

Skema Pendanaan: Dosen Pemula

LAPORAN AKHIR PENELITIAN



IDENTIFIKASI CERDAS HOAKS DENGAN METODE NAIVE BAYES DAN DECISION TREE

TIM PENELITI:

Ketua : Teja Endra Eng Tju, S.T., M.Kom. (190046)
Anggota : Arman Yusuf, S.Kom., M.M. (160015)

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS BUDI LUHUR

AGUSTUS 2023

HALAMAN PENGESAHAN

Judul Penelitian : Identifikasi Cerdas Hoaks dengan Metode Naive Bayes dan Decision Tree

Bidang Penelitian : ICT

Ketua Peneliti

a. Nama Lengkap : Teja Endra Eng Tju, S.T., M.Kom.
b. NIP/NIDN/ID-SINTA : 190046/0407127201/6713276
c. Jabatan Fungsional : Asisten Ahli
d. Program Studi : Sistem Informasi
e. Nomor HP : +628998909707
f. Alamat e-mail : teja.endraengtju@budiluhur.ac.id

Anggota Peneliti

a. Nama Lengkap : Arman Yusuf, S.Kom, M.M.
b. NIP/NIDN/ID-SINTA : 160015/0330087506/6693713

Mahasiswa yang terlibat

a. Nama Lengkap : Muhammad Umar Shalih
b. NIM : 2011501091

Lama Penelitian : 4 bulan

Biaya Penelitian : Sumber Universiastasi Budi Luhur Rp. 5 Juta

Jakarta, 17 Agustus 2023

Peneliti,

Mengetahui,
Dekan Fakultas Teknologi Informasi



(Dr. Ir. Denti Mahdiana, S.Kom, M.M, M.Kom)
NIP. 960012


(Teja Endra Eng Tju, S.T., M.Kom.)
NIP. 190046

Menyetujui,

Direktur Riset dan Pengabdian Kepada Masyarakat



(Dr. Krisna Adiyarta)
NIP. 890001

No. Registrasi	:	0	3	7	0	1	LPJ	0	9	2	3
Tanggal	:	0	5	0	9	2	3	Paraf:			

RINGKASAN

Perkembangan teknologi informasi dan media sosial telah memberikan akses luas untuk menyebarkan informasi secara cepat dan mudah. Namun, dampak negatifnya adalah penyebaran hoaks yang dapat menyesatkan dan mempengaruhi opini publik. Identifikasi hoaks kompleks dan menantang karena berbagai alasan, seperti keanekaragaman hoaks dalam bentuk dan konten, pola perubahan yang cepat, dan kesulitan verifikasi. Untuk mengatasi permasalahan ini, dilakukan penelitian menggunakan pemodelan kecerdasan buatan dengan teknik klasifikasi. Penelitian ini penting untuk membantu masyarakat terhindar dari dampak buruk hoaks, seperti penipuan, penghinaan, dan konflik sosial.

Penelitian ini terdiri dari sembilan langkah atau tahapan, yaitu tinjauan pustaka, pengumpulan data, preprocessing data, pemilihan fitur, penulisan proposal, pembangunan model, pengujian dan evaluasi kinerja model, analisis hasil dan interpretasi, serta penulisan laporan dan publikasi. Penelitian ini bertujuan untuk mengidentifikasi hoaks dalam berita dengan menggunakan pemodelan kecerdasan buatan dan teknik klasifikasi. Metodologi penelitian diadopsi dari *CRISP-DM* dan SKKNI No. 299 tahun 2020 yang disesuaikan dengan kebutuhan penelitian sehingga menjadi lima tahapan yaitu data *understanding*, data *preparation*, *modeling*, *evaluation*, dan *deployment*.

Data diperoleh dari Mafindo dan digunakan sebanyak 9.756 data yang dibagi menjadi 7.804 data latih dan 1.952 data uji. Terdapat enam fitur yaitu sumber, kapital, *keyword*, sentimen, *factcheck*, dan klasifikasi sebagai label supervisi. Dua fitur sentimen dan *factcheck* dibangun dengan metode *Multinomial Naïve Bayes*, selanjutnya dilakukan pemodelan pada *dataset* dengan metode *Decision Tree*. Pemodelan dilakukan pula dengan variasi kuantitas *dataset* sebanyak 2.000, 4.000, 6.000, 8000, juga dengan perbandingan masalah *imbalance dataset*. Hasil pemodelan dengan teknik *Confusion Matrix* diperoleh akurasi 93,5% dan skor F1 0,935 dan diperoleh bahwa *imbalance dataset* tidak terlalu berpengaruh pada hasil akurasi dan skor F1 namun memberikan kestabilan model dalam hal kuantitas besarnya data dengan label tertentu.

Kata kunci: berita_bohong; *machine_learning*; *confusion_matrix*; identifikasi_hoaks; kecerdasan_artifisial.

PRAKATA

Puji syukur kami panjatkan ke hadirat Allah yang Maha Kuasa, karena atas berkat-Nya kami dapat menyelesaikan penelitian ini dengan judul "Identifikasi Cerdas Hoaks dengan Metode *Naive Bayes* dan *Decision Tree*". Penelitian ini kami lakukan sebagai salah satu bentuk kontribusi kami dalam menghadirkan solusi dalam menghadapi tantangan penyebaran hoaks di era digital yang semakin kompleks.

Penyusunan penelitian ini tidak lepas dari arahan, dukungan, dan kerjasama berbagai pihak. Oleh karena itu, kami mengucapkan rasa terima kasih yang setinggi-tingginya kepada:

1. Ketua dan pengurus Yayasan Budi Luhur Chakti atas arahan dan dukungan yang diberikan sehingga terlaksananya program penelitian ini;
2. Rektor dan seluruh jajaran Universitas Budi Luhur atas kesempatan dan fasilitas yang diberikan untuk melaksanakan penelitian ini;
3. Keluarga kami atas doa, dukungan, dan semangat yang memberi kami kekuatan dalam menyelesaikan penelitian ini.
4. Teman-teman seperjuangan yang telah memberikan dukungan, motivasi, dan bantuan selama penelitian berlangsung.

Penelitian ini merupakan bagian dari upaya kami untuk menghadirkan sumbangsih ilmiah yang bermanfaat bagi perkembangan ilmu pengetahuan dan teknologi di Universitas Budi Luhur dan masyarakat luas. Kami menyadari bahwa penelitian ini belum sempurna dan masih memiliki keterbatasan, namun kami berharap bahwa hasil dari penelitian ini dapat memberikan wawasan yang berguna dalam mengatasi permasalahan penyebaran hoaks di era digital.

Semoga penelitian ini dapat memberikan manfaat dan menjadi bahan rujukan bagi penelitian selanjutnya. Akhir kata, kami berharap semoga penelitian ini dapat memberikan kontribusi positif dan menjadi bagian dari upaya mencerdaskan bangsa.

Jakarta, 17 Agustus 2023

Tim Peneliti dan Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN PENGESAHAN	ii
RINGKASAN.....	iii
PRAKATA.....	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN.....	ix
BAB 1. PENDAHULUAN	1
BAB 2. TINJAUAN PUSTAKA	2
BAB 3. METODE PENELITIAN	4
3.1. Data Understanding	5
3.1.1. Data Collection	5
3.1.2. Data Exploration	6
3.2. Data Preparation	6
3.2.1. Data Selection	6
3.2.2. Data Construction	6
3.3. Modeling.....	7
3.3.1. Data Testing Designed.....	7
3.4. Evaluation	8
3.4. Deployment.....	8
3.4.1. Save Model	8
3.4.2. User Interface.....	8
BAB 4. HASIL DAN ANALISA	9
4.1. Data Preparation	9
4.2. Modeling.....	13
4.3. Evaluation	15
4.4. Deployment.....	16
4.5. Analisa	17
BAB 5. KESIMPULAN DAN SARAN	19

DAFTAR PUSTAKA	20
LAMPIRAN-LAMPIRAN	25

DAFTAR TABEL

Tabel 1. Ringkasan tinjauan pustaka	2
Tabel 2. Percobaan dengan Variasi Jumlah Data.....	8
Tabel 3. Hasil Pengujian Tanpa SMOTE	15
Tabel 4. Hasil Pengujian dengan SMOTE.....	15

DAFTAR GAMBAR

Gambar 1. Tahapan penelitian	4
Gambar 2. Tahapan Penelitian	5
Gambar 3. Sampel Data dari Mafindo	5
Gambar 4. Pemilihan dan Konstruksi Fitur	6
Gambar 5. Deployment Environment	8
Gambar 6. Algoritma Data Fitur Sumber	9
Gambar 7. Algoritma Data Fitur Kapital	9
Gambar 8. Algoritma Data Fitur <i>Keyword</i>	10
Gambar 9. Algoritma <i>Text Preprocessing</i>	10
Gambar 10. Algoritma Pemodelan Data Sentimen <i>GitHub</i>	10
Gambar 11. Algoritma Data Fitur Sentimen	11
Gambar 12. Algoritma Pemodelan Data <i>Text Preprocessing</i>	11
Gambar 13. Algoritma Data Fitur <i>FackCheck</i>	12
Gambar 14. Pengkodean dengan <i>VSCode</i>	12
Gambar 15. Sampel Dataset Penelitian	13
Gambar 16. Algoritma Pemodelan	13
Gambar 17. Algoritma Pengujian	14
Gambar 17. <i>Library</i> yang Digunakan dengan <i>Python</i>	14
Gambar 19. Visualisasi <i>Accuracy</i> dan <i>F1 Score</i>	16
Gambar 20. Isian Konten Berita Baru	16
Gambar 21. Pilihan Sumber Berita Baru	17
Gambar 22. Hasil Prediksi suatu Berita	17

DAFTAR LAMPIRAN

Lampiran 1. Realisasi Biaya	25
Lampiran 2. Instrumen Penelitian.....	26
Lampiran 3. Biodata Peneliti	27
Lampiran 4. Surat Perjanjian Kontrak Penelitian	31
Lampiran 5. Catatan Harian	33
Lampiran 6. Artikel Ilmiah	34

BAB 1. PENDAHULUAN

Perkembangan teknologi informasi dan media sosial melalui internet telah memberikan akses yang luas bagi individu untuk menyebarkan informasi secara cepat dan mudah (Kumparan.com, 2021). Namun, dampak negatif dari kemudahan ini adalah penyebaran hoaks (berita bohong) yang dapat menyesatkan dan mempengaruhi opini publik. Dari perspektif hukum pidana, pelaku penyebaran hoaks juga dapat di pidana (Sarwan, 2023), sebagai pertanggungjawaban hukum individu yang sengaja menyebarkan atau menciptakan berita palsu dengan niat menyesatkan dan merugikan orang lain.

Identifikasi hoaks merupakan hal yang kompleks dan menantang. Beberapa permasalahan yang sering dihadapi dalam identifikasi hoaks adalah: keanekaragaman hoaks dalam berbagai bentuk dan konten melalui berbagai platform dan saluran komunikasi; pola perubahan yang cepat dalam narasi, strategi, dan teknik manipulasi informasi; kecepatan penyebaran dapat menjangkau ribuan atau bahkan jutaan orang dalam waktu singkat; penggunaan teknologi seperti kecerdasan buatan, pemrosesan bahasa alami, atau teknik manipulasi gambar dan video yang canggih; kesulitan verifikasi terutama dalam konteks berita yang cepat berubah atau kontroversial, mencari sumber terpercaya, memverifikasi fakta, dan mendapatkan konteks yang tepat membutuhkan waktu dan upaya yang signifikan; tantangan skala dengan pendekatan skala besar, termasuk penggunaan teknologi dan kecerdasan buatan, untuk mengatasi tantangan ini.

Untuk mengatasi permasalahan identifikasi hoaks, dalam penelitian digunakan pemodelan kecerdasan artifisial dengan teknik klasifikasi. Berita atau informasi yang diolah terbatas pada bentuk teks bahasa Indonesia yang berasal dari media sosial dan berita melalui internet. Tujuan khusus penelitian adalah didapatkan model klasifikasi dengan hasil pengukuran tingkat akurasi dan perbandingan hasil prediksi terhadap fakta dari hoaks. Penelitian ini penting dilakukan untuk mendapatkan algoritma dalam identifikasi hoaks sehingga dapat diaplikasikan dalam membantu masyarakat agar terhindar dari dampak buruk hoaks yaitu penipuan, fitnah, penghinaan, pencemaran nama baik, pelanggaran hukum informatika, polarisasi politik, konflik sosial, persaingan bisnis sehingga menciptakan sentimen negatif terhadap kelompok atau individu tertentu, serta memperdalam kesenjangan dan ketegangan dalam masyarakat.

BAB 2. TINJAUAN PUSTAKA

Tinjauan pustaka dilakukan dengan mencari, membaca, dan mempelajari artikel jurnal yang telah dilakukan oleh para peneliti. Artikel yang ditinjau difokuskan pada teknik klasifikasi hoaks menggunakan *machine learning* (Rebala, Ravi and Churiwala, 2019; Potentia Analytics, 2021), dengan ringkasan hasil tinjauan ditunjukkan pada Tabel 1.

Tabel 1. Ringkasan tinjauan pustaka

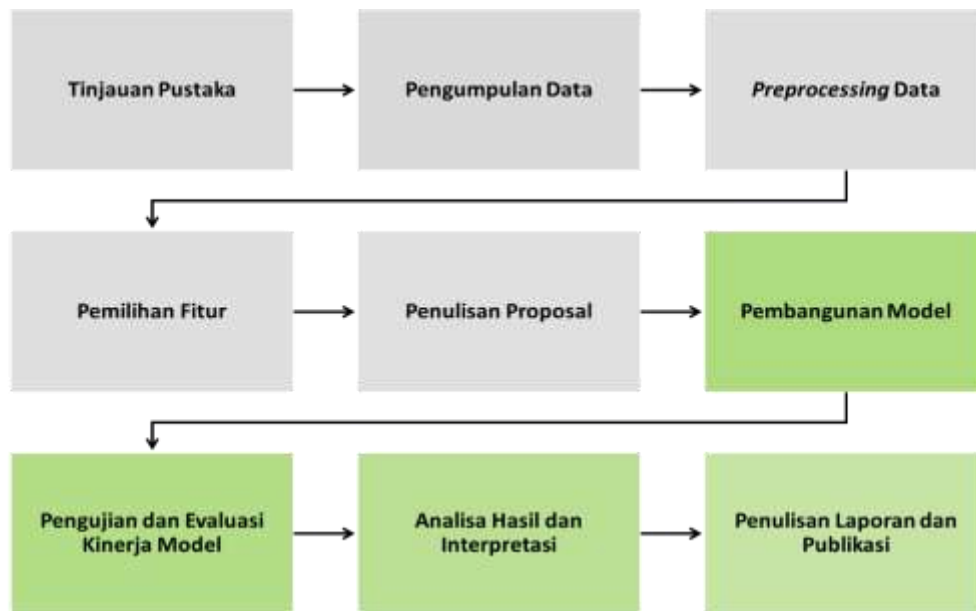
No	Penulis	Sumber (Jumlah Data)	Metode	Akurasi	Topik
1	(Agustina, Adrian and Hermawati, 2021)	kumparan.com (1,000)	Naive Bayes	81,00%	-
2	(Edward, 2021)	detik.com, turnbackhoax.id (1,525)	Decision Tree	99,60%	-
3	(Hasanah, Suciati and Purwitasari, 2021)	Twitter (12,000)	Neural Network	97,30%	Covid-19
4	(Irena and Setiawan, 2020)	Twitter (50,610)	Decision Tree	72,91%	6 topik
5	(Kencana, Setiawan and Kurniawan, 2020)	Twitter (50,646)	Neural Network	78,76%	6 topik
6	(Kesumawati and Thalib, 2018)	detik.com, turnbackhoax.id (1,849)	Naive Bayes	85,09%	6 topik
7	(Khan and Alhazmi, 2020)	Kaggle (20,008)	SVM	90,70%	-
8	(Kirana, Prasetyo and Lestari, 2023)	Twitter (720)	Naive Bayes	88,80%	Covid-19
9	(Muhabatin <i>et al.</i> , 2021)	turnbackhoax.id (110)	Naive Bayes	91,82%	-
10	(Mustofa and Mahfudh, 2019)	turnbackhoax.id (300)	Naive Bayes	85,28%	Covid-19
11	(Nugraha and Abdulloh, 2022)	detik.com, turnbackhoax.id (300)	Naive Bayes	93,00%	-
12	(Prasetya and Ferdiansyah, 2022)	liputan6.com, turnbackhoax.id (300)	Naive Bayes	86,30%	Covid-19
13	(Prayoga, Hadiana and Umbara, 2021)	Kaggle (11,000)	Naive Bayes	98,50%	-
14	(Putri <i>et al.</i> , 2019)	Artikel Berita (251)	Random Forest	76,47%	-
15	(Rahutomo, Pratiwi and Ramadhani, 2019)	Artikel Berita (600)	Naive Bayes	82,60%	-
16	(Rasywir and Purwarianti, 2015)	Artikel Berita (220)	Naive Bayes	91,36%	22 topik
17	(Sani <i>et al.</i> , 2022)	Artikel Berita (287)	SVM	95,60%	Kesehatan
18	(Santiyasa <i>et al.</i> , 2022)	Artikel Berita (740)	KNN	75,40%	-
19	(Santoso <i>et al.</i> , 2020)	Artikel Berita (30)	Naive Bayes	91,00%	-
20	(Soleman, 2021)	Artikel Berita (250)	Naive Bayes	72,00%	10 topik
21	(Sudrajat, Wulandari and Syafwan, 2022)	data.mendeley.com (600)	Naive Bayes	82,00%	-
22	(Utami <i>et al.</i> , 2021)	Jabar Saber Hoaks, Jala Hoaks (599)	KNN	75,89%	Covid-19
23	(Wati, 2020)	turnbackhoax.id (150)	Naive Bayes	85,19%	-

Berdasarkan Tabel 1, tampak bahwa penelitian terkait hoaks kebanyakan didominasi metode klasifikasi *Naïve Bayes*, sedangkan metode lainnya adalah *Decision Tree*, *SVM (Support Vector Machine)*, *KNN (K-Nearest Neighbour)*, dan *Random Forest*. Data berita ada yang berasal dari satu atau dua sumber tertentu atau dari berbagai sumber yang dituliskan sebagai Artikel Berita. Jumlah data bervariasi dari 30 saja sampai 50,610. Topik berita ada yang mengacu pada satu topik tertentu, beberapa topik, atau tanpa memperhatikan topik.

Dalam penelitian ini sebelumnya dalam proposal dipilih metode Naive Bayes karena fitur-fitur dari data bersifat independen dan cukup memadai dengan jumlah data yang tidak terlalu besar (Suryadewiansyah and Tju, 2022). Namun, karena tidak memberikan hasil yang bagus dikarenakan data nya ada yang kategorikal dan ada yang kontinu maka metode klasifikasi yang digunakan adalah *Decision Tree* (Wu *et al.*, 2008) dengan algoritma *CART (Classification and Regression Trees)* (Chandraveni, 2022) karena lebih fleksibel pada independensi fitur dan tipe data, serta banyaknya data yang diolah. Dataset bersumber dari Mafindo (Masyarakat Anti Fitnah Indonesia) (Mafindo.or.id, 2023) dengan jumlah data sekitar 10,000. Perbedaan dengan penelitian sebelumnya, selain dari data dan jumlah data yang digunakan, adalah percobaan pemilihan fitur dan teknik pemberian kategori dari setiap fiturnya, sehingga dapat memberikan akurasi cukup tinggi.

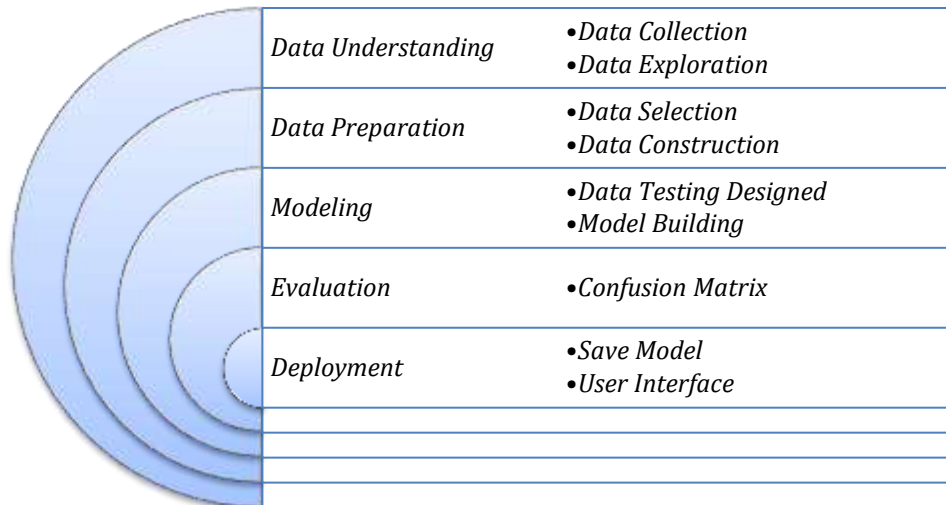
BAB 3. METODE PENELITIAN

Penelitian yang akan dilakukan terdiri dari sembilan langkah atau tahapan, seperti ditunjukkan pada Gambar 1. Empat tahap pertama sudah dan sedang dilaksanakan yaitu tinjauan pustaka, pengumpulan data, *preprocessing* data, dan pemilihan fitur. Setelah pembuatan proposal akan dilanjutkan dengan tiga tahap berikutnya yaitu pembangunan model, pengujian dan evaluasi kinerja model, analisa hasil dan interpretasi. Penelitian diakhiri dengan pembuatan laporan dan publikasi artikel di jurnal nasional terakreditasi Sinta.



Gambar 1. Tahapan penelitian

Tinjauan pustaka dilakukan dengan mencari artikel-artikel penelitian terkait di <https://www.mendeley.com>. Artikel yang relevan dimasukkan ke dalam *library* Mendeley Reference Manager untuk memudahkan dalam melakukan kutipan dan pembuatan daftar pustaka. Metodologi penelitian diadopsi dari CRISP-DM (*CRoss Industry Standard Process for Data Mining*) (Hotz, 2023) dan SKKNI (Standar Kompetensi Kerja Nasional Indonesia) No. 299 tahun 2020 (Kementerian Ketenagakerjaan Republik Indonesia, 2020) yang disesuaikan dengan kebutuhan penelitian sehingga menjadi lima tahapan yaitu *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*, seperti ditunjukkan pada Gambar 2.



Gambar 2. Tahapan Penelitian

3.1. Data Understanding

Tahapan penelitian dimulai dengan pemahaman data mulai dari pengambilan data hingga eksplorasi data. Sampel data dari Maifindo ditunjukkan pada Gambar 3.

Title	Content	Source	Classification
[SALAH] "CIA Bongkar jati diri Presiden Indonesia ...	"Di nilai Indonesia condong ke Blok China Komunis,...	facebook.com	Hoax
[SALAH] Hidayat Nur Wahid Akui PKS Tak Menganut As...	HIDAYAT NUR WAHID AKUI PKS TAK MENGANUT ASAS PANCA...	facebook.com	Hoax
[SALAH] "sakitnya apa di diagnosa nya covid"	Beredar sebuah foto dari akun Nanda Faturachman (f...	facebook.com	Hoax
Sebaran Kasus COVID-19 Di Kabupaten Temanggung	Jika sudah terjangkit covid-19 didaerah mana saja ...	whatsapp.com	Valid
FAQ			
Wabah Virus Corona			
Apa Saja yang Perlu Kita T...	Adakah corona virus bisa merebak melalui permukaan...	whatsapp.com	Valid
[SALAH] "Apa iyaaaa bisa hamil????"	Beredar postingan sejumlah foto pria tengah melaku...	facebook.com	Hoax
46 Tenaga Medis RS Kariadi Positif Corona, Lanjut ...	Dokter dan perawat rumah sakit karya di positif cov...	whatsapp.com	Valid
[SALAH] "bupati Luwu Utara terbaning di RS karena ...	Beredar foto bupati Luwu Utara terbaning di RS kar...	facebook.com	Hoax
[SALAH] Alhamdulillah Akhirnya Aceh Bisa Berangkat...	"Alhamdulillah Akhirnya Aceh Bisa Berangkat Haji ...	youtube.com	Hoax
[SALAH] Wapres Pemerintah Gak Sengaja Memakai Dane...	Akun Facebook Putra Inka membagikan gambar judul s...	facebook.com	Hoax
[SALAH] "dokter gigi di Surabaya stres telanjang d...	Akun facebook Samuel Moeladji mengunggah sebuah g...	facebook.com	Hoax
[SALAH] "Ibu Kota Pindah, Anies Bakal Jual Gedung ...	Beredar artikel berjudul "Ibu Kota Pindah, Anies B...	facebook.com	Hoax
[SALAH] "3 hari kedepan Arus angin dari Utara ke a...	Arus angin dari utara dari utara ke arah selatan y...	whatsapp.com	Hoax
[SALAH] Minggu 21 Juni 2020 Bam Wong Rilis Hadiah...	Beredar postingan Facebook dengan gambar poster ar...	facebook.com	Hoax
[SALAH] "Amien Rais Kalau Jokowi Sampai Diturunka...	Akun Afrizal (fb.com/100035229911489) mengunggah s...	facebook.com	Hoax
[SALAH] Artikel "#MakzulkanJKWBubarkanPDIP Trondin...	#MakzulkanJKWBubarkanPDIP Tronding, RUU HIP Ditudi...	facebook.com	Hoax
	Mana toriakan "Saya Pancasila mu ?"		
[SALAH] Tulisan "Mana teriakan "Saya Pancasila mu? ...		facebook.com	Hoax
	Oleh : Dr Abd		
Tiga Orang Dekat Menpan RB Tjahjo Kumolo Positif C...	Tjahjo Kumolo positif covid-19 breaking news	whatsapp.com	Valid
[SALAH] Kapal Tenggelam di Perairan Makassar, 13 O...	Innalillahi wa innalilahi Rojiun, Turut berduka c...	whatsapp.com	Hoax

Gambar 3. Sampel Data dari Mafindo

3.1.1. Data Collection

Data berita dalam format *JSON (JavaScript Object Notation)* (Pezoa et al., 2016), diambil dengan *API (Application Programming Interface)* yang disediakan dan diberikan akses oleh Mafindo, selanjutnya data disimpan dalam database *MySQL* dan ditampilkan dengan *phpMyAdmin*.

3.1.2. Data Exploration

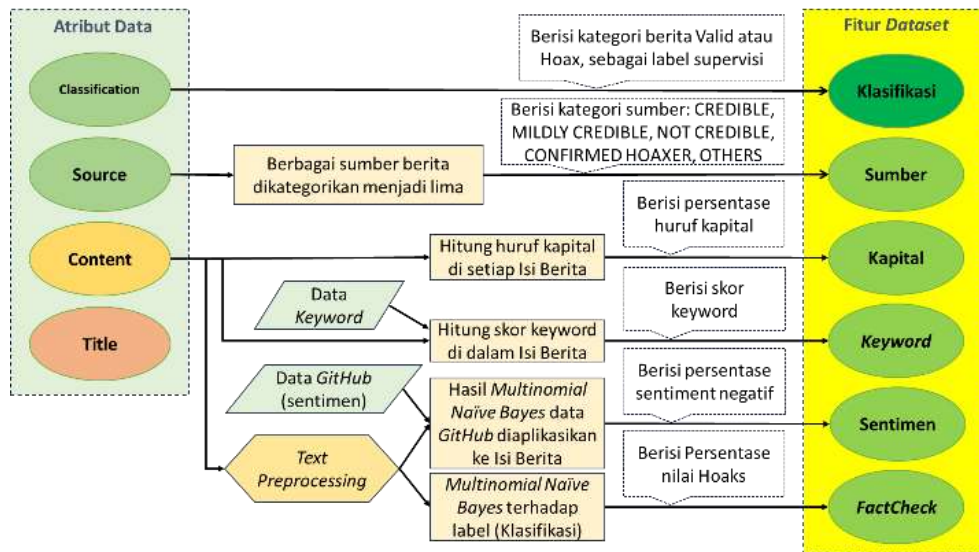
Setiap berita terdiri dari atribut *Title* (judul berita), *Content* (isi berita), *Source* (sumber berita), dan *Classification* (klasifikasi berita yang berisi label *Valid* atau *Hoax*), seperti sampel berita yang disajikan pada Gambar 3.

3.2. Data Preparation

Persiapan data dengan pemilihan dan rekayasa dari atribut data menjadi fitur-fitur yang diperlukan sehingga didapat *dataset* untuk pemodelan.

3.2.1. Data Selection

Data dianalisa dan ditentukan atribut-atribut yang akan digunakan sebagai fitur pada *dataset* pemodelan. Fitur klasifikasi diambil secara langsung dari atribut *classification*. Fitur klasifikasi ini merupakan label yang digunakan sebagai supervisi, sedangkan fitur lainnya diolah dari atribut *source* dan *content*. Pemilihan dan transformasi fitur ini bisa dilihat pada Gambar 4.



Gambar 4. Pemilihan dan Konstruksi Fitur

3.2.2. Data Construction

Atribut *source* yang berisi berbagai data media online dikategorikan menjadi lima (CREDIBLE, MILDLY CREDIBLE, NOT CREDIBLE, CONFIRMED HOAXER, OTHERS). Dua kategori pertama diadopsi berdasarkan tingkatan status perusahaan pers (Dewan Pers, 2023) yaitu berturut-turut “Terverifikasi Administrasi dan Faktual” dan “Terverifikasi Administrasi”, sedangkan NOT CREDIBLE adalah sumber yang tidak terdaftar di dewan pers. CONFIRMED HOAXER adalah

sumber berita yang masuk dalam daftar hoaks Mafindo. Untuk OTHERS dikategorikan untuk berita-berita yang tidak diketahui sumbernya.

Selain fitur Sumber, empat fitur dikonstruksi dari atribut isi berita dan diberi nama Kapital, *Keyword*, Sentimen, dan *Factcheck*. Fitur kapital diperoleh dari persentase banyaknya huruf besar/kapital dalam setiap isi berita. Fitur *keyword* berupa total skor dari kata kunci hoaks (milik Mafindo) yang muncul dalam berita. Fitur sentimen dihasilkan dengan *Multinomial Naive Bayes* menggunakan data kata sentimen positif (Prakoso, 2017b) dan negatif (Prakoso, 2017a) dari *repository* di *GitHub* dan diaplikasikan pada setiap isi berita sehingga diperoleh dan digunakan persentase sentimen negatif. Demikian juga untuk fitur *FactCheck* dihasilkan dengan *Multinomial Naive Bayes* pada hasil *text preprocessing* seluruh isi berita kemudian terhadap label dilakukan perhitungan probabilitas hoaks setiap berita sehingga diperoleh dan digunakan persentase hoaks. Konstruksi fitur juga disajikan pada Gambar 4.

3.3. Modeling

Pemodelan dilakukan dengan terlebih dahulu *dataset* dipisahkan menjadi data latih dan data uji. Semua proses perhitungan menggunakan bahasa pemrograman *Python* (Pilgrim and Willison, 2009; Van Rossum and Drake, 2009) dengan *Visual Studio Code (VSCode)* sebagai editor.

3.3.1. Data Testing Designed

Total data yang diperoleh sebanyak 11.934 , namun ada 2.178 data yang belum ada label (klasifikasi) karena belum dilakukan pemeriksaan fakta oleh tim Mafindo, sehingga *dataset* yang dipakai terdiri dari total 9.756 data yang terdiri dari 8.940 hoaks dan 816 valid. *Dataset* dibagi menjadi 80% data latih dan 20% data uji, sedangkan 2.178 data tanpa label tidak digunakan. Pelatihan juga dilakukan dengan jumlah data random yang bervariasi untuk mengetahui pengaruh besarnya data dengan pemodelan *Decision Tree Classifier CART* (Chandraveni, 2022), dengan distribusi data disajikan pada Tabel 2. Untuk membuktikan pengaruh *imbalance dataset*, pelatihan dilakukan pula dengan dan tanpa teknik SMOTE (Chawla *et al.*, 2002).

3.3.2. Model Building

Pemodelan dilakukan dengan metode *Decision Tree Classification* (Wu *et al.*, 2008) yang mengimplementasikan algoritma CART.

Tabel 2. Percobaan dengan Variasi Jumlah Data

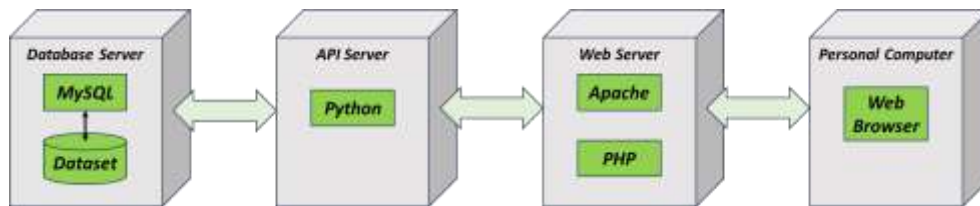
Total Data Pemodelan	Dataset	
	Data Latih	Data Uji
9.756	7.804	1.952
8.000	6.400	1.600
6.000	4.800	1.200
4.000	3.200	800
2.000	1.600	400

3.4. Evaluation

Hasil pemodelan dievaluasi dengan *confusion matrix* berikut *accuracy* dan *F1 Score* sebagai metrik pengukurannya (Suryadewiansyah and Tju, 2022) pada beberapa *dataset* dengan jumlah data berbeda-beda seperti di Tabel 2.

3.4. Deployment

Proses dan hasil pemodelan diimplementasikan berbasis *web* dengan arsitektur seperti pada Gambar 5.



Gambar 5. Deployment Environment

3.4.1. Save Model

Hasil pemodelan, digunakan *pickle library* untuk diubah menjadi *binary file*, disimpan ke dalam *storage API Server*, kemudian dengan *Python API* diumpankan ke *web server*.

3.4.2. User Interface

Aplikasi pengguna dibuat dengan *HTML* dan *CSS* yang diakses melalui *web browser*.

BAB 4. HASIL DAN ANALISA

Setelah dilakukan tahapan dalam metodologi penelitian, didapatkan hasil yang secara komprehensif dijelaskan ke dalam empat tahap berikut ini.

4.1. Data Preparation

Data untuk fitur Sumber diperoleh dengan algoritma pada Gambar 6.

```
Calculate Score Sumber(string sumber)

source_rankings = get_source_rankings()
sumber_input = lower_case(sumber)

for each source_ranking in source_rankings do
    list_source = source_ranking['list_source']
    the_score_of_the_source = source_ranking['ranking']

    for each source in list_source do
        if source is found in sumber_input then
            return the score of the source
        end if
    end for

return source is not found
```

Gambar 6. Algoritma Data Fitur Sumber

Untuk mendapatkan data fitur Kapital dari data atribut *Content* digunakan algoritma yang ditunjukkan pada Gambar 7.

```
Calculate Score Capital(string teks)

count_capital := 0
count_lower := 0
total_lower_capital := 0

for each character c in teks do
    if c is uppercase then
        increment count_capital by 1
    else if c is lowercase then
        increment count_lower by 1
    end if
end for

set total_lower_capital as the sum of count_capital and count_lower

if total_lower_capital is 0 then
    set percentage_capital as 0
else
    set percentage_capital as count_capital divided by total_lower_capital
end if

return percentage_capital
```

Gambar 7. Algoritma Data Fitur Kapital

Selanjutnya fitur *Keyword*, data diperoleh dari atribut *Content* dan daftar kata kunci hoaks dari Mafindo menggunakan algoritma seperti pada Gambar 8.

Calculate Score Keyword (list KEYWORDS_SCORES, string teks)

```
score_teks = 0

for each keyword and score from KEYWORDS_SCORES items do
    occurrence = find occurrence of keyword in teks
    score_teks += (occurrence * score)

return score_teks
```

Gambar 8. Algoritma Data Fitur *Keyword*

Sebelum mendapatkan data untuk fitur Sentimen dan *FactCheck*, dilakukan *text preprocessing* data pada atribut *Content* dengan algoritma di Gambar 9.

PreProcess Text(string teks)

```
CaseFoldedText = caseFolding(teks)
SymbolRemovedText = symbolRemoval(teks)
TokenizedWords = WordTokenization(teks)
StopWordRemovedText = stopWordRemoval(teks)
ClearSingleCharacterText = removeSingleCharacterWords(teks)
CleanText = removeWhitespace(teks)

Return CleanText
```

Gambar 9. Algoritma *Text Preprocessing*

Dengan algoritma pada Gambar 10, dilakukan *Multinomial Naïve Bayes* dengan data kata sentimen (Prakoso, 2017b), (Prakoso, 2017a) dari *GitHub*, hasil pemodelannya digunakan algoritma pada Gambar 11 untuk menghitung probabilitas negatif pada hasil *text preprocessing* data atribut *Content* untuk mengisi data fitur Sentimen.

Train Data Sentimen 0

```
negative_keywords = open file of negative keywords
positive_keywords = open file of positive keywords

keywords = negative_keywords combined with positive_keywords
labels = add 0 for count of negative_keywords times
labels += add 1 for count of positive_keywords times

mnb = create MultinomialNaiveBayes object
mnb.fit(keywords, labels)

store mnb in a binary file using pickle
```

Gambar 10. Algoritma Pemodelan Data Sentimen *GitHub*

```

Predict Sentimen With Probability (string teks)

trained_model = load trained model from file (FILE_SENTIMEN_MODEL)

text to predict = preprocessing(teks)

new_text_features = Count Vectorize (new_text)
probabilities = trained_model.Predict With Probabilities(new_text_features)
predictions = trained_model.Predict With Predictions(new_text_features)

not_hoax_percentage = probabilities of not hoax * 100 and round it for last 5
digit behind decimal
hoax_percentage = probabilities of hoax * 100 and round it for last 5 digit
behind decimal

probabilities_percentage = [not_hoax_percentage, hoax_percentage]

return predictions, probabilities_percentage

```

Gambar 11. Algoritma Data Fitur Sentimen

Sedangkan fitur *FactCheck* diperoleh dengan algoritma pada Gambar 12, yaitu dilakukan pemodelan *Multinomial Naïve Bayes* dengan data hasil *text preprocessing* terhadap label dan hasil pemodelannya digunakan pada algoritma pada Gambar 13 untuk menghitung probabilitas hoaks data fitur *FactCheck*.

```

Train After Text Preprocessing ()

train_text = load train text from database
train_label = load train label from database

texts_to_train = preprocessing(train_text)

balanced_data = SMOTE(texts_to_train)

new_text_features = TfIDF(balanced_data)
mnb = create MultinomialNaiveBayes object
mnb.fit(new_text_features, train_label)

store mnb in a binary file using pickle

```

Gambar 12. Algoritma Pemodelan Data *Text Preprocessing*

```

Predict Fact Check With Probability (string text)

trained_model = load trained model from file (FILE_FACT_CHECK_MODEL)

text to predict = preprocessing(teks)

new_text_features = TfIDF(new_text)
probabilities = trained_model.Predict With Probabilities(new_text_features)
predictions = trained_model.Predict With Predictions(new_text_features)

not_hoax_percentage = probabilities of not hoax * 100 and round it for last 5
digit behind decimal
hoax_percentage = probabilities of hoax * 100 and round it for last 5 digit
behind decimal

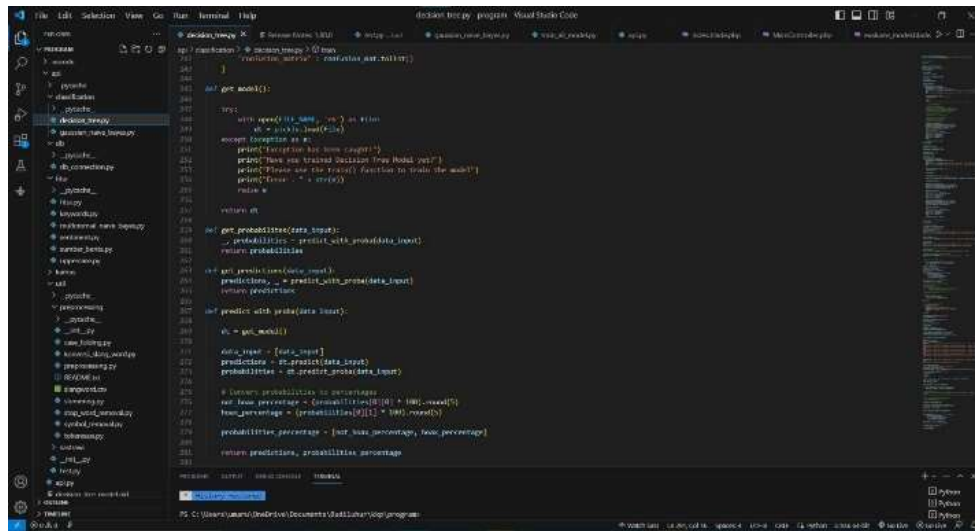
probabilities_percentage = [not_hoax_percentage, hoax_percentage]

return predictions, probabilities_percentage

```

Gambar 13. Algoritma Data Fitur *FackCheck*

Pengkodean *Python* dari algoritma untuk mendapatkan fitur-fitur *dataset* dilakukan dengan editor *VSCode* dengan tangkapan layar salah satu bagian ditunjukkan pada Gambar 14.



Gambar 14. Pengkodean dengan *VSCode*

Berdasarkan semua algoritma di atas maka sampel *dataset* yang diperoleh dari pemilihan dan kontruksi atribut data Mafindo ditunjukkan pada Gambar 15.

Keyword	Kapital	Sumber	Sentimen	FactCheck	Klasifikasi
0	4.093567281961441	NOT CREDIBLE	7.75922	100	Hoax
0	20.74074000120163	NOT CREDIBLE	61.8778	99.9998	Hoax
0	10.447761416435242	NOT CREDIBLE	51.1513	100	Hoax
0	0	CREDIBLE	69.5534	50	Valid
-1	7.737801969051361	NOT CREDIBLE	0.00001	100	Hoax
0	19.72789168357849	OTHERS	49.3178	99.9799	Hoax
0	8.457711338996887	OTHERS	16.8671	99.9985	Hoax
0	0.6060605868697166	NOT CREDIBLE	7.75922	100	Hoax
0	2.409638464450836	NOT CREDIBLE	87.3732	100	Hoax
0	18.292683362960815	NOT CREDIBLE	23.3328	98.7187	Hoax
0	0	NOT CREDIBLE	75.1756	100	Hoax
0	0	NOT CREDIBLE	69.5534	50	Valid
0	0	OTHERS	69.5534	50	Valid
0	14.049586653709412	NOT CREDIBLE	69.5534	90.9384	Hoax
0	2.9411764815449715	NOT CREDIBLE	9.75707	99.9798	Valid
0	3.375527262687683	NOT CREDIBLE	2.8024	100	Hoax
0	7.14285746216774	NOT CREDIBLE	69.5534	99.9752	Hoax
0	23.84105920791626	NOT CREDIBLE	72.7223	100	Hoax
0	9.48905125260353	NOT CREDIBLE	69.5534	17.0902	Hoax
0	21.052631735801697	NOT CREDIBLE	45.4688	99.9452	Hoax

Gambar 15. Sampel Dataset Penelitian

4.2. Modeling

Pemodelan *Decision Tree* dengan data latih dilakukan dengan algoritma pada Gambar 16.

```

Train Model ()

all_train_scores = load all train scores from database
train_labels = load labels from database

balanced_train_scores, balanced_train_labels = SMOTE(all_train_scores,
train_labels )

dt = create DecisionTree object
dt.fit(balanced_train_scores, balanced_train_labels )

store dt in a binary file using pickle

```

Gambar 16. Algoritma Pemodelan

Setelah dilakukan pemodelan dilanjutkan pengujian model pada data latih dengan algoritma pada Gambar 17.

```
Predict With Probability (float score_capital, int score_keyword, string score_sumber, float score_sentimen_negatif, float score_factcheck)  
  
trained_model = load trained model from file  
(FILE_HOAX_DETECTOR_MODEL)  
  
all_scores = [score_capital, score_keyword, score_sumber,  
score_sentimen_negatif, score_factcheck]  
  
probabilities = trained_model.Predict With Probabilities(all_scores )  
predictions = trained_model.Predict With Predictions(all_scores )  
  
not_hoax_percentage = probabilities of not hoax * 100 and round it for last 5  
digit behind decimal  
hoax_percentage = probabilities of hoax * 100 and round it for last 5 digit  
behind decimal  
  
probabilities_percentage = [not_hoax_percentage, hoax_percentage]  
  
return predictions, probabilities_percentage
```

Gambar 17. Algoritma Pengujian

Algoritma pemodelan diproses dengan *Python* dengan semua *library* yang dibutuhkan. Beberapa *library* penting ditunjukkan hasil tangkapan layar pada Gambar 18.

```
# Library NLTK untuk preprocessing teks  
from nltk.tokenize import word_tokenize  
from nltk.corpus import stopwords  
  
# Untuk mengambil dataset dan melatih model  
import numpy as np  
import pandas as pd  
from sklearn.feature_extraction.text import CountVectorizer  
from sklearn.feature_extraction.text import TfidfVectorizer  
from sklearn.naive_bayes import MultinomialNB  
from sklearn.tree import DecisionTreeClassifier  
from imblearn.over_sampling import SMOTE  
from sklearn.model_selection import train_test_split  
  
# Untuk perhitungan akurasi model  
from sklearn.metrics import accuracy_score  
from sklearn.metrics import confusion_matrix  
from sklearn.metrics import f1_score  
from sklearn.metrics import precision_score  
from sklearn.metrics import recall_score  
  
# Untuk penyimpanan model hasil train dalam bentuk file biner  
import pickle  
  
# Untuk mengambil dan menyimpan hasil skoring dataset di database MySQL  
import mysql.connector
```

Gambar 18. *Library* yang Digunakan dengan *Python*

4.3. Evaluation

Metode *Confusion Matrix* (Suryadewiansyah and Tju, 2022) digunakan untuk menampilkan hasil pengujian. Dalam penelitian ini, *P* (*Positive*) digunakan sebagai hoaks dan *N* (*Negative*) sebagai valid (bukan hoaks), dengan demikian *TP* (*True Positive*) adalah kondisi hasil prediksi dan label sama-sama hoaks, *TN* (*True Negative*) adalah kondisi hasil prediksi dan label sama-sama valid, *FP* (*False Positive*) adalah kondisi hasil prediksinya hoaks namun label berisi valid, *FN* (*False Negative*) adalah kondisi hasil prediksinya valid namun label berisi hoaks. Pengukuran dilakukan dengan *Accuracy* ($Acc.=\frac{TP+TN}{TP+TN+FP+FN}$) dan *F1 Score* ($F1=\frac{2*TP}{2*TP+FP+FN}$)

Hasil pengujian terhadap *dataset* (*Set*) dengan jumlah data (*N*) yang berbeda-beda tanpa SMOTE (Chawla *et al.*, 2002) ditampilkan pada Tabel 3. Pengujian ini dilakukan juga pada data latih, tidak hanya pada data uji, sebagai percobaan untuk perbandingan.

Tabel 3. Hasil Pengujian Tanpa SMOTE

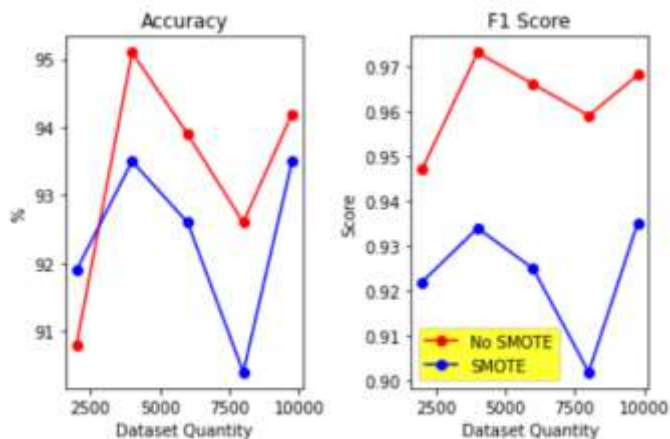
Set (N)	Pengujian		TP	FP	FN	TN	Acc.	F1
	Data	N						
9.756	Latih	7.804	7.180	0	32	664	99.6%	0,998
	Uji	1.952	1.744	58	56	94	94.2%	0,968
8.000	Latih	6.400	5.729	0	29	642	99.5%	0,997
	Uji	1.600	1.386	62	56	96	92.6%	0,959
6.000	Latih	4.800	4.290	0	30	480	99.4%	0,997
	Uji	1.200	1.034	27	46	93	93.9%	0,966
4.000	Latih	3.200	2.845	0	28	327	99.1%	0,995
	Uji	800	702	14	25	59	95.1%	0,973
2.000	Latih	1.600	1.421	0	21	158	98.7%	0,993
	Uji	400	333	12	25	30	90.8%	0,947

Sedangkan hasil pengujian pemodelan dengan menerapkan SMOTE (untuk menyeimbangkan jumlah data hoaks dan label) ada di Tabel 4.

Tabel 4. Hasil Pengujian dengan SMOTE

Set (N)	Pengujian		TP	FP	FN	TN	Acc.	F1
	Data	N						
9.756	Latih	7.804	7100	0	33	7171	99.8%	0,998
	Uji	1.952	1681	107	126	1662	93.5%	0,935
8.000	Latih	6.400	5738	0	33	5749	99.7%	0,997
	Uji	1.600	1275	123	154	1328	90.4%	0,902
6.000	Latih	4.800	4289	0	30	4321	99.7%	0,997
	Uji	1.200	993	72	88	1007	92.6%	0,925
4.000	Latih	3.200	2845	0	28	2878	99.5%	0,995
	Uji	800	667	43	51	679	93.5%	0,934
2.000	Latih	1.600	1395	0	20	1465	99.3%	0,993
	Uji	400	344	17	41	318	91.9%	0,922

Sebagai visualisasi evaluasi data uji, dari Tabel 3 dan 4 ditampilkan dalam bentuk grafik seperti pada Gambar 19.



Gambar 19. Visualisasi *Accuracy* dan *F1 Score*

4.4. Deployment

Hasil deployment purwarupa aplikasi berbasis *web* sederhana dengan memasukkan suatu berita baru untuk diprediksi hoaks atau valid, dengan cara *copy paste* ke dalam isian Konten Berita pada Gambar 20.

The screenshot shows a web form with a dark background. At the top, there is a label 'Sumber Berita' followed by a dropdown menu with the text 'Pilih sumber berita'. Below this is a large text area labeled 'Konten Berita'. At the bottom of the form, there are two buttons: a blue button labeled 'Submit' and a grey button labeled 'Evaluasi Model'.

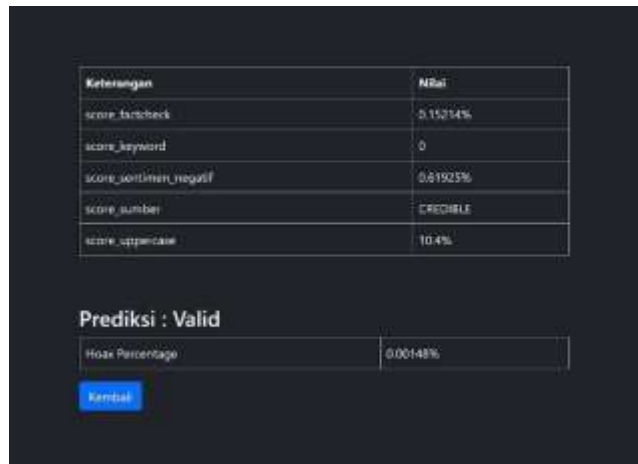
Gambar 20. Isian Konten Berita Baru

Selanjutnya untuk sumber berita bisa dipilih sesuai dengan sumber yang ada di data pelatihan, seperti pada Gambar 21, jika tidak ada yang sesuai maka dipilih sebagai data OTHERS untuk fitur Sumber.



Gambar 21. Pilihan Sumber Berita Baru

Selanjutnya tombol *Submit* ditekan untuk mendapatkan hasil prediksi dari model yang dibuat. Hasil prediksi disertakan informasi perhitungan data fitur hasil perhitungan model pada berita yang dimasukkan, seperti ditampilkan pada Gambar 22.



Gambar 22. Hasil Prediksi suatu Berita

4.5. Analisa

Pengujian yang dilakukan terhadap data latih dan uji memberikan hasil akurasi dan F1 yang sudah tepat, karena pengujian pada data latih cenderung lebih baik namun dengan pemodelan *Decision Tree* tidak terjadi *overfit*. Perbedaan hasil pengujian data uji yang sedikit lebih rendah menunjukkan hasil pemodelan cukup baik.

Perbandingan pengujian pada kuantitas *dataset* yang berbeda-beda, tidak menunjukkan tren yang berarti pada nilai akurasi dan skor F1 karena terjadi naik turun tidak *significant*. Mengacu pada hasil visualisasi pada Gambar 18, maksimal selisih perbedaan akurasi tanpa SMOTE adalah 4,3% dan dengan SMOTE adalah 3.1%, sedangkan skor F1 maksimal selisih perbedaan tanpa SMOTE adalah 0.026 dan dengan SMOTE adalah 0,033.

Lebih lanjut pengaruh *imbalance dataset* dengan *SMOTE*, yaitu melakukan *oversampling* pada data yang berlabel lebih sedikit, secara umum memberikan kinerja yang sedikit lebih rendah dibandingkan tanpa *SMOTE*. Perbedaan selisih akurasi maksimal adalah 2,2% dan selisih F1 maksimal adalah 0,057, keduanya terjadi pada *dataset* 8000. Namun demikian model dengan *SMOTE* memiliki kecenderungan yang lebih stabil atau adil dalam mengklasifikasikan data, dengan perbandingan False Negative (FN) dan False Positive (FP) yang lebih seimbang dibandingkan model tanpa *SMOTE*, yang berarti tidak didominasi oleh label yang lebih banyak.

Pengaruh pemakaian fitur dari beberapa aspek dan diproses dengan sub pemodelan *Multinomial Naïve Bayes* pada fitur Sentimen dan *FactCheck* diduga memberikan dampak positif diperolehnya model yang stabil, namun perlu dilakukan penelitian lebih lanjut dalam hal ini.

BAB 5. KESIMPULAN DAN SARAN

Dalam penelitian ini, pemodelan *Decision Tree Classifier* pada data berita dengan fitur sumber berita, huruf kapital, kata kunci hoaks, sentimen negatif, *fact check* hoaks, dengan label klasifikasi valid dan hoaks, memberikan hasil yang memuaskan dengan nilai akurasi 93,5% dan skor F1 0,935 (angka tersebut secara kebetulan mirip). Penggunaan *SMOTE* dalam pemodelan *Decision Tree* untuk mengatasi imbalance dataset tidak terlalu berpengaruh pada hasil akurasi dan skor F1 namun memberi dampak kestabilan model terhadap kuantitas *dataset*.

DAFTAR PUSTAKA

- Agustina, N., Adrian, A. and Hermawati, M. (2021) 'Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media', *Faktor Exacta*, 14(4).
- Chandraveni, T. (2022) *CART (Classification And Regression Tree) in Machine Learning*, *GeeksforGeeks*. Available at: <https://www.geeksforgeeks.org/cart-classification-and-regression-tree-in-machine-learning/> (Accessed: 8 July 2023).
- Chawla, N. V *et al.* (2002) 'SMOTE: synthetic minority over-sampling technique', *Journal of artificial intelligence research*, 16, pp. 321–357.
- Dewan Pers (2023) *Data Perusahaan Pers, Lembaga Dewan Pers*. Available at: <https://dewanpers.or.id/data/perusahaanpers> (Accessed: 6 July 2023).
- Edward, E. (2021) 'Identifikasi Berita HOAX Berbasis Web Menggunakan Algoritma C4.5', *Jurnal Ilmu Komputer dan Sistem Informasi*, 9(1). Available at: <https://doi.org/10.24912/jiksi.v9i1.11558>.
- Hasanah, N.A., Suciati, N. and Purwitasari, D. (2021) 'Identifying degree-of-concern on covid-19 topics with text classification of twitters', *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 7(1). Available at: <https://doi.org/10.26594/register.v7i1.2234>.
- Hotz, N. (2023) *What is CRISP DM? - Data Science Process Alliance*. Available at: <https://www.datascience-pm.com/crisp-dm-2/> (Accessed: 1 July 2023).
- Irena, B. and Setiawan, E.B. (2020) 'Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(4), pp. 711–716. Available at: <https://doi.org/10.29207/resti.v4i4.2125>.
- Kementerian Ketenagakerjaan Republik Indonesia (2020) *SKKNI Keahlian Artificial Intelligence (Data Science)*. Available at: <https://skkni.kemnaker.go.id/tentang-skkni/dokumen?area=data%20science&limit=20&page=1%20> (Accessed: 1 July 2023).
- Kencana, C.W., Setiawan, E.B. and Kurniawan, I. (2020) 'Hoax Detection System on Twitter using Feed-Forward and Back-Propagation Neural Networks Classification Method', *Jurnal*

RESTI (Rekayasa Sistem dan Teknologi Informasi), 4(4). Available at: <https://doi.org/10.29207/resti.v4i4.2038>.

Kesumawati, A. and Thalib, A.K. (2018) 'Hoax classification with Term Frequency - Inverse Document Frequency using non-linear SVM and Naïve Bayes', *International Journal of Advances in Soft Computing and its Applications*, 10(3).

Khan, M.Z. and Alhazmi, O.H. (2020) 'Study and analysis of unreliable news based on content acquired using ensemble learning (prevalence of fake news on social media)', *International Journal of System Assurance Engineering and Management*, 11(S2), pp. 145–153. Available at: <https://doi.org/10.1007/s13198-020-01016-4>.

Kirana, A.P., Prasetyo, G.B. and Lestari, E.W. (2023) 'The Detection of Indonesian Hoax Content about COVID-19 Vaccine using Naive Bayes Multinomial Method', *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 5(1). Available at: <https://doi.org/10.35882/ijeeemi.v5i1.262>.

Kumparan.com (2021) *Dampak dari Pesatnya Perkembangan Teknologi di Era Digital*. Available at: <https://kumparan.com/berita-update/dampak-dari-pesatnya-perkembangan-teknologi-di-era-digital-1vBkPOYNffj> (Accessed: 2 June 2023).

Mafindo.or.id (2023) *MAFINDO – Masyarakat Anti Fitnah Indonesia*. Available at: <https://www.mafindo.or.id/> (Accessed: 11 June 2023).

Muhabatin, H. *et al.* (2021) 'Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO', *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, 5(2), p. 156. Available at: <https://doi.org/10.51211/itbi.v5i2.1531>.

Mustofa, H. and Mahfudh, A.A. (2019) 'Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes', *Walisongo Journal of Information Technology*, 1(1), p. 1. Available at: <https://doi.org/10.21580/wjit.2019.1.1.3915>.

Nugraha, A.Y.A. and Abdulloh, F.F. (2022) 'Optimasi Naive Bayes dan Cosine Similarity Menggunakan Particle Swarm Optimization Pada Klasifikasi Hoax Berbahasa Indonesia', *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3). Available at: <https://doi.org/10.30865/mib.v6i3.4170>.

Pezoa, F. *et al.* (2016) ‘Foundations of JSON schema’, in *Proceedings of the 25th International Conference on World Wide Web*, pp. 263–273.

Pilgrim, M. and Willison, S. (2009) *Dive Into Python 3*. Springer.

Potentia Analytics (2021) *What Is Machine Learning: Definition, Types, Applications and Examples*. Available at: <https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/> (Accessed: 10 June 2023).

Prakoso, R. (2017a) *analisis-sentimen/kamus/negatif_ta2.txt*, *GitHub*. Available at: https://github.com/ramaprakoso/analisis-sentimen/blob/master/kamus/negatif_ta2.txt (Accessed: 5 July 2023).

Prakoso, R. (2017b) *analisis-sentimen/kamus/positif_ta2.txt*, *GitHub*. Available at: https://github.com/ramaprakoso/analisis-sentimen/blob/master/kamus/positif_ta2.txt (Accessed: 5 July 2023).

Prasetya, F. and Ferdiansyah, F. (2022) ‘Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes’, *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(1). Available at: <https://doi.org/10.30865/json.v4i1.4852>.

Prayoga, A.Y., Hadiana, A.I. and Umbara, F.R. (2021) ‘Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naive Bayes dengan Ekstraksi Fitur Tf-Idf’, *Jurnal Health Sains*, 2(10). Available at: <https://doi.org/10.46799/jsa.v2i10.327>.

Putri, T.T.A. *et al.* (2019) ‘ANALYSIS AND DETECTION OF HOAX CONTENTS IN INDONESIAN NEWS BASED ON MACHINE LEARNING’, *Journal Of Informatics Pelita Nusantara*, 4(1).

Rahutomo, F., Pratiwi, I.Y.R. and Ramadhani, D.M. (2019) ‘Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia’, *JURNAL PENELITIAN KOMUNIKASI DAN OPINI PUBLIK*, 23(1). Available at: <https://doi.org/10.33299/jpkop.23.1.1805>.

Rasywir, E. and Purwarianti, A. (2015) ‘Eksperimen pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin’, *Jurnal Cybermatika*, 3(2).

Rebala, G., Ravi, A. and Churiwala, S. (2019) ‘Machine Learning Definition and Basics’, in *An Introduction to Machine Learning*. Cham: Springer International Publishing, pp. 1–17. Available at: https://doi.org/10.1007/978-3-030-15729-6_1.

Van Rossum, G. and Drake, F.L. (2009) *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Sani, R.R. *et al.* (2022) ‘Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia’, *JURNAL MASYARAKAT INFORMATIKA*, 13(2). Available at: <https://doi.org/10.14710/jmasif.13.2.47983>.

Santiyasa, I.W. *et al.* (2022) ‘IDENTIFICATION OF HOAX BASED ON TEXT MINING USING K-NEAREST NEIGHBOR METHOD’, *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 10(2). Available at: <https://doi.org/10.24843/jlk.2021.v10.i02.p04>.

Santoso, H.A. *et al.* (2020) ‘Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization’, *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(2), p. 799. Available at: <https://doi.org/10.12928/telkomnika.v18i2.14744>.

Sarwan (2023) *Perspektif Hukum Pidana Mengenai Berita Hoaks*. Available at: <https://www.kompasiana.com/inggamaulana45747/64785a338221996cf1383c52/perspektif-hukum-pidana-mengenai-berita-hoax-tentang-modus-pemerasan-pemotor-tabrakan-diri-kemobil-di-tangerang> (Accessed: 2 June 2023).

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). Available at: <https://doi.org/10.24014/coreit.v7i2.14290>.

Sudrajat, A., Wulandari, R.R. and Syafwan, E. (2022) ‘Indonesian Language Hoax News Classification Based on Naïve Bayes’, *Journal of Applied Intelligent System*, 7(1). Available at: <https://doi.org/10.33633/jais.v7i1.5985>.

Suryadewiansyah, M.K. and Tju, T.E.E. (2022) ‘Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert’, *Jurnal Nasional Teknologi dan Sistem Informasi*, 8(2). Available at: <https://doi.org/10.25077/teknosi.v8i2.2022.81-88>.

Utami, E. *et al.* (2021) ‘Covid-19 Hoax Detection Using KNN in Jaccard Space’, *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3). Available at: <https://doi.org/10.22146/ijccs.67392>.

Wati, R. (2020) ‘PENERAPAN ALGORITMA NAIVE BAYES DAN PARTICLE SWARM OPTIMIZATION UNTUK KLASIFIKASI BERITA HOAX PADA MEDIA SOSIAL’, *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 5(2). Available at: <https://doi.org/10.33480/jitk.v5i2.1034>.

Wu, X. *et al.* (2008) ‘Top 10 algorithms in data mining’, *Knowledge and information systems*, 14(1), pp. 1–37.

LAMPIRAN-LAMPIRAN

Lampiran 1. Realisasi Biaya

1. Honorarium				
Pelaksana	Honor/jam (Rp)	Waktu (jam/minggu)	Minggu	Honor (Rp)
Ketua	15.000,00	4,5	16	1.080.000,00
Anggota	10.000,00	3,0	14	420.000,00
Sub Total (Rp)				1.500.000,00
2. Bahan Habis Pakai dan Peralatan Penunjang				
Bahan/Alat	Justifikasi Pemakaian	Kuantitas	Harga Satuan (Rp)	Harga (Rp)
Sewa Laptop	Kegiatan dari awal sampai selesai	4 bulan	300.000,00	1.200.000,00
Kuota Internet	Kegiatan dari awal sampai selesai	4 bulan	200.000,00	800.000,00
<i>Dataset</i>	Konsultan Sumber Data	1 paket	500.000,00	500.000,00
Sub Total (Rp)				2.500.000,00
3. Operasional Kegiatan Penelitian				
Kegiatan	Justifikasi Pemakaian	Kuantitas	Harga Satuan (Rp)	Biaya (Rp)
Perjalanan	Transportasi	14 pp	40.000,00	560.000,00
Konsumsi	Makanan dan Minuman	14 kali	30.000,00	420.000,00
Materai	Kontrak Penelitian	2 buah	10.000,00	20.000,00
Sub Total (Rp)				1.000.000,00
4. Hak Atas Kekayaan Intelektual				
Uraian	Justifikasi	Kuantitas	Harga Satuan (Rp)	Biaya (Rp)
-	-	-	0,00	0,00
Sub Total (Rp)				0,00
5. Lain-Lain				
Uraian	Justifikasi	Kuantitas	Harga Satuan (Rp)	Biaya (Rp)
-	-	-	0,00	0,00
Sub Total (Rp)				0,00
Total Anggaran (Rp)				5.000.000,00

Lampiran 2. Instrumen Penelitian

Data berita dalam format JSON (JavaScript Object Notation), diambil dengan API (Application Programming Interface) yang disediakan dan diberikan akses oleh Mafindo, selanjutnya data disimpan dalam database MySQL dan ditampilkan dengan phpMyAdmin. Hasil contoh data ditampilkan berikut ini:

Title	Content	Source	Classification
[SALAH] "CIA Bongkar jati diri Presiden Indonesia..."	"Di nilai Indonesia condong ke Blok China Komunis,..."	facebook.com	Hoax
[SALAH] Hidayat Nur Wahid Akui PKS Tak Menganut As...	HIDAYAT NUR WAHID AKUI PKS TAK MENGANUT ASAS PANCA...	facebook.com	Hoax
[SALAH] "sakitnya apa di diagnosa nya covid"	Beredar sebuah foto dari akun Nanda Faturachman (f...	facebook.com	Hoax
Sebaran Kasus COVID-19 Di Kabupaten Temanggung	Jika sudah terjangkit covid-19 didaerah mana saja	whatsapp.com	Valid
FAQ			
Wabah Virus Corona	Adakah corona virus bisa merebak melalui permukaan...	whatsapp.com	Valid
Apa Saja yang Perlu Kita T...			
[SALAH] "Apa iyaaaa bisa hamil????"	Beredar postingan sejumlah foto pria tengah melaku...	facebook.com	Hoax
40 Tenaga Medis RS Kariadi Positif Corona, Lanjut ...	Dokter dan perawat rumah sakit karyadi positif cov...	whatsapp.com	Valid
[SALAH] "bupati Luwu Utara terbaring di RS karena ...	Beredar foto bupati Luwu Utara terbaring di RS kar...	facebook.com	Hoax
[SALAH] Alhamdulillah Akhirnya Aceh Bisa Berangkat...	"Alhamdulillah Akhirnya Aceh Bisa Berangkat Haji ...	youtube.com	Hoax
[SALAH] Wapres Pemerintah Gak Sengaja Memakai Dand...	Akun Facebook Putra Inka membagikan gambar judul s...	facebook.com	Hoax
[SALAH] "dokter gigi di Surabaya stres telanjang d...	Akun facebook Samuel Moeliadji mengunggah sebuah g...	facebook.com	Hoax
[SALAH] "Ibu Kota Pindah, Anies Bakal Jual Gedung ...	Beredar artikel berjudul "Ibu Kota Pindah, Anies B...	facebook.com	Hoax
[SALAH] "3 hari kedepan Arus angin dari Utara ke a...	Arus angin dari utara dari utara ke arah selatan y...	whatsapp.com	Hoax
[SALAH] Minggu 21 Juni 2020 Baim Wong Rilis Hadiah...	Beredar postingan Facebook dengan gambar poster ar...	facebook.com	Hoax
[SALAH] "Amien Rais : Kalau Jokowi Sampai Dilurunka...	Akun Afrizal (fb.com/100035229911489) mengunggah s...	facebook.com	Hoax
[SALAH] Artikel "#MakzulkanJKWBubarkanPDIP Trendin...	#MakzulkanJKWBubarkanPDIP Trending, RUU HIP Ditudi...	facebook.com	Hoax
	Mana toriakan "Saya Pancasila mu?"		
[SALAH] Tulisan "Mana teriakan "Saya Pancasila mu?..."	Oleh : Dr Abid...	facebook.com	Hoax
Tiga Orang Dekat Menpan RB Tjahjo Kumolo Positif C...	Tjahjo Kumolo positif covid-19 breaking news	whatsapp.com	Valid
[SALAH] Kapal Tenggelam di Perairan Makassar, 13 O...	Innalillahi wa innailahi Rojiun, Turut berduka c...	whatsapp.com	Hoax

Lampiran 3. Biodata Peneliti

I. Biodata Ketua Peneliti

A. Identitas Diri

1. Nama Lengkap : Teja Endra Eng Tju, S.T., M.Kom.
2. Jenis Kelamin : Laki-laki
3. Jabatan Fungsional : Asisten Ahli
4. NIP/NIDN/ID-SINTA : 190046/0407127201/6713276
5. Tempat, Tanggal Lahir : Malang, 07 Desember 1972
6. E-mail : teja.endraengtju@budiluhur.ac.id
7. Nomor Handphone : +628998909707
8. Alamat : Alam Sutera, Tangerang Selatan

B. Riwayat Pendidikan

	S1	S2
Perguruan Tinggi	ITS Surabaya	UBL Jakarta
Bidang Ilmu	Teknik Elektro - Telekomunikasi	Ilmu Komputer - E-Commerce
Tahun Masuk-Lulus	1991 - 1996	2003 - 2009

C. Pengalaman Penelitian (5 Tahun Terakhir)

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jumlah (Rp)

D. Publikasi Artikel Ilmiah dalam Jurnal (5 Tahun Terakhir)

No	Judul Artikel Ilmiah	Nama Jurnal	Volume/Nomor/Tahun
1	Pelatihan Mendeley 2022 Bersama Ngampooz	Sinergitas PKM & CSR	Vol.6/No.2/2022
2	Creation of Questionnaire Keywords with the CISE Method for KMS User Satisfaction Evaluation	METHOMIKA	Vol.6/No.2/2022
3	<i>Naive Bayes dan Confusion Matrix</i> untuk Efisiensi Analisa <i>Intrusion Detection System Alert</i>	TEKNOSI	Vol.8/No.2/2022
4	Diseminasi Teknik Klasifikasi <i>Naive Bayes</i> pada <i>Intrusion Detection System</i> di Perusahaan Artajasa	Literasi	Vol.2/No/2/2022
5	Peningkatan Kesadaran Masyarakat pada <i>Covid-19</i> di RT01/RW01 Kelurahan Poris Plawad Indah Kota Tangerang	Igkojei	Vol.3/No.1/2022

6	<i>Prediction of the COVID-19 Vaccination Target Achievement with Exponential Regression</i>	JISA	Vol.4/No.2/2021
7	Pendampingan UMKM saat Pandemi Covid-19 dengan Teknologi Informasi Berbasis CMS WordPress pada Alphonse Shoe di Jakarta Selatan	Selaparang	Vol.4/No.3/2021
8	Pemberdayaan UMKM Sepatu dengan Aplikasi E-Commerce sebagai Pendukung Pemasaran saat Pandemi Covid-19	Madaniya	Vol.2/No.1/2021

E. Pemakalah Seminar Ilmiah (5 Tahun Terakhir)

No	Nama Temu Ilmiah/ Seminar	Judul Artikel Ilmiah	Waktu dan Tempat
1	SISFOTEK 2019, IAII	Kajian Penerapan MoT sebagai Inisiasi CRM pada Penyedia Layanan Akses Broadband: FirstMedia	25 September 2019 Jakarta, Indonesia

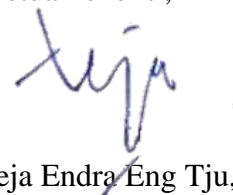
F. Karya Buku (5 Tahun Terakhir)

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit

G. Perolehan HKI (10 Tahun Terakhir)

No	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID
1	Sistem Informasi Manajemen Bank Sampah	2020	Program Komputer	000177935

Jakarta, 17 Agustus 2023
Ketua Peneliti,



Teja Endra Eng Tju, S.T., M.Kom.

II. Biodata Anggota peneliti

A. Identitas Diri

9. Nama Lengkap : Arman Yusuf, S.Kom., M.Kom.
10. Jenis Kelamin : Laki-laki
11. Jabatan Fungsional : Tenaga Pengajar
12. NIP/NIDN/ID-SINTA : 160015/0330087506/6693713
13. Tempat, Tanggal Lahir : Jakarta, 30 Agustus 1975
14. E-mail : arman.yusuf@budiluhur.ac.id
15. Nomor Handphone : +628979135967
16. Alamat : Kebon Jeruk, Jakarta Barat

B. Riwayat Pendidikan

	S1	S2
Perguruan Tinggi	Institut Keuangan Perbanas	Universitas Budi Luhur
Bidang Ilmu	Teknik Informatika	Managemen
Tahun Masuk-Lulus	1998 - 2001	2009 - 2013

C. Pengalaman Penelitian (5 Tahun Terakhir)

No	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jumlah (Rp)

D. Publikasi Artikel Ilmiah dalam Jurnal (5 Tahun Terakhir)

No	Judul Artikel Ilmiah	Nama Jurnal	Volume/Nomor/Tahun

E. Pemakalah Seminar Ilmiah (5 Tahun Terakhir)

No	Nama Temu Ilmiah/ Seminar	Judul Artikel Ilmiah	Waktu dan Tempat
1	SISFOTEK 2019, IAI	Pemanfaatan AI Perwakilan Manajer dalam Implementasi Persetujuan Elektronik Berbasis SVM di Bidang SIM-SDM	25 September 2019 Jakarta, Indonesia

F. Karya Buku (5 Tahun Terakhir)

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit


G. Perolehan HKI (10 Tahun Terakhir)

No	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID

Jakarta, 17 Agustus 2023
Anggota Peneliti,

Arman Yusuf, S.Kom., M.M.

Lampiran 4. Surat Perjanjian Kontrak Penelitian

	UNIVERSITAS BUDI LUHUR FAKULTAS TEKNOLOGI INFORMASI KAMPUS PUSAT : Jl. Raya Ciledug, Petukangan Utara, Jakarta Selatan 12260 Telp. : (021) 5853753 (Hunting) Fax : (021) 7371164, 5853752 Website : http://www.budiluhur.ac.id
	92

SURAT PERJANJIAN KONTRAK PENELITIAN
Nomor: A/UBL/DRPM/00.1/005/06/23

Pada hari ini Selasa, 27 Juni 2023 Semester Genap Tahun Ajaran 2022/2023, kami yang bertandatangan di bawah ini:

1. **Krisna Adiyarta M, Ph.D.**, selaku Direktur Riset dan Pengabdian kepada Masyarakat Universitas Budi Luhur, selanjutnya disebut PIHAK PERTAMA.
2. **Teja Endra Eng Tju, S.T., M.Kom.**, selaku Peneliti selanjutnya disebut PIHAK KEDUA.

Kedua belah pihak menyatakan bersepakat untuk membuat perjanjian kontrak penelitian sebagai berikut:

Pasal 1
Judul Penelitian

PIHAK PERTAMA dalam jabatannya tersebut di atas, memberikan tugas kepada PIHAK KEDUA untuk melaksanakan penelitian yang berjudul: **Identifikasi Cerdas Hoaks Dengan Metode Klasifikasi Naïve Bayes.**

Pasal 2
Personalia Penelitian

Peneliti Utama : Teja Endra Eng Tju, S.Kom., M.Kom
Anggota : Arman Yusuf, S.Kom., M.Kom

Pasal 3
Waktu dan Biaya Penelitian

1. Waktu Penelitian adalah 3 (tiga) bulan, terhitung sejak tanggal 1 Juni 2023 sampai dengan 30 Agustus 2023.
2. Biaya pelaksanaan penelitian ini dibebankan pada Yayasan Pendidikan Budi Luhur Cakti tahun 2023 dengan nilai kontrak sebesar Rp. 5.000.000,- (Lima Juta rupiah).

Pasal 4
Cara Pembayaran

Pembayaran biaya penelitian diberikan bertahap, sebagai berikut

1. Tahap pertama sebesar 40% dari nilai kontrak, setelah perjanjian kontrak penelitian ini ditandatangani oleh kedua belah pihak.
2. Tahap kedua sebesar 30% dari nilai kontrak, setelah PIHAK KEDUA menyerahkan Laporan Hasil Penelitian kepada PIHAK PERTAMA.
3. Tahap ketiga sebesar 30% dari nilai kontrak, setelah PIHAK KEDUA menyerahkan Bukti Luaran Penelitian kepada PIHAK PERTAMA.

Pasal 5
Keaslian Penelitian

1. PIHAK KEDUA bertanggungjawab atas keaslian penelitian sebagaimana disebutkan dalam pasal 1 Surat Perjanjian Kontrak Penelitian ini dari penelitian orang lain.
2. PIHAK KEDUA bertanggungjawab atas keterikatan dengan pihak lain sebagaimana yang tercantum dalam Perjanjian Kerja Sama yang dijadikan rujukan dalam kegiatan penelitian.
3. PIHAK KEDUA menjamin bahwa penelitian bukan merupakan penelitian yang SUDAH atau SUDAH selesai dikerjakan, baik didanai oleh pihak lain ataupun oleh diri sendiri.
4. PIHAK PERTAMA tidak bertanggungjawab terhadap tindakan plagiat yang dilakukan oleh PIHAK KEDUA.

KAMPUS ROXY MAS : Pusat Niaga Roxy Mas Blok E.2 No. 38-39 Telp : (021) 6328709, 6328710, Fax : (021) 6322872
KAMPUS SALEMBA MAS : Sentra Salemba Mas Blok 5-T, Telp : (021) 3928688, 3928689, Fax : (021) 3181636



UNIVERSITAS BUDI LUHUR

FAKULTAS TEKNOLOGI INFORMASI

KAMPUS PUSAT : Jl. Raya Ciledug, Pelukangan Utara, Jakarta Selatan 12260

Telp : (021) 5853753 (Hunting) Fax : (021) 7371164, 5853752

Website : <http://www.budiluhur.ac.id>

- 5) Apabila dikemudian hari diketahui ketidakbenaran pernyataan ini, maka kontrak penelitian DINYATAKAN BATAL dan PIHAK KEDUA wajib mengembalikan dana yang telah diterima kepada Yayasan Pendidikan Budi Luhur Cakti sebagai pemberi dana

Pasal 6

Laporan Akhir Penelitian

PIHAK KEDUA wajib menyerahkan laporan akhir dalam bentuk softcopy, paling lambat tanggal 30 Agustus 2023.

Pasal 7

Sanksi

Segala kelalaian baik disengaja maupun tidak, sehingga menyebabkan keterlambatan menyerahkan laporan hasil penelitian dengan batas waktu yang telah ditentukan akan mendapatkan sanksi sebagai berikut:

- 1) Tidak diperbolehkan mengajukan usulan penelitian pada semester berikutnya bagi ketua dan anggota peneliti.
- 2) PIHAK KEDUA diberikan kesempatan perpanjangan waktu penelitian selama 2 (dua) minggu sampai dengan tanggal 14 September 2023.
- 3) Jika setelah masa perpanjangan tersebut PIHAK KEDUA tidak dapat menyelesaikan penelitiannya, PIHAK KEDUA diwajibkan mengembalikan dana yang sudah diterima kepada Yayasan Pendidikan Budi Luhur Cakti.

Pasal 8

Penutup

Perjanjian ini berlaku sejak ditandatangani dan disetujui oleh PIHAK PERTAMA dan PIHAK KEDUA.

PIHAK PERTAMA

Krisna Adiyarta M, Ph.D
NIP. 890001

Jakarta, 27 Juni 2023

PIHAK KEDUA

Teja Endra Eng Tju, S.T., M.Kom
NIP. 190046

Lampiran 5. Catatan Harian

No	Tanggal	Kegiatan
1	01/06/2023	Penyusunan proposal penelitian
2	07/06/2023	Pengumpulan dan pengesahan proposal penelitian
3	19/06/2023	Pengumpulan data penelitian
4	26/06/2023	Pemilihan dan pembangunan fitur-fitur dataset
5	27/06/2023	Pemodelan dengan metode <i>Naive Bayes</i>
6	03/07/2023	Pemodelan dengan metode <i>Decision Tree</i>
7	04/07/2023	Analisa hasil pemodelan
8	09/07/2023	Penulisan artikel untuk publikasi jurnal
9	31/07/2023	Reformat proposal penelitian

Lampiran 6. Artikel Ilmiah

Jurnal Ilmiah Pengabdian Masyarakat (JITEKSI) Nomor 31 - 2023/01/2023 tanggal 1 Juli 2023

JURNAL NASIONAL TEKNOLOGI & SISTEM INFORMASI

TEKNOLOGI

p-Hen 3490-3495
e-Issn 3476-8802

PENGUNA

Anda login sebagai...

teja

- [Email Saya](#)
- [Lupa Sandi](#)

Dewan Editorial

Revisi

Buat Unggah dan Publikasi

Publication Office

Pembaca untuk Penulis

Pembaca untuk Pembaca

Tajuk

Ulasan

Pengantar

Matahari Pengantar

TEMPLATE ARTIKEL

Journal
Template

TERHADAP DENGAN

TOOLS

ISSN

5 772476 881001

BERANDA
TENTANG JAWA
BERANDA PENGUNYA
CARI
TERMIN
ARSIP
INFORMASI
STATISTIK

Beranda > Pengantar > Penulis > Naskah > #2397 > Ringkasan

#2397 Ringkasan

[BROKURAH](#) | [TRUKAH](#) | [PENGANTAR](#)

Naskah

Penulis	Muhammad Umar Shalih, Anhan Yusuf, Teja Endra Eng Tju
Judul	Identifikasi Cerdas Hoaks dengan Metode Naive Bayes Ganda dan Decision Tree
File Asli	0107-3383-1-504.DOCX 2023-07-09
Indikasi File Terlampir	Tidak Ada TAMBAH FILE TAMBAHAN
Naskah	Teja Endra Eng Tju
Tanggal diunggah	July 8, 2023 - 06:17 AM
Bagian	
Editor	Huang Hanzi

Status

Status	Dalam Review
Dimulai	2023-07-09
Tersedia Dimodifikasi	2023-07-09

Metadata Naskah

BENTUK METADATA

Penulis

Nama	Muhammad Umar Shalih
Afiliasi	Universitas Budi Luhur
Negara	Indonesia
Biografi	—
Nama	Anhan Yusuf
Afiliasi	Universitas Budi Luhur
Negara	Indonesia
Biografi	—
Nama	Teja Endra Eng Tju
ID ORCID	https://orcid.org/0000-0001-0149-8802
Afiliasi	Universitas Budi Luhur
Negara	Indonesia
Biografi	—

Kontak Utama untuk Korespondensi Editorial:

Judul dan Sari

Judul	Identifikasi Cerdas Hoaks dengan Metode Naive Bayes Ganda dan Decision Tree
Abstrak	Identifikasi hoaks cukup kompleks dan menantang dengan permasalahan seperti heterogenitas hoaks, perubahan naskah yang cepat, kecepatan penyebaran yang luas, tanggapan teknologi canggih, kesulitan verifikasi, dan tantangan skala yang dihadapi. Sebagai kepedulian dengan hoaks pada masyarakat, penelitian identifikasi cerdas hoaks perlu dilakukan. Metodologi berbasis dan CRISP-DM dan Scikit-learn, 269 tahun 2020 yang dilaksanakan dengan ketekunan penelitian sehingga menjadi lima tahapan yaitu data understanding, data preparation, modeling, evaluation, dan deployment. Data diperoleh dari Mafindo dan digunakan sebanyak 2.750 data yang dibagi menjadi 7.004 data train dan 1.052 data uji. Terdapat enam fitur yaitu jumlah, kapital, keyword, sentimen, factcheck dan klasifikasi sebagai label supervisi. Dua fitur sentimen dan factcheck dibangun dengan metode Multinomial Naive Bayes, selanjutnya dilakukan pemodelan pada dataset dengan metode Decision Tree. Pemodelan dilakukan pula dengan variasi jumlah dataset sebanyak 1.000, 4.000, 8.000, 2000 juga dengan pertambahan masalah imbalance dataset. Hasil pemodelan dengan teknik Confusion Matrix diperoleh akurasi 63,5% dan skor F1 0,635 dan diperoleh bahwa imbalance dataset tidak terlalu berpengaruh pada hasil akurasi dan skor F1 namun menambah ketepatan model dalam hal jumlah besarnya data dengan label tertentu.

Pengindeksan

Daftar Ilmu dan Sub-Daftar	Informasi Sistem, Informasi Teknologi
Kata Kunci	CART, SMOTE, Confusion_Matrix, Fact_Check, Matrices
Bahasa	id