KLASIFIKASI HOAX PADA APLIKASI X MENGENAI PEMILIHAN UMUM CALON PRESIDEN DENGAN METODE ADABOOST

FELIX OSHWALT CHRISTIAN RUMAHORBO

201402101

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS SUMATERA UTARA

MEDAN

2024

KLASIFIKASI HOAX PADA APLIKASI X MENGENAI PEMILIHAN UMUM CALON PRESIDEN DENGAN METODE ADABOOST

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi

FELIX OSHWALT CHRISTIAN RUMAHORBO

201402101

PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024

PERSETUJUAN

Judul	: Klasifikasi Hoax Pada Aplikasi X Mengenai Pemilihan Umum Calon Presiden Menggunakan Adaboost
Kategori	: Skripsi
Nama Mahasiswa	: Felix Oshwalt Christian Rumahorbo
Nomor Induk Mahasiswa	: 201402101
Program Studi	: Sarjana (S-1) Teknologi Informasi
Fakultas	: Ilmu Komputer Dan Teknologi Informasi Universitas Sumatera Utara

Medan, 04 Juli 2024

Komisi Pembimbing :

Pembimbing 2,

Annisa Fadhillah Pulungan S.Kom., M.Kom
NIP. 199308092020012001

Pembimbing 1,

Sarah Purnamawati, S.T., M.Sc.
NIP. 198302262010122003

Diketahui/disetujui oleh
Program Studi S1 Teknologi Informasi Ketua,

Dedy Arisandi, S.T., M.Kom.
NIP. 197908312009121002

PERNYATAAN

KLASIFIKASI HOAX PADA APLIKASI X MENGENAI
PEMILIHAN UMUM CALON PRESIDEN
DENGAN METODE ADABOOST

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 04 Juli 2024

Felix Oshwalt Christian Rumahorbo
201402101

UCAPAN TERIMA KASIH

Puji dan syukut penulis panjatkan kepada Tuhan yang Maha Esa atass Rahmat dan karunia-Nya yang diberikan, yang telah menuntun penulis dalam menyelesaikan skripsi berjudul “Klasifikasi Hoax Pada Aplikasi X Mengenai Pemilihan Umum Calon Presiden Dengan Metode Adaboost” sebagai salah satu syarat untuk memperoleh gelar Sarjan Komputer di Program Studi Teknologi Indoemasi Universitar Sumatera Utara.

Disamping itu, penulis juga menyampaikan rasa terimakasih sebesar-besarnya kepada ber Bagai pihak, atas dukungan dan doa, sehinggal penulis dapat menyusun skripsi sampai selesai. Oleh karena itu pada kesempatan ini, penulis mengucapkan terima kasih dengan ketulusan hati kepada:

1. Tuhan Yesus Kristus yang selalu memberikan Kesehatan, kesembuhan, kemudahan dan kemampuan kepada penulis dalam proses penelitian dari awal hingga tahap akhir penulisan.
2. Keluarga penulis Ayah, Ibu, Adik, seluruh keluarga Rumahorbo, Haloho dan Saragih yang selalu memberikan dukungan, doa dan semangat hingga akhirnya penulis dapat menyelesaikan skripsi ini.
3. Ibu Dr. Maya Silvi Lydia, M.Sc selaku Dekan Fasilkomk-TI USU
4. Bapak Dedy Arisandi ST., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara dan Bapak Ivan Jaya S.Si., M.Kom., selaku Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
5. Ibu Sarah Purnamawati S.T., M.Sc., selaku Dosen Pembimbing I dan Ibu Annisa Fadhillah Pulungan S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan bimbingan bagi penulis dengan memberikan saran dan masukan serta waktu yang berharga untuk membimbing penulis.
6. Dosen pengajar di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara
7. Staff dan pegawai di Program Studi S1 Teknologi Informasi Universitas Sumatera Utara
8. Teman-Teman Angkatan 20230 Teknologi Informasi Universitas Sumatera Utara yang memberikan dukungan dan motivasi kepada penulis

9. Teman-teman seperjuangan pennulis dalam masa perkuliahan penulisan skripsi Yeftha El Imani Pardede, Kevin Aryanda Yeremua Bangun, Vicky Natanael Sinaga, Albert Lukas Talupan Pangaribuan, Pretty Ohara Hutasoit, Kelvin Nathanael Lumbanraja, Kevin Tulus Silitonga, Arief Fadhlan, Yericho Natanael.
10. Abang kakak Angkatan 2019 selaku sahabat yang penulis kasihi yang selalu memberikan saran, semangat dan motivasi kepada penulis Sarah Theodora Sinurat, Vania Miranda Emmanuella Siahaan, Dhea Novianty Sitompul.
11. Erika Yolanda Silalahi, S.E selaku orang yang selalu memberikan saran, dukungan dan bantuan serta motivasi dan menjadi teman seperjuangan dalam penulisan skripsi dari tahap awal hingga akhir.
12. Semua pihak berperan secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu persatu yang telah banyak membantu penulis dalam penyelesaian skripsi ini.
13. Kepada diri sendiri yang sudah semangat, kaut dan tidak menyerah dalam mengerjakan pennulisan skripsi ini dari awal hingga akhir

Akhir kata, Penulis berharap agar skripsi ini dapat bermanfaat bagi semua orang yang membaca dan kiranya Tuhan Yang Maha Esa selalu memberkati seluruh pihak yang turut serta membantu dan mendukung penulis dalam penyelesaian skripsi ini.

Medan, 04 Juli 2024

Penulis,

Felix Oshwalt Christian Rumahorbo

ABSTRAK

Proses klasifikasi hoaks di aplikasi X terkait pemilihan presiden umumnya dilakukan secara manual oleh para ahli. Keterbatasan sumber berita yang valid untuk memverifikasi kebenaran suatu cuitan membuat pembaca mudah mempercayai informasi yang belum tentu benar. Teknologi informasi mempercepat penyebaran informasi, tetapi juga dapat membantu pembaca dalam memverifikasi kebenaran informasi melalui media sosial, televisi, dan aplikasi pesan. Penelitian ini bertujuan untuk menentukan keakuratan informasi dan mengidentifikasi jenis hoax yang beredar. Banyak orang masih menganggap berita sebagai hoax jika informasi tersebut tidak benar. Oleh karena itu, penelitian ini juga bertujuan untuk memberikan informasi kepada masyarakat tentang berbagai jenis hoax yang banyak beredar. Penelitian ini menggunakan metode Adaptive Boosting dengan Decision Tree sebagai base estimator untuk meningkatkan analisis dan pengolahan data. Data yang digunakan berasal dari cuitan di aplikasi X mengenai topik pemilihan presiden dan wakil presiden dalam bahasa Indonesia. Data sebanyak 954 dibagi menjadi 80% untuk data training dan 20% untuk data testing. Model ini mencapai tingkat akurasi 91.15% untuk identifikasi hoax dan non-hoax, serta 83.77% untuk klasifikasi informasi, dengan evaluasi cross validation dan matriks evaluasi.

Kata Kunci : hoax, non-hoax, klasifikasi hoax, *Adaptive Boosting, Decision Tree,* tweet hoax.

CLASSIFICATION OF HOAX ON THE X APPLICATION REGARDING PRESIDENTIAL ELECTION USING THE ADABOOST METHOD

ABSTRACT

The process of classifying hoax news on the X application related to the presidential election is generally carried out manually by experts. The limitation of valid news sources to verify the truth of a tweet makes readers easily believe information that may not be true. Information technology accelerates the spread of information but can also help readers verify the truth of information through social media, television, and messaging applications. This study aims to determine the accuracy of information and identify the types of hoaxes circulating. Many people still consider news as a hoax if the information is not true. Therefore, this study also aims to inform the public about various types of hoaxes that are widely circulated. This research uses the Adaptive Boosting method with Decision Tree as the base estimator to improve data analysis and processing. The data used comes from tweets on the X application regarding the topic of the presidential and vice-presidential elections in Indonesian. A total of 954 data points were divided into 80% for training data and 20% for testing data. This model achieves an accuracy rate of 91.15% for identifying hoaxes and non-hoaxes, and 83.77% for information classification, with evaluation using cross-validation and evaluation matrics

Keywords : *hoax, non-hoax, hoax classification, Adaptive Boosting, Decision Tree, hoax tweet.*

DAFTAR ISI

PERSETUJUAN i
PERNYATAAN ii
UCAPAN TERIMA KASIH iii
ABSTRAK v
ABSTRACT vi
DAFTAR ISI vii
DAFTAR TABEL ix
DAFTAR GAMBAR x
DAFTAR PSEUDOCODE xi
BAB 1 PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 3
1.3. Tujuan Penelitian 3
1.4. Batasan Masalah 4
1.5. Manfaat Penelitian 4
1.6. Metodolgi Penelitian 4
1.7. Sistematikan Penulisan 6
BAB 2 LANDASAN TEORI 8
2.1.Hoaks 8
2.1.1. Misleading Content 8
2.1.2. Fabricated Content 9
2.1.3. Manipulated Content 9
2.2. Website 10
2.3. Flask 10
2.4. Text Preprocessing 11
2.5. N-Gram 11
2.6. Term Frequency Inverse Document Frequency (TF-IDF) 12
2.7. Adaptive Boosting (AdaBoost) 13
2.8. Confusion Matrix 15
2.9. Cross Validation 16

2.10. Penelitian Terdahulu 16
2.11. Perbedaan Penelitian 29
BAB 3 METODOLOGI PENELITIAN 30
3.1. Dataset 30
3.2. Arsitektur Umum 34
3.2.1. Input Data 35
3.2.2. Pre-processing 36
3.2.1.1. Case Folding 37
3.2.1.2. Cleaning 39
3.2.1.3. Tokenisasi 40
3.2.1.4. Stopword Removal 41
3.2.3. TF-IDF 42
3.2.4. SelectKbest 45
3.2.5. Adaptive Boosting 47
3.3. Perancangan Aplikasi Sistem 51
3.3.1. Gambaran Tampulan Halaman Depan 51
3.3.2. Rancangan Tampilan Training dan Testing 52
3.3.3. Rancangan Tampilan User Input 53
3.4. Metode Evaluasi 54
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM 56
4.1. Imlementasi Sistem 56
4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak 56
4.1.2. Penerapan Sistem Berbasis Web 56
4.2. Hasil Pelatihan Model 60
4.3. Hasil Pengujian Sistem 64
4.4. Metode Evaluasi 67
BAB 5 KESIMPULAN DAN SARAN 77
5.1. Kesimpulan 77
5.2. Saran 78
DAFTAR PUSTAKA 79

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu 20
Tabel 3.1 Dataset 30
Tabel 3.2. Kata Kata Pada Fabricated 32
Tabel 3.3. Kata Kata Pada Misleading 33
Tabel 3.4 Kata Kata Pada Manipulated 33
Tabel 3.5 Contoh Tweet Hoax dan Non-hoax 36
Tabel 3.6 Implementasi Proses Cleaning 37
Tabel 3.7 Implementasi Proses Cleaning 40
Tabel 3.8 Implementasi Proses Tokenize 41
Tabel 3.9 Implementasi Stopwords Removal 42
Tabel 3.10 Implementasi TF-IDF Vectorizer 44
Tabel 3.11 Implementasi Fitur N-Gram 45
Tabel 3.12 Hasil Pemilihan Fitur SelectKBest 46
Tabel 3.13 Perhitungan Accuracy 55
Tabel 4.1 Daftar parameter 60
Tabel 4.2 Hasil Akurasi Kombinasi Parameter 61
Tabel 4.3 Hasil Pengujian Sistem pada Data Testing 64
Tabel 4.4 Keterangan Confusin Matrix Label Non-hoax 68
Tabel 4.5 Keterangan Confusin Matrix Label Hoax 69
Tabel 4.6 Keterangan Confusin Matrix Klasifikasi Non-Hoax 70
Tabel 4.7 Keterangan Confusin Matrix Klasifikasi Fabricated 71
Tabel 4.8 Keterangan Confusin Matrix Klasifikasi Misleading 72
Tabel 4.9 Keterangan Confusin Matrix Klasifikasi Manipulated 73
Tabel 4.10 Hasil Perhitungan Evaluasi Pada Label 75
Tabel 4.11 Hasil Perhitungan Evaluasi Pada Klasifikasi 75

DAFTAR GAMBAR

Gambar 2.1 Contoh tweet hoaks (x.com) 8
Gambar 2.2 Contoh Misleading Content 9
Gambar 2.3 Contoh Fabricated Content 9
Gambar 2.4 Contoh Manipulated Content 10
Gambar 2.5 Ilustrasi Adaboost 14
Gambar 3.1 Arsitektur Umum 35
Gambar 3.2 Alur Kerja Adaboost 48
Gambar 3.3 Rancangan Tampilan Home 52
Gambar 3.4 Rancangan Tampilan Training dan Testing 53
Gambar 3.5 Rancangan Tampilan User Input 54
Gambar 4.1 Halaman Home 57
Gambar 4.2 Halaman Home Info Hoaks 57
Gambar 4.3 Halaman Home Contoh Hoax 57
Gambar 4.4 Input Teks 58
Gambar 4.5 Output Klasifikasi 58
Gambar 4.6 Halaman Training dan Testing 59
Gambar 4.7 Hasil Training 59
Gambar 4.8 Hasil Testing 60
Gambar 4.9 Confussion Matrix Non-Hoax pada Label 68
Gambar 4.10 Confussion Matrix Pada Label Hoax 69
Gambar 4.11 Confussion Matrix Pada Klasifikasi Non-hoax 70
Gambar 4.12 Confussion Matrix Fabricated Pada Klasifikasi 71
Gambar 4.13 Confussion Matrix Misleading Pada Klasifikasi 72
Gambar 4.14 Confussion Matrix Manipulated Pada Klasifikasi 73
Gambar 4.15 Confusion Matrix Pada Gabungan Klasifikasi 76

DAFTAR PSEUDOCODE

Pseudocode 3.1 Case Folding 37
Pseudocode 3.2 Clean Text 40
Pseudocode 3.3 Tokenisasi 41
Pseudocode 3.4 Stopwords Removal 42
Pseudocode 3.5 TF-IDF Vectorizer 44
Pseudocode 3.6 KBest Selector 46
Pseudocode 3.7 Model Adaboost pada Label 50
Pseudocode 3.8 Model Adaboost pada Klasifikasi 51

BAB 1
PENDAHULUAN

1.1. Latar Belakang

Saat ini, teknologi mengalami kemajuan yang sangat pesat, terutama di bidang teknologi informasi. Teknologi informasi adalah teknologi yang digunakan untuk mengolah data, termasuk dalam hal memproses, mendapatkan, menyusun, dan memanipulasi data untuk menghasilkan informasi yang berkualitas. Dengan adanya teknologi informasi, penyebaran informasi menjadi sangat cepat, baik di dunia nyata maupun di dunia maya, khususnya melalui media sosial yang kita gunakan sehari-hari. Media sosial merupakan platform daring yang memungkinkan interaksi sosial tanpa batasan ruang dan waktu. Media sosial memiliki pengaruh besar dalam kehidupan kita saat ini, seseorang yang awalnya memiliki karier kecil dapat menjadi sukses berkat media sosial, namun sebaliknya, seseorang dengan karier besar dapat kehilangan semuanya dalam hitungan detik.

Hoaks menurut Kamus Besar Bahasa Indonesia (KBBI) adalah kabar, informasi, atau berita bohong yang disampaikan untuk tujuan iseng, lelucon, atau menjatuhkan pesaing (kampanye hitam). Hoaks telah menimbulkan keresahan di kalangan masyarakat dan dapat memecah belah persatuan. Di Indonesia, media sosial menjadi sarana utama penyebaran hoaks dengan persentase 87.5%. Aplikasi chatting memiliki persentase 67%, website 28.2%, televisi 8.7%, media cetak 6.4%, dan email dengan tingkat terendah 2.6% .Berita hoaks dikategorikan menjadi beberapa jenis seperti konten yang sepenuhnya dibuat-buat (Fabricated Content), konten yang dimanipulasi (Manipulated Content), dan konten yang menyesatkan (Misleading Content). Saat ini, penyebaran hoaks mencakup berbagai bidang seperti olahraga, budaya, ekonomi, bahkan politik. Pemilihan umum merupakan sarana demokrasi untuk mewujudkan sistem pemerintahan yang berkedaulatan, di mana pemerintah dibentuk dari rakyat, dijalankan sesuai kehendak rakyat, dan diabdikan untuk

kesejahteraan rakyat.Pemilihan Umum di Indonesia dilakukan sejak 1955 untuk memilih anggota-anggota DPR, namun untuk pemilihan umum presiden dan wakil presiden pertama kali dilakukan pada tahun 2004.(Sriyano & Setiawan, 2021).

Meskipun pemilihan umum presiden dan wakil presiden telah selesai, masih banyak berita hoaks yang beredar di masyarakat, mulai dari masa kampanye hingga penghitungan akhir pemilu. Oleh karena itu, penulis melakukan penelitian ini untuk mengklasifikasikan berita hoaks yang tersebar di media sosial, khususnya di platform aplikasi X. Data tersebut dikategorikan berdasarkan kriteria yang telah ditentukan menggunakan metode Adaboost dan pembobotan Term Frequency Inverse Document Frequency (TF-IDF). Penulis memilih Adaptive Boosting karena memiliki tingkat akurasi yang tinggi, mudah diimplementasikan, waktu komputasi yang rendah, dan tingkat kesalahan yang minimal.

Penelitian ini bertujuan untuk memberikan pemahaman mengenai adanya klasifikasi hoaks yang beredar dikalangan masyarakat, selain itu juga bertujuan untuk mengetahui penggunaan metode Adaboost dalam mengklasifikasikan berita, serta bagaimana tingkat akurasi dari pendeteksian berita hoaks apakah termasuk kedalam *Fabricated Content, Manupilated Content dan Misleading content* dan menganalisis trend penyebaran hoaks pada pemilihan umum presiden dan calon presiden dimana termasuk kedalam klasifikasi hoaks yang akan diklasifikasikan berdasarkan jenisnya, mengetahui dampak penyebaran hoaks berdasarkan kaslifikasinya serta mengetahui tingkat akurasi klasifikasi hoaks menjadi beberapa jenis menggunakan metode *Adaboost*.

Penelitian ini juga memasukan kajian terhadap penelitian-penelitian sebelumnya untuk melakukan pembandingan dan memahami solusi-solusi yang telah diterapkan dalam menangani permasalahan yang askan diteliti. Oleh karena itu, peneliti merinci beberapa penelitian terdahulu, termasuk studi yang dilakukan oleh Candra Surya Sriyanto dan Budi Setiawan yang melalukan peneitain yang bertopik "Pendeteksian Berita hoaks Menggunakan *Naïve Bayes Multinomial* pada Twitter dengan fiitur pembobotan TF-IDF" penulis melihat bahwa penyebaran banyak penyebaran informasi yang tidak sesuai fakta, banyak yag tidak sesuai dengan kebenarannya, makan peneliti membangun sistem untuk mendeteksi hoaks pada media sosial menggunakan metode *Multinomial Naïve Bayes* dengan pembobotan TF-IDF,

dengan menggunakna metode-metode tersebut peneliti berhasil mendapat akurasi tertinggi sebesar 76%(Sriyano & Setiawan, 2021) .

Selanjutnya dalah penelitian dKlasifikasi Polusi Udara dengan Adaboost berbasis Pembelajaran Mesin. Dalam penelitian tersebut, penulis memiliki tujuan untuk mengkategorikan polusi udara, termasuk polutan seperti Konsentrat Partikulat, Karbon Monoksida, dan Sulfur Monoksida. Harapannya adalah dapat memantau dan mengklasifikasikan polusi udara di kota-kota besar, dan hasil klasifikasinya, yang diuji menggunakan program dengan model machine learning, mencapai tingkat akurasi sekitar 90% (Airell Rizqiawan et al., 2023).

Dengan uraian latar belakang diatas, maka dengan demikian peneliti mengajukan penelitian dengan judul "Klasifikasi Hoaks Pada Aplikasi X Mengenai Pemilihan Umum Calon Presiden Menggunakan Metode *Adaptive Boosting* "

1.2. Rumusan Masalah

Adanya sebuah informasi atau berita sudah sangat lekat dengan kehidupan sehari-hari dimana kita bisa mendapatkan informasi megenai suatu peristiwa atau kejadian yang dianggap penting oleh media atau sumber berita, dimana tujuan dari berita itu adalah menyampaikan fakta-fakta terkini dan memberikan pemahaman kepada masyarakat tentang kejadian disekitarnya. Berita juga mencakup beberapa bidang seperti politik, ekonomi, sosial, budaya, olahraga serta topik lainnya. Namun dewasa ini semakin cepatnya berita tersebar maka banyak informasi-informasi yang tidak pasti kebenarnnya atau yang biasa kita dengar dengan Hoaks (Sriyano & Setiawan, 2021).

Pada masa pemilihan umum presiden dan wakil presiden, banyak informasi yang beredar dikalangan publik baik secara nyata maupun melalui media sosial. Permasalahan yang sering terjadi, masyarakat masih susah dalam membedakan informasi fakta atau hoaks, dan jika masyarakat mengetahui informasi tersebut adalah hoaks, masyarakat belum mengetahui jenis hoaks tersebut.

1.3. Tujuan Penelitian

Adapun Tujuan penelitian ini adalah melakukan klasifikasi terhadap trend penyebaran informasi pada pemilihan umum calon presiden dan calon wakil presiden yang memungkinkan adanya hoaks dalam penyebaran informasi, dimana informasi hoaks akan diklasifikasikan berdasarkan jenisnya.

1.4. Batasan Masalah

Berdasarkan pemaparan latar belakang diatas, maka Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data yang akan di analisis adalah berupa teks.
2. Data yang akan di gunakan adalah bahasa Indonesia.
3. Pengumpulan data pada penelitian ini akan dilakukan dengan Crawiliing data.
4. Penelitian ini melakukan identfikasi tweet hoaks dan non-hoaks, dan jika datanya termasuk hoaks akan diklasifikasikan menjadi *Fabricated content, Misleading content, Manipulated conten.*
5. Penelitian ini mengambil data pada aplikasi X tentang pemilihan calon presiden dan wakil presiden

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

1. Memberikan pengetahuan kepada pengguna atau masyarakat agar mereka dapat menilai kebenaran atau ketidakbenaran sebuah tweet yang mereka baca.
2. Menyampaikan pemahaman kepada pembaca mengenai beragam jenis hoaks yang sering tersebar di media sosial, sehingga mereka lebih waspada dan terinformasi secara lebih baik.

1.6. Metodolgi Penelitian

Adapun metodologi penelitian yang akan dilakukan dalam penelitian ini, yaitu :Studi Literatur

1. Studi Literatur

   Peneliti melakukan penelitian dengan memulai dengan mencari dan mengumpulkan berbagai sumber referensi yang telah diteliti seperti artikel, jurnal, skripsi, internet maupun sumber informasi lainnya yang memiliki keterikatan tentang *Text Preprocessing*, *Crawling*, penggunaan metode *Adaptive Boosting* dan elemen-elemen lain yang akan digunakan pada penelitian ini.

2. Pengumpulan Data

Peneliti melakukan pengumpulan data dengan cara Crawling data tweet menggunakan API pada aplikasi X berdasarkan keyword keyword seperti ‘Pemilu curang’, ‘Anak haram konstitusi’ menggunakan *Google Colab,* lalu melakukan labeling manual dengan cara Crowdsourcing dan menggunakan bantuan ahli untuk menentukan data yang berlabel ‘hoax’ dan ‘non-hoax’ lalu label untuk klasifkasi ‘ *Fabricated* ’, ‘*Misleading*’ dan ‘*Manipulated* ’.

3. Analisis Permasalahan

Pada tahap ini, identifikasi dan klasifikasi akan dilakukan berdasarkan data yang digunakan untuk mendapatkan informasi yang beragam. Tujuan pada tahap ini untuk memahami konsep dari metode *Adaptive Boosting* dan elemen-elemen yang terlibat sehingga dapat diterapkan dalam melakukan identifikasi tweet hoaks dan klasifikasi hoaks berdasarkan jenis klasifikasi yang diberikan.

4. Perancangan Sistem

Pada tahap ini, peneliti akan merancang arsitektur umum untuk masalah yang diteliti. Perancangan ini dimulai dengan pelabelan data untuk mengidentifikasi label hoax dan non-hoax, kemudian memberikan label klasifikasi untuk jenis hoax sesuai dengan kategori yang telah ditentukan. Setelah itu, dilakukan tahapan preprocessing teks lainnya hingga mencapai pemodelan menggunakan Adaptive Boosting. Tahap ini bertujuan untuk membentuk kerangka kerja yang sistematis dan kuat, memastikan bahwa semua elemen yang diperlukan telah direncanakan dengan baik dan siap digunakan dalam pembuatan sistem.

5. Implementasi

Pada tahap ini peneliti akan mengimplementasi yang melibatkan prencanaan arsitektur umum menjadi suatu sistem yang konkret. Pada tahap ini akan menentukan website yang baru dibuat dan akan dioperasikan pada keadaan sebenarnya dan tahap ini juga akan diketahui mengenai efektifitas sistem akan diketahui seperti kelebihan dan kekurangan dari sistem dan website yang dibuat.

6. Pengujian

Tahapan selanjutnya adalah tahap pengujian, tahap ini melibatkan uji performa sistem yang telah direncanakan untuk mendapat hasil dari penerapan metode *Adaptive Boosting* dalam penelitian ini.

7. Tahap Laporan

Tahap selanjutnya adalah tahap laporan, pada tahap ini peneliti akan melakukan penyusunan laporan sebagai dokumentasi peneliti dan merupakah hasil akhir dari penelitian yang dilakukan.

1.7. Sistematikan Penulisan

Pada penelitian ini, sistematika penulisan mencakup 5 (lima) bagian , yaitu :

Bab 1 : Pendahuluan

Bab ini akan menjelaskan berbagai elemen penting dari penelitian ini. Dimulai dengan latar belakang yang menjadi dasar penelitian, kemudian rumusan masalah yang menjadi fokus utama. Tujuan penelitian akan diuraikan untuk mengidentifikasi apa yang ingin dicapai. Batasan masalah akan dijelaskan untuk memperjelas ruang lingkup penelitian. Manfaat penelitian akan diuraikan untuk menunjukkan kontribusi yang diharapkan, baik secara teoritis maupun praktis. Metodologi penelitian akan dijelaskan secara rinci, termasuk pendekatan, teknik, dan prosedur yang digunakan. Terakhir, sistematika penelitian akan disajikan untuk memberikan gambaran mengenai struktur dan alur penulisan penelitian ini.

Bab 2 : Landasan Teori

Landasan teori ini akan memuat berbagai teori yang diterapkan untuk memahami masalah dalam penelitian ini. Bab ini akan menjelaskan pengertian hoaks, berbagai jenis hoaks, *Text Preprocessing* teori yang relevan dalam proses identifikasi dan klasifikasi tweet hoaks.

Bab 3 : Analisis dan Perancangan Sistem

Pada tahap ini akan dijelaskan mengenai analisis dan perancangan sistem serta pembuatan desain untuk penerapan algoritma Adaptive Boosting dalam identifikasi dan klasifikasi tweet hoaks berdasarkan jenis-jenis yang telah ditentukan. Bab ini juga

akan menguraikan langkah-langkah penyelesaian masalah dan membahas arsitektur umum dari sistem yang dikembangkan.

Bab 4 : Implementasi dan Pengujian Sistem

Selanjutnya, bab ini akan membahas implementasi dan pengujian sistem. Bagian ini akan menjelaskan bagaimana analisis dan desain diimplementasikan berdasarkan rancangan yang telah dibuat sebelumnya, serta menguraikan hasil akhir sistem dan pelaksanaan evaluasi terhadap sistem yang telah dibangun.

Bab 5 : Kesimpulan dan Saran

Bab ini merupakan bab terakhir dari penelitian ini. Di dalamnya akan disajikan kesimpulan yang didasarkan pada keseluruhan pembahasan dari bab-bab sebelumnya, serta saran untuk penelitian dan pengembangan lebih lanjut.

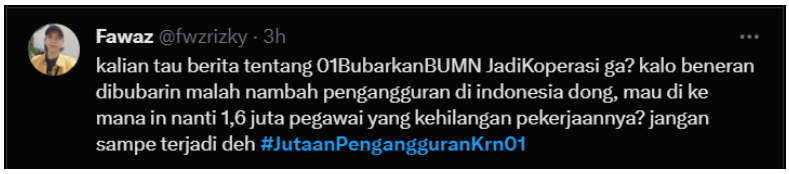
BAB 2
LANDASAN TEORI

2.1. Hoaks

Hoaks adalah informasi atau berita palsu yang disebarkan untuk menipu atau memanipulasi orang yang menerimanya. Tujuannya beragam, seperti mempengaruhi opini publik, menciptakan ketidakpercayaan terhadap individu atau kelompok tertentu, atau merugikan pesaing melalui kampanye hitam. Selain itu, hoaks juga bisa digunakan untuk promosi dengan cara yang tidak jujur, menciptakan persepsi palsu untuk mendukung agenda tertentu. Penyebaran hoaks dapat berdampak negatif signifikan pada informasi yang benar dan integritas komunikasi (Sriyano & Setiawan, 2021).

Menurut Dewan Pers, ciri-ciri hoaks adalah sebagai berikut:

1. Mengakibakan kecemasan, kebencian dan Permusuhan
2. Sumber berita tidak jelas, biasanya hoaks yang beredar dimedia sosial tidak terverifikasi, tidak berimbang dan menyudutkan pihak tertentu.
3. Bermuatan fanatisme atas nama ideologi, dan pengantarnya provokatif dan memberikan penghukuman serta menyembunyikan fakta atau data.



Gambar 2.1 Contoh tweet hoaks (x.com)

Berikut adalah beberapa jenis hoaks menurut Kementrian Keuangan Republik Indonesia:

2.1.1. Misleading Content

Konten yang menyesatkan merujuk pada penggunaan informasi untuk mengarahkan pemahaman suatu isu atau individu dengan cara yang tidak objektif. Pembuat konten yang bersifat menyesatkan dapat dengan sengaja menciptakan informasi dengan

mengutip pernyataan tanpa menyertakan konteksnya, atau menampilkan data statistik yang hanya mendukung pandangan tertentu, sambil mengabaikan informasi lain yang relevan. Dengan demikian, informasi tersebut menjadi tidak lengkap dan dapat mempengaruhi opini pembaca sesuai dengan keinginan pembuat konten.

Gambar 2.2 Contoh Misleading Content
Sumber : https://x.com

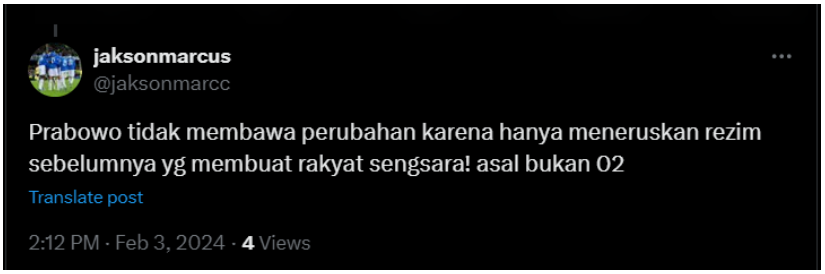
2.1.2. Fabricated Content

Fabricated content dianggap sebagai jenis konten palsu yang memiliki potensi bahaya yang tinggi. Jenis konten ini sepenuhnya dibuat-buat dan tidak dapat dipertanggungjawabkan secara faktual. Umumnya, fabricated content muncul dalam bentuk informasi palsu terkait lowongan pekerjaan dan topik lainnya yang dimana informasi tersebut tidak ada sama sekali.

Gambar 2.3 Contoh Fabricated Content
Sumber : https://x.com

2.1.3. Manipulated Content

Manipulated Content adalah hasil editan dari informasi yang sebelumnya telah diterbitkan oleh media besar dan kredibel. Dengan kata lain, konten jenis ini terbentuk melalui pengeditan terhadap materi yang sudah ada, dengan tujuan menyesatkan publik. Sumber dari jenis konten ini adalah materi asli yang mengalami manipulasi, baik sebagai lelucon, untuk memprovokasi pembaca, menyebarkan propaganda, atau untuk kepentingan politik tertentu.



Gambar 2.4 Contoh Manipulated Content
Sumber : https://x.com

2.2. Website

Website pertama kali ditemukan oleh Sir Tim Berners-Lee pada tahun 1989 di CERN, Swiss, untuk memudahkan pertukaran informasi antar ilmuwan. Pada tahun 1990, ia mengembangkan tiga teknologi dasar: HTML, URI, dan HTTP, serta menciptakan halaman web dan server web pertama. Website pertama kali diperkenalkan kepada publik pada tahun 1991

Pada tahun 1993, browser grafis pertama, Mosaic, dirilis, memungkinkan tampilan teks dan gambar secara bersamaan, yang membuat web lebih menarik dan mudah digunakan. Perkembangan teknologi seperti CSS dan JavaScript pada pertengahan 1990-an memungkinkan kontrol yang lebih baik atas desain dan interaktivitas halaman web Pengembangan website melibatkan berbagai bahasa pemrograman dan teknologi. Untuk frontend, digunakan bahasa seperti HTML, CSS, dan JavaScript serta framework seperti React, Angular, dan Vue.js. Di sisi backend, bahasa seperti Python, PHP, Ruby, dan Java, serta teknologi server seperti Node.js sering digunakan. Database seperti MySQL, PostgreSQL, dan MongoDB biasa digunakan untuk penyimpanan data, memungkinkan website tidak hanya menyampaikan informasi tetapi juga mendukung aplikasi web yang kompleks dan interaktif. Implementasi ini memungkinkan pembuatan website yang dinamis dan mampu menangani berbagai fungsi interaktif.

2.3. Flask

Flask adalah sebuah web framework minimalis yang ditulis dalam bahasa Python. Framework ini dirancang untuk mempermudah pembuatan aplikasi web yang ringan dan mudah dikembangkan. Flask tidak memerlukan alat atau library tertentu untuk mulai digunakan, menjadikannya fleksibel dan mudah disesuaikan. Dua komponen

utama yang mendukung Flask adalah WSGI (Web Server Gateway Interface), yang memungkinkan interaksi antara server dan aplikasi web, serta Jinja2, yang merupakan engine template untuk rendering HTML. Flask juga mendukung penggunaan ekstensi pihak ketiga untuk menambah fungsionalitas, seperti validasi formulir, pengelolaan unggahan file, dan integrasi basis data(Darmawan C A, 2023).

2.4. Text Preprocessing

Pemrosesan Teks adalah tahap awal yang penting dalam ekstraksi dokumen, di mana teks diubah menjadi informasi yang siap digunakan. Proses ini melibatkan pemecahan teks menjadi unit-unit kecil yang disebut token. Selain itu, simbol-simbol yang tidak relevan dan angka-angka juga dihapus untuk mengurangi kebisingan dalam data. Langkah-langkah seperti penghapusan stop words, stemming, dan lemmatization juga dilakukan untuk menyempurnakan teks. Semua ini bertujuan untuk mempersiapkan teks mentah agar dapat dianalisis lebih efektif dan akurat, sehingga memungkinkan pengambilan keputusan yang lebih baik berdasarkan data teks yang telah dibersihkan dan disiapkan. (Listiana E & Muslim A M, 2017)

Text processing merupakan salah satu teknik untuk mengolah data dari teks yang tidak terstruktur menjadi suatu data teks tyang terstruktur. Pada *text processing* ini melewati beberapa tahapan seperti *cleaning, case folding, stemming, stopwords, normalization, tekonization* dan sebagainya. Selain dari itu tujuan text processing ini adalah tahap awal untuk mepersiapkan text menjadi data yang akan di olah selanjutnya(Feldman & Sanger, 2007) ,(Herwijayanti et al., 2018).

2.5. N-Gram

N-Gram adalah sekumpulan item berurutan dalam konteks tertentu yang dapat berupa angka, kata, simbol, atau tanda baca. Model ini sangat berguna dalam analisis teks di mana urutan kata-kata memiliki relevansi, seperti dalam analisis sentimen, klasifikasi teks, dan deteksi teks. N-Gram juga merupakan model probabilistik yang dikembangkan untuk memprediksi item berikutnya dalam sebuah urutan. Dalam aplikasi koreksi ejaan, N-Gram merupakan urutan sejumlah N kata dalam suatu kata atau string.

Ada beberapa jenis N-Gram yang sering digunakan, di antaranya:

1. Uni-gram adalah sekumpulan kata atau karakter dalam teks misalnya, “saya”, “ sedang”, “makan”, “buah”.
2. Bi-gram adalah sekumpulan duakata atau karakter dalam teks misalnya, “saya melihat”, “rumah besar”, “jalan raya”.
3. Tri-gram adalah tiga kata atau karakter dalam teks. misalnya, “saya melihat dia”, “jokowi adalah presiden”, “ anis adalah gubernur”.

2.6. Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode yang digunakan untuk menentukan frekuensi kemunculan sebuah kata dalam sebuah dokumen atau artikel, serta frekuensi dokumen itu sendiri (Evan, 2014). Algoritma TF-IDF memberikan pembobotan pada setiap kategori untuk menemukan kemiripan dengan kategori lain, namun sebelumnya dilakukan tahap preprocessing. TF-IDF berasal dari perhitungan antara TF (Term Frequency) dan IDF (Inverse Document Frequency). TF memberikan informasi tentang seberapa sering istilah muncul dalam dokumen. Karena dokumen bisa memiliki panjang yang berbeda, frekuensi kata akan dibagi dengan panjang dokumen. Misalnya, jika dokumen lebih panjang, jumlah kemunculan kata ‘x’ akan dibagi dengan panjang dokumen ‘y’, sehingga hasilnya berbeda dengan dokumen yang pendek (Widaningrum et al., 2022).

IDF memberi informasi tentang kelangkaan relative suatu istilah dalam sebuah dokumen. Jika nilai *IDF* semakin kecil maka kata tersebut akan dianggap tidak penting, begitu juga jika data semakin banyak maka data tersebut dianggap penting. Contoh kata yang sering muncul namun tidak penting adalah seperti “dan”, “yang”, “ini”, “adalah”, “merupakan”, dan lain lain. Berikut adalah rumus *IDF :*(Komang N et al, 2018)

$$TF_{(kd)} = \frac{f_{(k)}}{\max f_{(d)}} \qquad (2.1)$$

$$IDF_{(k)} = log\frac{N}{DF} \qquad (2.2)$$

$$TF - IDF_{(kd)} = TF_{(kd)} x IDF_{(k)} \qquad (2.3)$$

Keterangan	:	
TF(kd)	:	Term Frequency untuk kata i dalam dokumen j
f(k)	:	Frekuensi kemunculan k dalam dokumen d
max f(d)	:	Frekuensi maksimum dari kata yang paling sering muncul dalam dokume d
IDF (kd)	:	Inverse Document Frequency untuk kata i.
N	:	Jumlah total dokumen dalam koleksi.
DF (k)	:	Jumlah dokumen yang mengandung kata i
TF-IDF kd	:	Nilai TF-IDF untuk kata i dalam dokumen j.
k	:	Kata
d	:	Dokumen
F (k)	:	Frekuensi kata
F (d)	:	Frekuensi dalam dokumen
DF (k)	:	Jumlah dokumen yang mengandung kata k

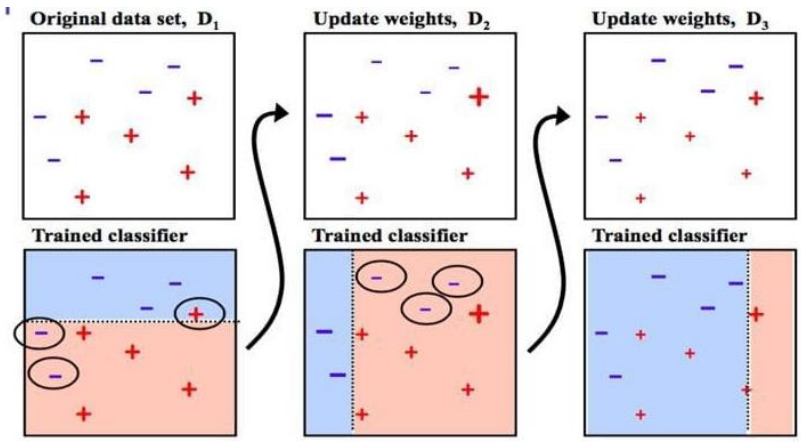
2.7. Adaptive Boosting (AdaBoost)

AdaBoost adalah metode pembelajaran mesin yang digunakan dalam data mining untuk klasifikasi data, awalnya dikembangkan oleh Yoav Freund dan Robert Schapire pada tahun 1995. Metode ini merupakan bagian dari Ensamble Learning, yang menggabungkan beberapa algoritma classifier untuk menghasilkan prediksi yang lebih akurat dibandingkan dengan menggunakan satu algoritma saja. Dalam penelitian ini, Adaboost dikombinasikan dengan algoritma Decision Tree sebagai base estimator.

Decision Tree adalah model machine learning yang digunakan untuk klasifikasi, tetapi memiliki kelemahan seperti sensitivitas terhadap perubahan kecil dalam data dan masalah overfitting. Oleh karena itu, Adaboost digunakan untuk mengatasi masalah ini dengan memberikan bobot pada tingkat kesalahan klasifikasi, yang memungkinkan untuk mengubah distribusi data sehingga meningkatkan akurasi dan mengurangi kesalahan klasifikasi. Adaboost bekerja dengan memberikan bobot lebih pada data yang salah diklasifikasikan oleh model sebelumnya, sehingga pada setiap iterasi, model baru difokuskan pada data yang lebih sulit untuk diklasifikasikan.

Algoritma ini menggabungkan model-model lemah (*Weak Classifiers*) menjadi satu model kuat (*Strong Classifier*) dengan cara menghitung rata-rata tertimbang dari prediksi semua model lemah. Untuk meminimalkan kesalahan dan meningkatkan

akurasi, Adaboost menggunakan fungsi eksponensial sebagai loss function. Dengan menggabungkan banyak model lemah, Adaboost dapat menghasilkan prediksi yang lebih akurat dan robust terhadap overfitting. Algoritma ini didukung oleh teori matematika yang kuat, menjadikannya pilihan yang dapat diandalkan dalam berbagai aplikasi. Secara sederhana, Adaboost memperbaiki kelemahan model Decision Tree dengan berfokus pada data yang sulit diklasifikasikan dan menggabungkan kekuatan dari beberapa model untuk membuat prediksi yang lebih kuat dan akurat.



Gambar 2.5 Ilustrasi Adaboost

Kelebihan dari penerapan metode Adaboost termasuk sederhana, karena untuk membangun stump hanya menggunakan satu kali splitting. Bobot yang berbeda diberikan pada prediksi akhir dari dua pohon yang dibangun. J ika error yang dihasilkan stumps cukup besar, dampaknya pada pengambilan keputusan menjadi lebih kecil.(Trivusi, 2023)

Cara kerja Adaptive Boosting adalah sebagai berikut :

1. Input : Kumpulan training sample dengan label $\{(x_1,y_1), ...(x_N,y_N)\}$; component learn algoritma, dengan jumlah perputaran T.
2. Initialize : Bobot sample training w_i^1= 1/ N, untuk semua nilai i =1,….,N
3. For t = 1, …., T ;
4. Menggunakan component learn algortima untuk melatih komponen klasifikasi, , pada sample bobot training

5. Hitung error training pada

$$h_t : \varepsilon_t = \sum_{i=1}^{N} w_i^t, y_i \neq h_t(x_i)$$

6. Tetapkan bobot untuk component classifier

$$h_t = \alpha_t = \frac{1}{2} In \left(\frac{1-\varepsilon t}{\varepsilon t}\right)$$

7. Update sample training dengan

$$w_i^{t+1} = \frac{w_i^t \exp\{-a_t y_i h_t(x_i)\}}{Ct}, i = 1, \ldots, N \; C_t$$ ialah konstanta normalisasi

8. Output : $F(x) = sign\left(\sum_{t=1}^{t} a_t, h_t(x)\right)$

Dengan Keterangan :

α_t = Tingkat pembelajaran (Learning Rate)

$h_t(x)$ = Klasifikasi dasar atau lemah

$F(x)$ = Hasil akhir berupa klasfikasi yang kuat

2.8. Confusion Matrix

Confusion Matrix adalah alat penting dalam mengukur performa model klasifikasi, yang berfungsi untuk menganalisis kemampuan classifier dalam mengidentifikasi tuple dari berbagai kelas secara akurat (Han dan Kamber, 2021). Confusion Matrix tidak hanya memberikan pemahaman yang lebih mendalam tentang sejauh mana model berhasil atau gagal dalam memprediksi kelas target, tetapi juga memungkinkan perhitungan berbagai metrik performa seperti Accuracy, Precision, Recall, dan F1-score. Dalam perhitungan nilai Precision, Recall, F1-Score dan Accurascy dapat menggunakan persamaan 2.1, 2.2, 2.3,2.4 yang di uraikan berikut.

$$Precision = \frac{TP}{TP+FP} \; x \; 100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

$$Recall = \frac{TP}{TP+FN} \; x \; 100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

$$F-1 \; Score = \frac{2 \; x \; Precision \; x \; recall}{Precision \; Recall} \; x \; 100\% \ldots (2.3)$$

$$Accuracy = \frac{TP}{Total \; Data} \; x \; 100\% \ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

2.9. Cross Validation

Cross validation adalah metode statistik yang digunakan dalam machine learning dan pemodelan prediktif untuk menilai kinerja dan kemampuan generalisasi model. Menurut geeksforgeeks.org, teknik ini membagi data yang tersedia menjadi beberapa subset atau fold, memungkinkan model dilatih dan diuji berkali-kali. Cross validation memberikan estimasi performa model yang lebih akurat pada data yang belum pernah dilihat sebelumnya dan membantu mengatasi masalah overfitting, yaitu ketika model terlalu spesifik pada data pelatihan dan berkinerja buruk pada data baru.

Dalam praktiknya, cross validation bekerja dengan membagi data menjadi beberapa bagian yang sama besar. Metode yang paling umum adalah k-fold cross validation, di mana data dibagi menjadi k bagian. Model dilatih pada k-1 bagian dan diuji pada bagian yang tersisa, dan proses ini diulang sebanyak k kali, dengan bagian yang berbeda digunakan untuk pengujian setiap kali. Hasil dari setiap pengujian kemudian dirata-rata untuk memberikan estimasi kinerja model yang lebih andal.

Keunggulan utama cross validation adalah kemampuannya memberikan gambaran yang lebih komprehensif tentang kinerja model, terutama saat data yang tersedia terbatas. Setiap bagian data digunakan untuk pelatihan dan pengujian, memungkinkan penyesuaian parameter model secara lebih sistematis dan objektif, serta mengurangi risiko optimasi berlebihan pada satu set data tertentu. Selain itu, cross validation membantu dalam memilih model terbaik untuk dataset tertentu dengan membandingkan kinerja berbagai model atau konfigurasi parameter.

Secara keseluruhan, cross validation adalah alat yang sangat penting dalam pengembangan model machine learning yang andal dan mampu melakukan generalisasi dengan baik, memastikan model tidak hanya bekerja baik pada data pelatihan tetapi juga pada data baru yang belum pernah dilihat sebelumnya.

2.10. Penelitian Terdahulu

Penelitian terdahulu yang yang menjadi refrensi dalam menemukan inspirasi yang baru sudah dilakukan oleh beberapa peneliti dengan menerapkan metode – metode yang berbeda. Berikut adalah penelitian yang dilakukan oleh Candra Surya Sriyanto dan budi setiawan yang melalukan peneitain yang bertopik “Pendeteksian Beria Hoaks Menggunakan Naïve Bayes Multinomial pada Twitter dengan fiitur pembobotan TF-IDF ”penulis melihat bahwa penyebaran banyak penyebaran informasi yang tidak

sesuai fakta, banyak yag tidak sesuai dengan kebenarannya makan peneliti membangun sistem untuk mendeteksi Hoaks pada media sosial menggunakan metode Multinomial Naïve Bayes dengan pembobotan TF-IDF, dengan menggunakna metode-metode tersebut peneliti berhasil mendapat akurasi tertinggi sebesar 76% (Sriyano & Setiawan, 2021)

"Klasifikasi berita hoaks Pilpres Menggunakan Metode Modified K-Nearest Neighbour dan Pembobotan TF-IDF" yang dilakuka oleh Faizal Nur Rozi dan Dwi Harini Sulistywati melakukan klasifikasi berita menjadi berita hoaks dan non hoaks, peneliti juga menggunakan data yang diperoleh dari komingo untuk berita Hoaks dan media sosial lainnya untuk berita yang non Hoaks dan dari metode yang digunakan, peneliti mendapat tingkat akurasi sebesar 92.31(Nur Rozi & Harini Sulistyawati, 2019).

Penelitian yang dilakukan oleh Bening Herwijayanti, Dian Eka Ratnawati, dan Lailil Muflikhah berfokus pada klasifikasi berita online dengan menggunakan metode pembobotan TF-IDF dan Cosine Similarity. Mereka mengamati bahwa sering kali terdapat banyak berita yang diposting di media sosial yang tidak sesuai dengan kategorinya, di mana kesalahan manusia (*human error*) menjadi penyebab utama berita tersebut salah penempatan. Oleh karena itu, mereka menerapkan klasifikasi berita online dengan memanfaatkan TF-IDF dan *Cosine Similarity* yang melalui tahap-tahap pengolahan seperti tokenisasi, stopword, dan stemming untuk mengurangi jumlah term dan mempercepat proses perhitungan pembobotan term menggunakan TF-IDF, serta mempercepat proses Cosine Similarity. Tujuan utama dari penelitian ini adalah untuk mengurangi kesalahan dalam pengkategorian berita dan meminimalkan human error. Hasil dari penelitian ini menunjukkan akurasi sebesar 91.25% dalam klasifikasi berita (Herwijayanti et al., 2018).

Berikut adalah penelitian yang dilakukan oleh Risa Wati yang mengangkat topik "Penerapan Algoritma Naïve Bayes dan Particle Swarm Optimization untuk Klasifikasi Berita Hoaks pada Media Sosial." Dalam penelitian ini, Risa Wati menemukan bahwa media sosial dapat memberikan dampak negatif akibat penyalahgunaan oleh pengguna, salah satunya adalah penyebaran berita hoaks. Penyebaran berita hoaks ini dilakukan oleh individu yang tidak bertanggung jawab dan dapat memicu perselisihan serta masalah lainnya. Oleh karena itu, penelitian ini fokus pada klasifikasi berita hoaks dan non-hoaks di media sosial. Metode yang

digunakan dalam penelitian ini adalah algoritma *Naïve Bayes*, yang menghasilkan tingkat akurasi sebesar 74,67%. Selain itu, ketika algoritma *Naïve Bayes* digabungkan dengan *Particle Swarm Optimization*, akurasi meningkat menjadi 85,19%. Tujuan dari penelitian ini adalah untuk mengevaluasi pengaruh Particle Swarm Optimization dalam meningkatkan akurasi klasifikasi yang sebelumnya dilakukan menggunakan Naïve Bayes. Hasilnya menunjukkan bahwa penggunaan Particle Swarm Optimization meningkatkan akurasi klasifikasi sebesar 10,52%.(Wati, 2020)

Penelitian selanjutnya dilakukan oleh Hanny Nurrohmah pada penelitiannya mengenai "Klasifikasi Berita Hoaks Berbahsa Indonesia Menggunakan Bidirectional *Long Short Term Memory* (Bi-LSTM)" tujuan dari penelitian yang dibuat adalah untuk mencari model klasifikasi yang optimal dengan menggunakan metode Bidirectional Long Short Term Memory dengan jumlah data yang digunakan sebanyak 8.744. Dari pengujian tersebut didapatkan performa model terbaik dari parameter neuron sebanyak 16 dan jumlah batch size sebanyak 512 dan untuk menangani ketidakseimbangan data dilakukan resampling data dengan menggunakan SMOTE, Random Undesampling dan Random Oversampling dan dari model dengan performa terbaik adalah model tanpa menggunakan resampling dengan akurasi 93,65% dan Recall sebesar 89% dan nilai AUC yang sudah cukup baik sebesar 0.97(Nurrohmah H, 2022).

Selanjutnya adalah penelitian dari Roy Arjun, et.al dari Cornell University Dimana penulis melakukan penelitian mengenai "A Deep Ensamble Framework For Fabricated content Detection and Classification" yang mana penulis bertujuan untuk melakukan identifikasi berita Hoaks dan melakukan klasifikasi kedalam kategori yang terperinci sebelumnya, dan penulis menggunakan metode Convolutional Neural Network (CNN) dan BiDirectional Long Short Therm Memory (Bi-LSTM) lalu representasi yang didapat dari kedua model ini dimasukan kedalam Multi-layer Perceptron Model untuk klasifikasi akhir dan memperoleh hasil akurasi sebesar 44.87% mengungguli nilai akurasi dari yang ada saat ini(Roy et al., 2018).

Berikutnya adalah penelitian yang dilakukan oleh Qureshi Ahmed Khabib.et al (2021) melakukan penelitian dengan judul "Complex Network and Source Inspired COVID-19" pada penlitain ini penulis melakukan klasifikasi berita Hoaks dengan sumber jaringan dan sumber profil untuk mengklasifikasikan berita Hoaks tentang Covid-19 di twitter. Penelitian ini di latar belakangi karena banyaknya penyebaran

hoaks dimedia sosial seperti facebook, youtube, dan twitter serta platform lain, maka dari itu penulis melakukan klaisfikasi menggabungkan beberapa metode seperti Catboost, Light GBM, Adaboost, Gradient Boosting, Random Forest dan Decision Tree lalu menggabungkan hasil dari beberapa metode tersebut dan mendapatkan Tingkat akurasi masing-masing sebesar 98,4%, 98,1% dan 98%(Qureshi et al., 2021),(Roy et al., 2018).

Penelitian yang dilakukan oleh Muhammad Airell Rizqiawan mengenai Klasifikasi Polusi Udara dengan Adaboost berbasis Pembelajaran Mesin. Dalam 20 penelitian tersebut, penulis memiliki tujuan untuk mengkategorikan polusi udara, termasuk polutan seperti Konsentrat Partikulat, Karbon Monoksida, dan Sulfur Monoksida. Harapannya adalah dapat memantau dan mengklasifikasikan polusi udara di kota-kota besar, dan hasil klasifikasinya, yang diuji menggunakan program dengan model machine learning, mencapai tingkat akurasi sekitar 90%(Airell Rizqiawan et al., 2023).

Selanjutnya adalah penelitian yang dilakukan oleh Lidia Pebrianti Penelitian yang dilakukan berfokus pada penggunaan Adaboost untuk meningkatkan akurasi algoritma Naïve Bayes dalam pengklasifikasian data diabetes. Mereka menemukan bahwa diabetes melitus adalah gangguan metabolik yang disebabkan oleh ketidakmampuan tubuh memanfaatkan insulin, menyebabkan kadar gula darah tidak terkontrol. Menurut proyeksi IDF, jumlah penderita diabetes di Indonesia bisa mencapai 28,57 juta pada tahun 2045.Para peneliti menggunakan algoritma Naïve Bayes dengan pembagian data 60/40, menghasilkan akurasi 76%. Setelah menerapkan Adaboost, akurasi meningkat menjadi 0,77%. Penelitian ini menunjukkan bahwa penggunaan Adaboost dapat memperbaiki kinerja Naïve Bayes dalam memproses data besar dan multi-kelas, sehingga lebih efektif dalam prediksi kondisi kesehatan.

Penelitian ini dilakukan oleh Abdulrahman Awf, Bayakara melakukan klasifikasi berita palsu dengan judul “Fake News Detection Using Machine Learning and Deep Learning Algorithms” peneliti melakukkan klasifikasi dengan beberapa metode seperti TF-IDF, Count Vector, Character Vector dan N-Gram level Vector dan menghasilkan akurasi sebesar 81% (Abdulrahman & Baykara, 2020).

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Metode	Judul	Keterangan
1.	Sriyano C, *et.al* (2021)	Naïve Bayes Multinomal	Pendeteksian Berita Hoaks Menggunakan Naive Bayes Multinomial Pada Twitter Dengan Fitur Pembobotan Tf-Idf	Peneliti memperhatikan penyebaran banyak informasi yang tidak sesuai fakta di media sosial. Oleh karena itu, penelitian ini mengembangkan sistem deteksi Hoaks menggunakan metode Multinomial Naïve Bayes dengan pembobotan TF-IDF. Dengan penerapan metode ini, peneliti berhasil mencapai akurasi tertinggi sebesar 76%.
2.	Nur Rozi, F, *et.al* (2019)	K-Nearest Network	Klasifikasi Berita Hoaks Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf	klasifikasi berita menjadi berita Hoaks dan non Hoaks, peneliti juga menggunakan data yang

				diperoleh dari kominfo untuk berita Hoaks dan media sosial lainnya untuk berita yang non Hoaks dan dari metode yang digunakan, peneliti mmendapat tingkat akurasi sebesar 92.31%.
3.	Herwijayanti,B,*et.al* (2018)	TF-IDF & Cosine Similarity	Klasifikasi Berita Online Dengan Menggunakan Pembobotan Tf-Idf Dan Cosine Similarity	Penelitian mengatasi ketidaksesuaian kategori data di media sosial dengan menerapkan klasifikasi berita menggunakan metode TF-IDF dan Cosine Similarity. Fokus pada mengurangi human error, tahapan pemrosesan termasuk tokenisasi, penghilangan stopword, dan

				stemming. Tujuan utama adalah mempercepat perhitungan dan memfasilitasi tugas manusia. Hasilnya, penelitian mencapai akurasi sebesar 91.25%.
4.	Wati, R , *et.al(2020)*	Naïve Bayes & Particle Swarm Optimization	Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoaks Pada Media Sosial	Penelitian menunjukkan bahwa penggunaan media sosial dapat berdampak negatif, terutama melalui penyebaran berita Hoaks yang memicu konflik. Peneliti melakukan klasifikasi berita Hoaks dan nonHoaks menggunakan metode naïve bayes, dengan tingkat akurasi 74.67%. Saat menggunakan Naïve Bayes

				berbasis Particle Swarm Optimization, tingkat akurasi meningkat menjadi 85.19%, menunjukkan peningkatan sebesar 10.52% dan membuktikan dampak positif dari penggunaan Particle Swarm Optimization pada klasifikasi berita Hoaks.
5.	Nurrohmah, H, *et.al*(2022)	Bidirectional Long Short Term Memory (Bi-LSTM)	Klasifikasi Berita Hoaks Berbahasa Indonesia Menggunakan Bidirectiona Long Short Term Memory (Bi-Lstm)	Penelitian ini bertujuan mencari model klasifikasi optimal dengan memanfaatkan metode Bidirectional Long Short Term Memory (BiLSTM) pada dataset berjumlah 8.744 data. Dari pengujian, ditemukan bahwa model mencapai performa terbaik dengan parameter

				16 neuron dan batch size sebanyak 512. Untuk mengatasi ketidakseimbangan data, dilakukan resampling menggunakan metode SMOTE, Random Undersampling, dan Random Oversampling. Model terbaik, yang tidak menggunakan resampling, berhasil mencapai akurasi sebesar 93,65%, Recall 89%, dan nilai AUC yang cukup baik yaitu 0,97.
6.	Roy, A, *et al(2018)*	Convolutional Neural Network (CNN) dan Bi-Directional Long Short Therm Memory (Bi-LSTM)	A Deep Ensemble Framework for Fabricated content Detection and Classification	Penulis bertujuan untuk mengidentifikasi dan mengklasifikasikan berita hoaks dengan menggunakan metode

				Convolutional Neural Network (CNN) dan Bi-Directional Long Short-Term Memory (Bi-LSTM). Setelah itu, representasi hasil dari kedua model tersebut diintegrasikan ke dalam model Multi-layer Perceptron untuk melakukan klasifikasi akhir. Hasil penelitian menunjukkan tingkat akurasi sebesar 44.87%, yang mengungguli nilai akurasi yang saat ini tersedia.
7.	Qureshi K, *et al* (2021)	Catboost, Light GBM, Adaboost , Gradientboost, EVC, Decision Tree Naïve Bayes	Complex Network and Source Inspired COVID-19 Fabricated content Classification on Twitter	Penelitian Qureshi Ahmed Khabib et al. (2021) berjudul "Complex Network and Source Inspired COVID-19" fokus pada klasifikasi

				berita Hoaks Covid-19 di Twitter. Dengan menggabungkan metode Catboost, Light GBM, Adaboost, Gradient Boosting, Random Forest, dan Decision Tree, penelitian ini mencapai tingkat akurasi 98,4% pada Catboost, 98,1% Light GBM, 97.92% untuk Gradientboosting, EVC sebesar 80% dan 98% RNN, dan untuk Adaboost 90%.
8.	Abdulrahman, A , *et.al(2020)*	N-Gram Level vector	Fake News Detection Using Machine Learning And Deep Learning Algorithms	Penelitian ini dilakukan oleh Abdulrahman Awf, Bayakara melakukan klasifikasi berita palsu dengan judul “Fake News Detection Using Machine Learning

				and Deep Learning Algorithms" peneliti melakukkan klasifikasi dengan beberapa metode seperti TF-IDF, Count Vector , Character Vector dan N-Gram level Vector dan menghasilkan akurasi sebesar 81%.
9.	Airell Rizqiawan , M, *et.al* (2023)	AdaBoost	Klasifikasi Polusi Udara Menggunakan Adaboost Berbasis Pembelajaran Mesin	Penelitian ini bertujuan mengkategorikan polusi udara, termasuk Konsentrat Partikulat, Karbon Monoksida, dan Sulfur Monoksida. Dengan menggunakan program dan model machine learning, penelitian ini berharap mencapai tingkat

				akurasi sekitar 90% dalam mengklasifikasikan polusi udara di kota-kota besar.
10.	Pebrianti, L,et.al (2022)	Adaboost	ImplementasiMetode Adaboostuntuk Mengoptimasi KlasifikasiPenyakit Diabetes dengan Algoritma Naïve Bayes	Penelitian ini menggunakan Adaboost untuk meningkatkan akurasi Naïve Bayes dalam pengolahan data diabetes. Temuan menunjukkan bahwa dengan menerapkan Adaboost, akurasi meningkat dari 0,7608 menjadi 0,7694. Ini menunjukkan bahwa Adaboost dapat memperbaiki kinerja Naïve Bayes, menjadikannya lebih efektif dalam klasifikasi data besar dan multi-kelas. .

2.11. Perbedaan Penelitian

Beberapa aspek yang membedakan penelitian ini dari penelitian-penelitian sebelumnya adalah metodologi dan pendekatan yang digunakan dalam proses identifikasi dan klasifikasi hoax. Penelitian ini dimulai dengan langkah identifikasi untuk membedakan antara berita hoax dan non-hoax. Setelah proses identifikasi tersebut selesai, dilanjutkan dengan klasifikasi jenis hoax pada data yang telah diidentifikasi sebagai hoax. Adapun tiga jenis klasifikasi hoax yang digunakan dalam penelitian ini adalah Fabricated (dibuat-buat), Misleading (menyesatkan), dan Manipulated (dimanipulasi).

Metode yang diterapkan dalam penelitian ini adalah Adaptive Boosting, dengan menggunakan decision tree sebagai base estimator yang memiliki maksimal kedalaman (max_depth) 2. Parameter yang digunakan juga cukup spesifik, yaitu jumlah n_estimator sebanyak 100 untuk prosess identifikasi hoax dan non-hoax, serta 1000 untuk klasifikasi jenis hoax.

Selain itu, penelitian ini juga menerapkan KBestSelector untuk menentukan fitur-fitur terbaik yang akan digunakan dalam pemodelan topik. Fitur-fitur ini dianggap penting untuk meningkatkan akurasi model dalam mengklasifikasikan jenis hoax.

Sumber data yang digunakan dalam penelitian ini diperoleh dari aplikasi X melalui proses *Crawling* menggunakan bahasa pemrograman Python yang dijalankan pada platform *Google Colab*. Metode *Crawling* ini memungkinkan peneliti untuk mengumpulkan data dalam jumlah besar secara efisien, yang kemudian diolah dan dianalisis lebih lanjut sesuai dengan tujuan penelitian.

Pendekatan yang komprehensif dan penggunaan teknik serta parameter yang spesifik ini menunjukkan inovasi dan perbedaan signifikan dibandingkan dengan penelitian-penelitian sebelumnya dalam bidang yang sama.

BAB 3
METODOLOGI PENELITIAN

3.1. Dataset

Data yang digunakan pada penelitian ini merupakan data tweet tentang pemilihan presiden dan wakil presiden yang diambil dari platform X dengan cara *crawling data*. Data yang digunakan sebanyak 954 data tweet yang dibagi menjadi 80% data training dan 20% data testing. Sehingga pembagian data berjumlah 763 untuk data trainig dan 191 untuk data testing. Data tweet yang crawling dan dipilih dimasukan kedalam ekstensi .xlsx.

Tabel 3.1 Dataset

No	Tweet	Label	Klasifikasi
1.	Mahfud menjelaskan secara hukum tertulis Gibran sudah sah menjadi Cawapres. Menurutnya hukuman sosial dan hukuman moral bisa dilakukan oleh masyarakat. Itu adalah cibiran masyarakat anak haram konstitusi yang tdk akan pernah terhapuskan selama hidupnya," ujar Mahfud https://t.co/9Bcg2Jrbnk"	non-hoax	non-hoax
2.	Dari seluruh Capres-Cawapres, cuma Prabowo-Gibran yang melalui proses pelanggaran berat etik MK. Satu produk gagal reformasi, satu lagi anak haram konstitusi. 😜 https://t.co/7g2rnhcK9w	non-hoax	non-hoax
3.	PDIP Desak Pimpinan DPR Setujui Hak Angket untuk Usut Dugaan Pemilu Curang https://t.co/7QuqwJDwbz	non-hoax	non-hoax

4.	Kondisi indonesia sudah hancur Jadi harus ambil keputusan YANG TERBURUK UNTUK HASIL YANG TERBAIK Dan yang Terbaik saat ini JOKOWI END baik itu digulingkan atau mundur sendiri.. Karna sumber dari segala macam masalah diindonesia ada di dia ... #GoodByeJokowi	hoax	fabricated
5.	Demo Pemilu Curang Besok, Mantan Danjen Kopassus Soenarko: Campur Tangan Jokowi Paling Menonjol	hoax	misleading
6.	@Muhammad_Saewad @gunung3tilu Gibran saat ditanya terkait Putusan MK (yng Inkonstitusional) menjawab : Biar Rakyat yng Menilai, padahal dia tahu kl itu memang dibuat-buat Salah tp belagak Pilon. Dasar anak Haram Konstitusi!	hoax	manipulated
7.	KPU memihak kepada anak haram mengejutkan konstitusi. Rakyat harus siap revolusi #GibranTiruCaraLicikBabenya #GibranTiruCaraLicikBabenya	hoax	manipulated
8.	Pernyataan Tim Hukum 03: Pemilu 2024 pemilu curang paling brutal, pemilu Desain bin Najis https://t.co/m3zguklpb6	hoax	misleading
9.	JOKOWI SUMBER SEGALA MASALAH BANGSA DIA HARUS DITUMBANGKAN.!!!! Selanjutnya simak di https://t.co/CxV7aKwkdU Sebarkan Jika Kalian Menjunjung Tinggi NilaiÂ² Kejujuran dan Keadilan. Jangan lupa Subscribe untuk membangun chanel Perjuangan Ini. https://t.co/dChdfUNNWc	hoax	fabricated

10.	Sumatera Selatan, Kota Palembang, Kecamatan Sematang Borang, Desa Karyamulya, TPS 013: C1: Anies: 126. Prabowo: 100. Ganjar : 16. Sirekap: Anies: 0. Prabowo: 400. Ganjar: 0. Gantung mati ketua KPU curang! #GantungMatiKetuaKPU #GantungMatiKetuaKPU https://t.co/rJl0SFLAD3	hoax	manipulated

Dalam dataset yang digunakan, klasifikasi seperti fabricated, misleading, dan manipulated memiliki ciri khas berupa kata-kata atau frasa yang sering muncul. Kata-kata ini membantu membedakan antara ketiga kategori tersebut. Berikut adalah beberapa contoh dari kata-kata yang umum ditemukan di masing-masing klasifikasi.

Tabel 3.2. Kata Kata Pada Fabricated

No	Kata-kata / Kalimat
1	"Ingat Akhirat Woi !!"
2	"Horeeee Indonesia mendunia..."
3	"Gila... 👺 "
4	"SUMBER DARI SEGALA SUMBER MASALAH"
5	"Pemilu curang akhirnya mendunia"
6	"Jokowi Sumber Masalah Bangsa Ini"
7	"KPU Ogah Buka-Bukaan soal Dana Kampanye"
8	“indikasikecurangan"
9	"Nepotisme dipertontonkan terang terangan"
10	"ribuan KTP tercecer"

Tabel 3.2 menunjukkan kata-kata yang sering kali digunakan dalam konten yang dibuat-buat. Kata-kata ini cenderung bersifat sensasional, menggunakan klaim ekstrim, dan sering kali tidak didasarkan pada fakta yang terverifikasi. Mereka dirancang untuk menarik perhatian pembaca dengan cara yang emosional dan

provokatif, seringkali tanpa mempertimbangkan keakuratan atau integritas informasi yang disampaikan.

Tabel 3.3. Kata Kata Pada Misleading

No	Kata-kata / Kalimat
1.	"pemilu curang paling brutal"
2.	"pemilu Desain bin Najis"
3.	"TOLAK PEMILU CURANG, Tangkap Jokowi...!"
4.	"Pemilu 2024 terburuk dalam sejarah Indonesia"
5.	“Seruan aksi mahasiswa bersama rakyat”
6.	“Bangkai Tikus dilemparkan ke @KPU_ID”
7.	“Dugaan Pemilu Curang. Golkar dan Gerindra Tolak Hak Angket.”
8.	“puncak kezaliman rezim Jokowi”
9.	“Saatnya rakyat unjuk kekuatan. Tolak Pemilu curang. Cabut mandat Jokowi.”
10.	“"komandoi pemilu curang"”

Tabel 3.3. menunjukan beberapa kata yang sering muncul pada miselading content, kalimat kaliamt yang memberikan pemahaman sendiri tanpa melakukan research terlebih dahulu untuk mengetaui kebenarannya.

Tabel 3.4 Kata Kata Pada Manipulated

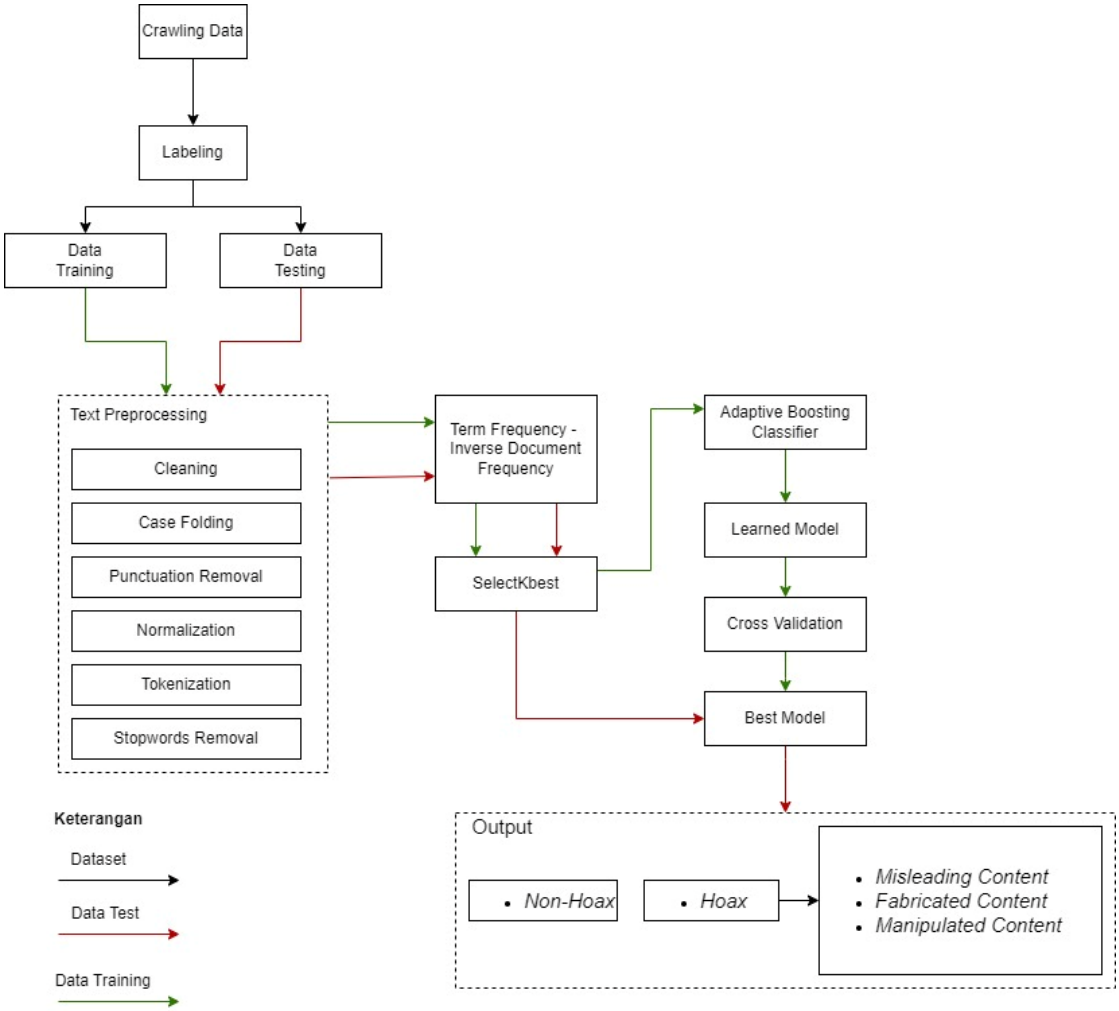
No	Kata-kata / Kalimat
1.	“menggunakan BANSOS untuk dukung ANAK HARAM KONSTITUSI”
2.	"Kebenaran terungkap akhirnya!!!"
3.	"Ekonomi akan runtuh jika Y terpilih!"
4.	"Semua media berbohong!"
5.	"Ada yang mengatakan bahwa kebocoran data ini disengaja"
6.	"Dipercaya bahwa ada peningkatan besar dalam kriminalitas"
7.	“@jokowi malah memproduksi ANAK HARAM KONSTITUSI”
8.	“Udah mulai menunjukkan kekonyolan dan ketidakberadaban”
9.	“bapak penguasa melahirkan anak haram konstitusi lewat Paman MK”
10.	“Sdh ada tanda2 Nasdem mau menolak hak angket”

Pada Tabel 3.4. menunjukan beberapa kalimat atau kata yang sering muncul dan menjadi acuan mengapa kalimat tersebut dikatakan menjadi manipulated content karena melakukan perubahan atau memanipulasi terhadap konten sebenarnya.

3.2. Arsitektur Umum

Pada penelitian ini, terdapat beberapa tahapan yang dilakukan. Adapaun tahapan tahapan tersebut adalah mengumpulkan data, dimana data yang diperoleh merupakan data yang diambil dari aplikasi X dengan proses crawling dan disimpan dalam format .xlsx . setelah data melewati tahapan preprocessing yang terdiri dari cleaning , case folding, punctual removal dan stopword removal, dan tokenisasi. Setelah melewati tahapan preprocessing selanjutnya akan masuk ketahapan TF-IDF dimana pada tahap TD-IDF ini akan memberikan bobot kepada teks agar bisa di proses oleh komputer.

pada penelitian ini penulis menggunakan fitur N-gram yaitu bigram dan unigram dan normlallisasi l2 untuk memastikan semua panjang vektor TF-IDF sama untuk semua dokumen. Setelah memberikan pembobotan dengan TF-IDF penulis melakukan seleksi fitur menggunakan alat lain yaitu SelectKBest untuk memilih fitur bigram yang terbaik, pada penelitian kali ini penulis memilih 10.000 fitur kata dengan menggunakan metode chi2 yang bisa membantu memilih fitur-fitur yang relevan denan tabel yang ingin di prediksi. Lalu melalui *Adaptive Boosting* berguna sebagai input untuk model training , pada penelitian kali penulis menggunakan *Adaptive Boosting* sebagai model nya dan menggunakan *Decision Tree* sebagai base estimator dimana hasil dari *Decision Tree* akan digunakan pada model *Adaboost* untuk meningkatkan model model yang lemah. Hasil dari model akan melewati proses pengujian kemampuan dengan menggunakan data testing dan pada data proses testing akan melewati proses evaluasi yaitu confusion matrix untuk memahami jenis kesalahan. Arstiektur umum pada penelitian ini digambarkan pada Gambar 3.1. sebagai berikut.



Gambar 3.1 Arsitektur Umum

3.2.1. Input Data

Pada penelitian ini, input data berisi mengenai pemilihan presiden dan wakil presiden 2024 yang diambil dari cuitan tweet pada aplikasi X dengan melakukan crawling data dalam bahasa indonesia dan mengabil data hoax dan non hoax. Data tweet akan diberikan 2 label yang pertama label dengan cara manual menjadi hoax dan non hoax, lalu label kedua yaitu menentukan klasifikasi hoax nya menjadi *fabricated*, *misleading* dan *manipulated.*

Tabel 3.5 Contoh Tweet Hoax dan Non-hoax

No	Tweet	Label	Klasifikasi
1.	Warga Medan tanpa tedeng aling-aling!! #PemiluCurang https://t.co/Y0z2H2VMFD	hoax	fabricated
2.	Bakal Calon Presiden (Bacapres) Ganjar Pranowo telah menjadi inspirasi bagi warga Yogyakarta dalam usaha untuk mempertahankan keberlanjutan lingkungan alam di Indonesia. Dengan semangat dan antusiasme yang di berikan warga sekitar, kegiatan bersih-bersih di Embung Bangeran,… https://t.co/mOb4GVCoJa	non-hoax	non-hoax
3.	KPU memperpanjang jdwl rekap pemilu. Artinya KPU SUDAH MELANGGAR UU PEMILU....Halloo KPU Takut ya tgl 19-20 mo di tongkrongin di KPU. Dan mau ada Gerakan Rakyat Tolak Pemilu Curang.. jadi di undur nih pengumuman hasil PEMILU 😳🤔ᴉᴅ atau mau di umumkan PD saat mlm lebaran ? https://t.co/QoBilwbfM3	hoax	misleading
4.	Gaya Anak Haram Konstitusi" yang tengil dan melecehkan Prof Mahfud ini semakin video viral membuka mata publik bahwa branding muda dan sopan itu hanya gimmick. #DebatCawapres.	hoax	manipulated

3.2.2. Pre-processing

Preprocessing dilakukan untuk mengklasifikasikan dan membuahkan pemrosesan data. Pada penelitian ini penulis menggunakan 5 tahap preprocessing yaitu cleaning, case folding, stop removal, normalization, dan tokenization.

3.2.1.1. Case Folding

Tahap ini adalah tahap awal yang akan mengubah seluruh isi apda teks menjadi huruf kecil atau Lower Case sebelum ketahap preprocessing selanjutnya, untuk implementasi dari Case Folding ini dapat dilihat dari pseudocode 3.1. dibawah ini

```
FUNCTION toLowercase(inputString)
   lowerString ← ""
   FOR each character IN inputString DO
      IF character IS uppercase THEN
         lowerCharacter ← convert character TO lowercase
      ELSE
         lowerCharacter ← character
      END IF
      lowerString ← lowerString + lowerCharacter
   END FOR
   RETURN lowerString
END FUNCTION
```

Pseudocode 3.1 Case Folding

case folding memainkan peran penting dalam praproses data untuk pembelajaran mesin, karena mengurangi jumlah fitur unik dalam korpus teks, sehingga meningkatkan efisiensi dan akurasi model. Proses ini juga memfasilitasi tokenisasi dan teknik pengolahan bahasa alami lainnya seperti lemmatization dan stemming. Hasil implementasi *Case Folding* dapat dilihat pada tabel 3.6.

Tabel 3.6 Implementasi Proses Case Folding

Sebelum Case Folding	Sesudah Case Folding
Sudah Resmi Pasangan capres-cawapres nomor urut 01 Anies Baswedan dan Muhaimin Iskandar jadi paslon pertama yang mendaftarkan gugatan perkara Perselisihan Hasil Pemilihan Umum (PHPU) Presiden dan Wakil Presiden tahun 2024 ke Mahkamah Konstitusi (MK). Gugatan ini telah diterima MK.…	sudah resmi pasangan capres-cawapres nomor urut 01 anies baswedan dan muhaimin iskandar jadi paslon pertama yang mendaftarkan gugatan perkara perselisihan hasil pemilihan umum (phpu) presiden dan wakil presiden tahun 2024 ke mahkamah konstitusi (mk). gugatan ini telah diterima mk.…

https://t.co/QLPCtcHEKI	https://t.co/qlpctcheki
Sumatera Selatan, Kota Palembang, Kecamatan Sematang Borang, Desa Karyamulya, TPS 013: C1: Anies: 126. Prabowo: 100. Ganjar : 16. Sirekap: Anies: 0. Prabowo: 400. Ganjar: 0. Gantung mati ketua KPU curang! #GantungMatiKetuaKPU #GantungMatiKetuaKPU https://t.co/rJl0SFLAD3	sumatera selatan, kota palembang, kecamatan sematang borang, desa karyamulya, tps 013: c1: anies: 126. prabowo: 100. ganjar : 16. sirekap: anies: 0. prabowo: 400. ganjar: 0. gantung mati ketua kpu curang! #gantungmatiketuakpu #gantungmatiketuakpu https://t.co/rjl0sflad3
Pak Anies udah bicara soal isu kemacetan sejak taun 2012, dan selama menjabat jd Gubernur berhasil melanjutkan dan mengembangkan sistem transportasi terintegrasi seperti Jaklingko. Bayangin orang dgn kapabilitas seperti belio direduksi jadi sekadar omon-omon. 😤	pak anies udah bicara soal isu kemacetan sejak taun 2012, dan selama menjabat jd gubernur berhasil melanjutkan dan mengembangkan sistem transportasi terintegrasi seperti jaklingko. bayangin orang dgn kapabilitas seperti belio direduksi jadi sekadar omon-omon. 😤
Prabowo Ingin Naikkan Upah Pejabat demi Cegah Korupsi, ICW: Koruptor bukan Orang Miskin https://t.co/qrpFwYLjJ1 #TempoNasional	prabowo ingin naikkan upah pejabat demi cegah korupsi, icw: koruptor bukan orang miskin https://t.co/qrpfwyljj1 #temponasional
Bersatunya Kubu 01 dan 03 Bukan Sekedar Omon-omon, JK dan Megawati Segera Bertemu! https://t.co/IJYwZI8SvW	bersatunya kubu 01 dan 03 bukan sekedar omon-omon, jk dan megawati segera bertemu! https://t.co/ijywzi8svw
Anies Baswedan ingin kembalikan kebebasan berbicara:“Di Jakarta kurang apa coba? Bkn hny kritik tp fitnah,diputar balikan. Apakah ada yg dilaporkan ke polisi? Tidak ada!” Ini baru namanya pemimpin, tidak anti kritik,klo wowo bisa masuk penjara semua yg ga pro dia! #AsalBukan02 https://t.co/H37pkosRlV	anies baswedan ingin kembalikan kebebasan berbicara:“di jakarta kurang apa coba? bkn hny kritik tp fitnah,diputar balikan. apakah ada yg dilaporkan ke polisi? tidak ada!” ini baru namanya pemimpin, tidak anti kritik,klo wowo bisa masuk penjara semua yg ga pro dia! #asalbukan02 https://t.co/h37pkosrlv

3.2.1.2. Cleaning

Tahap ini cleaning ini bertujuan untuk membersihkan teks dari elemen-elemen yang tidak relevan atau mengganggu pada proses analisis. Untuk implementasinya dapat dilihat pada pseudocode 3.2 dibawah ini.

```
FUNCTION cleanText(inputString)
   inputString ← removeURLs(inputString)
   cleanedString ← ""
   FOR each character IN inputString DO
      IF character IS alphanumeric OR character IS space THEN
         cleanedString ← cleanedString + character
      END IF
   END FOR
   cleanedString ← removeHashtags(cleanedString)
   finalString ← ""
   FOR each character IN cleanedString DO
      IF character IS NOT number THEN
         finalString ← finalString + character
      END IF
   END FOR
   finalString ← removeExtraSpaces(finalString)
    RETURN finalString
END FUNCTION

FUNCTION removeURLs(text)
   // Use a regex pattern to find and remove URLs
   regex ← "http[s]?://\S+|www\.\S+"
   text ← regexReplace(text, regex, "")
   RETURN text
END FUNCTION

FUNCTION removeHashtags(text)
   // Use a regex pattern to find and remove hashtags
   regex ← "#\S+"
   text ← regexReplace(text, regex, "")
```

```
    RETURN text
END FUNCTION

FUNCTION removeExtraSpaces(text)
    text ← regexReplace(text, "\s+", " ")
    text ← trim(text)
    RETURN text
END FUNCTION
```

Pseudocode 3.2 Clean Text

Berikut merupakan hasil dari data cleaning pada dataset yang ditampilkan pada tabel 3.7.

Tabel 3.7 Implementasi Proses Cleaning

Sebelum Cleaning	**Setelah Cleaning**
Gugat Hasil Pilpres ke MK, Tim Hukum Ganjar-Mahfud Bawa 4 Kontainer Berkas https://t.co/W3Bh64qI2L	gugat hasil pilpres ke mk tim hukum ganjar mahfud bawa 4 kontainer berkas

3.2.1.3. Tokenisasi

Pada tahap ini merupakan proses yang akan membagi teks dengan tujuan untuk memecah setiap kalimat menjadi kata yang tunggal dan disebut token. Unutk mempermudah proses dilakuakan perhitungan disetiap kata-kata, oleh karena itu maka akan dilakukan proses pemisahan pada tahap tokenisasi pada pseudocode 3.3 dibawah ini

```
FUNCTION tokenizeText(inputString)
    tokens ← []
    currentToken ← ""
    FOR each character IN inputString DO
        IF character IS alphanumeric OR character IS space THEN
            currentToken ← currentToken + character
        ELSE IF currentToken IS NOT empty THEN
```

```
        ADD currentToken TO tokens
        currentToken ← ""
      END IF
    END FOR

    IF currentToken IS NOT empty THEN
      ADD currentToken TO tokens
    END IF

    RETURN tokens
END FUNCTION
```

Pseudocode 3.3 Tokenisasi

Pseudocode diatas merupakan langkah yang dilakukan ke kode program, function tokenize(teks) dapat digunakan untuk memecah teks menjadi token. Setiap hasil token akan di masukan ke dalam daftar baru pada setiap dataframe seperti pada tabel 3.8.

Tabel 3.8 Implementasi Proses Tokenize

Sebelum Tokenize	**Setelah Tokenize**
dari pada dapet produk gagal reformasi dan anak haram konstitusi jauh lebih foto editan baik anies dgn segudang prestasi dan rekam jejak yg mumpuni	['dari', 'pada', 'dapet', 'produk', 'gagal', 'reformasi', 'dan', 'anak', 'haram', 'konstitusi', 'jauh', 'lebih', 'foto', 'editan', 'baik', 'anies', 'dgn', 'segudang', 'prestasi', 'dan', 'rekam', 'jejak', 'yg', 'mumpuni']

3.2.1.4. Stopword Removal

Pada proses ini dilalkukan dengan tujuan menghaus kata yang tidak digunakan namun sering muncul dan kata-kata yang tidak memiliki makna atau tidak berkaitan dengan topik pemilihan presiden dan wakil presiden.

```
FUNCTION removeStopwords(tokens)
    stopwords ← ["dan", "atau", "di", "itu", "dari", "yang", ...] // Daftar kata-kata stopwords
```

```
    filteredTokens ← []
    FOR each token IN tokens DO
        IF token NOT IN stopwords THEN
            ADD token TO filteredTokens
        END IF
    END FOR
    RETURN filteredTokens
END FUNCTION
```

Pseudocode 3.4 Stopwords Removal

Pada penelitian ini penulis menggunkan kamus pada sastrawi untuk melakukan stopwords. Data pada kamus tersebut diubah menjadi bentuk set efisiensi dalam pencarian, fungsi stopwords_removal digunakan untuk menghapus stopwords dari daftar kata-kata, hasil dari stopwords dapat dilihat pada tabel 3.9. dibawah ini

Tabel 3.9 Implementasi Stopwords Removal

Sebelum Stopwords	Sesudah Stopwords
dari pada dapet produk gagal reformasi dan anak haram konstitusi jauh lebih foto editan baik anies dgn segudang prestasi dan rekam jejak yg mumpuni	['dapet', 'produk', 'gagal', 'reformasi', 'anak', 'haram', 'konstitusi', 'foto', 'editan', 'anies', 'dgn', 'segudang', 'prestasi', 'rekam', 'jejak', 'yg', 'mumpuni']

3.2.3. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode untuk menilai seberapa penting sebuah kata dalam suatu dokumen dalam koleksi teks. Metode ini menghitung seberapa sering kata tersebut muncul dalam dokumen tertentu (TF), dibandingkan dengan seberapa umum kata tersebut muncul di seluruh koleksi dokumen (IDF).

Kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain diberi bobot yang lebih tinggi, menunjukkan bahwa kata-kata ini mungkin lebih signifikan dalam mewakili esensi dokumen tersebut. TF-IDF digunakan untuk membangun vektor representasi untuk setiap dokumen, membantu

dalam pengelompokan dokumen atau analisis lainnya dengan menggambarkan kemiripan antara dokumen dan kata-kata dalam korpus teks.

```
FUNCTION ComputeTFIDF(Documents):
  INPUT:
    - Documents: List of documents, where each document is a list of words
  OUTPUT:
    - TFIDF: Dictionary of TF-IDF values for each term in each document
  INITIALIZE:
    - TF ← {}  # Term Frequency dictionary
    - DF ← {}  # Document Frequency dictionary
    - N ← number of documents
  # Calculate Term Frequency (TF)
  FOR each document d in Documents DO:
    TF[d] ← {}
    numTerms ← total number of terms in d
    FOR each term t in d DO:
      IF t is not in TF[d] THEN
        TF[d][t] ← 0
      END IF
      TF[d][t] ← TF[d][t] + 1
    END FOR
    FOR each term t in TF[d] DO:
      TF[d][t] ← TF[d][t] / numTerms
    END FOR
  END FOR
  # Calculate Document Frequency (DF)
  FOR each document d in Documents DO:
    FOR each term t in TF[d] DO:
      IF t is not in DF THEN
        DF[t] ← 0
      END IF
      DF[t] ← DF[t] + 1
    END FOR
  END FOR
```

```
    # Calculate TF-IDF
    TFIDF ← {}
    FOR each document d in Documents DO:
        TFIDF[d] ← {}
        FOR each term t in TF[d] DO:
            IDF_t ← log(N / DF[t])
            TFIDF[d][t] ← TF[d][t] * IDF_t
        END FOR
    END FOR
    RETURN TFIDF
END FUNCTION
```

Pseudocode 3.5 TF-IDF Vectorizer

Dalam penelitian ini, penulis memanfaatkan fitur N-Gram sebagai salah satu teknik untuk menganalisis teks. N-Gram mengacu pada kombinasi kata-kata berurutan dalam teks, seperti Unigram (satu kata), Bigram (dua kata), dan Trigram (tiga kata). Pemilihan Bigram di sini penting karena memungkinkan model untuk menangkap konteks yang lebih lengkap dari kata-kata yang digunakan, sehingga membantu dalam memahami makna yang lebih akurat dari teks tersebut. Dengan implmenetasi pseudocode tf-idf vectorizer diatas berikut adalah hasil yang dapat dilihat dari hasil perhitugan tf-df pada tabel 3.10.

Tabel 3.10 Implementasi TF-IDF Vectorizer

Teks	Term Frequency (TF)	Inverse Document Frequency (IDF)	TF-IDF
abah anies	0.161774415	5.441343212	0.879878461
abadi rapat	0.118667114	6.5399555	0.775908701
abdi negara	0.16717099	6.945420609	1.160701517
abdullah koruptor	0.133349252	6.945420609	0.926200201

Salah satu keunggulan N-Gram, adalah kemampuannya untuk mengenali frasa atau ekspresi kata-kata yang khusus atau sering muncul dalam teks-teks tertentu, seperti dalam kasus tweet yang mengandung informasi hoax. Dengan menggunakan N-Gram, peneliti dapat mengidentifikasi pola-pola khusus ini dengan lebih baik, memungkinkan untuk deteksi dini atau analisis yang lebih mendalam terhadap teks-

teks yang mengandung informasi yang tidak benar atau menyesatkan. Penerapan fitur N-gram dapat dilihat pada tabel 3.11 dibawah ini

Tabel 3.11 Implementasi Fitur N-Gram

Label	Unigram	Bigram	Trigram
Non-Hoax	Dugaan	Dugaan Pemilu	Dugaan Pemilu Curang
	Bakal	Bakal calon	Bakal calon presiden
Hoax	viral	viral warga	viral warga solo
	Prabowo	Prabowo dulu	Prabowo dulu teriak

3.2.4. SelectKbest

SelectKBest adalah metode dari modul sklearn.feature_selection yang digunakan untuk pemilihan fitur, yaitu memilih sejumlah fitur terbaik berdasarkan skor statistik tertentu. Dalam penelitian ini, SelectKBest digunakan untuk mengurangi jumlah kata yang dipertimbangkan, sehingga dapat meningkatkan kinerja algoritma pembelajaran mesin, mengurangi overfitting, meningkatkan performa model, dan menyaring fitur yang tidak relevan.

Penulis menginisialisasi objek dengan menggunakan chi-squared sebagai fungsi skornya dan memilih 10000 fitur terbaik. Proses ini melibatkan perhitungan fungsi chi-squared antara setiap fitur dalam X_tfidf dan label target. Selanjutnya, dihitung tabel kontingensi untuk setiap fitur dan nilai chi-squared berdasarkan tabel kontingensi tersebut. Implementasi SelectKBest dapat dilihat pada pseduocode berikut

```
FUNCTION KBestSelector(X, y, k, score_function):
   INPUT:
      - X: Feature matrix (n_samples x n_features)
      - y: Target vector (n_samples)
      - k: Number of top features to select
      - score_function: Function to compute the score for each feature
   OUTPUT:
      - Selected_features: List of indices of the selected features
   INITIALIZE:
      - scores ← Array of zeros with length equal to the number of features in
X
   # Calculate scores for each feature
   FOR each feature i in X DO:
      scores[i] ← score_function(X[:, i], y)
   END FOR
   # Sort features by score in descending order
   sorted_indices ← Indices of scores sorted in descending order
   # Select top k features
   Selected_features ← sorted_indices[:k]
   RETURN Selected_features
END FUNCTION
```

Pseudocode 3.6 KBest Selector

Tujuan dari adalah memilih sejumlah fitur yang akan digunakan untuk mengurangi dimenasi data yang dapat meningkatkan efisiensi dan kecepatan dalam pelatihan model. Dapat dilihat pada tabel 3.12 merupakan hasil dari *SelectKBest* yang dipilih.

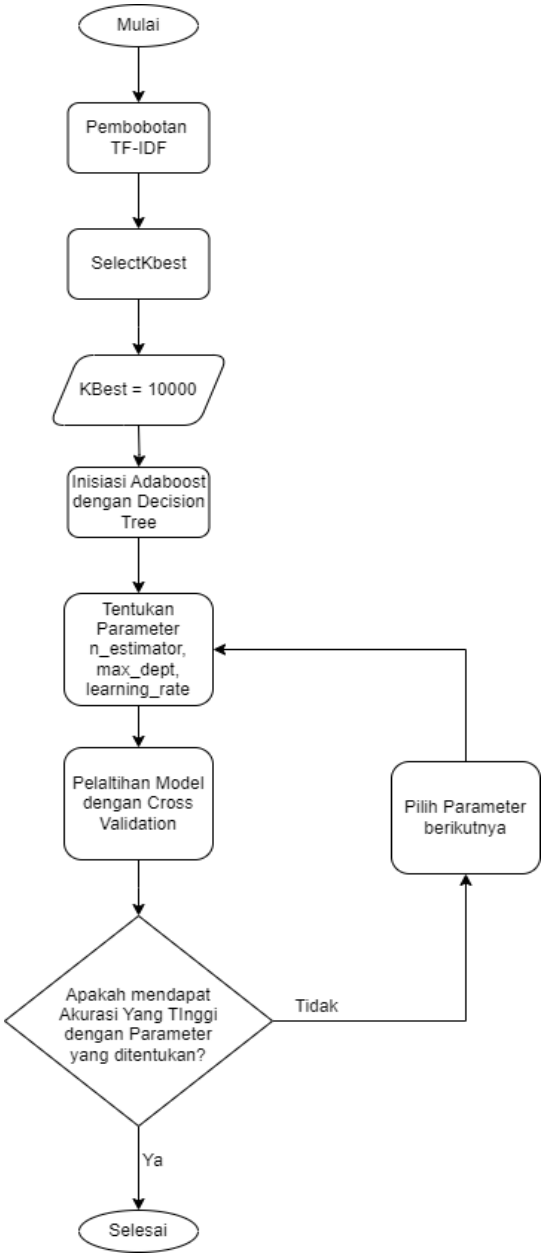
Tabel 3.12 Hasil Pemilihan Fitur SelectKBest

No	**Feature**	**Score**
1	presiden	9.893717471
2	anak haram	7.823326155
3	calon presiden	8.729956745
4	anies	9.396308566
5	muhaimin iskandar	3.463188115

3.2.5. Adaptive Boosting

Pada penelitian ini, penulis menggunakan algoritma *Adaptive Boosting* (Adaboost) untuk membangun model, dengan Decision Tree sebagai estimator dasarnya. Adaboost adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi. Algoritma ini bekerja dengan menggabungkan beberapa model prediksi sederhana yang disebut "weak learners" untuk menciptakan model prediksi yang lebih kuat. Alur kerja Adaboost dapat dilihat pada Gambar 3.2.

Adaboost memulai dengan melatih weak learner pertama pada data pelatihan, di mana setiap data pelatihan diberi bobot yang sama. Setelah pelatihan, data yang salah diklasifikasikan oleh weak learner pertama akan diberi bobot lebih besar, sementara data yang diklasifikasikan dengan benar akan diberi bobot lebih kecil. Proses ini diulangi untuk sejumlah iterasi, di mana setiap weak learner baru dilatih pada data yang telah diberi bobot ulang. Pada akhir setiap iterasi, model weak learner baru ditambahkan ke dalam kumpulan model, dan model akhir adalah kombinasi dari semua weak learners, di mana setiap model diberi bobot berdasarkan akurasi mereka. Metode ini memungkinkan Adaboost untuk fokus pada data yang sulit diklasifikasikan, memperbaiki kesalahan model sebelumnya, dan secara bertahap meningkatkan akurasi keseluruhan. Adaboost menggunakan fungsi eksponensial sebagai loss function untuk meminimalkan kesalahan dan meningkatkan akurasi model prediksi. Dengan demikian, Adaboost mengatasi kelemahan Decision Tree yang rentan terhadap overfitting dan sensitivitas terhadap perubahan kecil dalam data, menghasilkan model yang lebih robust dan akurat.



Gambar 3.2 Alur Kerja Adaboost

Gambar 3.2 menunjukkan alur kerja Adaboost, dimulai dengan persiapan data yang akan dianalisis. Dataset yang telah disiapkan akan melalui beberapa tahap preprocessing untuk mengubah data mentah menjadi data yang lebih mudah dipahami oleh sistem. Langkah berikutnya adalah pemberian bobot menggunakan tf-idf dengan tambahan fitur n-gram, yaitu bigram. Setelah itu, dilakukan pemilihan fitur menggunakan SelectKBest untuk memilih jumlah fitur yang akan digunakan agar menghindari overfitting.

Kemudian, model Adaboost dilatih dengan data yang telah diproses dan fitur-fitur yang telah dipilih. Adaboost menggabungkan beberapa weak learners dan menggunakan Decision Tree Classifier sebagai base estimator. Decision Tree ini menghasilkan weak learners yang digabungkan untuk menciptakan model yang lebih kuat. Setelah model Adaboost dilatih, model tersebut divisualisasikan dan dievaluasi menggunakan metode cross-validation untuk memastikan generalisasi yang baik.

Setelah hasil pelatihan memadai, model tersebut diuji dengan data baru yang belum pernah dilihat sebelumnya. Kinerja model kemudian dievaluasi berdasarkan metrik seperti akurasi, presisi, dan recall untuk memastikan hasil yang memuaskan. Proses ini berakhir setelah semua langkah selesai dilakukan, dan model Adaboost dengan Decision Tree sebagai base estimator siap digunakan untuk klasifikasi atau prediksi data baru.

Pada tahap ini model menggunakan decision tree dengan max_depth sebanyak 2 sebagai base esstimator, dan menggunakan parameter algorithm “SAMME” serta menggunakan n_estimar pada Adaboost. Berikut penerapan model Adaboost dapat dilihat pada Pseudocode 3.6. berikut:

```
FUNCTION AdaBoost(X, y, T):
  INPUT:
    - X: Feature matrix
    - y: Label vector
    - T: Number of iterations
  OUTPUT:
    - H: Final strong classifier
  INITIALIZE:
    - N ← number of samples in X
    - Initialize weights: w[i] ← 1 / N for i = 1 to N
  FOR t = 1 TO T DO:
    1. Train weak learner h_t using weights w
    2. Calculate error rate: err_t ← Σ(w[i] * I(h_t(X[i]) ≠ y[i])) / Σ(w)
    3. Compute weight of weak learner: alpha_t ← 0.5 * ln((1 - err_t) / err_t)
    4. Update weights:
      FOR i = 1 TO N DO:
        w[i] ← w[i] * exp(-alpha_t * y[i] * h_t(X[i]))
```

```
        END FOR
      5. Normalize weights: w ← w / Σ(w)
    END FOR
    DEFINE:
      FUNCTION FinalClassifier(X_new):
        sum_alpha_h ← 0
        FOR t = 1 TO T DO:
          sum_alpha_h ← sum_alpha_h + alpha_t * h_t(X_new)
        END FOR
        RETURN sign(sum_alpha_h)
    RETURN FinalClassifier
END FUNCTION
```

Pseudocode 3.7 Model Adaboost pada Label

```
FUNCTION AdaBoost(X, y, T):
  INPUT:
    - X: Feature matrix
    - y: Label vector
    - T: Number of iterations
  OUTPUT:
    - H: Final strong classifier
  INITIALIZE:
    - N ← number of samples in X
    - Initialize weights: w[i] ← 1 / N for i = 1 to N
  FOR t = 1 TO T DO:
    1. Train weak learner h_t using weights w
    2. Calculate error rate: err_t ← Σ(w[i] * I(h_t(X[i]) ≠ y[i])) / Σ(w)
    3. Compute weight of weak learner: alpha_t ← 0.5 * ln((1 - err_t) / err_t)
    4. Update weights:
      FOR i = 1 TO N DO:
        w[i] ← w[i] * exp(-alpha_t * y[i] * h_t(X[i]))
      END FOR
    5. Normalize weights: w ← w / Σ(w)
  END FOR
  DEFINE:
```

```
    FUNCTION FinalClassifier(X_new):
        sum_alpha_h ← 0
        FOR t = 1 TO T DO:
            sum_alpha_h ← sum_alpha_h + alpha_t * h_t(X_new)
        END FOR
        RETURN sign(sum_alpha_h)
  RETURN FinalClassifier
END FUNCTION
```

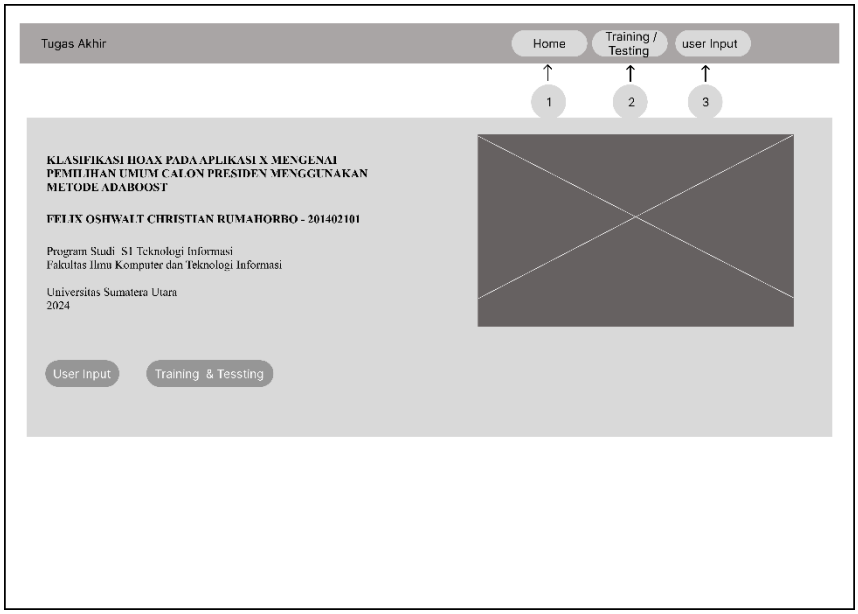
Pseudocode 3.8 Model Adaboost pada Klasifikasi

3.3. Perancangan Aplikasi Sistem

Perancangan aplikasi sistem adalah proses yang terstruktur dan sistematis untuk membangun aplikasi yang memenuhi kebutuhan pengguna. Tahapan ini meliputi penentuan persyaratan, analisis, desain, implementasi, pengujian, dan pemeliharaan. Tujuan utamanya adalah menggambarkan keseluruhan sistem dan operasinya agar sesuai dengan kebutuhan yang telah diidentifikasi.

3.3.1. Gambaran Tampulan Halaman Depan

Halaman depan ini akan menampilkan beranda. Pada halaman tersebut akan disajikan informasi dari situs web, yang mencakup judul penelitian, identitas peneliti, dan tombol menu yang berfungsi sebagai penghubung antar halaman. Brtikut merupakan gambaran tampilan halaman depan sistem web dalam gambar 3.3.



Gambar 3.3 Rancangan Tampilan Home

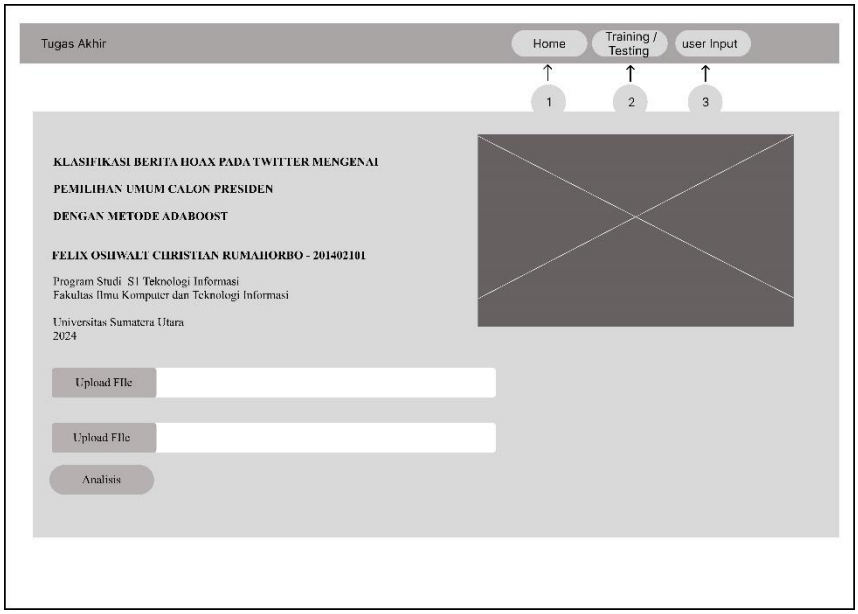
Keterangan :

1. Tombol Home (1) merupakan tombol yang mengarah ke halaman utama atau halaman beranda
2. Tombol training (2) merupakan Tombol yang mengarah ke halaman proeses data training dan testing
3. Tombol Prediksi (3) merupakan tombol yang mengarah ke halaman untuk memprediksi dari teks yang di input.

3.3.2. Rancangan Tampilan Training dan Testing

Halaman ini dirancang untuk proses pelatihan dan pengujian dataset ke dalam model. Pengguna dapat mengunggah file data pelatihan dalam format .xlsx, lalu mengklik "Upload File" untuk memulai proses pelatihan. Setelah proses pelatihan selesai, hasil data sebelum dan sesudah pelatihan akan ditampilkan. Selanjutnya, pengguna dapat menguji model dengan mengunggah file data pengujian dalam format .xlsx dan mengklik tombol "Upload File" untuk memulai proses pengujian. Setelah proses pengujian selesai, akan ditampilkan hasil data yang telah diproses dan prediksi terhadap kedua tabel.

Selain itu, akan ditampilkan matriks kebingungan (confusion matrix) yang mencakup nilai precision, recall, f1-score, dan support. Rancangan tampilan halaman untuk pelatihan dan pengujian ini dapat dilihat pada Gambar 3.4.



Gambar 3.4 Rancangan Tampilan Training dan Testing

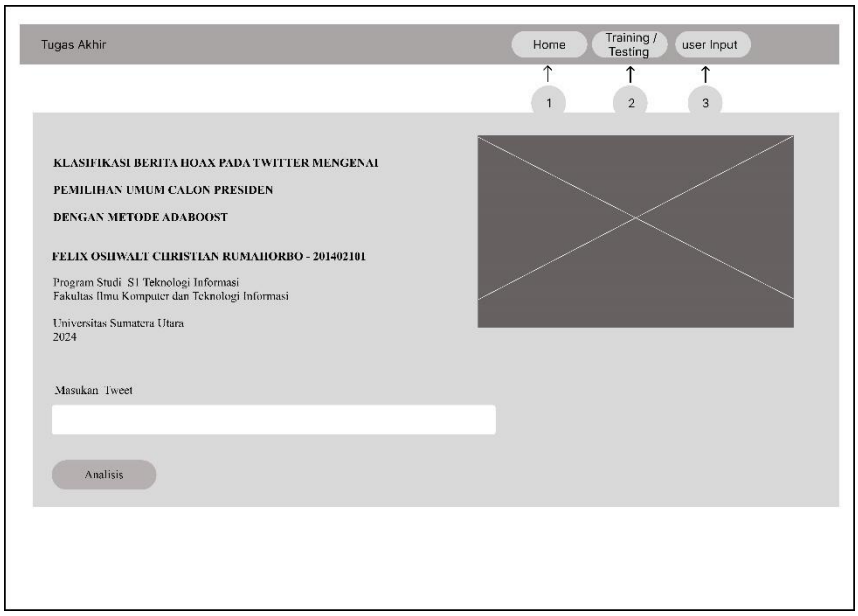
Keterangan :

1. Label A merupakan button untuk memilih data yang ingin di training
2. Label B merupakan button untuk memulai proses training
3. Tabel hasil training akan menampilkan hasil training yang berupa hebelum training dan setelah proses training
4. Label C merupakan button untuk memilih data yang ingin di testing
5. Label D merupakan button untuk memulai proses testing
6. Tabel hasil testing akan menampilkan hasil training yang berupa hebelum training dan setelah proses testing

3.3.3. Rancangan Tampilan User Input

Tampilan ini adalah halaman yang digunakan oleh user untuk melakukan proses identifikasi dan klasifikasi tweet hoax atau non-hoax tanpa input file data set uang berformat xlsx, user hanya cukup untuk menginput satu data tweet berbahasa indonesia dan akan memberikan output berupa hasil identifikasi dan klasifikasi yang

akan mengenerate tabel hasil indentifikasi dan klasifikasi. Rancangan tampilan halaman user input dapat dilihat pada gambar 3.5.



Gambar 3.5 Rancangan Tampilan User Input

3.4. Metode Evaluasi

Metode evaluasi yang digunakan dalam penelitian ini adalah Cross Validation dan Confusion Matrix. Cross Validation merupakan tahap evaluasi model yang digunakan untuk mengukur seberapa baik model pembelajaran mesin bekerja pada data yang belum pernah dilihat sebelumnya. Manfaat dari cross validation ini adalah untuk mencegah overfitting dan memberikan evaluasi yang lebih stabil. Sementara itu, Confusion Matrix digunakan dalam penelitian ini sebagai alat untuk menilai kinerja model klasifikasi. Matriks ini menunjukkan perbandingan antara prediksi yang dihasilkan oleh model dan hasil yang sebenarnya. Manfaat dari confusion matrix ini adalah memberikan rincian mengenai kinerja model dan mengukur tingkat kesalahan. Pengukuran confusion matriks dapat dilihat pada tabel 3.11 dalam perhitungan accuracy, precision, recall dan f1-score

Tabel 3.13 Perhitungan Accuracy

Classifiacation		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Keterangan :

TP : Jumlah banyaknya data yang diprediksi postif oleh model dan kelas aktual bernilai positif.

FN : Jumlah banyaknya data yang diprediksi negatif oleh model dan kelas akutualnya bernilai negatif.

FP : Jumlah banyaknnya data yang di prediksi positifi oleh model namun kelas akutualnya bernilai negatif

FN : Jumlah banyaknya data yang di prediksi negatif oleh model namun kelas aktualnya positif.

BAB 4
IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Imlementasi Sistem

Dalam merancang sistem untuk klasifikasi hoax pada aplikasi X terkait pemilihan presiden dan calon presiden menggunakan metode Adaboost, diperlukan dukungan dari perangkat keras (hardware) dan perangkat lunak (software) sebagai sarana pendukung, yaitu:

4.1.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Berikut merupakan spesifikasi dari perangkat keras yang digunakna dalam penelitian ini :

1. Laptop Asus TUF GAMING F15
2. Processor Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz 2.50 GHz
3. Daya tampung 16 GB

Adapun spesifikasi dari perangkat lunak yang digunakan dalam penelitian ini meliputi:

1. OS Windows 11 Home Single Language
2. Bahasa pemrograman Python versi 1.9.13
3. IDE : *Jupiter Notebook* dan *Microsoft Visual Studio Code*

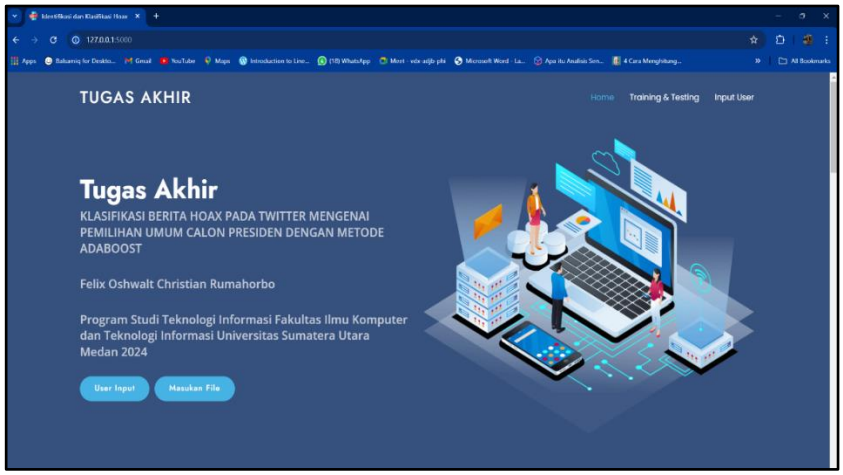
4.1.2. Penerapan Sistem Berbasis Web

Berikut merupakan implementasi dari rancangan sistem berbasis web yang telah di jelaskan pada bab sebelumnya. Sistem ini terdiri dari 3 menu yaitu Home, Input Teks dan Training/Testing.

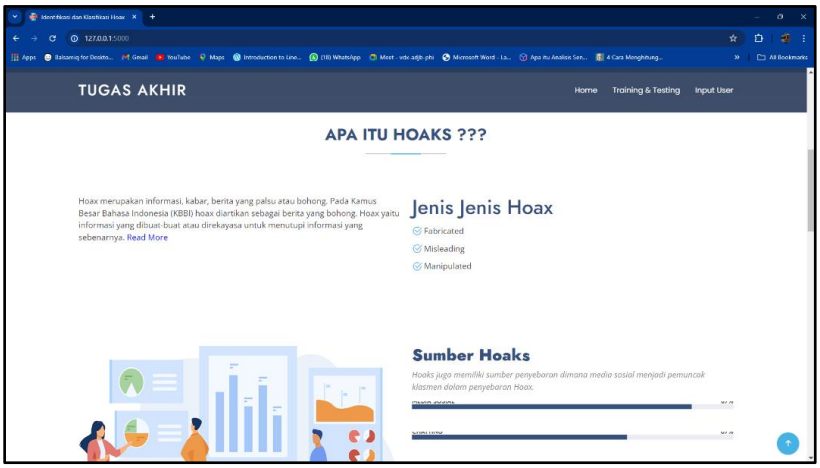
1. Halaman Home

Pada halaman ini, pengguna dapat melihat tampilan beranda yang memuat informasi umum tentang hoax, berbagai jenis hoax yang dibahas dalam penelitian ini, serta beberapa contoh hoax berdasarkan klasifikasi yang telah ditetapkan. Untuk masuk

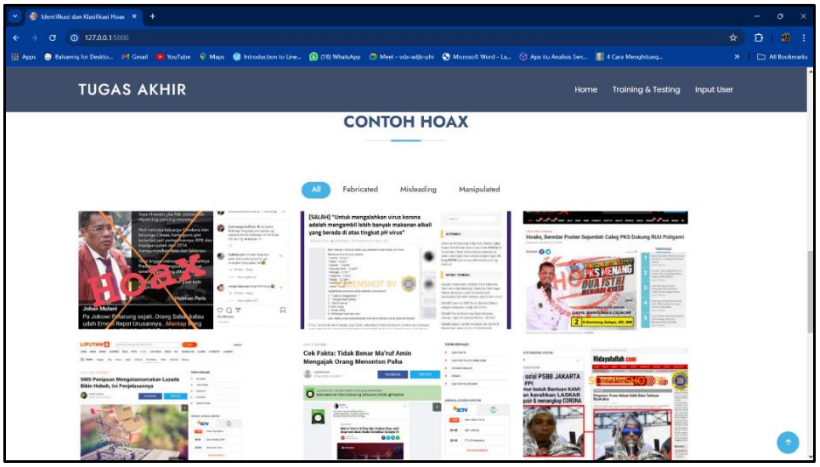
kehalaman yang selanjutnya, penulis sudah menyediakan button “Input Teks” dan button “Training dan Testing” (Gambar 4.1).



Gambar 4.1 Halaman Home



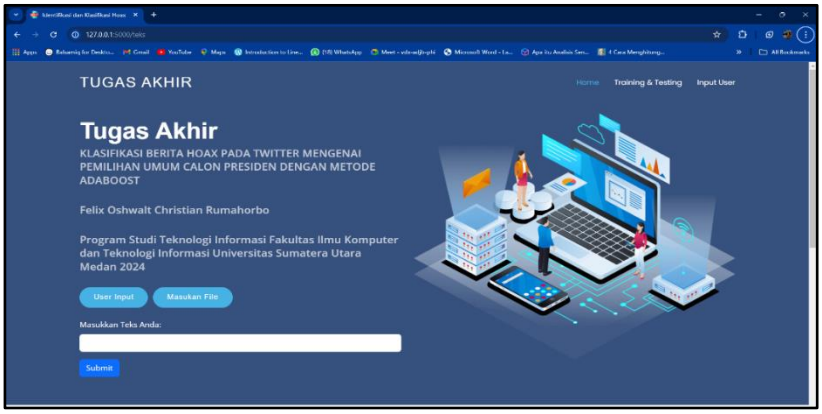
Gambar 4.2 Halaman Home Info Hoaks



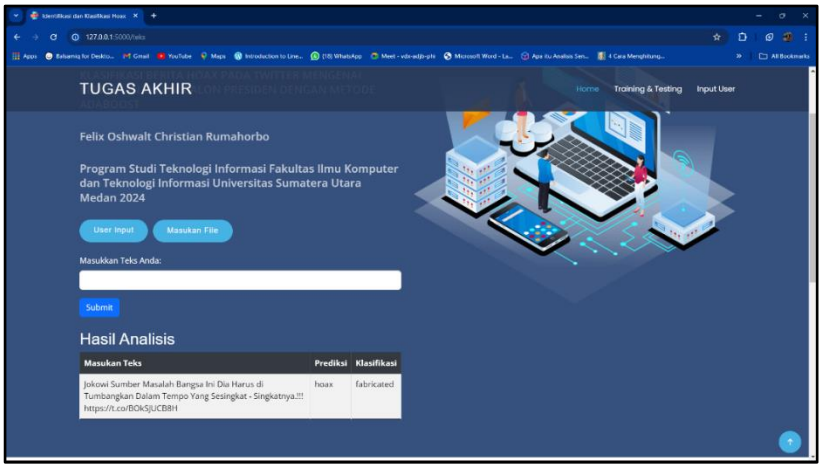
Gambar 4.3 Halaman Home Contoh Hoax

2. Halaman Input Teks

Halaman Input Teks memungkinkan pengguna untuk menginput sebuah teks yang diambil dari aplikasi X atau sebuah tweet. Pengguna dapat memasukan teks yang ingin di klasifikasi ke dalam teksbox yang telah di sediakan (Gambar 4.4). Setelah menginput teks, maka akan dilakukan pengecekan. Pengguna dapat mengklik tombol " Analisis". Selanjutnya teks yang dimasukan akan di proses dan menampilkan output dibagian bawah button berupa hasil prediksi yang berada di dalam tabel yang berisikan full_text, label dan klasifikasi (Gambar 4.5).



Gambar 4.4 Input Teks

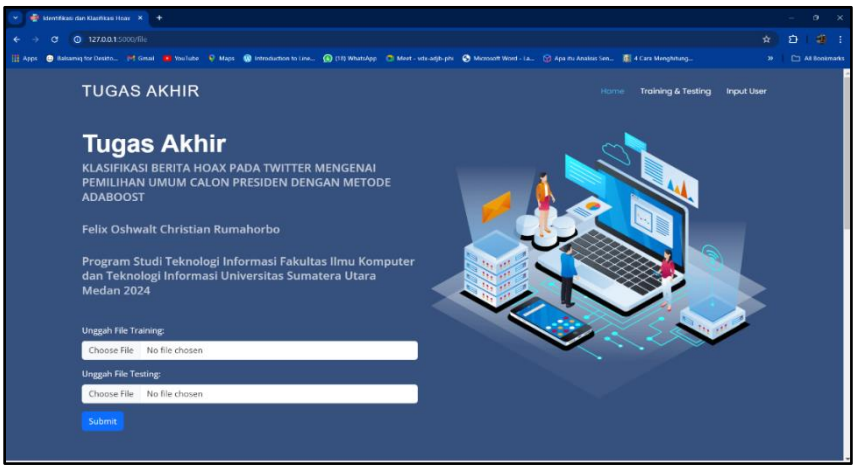


Gambar 4.5 Output Klasifikasi

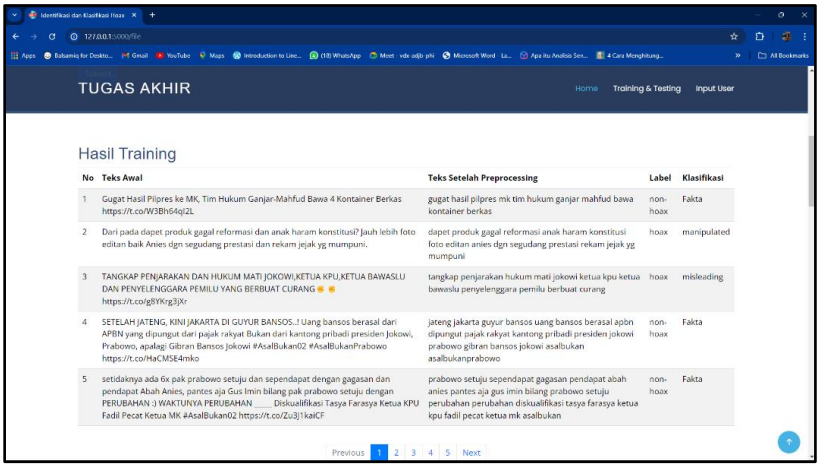
3. Halaman Training dan Testing

Halaman Training Testing ini memungkinkan pengguna untuk mengupload file training dan file testing untuk melakukan klasifikasi hoax menggunakan file training dan testing (gambar 4.6). Selanjutnya sistem akan melakukakn proses training dan

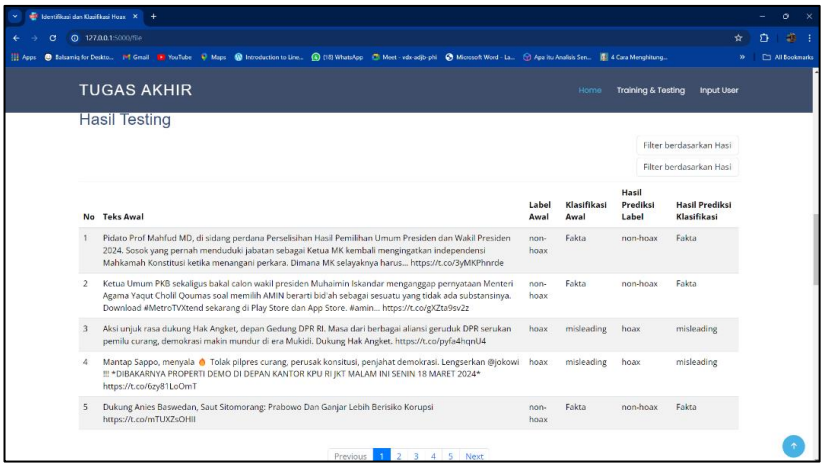
testing dengan file tersebut. Kemudian setelah proses training dan testing selesai maka akan menampilkan hasil training dan hasil testing, untuk hasil training akan menampilkan teks sebelum dilakukannya preprocesing, teks setelah preprocessing, label dan klasifikasi (gambar 4.7). Sedangkan untuk hasil testing akan menampilkan full_text, label awal, klasifikasi awal, hasil prediksi label dan hasil prediksi klasfikasi. Selain itu pada data testing akan menampilkan confusion matrix label dan confusion matriks klasifikasi. Penulis juga menambahkan fitur pagination dan fitur filter agar memudahkan pengguna untuk mencari hasil prediksi (gambar 4.8.).



Gambar 4.6 Halaman Training dan Testing



Gambar 4.7 Hasil Training



Gambar 4.8 Hasil Testing

4.2. Hasil Pelatihan Model

Penelitian ini menggunakan metode Decision Tree sebagai model dasar dan menggunakan Adaptive Boosting sebagai metodenya untuk meningkatkan kinerja dalam melakukan prediksi. Decision Tree digunakan sebagai base estimator karena model ini merupakan model yang sederhana dan sangat efektif dalam konteks boosting, kesederhanaan itu memungkinkan model dengan cepat diadaptasi dan ditingkatkan melalui iterasi boosting. Sedangkan Adaptive Boosting digunakan sebagai metode utamanya adalah Adaboost meningkatkan akurasi model dengan menggabungkan beberapa classifier yang lemah seperti hasil dari decision tree untuk membentuk classifier yang kuat, selain itu Adaboost memiliki kemampuan untuk menangani overfitting dan peningkatan akurasi (Freund dan Schapire , 1997). Peneliti menggunakan beberapa parameter yang dikombinasi untuk menghasilkan akurasi yang tinggi dan akurat. Adapun parameter yang digunakan dalam penelitian ini adalah sebagai berikut :

Tabel 4.1 Daftar parameter

Parameter	Value
Base Estimator “max_depth”	2, 4, 6
Adaboost “algorithm”	“SAMME”
Adboost “n_estimator” label	50, 100, 500
Adboost “n_estimator” klasifikasi	100, 500, 1000
Adboost “learning_rate”	0.5, 1.0, 1.5

Dari parameter pada tabel 4.1. tersebut, selajutnya dilakukan kombinasi dan hasilnya dapat dilihat pada tabel 4.2 :

Tabel 4.2 Hasil Akurasi Kombinasi Parameter

max_depth	algorithm	n_estimators_label	n_estimators_classification	learning_rate	accuracy_label	Accuracy_klasifikasi
2	SAMME	50	100	0.1	89%	82%
2	SAMME	50	100	0.5	88%	83%
2	SAMME	50	100	1	88%	80%
2	SAMME	50	100	1.5	86%	79%
2	SAMME	50	500	0.1	89%	83%
2	SAMME	50	500	0.5	88%	82%
2	SAMME	50	500	1	88%	81%
2	SAMME	50	500	1.5	90%	82%
2	SAMME	50	1000	0.1	89%	83%
2	SAMME	50	1000	0.5	88%	84%
2	SAMME	50	1000	1	88%	81%
2	SAMME	50	1000	1.5	91%	79%
2	SAMME	100	100	0.1	89%	82%
2	SAMME	100	100	0.5	90%	83%
2	SAMME	100	100	1	90%	80%
2	SAMME	100	100	1.5	91%	79%
2	SAMME	100	500	0.1	89%	83%
2	SAMME	100	500	0.5	90%	82%
2	SAMME	100	500	1	89%	81%
2	SAMME	100	500	1.5	88%	82%
2	SAMME	100	1000	0.1	89%	83%
2	SAMME	100	1000	0.5	91%	84%
2	SAMME	100	1000	1	89%	81%
2	SAMME	100	1000	1.5	91%	79%
2	SAMME	500	100	0.1	88%	82%
2	SAMME	500	100	0.5	88%	83%
2	SAMME	500	100	1	89%	80%

2	SAMME	500	100	1.5	87%	79%
2	SAMME	500	500	0.1	88%	83%
2	SAMME	500	500	0.5	88%	82%
2	SAMME	500	500	1	91%	81%
2	SAMME	500	500	1.5	88%	81%
2	SAMME	500	1000	0.1	88%	83%
2	SAMME	500	1000	0.5	88%	84%
2	SAMME	500	1000	1	89%	81%
2	SAMME	500	1000	1.5	87%	79%
4	SAMME	50	100	0.1	86%	86%
4	SAMME	50	100	0.5	87%	84%
4	SAMME	50	100	1	87%	84%
4	SAMME	50	100	1.5	87%	83%
4	SAMME	50	500	0.1	86%	85%
4	SAMME	50	500	0.5	88%	84%
4	SAMME	50	500	1	89%	84%
4	SAMME	50	500	1.5	86%	83%
4	SAMME	50	1000	0.1	87%	85%
4	SAMME	50	1000	0.5	88%	83%
4	SAMME	50	1000	1	90%	83%
4	SAMME	50	1000	1.5	88%	82%
4	SAMME	100	100	0.1	88%	86%
4	SAMME	100	100	0.5	88%	84%
4	SAMME	100	100	1	89%	83%
4	SAMME	100	100	1.5	90%	83%
4	SAMME	100	500	0.1	87%	84%
4	SAMME	100	500	0.5	90%	83%
4	SAMME	100	500	1	91%	84%
4	SAMME	100	500	1.5	87%	85%
4	SAMME	100	1000	0.1	88%	85%
4	SAMME	100	1000	0.5	88%	83%
4	SAMME	100	1000	1	90%	84%
4	SAMME	100	1000	1.5	87%	86%

4	SAMME	500	100	0.1	88%	85%
4	SAMME	500	100	0.5	88%	84%
4	SAMME	500	100	1	86%	82%
4	SAMME	500	100	1.5	87%	83%
4	SAMME	500	500	0.1	88%	84%
4	SAMME	500	500	0.5	90%	84%
4	SAMME	500	500	1	89%	84%
4	SAMME	500	500	1.5	88%	85%
4	SAMME	500	1000	0.1	88%	83%
4	SAMME	500	1000	0.5	88%	83%
4	SAMME	500	1000	1	87%	83%
4	SAMME	500	1000	1.5	87%	83%
6	SAMME	50	100	0.1	86%	84%
6	SAMME	50	100	0.5	90%	83%
6	SAMME	50	100	1	89%	84%
6	SAMME	50	100	1.5	86%	85%
6	SAMME	50	500	0.1	87%	84%
6	SAMME	50	500	0.5	88%	81%
6	SAMME	50	500	1	90%	83%
6	SAMME	50	500	1.5	88%	84%
6	SAMME	50	1000	0.1	88%	84%

Akurasi yang di dapatkan pada tabel 4.2. diatas merupakan akurasi rata-rata dari beberapa kombinasi parameter pada data testing yang telah ditentukank. Kesimpulan yang dapat diambil bahwa parameter terbaik dalam pelatihan Adaptive boosting pada adalah dengan menggunakan max_depth sebesar 2, algorithm “SAMME” ,n_estimato_label sebesar 100, n_estimator_klasifikasi sebesar 1000 dan learning rate 0.5. Kombinasi pada parameter tersebut berhasil mencapai akurasi pada label sebesar 91.15% dan pada klasifikasi sebesar 83.77% dalam mentukan label hoax dan non-hoax dan jenis hoaxnya.

4.3. Hasil Pengujian Sistem

Pada tahap ini, model yang telah dibangun dan disimpan selama proses pelatihan akan diuji dengan data uji untuk mengevaluasi kemampuannya dalam melakukan klasifikasi hoax terkait pemilihan presiden dan wakil presiden pada aplikasi X menggunakan metode Adaboost. Pengujian ini memanfaatkan data yang belum pernah dipelajari oleh model sebelumnya, dengan jumlah 191 data. Data ini juga akan melewati tahapan preprocessing seperti pada tahap pelatihan, diikuti dengan tahap pembobotan menggunakan TF-IDF, sebelum diuji menggunakan model Adaboost untuk menghasilkan prediksi. Hasil pengujian tersebut ditampilkan pada tabel 4.3.

Tabel 4.3 Hasil Pengujian Sistem pada Data Testing

Tweet	Label	Klasifikasi	Hasil Prediksi
Pidato Prof Mahfud MD, di sidang perdana Perselisihan Hasil Pemilihan Umum Presiden dan Wakil Presiden 2024. Sosok yang pernah menduduki jabatan sebagai Ketua MK kembali mengingatkan independensi Mahkamah Konstitusi ketika menangani perkara. Dimana MK selayaknya harus… https://t.co/3yMKPhnrde	Non-hoax	Non-hoax	Benar
Ketua Umum PKB sekaligus bakal calon wakil presiden Muhaimin Iskandar menganggap pernyataan Menteri Agama Yaqut Cholil Qoumas soal memilih AMIN berarti bid'ah sebagai sesuatu yang tidak ada substansinya. Download #MetroTVXtend sekarang di Play Store dan App Store. #amin… https://t.co/gXZta9sv2z	Non-hoax	Non-hoax	Benar
Aksi unjuk rasa dukung Hak Angket, depan	Hoax	misleading	Benar

Gedung DPR RI. Masa dari berbagai aliansi geruduk DPR serukan pemilu curang, demokrasi makin mundur di era Mukidi. Dukung Hak Angket. https://t.co/pyfa4hqnU4			
Mantap Sappo, menyala 💧 Tolak pilpres curang, perusak konsitusi, penjahat demokrasi. Lengserkan @jokowi !!! *DIBAKARNYA PROPERTI DEMO DI DEPAN KANTOR KPU RI JKT MALAM INI SENIN 18 MARET 2024* https://t.co/6zy81LoOmT	Hoax	misleading	Benar
Dukung Anies Baswedan, Saut Sitomorang: Prabowo Dan Ganjar Lebih Berisiko Korupsi https://t.co/mTUXZsOHII	Non-hoax	Non-hoax	benar
Apabila hasil pemilu curang ditetapkan juga oleh KPU, maka saya mengusulkan ormas2 besar di Indonesia (minimal NU &amp Muhammadiyyah) membuat deklarasi 'menolak hasil pemilu curang'. Bila ini tdk dilakukan berarti ormas2 tsb menerima hasil kecurangan. Ingat terus #6SyuhadaFPI #KM50	hoax	fabricated	Benar
@tvOneNews Arogan, songong, angkuh dan merasa tidak tersentuh hukum. Mentang2 didukung jokowi dia pikir bisa bertindak semena2 ? Semakin ngeri membayangkan kalau paslon ini menang, Capres agenda tersembunyi nya pelanggar HAM berat, Cawapresnya anak haram konstitusi.. apa yg bs diharapkan dr paslon spt ini ?	hoax	manipulated	Benar
Jokowi akan di kenang sejarah sebagai presiden yang telah merusak tatanan, bukan	hoax	fabricated	Benar

saja Demokrasi tapi juga merusak tatanan bernegara secara umum. #JokowiPerusakDemokrasi #PemiluCurang https://t.co/N4zLJidhR9			
JOROK SEKALI... TPN GANJAR MAHFUD MINTA JOKOWI STOP POLITIK JOROK LEWAT PEMILU CURANG....!! #APBNSubsidiTaipanCina #APBNSubsidiTaipanCina 🆘 🆗 https://t.co/Ay75d3x6FQ	hoax	misleading	Salah
Pengumuman Bakal Calon Presiden &amp Bakal Calon Wakil Presiden Republik Indonesia AMIN' @aniesbaswedan - @cakimiNOW ᴉᴅ Semoga Allah meridhoi 🙏 https://t.co/9hScPtVSCa"	non-hoax	non-hoax	Benar
Jakarta 19 Maret 2024 Opo Ora Ngeri Rek? Massa Tumpah Ruah Jadi Lautan Manusia. Kepung Gedung Kpu &Amp Dpr Ri👍 Massa Besar2an Kepung Gedung Dpr Ri Tuntut Hak Angket👍 Tolak Hasil Pemilu Curang. Ini Bikin Kakek Wowo Gagal Bro🤣 Bikin Kakek Wowo &Amp Pak De Wiwi Bingung &Amp Stres Bro😄🤣 Https://T.Co/1nmkpzklw5	hoax	misleading	Salah

Saat melakukan pengujian model, ditemukan adanya ketidaktepatan sistem dalam mengklasifikasikan data karena beberapa faktor. Penulis menganalisis beberapa faktor yang memungkinkan penyebab terjadinya kesalahan dalam prediksi, yaitu :

1. Dalam suatu tweet terdapat banyak kata yang cenderung mengarah kesuatu kelas tertentu, sehingga cenderung salah prediksi.

   Contoh :

   a. kl habibie nyapres dicap sok intelek omon omon
   b. nepotisme dipertontonkan terang terangan prabowo ngoceh berantas korupsi biji

(Seperti contoh diatas, sistem memprediksi bahwa tweet tersebut adalah “non-hoax” karena banyaknya kata kata yang mengacu pada kurangnya informasi pada model atau konteks tentang kejadian tersebut yang mengarahkan sistem untuk memprediksi menjadi “non-hoax”)

2. Kesalahan prediksi juga disebabkan karena ada kesamaan kata atau ciri kata pada setiap kelas. Pada tweet non-hoax terdapat kata seperti “ ngeri”, “kotor”, namun juga banyak ditemukan di tweet hoax.
3. Selanjutnya kelasahan prediksi pada input user masih bisa terjadi dikarenakan data yang digunakan masih kurang bervarisasi yang mengakibatkan model kurang mengenal pola-pola pada setiap kalimat.

   Contoh :

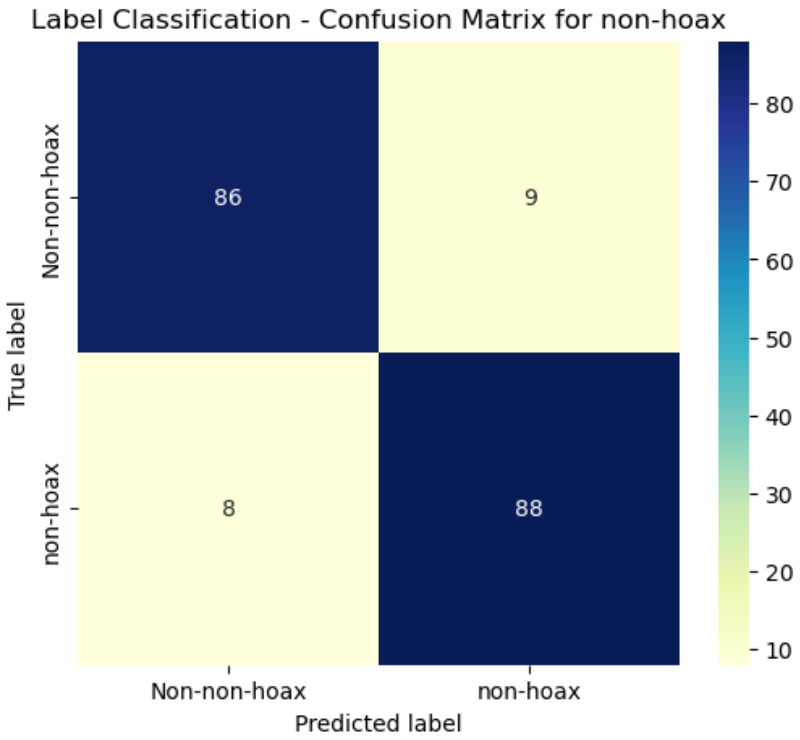
   a. Prabowo memaksa gibran untuk maju jadi cawapres

   (Seperti contoh diatas, sistem memiliki prediksi bahwa tweet tersebut adalah “non hoax” karena masih kurangnya data yang mengandung konteks atau kejadian tersebut)

4.4. Metode Evaluasi

Metode evaluasi adalah cara untuk menilai seberapa baik model klasifikasi bekerja, termasuk untuk mengklasifikasikan tweet hoax dengan menggunakan metode Adaboost. Salah satu cara yang sering digunakan adalah dengan confusion matrix, yaitu tabel yang membandingkan hasil prediksi model dengan hasil yang sebenarnya. Dari tabel ini, kita bisa menghitung beberapa nilai penting seperti precision, recall, f1-score, dan accuracy.

Precision menunjukkan seberapa akurat model dalam memprediksi tweet hoax yang benar-benar hoax, sedangkan recall menunjukkan seberapa baik model dalam menemukan semua tweet hoax. F1-score adalah kombinasi dari precision dan recall yang memberikan gambaran seimbang tentang kinerja model. Accuracy adalah persentase dari semua prediksi yang benar. Dalam penelitian ini, hasil evaluasi ini akan ditampilkan dalam bentuk heatmap, yaitu gambar yang menggunakan warna untuk menunjukkan nilai-nilai tersebut, sehingga memudahkan kita untuk melihat dan memahami hasil evaluasi dengan cepat dan mudah.

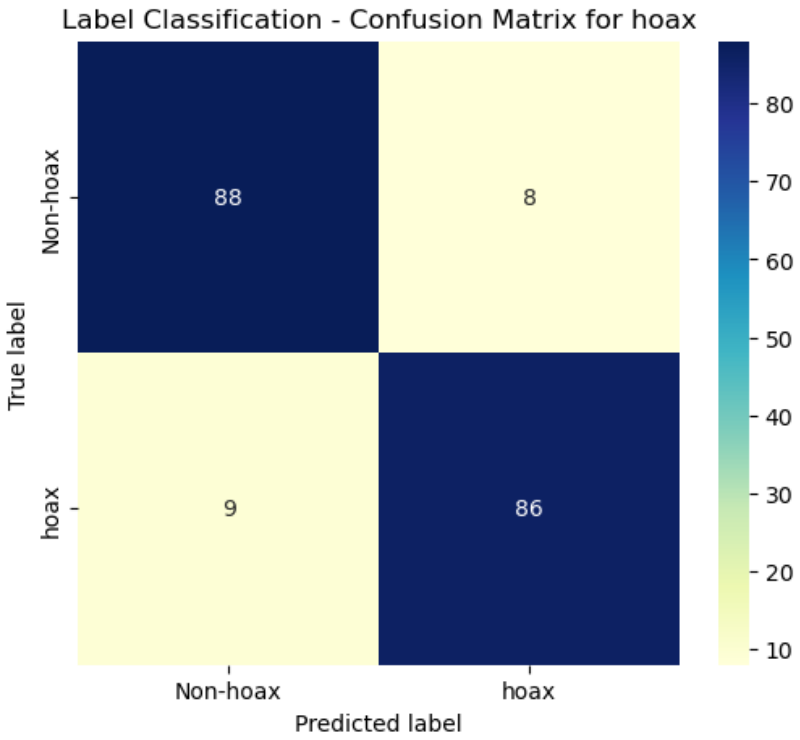


Gambar 4.9 Confussion Matrix Non-Hoax pada Label

Pada confussion matrix ini dilakukan perhitungan dan pengujian pada label non-hoax menunjukan bahwa data sebanyak 88 data benar di prediksi sebagai non-hoax dan 9 data yang memiliki label hoax dan diprediksi sebagai non-hoax. Sedangkan banyaknya data yang memiliki label hoax dan di klasifikasi sebagai hoax ada sebanyak 86 dan data yang memiliki label non-hoax dan di prediksi sebagai hoax prediksi ada sebanyak 8 data. Pada tabel 4.4 dapat dilihat informasi rincian mengenai label non-hoax dalam confusion matrix.

Tabel 4.4 Keterangan Confusin Matrix Label Non-hoax

No	Label non-hoax	Jumlah
1	TPT (True Positive)	88
2	FPT (False Positive)	8
3	TNT (True Negative)	86
4	FNT (False Negative)	9

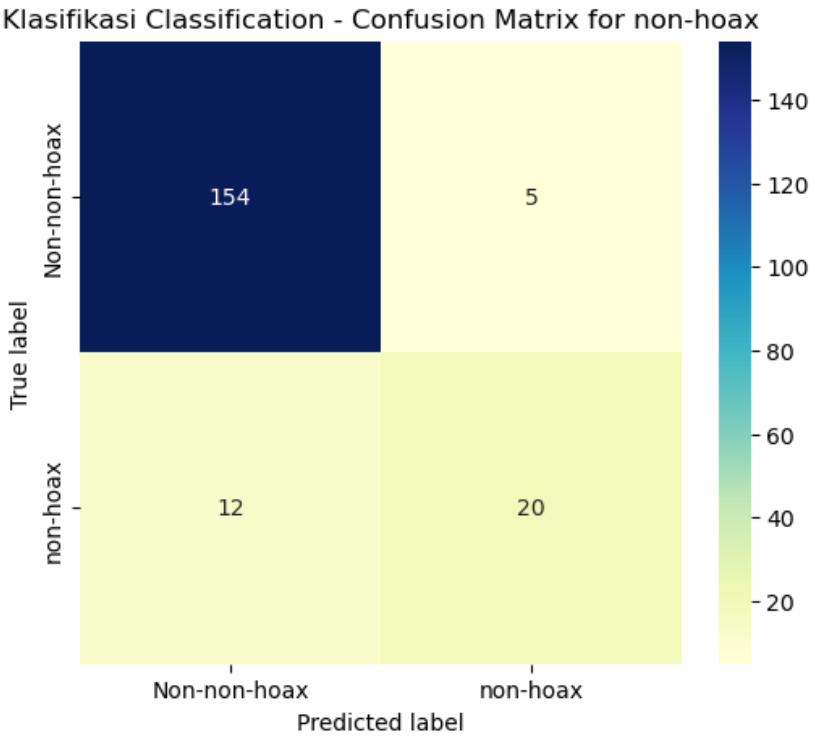


Gambar 4.10 Confussion Matrix Pada Label Hoax

Berdasarkan gambar confusion matrix untuk label hoax, dilakukan perhitungan dan pengujian yang menunjukkan hasil sebagai berikut: Data sebanyak 86 benar diprediksi sebagai hoax (True Positive). Namun, terdapat 9 data hoax yang salah diprediksi sebagai non-hoax (False Negative), serta 8 data non-hoax yang salah diprediksi sebagai hoax (False Positive). 88 data non-hoax yang benar diprediksi sebagai non-hoax (True Negative). Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi data berlabel hoax, namun masih terdapat kesalahan dalam memprediksi data non-hoax sebagai hoax dan data hoax sebagai non-hoax. Tabel 4.5 memberikan infoemasi mengenai rincian label hoax dalam confusion matrix

Tabel 4.5 Keterangan Confusin Matrix Label Hoax

No	Label non-hoax	Jumlah
1	TPT (True Positive)	86
2	FPT (False Positive)	8
3	TNT (True Negative)	88
4	FNT (False Negative)	9

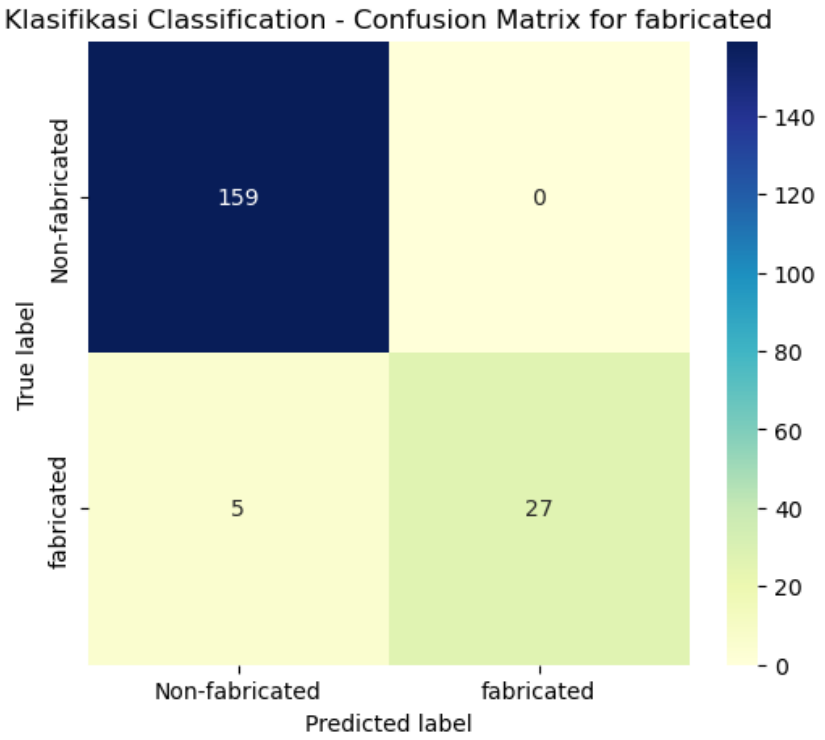


Gambar 4.11 Confussion Matrix Pada Klasifikasi Non-hoax

Berdasarkan perhitungan yang dilakukan pada confusion matrix dari pengujian aspek non-hoax pada klasifikasi, hasilnya menunjukkan bahwa 20 data benar diprediksi sebagai non-hoax, sementara 5 data hoax diprediksi salah sebagai non-hoax. Selain itu, terdapat 154 data yang benar-benar non-hoax dan diprediksi dengan benar, sedangkan 12 data non-hoax diprediksi salah sebagai hoax. Tabel 4.6 memberikan rincian lebih lanjut mengenai klasifikasi non-hoax dalam confusion matrix ini.

Tabel 4.6 Keterangan Confusin Matrix Klasifikasi Non-Hoax

No	Label non-hoax	Jumlah
1	TPT (True Positive)	20
2	FPT (False Positive)	5
3	TNT (True Negative)	154
4	FNT (False Negative)	12

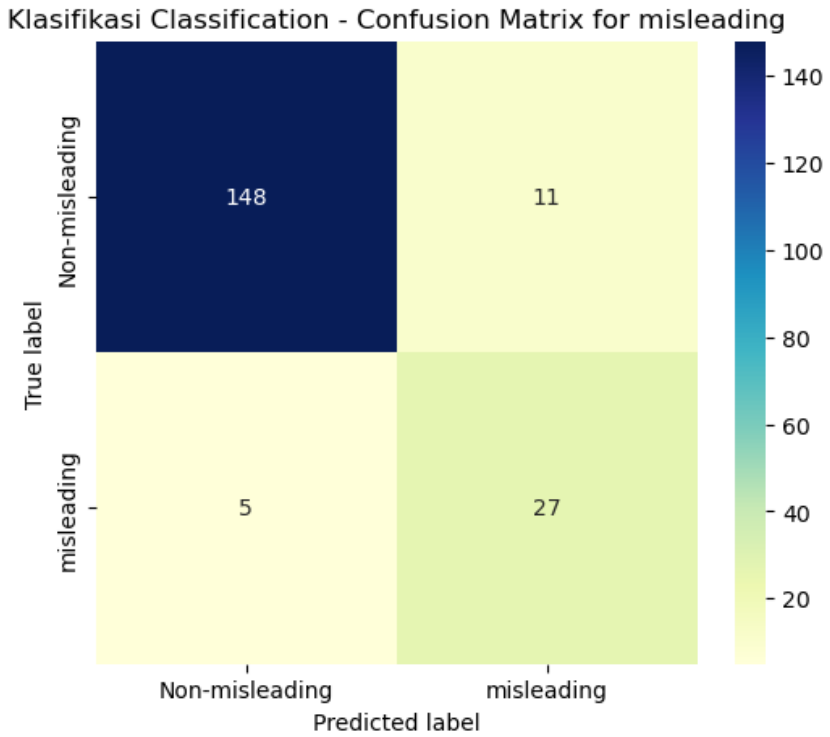


Gambar 4.12 Confussion Matrix Fabricated Pada Klasifikasi

Berdasarkan hasil perhitungan pada confusion matrix dari pengujian klasifikasi fabricated, ditemukan bahwa 27 data berhasil diprediksi dengan benar sebagai fabricated, sementara ada 5 data yang bukan fabricated yang salah diprediksi sebagai fabricated. Selain itu, ada 159 data yang bukan fabricated dan diprediksi dengan tepat, dan tidak ada data fabricated salah diprediksi bukan fabricated. Tabel 4.7 memberikan rincian lebih detail mengenai klasifikasi fabricated dalam confusion matrix ini.

Tabel 4.7 Keterangan Confusin Matrix Klasifikasi Fabricated

No	Label non-hoax	Jumlah
1	TPT (True Positive)	27
2	FPT (False Positive)	0
3	TNT (True Negative)	159
4	FNT (False Negative)	5

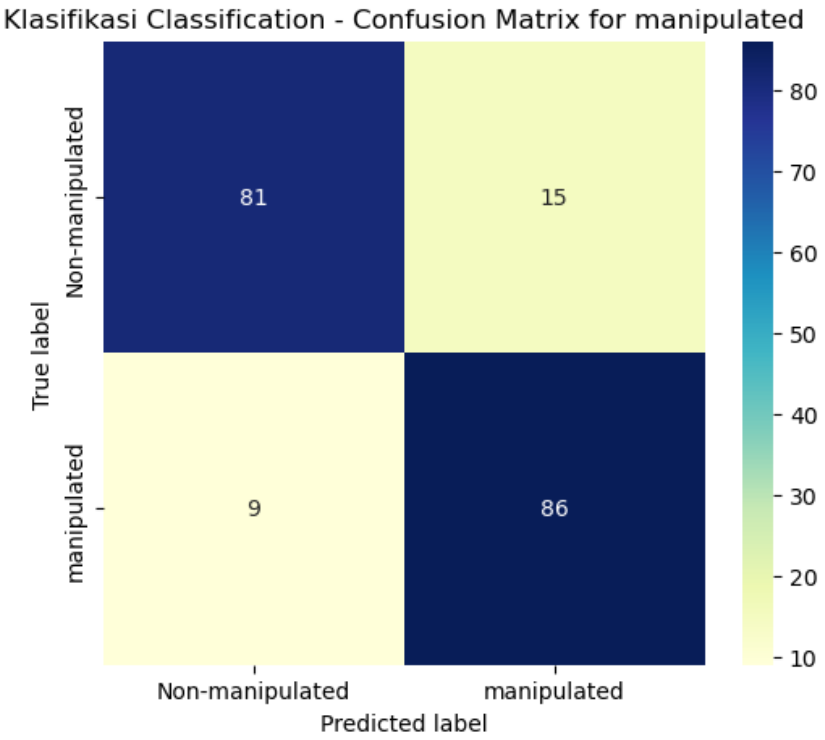


Gambar 4.13 Confussion Matrix Misleading Pada Klasifikasi

Berdasarkan hasil perhitungan pada confusion matrix dari pengujian aspek klasifikasi misleading, diketahui bahwa 27 data berhasil diprediksi dengan benar sebagai misleading, sementara tedapat 11 data yang bukan misleading yang salah diprediksi sebagai misleading. Selain itu, terdapat 148 data yang bukan misleading dan diprediksi dengan benar, sedangkan 5 data misleading diprediksi salah sebagai bukan misleading. Rincian lebih detail mengenai klasifikasi non-hoax dalam confusion matrix ini disajikan dalam Tabel 4.8.

Tabel 4.8 Keterangan Confusin Matrix Klasifikasi Misleading

No	Label non-hoax	Jumlah
1	TPT (True Positive)	27
2	FPT (False Positive)	11
3	TNT (True Negative)	148
4	FNT (False Negative)	5



Gambar 4.14 Confussion Matrix Manipulated Pada Klasifikasi

Berdasarkan perhitungan pada confusion matrix dari pengujian klasifikasi manipulated, diketahui bahwa 86 data berhasil diprediksi dengan benar sebagai manipulated, sementara terdapat 15 data yang sebenarnya bukan manipulated namun salah diprediksi sebagai manipulated. Selain itu, terdapat 81 data yang yang bukan misleading dan diprediksi dengan tepat, sedangkan 3 data manipulated diprediksi salah sebagai bukan manipulated. Rincian lebih detail mengenai klasifikasi non-hoax dalam confusion matrix ini dapat ditemukan pada Tabel 4.9.

Tabel 4.9 Keterangan Confusin Matrix Klasifikasi Manipulated

No	Label non-hoax	Jumlah
1	TPT (True Positive)	86
2	FPT (False Positive)	15
3	TNT (True Negative)	81
4	FNT (False Negative)	9

Berdasarkan tabel yang telah di urasikan diatas, evaluasi dapat dilakukan dengan mengitung *classification report* seperti *precision, recall dan f1-score* , berikut adalah hasil perhitungan pada kelas label adan klasifikasi :

a. Label Non-hoax

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{88}{88+8}\ x\ 100\% = \quad 91\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{88}{88+9}\ x\ 100\% = \quad 90\%$

F1-Score $= 2x\ \frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{0.91\ x\ 0.90}{0.91 + 0.90}\ x\ 100\%90\%$

b. Label Hoax

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{86}{86+9}\ x\ 100\% = \quad 90\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{86}{86+8}\ x\ 100\% = \quad 91\%$

F1-Score $= 2x\ \frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{0.90\ x\ 0.91}{0.90 + 0.91}\ x\ 100\% = \ 90\%$

c. Klasifikasi Non-hoax

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{20}{20\ +5}\ x\ 100\% = \ 80\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{20}{20+12}\ x\ 100\% = \quad 62\%\%$

F1-Score $=2x\frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{0.80\ x\ 0.62}{0.80 + 0.62}\ x\ 100\% = \ 70\%$

d. Klasifikasi Fabricated

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{27}{27+\ 0}\ x\ 100\% = \ 100\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{27}{27+\ 5}\ x\ 100\% = \quad 84\%\%$

F1-Score $=2x\frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{1.00x\ 0.84}{1.00 + 0.84}\ x\ 100\% = \quad 91\%$

e. Klasifikasi Misleading

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{27}{27+11}\ x\ 100\% = \ 71\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{27}{27+5}\ x\ 100\% = \ 84\%$

F1-Score $=2x\frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{0.71\ x\ 0.84}{0.71 + 084}\ x\ 100\% = \quad 77\%$

f. Klasifikasi Manipulated

Precission $= \frac{TP}{TP+FP}\ x\ 100\% = \frac{86}{86+15}\ x\ 100\% = \ 85\%$

Recall $= \frac{TP}{TP+FN}\ x\ 100\% = \frac{86}{86+9}\ x\ 100\% = \ 90\%$

F1-Score $= 2x\frac{Precision\ x\ Recall}{Precision + Recall}\ x\ 100\% = 2x\ \frac{0.85x\ 0.90}{0.85 + 0.90}\ x\ 100\% = \ 87\%$

Berdasarkan hasil pergitungan evaluasi diatas berikut adalah rincian yang dapat dilihat pada tabel 4.10.

Tabel 4.10 Hasil Perhitungan Evaluasi Pada Label

Jenis Label	Precision	Recall	F1-Score
Label Non-hoax	91%	90%	90%
Label Hoax	90%	91%	90%
Rata-Rata	90.5%	90.5%	90%

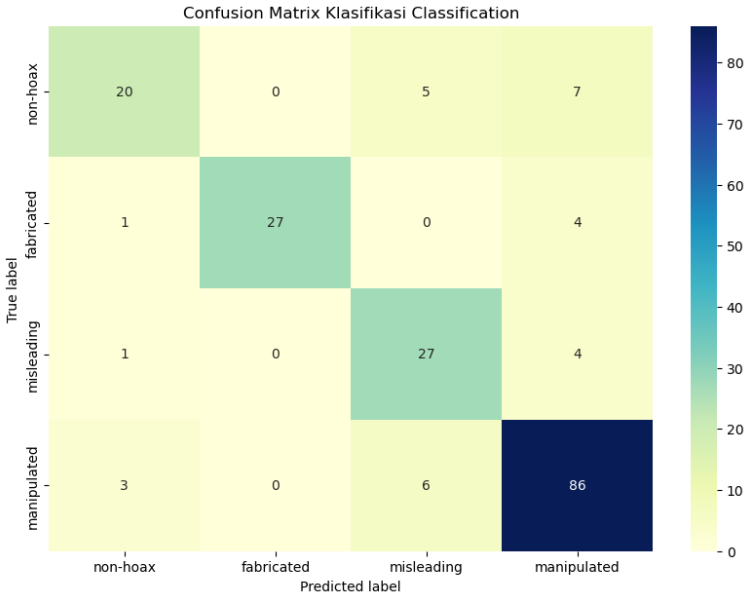
Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} x100\%$

$= \frac{174+174}{174+174+17+17} x100\%$

$= 91.15\ \%$

Tabel 4.11 Hasil Perhitungan Evaluasi Pada Klasifikasi

Jenis Klasifikasi	Precision	Recall	F1-Score
Klasifikasi Non-hoax	80%	62%	70%
Klasifikasi Fabricated	100%	84%	91%
Klasifikasi Misleading	71%	84%	77%
Klasifikasi Manipulated	85%	90%	87 %
Rata-Rata	84%	80%	81%



Gambar 4.15 Confusion Matrix Pada Gabungan Klasifikasi

Pada gambar 4.15 ditunjukan apabila menggabungkan confusin matrix dari tiap jenis klasifikasi. Dan berikut adalah hasil akurasi jika confusion matrix pada Klasifikasi hoax

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} x100\%$

$= \frac{20+27+27+86}{32+32+32+95} x100\%$

$= 83.77\ \%$

Berdasarkan hasil pengujian, model Adaptive Boosting yang diterapkan untuk mengklasifikasikan hoax dalam aplikasi X terkait pemilihan presiden dan wakil presiden menunjukkan kinerja yang memadai. Secara rata-rata, model ini mencapai precision, recall, dan f1-score sebesar 90% untuk label, serta akurasi sebesar 91%. Untuk klasifikasi, nilai rata-rata precision adalah 84%, recall 80%, dan f1-score 81%, dengan akurasi mencapai 84%. Dengan demikian, dapat disimpulkan bahwa model ini berhasil dalam tugasnya untuk mengenali konten hoax dengan baik dalam konteks yang diberikan.

BAB 5
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan dari penelitain ini untuk melakukan klasifikasi hoax pada aplikasi X mengenai pemilihan calon presiden dan wakil presiden menggunakn Metode Adaptive Boosting adalah sebagai berikut :

1. Penggunaan metode Adaptive Boosting terbukti efektif dalam melakukan klasifikasi hoax pada aplikasi X. Penelitian ini menggunakan total 954 data berupa tweet, yang dibagi menjadi 80% untuk training dan 20% untuk testing. Proses klasifikasi diawali dengan tahap preprocessing untuk membersihkan data, diikuti oleh pemberian bobot menggunakan metode TF-IDF dengan fitur n-gram (1, 2) dan normalisasi 'l2'. Setelah itu, dilakukan perhitungan klasifikasi menggunakan metode Adaptive Boosting.
2. Hasil akhir dari proses klasifikasi ini tidak hanya menentukan apakah sebuah tweet termasuk hoax atau tidak, tetapi juga mengklasifikasikan jenis hoax tersebut menjadi fabricated, misleading, atau manipulated. Penelitian ini menunjukkan bahwa metode Adaptive Boosting dapat memberikan hasil yang akurat dalam mengidentifikasi dan mengklasifikasikan hoax pada tweet, sehingga membantu meningkatkan keandalan aplikasi X dalam mendeteksi informasi yang salah.
3. Berdasarkan hasil pengujian klasifikasi tweet hoax menggunakan metode Adaboost, diperoleh nilai rata-rata precision sebesar 90%, recall 90%, dan f1-score sebesar 90% untuk identifikasi hoax dan non-hoax. Sedangkan, untuk klasifikasi jenis hoax, nilai rata-rata precision sebesar 84%, recall 80%, dan f1-score 81%. Selain itu, parameter-parameter seperti n_estimator, max_depth, dan learning_rate yang digunakan dalam algoritma ini mempengaruhi tingkat akurasinya.

4. Terdapat banyak tweet yang mengandung kata-kata yang mengarah ke suatu jenis klasifikasi tertentu dan juga banyak pola-pola kata yang memiliki kesamaan atau ciri kata pdada beberapa jenis klasifikasi yang berpengaruh pada hasil klasifiaksi pada penelitian ini.

5.2. Saran

Penulis memberikan saran pada pengembangan sistem pada penlitian ini sebagai berikut :

1. Menambah lebih banyak dataset yang lebih beragam dan pembagaian yang seimbang serta memiliki pola yang seimbang di setiap label yang diberikan agar meminimalisir terjadinya error pada saat sistem membaca pola-pola kata atau konteks tertentu.
2. Menggunakan proses normalisasi lebih banyak lagi seperti membuat kamus sendiri untuk membuat normalisasi lebih banyak.
3. Pada saat ingin melakukan identifiaksi tweet hoax dan klasifiaksi hoax gunakan pembanding untuk mengetahui fakta dari kebenaran hoax tersebut, dan buat pembanding untuk klasifikasi hoax tersebut.

DAFTAR PUSTAKA

Abdulrahman, A., & Baykara, M. (2020). Fake News Detection Using Machine Learning and Deep Learning Algorithms. *3rd International Conference on Advanced Scienceand Engineering, ICOASE 2020*, 18–23. https://doi.org/10.1109/ICOASE51841.2020.9436605

Airell Rizqiawan, M., Dwi Wibawa, Igp. S., & Kallista SSi, M. (2023). *Klasifikasi Polusi Udara Menggunakan Adaboost Berbasis Pembelajaran Mesin*. https://www.marlborough.govt.nz/environment/air-

Bauer, E., Chan, P., Stolfo, S., & Wolpert, D. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. In *Machine Learning* (Vol. 36).

Darmawan C A. (2023). Pengembangan Aplikasi Berbasis Web Dengan *Python Flask* Untuk Klasifikasi Data Menggunakan Metode *Decisi on Tree C4.5*.

Fajar Astuti Hermawati. (2013). *Data Mining* (P. Christian, Ed.; Putri Christian). Andi.

Feldman, R., & Sanger, J. (2007). *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press.

Freund, Y., & Schapire, R. E. (1997). *A Decision-Theoritic Generalization On-Line Learning and an Application to Boosting* (Vol. 55).

Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). *Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity* (Vol. 2, Issue 1). http://j-ptiik.ub.ac.id

Komang N et al. (2018). Seleksi Fitur Bobot Kata dengan Metode TF-IDF untuk Ringkasan Bahasa Indonesia. *MERPATI*, *6*(2).

Listiana E, & Muslim A M. (2017). Penerapan *Adaboost* Untuk Klasifikasi *Support Vector Machine* Guna Meningkatkan Akurasi Pada *Diagnosa Chronic Kidney Disease*.

Nur Rozi, F., & Harini Sulistyawati, D. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan metode *Modified K-Nearest* Neighbor Dan Pembobotan Menggunakan *TF-IDF* (Vol. 15, Issue 1).

Nurrohmah H. (2022). Klasifikasi Berita Hoax Berbahasa Indonesia Menggunakan *Bidirectional Long Short Term Memory (Bi-LSTM)*.

Qureshi, K. A., Malick, R. A. S., Sabih, M., & Cherifi, H. (2021). Complex Network and Source Inspired COVID-19 Fake News Classification on Twitter. *IEEE Access*, *9*, 139636–139656. https://doi.org/10.1109/ACCESS.2021.3119404

Roy, A., Basak, K., Ekbal, A., & Bhattacharyya, P. (2018). *A Deep Ensemble Framework for Fake News Detection and Classification*. http://arxiv.org/abs/1811.04670

Sriyano, C. S., & Setiawan, E. B. (2021). Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF.

Trivusi. (2023, July 12). *Algoritma AdaBoost: Pengertian, Cara Kerja, dan Kegunaannya*. Https://Www.Trivusi.Web.Id/2023/07/Algoritma-Adaboost.Html.

Wati, R. (2020). Penerapan Algoritma Naive Bayes Dan Particle Swarm Optiization Untuk Klasifikasi Berita Hoax Pada Media Sosial. www.bsi.ac.id

Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., & Fatmawati, D. (2022). *Algoritma Term Frequency – Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen*.

McNair, B. (2017). *Fake news: Falsehood, fabrication and fantasy in journalism* (1st ed.). Routledge. https://doi.org/10.4324/9781315142036

Trninić, D., Kuprešanin Vukelić, A., & Bokan, J. (2022). Perception of “Fake News” and Potentially Manipulative Content in Digital Media—A Generational Approach. *Societies*, 12(1), 3. https://doi.org/10.3390/soc12010003.

Julianto, Y., Setiabudi, D. H., & Rostianingsih, S. (2022). Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine. Program Studi Informatika, Universitas Kristen Petra. Surabaya, Indonesia.

Putra, M. G. L., & Putera, M. I. A. (2019). Analisis Perbandingan Metode SOAP dan REST yang Digunakan pada Framework Flask untuk Membangun Web Service. SCAN: Jurnal Teknologi Informasi dan Komunikasi, XIV(2). ISSN: 1978-0087.

Kemenkeu. (2021/Update). Judul Artikel. Diakses pada [2024], dari [https://www.djkn.kemenkeu.go.id/kanwil-lampungbengkulu/baca-artikel/13912/Selamat-Tinggal-Hoax.html]

Neapolitan, Richard E. (2004). *Foundations of Algorithms Using Java Pseudocode*. Jones & Bartlett Publishers.Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, dan Clifford Stein. (2001). Introduction to Algorithms. MIT Press.

M. Ayyanar, et al.(2022), "Predicting the Cardiac Diseases using SelectKBest Method Equipped Light Gradient Boosting Machine," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 117-122, doi: 10.1109/ICOEI53556.2022.9777224.

Utami, L. D., & Wahono, R. S. (2015). Integrasi metode *Information Gain* untuk seleksi fitur dan Adaboost untuk mengurangi bias pada analisis sentimen review restoran menggunakan algoritma *Naïve Bayes*. Journal of Intelligent Systems, 1(2), 120. ISSN 2356-3982. Retrieved from http://journal.ilmukomputer.org

Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naïve Bayes. *JAMBURA Journal of Informatics*, 6(1), 26-39. https://doi.org/10.37905/jji.v6i1.24718