

Rancang Bangun Model *Name Entity Recognition* Menggunakan Metode *Backpropagation* dalam Klasifikasi Berita Hoaks Seputar Vaksin Covid-19

Muhammad Fadil Khairunnas¹, Muhammad Arhami², M.Khadafi^{3*}

^{1,2,3} *Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA*

¹mfadilk22@gmail.com

²muhammad.arhami@pnl.ac.id

^{3*}mkhadafi@pnl.ac.id

Abstrak — Arus berita mengenai perkembangan Covid-19 mendominasi berbagai saluran informasi di Indonesia dalam 2 tahun terakhir, baik melalui media cetak atau media digital. Berbagai jenis berita yang berhubungan dengan Covid-19 terus beredar, tidak terkecuali berita *hoaks*. Salah satu berita hoaks yang banyak beredar adalah berita tentang vaksin Covid-19. Maraknya informasi yang berisi berita hoaks dan rumor yang tidak benar tentang vaksin Covid-19 di lingkungan masyarakat dapat memperburuk situasi pandemi. Saat ini belum ada sistem cerdas yang mampu mengklasifikasi beritahoaks seputar vaksin Covid-19. Untuk memaksimalkan pencegahan penyebaran berita hoaks seputar vaksin Covid-19 serta mengatasi permasalahan yang dihadapi, maka penulis merancang sistem klasifikasi berita hoaks seputar vaksin Covid-19 dengan pendekatan machine learning. Sistem yang dibangun dapat mengklasifikasi berita dengan kombinasi algoritma Name Entity Recognition (NER) dan Backpropagation. Dataset yang digunakan berjumlah 600 data berita vaksin Covid-19 yang diperoleh dari situs <https://turnbackhoax.id/> dan <https://www.kompas.com/> dengan kata kunci “vaksin covid”. Dataset dibagi menjadi dua, data latih dan data uji. Data latih dilakukan proses preprocessing kemudian digunakan dalam perancangan model. Data uji digunakan untuk mengevaluasi hasil perancangan model. Proses tersebut menghasilkan model machine learning dengan tingkat akurasi yang baik yaitu 97,62%. Dari hasil tersebut, sistem mampu mengklasifikasi berita seputar vaksin Covid-19.

Kata kunci — *Backpropagation*, Berita Hoaks, NER

Abstract — *The flow of news regarding the development of Covid-19 has dominated various information channels in Indonesia in the last 2 years, either through print or digital media. Various types of news related to Covid-19 continue to circulate, including hoax news. One of the most widely circulated hoax news is the news about the Covid-19 vaccine. The rise of information containing hoax news and untrue rumors about the Covid-19 vaccine in the community can worsen the pandemic situation. Currently, there is no intelligent system capable of classifying hoaxes about the Covid-19 vaccine. To maximize the prevention of the spread of hoax news about the Covid-19 vaccine and overcome the problems faced, the researchers designed a classification system for hoax news about the Covid-19 vaccine with a machine learning approach. The system built can classify news with a combination of the Backpropagation Name Entity Recognition (NER) algorithm. Dataset used is 600 Covid-19 vaccine news data obtained from the sites <https://turnbackhoax.id/> and <https://www.kompas.com/> with the keyword "vaksin covid". Dataset divided into two, training data and test data. The training data is preprocessed and then used in model design. Test data is used to evaluate the results of model design. This process produces a machine learning model with accuracy rate of 97,62%. From these results, the system is able to classify news texts about Covid-19 vaccine. From these results, the system is able to classify news texts about Covid-19 vaccine.*

Keywords — *Backpropagation*, Hoax news, NER

I. PENDAHULUAN

Pandemi Covid-19 telah melanda berbagai belahan dunia selama 2 tahun terakhir. Penyakit Covid-19 disebabkan oleh virus corona yang pertama kali muncul di kota Wuhan, China pada akhir tahun 2019. Wabah Covid-19 telah melanda berbagai belahan dunia selama 2 tahun terakhir tidak terkecuali Indonesia. Kasus pertama Covid-19 di Indonesia tercatat pada tanggal 02 Maret 2021. Presiden Joko Widodo pada tanggal tersebut mengumumkan kasus Covid-19 pertama di Indonesia, yakni perempuan berusia 31 tahun dan ibunya yang berusia 64 tahun. Sejak saat itu, lonjakan kasus Covid-19 di Indonesia terus meningkat setiap harinya dan belum

menunjukkan tanda-tanda akan berakhir. Berdasarkan data dari <https://covid19.go.id/> per tanggal 21 November 2021, jumlah total penderita Covid-19 mencapai 4.253.412. Dari jumlah tersebut 4.101.547 dinyatakan sembuh dan 143.739 dinyatakan meninggal. Arus berita mengenai perkembangan Covid-19 mendominasi berbagai saluran informasi di Indonesia dalam 2 tahun terakhir, baik melalui media cetak atau media digital. Berbagai jenis berita yang berhubungan dengan Covid-19 terus beredar, tidak terkecuali berita hoaks. Berita hoaks dengan tujuan dan narasinya masing-masing ikut meramalkan arus berita setiap harinya. Salah satu berita hoaks yang banyak beredar adalah berita tentang vaksin Covid-19. Maraknya informasi yang berisi berita hoaks dan rumor yang

tidak benar tentang vaksin Covid-19 di lingkungan masyarakat dapat memperburuk situasi pandemi. Lingkungan demografi yang berbeda, serta tingkat pendidikan masyarakat yang beragam menjadi peluang penyebaran berita hoaks seputar vaksin Covid-19.

Data dari Kementerian Komunikasi dan Informatika (Kemenkominfo) periode 23 Januari 2020 sampai 4 Desember 2021 menemukan adanya 2.015 isu hoaks Covid-19 yang tersebar melalui berbagai platform sosial media, seperti Facebook, Instagram, Twitter, Youtube, dan TikTok. Sementara itu hingga 20 Juli 2021, Kemenkominfo telah menurunkan 1.870 isu hoaks seputar vaksin Covid-19. Hoaks vaksin Covid-19 paling banyak tersebar di media sosial seperti Facebook dengan 1.691 konten hoaks, Twitter dengan 98 konten hoaks, Youtube dengan 41 konten hoaks, TikTok dengan 17 konten hoaks, dan Instagram dengan 11 konten hoaks[1]. Banyaknya informasi yang beredar menyebabkan masyarakat ragu dan berada diposisi yang tidak bisa membedakan hoaks atau tidak [2].

Berdasarkan uraian di atas, terdapat sejumlah permasalahan yang dihadapi oleh masyarakat mengenai penyebaran berita hoaks, khususnya berita hoaks seputar vaksin covid-19. Beberapa permasalahan yang dihadapi adalah sebagai berikut:

1. Penyebaran berita hoaks memperburuk keadaan dan situasi selama pandemi Covid-19.
2. Peredaran berita hoaks yang cepat dan masif belum diimbangi proses identifikasi berita yang cepat.
3. Belum adanya sistem peringatan dini yang dapat mendeteksi dan mengklasifikasi berita hoaks seputar vaksin Covid-19.

Untuk memaksimalkan pencegahan penyebaran berita hoaks seputar vaksin Covid-19 serta mengatasi permasalahan yang dihadapi, maka penulis mengembangkan sistem yang mampu mendeteksi dan mengklasifikasi berita hoaks seputar vaksin Covid-19 yaitu sistem klasifikasi berita hoaks seputar vaksin Covid-19 menggunakan metode Name Entity Recognition (NER) dan metode Backpropagation. Sistem diharapkan memberikan dampak positif dan menjadi solusi atas permasalahan yang diuraikan, di antaranya:

1. Mengklasifikasi berita hoaks seputar vaksin Covid-19 yang beredar di masyarakat.
2. Menjadi peringatan dini untuk mencegah dan memutus penyebaran berita hoaks seputar vaksin Covid-19.

Penelitian ini mengklasifikasi kebenaran pada berita tentang vaksin Covid-19. Named Entity Recognition digunakan untuk mengekstrak kata kunci dan informasi penting sehingga dapat mengenali entitas pada teks berita. Algoritma Backpropagation digunakan pada proses pembelajaran terstruktur sehingga dapat melakukan klasifikasi berita dengan baik. Hasil pembelajaran akan menghasilkan model machine learning yang digunakan untuk mengklasifikasi tingkat hoaks berita seputar vaksin Covid-19.

Natural Language Processing (NLP) adalah cabang ilmu dari *Artificial Intelligence* yang memungkinkan mesin dapat mengenali struktur bahasa alami manusia. NLP menerapkan kombinasi linguistik dan sains untuk mempelajari struktur kaidah bahasa sehingga mesin dapat mengerti, menganalisa, dan mengevaluasi makna yang terdapat dalam teks. NLP berkaitan dengan bagaimana komputer dapat memahami dan

mengerti bahasa manusia. Struktur bahasa manusia yang kompleks menjadi tantangan tersendiri bagi komputer untuk dapat mengerti dan memahami bahasa manusia.

NLP menggunakan dua teknik untuk membantu komputer memahami bahasa manusia, yaitu *Syntactic Analysis* dan *Semantic Analysis*. *Syntactic Analysis* melakukan pendekatan aturan tata bahasa dan penyusunan kalimat dalam mengidentifikasi struktur kalimat, penyusunan kalimat, dan keterkaitan antar kata dalam kalimat. Beberapa proses yang terdapat dalam *Syntactic Analysis* adalah:

1. Tokenisasi

Tokenisasi adalah pemisahan kata dalam teks lalu mengonversinya ke dalam angka. Setiap kata yang telah dikonversi disebut sebagai token. Tokenisasi memiliki peranan yang penting dalam *preprocessing* teks. Hal tersebut karena tokenisasi dapat memecah data teks yang tidak terstruktur menjadi bagian-bagian kecil yang dapat dikonversi ke dalam angka, sehingga dapat dipahami oleh mesin.

2. Stemming

Stemming adalah proses menemukan kata dasar dengan menghapus awalan dan imbuhan dari sebuah kata yang sudah mengalami perubahan morfologi [3]. Contohnya “membuka” dan “bukalah” menjadi “buka”. Proses ini bertujuan untuk menemukan makna walaupun sudah dalam bentuk yang berbeda [4].

3. Stopword Removal

Proses penghapusan kata-kata yang frekuensi kemunculannya tinggi namun tidak memiliki makna yang berarti, seperti “yang”, “pada”, “ke”, “di”, dan sebagainya.

Implementasi NLP mencakup area yang luas yang berhubungan dengan pemanfaatan bahasa manusia. NLP memungkinkan komputer berinteraksi dengan gaya bahasa manusia dan melakukan proses automasi seperti analisa sentimen, ekstraksi kata [5], klasifikasi teks [6].

Ekstraksi fitur adalah proses pengambilan ciri khas yang terdapat pada suatu data yang menggambarkan keseluruhan karakteristik data tersebut. Pada NLP, ekstraksi fitur diterapkan pada teks yang belum terstruktur sehingga menjadi terstruktur [7] untuk kemudian diberikan bobot sehingga dapat diketahui frekuensi kata. Sederhananya, ekstraksi fitur pada NLP berarti mengubah data tekstual menjadi data numerik, dalam hal ini vektor. Ekstraksi fitur merupakan bagian penting untuk pemrosesan data berdimensi tinggi dan sudah diterapkan pada *speech recognition*, bioteknologi, *targeted ads*, dan bidang lainnya [8]. Selama proses ekstraksi fitur, kata yang tidak berkorelasi dan tidak menghasilkan makna yang penting akan dihapus dan diabaikan [9].

Named Entity Recognition (NER) adalah sebuah teknik yang secara otomatis dapat mengidentifikasi serta mengklasifikasi kata pada teks atau korpus menurut kategorinya masing-masing., seperti lokasi, nama, lokasi, waktu, dan sebagainya. NER memungkinkan ekstraksi informasi penting dari sekumpulan data teks yang tidak terstruktur [10]. Penerapan NER dapat mencakup berbagai bidang, seperti mesin penerjemah, pengenalan suara, chat bots, dan mesin pencari [11].

Backpropagation adalah versi modifikasi artificial neural network yang cara kerjanya adalah dengan menyesuaikan bobot dan eror sembari bergerak mundur dari output ke input.

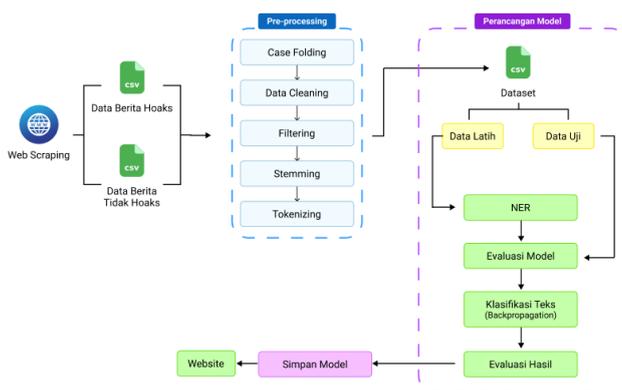
Ada dua bagian utama dari backpropagation, yaitu forward pass dan backward pass. Forward Pass adalah proses yang bergerak maju dari neuron input menuju neuron output. Proses forward pass akan menghasilkan nilai eror yang akan digunakan untuk penyesuaian bobot dan eror pada proses selanjutnya, yaitu backward pass. Nilai eror yang dihasilkan forward pass digunakan untuk memperbaiki diri dengan menyesuaikan bobot dan bias model. Penyesuaian bobot yang tepat mampu mengurangi nilai eror dan menjadikan model lebih efektif.

Proses forward pass dan backward pass dilakukan secara berulang hingga menghasilkan nilai bobot dan bias yang dapat menghasilkan nilai eror terendah. Cara kerja backpropagation dalam neural network adalah dengan menghitung gradien dari loss function terhadap semua parameter dengan cara mencari nilai turunan parsial dari fungsi tersebut. Algoritma backpropagation sangat efektif dalam pengenalan pola kompleks pada jaringan multilayer.

II. METODOLOGI PENELITIAN

A. Arsitektur Umum

Perancangan sistem klasifikasi terdiri dari beberapa tahapan seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur Umum

Gambar 1 menjelaskan arsitektur umum perancangan sistem. proses dimulai dari *scraping* untuk memperoleh data hoaks dan tidak hoaks Masing-masing data kemudian dilakukan proses *preprocessing*. Tahap *preprocessing* adalah tahapan untuk membersihkan dan mempersiapkan data yang belum terstruktur menjadi data terstruktur melalui beberapa tahap, yaitu, *case folding*, *data cleaning*, *filtering*, *stemming*, dan *tokenizing*. Dataset selanjutnya dipisah menjadi data latih dan data uji. Data latih digunakan untuk melatih algoritma, sedangkan data uji digunakan untuk menguji, mengevaluasi, dan mengukur kinerja model *machine learning*. Kemudian merancang model *machine learning* dan mengimplementasikannya di dalam website.

B. Dataset

Dataset yang digunakan dalam penelitian berjumlah 600 data. Dataset terbagi dua, yaitu data berita hoaks dan data berita tidak hoaks. Data berita hoaks berjumlah 300 data yang diambil dari situs www.turnbackhoax.id/ dengan kata kunci “vaksin covid”. Data berita tidak hoaks berjumlah 300 data

yang diambil dari situs www.kompas.com/ dengan kata kunci “vaksin covid”.

C. Preprocessing

Proses preprocessing dilakukan untuk membersihkan data yang masih belum terstruktur sehingga data tersebut siap digunakan untuk merancang model machine learning. Tahapan-tahapan yang dilakukan meliputi:

1) Case Folding

Case folding mengubah seluruh huruf besar menjadi huruf kecil. Tahap ini dilakukan agar membantu proses normalisasi dengan membuat semua karakter pada dataset seragam.

2) Data Cleaning

Kalimat pada dataset dibersihkan dari segala karakter yang tidak berguna dan dapat mempengaruhi performa seperti angka, emoji, tanda baca, link, dan simbol.

3) Filtering

Kata-kata umum dalam jumlah besar yang dianggap tidak penting *dihapus* dari dataset, seperti kata hubung “dan”, “atau”, “di”, “dari”, dan sebagainya. Kata-kata tersebut biasa disebut dengan *stopword*.

4) Stemming

Stemming dilakukan untuk mengembalikan kata yang sudah bertransformasi menjadi bentuk asalnya dengan cara menghilangkan awalan dan imbuhan. Pada proses ini, penulis menggunakan *library* Sastrawi.

5) Tokenisasi

Tokenisasi adalah proses memecah kalimat menjadi kepingan kata yang disebut token. Tokenisasi berguna saat ingin mengekstrak makna dari sebuah kata. Pada proses ini, penulis menggunakan function *word_tokenize* yang disediakan oleh *library* NLTK.

D. Perancangan Model Machine Learning

Tahap merancang model machine learning diawali dengan mendefinisikan data latih (x) dan data uji (y). Data latih (x) bersifat independent yang digunakan untuk melatih algoritma, sedangkan data uji (y) bersifat dependen yang digunakan untuk menguji, mengevaluasi, dan mengukur kinerja model machine learning. Proporsi data latih sebesar 70% dan data uji sebesar 30%. Sebelum melatih model *backpropagation*, dataset harus dikonversi ke bilangan numerik yang dapat dimengerti oleh komputer sebagai tahap persiapan representasi input. Tahapan-tahapan representasi *input* adalah:

1) Tokenizer

Tokenizer adalah proses mengonversi setiap kata menjadi bilangan numerik. Proses tokenizer dilakukan agar kata-kata tersebut dapat dipahami dan dimengerti oleh komputer.

2) Sekuensial

Sekuensial adalah proses mengubah setiap kalimat dalam teks menjadi sebuah larik bilangan yang di dalamnya berisi token yang nilainya sesuai dengan setiap kata pada sebuah teks.

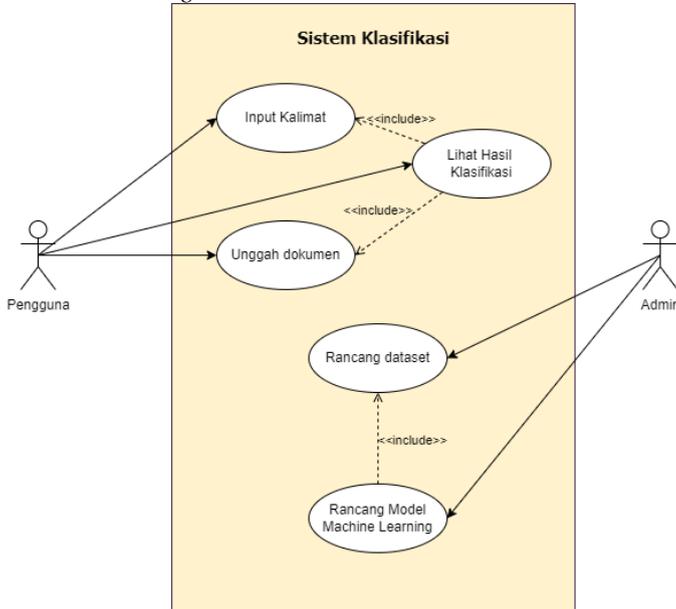
3) Padding

Backpropagation menerima input teks dengan dimensi dan ukuran yang sama. Oleh karena itu, dibutuhkan sebuah proses penyeragaman dimensi pada kata yang disebut

dengan padding.

E. Use Case Diagram

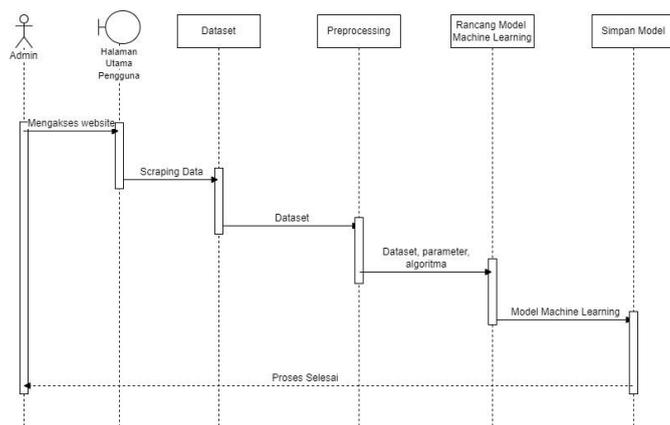
Pengguna dapat menginput teks kalimat atau mengunggah dokumen, serta melihat hasil klasifikasi keduanya. Admin bertugas untuk merancang dataset dan merancang model machine learning.



Gambar 2. Use Case Diagram

F. Sequence Diagram Admin

Admin melakukan *scraping* data untuk mengumpulkan dataset yang akan digunakan untuk pelatihan model. Setelah data terkumpul, *admin* kemudian melakukan proses *preprocessing* pada dataset tersebut. Selanjutnya, admin merancang model *machine learning* sembari menyesuaikan parameter-parameter yang dibutuhkan untuk menghasilkan model dengan performa yang baik. Langkah terakhir yang dilakukan *admin* adalah menyimpan *model machine learning* tersebut.

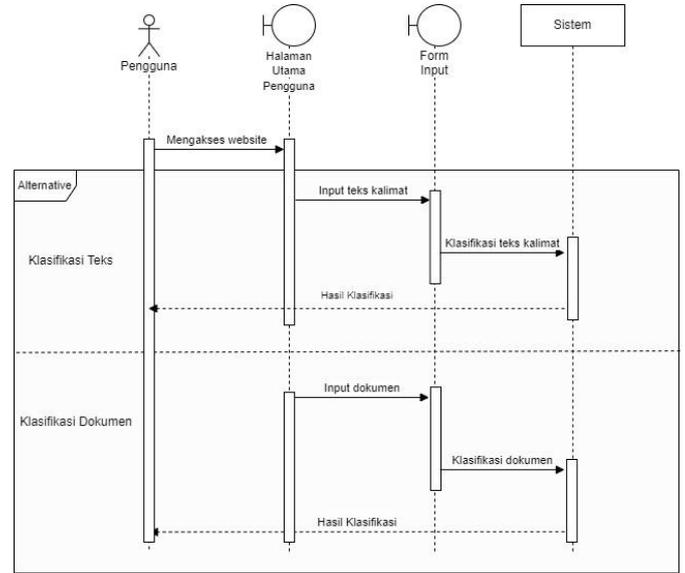


Gambar 3. Sequence Diagram Admin

G. Sequence Diagram Pengguna

Terdapat dua alternatif skenario, yaitu pengguna dapat melakukan klasifikasi teks dan klasifikasi dokumen. Pada proses klasifikasi teks, pengguna harus meng-input teks kalimat. Sistem kemudian melakukan klasifikasi teks kalimat

lalu menampilkan hasil klasifikasi teks kalimat tersebut. Pada proses klasifikasi dokumen, pengguna mengunggah sebuah dokumen berformat .txt. sistem kemudian melakukan klasifikasi pada dokumen dan menampilkan hasil klasifikasi dokumen tersebut.



Gambar 4. Sequence Diagram Pengguna

III. HASIL DAN PEMBAHASAN

A. Implementasi Proses

Implementasi proses menjabarkan prosedur dan tahapan yang terlibat berdasarkan perancangan model *machine learning* yang telah dijabarkan.

1) Web Scraping

Proses web scraping dilakukan dengan menggunakan pustaka Beautiful Soup 4. Total data yang dikumpulkan adalah 300 data berita hoaks dan 300 data berita tidak hoaks. Masing-masing dataset dikelompokkan menjadi 5 kolom yaitu "Judul", "Link", "Tanggal", "Berita", "Kategori". Dataset kemudian disimpan dalam file berformat .csv.

1	Judul	Tanggal	Link	Berita	Kategori
2	[SALAH] Vaksin	April 11, 2022	https://humbackhoax.id/2022/04/11/salah-vaksin-covid-19-mengandung-luciferasa-4		Hoaks
3	[SALAH] Kult B Maret 8, 2022	2022	https://humbacki.com/narasit/		Hoaks
4	[SALAH] Vaksin Februari 28, 2022	2022	https://humbacki.com/narasit/	====NARASI: Mohon yg sdh divaksin baca betul informasi ini. t	Hoaks
5	[SALAH] Anjuran Februari 28, 2022	2022	https://humbacki.com/narasit/	#ERROR!	Hoaks
6	[SALAH] Menengg Februari 24, 2022	2022	https://humbacki.com/narasit/	====NARASI: Ada berita kejadian yg fatal Bagi siapa saja yg ba	Hoaks
7	[SALAH] Mayoral Februari 14, 2022	2022	https://humbacki.com/narasit/	NARASI DO NOT LET THEM FOOL YOU Most of the world's pop	Hoaks
8	[SALAH] Aplikasi Februari 11, 2022	2022	https://humbacki.com/narasit/	NARASI:	Hoaks
9	[SALAH] Video " Februari 10, 2022	2022	https://humbackhoax.id/2022/02/10/salah-video-nigeria-minggu-23-01-2022-mengh		Hoaks
10	[SALAH] Atlet Ri Februari 3, 2022	2022	https://humbackhoax.id/2022/02/03/salah-atlet-ri-februari-3-2022	NARASI:	Hoaks
11	[SALAH] Micros Januari 26, 2022	2022	https://humbackhoax.id/2022/01/26/salah-microsoft-mengembangkan-implan-chip-4		Hoaks
12	[SALAH] Bill Gat Januari 23, 2022	2022	https://humbacki.com/narasit/	NARASI: BILL GATES CONFESIONBill Gates has demanded th	Hoaks
13	[SALAH] Atlet Te Januari 19, 2022	2022	https://humbacki.com/narasit/	NARASI:	Hoaks
14	[SALAH] Daftar J Januari 16, 2022	2022	https://humbacki.com/narasit/	NARASI:	Hoaks
15	[SALAH] Video F Januari 7, 2022	2022	https://humbacki.com/narasit/	NARASI:	Hoaks
16	[SALAH] Pemes Desember 31, 21	2021	https://humbacki.com/narasit/	NARASI:	Hoaks
17	[SALAH] Gejala Desember 12, 21	2021	https://humbacki.com/narasit/	NARASI:	Hoaks
18	[SALAH] Mandi I Desember 10, 21	2021	https://humbackhoax.id/2021/12/10/salah-mandi-detox-menggunakan-campuran-gi		Hoaks
19	[SALAH] Terpad November 30, 21	2021	https://humbackhoax.id/2021/11/30/salah-terpad-parasi-yang-mematikan-di-dalar		Hoaks
20	[SALAH] Bayi Di November 15, 21	2021	https://humbacki.com/narasit/	NARASI: Mengerikan bayi baru lahir dengan satu mata dan ekor.	Hoaks
21	[SALAH] Vaksin November 14, 21	2021	https://humbacki.com/narasit/	NARASI:	Hoaks
22	[SALAH] 13 Ana November 9, 20	2021	https://humbacki.com/narasit/	NARASI: 13 anak sekolah di Afrika Selatan meninggal setelah me	Hoaks

Gambar 5. Dataset Berita Hoaks

1	Judul	Link	Tanggal	Berita	Kategori
2	Warga Riau Antusias Ikuti Vaksinasi, Jumlah Ru...	http://regional.kompas.com	10/04/2022	Rumah Vaksin 24 Jam yang baru itu terletak di Rumah sakit (RS) Tampan di Jalan HF Benar	
3	Pria 61 Tahun Disuntik 90 Dosis Vaksin Covid-1...	http://www.kompas.com	09/04/2022	Dianisi Oddy Central, Selasa (5/4/2022), pria yang tidak disebutkan namanya tersebut Benar	
4	Usai Tarawih, Ratusan Jemaah Masjid Agung Tulu...	http://www.kompas.com	09/04/2022	"Ini merupakan kegiatan vaksinasi keempat selama Ramadhan ini," kata petugas MI Benar	
5	[HOAKS] Vaksin Covid-19 Mengandung Komponen Vi...	http://www.kompas.com	09/04/2022	Berdasarkan verifikasi Kompas.com sejauh ini, informasi ini tidak benar. Diabukan itu Benar	
6	Apa Jenis Vaksin Booster untuk Penerima Vaksin...	http://www.kompas.com	09/04/2022	Sejauh ini, vaksin Janssen hanya diberikan untuk kelompok usia 18 tahun ke atas. Val Benar	
7	Tinjau Produksi Vaksin Zifivax, BPOM: Pandemi ...	http://nasional.kompas.com	09/04/2022	"Saya selalu mengatakan bahwa di balik musibah pandemi ini pasti ada hal positif. Se Benar	
8	Update Corona 9 April 2022: Moderna Tarik Ratu...	http://www.kompas.com	09/04/2022	Berdasarkan catatan Worldometer, update kasus virus corona secara global adalah » Benar	
9	Dinkes DKI: 2.864.484 Orang Sudah Divaksinasi ...	http://megapoliteka.com	08/04/2022	Kepala Bidang Pencegahan dan Pengendalian Penyakit Dinkes DKI Jakarta Dwi Oktu Benar	
10	Jadwal, Lokasi dan Syarat Vaksin Covid-19 di K...	http://megapoliteka.com	08/04/2022	Hal ini dilakukan mengingat bahwa vaksin dosis penguat atau booster akan digunakan Benar	
11	Vaksin Booster - Update Cakupan Vaksinasi Covid-19 Dosis Kedua...	http://megapoliteka.com	08/04/2022	Vaksinasi massal diadakan di beberapa lokasi mulai dari vaksin dosis pertama, kedua Benar	
12	Bukti Sudah Vaksin Covid-19 Booster Jadi Syara...	http://regional.kompas.com	07/04/2022	Pada Kamis (7/4/2022) sejak 09.00 WIB, para warga yang telah melakukan vaksinasi Benar	
13	16 Lokasi Vaksin Booster untuk Penerima Vaksin...	http://megapoliteka.com	07/04/2022	"Kegiatan ini bertujuan untuk mendukung program pemerintah dan Kapolda Metro Ja; Benar	
14	Kemenkes Sebati: 2.864.484 Orang Sudah Divaksinasi ...	http://nasional.kompas.com	07/04/2022	"Yang harus kita pahami bersama bahwa booster ini bukan sesuatu yang meropatkan Benar	
15	Sertifikat Vaksin Booster Belum Muncul di Pedu...	http://www.kompas.com	07/04/2022	Bagi pelaku perjalanan yang telah memperoleh vaksinasi dosis ketiga atau booster, ti Benar	
16	Soal Vaksin Covid-19 Anak Usia di Bawah 6 Tahu...	http://nasional.kompas.com	06/04/2022	Charles menilai, sudah saatnya Bio Farma mengajukan EUA vaksin Covid-19 untuk u Benar	
17	Bio Farma: Pemerintah Belum Tugaskan Beli Vaks...	http://nasional.kompas.com	06/04/2022	Sebab, kata dia, kebutuhan vaksin dalam negeri tahun ini bersifat carry over atau bau Benar	
18	Kepala BPOM Dorong Industri Farmasi Swasta Pro...	http://nasional.kompas.com	06/04/2022	Manuru dia, hal tersebut untuk membantu PT Bio Farma yang dikenal menjadi satu-s Benar	
19	Cerita Tim Vakasinotor Polres Boyolali Terjeba...	http://regional.kompas.com	06/04/2022	"Kalau kita melihat dari status bulan April akan ada potensi sebesar 1,53 juta dosis itu Benar	
20	Vaksinasi Jadi Syarat Mudik, Peminat Vaksin Co...	http://rahayana.kompas.com	05/04/2022	Hal ini diduga terkait dengan ketentuan pemerintah pusat yang menjadikan vaksinasi Benar	

Gambar 6. Dataset Berita Tidak Hoaks

2) Case Folding

Proses *case folding* menjadikan semua karakter menjadi huruf kecil. Hasil proses *case folding* ditunjukkan pada Gambar 7.

	Judul	judul_bersih
0	Warga Riau Antusias Ikuti Vaksinasi, Jumlah Ru...	warga riau antusias ikuti vaksinasi, jumlah ru...
1	Pria 61 Tahun Disuntik 90 Dosis Vaksin Covid-1...	pria 61 tahun disuntik 90 dosis vaksin covid-1...
2	Usai Tarawih, Ratusan Jemaah Masjid Agung Tulu...	usai tarawih, ratusan jemaah masjid agung tulu...
3	[HOAKS] Vaksin Covid-19 Mengandung Komponen Vi...	[hoaks] vaksin covid-19 mengandung komponen vi...
4	Apa Jenis Vaksin Booster untuk Penerima Vaksin...	apa jenis vaksin booster untuk penerima vaksin...
5	Tinjau Produksi Vaksin Zifivax, BPOM: Pandemi ...	tinjau produksi vaksin zifivax, bpm: pandemi ...

Gambar 7. Hasil Case Folding

3) Data Cleaning

Data dibersihkan dari segala karakter yang tidak berguna dan dapat mempengaruhi performa seperti angka, emoji, tanda baca, link, simbol. Hasil proses *data cleaning* ditunjukkan pada Gambar 8.

	Judul	judul_bersih
5	Tinjau Produksi Vaksin Zifivax, BPOM: Pandemi ...	tinjau produksi vaksin zifivax bpm pandemi co...
6	Update Corona 9 April 2022: Moderna Tarik Ratu...	update corona april moderna tarik ratusan ri...
7	Dinkes DKI: 2.864.484 Orang Sudah Divaksinasi ...	dinkes dki orang sudah divaksinasi covid d...
8	Jadwal, Lokasi dan Syarat Vaksin Covid-19 di K...	jadwal lokasi dan syarat vaksin covid di kota...
9	Vaksin Booster Jabodetabek 8 April 2022	vaksin booster jabodetabek april
10	UPDATE: Cakupan Vaksinasi Covid-19 Dosis Kedua...	update cakupan vaksinasi covid dosis kedua ...
11	Bukti Sudah Vaksin Covid-19 Booster Jadi Syara...	bukti sudah vaksin covid booster jadi syarat ...

Gambar 8. Hasil Data Cleaning

4) Filtering

Data dibersihkan dari kata yang tidak memiliki makna berarti dan berjumlah banyak.

	Judul	judul_bersih
14	Sertifikat Vaksin Booster Belum Muncul di Pedu...	sertifikat vaksin booster muncul pedulilindung...
15	Soal Vaksin Covid-19 Anak Usia di Bawah 6 Tahu...	vaksin covid anak usia bio farma kumpulan data
16	Bio Farma: Pemerintah Belum Tugaskan Beli Vaks...	bio farma pemerintah tugaskan beli vaksin covid
17	Kepala BPOM Dorong Industri Farmasi Swasta Pro...	kepala bpm dorong industri farmasi swasta pro...
18	Bio Farma: 1,53 Juta Dosis Vaksin Covid-19 Ber...	bio farma juta dosis vaksin covid berpotensi k...
19	Cerita Tim Vakasinotor Polres Boyolali Terjeba...	cerita tim vakasinotor polres boyolali terjeba...
20	Vaksinasi Jadi Syarat Mudik, Peminat Vaksin Co...	vaksinasi syarat mudik peminat vaksin covid ke...

Gambar 9. Hasil Filtering

5) Stemming

Kata yang sudah mengalami transformasi, seperti penambahan awalan dan/atau akhiran, dikembalikan ke bentuk dasarnya. Hasil dari proses *stemming* ditunjukkan pada Gambar 10.

	Judul	judul_bersih
30	Stok Vaksin Booster di Puskesmas Sempat Menipi...	stok vaksin booster puskesmas tip dinkes depok...
31	Koster: Tidak Ada Vaksin Kedaluwarsa di Bali	koster vaksin kedaluwarsa bal
32	63,4 Juta Dosis Vaksin Hibah Akan Diterima Ind...	juta dosis vaksin hibah terima indonesia pfize...
33	Pemprov DKI: 2.459.498 Orang Sudah Divaksinasi...	pemprov dki orang vaksinasi covid dosis tiga
34	Evaluasi Pemerintah soal 19,3 Juta Vaksin Covi...	evaluasi perintah juta vaksin covid hibah keda...
35	[POPULER JABODETABEK] Lokasi Vaksin Jabodetabe...	populer jabodetabek lokasi vaksin jabodetabek ...

Gambar 10. Hasil Stemming

6) Tokenizing

Tokenizing merupakan proses terakhir dalam tahapan *preprocessing*. Melalui proses *tokenizing* kalimat dipecah menjadi penggalan kata. Keluaran proses *tokenizing* ditunjukkan pada Gambar 11.

	Judul	judul_bersih
30	Stok Vaksin Booster di Puskesmas Sempat Menipi...	[stok, vaksin, booster, puskesmas, tip, dinkes...
31	Koster: Tidak Ada Vaksin Kedaluwarsa di Bali	[koster, vaksin, kedaluwarsa, bal]
32	63,4 Juta Dosis Vaksin Hibah Akan Diterima Ind...	[juta, dosis, vaksin, hibah, terima, indonesia...
33	Pemprov DKI: 2.459.498 Orang Sudah Divaksinasi...	[pemprov, dki, orang, vaksinasi, covid, dosis,...
34	Evaluasi Pemerintah soal 19,3 Juta Vaksin Covi...	[evaluasi, perintah, juta, vaksin, covid, hiba...
35	[POPULER JABODETABEK] Lokasi Vaksin Jabodetabe...	[populer, jabodetabek, lokasi, vaksin, jabodet...

Gambar 11. Hasil Tokenizing

7) Named Entity Recognition (NER)

Named entity recognition (NER) mampu mengklasifikasi entitas pada teks ke dalam kategori yang telah ditentukan. Pada penelitian ini, penulis mengelompokkan entitas ke dalam kategori "ORGANISASI", "ORANG", "LOKASI", "PENYAKIT", dan "VAKSIN". Tahapan proses *Named Entity Recognition* (NER) dilakukan menggunakan *library SpaCy*. Sebelum melakukan proses *training*, dataset terlebih dahulu dilakukan anotasi kemudian disesuaikan dengan format yang dapat diterima *SpaCy*.

```
{
  "classes": [
    "ORGANISASI", "ORANG", "LOKASI", "PENYAKIT", "VAKSIN"
  ],
  "annotations": [
    [
      "Vaksin Covid-19 Dapat Sebabkan AIDS Sampai Cacar Monyet",
      {
        "entities": [
          [0, 6, "VAKSIN"],
          [7, 15, "PENYAKIT"],
          [23, 28, "PENYAKIT"]
        ]
      }
    ]
  ]
}
```

Gambar 12. Hasil Anotasi Dataset

Gambar 12 menunjukkan dataset yang sudah dilakukan proses anotasi dan penyesuaian struktur yang dapat diterima *SpaCy*. Pada contoh di atas, karakter dengan indeks 0 sampai 6 merupakan kata yang tergolong entitas "VAKSIN", karakter indeks 7 sampai 15 merupakan kata yang tergolong entitas "PENYAKIT", dan seterusnya. *Pustaka SpaCy* tidak menyediakan model berbahasa Indonesia. Namun *SpaCy* memungkinkan untuk mengembangkan model sesuai dengan kebutuhan. Langkah-langkah membuat model NER pada *SpaCy* adalah:

- Membuat model kosong berbahasa Indonesia.
- Menambahkan *pipeline* NER pada model.
- Menambahkan label anotasi yang sudah dipersiapkan ke dalam *pipeline* NER.
- Menyisihkan modul selain "NER" pada *pipeline*.
- Sebelum proses pembelajaran dimulai, data latihan diacak agar model tidak menggeneralisasi berdasarkan urutan data tertentu.

- Melatih model sebanyak 300 iterasi. Pada setiap iterasi, model akan memperbarui dan menyesuaikan bobotnya hingga nilai *loss* mengecil dan akurasi meningkat.

Presiden Joko Widodo (**Jokowi** **ORANG**) kembali menegaskan pentingnya **vaksin** **VAKSIN** Covid-19 **PENYAKIT** . Meskipun kondisi pandemi Covid-19 **PENYAKIT** di Tanah Air saat ini melandai, **masyarakat** **ORANG** diminta untuk segera melakukan vaksinasi dosis lengkap dan booster **VAKSIN** .

Gambar 13. Hasil Named Entity Recognition

8) Backpropagation

Tahap merancang model *machine learning* diawali dengan mendefinisikan data latih (*x*) dan data uji (*y*). Data latih (*x*) bersifat independent yang digunakan untuk melatih algoritma, sedangkan data uji (*y*) bersifat dependen yang digunakan untuk menguji, mengevaluasi, dan mengukur kinerja model *machine learning*. Proporsi data latih adalah 70% dari data, yaitu berjumlah 420 data dan proporsi data uji adalah 30% dari data, yaitu berjumlah 180 data.

Total terdapat tiga *layer* pada model, yaitu satu *input layer* dengan jumlah 128 *neuron*, satu *hidden layer* dengan jumlah 24 *neuron*, dan satu *output layer* dengan jumlah 1 *neuron*. Setiap *layer* menggunakan fungsi aktivasi yang sama, yaitu sigmoid. Pemilihan fungsi aktivasi sigmoid dilakukan karena sigmoid memiliki rentang 0 hingga 1 sehingga sesuai digunakan untuk klasifikasi 2 kelompok data. Ringkasan struktur *layer* ditunjukkan Gambar 14.

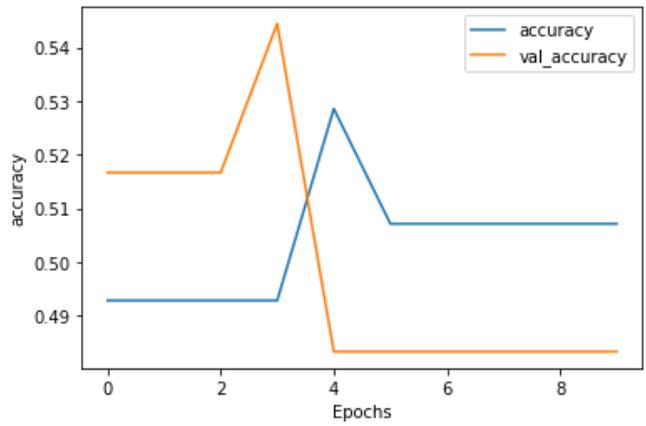
Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 20, 32)	47392
flatten_2 (Flatten)	(None, 640)	0
dense_6 (Dense)	(None, 128)	82048
dense_7 (Dense)	(None, 16)	2064
dense_8 (Dense)	(None, 1)	17

Total params: 131,521
Trainable params: 131,521
Non-trainable params: 0

Gambar 14. Ringkasan Struktur Layer Model

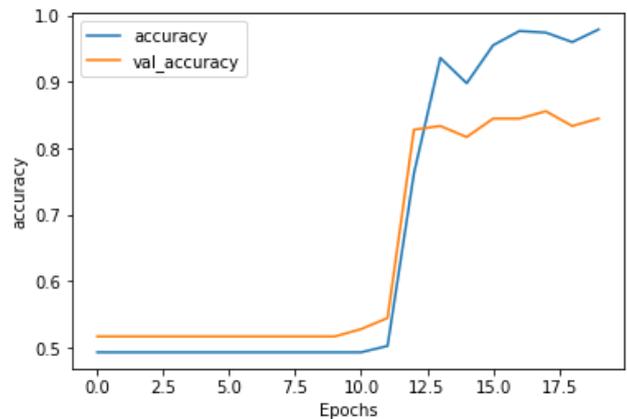
9) Pengujian dan Evaluasi

Pada pengujian ini penulis menguji dengan 10 *epoch*, 20 *epoch*, dan 40 *epoch*. Gambar 15 menunjukkan hasil akurasi 10 *epoch*. Gambar 16 menunjukkan hasil akurasi 20 *epoch*. Gambar 17. menunjukkan hasil akurasi 40 *epoch*.



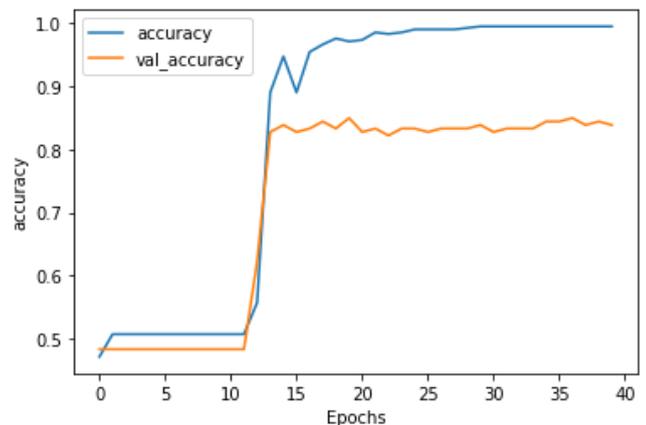
Gambar 15. Grafik Akurasi 10 epoch

Tingkat akurasi yang dihasilkan 10 *epoch* fluktuatif dan berhenti di angka 50,1%.



Gambar 16. Grafik Akurasi 20 epoch

Tingkat akurasi yang dihasilkan 20 *epoch* cenderung stagnan dari *epoch* 1 sampai 10 lalu mengalami kenaikan yang signifikan ketika melewati *epoch* ke-10. Tingkat akurasi yang dihasilkan adalah 97,62% dan validasi akurasi 83,33%.



Gambar 17. Grafik Akurasi 40 epoch

Tingkat akurasi dan validasi akurasi yang dihasilkan 40 *epoch* mengalami kenaikan yang signifikan ketika melewati *epoch* ke-10 dan cenderung stagnan sampai *epoch* 40. Namun terdapat jarak antara nilai akurasi dan validasi akurasi yang

mengindikasikan adanya *overfitting* pada model. *overfitting*, yaitu keadaan model terlalu fokus untuk mempelajari seluruh variabel pada data, termasuk *noise*, sehingga proses klasifikasi tidak berjalan baik.

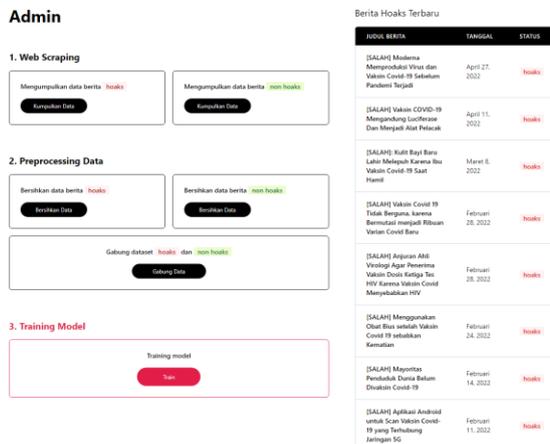
Berdasarkan hasil pengujian terhadap ketiga percobaan *epoch* tersebut, penulis menggunakan 20 *epochs*. Pemilihan 20 *epoch* dilakukan karena memiliki tingkat akurasi yang lebih tinggi dibanding 10 *epoch* dan kecenderungan *overfitting* yang lebih rendah dibanding 40 *epoch*. Kecenderungan *overfitting* yang rendah menandakan bahwa model *backpropagation* memiliki performa yang lebih stabil.

B. Implementasi Sistem

Implementasi sistem menjabarkan implementasi dari perancangan sistem.

1) Halaman Admin

Halaman admin berfungsi untuk mengumpulkan data, preprocessing data, dan training model. Setiap langkah-langkah disediakan tombol dan urutan masing-masing. Halaman admin ditunjukkan oleh Gambar 18.

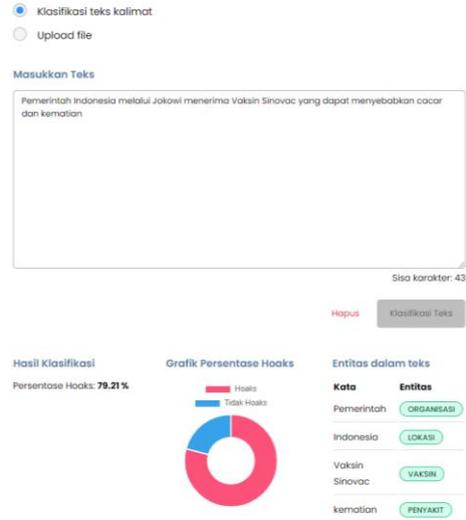


Gambar 18. Halaman Admin

2) Halaman Klasifikasi Teks

Halaman ini memungkinkan pengguna untuk menginput teks kalimat dan melihat hasil klasifikasinya disertai dengan visualisasi data. Halaman klasifikasi teks ditunjukkan oleh Gambar 19.

Klasifikasi Hoaks Seputar Vaksin Covid-19

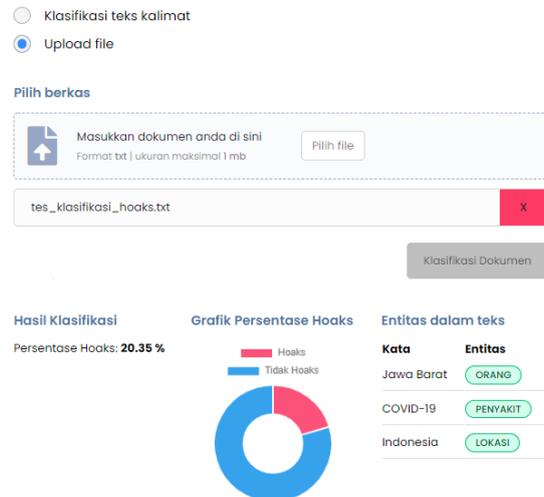


Gambar 19. Halaman Klasifikasi Teks

3) Halaman Klasifikasi Dokumen

Halaman klasifikasi dokumen memungkinkan pengguna untuk meng-upload sebuah dokumen berformat .txt untuk kemudian melihat hasil klasifikasinya disertai dengan visualisasi data.

Klasifikasi Hoaks Seputar Vaksin Covid-19



Gambar 20. Halaman Klasifikasi Dokumen

IV. KESIMPULAN

Setelah melalui proses pengujian dan implementasi, dapat disimpulkan bahwa sistem mampu mengklasifikasi teks singkat dengan jumlah kurang dari 144 karakter, seperti judul berita. Namun, sistem belum optimal dalam mengklasifikasi teks panjang, seperti isi berita. Klasifikasi berita hoaks seputar vaksin Covid-19 menggunakan metode *Backpropagation* menghasilkan akurasi sebesar 97,62% dan *val_accuracy* sebesar 83,33% menggunakan *hyperparameters epochs* 20. Proses *Preprocessing* dataset yang tidak optimal dan masih menyisakan banyak *noise* memengaruhi akurasi yang

diperoleh saat mengimplementasikan *Backpropagation*. Semakin banyak *noise* pada dataset mengakibatkan penurunan akurasi.

REFERENSI

- [1] D. H. Jayani, "Kemenkominfo Turunkan 1.870 Konten Hoaks Vaksin Covid-19 di Media Sosial | Databoks." <https://databoks.katadata.co.id/datapublish/2021/07/21/kemenkominfo-turunkan-1870-konten-hoaks-vaksin-covid-19-di-media-sosial> (accessed Dec. 05, 2021).
- [2] C. Juditha, "Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya Hoax Communication Interactivity in Social Media and Anticipation," 2018.
- [3] M. Aziz Assuja, "ANALISIS SENTIMEN TWEET MENGGUNAKAN BACKPROPAGATION NEURAL NETWORK," *Jurnal TEKNOINFO*, vol. 10, no. 2, pp. 1693–1703, 2016.
- [4] A. Tabitha, B. Panjaitan, and D. I. Santoso, "Deteksi Hoaks Pada Berita Berbahasa Indonesia Seputar COVID-19," *Jurnal Format*, vol. 10, no. 1, pp. 76–85, 2021.
- [5] S. Kim, S. Choi, and J. Seok, "Keyword Extraction in Economics Literatures using Natural Language Processing," *International Conference on Ubiquitous and Future Networks, ICUFN*, vol. 2021-August, pp. 75–77, Aug. 2021, doi: 10.1109/ICUFN49451.2021.9528546.
- [6] C. Wintang Kencana, E. B. Setiawan, and I. Kurniawan, "Hoax Detection on Twitter using Feed-forward and Back-propagation Neural Networks Method," *masa berlaku mulai*, vol. 1, no. 3, pp. 648–654, 2017.
- [7] I. Budiman, R. Faisal, and D. T. Nugrahadi, "Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah," *Jurnal Komputasi*, vol. 8, no. 1, pp. 62–69, Apr. 2020.
- [8] D. Wang, J. Su, and H. Yu, "Feature extraction and analysis of natural language processing for deep learning english language," *IEEE Access*, vol. 8, pp. 46335–46345, 2020, doi: 10.1109/ACCESS.2020.2974101.
- [9] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," *EURASIP J Wirel Commun Netw*, vol. 2017, no. 1, Dec. 2017, doi: 10.1186/S13638-017-0993-1.
- [10] I. Roldos, "Introduction to Entity Extraction: What Is It And How It Works," Nov. 09, 2020. <https://monkeylearn.com/blog/entity-extraction/> (accessed Dec. 06, 2021).
- [11] S. Kannan *et al.*, "Big Data Analytics for Social Media," *Big Data: Principles and Paradigms*, pp. 63–94, Jun. 2016, doi: 10.1016/B978-0-12-805394-2.00003-9.