engineering and Computer Science (IJSECS)*, 4(1), 40–48. <https://doi.org/10.35870/ijsecs.v4i1.2068>
- Rahman, A., Wiranto, & Doewes, A. (2017). Online News Classification Using Multinomial Naive Bayes. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 6(1), 32–38. www.kompas.com
- Saputra, I. dan D. A. K. (2022). *MACHINE LEARNING UNTUK PEMULA*. Penerbit Informatika.
- Serota, K. B., Levine, T. R., Zvi, L., Markowitz, D. M., & Docan-Morgan, T. (2024). The ubiquity of long-tail lie distributions: seven studies from five continents. *Journal of Communication*, 74(1), 1–11. <https://doi.org/10.1093/joc/jqad040>
- Setianingrum, A. H., Kalokasari, D. H., & Shofi, I. M. (2018). IMPLEMENTASI ALGORITMA MULTINOMIAL NAIVE BAYES CLASSIFIER. *JURNAL TEKNIK INFORMATIKA*, 10(2), 109–118.

<https://doi.org/10.15408/jti.v10i2.6822>

Soleman, S. (2021). Pemanfaatan Metode Klasifikasi Naïve Bayes Untuk Pendeteksi Berita Hoax Pada Artikel Berbahasa Indonesia. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7(2), 83. <https://doi.org/10.24014/coreit.v7i2.14290>

Yodi Prayoga, A., Id Hadiana, A., & Rakhmat Umbara, F. (2021). Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf. *Jurnal Syntax Admiration*, 2(10), 1808–1823. <https://doi.org/10.46799/jsa.v2i10.327>

LAMPIRAN

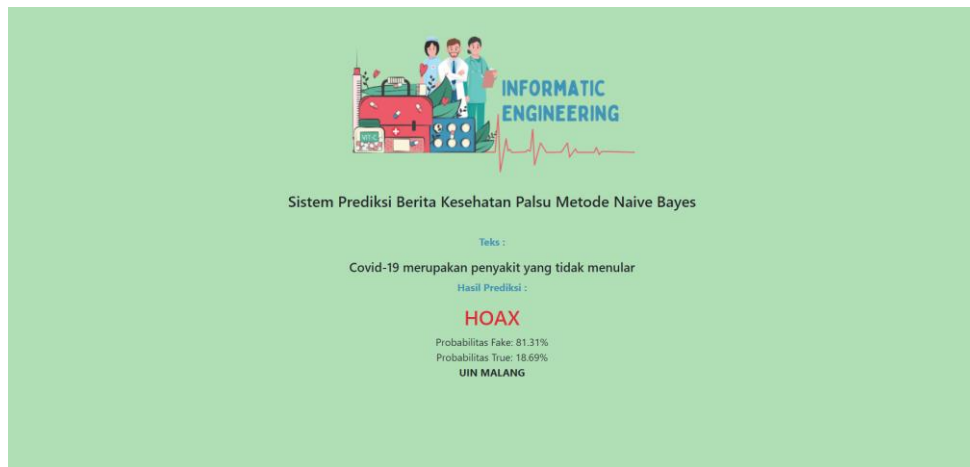
GAMBAR HASIL UJI PREDIKSI PADA WEB TABEL 4.3.



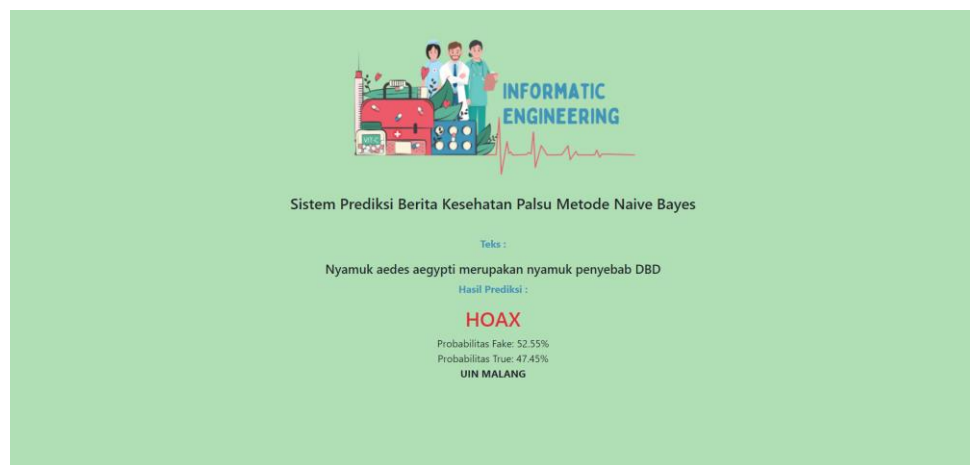
Gambar 4. 5 Hasil Uji Prediksi pada Program Web



Gambar 4. 6 Hasil Uji Prediksi pada Program Web



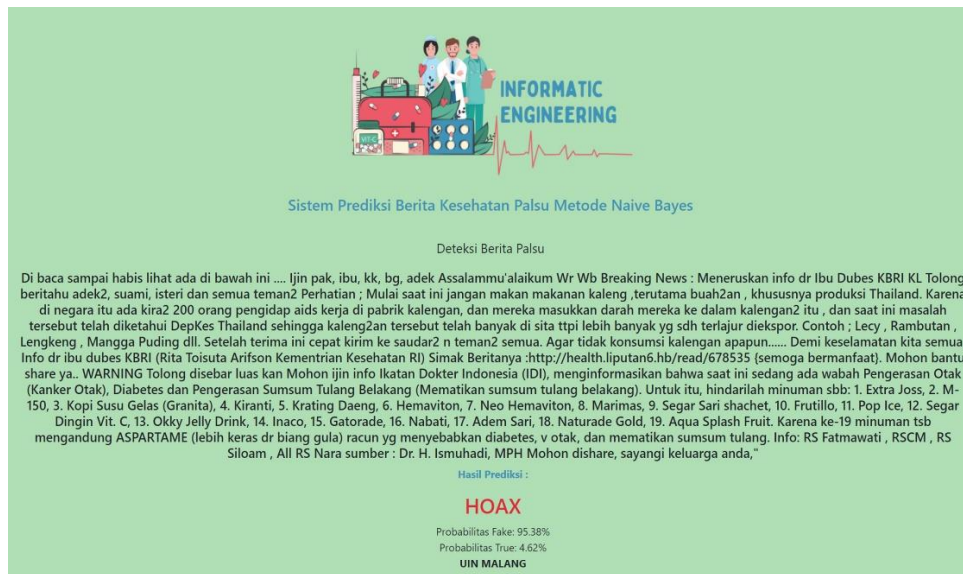
Gambar 4. 7 Hasil Uji Prediksi pada Program Web



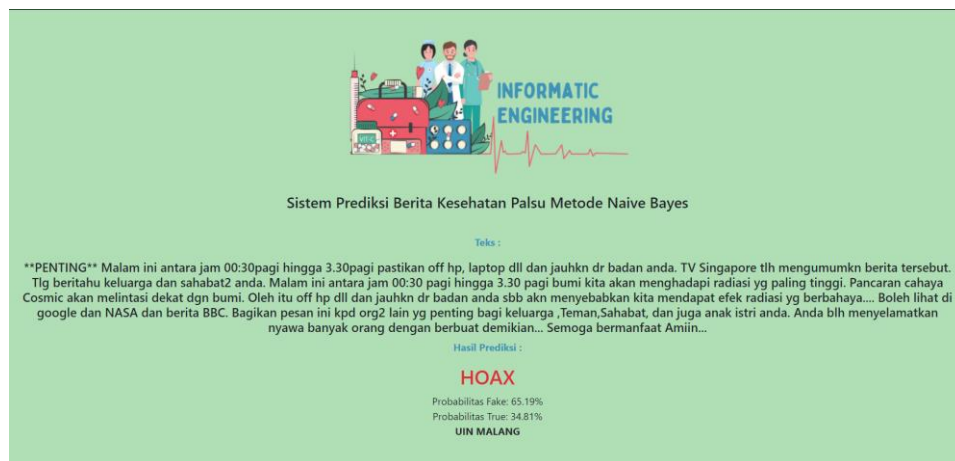
Gambar 4. 8 Hasil Uji Prediksi pada Program Web



Gambar 4. 9 Hasil Uji Prediksi pada Program Web



Gambar 4. 10 Hasil Uji Prediksi pada Program Web



Gambar 4. 11 Hasil Uji Prediksi pada Program Web



Gambar 4. 12 Hasil Uji Prediksi pada Program Web



Sistem Prediksi Berita Kesehatan Palsu Metode Naive Bayes

Teks :

jika disekitar kalian terdapat tower dengan hiasan seperti daun palem, itu adalah TOWER 5g dengan frekuensi yg dpt menimbulkan radiasi yg tdk baik utk tubuh manusia. Semakin pendek gelombang semakin tinggi frekuensinya, semakin tinggi frekuensi semakin berbahaya radiasinya.

Hasil Prediksi :

HOAX

Probabilitas Fake: 51.71%
Probabilitas True: 48.29%
UIN MALANG

Gambar 4. 13 Hasil Uji Prediksi pada Program Web



Sistem Prediksi Berita Kesehatan Palsu Metode Naive Bayes

Deteksi Berita Palsu

Teks :

KOMPAS.com - Air rebusan serai atau teh serai termasuk minuman herbal yang populer. Mengonsumsi air serai sebelum tidur dapat memberi efek relaksasi dan meredakan kecemasan. Efek serai tersebut dapat membantu orang yang memiliki masalah tidur, seperti insomnia sekaligus meningkatkan kualitas tidur. Untuk mengetahui lebih lanjut aneka potensi manfaat minum air serai sebelum tidur, simak ulasan berikut. Apa manfaat minum air serai sebelum tidur? Disarikan dari Healthline dan Tua Saude, berikut manfaat minum air serai sebelum tidur. Mengatasi susah tidur. Serai juga mengandung senyawa sirtral yang bertindak sebagai obat penenang yang dapat mengatasi insomnia (susah tidur) dan meningkatkan kualitas tidur. Relaksasi. Serai dapat membantu mengendurkan otot-otot tubuh dan memiliki efek penenang sehingga membuat tubuh lebih relaks. Meredakan kecemasan. Efek relaksasi dan penenang yang didapat dari teh serai dapat membantu meredakan stres dan kecemasan. Punya efek dekonjestan. Air serai juga dapat bertindak sebagai dekonjestan untuk mengatasi hidung tersumbat akibat flu, pilek, alergi, sinusitis, dan bronkitis. Mengatasi gejala PMS. Keluhan PMS kerap kali membuat wanita susah tidur. Nah, salah satu cara alami untuk mengatasi gejala PMS yaitu dengan mengonsumsi air rebusan serai. Air rebusan serai membantu meredakan kram, kembung, dan sensasi panas di perut akibat PMS sehingga wanita dapat tidur dengan nyenyak. Baca juga: Kapan Waktu yang Tepat untuk Minum Air Rebusan Serai? Mengurangi nyeri dan peradangan pada tubuh. Manfaat minum air serai berikutnya adalah mengurangi nyeri dan peradangan tubuh akibat asam urat, artritis atau radang sendi, rematik, dan efek kelelahan. Hal ini karena serai mengandung senyawa bioaktif yaitu myrcene dan citral yang bersifat analgesik. Meski begitu, air serai memiliki efek diuretik yang meningkatkan frekuensi urine sehingga bisa membuat tidur di malam hari menjadi terganggu. Untuk mencegah kondisi tersebut, hindari minum air serai berlebihan. Bagaimana cara membuat air rebusan serai? Cara membuat air serai cukup sederhana. Anda dapat menyiapkan satu gelas air putih dan merebusnya sampai mendidih. Kemudian, masukkan beberapa batang dan daun serai ke dalamnya. Tunggu selama 3-4 menit, lalu matikan api dan pindai air serai ke gelas atau cangkir. Anda juga dapat memadukan air serai dengan jahe, lemon, atau gula batu secukupnya untuk menambah kesegaran dan kenikmatan minuman herbal ini. Setelah menyimak manfaat minum air serai sebelum tidur dan cara menyiapkannya, Anda dapat mulai mengonsumsi minuman herbal ini untuk mendukung kesehatan.

Hasil Prediksi :

FAKTA

Probabilitas Fake: 42.85%
Probabilitas True: 57.15%
UIN MALANG

Gambar 4. 14 Hasil Uji Prediksi pada Program Web