

**DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA
MENGGUNAKAN ALGORITMA NAIVE BAYES
DAN SUPPORT VECTOR MACHINE**

THESIS

Oleh :
NUR ELYTA FEBRIYANTY
NIM. 220605210006



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA
MENGGUNAKAN ALGORITMA NAIVE BAYES
DAN SUPPORT VECTOR MACHINE**

THESIS

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)

Oleh :
NUR ELYTA FEBRIYANTY
NIM. 220605210006

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023

**DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA
MENGGUNAKAN ALGORITMA NAIVE BAYES
DAN SUPPORT VECTOR MACHINE**

THESIS

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
NUR ELYTA FEBRIYANTY
NIM. 220605210006**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

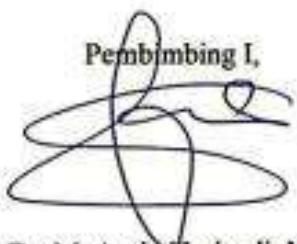
**DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA
MENGGUNAKAN ALGORITMA NAÏVE BAYES
DAN SUPPORT VECTOR MACHINE**

THESIS

Oleh :
NUR ELYTA FEBRIYANTY
NIM. 220605210006

Telah diperiksa dan disetujui untuk di uji:
Tanggal 30 November 2023

Pembimbing I,



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

Pembimbing II,


Dr. Cahyo Crysian
NIP. 19740423 200901 1 008

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



Cahyo Crysian
NIP. 19740424 200901 1 008

DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

THEESIS

Oleh :
NUR ELYTA FEBRIYANTY
NIM. 220605210006

**Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)**
Tanggal 30 November 2023

Susunan Dewan Penguji

Penguji Utama : Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001

Ketua Pengudi : Dr. Ririen Kusumawati, S.Si.,M.Kom
NIP. 19720309 200501 2 002

Sekretaris Pengujian : Dr. M. Amin Hariyadi, M.T
NIP. 19670018 200501 1 008

Anggota Pengudi : Dr. Cahyo Crysdiyan
NIP. 19740424 200901 1 008

er (M.Kom)

Tanda Tangan

()

()

()

()

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



NP 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Nur Elyta Febriyanty

NIM : 220605210006

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

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

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

Malang, 30 November 2023

Yang Membuat Pernyataan



Nur Elyta Febriyanty
NIM. 220605210006

MOTTO

Selalu berbaik sangkalah kepada Allah SWT, karena apa yang sudah ditakdirkan,
Allah SWT akan selalu bantu.

LEMBAR PERSEMPAHAN

Thesis ini akan saya persembahkan untuk kedua orang tuaku, mertua, suami dan kedua anakku. Doa dan restu yang tiada henti serta selalu membersamaai penulis selama menyelesaikan studi ini. Dan penulis tak lupa ucapan terima kasih kepada teman-teman angkatan masuk 2022/2023 ganjil atas semua waktu, support dan doa sampai penyelesaian studi ini.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Puji syukur penulis panjatkan kepada Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya sehingga saya dapat menyelesaikan thesis ini dan menyelesaikan studinya di Program Studi Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Selain itu, penulis ingin menyampaikan penghargaan, doa dan harapannya, kepada semua pihak yang telah membantu penyelesaian thesis. Penulis mengucapkan terima kasih kepada:

1. Selaku pembimbing thesis, Bapak Dr. M. Amin Hariyadi, M.T dan Dr. Cahyo Crysdiyan yang telah memberikan luasnya kesabaran dalam membimbing dan memberikan pengalaman ilmu yang sangat berharga
2. Saya mengucapkan terima kasih kepada seluruh civitas akademika yang tergabung dalam Program Studi Magister Informatika UIN Maulana Malik Ibrahim Malang, khususnya kepada seluruh dosen.
3. Orang tua, Mertua, suami dan anak-anakku tercinta yang selalu memberikan doa, restu dan menyemangati penulis dalam menempuh pendidikan,
4. Semua civitas akademik Universitas Insan Budi Utomo, khususnya Biro BAAK yang telah memberikan doa dan dukungan tiada henti kepada penulis sampai penyelesaian studi ini.
5. Semua pihak yang telah memberikan kontribusi, moral dan materiil, untuk penyelesaian thesis ini.

Penulis berharap untuk thesis ini dapat memberikan manfaat bagi para pembaca, khususnya penulis sendiri, meskipun masih terdapat kekurangan dalam proses penyusunannya. Amin ya, Rabbal Alamin.

Wassalamu'alaikum, Wr. Wb.

Malang,
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
ABSTRAK	xiv
ABSTRACT	xv
خلاصة.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Identifikasi Berita Hoax	5
2.2 Kerangka Teori.....	7
BAB III METODOLOGI PENELITIAN.....	9
3.1 Prosedur Penelitian.....	9
3.1.1 Pengumpulan Data	9
3.1.2 Data Engineering.....	10
3.2 Perancangan Sistem.....	17
3.3. Eksperiment.....	18
BAB IV METODE NAÏVE BAYES	27
4.1 Desain Metode Naïve Bayes	27
4.2 Implementasi Metode Naïve Bayes.....	29
BAB V METODE SUPPORT VECTOR MACHINE.....	44
5.1 Desain Metode Support Vector Machine	44
5.2 Implementasi Metode Support Vector Machine	52
BAB VI PEMBAHASAN	68
6.1 Pengujian Metode Naïve Bayes	69

6.2 Pengujian Metode Support Vector Machine	72
7.3 Hasil Perbandingan Metode Naïve Bayes dan Support Vector Machine.....	75
BAB VII KESIMPULAN	79
8.1 Kesimpulan	79
8.2 Saran.....	79
DAFTAR PUSTAKA.....	80

DAFTAR TABEL

Tabel 3.1 Jumlah Dataset	9
Tabel 3.2 Pelabelan Dataset.	10
Tabel 4.1 Pembagian Data Latih dan Data Uji.....	33
Tabel 4.2 Perbandingan Data	34
Tabel 4.3 Pengujian 50% : 50%	34
Tabel 4.4 Hasil Perhitungan 50% : 50%	36
Tabel 4.5 Pengujian 60% : 40%	36
Tabel 4.6 Hasil Perhitungan 60% : 40%	37
Tabel 4.7 Pengujian 70% : 30%	38
Tabel 4.8 Hasil Perhitungan 70% : 30%	39
Tabel 4.9 Pengujian 80% : 20%	40
Tabel 4.10 Hasil Perhitungan 80% : 20%	41
Tabel 4.11 Pengujian 90% : 10%	41
Tabel 4.12 Hasil Perhitungan 90% : 10%	43
Tabel 4.13 Nilai Akurasi, Presisi, Recall dan F1 Score Metode Naïve Bayes	43
Tabel 5.1 Sample Data.....	45
Tabel 5.2 Pembobotan Sample Data	46
Tabel 5.3 Bobot Kata	49
Tabel 5.4 Perhitungan Hyperplane.....	49
Tabel 5.5 Pembagian Data Latih dan Data Uji.....	56
Tabel 5.6 Perbandingan Data	56
Tabel 5.7 Pengujian 50% : 50%	57
Tabel 5.9 Hasil Perhitungan 50% : 50%	58
Tabel 5.10 Pengujian 60% : 40%	59
Tabel 5.11 Hasil Perhitungan 60% : 40%	60
Tabel 5.12 Pengujian 70% : 30%	61
Tabel 5.13 Hasil Perhitungan 70% : 30%	62
Tabel 5.14 Pengujian 80% : 20%	63
Tabel 5.15 Hasil Perhitungan 80% : 20%	64
Tabel 5.16 Pengujian 90% : 10%	65
Tabel 6.17 Hasil Perhitungan 90% : 10%	66
Tabel 5.18 Nilai Akurasi, Presisi, Recall dan F1 Score Metode SVM.....	67
Tabel 6.1 Perbandingan Data Latih dan Data Uji	68
Tabel 6.2 Data Latih dan Data Uji	69
Tabel 6.3 Hasil Perbandingan 70%:30%	70
Tabel 6.4 Hasil Perbandingan 90%:10%	70
Tabel 6.5 Hasil Pengujian Naïve Bayes	71

Tabel 6.6 Hasil Perbandingan 70%:30%.....	73
Tabel 6.7 Hasil Perbandingan 90%:10%.....	73
Tabel 6.8 Hasil Pengujian Support Vector Machine	74
Tabel 6.9 Hasil Keseluruhan Metode.....	75

DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori.....	8
Gambar 3.1 Prosedur Penelitian.....	9
Gambar 3.2 Persiapan Data	10
Gambar 3.3 Alur Case Folding	11
Gambar 3.4 Proses Lowercase	12
Gambar 3.5 Alur Stemming	13
Gambar 3.6 Proses Stemming	13
Gambar 3.7 Alur Stopword Removal.....	14
Gambar 3.8 Proses Stopword Removal.....	15
Gambar 3.9 Proses Tokenazation.....	16
Gambar 3.10 Rancangan Sistem	17
Gambar 3.11 Ujicoba Sistem	18
Gambar 3.12 Proses Pembobotan TF IDF	21
Gambar 3.13 Implementasi TF IDF	22
Gambar 3.14 Alur Algoritma Naïve Bayes	24
Gambar 3.15 Alur Algoritma Support Vector Machine.....	26
Gambar 4.1 Tahap Metode Naïve Bayes	27
Gambar 4.2 Input Judul Berita	29
Gambar 4.3 Preprocessing.....	30
Gambar 4.4 Proses TF IDF	31
Gambar 4.5 Prediksi Metode Naïve Bayes.	31
Gambar 5.1 Alur Algoritma Support Vector Machine.....	44
Gambar 5.2 Visualisasi SVM.....	51
Gambar 5.3 Input Judul Berita	52
Gambar 5.4 Preprocessing.....	53
Gambar 5.5 Proses TF IDF.	54
Gambar 5.6 Prediksi Metode Support Vector Machine	55
Gambar 6.1 Hasil Keseluruhan Pengujian.	76

ABSTRAK

Febriyanty, Nur Elyta. 2023. **Deteksi Berita Hoax Dari Media Online Indonesia Menggunakan Algoritma Naïve Bayes dan Support Vector Machine.** Thesis Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, MT., (II) Dr. Cahyo Crysdiann.

Kata Kunci : Deteksi Berita *Hoax*, Naive Bayes, Support Vector Machine

Website dan blog terkenal sebagai media penyiaran berita di berbagai bidang seperti penyiaran berita. Validitas artikel berita bisa valid atau palsu. Berita palsu disebut juga dengan berita *Hoax*. Tujuan pembuatan berita *Hoax* adalah untuk membujuk, memanipulasi, dan mempengaruhi pembaca berita agar melakukan hal-hal yang bertentangan atau menghalangi tindakan yang benar. Penelitian ini mengusulkan untuk melakukan eksperimen klasifikasi Support Vector Machine dan Naïve Bayes untuk mendeteksi berita *Hoax* berbahasa Indonesia. Penelitian ini menggunakan dataset dari data publik yaitu berita antara berita Valid dan *Hoax*. Sistem dapat mengklasifikasikan berita online berbahasa Indonesia dengan istilah frekuensi fitur algoritma Support Machine Vector dan klasifikasi Naïve Bayes. Sedangkan model evaluasi yang digunakan adalah *Confusion Matrix*. Hasil perbandingan kedua model sebagai Support Vector Machine mempunya hasil rata – rata *Naive Bayes* Akurasi 93%, Presisi 100%, *Recall* 93% dan *F1 Score* 93%, hasil rata – rata metode *Support Vector Machine* Akurasi 100%, Presisi 99%, *Recall* 100% dan *F1 Score* 91%. Oleh karena itu untuk klasifikasi berita *Hoax* kami merekomendasikan model Naive Bayes karena memiliki F1 score yang lebih baik dibandingkan Support Vector Machine.

ABSTRACT

Febriyanty, Nur Elyta. 2023. **Hoax News Detection From Indonesian Online Media Using Naïve Bayes Algorithm and Support Vector Machine.** Thesis of Master's Degree Program in Informatics, Faculty of Science and Technology, Islamic State University Maulana Malik Ibrahim Malang. Supervisor: (I) Dr. M. Amin Hariyadi, MT., (II) Dr. Cahyo Crysdiyan

Keywords : Detection Hoax News, Naive Bayes, Support Vector Machine

Websites and blogs are well-known as media for broadcasting news in various fields such as broadcasting news. The validity of news articles can be valid or fake. Fake news is also known as hoax news. The purpose of making hoax news is to persuade, manipulate, and influence news readers to do things that contradict or prevent correct action. This study proposes to experiment with the Support Vector Machine and Naïve Bayes classifications to detect hoax news in Indonesian. This study uses a dataset from public data, namely news between valid news and hoaxes. The system can classify online news in Indonesian with the term frequency feature the machine vector Support algorithm and naïve Bayes classification. The results of the comparison of the two models as a Support Vector Machine have an average result of Naïve Bayes accuracy of 93%, the precision of 100%, recall of 93%, and F1 score of 93%, the average results of the Support Vector Machine method are 100% Accuracy, Precision 99%, Recall 100 % and F1-Score 91%. Therefore, for the classification of hoax news, we recommend the Naive Bayes model because it has a better level of accuracy than the Support Vector Machine.

خلاصة

فيريانتي، نور إلينا. ٢٠٢٣. الكشف عن الأخبار الخادعة من وسائل الإعلام الإندونيسية عبر الإنترن特 باستخدام خوارزمية نايف بايز وآلة التوجهات الدعمة. الأطروحة. برنامج دراسة المعلوماتية كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (أ) الدكتور محمد أمين هاريادي، الماجستير (ب) الدكتور جاهيو كريسيديان.

الكلمات الرئيسية: كشف الأخبار الخادعة، نايف بايز، سوببورت فيكتور مشين

تعرف موقع الويب والمدونات بأنها وسائل بث إخبارية في مجالات مختلفة مثل بث الأخبار. صحة المقالات الإخبارية يمكن أن تكون صحيحة أو مزيفة. وتسمى الأخبار المزيفة أيضاً بالأخبار الخادعة (هاوكس). والغرض من إنشاء أخبار خادعة هو إقناع قراء الأخبار والتلاعب بهم والتأثير عليهم للقيام بأشياء تتعارض مع الإجراءات الصحيحة أو تعيقها. يقترح هذا البحث إجراء تجربة لكشف عن الأخبار الخادعة باللغة (Naïve Bayes) تصنف آلة التوجهات الداعمة (سوببورت فيكتور مشين) و نايف بايز الإندونيسية. يستخدم هذا البحث مجموعة بيانات من البيانات العامة، وهي الأخبار بين الأخبار الصحيحة والأخبار الخادعة. يمكن للنظام تصنيف الأخبار عبر الإنترنوت باللغة الإندونيسية باستخدام تردد ميزة خوارزمية آلة التوجهات الداعمة وتصنيف نايف بايز نتائج المقارنة بين النموذجين كلة التوجهات (*Confusion Matrix*). وأما نموذج التقييم المستخدم فهو كونفوشيون ماتريكس الداعمة لديه متوسط نايف بايز دقة ٩٣٪، وبريسبي ١٠٠٪، وريكول ٩٣٪، وف١-سكور ٩٣٪، ومتوسط نتائج نموذج آلة التوجهات الداعمة دقة ١٠٠٪، وبريسبي ٩٩٪، وريكول ١٠٠٪، وف١-سكور ٩١٪. لذلك، بالنسبة لتصنيف الأخبار الخادعة، نوصي بنموذج نايف بايز لأنه يتمتع بف١-سكور أفضل من آلة التوجهات الداعمة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Persebaran informasi melalui internet saat ini sangat cepat dan mudah, tanpa terbatas oleh waktu dan jarak. Dengan jumlah pengguna internet di Asia mencapai 212,35 juta jiwa pada Maret 2021, Indonesia menempati peringkat ketiga dalam hal jumlah pengguna internet di kawasan ini. Peningkatan pengguna internet memungkinkan masyarakat mengakses informasi dengan cepat, namun juga membawa dampak negatif seperti penyebaran informasi palsu atau *Hoax*.

Hoax adalah informasi atau berita yang belum terverifikasi dan bukan fakta yang sebenarnya, sering kali dibuat dan disebar dengan tujuan pribadi atau untuk mendapatkan keuntungan. *Hoax* cenderung memiliki judul yang sensasional dan provokatif untuk menarik perhatian pembaca. Sebanyak 30% hingga hampir 60% masyarakat Indonesia mengalami paparan terhadap *Hoax* ketika berinteraksi online, sementara hanya 21% hingga 36% yang mampu mengenali atau mendeteksi keberadaan *Hoax*. Isu-isu terkait politik, kesehatan, dan pendidikan sering menjadi subjek utama dari *Hoax* yang ditemukan. Dampaknya mencakup keraguan, kebingungan, dan potensi merusak citra individu dan kelompok.

Teks klasifikasi mengidentifikasi pola atau informasi baru dengan memanfaatkan metode *Machine Learning*, proses mengubah teks yang tidak terstruktur menjadi format terstruktur. Ada berbagai algoritma yang dapat digunakan dalam teks klasifikasi, termasuk *Decision Tree*, *K-Means Clustering*, *Naïve Bayes Classifier*, *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan sebagainya. Dalam penelitian ini, peneliti memilih menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM). Pemilihan SVM disebabkan oleh kinerja yang baik ketika terdapat margin pemisahan yang jelas antar kelas dan penggunaan memori yang efisien.

Di dalam al - Qur'an jelas diterangkan bahwa berita bohong atau *Hoax* adalah modal orang-orang munafik untuk merealisasikan niat jahat kepada umat muslim. Sebagaimana dijelaskan dalam Al Qur'an surah A Nur 15:20

إِذْ تَقُولُنَّ بِالسِّتْكِ وَتَقُولُنَّ بِأَفْوَاهِكُمْ مَا لَيْسَ لَكُمْ بِهِ عِلْمٌ وَتَحْسِبُونَهُ هَيْئًا وَهُوَ عِنْدَ اللَّهِ عَظِيمٌ ١٥

وَلَوْلَا إِذْ سَمِعْتُمُوهُ قُلْتُمْ مَا يَكُونُ لَنَا أَنْ تَنْكَلَمْ بِهَذَا سِبْحَانَكَ هَذَا بُهْتَانٌ عَظِيمٌ ١٦

يَعْظُكُمُ اللَّهُ أَنْ تَغُرُّوا لِمَثْلَهِ أَبْدًا إِنْ كُنْتُمْ مُؤْمِنِينَ ١٧

وَبَيْنَ أَنَّ اللَّهَ لَكُمُ الْأَيَاتِ وَأَنَّ اللَّهَ عَلِيمٌ حَكِيمٌ ١٨

إِنَّ الَّذِينَ يُجْبِيْنَ أَنْ تَشْيِعَ الْفَحْشَةَ فِي الْأَدْنِيَاءِ إِنَّمَّا لَهُمْ عَذَابٌ لِمَا فِي الْأَدْنِيَاءِ وَالْآخِرَةِ وَاللَّهُ يَعْلَمُ

وَأَنْتُمْ لَا تَعْلَمُونَ ١٩

وَلَوْلَا فَضْلُ اللَّهِ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ رَغْفَةٌ رَّحِيمٌ ٢٠

"Ketika kamu menyebarkannya dari satu lidah ke lidah yang lain, dan mengucapkan dengan mulutmu apa yang tidak kamu ketahui, menganggapnya enteng padahal itu 'sangat' serius di sisi Allah(15). Andai saja ketika mendengarnya Anda berkata, "Bagaimana kita bisa membicarakan hal seperti itu!" Maha Suci Engkau 'Ya Tuhan'! Ini adalah fitnah yang keji!(16). Allah melarang Anda melakukan hal seperti ini lagi, jika Anda benar-benar beriman (17). Dan Allah menerangkan perintah-Nya kepadamu, sebab Allah Maha Mengetahui lagi Maha Bijaksana(18). Sesungguhnya orang-orang yang senang melihat kecabulan tersebar di kalangan orang-orang beriman, akan mendapat siksa yang pedih di dunia dan di akhirat. Allah mengetahui dan kamu tidak mengetahui (19). Kamu pasti akan menderita, kalau saja bukan karena karunia dan rahmat Allah kepadamu, dan seandainya Allah tidak Maha Pemurah lagi Maha Penyayang(20).

Dalam tafsir Departemen Agama RI Jilid 10 tahun 2008, disebutkan bahwa kata "Buhtān" merujuk pada kebohongan yang sangat besar. Asal usulnya berasal dari "buhita," yang menggambarkan perasaan tercengang dan bingung tanpa tahu apa yang harus dilakukan. Penyebaran gosip dianggap sebagai buhtān karena merupakan ucapan yang disengaja, tanpa alasan atau bukti, dan terkait dengan kehormatan manusia, termasuk rumah tangga Nabi yang dianggap sebagai pilihan Allah.

Pada ayat sebelumnya, Allah mencela munculnya fitnah di kalangan Muslim dan menjelaskan konsekuensi bagi penuduh palsu. Ayat-ayat berikutnya menjelaskan cara penyebaran fitnah, cara menghentikannya, dan konsekuensi

yang akan dialami oleh pelaku penyebaran fitnah. Tafsir ayat (15) menjelaskan bahwa tanpa karunia dan rahmat Allah, penyebar berita bohong akan mendapat azab. Penyebaran berita bohong dapat terjadi melalui percakapan mulut ke mulut, penyebarluasan tanpa pengetahuan yang jelas, dan menganggap enteng berita bohong.

Ayat (16) menyayangkan sikap sebagian kaum Muslim yang tidak menghentikan pembicaraan fitnah terkait rumah tangga Nabi. Allah memperingatkan agar tidak mengulangi perbuatan jahat tersebut jika mereka beriman. Ayat (18) menjelaskan bahwa Allah telah merinci syariat, akhlak, dan adab-Nya dalam kitab-Nya. Dia Maha Mengetahui segala sesuatu dan membala sesuai perbuatan baik atau jahat.

Ayat (19) menegaskan bahwa orang yang menyebarkan perbuatan keji seperti zina akan mendapat hukuman di dunia dan akhirat, kecuali jika mereka bertaubat. Allah Maha Mengetahui hal-hal yang tersembunyi. Ayat (20) menjelaskan bahwa larangan terhadap pornografi dan zina merupakan tindakan kasih sayang Allah terhadap umat manusia. Allah memberikan karunia kepada penyebar berita bohong, namun Dia senantiasa berbuat baik kepada hamba-Nya yang mendatangkan manfaat bagi mereka, meskipun mereka melakukan pelanggaran. Dari kutipan ayat-ayat tersebut dapat diketahui bahwa fitnah atau berita bohong tersebar tanpa adanya validitas berita. Berita bohong atau fitnah dapat dicegah dengan tidak menyebarkan berita yang tidak berdasar.

Untuk mengidentifikasi kebenaran suatu informasi atau berita, terutama apakah berisi *Hoax* atau Valid, diperlukan alat bantu yang mampu melakukan deteksi tersebut. Salah satu pendekatan yang digunakan untuk mengurangi penyebaran berita berisi *Hoax* adalah metode klasifikasi teks. Penelitian ini bertujuan untuk melakukan eksperimen pada berbagai metode klasifikasi teks guna menentukan yang paling optimal dalam menganalisis berita *Hoax*.

1.2 Pernyataan Masalah

Bagaimana mengidentifikasi berita hoax dengan menggunakan pendekatan metode klasifikasi serta berapa tingkat akurasinya?

1.3 Tujuan Penelitian

Penelitian bertujuan untuk mengevaluasi kinerja metode klasifikasi dalam mengidentifikasi berita *Hoax*, serta membandingkan rata-rata akurasi dari kedua metode tersebut.

1.4 Batasan Masalah

1. Berita *online* yang dianalisa dalam bahasa Indonesia
2. Menggunakan dataset yang dikumpulkan dalam periode Januari 2020 – Juli 2023.
3. Dataset berita *Hoax* menggunakan berita dari platform media sosial dari sumber berita tertentu.

1.5 Manfaat Penelitian

Output penelitian dapat dimanfaatkan oleh :

1. Penelitian ini diharapkan dapat memberikan kontribusi positif untuk pengembangan penyedia layanan berita publik di lingkungan KOMINFO.
2. Penelitian ini diharapkan bisa memberikan sumbangan penting sebagai masukan untuk pemerintah dalam mengembangkan aspek-aspek yang terkait dengan kebijakan keterbukaan informasi publik.
3. Penelitian ini diharapkan memberi manfaat dalam edukasi masyarakat, mempermudah proses verifikasi dalam mencegah penyebaran berita *Hoax*.

BAB II

TINJAUAN PUSTAKA

2.1 Identifikasi Berita *Hoax*

Penyebaran berita palsu atau *Hoax* secara luas dapat berdampak negatif yang signifikan terhadap individu dan masyarakat. Perkembangan internet menjadi salah satu faktor penyebab terjadinya ledakan informasi Pratiwi et al., (2022). Beberapa jurnal membahas permasalahan deteksi berita *Hoax*, diantaranya yaitu

Penelitian teks klasifikasi dari Abid (2020) mengenai korpus berita yang terdiri dari berita kehilangan manusia dari berbagai kategori telah dikembangkan dengan mengumpulkan data dari berbagai sumber situs berita terkenal dan otentik. Korpus itu juga memuat sejumlah berita rangkap. Kedua membandingkan pendekatan klasifikasi yang berbeda secara empiris menemukan pengklasifikasi teks yang paling cocok untuk kategorisasi berbagai sub-kategori berita, dan ketiga metode telah diusulkan dan dibandingkan untuk mendeteksi duplikat berita dari korpus dengan melibatkan berbagai teknik pra-pemrosesan dan ukuran kesamaan yang banyak digunakan, kesamaan kosinus, dan koefisien Jaccard. Hasilnya menunjukkan bahwa pengklasifikasi teks konvensional masih relevan dan bekerja dengan baik dalam tugas klasifikasi teks karena *Multinomial Naive B* telah memberikan hasil Akurasi 89,5%. Sedangkan koefisien Jaccard menunjukkan hasil yang jauh lebih baik daripada *Cosine Similarity* untuk pendekripsi berita duplikat dengan variasi *Pre-processing* yang berbeda dengan Akurasi rata-rata 83,16%.

Arsitektur *Hybrid Neural Network*, yang menggabungkan kemampuan CNN dan LSTM diimplementasikan dalam penelitian Umer et al., (2020) dengan menggunakan dua pendekatan reduksi dimensi yang berbeda, Prinsip Analisis Komponen (PCA) dan Chi-Square. Penelitian ini mengusulkan untuk menggunakan teknik reduksi dimensi untuk mengurangi dimensi vektor fitur sebelum meneruskannya ke pengklasifikasi. Untuk mengembangkan penalaran, studi ini memperoleh dataset dari situs web *Fake News Challenges* (FNC) yang memiliki empat jenis sikap, setuju, tidak setuju, berdiskusi, dan tidak terkait. Fitur

nonlinear diumpulkan ke PCA dan Chi-Square yang menyediakan lebih banyak fitur kontekstual untuk deteksi berita palsu. Motivasi dari penelitian ini adalah untuk menentukan sikap relatif sebuah artikel berita terhadap judulnya. Model yang diusulkan meningkatkan hasil sebesar ~ 4% dan ~ 20% dalam hal Akurasi dan *F1-Score*. Hasil percobaan menunjukkan bahwa PCA mengungguli *Chi-Square* dan metode canggih dengan Akurasi 97.8%.

Kausika et al., (2021) mempelajari kemungkinan penggunaan teknik *Deep Learning* untuk mendiskriminasi berita palsu di Internet hanya dengan menggunakan teksnya. Dalam penelitian ini aplikasi untuk mendeteksi 'berita palsu', yang merupakan berita menyesatkan dari sumber terpercaya dengan metode NLP (*Natural Language Processing*). Pendekatan ini telah diimplementasikan dan diperiksa dalam bentuk sistem perangkat lunak. Pendekatan pendekripsi berita palsu dengan SVM mencapai Akurasi 92% dan *Naive Bayes* mencapai Akurasi 73%. Dalam penelitian ini SVM dapat lebih baik daripada model pengklasifikasi *Naive Bayes* dalam pendekatan prediksi baru.

Dalam penelitian P. Meel dan DK Vishwakarma (2021) merancang kerangka kerja berdasarkan lima pengklasifikasi pembelajaran mesin yang diawasi secara independen, *Support Vector Machine*, *K-Nearest Neighbor*, *Logistic Regression*, *Naïve Bayes*, dan *Random Forest* untuk penilaian kepercayaan konten informasi web. Pengklasifikasi data dilatih dan diuji pada dua set data yang berbeda. Deteksi Berita Palsu (Jruvika/FND) dan Berita nyata atau palsu yang berisi artikel berita lengkap dalam bentuk tajuk utama dan isi. Eksperimen dan analisis hasil memverifikasi bahwa Akurasi tertinggi dicapai oleh metode yang diproyeksikan adalah 96,61% pada dataset deteksi berita palsu menggunakan pengklasifikasi *Support Vector Machine* (SVM).

Ferdiansyah dan Wella (2022) mengatasi masalah *Hoax* di internet dengan menguji deteksi menggunakan Algoritma *Support Vector Machine*, *Naive Bayes*, *Random Forest*, dan *K-Nearest Neighbor*. Penelitian menggunakan 1000 berita Indonesia, 500 non-*Hoax*, dan 500 berita *Hoax* sebagai data pelatihan, dengan tambahan 500 data dari Twitter. Data diperoleh dari portal berita online Indonesia dengan metode text mining. Hasilnya menunjukkan Akurasi tertinggi 89,27%

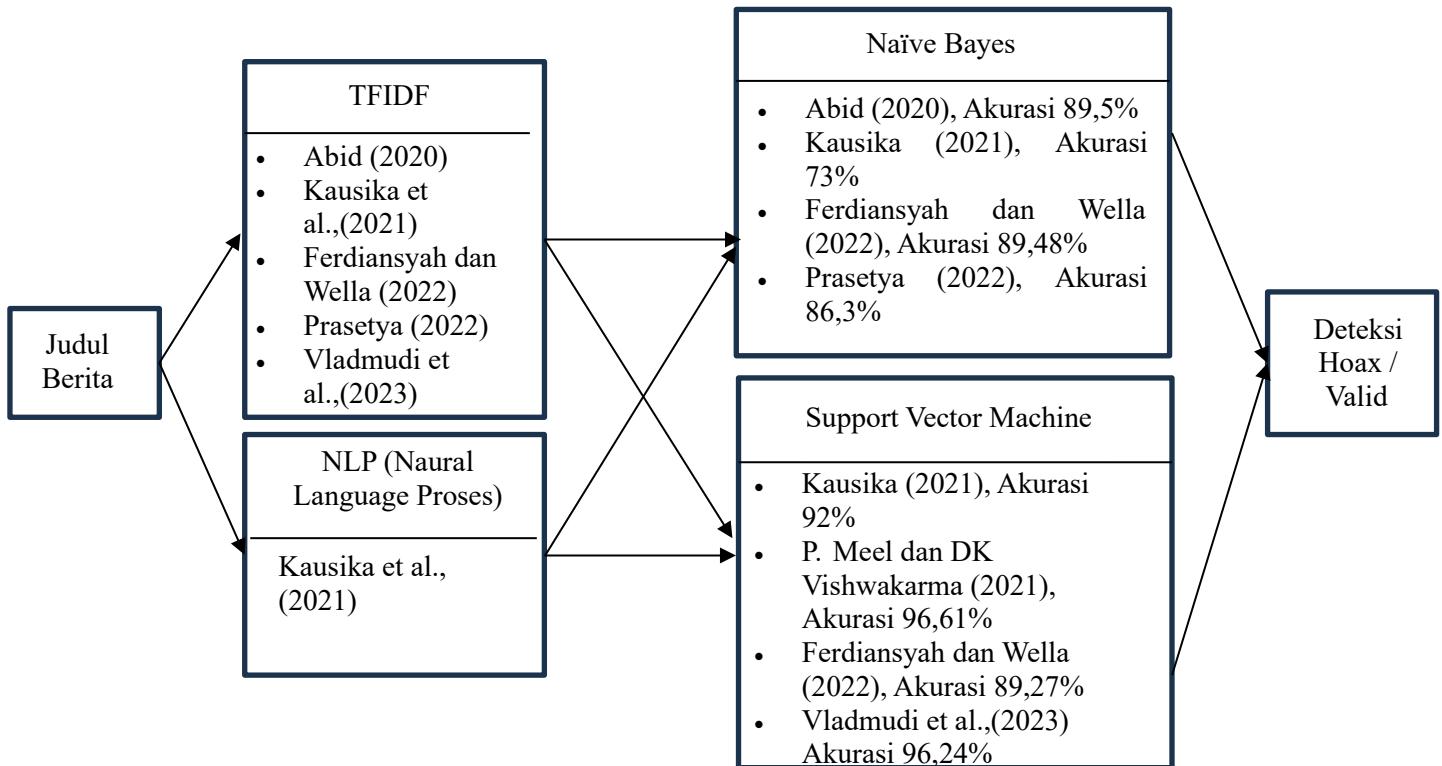
dengan SVM dan 89,48% dengan *Naïve Bayes*. Pemodelan ini mendukung deteksi berita *Hoax* di Indonesia dan dapat diterapkan oleh pemerintah untuk menyaring informasi kepada masyarakat.

Prasetya (2022) melakukan eksperimen pada algoritma klasifikasi *Naïve Bayes* untuk mengklasifikasikan berita *Hoax* Covid-19. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* dan Cross Validation dapat melakukan klasifikasi berita *Hoax* dengan baik, mencapai Akurasi 86,3%, yang termasuk dalam kriteria klasifikasi baik (80-90%). Jumlah prediksi yang salah relatif kecil, hanya 41 dari total 300 dataset, sehingga model ini dapat dijadikan referensi untuk pengembangan model prediksi yang lebih kompleks.

Penelitian lain Vadlamudi et al., (2023) metode klasifikasi *Naïve Bayes*, dan penelitian yang relevan untuk mengidentifikasi berita palsu. *Naïve Bayes Classifier* dan SVM dengan linear kernel trick adalah dua metode populer yang digunakan dalam percobaan ini sebagai bagian dari pendekatan *Hybrid*. Ukuran sampel untuk setiap Classifier adalah 41. Hasilnya, tingkat Akurasi SVM menjadi 96,24% lebih tinggi dibandingkan tingkat Akurasi *NB Classifier* sebesar 87,80%. Tidak ada perbedaan yang signifikan secara statistik antara kedua kelompok, menurut analisis statistik dan independent sample T-test dengan nilai $p=0,433$ ($p>0,05$). Teknik Hibrid berdasarkan mesin vektor dukungan dengan trik kernel linier mengungguli pengklasifikasi *Naïve Bayes*. URL tingkat Akurasi *Support Vector Machine* adalah 96,24% lebih tinggi dari tingkat Akurasi *Naïve Bayes Classifier* sebesar 87,80%.

2.2 Kerangka Teori

Dalam kerangka kerja ini, peneliti menyajikan dimensi sistem informasi utama yang disesuaikan dengan sistem. Kerangka Teoritis diberikan pada Gambar 2.1



Gambar 2.1 Kerangka Teori

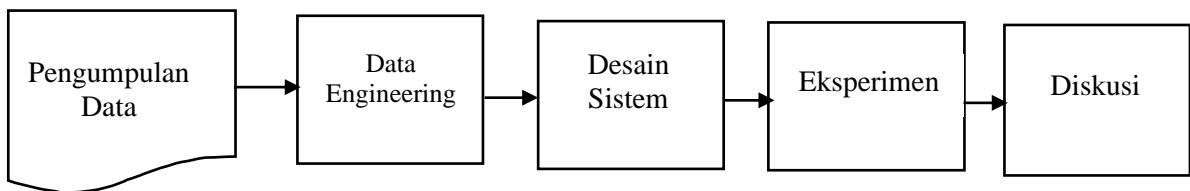
Dalam melakukan analisa dan menentukan metode Akurasi tinggi pada penelitian ini, peneliti mengklasifikasi pada Gambari 2.1 diperoleh hasil Akurasi tertinggi yaitu *Support Vector Machine* dengan nilai 96,61% dan metode yang paling banyak digunakan adalah Naïve Bayes 89,27%. Kedua metode ini akan digunakan sebagai rujukan dalam penelitian mendatang.

BAB III

DESAIN PENELITIAN

3.1. Prosedur Penelitian

Tahap-tahap dalam penelitian ini melibatkan serangkaian langkah, mulai dari konsepsi hingga implementasi, yang dirancang untuk mencapai hasil optimal melalui proses yang sistematis. Kerangka kerja dalam penelitian ini diusulkan diilustrasikan pada Gambar 3.1



Gambar 3.1 Prosedur Penelitian

3.1.1 Pengumpulan Data

Sumber data penelitian ini berupa data kualitatif, yaitu data yang memberikan informasi tentang atribut atau kategori dari suatu informasi. Data diambil dari kumpulan data publik yang dapat diakses oleh masyarakat umum.. Peneliti mengumpulkan dataset berita Valid secara manual melalui mesin pencarian dari portal berita dengan teknik *Scraping* data Kominfo, Tribunnews, Detik.com, Kompas.com dan CNN.com. Dataset diperoleh dari berita yang muncul selama periode Januari 2020 – Juli 2023 sejumlah 2000 judul berita. Jumlah pembagian dataset pada penelitian ditunjukkan pada Tabel 3.1

Tabel 3.1 Jumlah Dataset

Kategori	Data Latih	Data Uji
Valid	458 Text	256 Text
<i>Hoax</i>	1452 Text	144 Text
Total	2.000 Text	400 Text

Data berupa judul dari berita, setiap berita dikelompokan dengan label berita Valid dan berita *Hoax*. Pelabelan dalam contoh sampel data yang akan digunakan yang di sajikan dalam Tabel 3.2

Tabel 3.2 Pelabelan Dataset

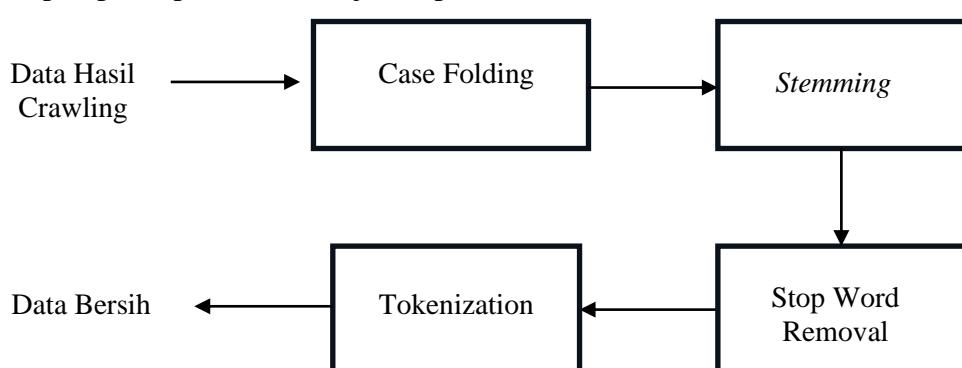
Judul Berita	Kategori	Sumber Data	Kategori Berita
Warga Australia Disarankan Tetap Mengenakan Masker Hadapi Penyebaran Flu di Musim Dingin	Valid	Detik.com	Health News
Naik KA. Tetap Wajib Booster dan Pakai Masker	Valid	Tempo.co.id	Social
Presiden Jokowi Sapa Masyarakat Indonesia Hingga Bertemu Kanselir Jerman	Valid	Kompas.com	Politik
Berdamai Dengan Corona yang Dimaksud Jokowi adalah Pemerintah telah Kalah dan Menyerah	<i>Hoax</i>	<i>Hoax</i> Dataset	Health News
Penerima Stiker Pada Pesan Whatsapp Dikenakan Sejumlah Biaya	<i>Hoax</i>	<i>Hoax</i> Dataset	Business
Mesin Mobil Sering Mati Diatas Rel Kereta Api	<i>Hoax</i>	<i>Hoax</i> Dataset	Travel

Peneliti mengumpulkan dataset dari portal berita resmi Indonesia sebagai referensi untuk mengukur validitas sebuah berita berdasarkan artikel yang diterbitkan di situs-situs berita tersebut. Penandaan dilakukan secara manual dengan memeriksa keterangan dan informasi yang terdapat pada sumber berita.

3.1.2 *Data Engineering*

Data Engineering atau persiapan data merupakan langkah penting dalam klasifikasi berita karena mempengaruhi kualitas dan performa model klasifikasi. Setiap dataset berita pelatihan dan pengujian melewati proses pra-pemrosesan. Langkah dilakukan untuk memperoleh model yang diinginkan.

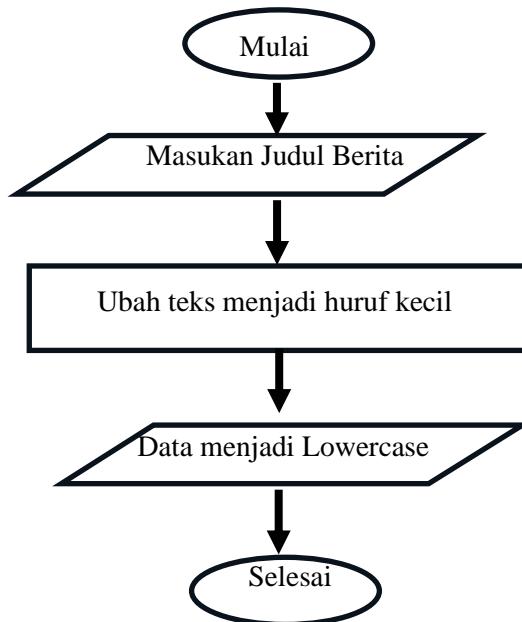
Tahapan persiapan data ditunjukkan pada Gambar 3.2



Gambar 3.2 Persiapan Data

Beberapa tahap dalam *Preprocessing* atau persiapan data yang dilakukan proses teks *Preprocessing* yaitu proses :

1. *Case folding*, *Preprocessing* data teks yang melibatkan konversi seluruh karakter huruf dalam teks menjadi huruf kecil (*lowercase*). Tujuan dari proses ini adalah untuk mengubah teks menjadi bentuk yang seragam dalam hal kapitalisasi huruf sehingga mempermudah analisis teks. Selanjutnya alur *Case Folding* ditunjukan pada Gambar 3.3



Gambar 3.3 Alur *Case Folding*

Dengan mengabaikan perbedaan huruf besar dan kecil, proses ini membuat teks lebih konsisten dan mudah diolah dalam berbagai operasi pemrosesan bahasa alami dan analisis teks. *Case Folding* membantu mengurangi ambiguitas dalam operasi pencarian dan pembandingan teks. Ketika mencari kata kunci dalam teks atau membandingkan dua string, memastikan bahwa perbedaan huruf besar dan kecil diabaikan dapat meningkatkan akurasi dan konsistensi. Implementasi sistem dari proses *Case Folding* ditunjukan pada Gambar 3.4

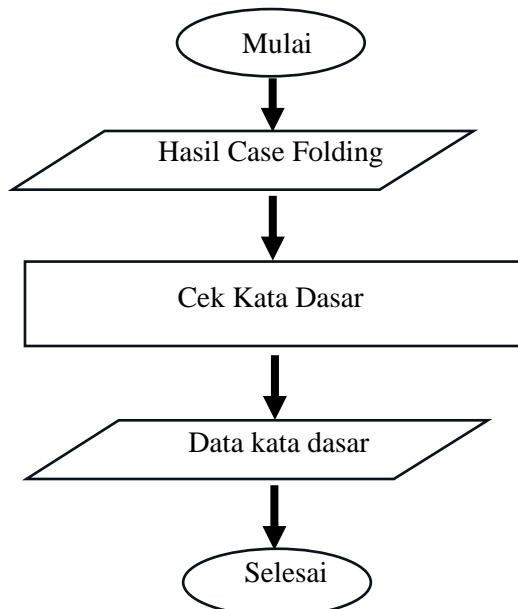
```
# Membaca file CSV dan menyimpannya ke dalam dataframe
dataframe = pd.read_csv("csv_data_source/" + sys.argv[1] +
".csv", on_bad_lines = 'skip')
dataframe['Judul'].str.lower()
```

Gambar 3.4 Proses *Lowercase*

Dalam Gambar tersebut menggunakan Pandas dan kemudian melakukan transformasi pada kolom 'Judul' untuk mengonversinya menjadi huruf kecil (*Lowercase*). Fungsi pd.read_csv() dari Pandas digunakan untuk membaca data dari file CSV. sys.argv[1] digunakan untuk mengambil argumen dari baris perintah yang dijalankan. Fungsi ini seharusnya berisi nama file CSV yang ingin dibaca. str.lower(), Metode ini digunakan untuk mengonversi semua teks dalam kolom 'Judul' menjadi huruf kecil yang merupakan tahap *Case Folding* dan selanjutnya dilakukan proses *Stemming*.

4.2 *Stemming*

Tahap *Stemming* digunakan untuk mengubah kata dalam dokumen menjadi kata dasar, dengan menghilangkan imbuhan pada kata. Tujuan dari proses *Stemming* adalah menghasilkan bentuk dasar atau akar kata, untuk menyederhanakan kata-kata menjadi bentuk dasarnya sehingga kata-kata yang memiliki akar kata yang sama dianggap setara. Langkah ini membantu dalam analisis teks dan efisiensi informasi. Penerapannya menggunakan *library Sastrawi*, dan *Flowchart Stemming* dapat dilihat pada Gambar 3.5

Gambar 3.5 Alur *Stemming*

Stemming membantu mengatasi variasi morfologi dalam bahasa. Bahasa seringkali memiliki variasi kata seperti pembentukan kata berimbahan, pembentukan kata ulang, dan sebagainya. *Stemming* membantu mengidentifikasi dan mengatasi variasi Bahasa tersebut. Proses *Stemming* menghasilkan kata dasar seperti ditunjukkan dalam implementasi sistem pada Gambar 3.6

```

# create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
print('memulai proses stemming ....')
file_exists = exists('csv_process/' + sys.argv[1] + '_stemming.csv')
if file_exists is True: os.remove('csv_process/' + sys.argv[1] +
'_stemming.csv')
dataframe['Judul'] = dataframe.apply(lambda row:
stemmer.stem(row['Judul']), axis=1)
dataframe.to_csv('csv_process/' + sys.argv[1] + '_stemming.csv',
index=False)
print("Dataset stemming berhasil disimpan!")
  
```

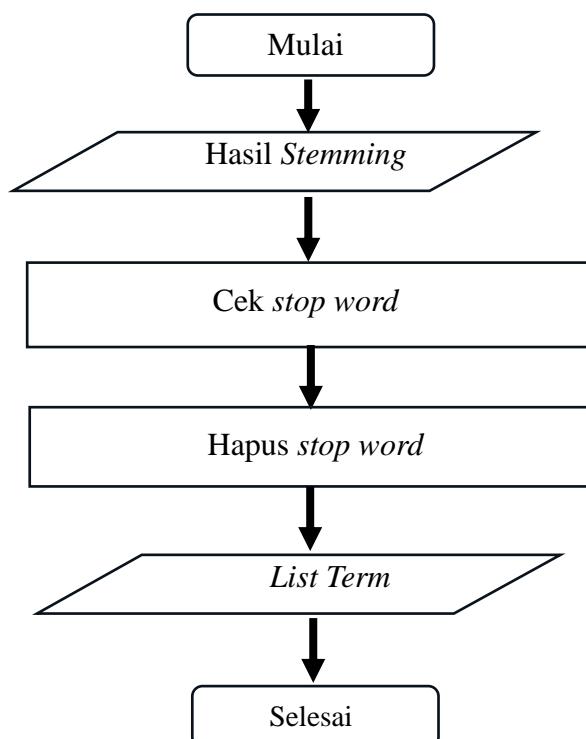
Gambar 3.6 Proses *Stemming*

Implementasi ini menggunakan StemmerFactory dari modul Sastrawi untuk membuat objek stemmer. Stemmer digunakan untuk melakukan stemming pada kata-kata stemmer untuk melakukan *Stemming* pada kolom 'Judul' dari suatu DataFrame, kemudian menyimpan dataset hasil *Stemming* ke dalam file CSV. Setelah proses *Stemming* selanjutnya dilakukan tahap *Stopword Removal*.

4.3 Stopword Removal

Proses penghapusan *Stopword* adalah langkah dalam pemrosesan teks yang melibatkan eliminasi kata-kata yang dianggap tidak memberikan informasi signifikan dalam konteks analisis teks. Kata-kata ini disebut "*Stop Word*" karena mereka sering muncul dalam teks namun tidak memberikan nilai tambah dalam pemahaman makna atau konten. Penghapusan *Stopword* bertujuan untuk menyederhanakan teks dan meningkatkan efisiensi dalam analisis teks. Tahap untuk proses *Stopword Removal* ditunjukkan pada Gambar

3.7



Gambar 3.7 Alur Stopword Removal

Tujuan utama dari penghapusan *Stopword* adalah menghilangkan kata-kata yang sering muncul dan umumnya tidak memberikan kontribusi signifikan terhadap pemahaman konten teks. Hal ini membantu mengurangi *noise* dan fokus pada kata-kata yang lebih informatif. Implementasi pada sistem ditunjukkan pada Gambar 3.8 untuk proses *Stopword Removal*.

```
print("memulai proses stopword ....")
file_exists2 = exists('csv_process/' + sys.argv[1] + '_stopword.csv')
if file_exists2 is True:
    os.remove('csv_process/' + sys.argv[1] + '_stopword.csv')
dataframe = pd.read_csv("csv_process/" + sys.argv[1] +
                       "_stemming.csv", on_bad_lines = 'skip')
stop_factory = StopWordRemoverFactory()
data = stop_factory.get_stop_words()
stopword = stop_factory.create_stop_word_remover()
dataframe['Judul'] = dataframe.apply(lambda row:
                                      stopword.remove(row['Judul']), axis=1)
dataframe.to_csv('csv_process/' + sys.argv[1] + '_stopword.csv',
                 index=False)
print("Dataset stopword berhasil disimpan!")
```

Gambar 3.8 Implmentasi *Stopword Removal*

`StopWordRemoverFactory()` dari modul Sastrawi digunakan untuk membuat objek untuk mengelola proses penghapusan *stopword*. `get_stop_words()` digunakan untuk mendapatkan daftar *stopword* bahasa Indonesia. `create_stop_word_remover()` digunakan untuk membuat objek *remover stopword*. Menggunakan fungsi `apply` untuk menerapkan penghapusan *Stopword* pada setiap baris dalam kolom 'Judul'. Menggunakan `to_csv` dari Pandas untuk menyimpan DataFrame yang sudah dihapus *stopword* ke dalam file CSV baru. Penghapusan *Stopword* adalah salah satu langkah penting dalam persiapan data teks yang membantu memfokuskan analisis pada kata-kata yang lebih bermakna dan relevan.

4.4 Tokenazation

Proses Tokenisasi (*Tokenization*) adalah langkah awal dalam pemrosesan teks yang melibatkan pembagian teks menjadi unit diskrit yang lebih kecil, yang disebut dengan token. Token dapat berupa kata, frasa, atau entitas lain yang menjadi unit dasar dalam analisis teks. Proses tokenisasi bertujuan untuk menyederhanakan teks agar dapat diolah dengan lebih mudah dalam analisis bahasa alami, pemrosesan bahasa alami (NLP), atau tugas-tugas pemrosesan teks. Token menjadi unit dasar yang dapat dihitung dan diukur dalam berbagai analisis. Proses tokenisasi merupakan langkah dasar dalam pemrosesan bahasa alami dan analisis teks. Data teks seringkali diubah menjadi token sebelum diolah lebih lanjut oleh model atau algoritma. Pada Gambar 3.9 ditunjukkan hasil dari proses *Tokenazation*.

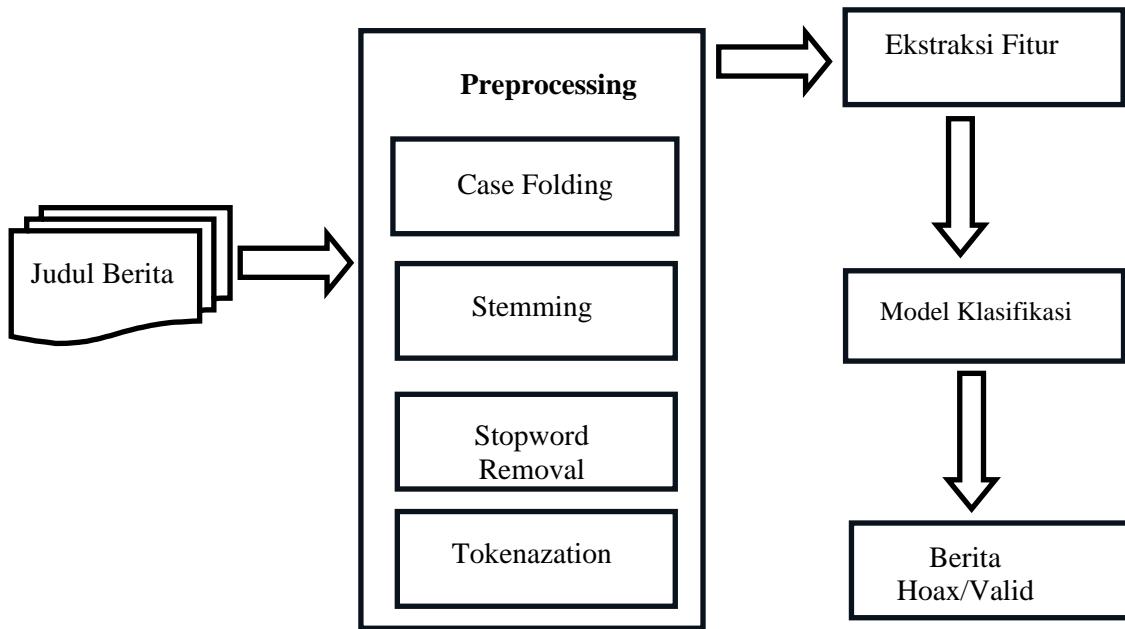
```
file_exists3 = exists('csv_process/' + sys.argv[1] +  
'_tokenized.csv')  
if file_exists3 is True:  
    os.remove('csv_process/' + sys.argv[1] + '_tokenized.csv')
```

Gambar 3.9 Proses *Tokenazation*

‘exists’ dari modul os.path digunakan untuk memeriksa apakah file CSV hasil proses tokenisasi sudah ada. Jika file tersebut sudah ada (file_exists3 is True), maka os.remove dari modul os akan menghapus file tersebut. Tujuan dari penghapusan ini mungkin adalah untuk menghindari penumpukan data hasil proses pada file yang sama jika program dijalankan beberapa kali.

3.1.3 Perencanaan Sistem

Rencana sistem yang diusulkan dalam penelitian ini, menggunakan tiga komponen utama *Preprocessing*, Ekstraksi Fitur dan model klasifikasi. Sistem yang dirancang pada penelitian terlihat seperti pada Gambar 3.10



Gambar 3.10 Rancangan Desain Sistem

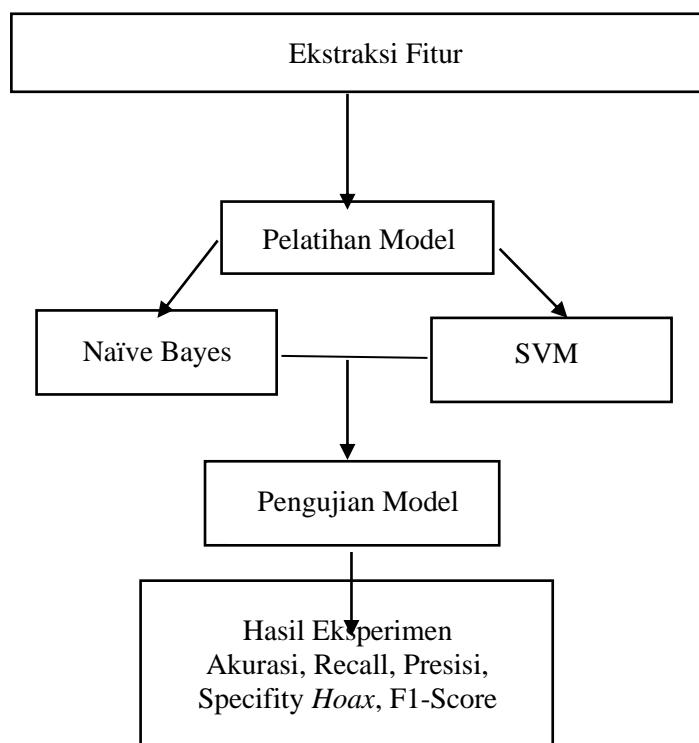
Pada Gambar 3.10 rancangan desain sistem dapat dijelaskan sebagai berikut :

1. Kumpulan judul berita berita valid dan berita *Hoax* dari hasil crowling,
2. Setelah itu dilakukan proses *Text processing* seperti proses *Case Folding* (Merubah karakter menjadi huruf kecil), *Stemming* (mendapatkan kata dasar), *Stopword Removal* (menghilangkan kata yang tidak penting), *Tokenization* (kalimat dipecah menjadi token).
3. Dari proses teks *Preprocessing* mendapatkan kata dasar.
4. Dalam konteks ini, pembobotan dilakukan dengan menggunakan TF IDF untuk memberikan bobot pada term dalam dataset atau dokumen. Bobot ini diberikan berdasarkan urutan token unigram, sehingga jumlah token dari TF IDF hanya satu kata saja..

5. Dataset pada penelitian ini dibagi menjadi 4 set, Latih-Uji dan tes-set untuk dimasukkan kedalam algoritma pengklasifikasian dengan perbandingan Latih-Uji 90% dan tes-set 10%, Latih-Uji 80% dan tes-set 20%, Latih-Uji 70% dan tes-set 30%, Latih-Uji 60% dan tes-set 40% dari keseluruhan dataset. Algoritma *Naïve Bayes* dan SVM, Setelah itu, dilakukan pelatihan data yang menghasilkan pembelajaran yang digunakan sebagai referensi dalam proses pengujian algoritma.
6. Mengukur kehandalan algoritma dengan *Confusion Matrix* untuk mendapatkan hasil Akurasi, Presisi dan *Recall* dari masing masing metode *Naïve Bayes* dan SVM.

3.1.4 Eksperimen

Pada uji coba, scenario yang coba diusulkan dalam penelitian ini, pada Gambar 3.11.



Gambar 3.11 Ujicoba Sistem

Pada alur ujicoba sistem dapat dijelaskan sebagai berikut :

1. Ekstraksi Fitur

Mengubah teks berita menjadi representasi numerik yang dapat digunakan oleh algoritma klasifikasi. Contoh ekstraksi fitur adalah menggunakan metode TF IDF untuk menghitung frekuensi kemunculan kata-kata dalam setiap dokumen.

2. Pelatihan Model

- a. *Naive Bayes*

Menggunakan data training dan label klasifikasi untuk melatih model *Naive Bayes*. Menghitung probabilitas prior untuk setiap kelas (*Hoax* dan non-*Hoax*). Menghitung probabilitas kondisional untuk setiap fitur dalam setiap kelas. Membangun model *Naive Bayes* berdasarkan probabilitas prior dan probabilitas kondisional

- b. SVM

Menggunakan data training dan label klasifikasi untuk melatih model SVM. Melakukan proses pelatihan SVM dengan mencari *Hyperplane* yang memaksimalkan margin antara kelas-kelas

3. Pengujian Model

Menggunakan data testing untuk menguji performa model yang telah dilatih. Melakukan prediksi klasifikasi berita menggunakan model *Naive Bayes* atau SVM. Menghitung metrik evaluasi seperti Akurasi, Presisi, *Recall*, dan F1-Score untuk mengevaluasi performa model.

Evaluasi Metriks

Untuk mengukur kinerja Akurasi, Presisi, *Recall*, F1-Score, dan bagaimana pendekatan relatif terhadap nilai yang tepat.

Akurasi = jumlah prediksi yang benar/jumlah total prediksi.

Pendekatan yang ada menganggap masalah berita palsu sebagai masalah klasifikasi yang memprediksi apakah sebuah artikel berita palsu (*Hoax*) atau tidak:

True Positive (TP): ketika judul berita palsu yang diprediksi sebenarnya diklasifikasikan sebagai berita palsu.

True Negative (TN): saat judul berita benar yang diprediksi benar-benar diklasifikasikan sebagai berita benar.

False Negative (FN): ketika prediksi judul berita sebenarnya diklasifikasikan sebagai berita palsu.

False Positive (FP): ketika judul berita palsu yang diprediksi sebenarnya diklasifikasikan sebagai berita benar.

Persamaan Akurasi diberikan oleh Persamaan. (3.1)

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.1)$$

Presisi adalah seberapa dekat pengukuran, yang diberikan oleh Persamaan. (3.2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.2)$$

Recall adalah berapa banyak positif aktual yang didefinisikan dengan benar, diberikan oleh Persamaan. (3.3).

$$\text{Recall|Sensitivity|TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.3)$$

Definisi F1 adalah bahwa jika biaya positif palsu dan negatif palsu bervariasi, kami membutuhkan presisi dan penarikan kembali, seperti yang diberikan oleh Persamaan. (3.4).

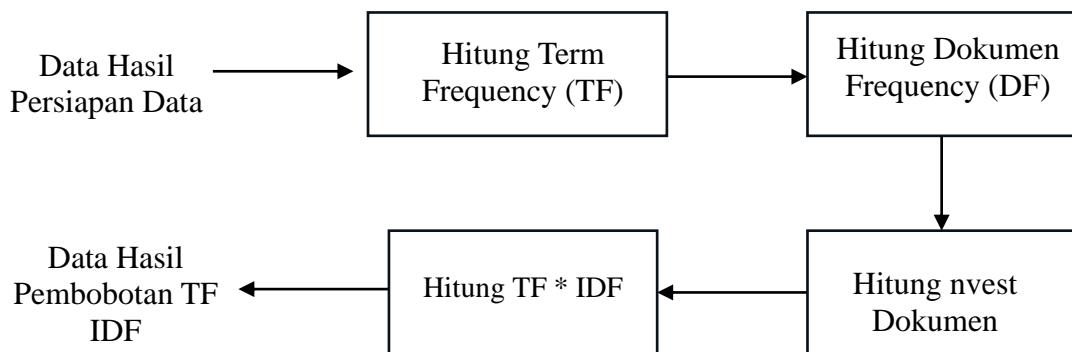
$$F1 = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FN} + \text{FP})} \quad (3.4)$$

Pembobotan TF IDF

Data bersih hasil dari proses persiapan data menghitung kata-kata kunci yang memiliki bobot tinggi berdasarkan frekuensi kemunculan kata tersebut dalam dokumen dan koleksi dokumen secara keseluruhan dengan TF IDF. Menggunakan *Term Frequency* (TF) untuk mengukur seberapa sering suatu kata muncul dalam suatu dokumen. Ini dihitung dengan membagi jumlah kemunculan kata tersebut dengan total kata dalam dokumen. Menggunakan *Inverse Document Frequency* (IDF) untuk mengukur seberapa unik suatu

kata dengan membandingkan jumlah dokumen dalam korpus yang mengandung kata tersebut. Ini dihitung dengan membagi total dokumen dalam korpus dengan jumlah dokumen yang mengandung kata tersebut. Untuk menghitung kemunculan term (TF) dalam suatu dokumen dan jumlah term dalam kumpulan dokumen (IDF) menggunakan Persamaan (3.5).

Pada Gambar 3.12 menunjukkan alur proses pembobotan.



Gambar 3.12 Proses Pembobotan TF IDF

Dalam tahap pengujian, data uji akan dicocokkan dengan model yang telah disiapkan dalam proses persiapan data TF IDF. TF IDF digunakan untuk menilai frekuensi istilah tertentu relatif terhadap sebuah kata dalam kumpulan dokumen, serta untuk menilai seberapa umum atau tidak umum suatu kata dalam sebuah korpus (kumpulan teks yang terstruktur). Proses TF IDF ini menggunakan urutan token Unigram, sehingga jumlah token dari TF IDF hanya satu kata.

Berikut rumus untuk perhitungan TF IDF :

$$F = \begin{cases} 1 + \log_{10}(ft, d), & ft, d > 0 \\ 0, & ft, d = 0 \end{cases} \quad (3.5)$$

$$\text{IDF}_j = \log\left(\frac{D}{D_{f_j}}\right) \quad (3.6)$$

$$w_{ij} = tf_{ij} \times \log\left(\frac{D}{D_{f_j}}\right) + 1 \quad (3.7)$$

Keterangan:

ft, d = frekuensi term (t) untuk dokumen (d)

D = jumlah keseluruhan dokumen dalam dataset

df_j = jumlah dokumen yang mengandung term (tj)
 w_{ij} = bobot term (tj) untuk dokumen (d)
 tf_{ij} = jumlah penemuan term (tj) dalam dokumen (d)

Proses TF IDF yang sudah dijelaskan dengan rumus, implementasi pada sistem ditunjukan dalam Gambar 3.13

```

def tf(stringDict):
    wordDict = stringDict
    if (isinstance(stringDict, str) == True):
        wordDict = literal_eval(stringDict)
    total = sum(list(wordDict.values()))
    for idx, x in enumerate(wordDict):
        wordDict[x] = wordDict[x] / float(total)
    return wordDict
def bagOfWords(dict): #pengambilan kata unique dipakai di perhitungan tfidf
    mergedDict = []
    for idx, x in enumerate(dict):
        wordDict = x
        if (isinstance(x, str) == True):
            wordDict = literal_eval(x)
        mergedDict = mergedDict + wordDict
    my_list = list(set(mergedDict))
    return my_list
def idf(words, documents, N):
    idfDict = dict.fromkeys(words, 0)
    for idx, word in enumerate(words):
        for idy, doc in enumerate(documents):
            if(word in doc):
                idfDict[word] += 1
        for word, val in idfDict.items():
            idfDict[word] = math.log10(N / float(val))
    return idfDict
def tfidf(tfBagOfWords, idfs):
    tfidf = {}
    for word, val in tfBagOfWords.items():
        tfidf[word] = val * idfs[word]
    return tfidf
  
```

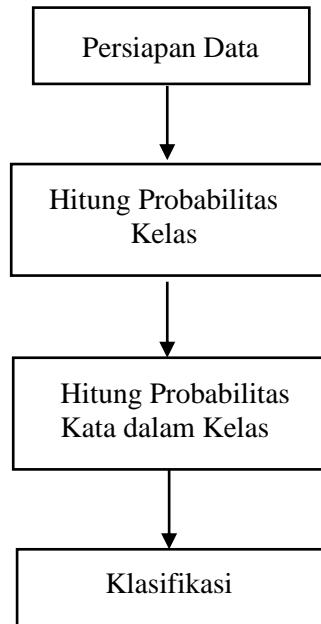
Gambar 3.13 Implementasi TF IDF

Gambar tersebut menjelaskan fungsi-fungsi ini mengambil string yang mewakili kamus (stringDict), kemudian mengonversinya menjadi kamus *Python* menggunakan literal_eval. Selanjutnya, fungsi menghitung nilai Total *Frequency* (TF) untuk semua kata dalam kamus dengan menjumlahkan frekuensi kemunculan setiap kata. Fungsi ini mengambil daftar kamus (dict) yang mewakili dokumen. Fungsi menggunakan loop untuk menggabungkan semua kamus menjadi satu list (mergedDict). Kemudian, fungsi membuat daftar kata unik dengan menggunakan list(set(mergedDict)). Hasilnya adalah daftar kata unik yang muncul dalam semua dokumen. Fungsi kemudian menghitung nilai TF untuk setiap kata dengan membagi frekuensi kemunculan kata tersebut dengan total frekuensi. Hasilnya adalah kamus yang berisi nilai TF untuk setiap kata dalam dokumen. Selanjutnya fungsi mengambil daftar kata (*words*) dan daftar dokumen (*documents*). Fungsi menginisialisasi kamus IDF (idfDict) dengan kata-kata dari daftar kata dan mengatur nilai awal menjadi 0. Fungsi menggunakan dua loop untuk menghitung jumlah dokumen yang mengandung setiap kata dalam daftar kata. Selanjutnya, fungsi menghitung nilai *Total Frequency* (TF) untuk semua kata dalam kamus dengan menjumlahkan frekuensi kemunculan setiap kata. Fungsi ini kemudian menghitung nilai TF untuk setiap kata dengan membagi frekuensi kemunculan kata tersebut dengan total frekuensi.

Hasilnya adalah kamus yang berisi nilai TF untuk setiap kata dalam dokumen. kemudian TF IDF yang mengambil dua parameter: tfBagOfWords merupakan representasi Term Frequency (TF) dari suatu dokumen dalam bentuk kamus (*dictionary*), di mana kunci adalah kata-kata dalam dokumen dan nilai adalah frekuensi kemunculan kata-kata tersebut. Idfs merupakan representasi *Inverse Document Frequency* (IDF) dalam bentuk kamus, di mana kunci adalah kata-kata dan nilai adalah nilai IDF dari setiap kata. Fungsi ini mengembalikan kamus baru yang mewakili nilai TF IDF untuk setiap kata dalam dokumen, dihitung dengan mengalikan nilai TF dengan nilai IDF.

Naïve Bayes

Model *Naïve Bayes* digunakan untuk membandingkan hasil dari setiap algoritma yang digunakan dengan memanfaatkan *library Scikit Learn*, yaitu *Sklearn Naïve_Bayes*, untuk mengimpor *Naïve Bayes* dalam bahasa pemrograman Python. Alur dalam pengujian menggunakan algoritma *Naïve Bayes* diusulkan dalam penelitian ini, pada Gambar 3.14 sebagai berikut :



Gambar 3.14 Alur Algoritma Naïve Bayes

Pada Gambar tersebut dijelaskan alur perhitungan, antara lain :

1. Persiapan Data, Persiapan data meliputi *Preprocessing* pada teks berita, mengubah teks berita menjadi representasi numerik dengan menggunakan metode TF IDF.
2. Perhitungan Probabilitas Kelas, hitung jumlah data training dengan label *Hoax* (*N_Hoax*) dan label bukan *Hoax* (*N_Valid*). Hitung probabilitas prior untuk kelas *Hoax* ($P(Hoax)$) dan kelas bukan *Hoax* ($P(Valid)$) dalam Persamaan (3.8) dan (3.9):

$$P(Hoax) = \frac{N_{Hoax}}{N_{Hoax} + N_{Valid}} \quad (3.8)$$

$$P(Valid) = \frac{N_{Valid}}{N_{Valid} + N_{Hoax}} \quad (3.9)$$

3. Perhitungan probabilitas kemunculan kata dalam kelas, *Hoax* ($P(\text{Word}|\text{Hoax})$) dan dalam kelas bukan *Hoax* ($P(\text{Word}|\text{Valid})$) dalam Persamaan (3.10) dan (3.11):

$$P(\text{Word}|\text{Hoax}) = \frac{\text{(jumlah kemunculan kata dalam kelas hoax +1)}}{\text{Total Kata dalam kelas hoax + jumlah kata unik}} \quad (3.10)$$

$$P(\text{Word}|\text{Valid}) = \frac{\text{(jumlah kemunculan kata dalam kelas valid +1)}}{\text{Total Kata dalam kelas valid + jumlah kata unik}} \quad (3.11)$$

4. Klasifikasi, Diberikan teks berita yang akan diklasifikasikan (teks_uji). Lakukan *Preprocessing* pada teks_uji yang sama seperti pada data training. Ubah teks_uji menjadi representasi numerik yang sesuai. Hitung probabilitas posterior untuk teks_uji dalam kelas *Hoax* ($P(\text{Hoax}|\text{teks_uji})$) dan dalam kelas bukan *Hoax* ($P(\text{Valid}|\text{teks_uji})$) menggunakan rumus Teorema Bayes dalam Persamaan (3.12) dan (3.13):

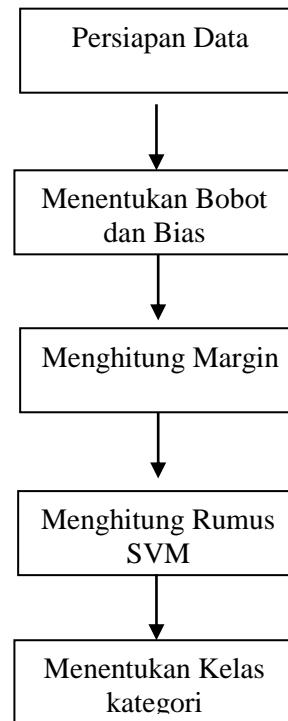
$$P(\text{Hoax}|\text{teks_uji}) = P(\text{teks_uji}|\text{hoax}) * \frac{P(\text{hoax})}{P(\text{teks_uji})} \quad (3.12)$$

$$P(\text{Valid}|\text{teks_uji}) = P(\text{teks_uji}|\text{valid}) * \frac{P(\text{valid})}{P(\text{teks_uji})} \quad (3.13)$$

Membandingkan nilai $P(\text{Hoax}|\text{teks_uji})$ dan $P(\text{Valid}|\text{teks_uji})$ untuk menentukan label kelas yang paling memdekati.

Support Vector Machine (SVM)

Kemudian, peneliti menggunakan kernel yang tersedia pada *Support Vector Machine* untuk membandingkan hasil dari setiap kernel yang digunakan. Berikut beberapa alur dalam tahap menghitung algoritma SVM. Terdapat dalam Gambar 3.15



Gambar 3.15 Alur Algoritma SVM

Dalam menghitung SVM untuk klasifikasi Persiapan Data, diberikan data training yang terdiri dari fitur-fitur dan label kelas (*Hoax* atau bukan *Hoax*). Lakukan *Preprocessing* pada data, seperti normalisasi atau pemilihan fitur yang relevan. Pelatihan SVM untuk mendapatkan nilai Akurasi, Presisi dan *Recall*. Klasifikasi, setelah mendapatkan vektor bobot dan bias yang optimal, gunakan rumus prediksi SVM untuk mengklasifikasikan data baru. Rumus prediksi dalam Persamaan (3.14).

$$\text{SVM} = \text{sign}(w * x + b) \quad (3.14)$$

Keterangan :

w = vektor bobot,

x = vektor fitur data baru

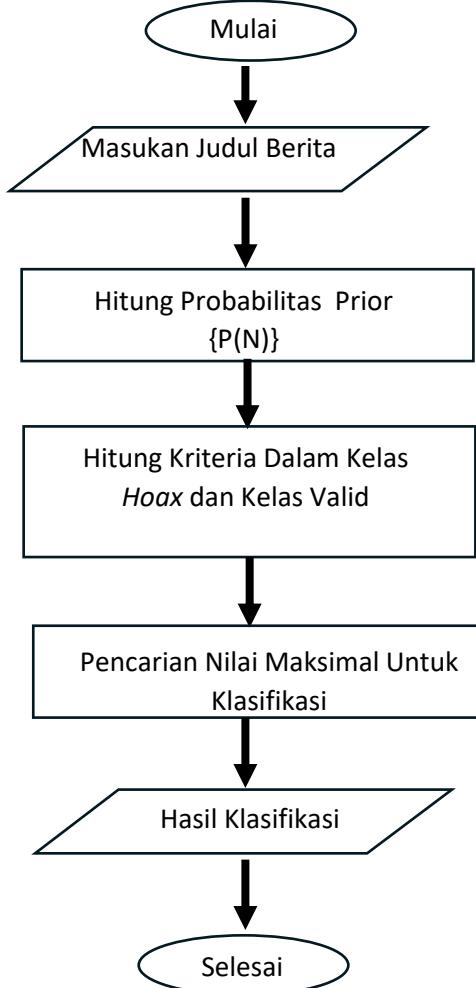
b = bias.

BAB IV

METODE NAIVE BAYES

4.1 Desain Metode Naive Bayes

Pada fase klasifikasi ini, digunakan jumlah kemunculan kata dari setiap kelas pada dataset. Bab ini akan menjelaskan pola dan desain klasifikasi yang dapat dilihat dalam proses pengujian menggunakan algoritma *Naïve Bayes* yang diusulkan dalam penelitian ini, sebagaimana tergambaran pada Gambar 4.1.



Gambar 4.1 Alur Algoritma *Naïve Bayes*

Rancangan *Naïve Bayes*, Menjelaskan dalam proses yang dikerjakan untuk menentukan kelas $h \in H$, H adalah judul berita. Sedangkan $x \in X$, X adalah kelas. Deskripsi $H = \{h_1, h_2, h_3, h_4, \dots, h_p\}$ dan deskripsi $X = \{x_1, x_2, x_3, x_4, \dots, x_p\}$ menentukan kelas dalam klasifikasi judul berita dilakukan dengan mendapatkan nilai maksimal dari $P(h|x)$ berdasarkan rumus Probabilitas

$$P = \{p(h|x) | h \in H \text{ dan } x \in X\} \quad (4.1)$$

Berdasarkan penjabaran algoritma *Naïve Bayes* untuk Persamaan perhitungan nilai dari Probabilitas $P(h|x)$ sebagai Persamaan berikut :

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}} \quad (4.2)$$

$$P(h|x) = \frac{P(x|h) * P(h)}{P(x)} \quad (4.3)$$

Keterangan :

$P(h|x)$ = Probabilitas berdasarkan kondisi peosterior

$P(h)$ = Probabilitas prior Probabilitas

$P(x|h)$ = Probabilitas berdasarkan kondisi Judul

$P(x)$ = Probabilitas Kelas

Dalam proses klasifikasi kelas menggunakan algoritma *Naïve Bayes* diperlukan persiapan data dan mementukan jumlah judul berita yang akan digunakan dalam proses perhitungan. Setelah data ditentukan, judul berita *Hoax* dan judul berita *Valid* didapatkan sejumlah 2.000 dataset yang terbagi menjadi 2 kelas yaitu kelas *Hoax* sebanyak 1542 dan kelas *Valid* sebanyak 458 judul berita. Dataset tersebut dibagi menjadi 2 dengan perbandingan 80:20, data training 80% (1.542 judul) dan 20% data uji (465 judul).

Menentukan Probabilitas setiap prior, selanjutnya menghitung probabilitas jumlah kelas *Hoax* dan kelas berita *Valid* dari dataset yang sudah ada. Nilai posterior dibandingkan dengan posterior di kelas lain untuk menentukan sampel kelas mana yang akan ditugaskan ke suatu kelas. Proses awal klasifikasi *Naïve Bayes* melibatkan perhitungan kemunculan fitur pada setiap kelas berita *Hoax* dan berita *Valid*. Dalam proses tersebut, fitur yang dihitung untuk jumlah kemunculan diambil dari peringkat TF IDF.

5.2 Implementasi Metode Naïve Bayes

Implementasi metode *Naïve Bayes* menggunakan pemrograman *Python* dan *Library* pendukung yaitu Numpy, Json, Math, Sklearn.naive_bayes dan menjalankan tahap perencanaan. Pada tahap ini dilakukan pemrosesan implementasi program sebagai berikut:

1. Input judul berita. Pada proses implementasi dengan sistem ditunjukkan dalam Gambar 4.2.

```
def index(): #halaman awal
    source_files = [os.path.splitext(filename)[0] for filename in
    os.listdir('csv_data_source')]
    return render_template('index.html', judul = '', dataset = '',
    datasetv = '', datatest = '', result = {}, resultcsv = [], hasil =
    "-", source_files = source_files, algo = '')
```

Gambar 4.2 Input Judul Berita

Dari kutipan implementasi tersebut, menjelaskan tentang proses membaca nama file dari direktori 'csv_data_source', kemudian merender template HTML ('index.html') dengan memberikan nilai-nilai default atau kosong untuk beberapa variable, kemudian dari judul yang dimasukan akan dilakukan proses persiapan data atau *Preprocessing*.

2. Melakukan *Preprocessing*. Setelah judul diinputkan, kemudian sistem melakukan *Preprocessing* tahap seperti *Case Folding*, *Stemming*, *Stopword Removal* dan *Tokenazation*. Berikut implementasi sistem yang disajikan dalam Gambar 4.3

```

if (request.form['judul']): #proses single text
    rtext_lower = request.form['judul'].lower() #lower case
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    rtext_stemming = stemmer.stem(rtext_lower) # stemming

    stop_factory = StopwordRemoverFactory()
    stopword = stop_factory.create_stop_word_remover()
    rtext_stopword = stopword.remove(rtext_stemming) # stopword

    nltk.download('punkt')
    rtext_token = nltk.word_tokenize(rtext_stopword) # tokenizing
    rtext_tcount = {i:rtext_token.count(i) for i in rtext_token} # generate term count
    rtext_dataframe_tcount = pd.DataFrame({ 'Judul': [json.dumps(rtext_tcount)] })
    rtext_tf = tf(rtext_tcount)
    rtext_dataframe = pd.DataFrame({ 'Judul': [json.dumps(rtext_token)] })
    rtext_raw_source = pd.DataFrame({ 'Judul': [request.form['judul']] })

```

Gambar 4.3 Proses *Preprocessing*

Pada Gambar tersebut menjelaskan tahap *Preprocessing* data sebelum data digunakan untuk proses pembobotan untuk mendapatkan nilai TF IDF.

3. Pembobotan TF IDF

Implementasi sistem pada proses pembobotan TF IDF dijelaskan dalam Gambar 4.4

```
# perhitungan tfidf

    dataframe1['Judul'] =  dataframe1['Judul'].apply(lambda x:
tfidf(x,idfWord)) #valid
    dataframe2['Judul'] =  dataframe2['Judul'].apply(lambda x:
tfidf(x,idfWord)) #hoax
    rtext_dataframe_tf['Judul'] =
    rtext_dataframe_tf['Judul'].apply(lambda x: tfidf(x,idfWord))
#hoax
# rtext_tfidf = tfidf(rtext_tf,idfWord)
```

Gambar 4.4 Proses TF IDF

Setiap baris kode di atas menggunakan fungsi TF IDF (tfidf) untuk menghitung nilai TF IDF dari setiap judul dalam kolom 'Judul' dari DataFrame yang sesuai (dataframe1, dataframe2, rtext_dataframe_tf). Nilai idfWord mungkin berisi nilai-nilai IDF yang telah dihitung digunakan untuk proses hitung klasifikasi dengan metode *Naïve Bayes*.

4. Tahap prediksi menggunakan metode *Naïve Bayes* disajikan dalam Gambar 4.5

```
# menyiapkan data untuk naive bayes
mergeDataframe = pd.concat([dataframe1,dataframe2])
dataset = mergeDataframe['Judul'].apply(lambda x:
list(x.values())) #valid & hoax
dataset = list(dataset)
vresult = np.full(len(dataframe1['Judul']), 'valid')
hresult = np.full(len(dataframe2['Judul']), 'hoax')
dataresult = np.append(vresult, hresult)
dataresult = dataresult.tolist()
# data yang mau di prediksi
datatest = rtext_dataframe_tf['Judul'].apply(lambda x:
list(x.values())) #valid & hoax
datatest = list(datatest)
# datatest = list(rtext_mapping.values())
```

Gambar 4.5 Prediksi Metode *Naïve Bayes*

Gambar tersebut menjelaskan mengenai proses model klasifikasi metode *Naïve Bayes*, mergeDataframe['Judul'] adalah cara mengakses kolom 'Judul' dari DataFrame yang telah digabungkan. .apply(lambda x: list(x.values())), Menggunakan fungsi apply untuk mengubah setiap judul menjadi daftar nilai (values). list(dataset) Mengonversi Series hasil dari apply ke dalam list. Hasil akhirnya adalah daftar yang berisi daftar judul yang telah diubah menjadi daftar nilai. dataresult.tolist(), Mengonversi array NumPy menjadi list. Hasil akhirnya adalah list yang berisi label 'valid' dan '*Hoax*' untuk setiap judul dalam dataset. list(datatest), Mengonversi Series hasil dari apply ke dalam list. Hasil akhirnya adalah daftar yang berisi daftar judul yang telah diubah menjadi daftar nilai.potongan kode ini menyatukan data dari dataframe1 dan dataframe2 menjadi satu dataset, menyiapkan label untuk dataset tersebut, dan menyiapkan data yang akan diprediksi (test data) dari rtext_dataframe_tf. Tahap tersebut dapat digunakan sebagai input untuk model klasifikasi *Naïve Bayes*.

5. Tahap terakhir adalah pengeluaran berupa hasil perbandingan Akurasi dari algoritma yang dijadikan penelitian, kemudian akan dicatat dan masuk ke tahap verifikasi dan Validasi menggunakan *Confusion Matrix*.

Selanjutnya menentukan parameter model *Naïve Bayes* yang akan digunakan. Menggunakan perhitungan Probabilitas karena dalam deteksi berita *Hoax*, data memiliki banyak fitur atau atribut yang digunakan untuk mengklasifikasikan berita, seperti kata-kata kunci, kata yang sering muncul , dan sebagainya. Perhitungan Probabilitas efektif dalam menangani data dengan banyak fitur. Adapun untuk pemrosesan antara lain :

Data Latih dan Data Uji

Model data dibentuk untuk algoritma klasifikasi pada data latih. Model ini berfungsi sebagai representasi pengetahuan untuk memprediksi data kelas baru yang tidak ada sebelumnya. Pembagian jumlah data latih dan uji dapat dilihat pada Tabel 4.1.

Tabel 4.1 Pembagian Data Latih dan Data Uji

Kelas	Data Latih	Data Uji
Valid	458 Teks	256 Teks
<i>Hoax</i>	1452 Teks	144 Teks
Total	2.000 Teks	400 Teks

Berdasarkan Tabel diatas perbandingan jumlah data latih dan uji. Dengan perbandingan data dari keseluruhan 2.000 judul berita, melakukan input data set dengan rasio perbandingan training 90% test : 10% dengan jumlah data 1800 data training dengan komposisi 458 data berita Valid, 1342 berita *Hoax*, dan 200 data sebagai data uji. Komposisi perbandingan 80 % : testing 20%, jumlah data 1600 training dengan komposisi 458 berita Valid dan 1142 berita *Hoax*, dan data uji sejumlah 400 data. Data latih pada perbandingan 70% uji 30%, jumlah data 1400 dengan komposisi 458 berita Valid dan 942 berita *Hoax*, dan 600 data sebagai data uji. Perbandingan 60 % : testing 40% dengan jumlah data 1200 dengan komposisi berita 458 Valid dan 742 berita *Hoax* dan 800 data uji. Perbandingan data latih 50% : testing 50%. Jumlah data 1000 dengan komposisi 458 berita Valid dan 542 berita *Hoax*, dan 100 data sebagai data uji. Berikut perbandingan data yang digunakan ditunjukan dalam Tabel 4.2

Tabel 4.2 Perbandingan Data

Data Latih	Perbandingan Data				
	50% : 50%	60% : 40%	70% : 30%	80% : 20%	90% : 10%
Berita Hoax	542	742	942	1142	1342
Berita Valid	458	458	458	458	458
Total Data Latih	1000	1200	1400	1600	1800
Total Data Uji	1000	800	600	400	200

Setelah mendapatkan perbandingan data latih dan data uji untuk selanjutnya dilakukan skenario pengujian masing masing perbandingan data. Proses pengujian dijabarkan sebagai berikut :

- Pengujian I

Pada pengujian I pembagian data latih 50% dan data uji 50% mendapatkan hasil penjabaran skrenario Tabel 4.3 pengujian *Confusion Matrix* sebagai berikut :

Tabel 4.3 Pengujian 50% : 50%

	Actual	
	TP	FN
Prediction	458	154
	FP	TN
	0	388

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 458 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 388 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan

kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 154 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian I perbandingan 50% : 50% sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{458+388}{458+388+0+145}$$

$$\text{Akurasi} = \frac{846}{1000} = 0,846 = 84,8\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{458}{458+0} = \frac{458}{458} = 1 = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{458}{458+154} = \frac{458}{612} = 0,748 = 74,8\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{458}{458+1/2(0+154)}$$

$$\text{F1 Score} = \frac{458}{458+77} = \frac{458}{535} = 0,856 = 85,6\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 4.4

Tabel 4.4 Hasil Perhitungan 50% : 50%

<i>Naïve Bayes</i>	Akurasi	Presisi	<i>Recall</i>	F1-Score
50% : 50%	84,6	100	74,8	85,6

Dari Tabel perhitungan perbandingan data 50% berita Valid dan 50% berita *Hoax* mendapatkan nilai Akurasi sebesar 84,5%, nilai Presisi 100% , nilai *Recall* 74,8% dan nilai F1-Score 85,6%.

- Pengujian II

Pada pengujian II pembagian data latih 60% dan data uji 40% didapatkan hasil penjabaran skrenario Tabel 4.5 pengujian *Confusion Matrix* sebagai berikut :

Tabel 4.5 Pengujian 60% : 40%

		Actual	
		TP	FN
Prediction	458	93	
	FP	TN	
	0	249	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 458 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 249 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 93 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian II perbandingan 60% : 40% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{458+249}{458+249+0+93}$$

$$\text{Akurasi} = \frac{707}{800} = 0,883 = 88,3\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{458}{458+0} = \frac{458}{458} = 1 = 100\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{458}{458+93} = \frac{458}{551} = 0,831 = 83,1\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{458}{458+1/2(0+93)}$$

$$\text{F1 Score} = \frac{458}{504,5} = 0,907 = 90,7\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 4.6

Tabel 4.6 Hasil Perhitungan 60% : 40%

<i>Naïve Bayes</i>	Akurasi	Presisi	Recall	F1-Score
60% : 40%	88,3%	100%	83,1%	90,7%

Dari Tabel diatas perhitungan perbandingan data 60% berita Valid dan 40% berita *Hoax* mendapatkan nilai Akurasi sebesar 88,3%, nilai Presisi 100% , nilai Recall 83,1% dan nilai F1-Score 90,7%.

- Pengujian III

Pada pengujian III pembagian data latih 70% dan data uji 30% didapatkan hasil penjabaran skenario Tabel 4.7 pengujian *Confusion Matrix*

Tabel 4.7 Pengujian 70% : 30%

		Actual	
		TP	FN
Prediction	TP	149	0
	FP		TN
	FN	0	460

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 149 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 460 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian III perbandingan 70% : 30% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{149+460}{149+460+0+0}$$

$$\text{Akurasi} = \frac{609}{609} = 1 = 100\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{149}{149+0} = \frac{149}{149} = 1 = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{149}{149+0} = 1 = 100\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{149}{149+1/2(0+0)}$$

$$\text{F1 Score} = \frac{149}{149} = 1 = 100\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 4.8

Tabel 4.8 Hasil Perhitungan 70% : 30%

<i>Naïve Bayes</i>	Akurasi	Presisi	Recall	F1-Score
70% : 30%	100%	100%	100%	100%

Dari Tabel perhitungan perbandingan data 70% berita Valid dan 30% berita *Hoax* mendapatkan nilai Akurasi sebesar 100%, nilai Presisi 100% , nilai Recall 100% dan nilai F1-Score 100%.

- Pengujian IV

Pada pengujian IV pembagian data latih 80% dan data uji 20% didapatkan hasil penjabaran skrenario Tabel 4.9 pengujian *Confusion Matrix* sebagai berikut :

Tabel 4.9 Pengujian 80% : 20%

		Actual	
Prediction	TP	FN	
	87	1	
	FP	TN	
	0	362	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 87 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 362 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 1 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian IV perbandingan 80% : 20% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{87+362}{87+362+0+1}$$

$$\text{Akurasi} = \frac{449}{450} = 0,997 = 99,7\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{87}{87+0} = 1 = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{87}{87+1} = 0,988 = 98,8\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{87}{87+1/2(0+1)}$$

$$\text{F1 Score} = \frac{87}{87,5} = 0,994 = 99,4\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel

4.10

Tabel 4.10 Hasil Perhitungan 80% : 20%

<i>Naïve Bayes</i>	Akurasi	Presisi	Recall	F1-Score
80% : 20%	99,7%	100%	98,8%	99,4%

Perhitungan perbandingan data 80% berita Valid dan 20% berita *Hoax* mendapatkan nilai Akurasi sebesar 99,7%, nilai Presisi 100% , nilai Recall 98,8% dan nilai F1-Score 99,4%.

- Pengujian V

Pada pengujian V pembagian data latih 90% dan data uji 10% mendapatkan hasil penjabaran skrenario Tabel 4.11 pengujian *Confusion Matrix* sebagai berikut :

Tabel 4.11 Pengujian 90% : 10%

		Actual	
		TP	FN
Prediction	30	1	
		FP	TN
	0	204	

Hasil *True Positive* (TP) metode berhasil mengidentifikasi 30 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 204 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 1 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian V perbandingan 90% : 10% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{30+204}{30+204+0+0}$$

$$\text{Akurasi} = \frac{234}{234} = 100\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{30}{30+0} = 1 = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{30}{30+0} = 1 = 100\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{30}{30+1/2(0+0)}$$

$$\text{F1 Score} = \frac{30}{30} = 1 = 100\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 4.12

Tabel 4.12 Hasil Perhitungan 90% : 10%

<i>Naïve Bayes</i>	Akurasi	Presisi	<i>Recall</i>	F1-Score
90% : 10%	100%	100%	100%	100%

Tabel perhitungan perbandingan data 90% berita Valid dan 10% berita *Hoax* mendapatkan nilai Akurasi sebesar 100%, nilai Presisi 100% , nilai *Recall* 100% dan nilai F1-Score 100%.

Dari hasil pengujian yang telah dilakukan sebanyak 5 kali pengujian didapatkan nilai rata-rata Akurasi, Presisi, *Recall* seperti ditampilkan pada Tabel 4.13 sebagai berikut

Tabel 4.13 Hasil Metode *Naïve Bayes*

Pengujian	Latih : Uji	<i>Naïve Bayes</i>			
		Akurasi	Presisi	<i>Recall</i>	F1-Score
Pengujian I	50:50	80%	100%	80%	100%
Pengujian II	60:40	88,3%	100%	83,1%	90,7%
Pengujian III	70:30	100%	100%	100%	100%
Pengujian IV	80:20	99,7%	100%	98,8%	99,4%
Pengujian V	90:10	100%	100%	100%	100%
Rata - Rata		93%	100%	93%	100%

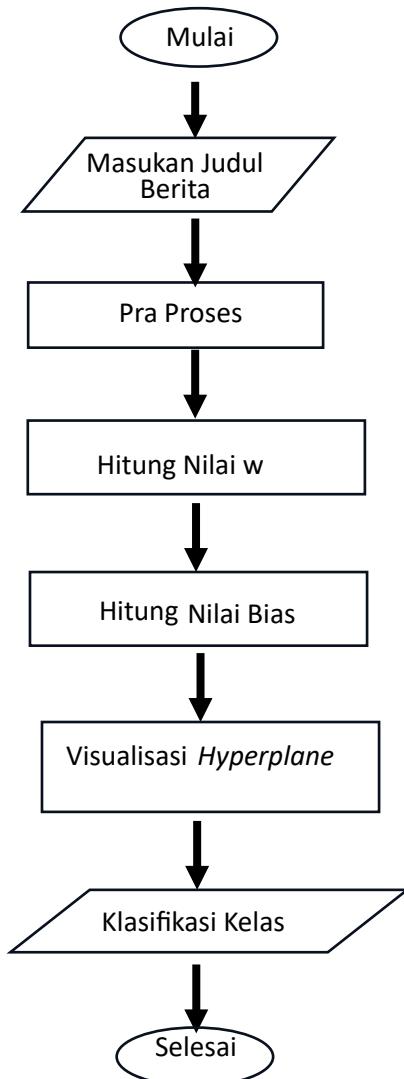
Pada Tabel terlihat bahwa metode *Naïve Bayes* mampu melakukan deteksi pada berita *Hoax*. Dengan 5 kali uji coba dengan melakukan perubahan data latih dan uji menunjukkan nilai rata-rata Akurasi sebesar **93%**, Presisi sebesar **100%**, *Recall* sebesar **93%** dan F1-Score sebesar **100%**.

BAB V

METODE SUPPORT VECTOR MACHINE (SVM)

5.1 Rancangan Metode Support Vector Machine (SVM)

Dalam bab ini, akan dijelaskan pola dan rancangan klasifikasi yang dapat dilihat pada alur pengujian menggunakan algoritma *Support Vector Machine* yang diusulkan dalam penelitian ini, seperti yang tergambar pada Gambar 5.1.



Gambar 5.1 Alur Algoritma SVM

Desain menggunakan metode SVM yang dijelaskan pada Gambar yaitu dari mulai masukan judul berita, pra proses, menentukan kata yang sering muncul dari tiap dokumen yang digunakan, menentukan nilai w dan bias, menentukan visualisasi *Hyperplane*, dan selanjutnya mengklasifikasikan data sesuai kelas. Sebagai contoh data untuk perhitungan ditunjukkan oleh Tabel 5.1.

Tabel 5.1 Sample Data

No	Judul	Jum_kata	Bobot_tf	Kelas
1	Pemakaian masker menyebabkan penyakit legionnaires	5	4	Hoax
2	Instruksi gubernur jateng tentang penilangan bagi yg tidak bermasker di muka umum rp.150.000 menggunakan e-tilang via apps pikobar	18	15	Hoax
3	Foto jokowi adalah presiden terbaik dlm sejarah bangsa indonesia	11	10	Hoax
4	ini bukan politik tapi kenyataan pak jokowi berhasil memulangkan 11000 triliun uang negara dari swiss.	15	12	Hoax
5	Foto kadrun kalo lihat foto ini panas dingin	8	7	Hoax
6	event promo smartphone jne 2 020 spesial di bulan juni	9	8	Hoax
7	mereka sudah mempersiapkan diri dengan baik untuk menguasai negri ini	10	6	Hoax
8	Warga Australia Disarankan Tetap Mengenakan Masker Hadapi Penyebaran Flu di Musim Dingin	12	11	valid
9	Naik KA Tetap wajib booster dan pakai masker	8	7	valid
10	Hadiri Hannover Messe 2023 Presiden Jokowi dan Ibu Iriana Bertolak ke Jerman	12	10	valid
11	Presiden Jokowi Sapa Masyarakat Indonesia hingga Bertemu Kanselir Jerman	9	9	valid

No	Judul	Jum_kata	Bobot_tf	Kelas
12	Presiden Jokowi Gelar Rapat Terkait Progres Pembangunan Hunian ASN di IKN	11	10	valid
13	Wujudkan Pemilu Berkualitas Kominfo - Polri Perkuat Kerja Sama	9	9	valid
14	Gubernur Jateng instruksikan pemberlakuan rekayasa lalu lintas jelang lebaran	9	9	valid

Dalam Tabel tersebut sudah dikelompokan untuk judul berita, jumlah kata, bobot *term frequency* dan kategori kelas Valid atau *Hoax*. Sehingga diperkonversi data menjadi angka. Tabel 5.2 Pembobotan sample data yang menunjukkan konversi data menjadi angka.

Tabel 5.2 Pembobotan Sample Data

i	yi	x1	x2
1	-1	5	4
2	-1	18	15
3	-1	11	10
4	-1	15	12
5	-1	8	7
6	-1	9	8
7	-1	10	6
8	1	12	11
9	1	8	7
10	1	12	10

i	yi	x1	x2
11	1	9	9
12	1	11	10
13	1	9	9
14	1	9	9

Dari Tabel konversi data dua dimensi dapat memvisualisasi sebagai garis yang memisahkan kelas -1 dari kelas +1. Rumus untuk *Hyperplane* pemisah diberikan oleh $wx+b=0$. Vektor w merupakan vektor bobot yang menentukan kemiringan *Hyperplane* dan koefisien b menentukan jarak dari titik asal. Untuk mendapatkan *Hyperplane* margin terbesar, Dalam hal ini memperkenalkan *Hyperplane* pendukung yang ditempatkan secara simetris di kedua sisi *Hyperplane* pemisah dan mencari vektor w dan b yang memaksimalkan margin. Pada metode SVM dimana ada 2 fitur yaitu fitur kelas Valid dan fitur kelas Hoax. Maka w akan memiliki 2 fitur yaitu (w_1 dan w_2). Persamaan (5.1) yang digunakan dalam memaksimal margin sebagai berikut :

$$\max 2\sigma \text{ dimana } \forall (x_i y_i) y = \left(\frac{wx_i + b}{\|w\|} \right)^N \geq \sigma \quad (5.1)$$

Dari Persamaan (5.1) dijelaskan dalam perhitungan berikut $x_1 = \text{jumlah kata}$ dan $x_2 = \text{Bobot_tf}$ sedangkan kelas Hoax = -1 dan kelas Valid = 1 Berikut hasil pembobotan untuk nilai w_1 dan w_2 dijelaskan pada Tabel 5.3.

Tabel 5.3 Bobot Kata

w1	w2
- 1.25	- 1.00
- 3.08	- 2.57

w1	w2
- 2.75	- 2.50
- 3.75	- 3.00
- 2.00	- 1.75
- 2.25	- 2.00
- 0.04	- 0.02
3.00	2.75
2.00	1.75
3.00	2.50
-	-
2.75	2.50
2.25	2.25
1.58	1.58

Dari Tabel tersebut diperoleh

Nilai w1 = Jumlah perkalian matrix x1

Nilai w2 = Jumlah perkalian matrix x2

$$b = y_1 - w \cdot x_1$$

Langkah berikutnya adalah menghitung nilai parameter untuk setiap fitur yang tersedia. Fungsi ini akan digunakan dalam pembuatan Hyperplane. Model Hyperplane yang digunakan dalam penelitian ini adalah model linear, seperti yang ditunjukkan pada Persamaan (5.4).

$$f(y) = \sum_{i=1}^{total_fitur} w_i \times x_i \quad (5.4)$$

Dalam Persamaan 5.4, w_i adalah bobot dari setiap fitur, dan x_i adalah nilai setiap fitur yang dihasilkan pada tahap vektorisasi TF-IDF. Persamaan ini menggambarkan proses perhitungan bobot untuk setiap fitur dalam model *Hyperplane*.

- a. Jika nilai dari $f(y_i) > 0$ maka $y_i = +1$
- b. Jika nilai dari $f(y_i) < 0$ maka $y_i = -1$

Hasil data latih tersebut akan digunakan dalam mencari nilai *Hyperplane* sebagai batas nilai positif (+1) dan negatif (-1). Dari batas tersebut terdapat jarak atau margin. Mencari nilai margin ada pada Persamaan (5.5):

$$\text{margin} = \frac{2}{\|w\|} \quad (5.5)$$

Dimana a merupakan nilai $f(y_i) > 0$ dan b merupakan nilai $f(y_i) < 0$. Dengan ini fungsi *Hyperplane* dapat dalam Persamaan (5.6) :

$$y = (\sum_{i=1}^{\text{total fitur}} w_i x_i) \geq \text{margin} \quad (5.6)$$

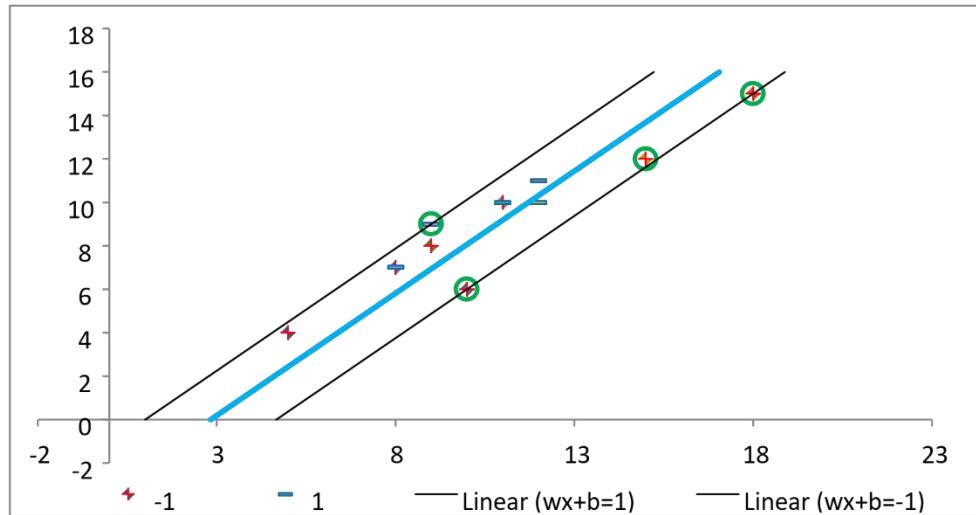
Setelah mengetahui nilai w_1 , w_2 dan nilai b , selanjutnya menentukan *Hyperplane* untuk data tersebut. Dari perhitungan bobot semua data maka nilai *Hyperplane* ditunjukkan dalam Tabel 5.4

Tabel 5.4 Perhitungan *Hyperplane*

<i>Hyperplane</i>			
x2	x1	-1	1
-	2.8	4.67	1.00
1.0	3.7	5.56	1.89

<i>Hyperplane</i>			
2.0	4.6	6.44	2.78
3.0	5.5	7.33	3.67
4.0	6.4	8.22	4.56
5.0	7.3	9.11	5.44
6.0	8.2	10.00	6.33
7.0	9.1	10.89	7.22
8.0	9.9	11.78	8.11
9.0	10.8	12.67	9.00
10.0	11.7	13.56	9.89
11.0	12.6	14.44	10.78
12.0	13.5	15.33	11.67
13.0	14.4	16.22	12.56
14.0	15.3	17.11	13.44
15.0	16.2	18.00	14.33
16.0	17.1	18.89	15.22

Dari hasil perhitungan tersebut, dapat memvisualisasikan *Hyperplane* dalam Gambar 5.1



Gambar 5.1 Visualisasi Metode SVM

Hyperplane berfungsi sebagai batas keputusan yang memisahkan antara fitur-fitur yang mewakili berita Valid dan berita *Hoax*. Jika berita memiliki fitur-fitur yang mendukung kevalidan posisinya akan lebih dekat ke sisi berlawanan dari *Hyperplane*. Sebaliknya, jika berita memiliki fitur-fitur yang menunjukkan ciri-ciri berita *Hoax* posisinya akan lebih dekat ke sisi yang sama dengan berita *Hoax*. Dengan menggunakan konsep *Hyperplane*, membayangkan upaya model untuk menemukan batas yang optimal dalam ruang fitur untuk memisahkan berita Valid dan berita *Hoax*.

5.2 Implementasi Metode Support Vector Machine

Implementasi pada metode *Support Vector Machine* menggunakan pemrograman Python dan beberapa *Library* pendukung yaitu Numpy, Json, Math, sklearn.svm dan menjalankan tahap perencanaan. Pada tahap ini dilakukan pemrosesan implementasi program, yang mana berikut adalah tahap pemrosesannya :

1. Input judul berita. Pada proses implementasi dengan sistem ditunjukan dalam Gambar 5.2

```
def index(): #halaman awal
    source_files = [os.path.splitext(filename)[0] for filename in
        os.listdir('csv_data_source')]
    return render_template('index.html', judul = '', dataset = '',
        datasetv = '', datatest = '', result = {}, resultcsv = [], hasil =
        "-", source_files = source_files, algo = '')
```

Gambar 5.2 Input Judul Berita

Pada implementasi tersebut menjelaskan tentang proses membaca nama file dari direktori 'csv_data_source', kemudian merender template HTML ('index.html') dengan memberikan nilai-nilai default atau kosong untuk beberapa variable, kemudian dari judul yang dimasukan akan dilakukan proses persiapan data atau *Preprocessing*.

2. Melakukan *Preprocessing*. Setelah judul diinputkan, kemudian sistem melakukan *Preprocessing* tahap seperti *Case Folding*, *Stemming*, *Stopword Removal* dan *Tokenazation* Berikut implementasi sistem yang disajikan dalam Gambar 5.3

```

if (request.form['judul']): #proses single text
    rtext_lower = request.form['judul'].lower() #lower case
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    rtext_stemming = stemmer.stem(rtext_lower) # stemming

    stop_factory = StopWordRemoverFactory()
    stopword = stop_factory.create_stop_word_remover()
    rtext_stopword = stopword.remove(rtext_stemming) # stopword

    nltk.download('punkt')
    rtext_token = nltk.word_tokenize(rtext_stopword) #
    tokenizing
        rtext_tcount = {i:rtext_token.count(i) for i in
    rtext_token} # generate term count
        rtext_dataframe_tcount = pd.DataFrame({ 'Judul':
    [json.dumps(rtext_tcount)] })
        rtext_tf = tf(rtext_tcount)
        rtext_dataframe = pd.DataFrame({ 'Judul':
    [json.dumps(rtext_token)] })
        rtext_raw_source = pd.DataFrame({ 'Judul':
    [request.form['judul']] })

```

Gambar 5.3 Proses *Preprocessing*

Pada Gambar tersebut menjelaskan tahap *Preprocessing* data sebelum data digunakan untuk proses pembobota untuk mendapatkan nilai TF IDF.

3. Pembobotan TF IDF

Implementasi sistem pada proses pembobotan TF IDF dijelaskan dalam Gambar 5.4

```

# perhitungan tfidf
dataframe1['Judul'] = dataframe1['Judul'].apply(lambda x:
tfidf(x,idfWord)) #valid
dataframe2['Judul'] = dataframe2['Judul'].apply(lambda x:
tfidf(x,idfWord)) #hoax
rtext_dataframe_tf['Judul']= rtext_dataframe_tf['Judul'].
apply(lambda x: tfidf(x,idfWord)) #hoax
# rtext_tfidf = tfidf(rtext_tf,idfWord)

```

Gambar 5.4 Proses TF IDF

Setiap baris kode diatas menggunakan fungsi TF-IDF (tfidf) untuk menghitung nilai TF IDF dari setiap judul dalam kolom 'Judul' dari DataFrame yang sesuai (dataframe1, dataframe2, rtext_dataframe_tf). Nilai idfWord mungkin berisi nilai-nilai IDF yang telah dihitung digunakan untuk proses hitung klasifikasi dengan metode *Support Vector Machine*.

4. Tahap prediksi menggunakan metode SVM disajikan dalam Gambar 5.5

```

# Support Vector Machine
clf = svm.SVC()
clf.fit(dataset, dataresult)
prediction_svc = clf.predict(datatest)

if (raw_test_filename): #output pembentukan csv
    print(rtext_dataframe_raw)
    rtext_raw =
rtext_dataframe_raw['Judul'].values.tolist() #valid
& hoax
    rtext_raw = list(rtext_raw)

response = {
    "csv": {}
}

```

Gambar 5.5 Prediksi Metode SVM

Membuat objek `clf` sebagai model SVM menggunakan `svm.SVC()`. Melatih model SVM dengan menggunakan `fit`, dengan data pelatihan (dataset) dan label (`dataresult`). Menggunakan model SVM yang sudah dilatih untuk melakukan prediksi terhadap data pengujian (`datatest`). Hasil prediksi disimpan dalam variabel `prediction_svc`. Mengambil data dari kolom 'Judul' dalam DataFrame `rtext_dataframe_raw`. Hasilnya disimpan dalam variabel `rtext_raw` sebagai daftar (list) judul. Jadi, Sistem membuat, melatih, dan menggunakan model SVM untuk melakukan prediksi pada data pengujian. Jika `raw_test_filename` terdefinisi, maka juga mengambil data dari DataFrame untuk keperluan pembentukan file CSV.

Pada tahap selanjutnya pemrosesan implementasi program, yang mana berikut adalah tahap pemrosesannya, input data set latih dan uji. Melakukan *preprocessing*. Melakukan pelatihan data uji dan latih menggunakan klasifikasi yang sudah ditentukan. Dan tahap terakhir adalah pengeluaran berupa hasil perbandingan Akurasi dari algoritma yang dijadikan penelitian, kemudian akan dicatat dan masuk ke tahap verifikasi dan Validasi menggunakan *Confusion Matrix*. Setelah membuat skenario pengujian, selanjutnya dilakukan tahap simulasi menggunakan algoritma yang telah ditentukan, pada tahap ini penulis melakukan uji klasifikasi dengan Algoritma SVM.

Data Latih dan Data Uji

Model data dibuat untuk algoritma klasifikasi menggunakan data latih. Model ini berfungsi sebagai representasi pengetahuan yang digunakan untuk memprediksi kelas data baru yang belum pernah ditemui sebelumnya. Pembagian jumlah data latih dan uji dapat ditemukan dalam Tabel 5.5.

Tabel 5.5 Pembagian Data Latih dan Data Uji

Kelas	Data Latih	Data Uji
Valid	458 Teks	256 Teks
<i>Hoax</i>	1452 Teks	144 Teks
Total	2.000 Teks	400 Teks

Berdasarkan perbandingan jumlah data latih dan uji . Dengan proporsi data dari keseluruhan 2.000 judul berita, melakukan input dataset dengan rasio perbandingan training 90% uji: 10% dengan jumlah data 1800 data latih dengan komposisi 458 data berita Valid, 1342 berita *Hoax* , dan 200 data sebagai data uji.Komposisi perbandingan 80 % : uji 20%, jumlah data 1600 latih dengan komposisi 458 berita Valid dan 1142 berita *Hoax*, dan data uji sejumlah 400 data. Data latih pada perbandingan 70% uji 30%, jumlah data 1400 dengan komposisi 458 berita Valid dan 942 berita *Hoax*, dan 600 data sebagai data uji. Perbandingan 60 % : uji40% dengan jumlah data 1200 dengan komposisi berita 458 Valid dan 742 berita *Hoax* dan 800 data uji. Perbandingan data latih 50% : uji50%. Jumlah data 1000 dengan komposisi 458 berita Valid dan 542 berita *Hoax*, dan 100 data sebagai data uji. Berikut perbandingan data yang digunakan ditunjukan dalam Tabel 5.6

Tabel 5.6 Perbandingan Data

Data Latih	Perbandingan Data				
	50% : 50%	60% : 40%	70% : 30%	80% : 20%	90% : 10%
Berita <i>Hoax</i>	542	742	942	1142	1342
Berita Valid	458	458	458	458	458

Total Data Latih	1000	1200	1400	1600	1800
Total Data Uji	1000	800	600	400	200

Setelah mendapatkan perbandingan data latih dan data uji untuk selanjutnya dilakukan skenario pengujian masing masing perbandingan data. Proses pengujian dijabarkan sebagai berikut :

- Pengujian I

Pada pengujian I pembagian data latih 50% dan data uji 50% didapatkan hasil penjabaran skenario Tabel 5.7 pengujian *Confusion Matrix* sebagai berikut :

Tabel 5.7 Pengujian 50% : 50%

		Actual	
		TP	FN
Prediction	Valid	432	27
	Hoax	29	566

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 432 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 566 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 29 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 27 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian I perbandingan 50% : 50% sebagai berikut

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{432+566}{432+566+29+27}$$

$$\text{Akurasi} = \frac{998}{1054} = 0,946 = 94,6\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{432}{432+29} = \frac{432}{461} = 0,937 = 93,7\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{432}{432+27} = \frac{432}{459} = 0,941 = 94,1\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{432}{432+1/2(29+27)}$$

$$\text{F1 Score} = \frac{432}{432+28} = \frac{432}{600} = 0,72 = 72\%$$

Dari perhitungan terebut mendapatkan hasil *Confusion Matrix* pada Tabel 5.8

Tabel 5.8 Hasil Perhitungan 50% : 50%

Naïve Bayes	Akurasi	Presisi	Recall	F1-Score
60% : 40%	94,6%	93,7%	94,1%	72%

Dari Tabel perhitungan perbandingan data 50% berita Valid dan 50% berita *Hoax* mendapatkan nilai Akurasi sebesar 94,6%, nilai Presisi 93,7% , nilai *Recall* 94,1% dan nilai F1-Score 72%.

- Pengujian II

Pada pengujian II pembagian data latih 60% dan data uji 40% didapatkan hasil penjabaran skenario Tabel 5.9 pengujian *Confusion Matrix* sebagai berikut :

Tabel 5.9 Pengujian 60% : 40%

		Actual	
		TP	FN
Prediction	458	3	
	FP	TN	
	9	329	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 458 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 329 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 9 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 3 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian II perbandingan 60% : 40% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{458+329}{458+329+9+3}$$

$$\text{Akurasi} = \frac{787}{799} = 0,984 = 98,4\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{458}{458+9} = \frac{458}{467} = 0,980 = 98,0\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{458}{458+3} = \frac{432}{459} = 0,993 = 99,3\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{458}{458+1/2(9+3)}$$

$$\text{F1 Score} = \frac{458}{458+6} = \frac{458}{464} = 0,987 = 98,7\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 5.10

Tabel 5.10 Hasil Perhitungan 60% : 40%

Naïve Bayes	Akurasi	Presisi	Recall	F1-Score
60% : 40%	98,4%	98%	99,3%	98,7%

Dari Tabel 5.10 perhitungan perbandingan data 60% berita Valid dan 60% berita *Hoax* mendapatkan nilai Akurasi sebesar 98,4%, nilai Presisi 98% , nilai *Recall* 99,3% dan nilai F1-Score 98,7%.

- Pengujian III

Pada pengujian III pembagian data latih 70% dan data uji 30% didapatkan hasil penjabaran skenario Tabel 5.11 pengujian *Confusion Matrix* sebagai berikut

Tabel 5.11 Pengujian 70% : 30%

		Actual	
		TP	FN
Prediction	16	0	
	FP	TN	
	0	593	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 16 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 593 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian III perbandingan 70% : 30% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{16+593}{432+566+29+27}$$

$$\text{Akurasi} = \frac{609}{609} = 1 = 100\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{16}{16+0} = \frac{16}{16} = 1 = 100\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{16}{16+0} = \frac{16}{16} = 1 = 100\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{16}{16+1/2(0+0)}$$

$$\text{F1 Score} = \frac{16}{16} = 1 = 100\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 5.12

Tabel 5.12 Hasil Perhitungan 70% : 30%

SVM	Akurasi	Presisi	Recall	F1-Score
70% : 30%	100%	100%	100%	100%

Dari Tabel 6.12 perhitungan perbandingan data 70% berita Valid dan 30% berita *Hoax* mendapatkan nilai Akurasi sebesar 100%, nilai Presisi 100% , nilai Recall 100% dan nilai F1-Score 100%.

- Pengujian IV

Pada pengujian IV pembagian data latih 80% dan data uji 20% didapatkan hasil penjabaran skenario Tabel 5.13 pengujian *Confusion Matrix* sebagai berikut :

Tabel 5.13 Pengujian 80% : 20%

		Actual	
		TP	FN
Prediction	16	1	
	FP	TN	
	0	433	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 16 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 433 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 1 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian IV perbandingan 80% : 20% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{16+443}{16+443+0+0}$$

$$\text{Akurasi} = \frac{449}{450} 0,997 = 99,7\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{16}{16+0} = \frac{16}{16} = 1 = 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{16}{16+1} = 0,941 = 94,1\%$$

$$\text{F1 Score} = \frac{TP}{TP+1/2(FP+FN)}$$

$$\text{F1 Score} = \frac{16}{16+1/2(0+1)}$$

$$\text{F1 Score} = \frac{16}{16,5} = 0,969 = 96,9\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 5.14

Tabel 5.14 Hasil Perhitungan 80% : 20%

Naïve Bayes	Akurasi	Presisi	Recall	F1-Score
80% : 20%	99,7%	100%	94,1%	96,9%

Dari Tabel perhitungan perbandingan data 80% berita Valid dan 20% berita *Hoax* mendapatkan nilai Akurasi sebesar 99,7%, nilai Presisi 100% , nilai Recall 94,1% dan nilai F1-Score 6,9%.

- Pengujian V

Pada pengujian V pembagian data latih 90% dan data uji 10% didapatkan hasil penjabaran skrenario Tabel 5.14 pengujian *Confusion Matrix* sebagai berikut :

Tabel 5.15 Pengujian 90% : 10%

		Actual	
		TP	FN
Prediction	13	1	
	FP	TN	
	0	221	

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 13 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 221 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Menghitung Akurasi dari skenario pengujian V perbandingan 90% : 10% sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Akurasi} = \frac{13+221}{13+221+0+0}$$

$$\text{Akurasi} = \frac{234}{234} = 1 = 100\%$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Presisi} = \frac{13}{13+0} = \frac{13}{13} = 1 = 100\%$$

$$Recall = \frac{TP}{TP+FN}$$

$$Recall = \frac{13}{13+0} = \frac{13}{13} = 1 = 100\%$$

$$F1\ Score = \frac{TP}{TP+1/2(FP+FN)}$$

$$F1\ Score = \frac{13}{13+1/2(0+0)}$$

$$F1\ Score = \frac{13}{13} = 1 = 100\%$$

Dari perhitungan tersebut mendapatkan hasil *Confusion Matrix* pada Tabel 5.16

Tabel 5.16 Hasil Perhitungan 90% : 10%

SVM	Akurasi	Presisi	Recall	F1-Score
90% : 10%	100%	100%	100%	100%

Dari Tabel 5.15 perhitungan perbandingan data 90% berita Valid dan 10% berita *Hoax* mendapatkan nilai Akurasi sebesar 100%, nilai Presisi 100% , nilai *Recall* 100% dan nilai F1-Score 100%.

Dari hasil pengujian yang telah dilakukan sebanyak 5 kali pengujian didapatkan Akurasi, Presisi, *Recall* seperti ditampilkan pada Tabel 5.16

Tabel 5.17 Akurasi, Presisi, Recall dan F1-Score metode SVM

Pengujian	Latih : Uji	Support Vector Machine			
		Akurasi	Presisi	Recall	F1-Score
Pengujian I	50:50	64,6%	93,7%	94,1%	72%
Pengujian II	60:40	98,4%	98%	99,3%	98,7%
Pengujian III	70:30	100%	100%	100%	100%
Pengujian IV	80:20	99,7%	100%	94,1%	100%
Pengujian V	90:10	100%	100%	100%	100%
Rata - Rata		100%	99%	100%	91%

Pada Tabel terlihat bahwa metode SVM mampu melakukan deteksi pada berita *Hoax*. Dengan 5 kali uji coba dengan melakukan perubahan pada data latih dan uji menunjukkan nilai rata-rata Akurasi sebesar **100%**, Presisi sebesar **99%**, *Recall* sebesar **100%** dan F1-Score sebesar **91%**.

BAB VI

PEMBAHASAN

Penelitian ini menggunakan data kualitatif, yang memberikan informasi tentang atribut atau kategori dari suatu informasi. Sebanyak 80% dari data diambil dari kategori *Hoax*, sementara 20% berasal dari kategori Valid. Sebelum diterapkan menggunakan metode *Naive Bayes* dan SVM, data harus melewati beberapa tahap *Preprocessing*. Setelah mendapatkan dataset, data dibagi menjadi dua bagian: data latih dan data uji. Terdapat 2.000 data teks dalam data latih, dengan 458 berita Valid dan 1.452 berita *Hoax*. Untuk data uji, terdapat 256 berita Valid dan 114 berita *Hoax*. Model data dibentuk untuk algoritma klasifikasi pada data latih, yang berfungsi sebagai representasi pengetahuan untuk memprediksi data kelas baru yang belum pernah ditemui sebelumnya. Pembagian jumlah data latih dan uji dapat dilihat pada Tabel 6.1.

Tabel 6.1 Pembagian Data Latih dan Data Uji

Kelas	Data Latih	Data Uji
Valid	458 Teks	256 Teks
<i>Hoax</i>	1452 Teks	114 Teks
Total	2.000 Teks	400 Teks

Berdasarkan Tabel 6.1 diketahui perbandingan jumlah data training dan testing. Dengan proporsi perbandingan sebesar 80% : 20% dari keseluruhan 2.000 judul berita, melakukan input data set dengan rasio perbandingan latih 90% uji : 10% training, 80 % : uji 20%, latih 70% : uji 30%, latih 60 % : uji 40% dan latih 50% : uji 50%. Berikut perbandingan data yang digunakan ditunjukan dalam Tabel 6.2.

Tabel 6.2 Data Latih dan Data Uji

Data Latih	Perbandingan Data				
	50% : 50%	60% : 40%	70% : 30%	80% : 20%	90% : 10%
Berita <i>Hoax</i>	542	742	942	1142	1342
Berita Valid	458	458	458	458	458
Total Data Latih	1000	1200	1400	1600	1800
Total Data Uji	1000	800	600	400	200

Perbandingan jumlah data training dan testing. Dengan proporsi data dari keseluruhan 2.000 judul berita, melakukan input data set dengan rasio perbandingan latih 90% uji : 10% dengan jumlah data 1800 data training dengan komposisi 458 data berita Valid, 1342 berita *Hoax*, dan 200 data sebagai data uji. Komposisi perbandingan 80 % : uji 20%, jumlah data 1600 training dengan komposisi 458 berita Valid dan 1142 berita *Hoax*, dan data uji sejumlah 400 data. Data latih pada perbandingan 70% uji 30%, jumlah data 1400 dengan komposisi 458 berita Valid dan 942 berita *Hoax*, dan 600 data sebagai data uji. Perbandingan 60 % : uji 40% dengan jumlah data 1200 dengan komposisi berita 458 Valid dan 742 berita *Hoax* dan 800 data uji. Perbandingan data latih 50% : uji 50%. Jumlah data 1000 dengan komposisi 458 berita Valid dan 542 berita *Hoax*, dan 1000 data sebagai data uji.

6.1 Hasil Pengujian Algoritma Naïve Bayes

Menggunakan lima skenario pengujian dan menghasilkan nilai Akurasi tertinggi pada pengujian ketiga dengan menggunakan skala uji 70% : 30% dan skala uji 90% dan 10%. Perbandingan data dengan 1400 data latih dan 600 data uji, dan perbandingan data 1800 data latih dan 200 data uji. Nilai dari skala uji

tersebut mencapai nilai Akurasi , Presisi, *Recall* dan *F1-Score* 100%. Pada Tabel 6.3 hasil prediksi dan hasil actual dari klasifikasi berita *Hoax* mendapatkan hasil sebagai berikut :

Tabel 6.3 Hasil Perbandingan 70%:30%

		Hasil Actual	
Hasil Prediksi		Kelas Valid	Kelas Hoax
	Kelas Valid	149	0
	Kelas Hoax	0	460

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 149 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 460 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Selanjutnya perbandingan skala 90% : 10% menghasilkan nilai Akurasi , Presisi, *Recall* dan *F1 Score* 100%. Pada Tabel 6.4 hasil prediksi dan hasil actual dari klasifikasi berita *Hoax* mendapatkan hasil sebagai berikut :

Tabel 6.4 Hasil Perbandingan 90%:10%

		Hasil Actual	
Hasil Prediksi		Kelas Valid	Kelas Hoax
	Kelas Valid	30	0
	Kelas Hoax	0	204

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 30 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 204 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*.

Dari kedua skala perbandingan pengujian mendapatkan Akurasi tertinggi pada pengujian data dengan perbandingan 70% data pelatihan dan 30% data uji dibandingkan dengan perbandingan 90% dan 10%, ada beberapa faktor yang mungkin berkontribusi pada hasil tersebut. Berikut beberapa kemungkinan seperti data uji yang lebih besar (30%) memberikan representasi yang lebih baik dari variasi yang mungkin ada di dalam data. Dataset cukup besar, membagi lebih banyak data untuk pengujian dapat membantu metode mendapatkan gambaran yang lebih baik tentang seberapa baik dapat melakukan generalisasi atau deteksi data. Pembagian data latih dan data uji secara seimbang sangat mempengaruhi dalam meningkatkan performa dari pengujian metode yang sudah dilakukan.

Hasil perhitungan *F1 Score* dari perhitungan *Confusion Matrix* dengan nilai $F1\ Score = 100\%$. Untuk keseluruhan hasil pengujian ddeteksi berita *Hoax* dengan menggunakan algoritma *Naïve Bayes* di jelaskan pada Tabel 6.5 berikut ini

Tabel 6.5 Hasil Pengujian Naïve Bayes

Judul Berita	Jumlah Perbandingan		Jumlah Judul Berita		Naïve Bayes			
	Data Latih	Data Uji	Data Latih	Data Uji	Akurasi	Presisi	Recall	<i>F1-Score</i>
	50	50	1000	1000	80%	100%	80%	100%
	60	40	1200	800	88,3%	100%	83,1%	90,7%
2000	70	30	1400	600	100%	100%	100%	100%

Judul Berita	Data Latih	Data Uji	Data Latih	Data Uji	Akurasi	Presisi	Recall	F1-Score
	80	20	1600	400	99,7%	100%	98,8%	99,4%
	90	10	1800	200	100%	100%	100%	100%
Rata - Rata					93%	100%	93%	100%

Pada Tabel 6.5 perbandingan hasil pengujian menyatakan pengamatan uji coba yang telah dilakukan 5 kali uji data yaitu dengan merubah prosentase data latih mendapatkan hasil yang menyatakan bahwa prosentase pembagian data latih dan data uji sangat mempengaruhi hasil Akurasi , Presisi, *Recall* dan *F1 Score* dalam proses pengujian. Pada hasil pengujian yang dilakukan menunjukkan bahwa nilai Akurasi tertinggi didapat pada pengujian ke 3 dan ke 5 dengan perbandingan skala pengujian 70% : 30% dan skala pengujian 90% : 10% dengan menghasilkan Akurasi 100%, Presisi 100%, *Recall* 100%, serta rata-rata Akurasi 100% dan nilai rata – rata dari kelima pengujian Akurasi **93%**, Presisi **100%**, *Recall* **100%** dan F1-Score **100%**. Dapat dijelaskan bahwa mendapatkan hasil dari uji coba menyatakan prosentase pembagian data latih dan data uji sangat mempengaruhi hasil Akurasi , Presisi, dan *Recall* dalam proses pengujian.

6.2 Pengujian Alogirtma Support Vector Machine

Menggunakan lima skenario pengujian dan menghasilkan nilai Akurasi tertinggi pada pengujian ketiga dengan menggunakan skala uji 70% : 30% dan skala uji 90% dan 10%. Perbandingan data dengan 1400 data latih dan 600 data uji, dan perbandingan data 1800 data latih dan 200 data uji. Nilai dari skala uji tersebut mencapai nilai Akurasi , Presisi, *Recall* dan *F1-Score* 100%. Pada Tabel 6.6 hasil prediksi dan hasil actual dari klasifikasi berita *Hoax* mendapatkan hasil

Tabel 6.6 Hasil Perbandingan 70%:30%

Hasil Prediksi		Hasil Actual	
		Kelas Valid	Kelas Hoax
		Kelas Valid	0
	Kelas Hoax	0	593

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 16 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 539 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai berita bukan Valid atau berita *Hoax*. Selanjutnya perbandingan skala 90% : 10% menghasilkan nilai Akurasi , Presisi, *Recall* dan *F1 Score* 100%. Pada Tabel 6.7 hasil prediksi dan hasil actual dari klasifikasi berita *Hoax* mendapatkan hasil sebagai :

Tabel 6.7 Hasil Perbandingan 90%:10%

Hasil Prediksi		Hasil Actual	
		Kelas Valid	Kelas Hoax
		Kelas Valid	0
	Kelas Hoax	0	221

Pada hasil *True Positive* (TP) metode berhasil mengidentifikasi 13 dari data berita Valid yang sebenarnya sebagai berita Valid. *True Negative* (TN) berhasil mengidentifikasi 221 data berita kelas berita *Hoax* sebagai kelas semestinya atau kelas berita *Hoax*. Hasil *False Positive* (FP) dengan jumlah 0 data diasumsikan sebagai kelas berita *Hoax* salah diklasifikasikan kelas berita Valid. Kemudian, *False Negatif* (FN) sejumlah 0 data berita Valid yang diklasifikasikan sebagai

berita bukan Valid atau berita *Hoax*.

Dari kedua skala perbandingan pengujian mendapatkan Akurasi tertinggi pada pengujian data dengan perbandingan 70% data pelatihan dan 30% data uji dibandingkan dengan perbandingan 90% dan 10%, ada beberapa faktor yang mungkin berkontribusi pada hasil tersebut. Berikut beberapa kemungkinan seperti data uji yang lebih besar (30%) memberikan representasi yang lebih baik dari variasi yang mungkin ada di dalam data. Dataset cukup besar, membagi lebih banyak data untuk pengujian dapat membantu metode mendapatkan gambaran yang lebih baik tentang seberapa baik dapat melakukan generalisasi atau deteksi data. Pembagian data latih dan data uji secara seimbang sangat mempengaruhi dalam meningkatkan performa dari pengujian metode yang sudah dilakukan.

Hasil perhitungan *F1 Score* dari perhitungan *Confusion Matrix* dengan nilai *F1 Score* = 100%. Untuk keseluruhan hasil pengujian deteksi berita *Hoax* dengan menggunakan algoritma SVM di jelaskan pada Tabel 6.8 berikut :

Tabel 6.8 Hasil Pengujian SVM

Judul Berita	Jumlah Perbandingan		Jumlah Judul Berita		Support Vector Machine			
	Data Latih	Data Uji	Data Latih	Data Uji	Akurasi	Presisi	Recall	F1-Score
	50	50	1000	1000	64,6%	93,7%	94,1%	72%
	60	40	1200	800	98,4%	98%	99,3%	98,7%
2000	70	30	1400	600	100%	100%	100%	100%
	80	20	1600	400	99,7%	100%	94,1%	100%
	90	10	1800	200	100%	100%	100%	100%
Rata - Rata					100%	99%	100%	91%

Perbandingan hasil pengujian menyatakan pengamatan uji coba yang telah dilakukan 5 kali uji data yaitu dengan merubah prosentase data latih mendapatkan hasil yang menyatakan bahwa prosentase pembagian data latih dan data uji sangat mempengaruhi hasil Akurasi , Presisi, Recall dan *F1 Score* dalam proses

pengujian. Pada hasil pengujian yang dilakukan menunjukkan bahwa nilai Akurasi tertinggi didapat pada pengujian ke 3 dan ke 5 dengan perbandingan skala pengujian 70% : 30% dan skala pengujian 90% : 10% dengan menghasilkan Akurasi 100%, Presisi 100%, *Recall* 100%, serta rata-rata Akurasi 100% mendapatkan hasil dari uji coba menyatakan bahwa prosentase pembagian data latih dan data uji sangat mempengaruhi hasil Akurasi , Presisi, dan *Recall* dalam proses pengujian.

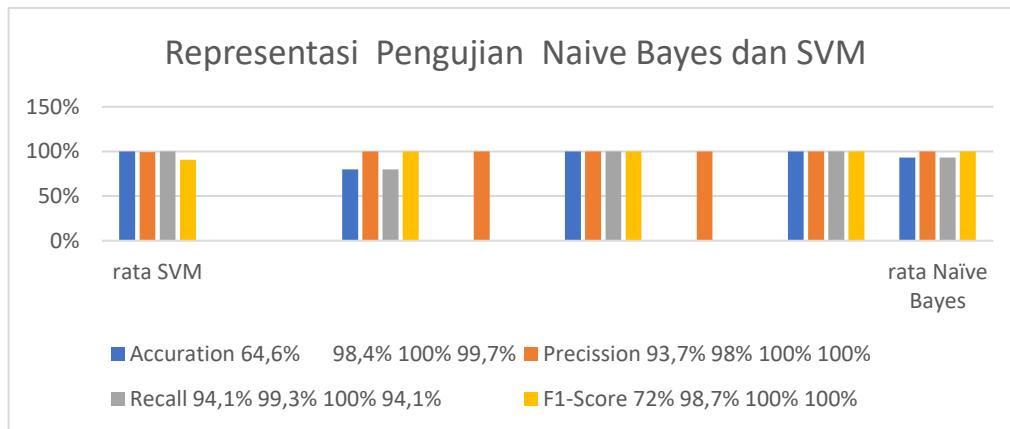
6.3 Hasil Pengujian Naïve Bayes dan Support Vector Machine

Dari hasil pengujian masing-masing metode yang dilakukan sebanyak 5 kali pengujian terhadap masing-masing algoritma, maka selanjutkan dilakukan perbandingan antara *Support Vector Machine* dengan metode *Naïve Bayes* seperti pada Tabel 6.9 sebagai berikut

Tabel 6.9 Hasil Keseluruhan Metode

Pengujian		Naïve Bayes				Support Vector Machine			
		Akurasi	Presisi	Recall	F1-Score	Akurasi	Presisi	Recall	F1-Score
Pengujian I	50%:50%	80%	100%	80%	100%	64,6%	93,7%	94,1%	72%
Pengujian II	60%:40%	88,3%	100%	83,1%	90,7%	98,4%	98%	99,3%	98,7%
Pengujian III	70% : 30%	100%	100%	100%	100%	100%	100%	100%	100%
Pengujian IV	80% : 20%	99,7%	100%	98,8%	99,4%	99,7%	100%	94,1%	100%
Pengujian V	90% :10%	100%	100%	100%	100%	100%	100%	100%	100%
Rata - Rata		93%	100%	93%	100%	100%	99%	100%	91%

Berdasarkan Tabel tersebut menunjukkan perbandingan data latih sangat mempengaruhi nilai Akurasi , Presisi dan *Recall* dari kedua algoritma. Dari hasil ujicoba yang dilakukan deteksi berita *Hoax* menggunakan algoritma SVM dan *Naïve Bayes* bahwa kedua algoritma tersebut dapat direkomendasikan sebagai metode deteksi berita *Hoax*. Hasil pengujian menunjukkan bahwa semakin bertambah data training maka semakin besar pula nilai Akurasi nya. Hasil Akurasi , Presisi, *Recall* keseluruhan dapat dilihat pada grafik seperti pada Gambar 6.1 di bawah ini



Gambar 6.1 Hasil Keseluruhan Pengujian

Pada Gambar 6.1 Algoritma *Naïve Bayes* hasil rata – rata *Naïve Bayes* Akurasi **93%**, Presisi **100%**, *Recall* **93%** dan *F1 Score* **100%**, hasil rata – rata metode SVM Akurasi **100%**, Presisi **99%**, *Recall* **100%** dan F1-score **91%**. Dapat dilihat bahwa dalam melakukan deteksi berita *Hoax* nilai *F1 Score* algoritma *Naïve Bayes* lebih tinggi dibandingkan algortima SVM, menunjukkan bahwa dalam penelitian deteksi berita *Hoax* menggunakan algoritma *Naïve Bayes* lebih unggul.

Berikut integrasi dalam Islam yang berhubungan dengan performa dan Akurasi data dalam mengambil keputusan atau timbangan amal kebaikan, seperti dijelaskan dalam Al Qur'an surah Al Kahfi ayat 49 sebagai berikut

وَوُضِعَ الْكِتَبُ فَتَرَى الْمُجْرِمِينَ مُشْفِقِينَ مِمَّا فِيهِ وَيَقُولُونَ يَا وَيَأْتِنَا مَالَ هَذَا الْكِتَبِ لَا يُغَادِرُ صَغِيرًاٌ
وَلَا كَبِيرًا إِلَّا أَحْصَاهَا وَوَجَدُوا مَا عَمِلُوا حَاضِرًاٌ وَلَا يَظْلِمُ رَبُّكَ أَحَدًا٤٩

“Dan catatan ‘perbuatan’ akan dibuka ‘terbuka’, dan kamu akan melihat orang-orang fasik ketakutan terhadap apa yang ‘tertulis’ di dalamnya. Mereka akan berseru, “Celakalah kami! Catatan macam apakah yang tidak meninggalkan dosa, kecil atau besar; tidak tercatat?” Mereka akan menemukan apa pun yang mereka lakukan hadir ‘di hadapan mereka’. Dan Tuhanmu tidak akan pernah menganiaya siapa pun” (QS. Al Kahfi 49)

Dalam ayat tersebut, menurut tafsir Departemen Agama RI Jilid 10 tahun 2008, Allah SWT memberikan keterangan tentang peristiwa di hari kiamat. Setiap orang akan menerima buku catatan amal perbuatannya selama hidup di dunia. Isi catatan tersebut mencakup perbuatan baik dan buruk, yang ditampilkan dari sebelah kanan atau sebelah kiri. Orang-orang mukmin dan beramal saleh menerima catatan dari sebelah kanan, dan setelah melihat isinya, mereka menyadari bahwa kebaikan mereka lebih besar daripada kejahatan. Allah segera mengampuni kejahatan tersebut, dan mereka diberi tempat di surga. Ayat tersebut menjelaskan bahwa semua perbuatan dan perkataan manusia selama hidup dicatat dalam buku catatan pribadinya, yang akan ditunjukkan pada Hari Perhitungan. Ini menegaskan bahwa setiap kebaikan dan keburukan akan tercatat secara rinci, dan catatan ini akan menentukan nasib seseorang, apakah masuk surga atau neraka. Ayat tersebut mengandung pesan, yaitu penekanan pada nilai-nilai kebaikan dan keburukan yang menjadi dasar pengambilan keputusan.

Dalam Surah As Sajdah ayat 5 secara implisit menyebutkan perhitungan dengan analisis tentang perhitungan metode :

يُدَبِّرُ الْأَمْرُ مِنَ السَّمَاءِ إِلَى الْأَرْضِ ثُمَّ يَعْرُجُ إِلَيْهِ فِي يَوْمٍ كَانَ مِقْدَارُهُ الْأَلْفُ سَنَةً مِمَّا تَعَدُّونَ ٥

“Dialah yang mengatur segala urusan mulai dari langit hingga bumi, lalu semuanya itu naik kepada-Nya pada suatu hari yang lamanya seribu tahun menurut perhitungannya.” (QS. As Sajdah:5)

Dalam ayat tersebut, menurut tafsir Departemen Agama RI Jilid 10 tahun 2008, dijelaskan bahwa Allah-lah yang mengatur, mengurus, dan menentukan segala yang ada di dunia ini. Semua peristiwa terjadi sesuai dengan kehendak dan ketetapan-Nya, tanpa ada yang melenceng dari kehendak-Nya. Pengaturan Allah mencakup dari langit hingga bumi, dan akhirnya semua urusan kembali kepada-Nya. Ayat ini memberikan gambaran tentang kebesaran dan kekuasaan Allah agar manusia dapat memahaminya dengan mudah. Selanjutnya, ayat menggambarkan bahwa Allah mengurus, mengatur, dan menyelesaikan urusan alam semesta ini dalam waktu sehari. Meskipun disebut "sehari," waktu ini setara dengan seribu tahun menurut ukuran waktu yang dikenal manusia di dunia ini. Pemakaian bilangan seribu tidak selalu mengacu pada nilai sebenarnya, tetapi kadang-kadang digunakan untuk menyatakan banyaknya sesuatu atau lamanya waktu yang diperlukan. Dalam konteks ini, seribu tahun digunakan untuk menggambarkan lamanya kehidupan alam semesta, mulai dari penciptaannya hingga kehancuran di hari Kiamat, ketika semua urusan akan kembali kepada Allah. Waktu yang sangat panjang ini sulit dihitung oleh manusia. Ayat ini juga menggunakan istilah ribuan untuk menjelaskan lamanya waktu yang dibutuhkan jika manusia ingin naik menghadap Allah, sementara para malaikat hanya memerlukan sehari. Ayat ini mengajarkan pentingnya menghitung waktu secara tepat dengan rumus dan metode yang berlaku.

BAB VII

KESIMPULAN

7.1 Kesimpulan

Hasil penelitian ini menunjukkan bahwa setelah menggunakan algoritma SVM dan *Naive Bayes* untuk mengklasifikasikan berita *Valid* dan *Hoax*, diperoleh hasil bahwa kinerja *Naive Bayes* lebih unggul dibandingkan SVM. Terbukti dari data pengujian yang digunakan *Naive Bayes* mempunyai nilai *F1 Score* sebesar 100%, dan SVM memiliki nilai sebesar 91%. Berdasarkan hasil penelitian dapat disimpulkan bahwa model ini dapat dijadikan acuan ketika melanjutkan dengan model deteksi berita *Hoax*. Dengan hasil rata – rata *Naive Bayes* Akurasi **93%**, Presisi **100%**, *Recall* **93%** dan *F1 Score* **93%**, hasil rata – rata metode SVM Akurasi **100%**, Presisi **99%**, *Recall* **100%** dan *F1 Score* **91%**. Kontribusi dalam penelitian ini mengembangkan metode validasi yang kuat untuk mengukur kinerja metode, termasuk menggunakan kumpulan data dengan jumlah yang cukup besar, data pengujian yang lebih beragam dan metrik evaluasi yang lebih komprehensif.

7.2 Saran

Saran berikutnya adalah memperluas dataset dengan memasukkan data dari berbagai bahasa dan meningkatkan model *Preprocessing* dokumen. Saat ini, penelitian hanya menggunakan *Preprocessing* tertentu dalam bahasa Indonesia, Disarankan untuk menyertakan metadata dengan mengintergrasikan analisis gambar, video bersama dengan teks, dan juga disarankan untuk melanjutkan dengan penerapan model algoritma pembelajaran mendalam dan metode pengujian yang lebih beragam.

DAFTAR PUSTAKA

- Abid, A., Ali, W., Farooq, M.S., Farooq, U., Khan, N.S., Abid, K., 2020. SemiAutomatic Classification and Duplicate Detection From Human Loss News Corpus. *IEEE Access* 8, 97737–97747. <https://doi.org/10.1109/ACCESS.2020.2995789>
- Bhalla, S., Kwan, P., Bedekar, M., Phalnikar, R., Sirsikar, S. (Eds.), 2020. Proceeding of International Conference on Computational Science and Applications: ICCSA 2019, Algorithms for Intelligent Systems. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-15-0790-8>
- Dey, A., Rafi, R.Z., Hasan Parash, S., Arko, S.K., Chakrabarty, A., 2018. Fake News Pattern Recognition using Linguistic Analysis, in: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (IcIVPR). Presented at the 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), IEEE, Kitakyushu, Japan, pp. 305–309. <https://doi.org/10.1109/ICIEV.2018.8641018>
- Elsaeed, E., Ouda, O., Elmogy, M.M., Atwan, A., El-Daydamony, E., 2021. Detecting Fake News in Social Media Using Voting Classifier. *IEEE Access* 9, 161909– 161925. <https://doi.org/10.1109/ACCESS.2021.3132022>
- Gilda, S., 2017. Evaluating Machine Learning Algorithms for Fake News Detection. th Student Conference on Research and Development 6.
- F. Ferdiansyah and Wella., 2022. "Comparative Study of *Hoax* Detection using Support Vector Machine Algorithm, Naive Bayes, Random Forest and K-Nearest Neighbor," 13th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, Republic of, 2022, pp. 1123-1126, doi: 10.1109/ICTC55196.2022.9952630.
- Yuditha, C., n.d., 2016. Interaksi Komunikasi Hoax di Media Sosial serta Antisipasinya Hoax Communication Interactivity in Social Media and Anticipation 3, 14. 23 Li, Z., Shang, W., Yan, M.,. News text classification model based on topic model, in: 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). Presented at the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), IEEE, Okayama, Japan, pp. 1–5. <https://doi.org/10.1109/ICIS.2016.7550929>

- Kousika, N., S, Deepa., C, Deepika., B M, Dhatchaiyine., J, Amrutha., 2021. A System for Fake News Detection by using Supervised Learning Model for Social Media Contents, in: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). Presented at the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, Madurai, India, pp. 1042–1047. <https://doi.org/10.1109/ICICCSS51141.2021.9432096>.
- Liu, C., Sheng, Y., Wei, Z., Yang, Y.-Q., 2018. Research of Text Classification Based on Improved TF-IDF Algorithm, in: 2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE). Presented at the 2018 IEEE International Conference of Intelligent Robotic and Control Engineering(IRCE),IEEE,Lanzhou,pp.218–222. <https://doi.org/10.1109/IRCE.2018.8492945>
- Nagel, T.W.S., 2022. Measuring fake news acumen using a news media literacy instrument. *JMLE* 14, 29–42. <https://doi.org/10.23860/JMLE-2022-14-1-3>
- Ozbay, F.A., Alatas, B., 2019. Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical MechanicsanditsApplications* 540, 123174. <https://doi.org/10.1016/j.physa.2019.123174>
- Prasetya, F., Ferdiansyah, F., 2022. Analisis Data Mining Klasifikasi Berita Hoax COVID 19 Menggunakan Algoritma Naive Bayes. *json* 4, 132. <https://doi.org/10.30865/json.v4i1.4852>
- P. Meel and D. K. Vishwakarma., 2021. "Machine Learned Classifiers for Trustworthiness Assessment of Web Information Contents," International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 29-35, doi: 10.1109/ICCCIS51004.2021.9397228.
- P. S. Vadlamudi, M. Gunasekaran and T. J. Nagalakshmi, "An Analysis of the Effectiveness of the Naive Bayes Algorithm and the *Support Vector Machine* for Detecting Fake News on Social Media," 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 2023, pp. 726-731, doi: 10.1109/IITCEE57236.2023.10090978.
- S. Raza and C. Ding., 2022. "Fake news detection based on news content and social contexts: a transformer-based approach," *Int J Data Sci Anal*, vol. 13, no. 4, pp. 335–362, May, doi: 10.1007/s41060-021-00302-z.

- Umer, M., Imtiaz, Z., 2020. Ullah, S., Mehmood, A., Choi, G.S., On, B.-W., 2020. Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). *IEEEAccess*8,156695–156706. <https://doi.org/10.1109/ACCESS.2020.3019735> 24
- Yanuar Risca Pratiwi, A. Ferdita Nugraha, Y. Pristyanto, R. Faticha Alfa Aziza, J. Kuswanto and I. Hadi Purwanto., 2022. "Machine Learning Model for Detecting Fake News Content in Indonesian-Language Online Media," International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, pp. 302-307, doi: 10.1109/ICIMCIS56303.2022.10017528.
- Venkatesh, Ranjitha, K.V., 2018. Classification and Optimization Scheme for Text Data using Machine Learning Naïve Bayes Classifier, in: 2018 IEEE World Symposium on Communication Engineering (WSCE). Presented at the 2018 IEEE World Symposium on Communication Engineering (WSCE), IEEE, Singapore,Singapore,pp.33–36. <https://doi.org/10.1109/WSCE.2018.8690536>
- V. B. Kusnandar,2021. “Pengguna Internet Indonesia Peringkat ke-3 Terbanyak di Asia,” databoks.katadata.co.id

LAMPIRAN – LAMPIRAN

Lampiran I Data Uji Berita *HOAX*

NO	JUDUL BERITA <i>HOAX</i>
1	angola resmi larang islam di negaranya
2	wasekjen mui, kh. tengku zulkarnain sebut ada pembakaran masjid di pangkalan brandan
3	produk toyota segera resmi hadir di indonesia?
4	tim putri indo vs tim putri rusia
5	bahaya memakan kepala ikan karena mengandung racun ciguatoksin
6	fetus soup eaten by asians? sop bayi dimakan oleh orang-orang asia?
7	ustadz abdul somad dihadiah mobil alphard
8	siswa-siswi smp n 12 berau kalimantan timur belajar sambil lesehan di lantai
9	film naura & genk juara lecehkan islam
10	jokowi dikatakan seperti anak kecil dan bodoh oleh media luar
11	pengusiran petani oleh pasukan barakuda
12	penculikan anak gelandangan di brebes
13	ada hio atau dupa di ruang sidang ahok yang dikaitkan dengan mistis
14	foto gus nuril dalam berita kompastv yang berjudul ummat islam curang mereka menggunakan sihir agar peserta aksi 212 dari ciamis agar terlihat banyak
15	raja salman danai kampanye pm israel netanyahu
16	foto pandemi flu 1918
17	kompas: infodemi membuat publik abai dengan bahaya covid-19
18	pidato lengkap jokowi, negara kita tidak pernah berutang, hanya melakukan pinjaman luar negeri
19	gambar dalam bayang-bayang luhut sampul majalah tempo
20	video sholat jumatan di china menjadi lautan manusia, orang islam di wuhan tidak satupun yang kena virus corona
21	dosen kedokteran gigi ui meninggal karena covid-19
22	polres pamekasan tidak selenggarakan kegiatan nikah gratis

NO	JUDUL BERITA HOAX
23	sebelum di setrum mati, permintaan pria ini membuat surat wasiat pada ibunya
24	jokowi minta maaf kepada pk1
25	ukiran sudah ratusan tahun
26	diam-diam david beckham ke wamena-papua
27	babi hutan melahirkan 3 anak mirip bayi manusia
28	merdeka.com membuat dan menayangkan berita ada wanita membawa anjing masuk masjid, polisi sebut pelaku alami gangguan jiwa, jangan dibesar-besarkan!!
29	foto capres jokowi mengenakan jaket yang berisi banyak kartu
30	karena sepi penonton, teman ahok diberi tiket gratis dan uang rp. 50 ribu agar menonton film a man called ahok
31	zumi zola jalan-jalan di bandara soekarno-hatta
32	pelaku bom surabaya pernah bentangkan #2019gantipresiden
33	program gebyar undian pt. whatsapp indonesia 2018
34	ternyata nkri sudah mati harga
35	kpk bantah keluarkan surat edaran terhadap universitas sam ratulangi
36	ternyata menu telah dipesan sebelumnya!
37	2 petani di rokan hilir tewas tersambar petir saat berteduh di pohon sawit
38	video habib rizieq disambut simpatisan di bandara bukan video kepulangannya ke indonesia
39	pers rilis sekretariat presiden menyebut sri mulyani menteri antek aseng
40	: beredar tarif tol bakauheni lampung
41	benarkah subsidi gas 3 kg dicabut, ini klarifikasi kementerian esdm
42	penangkapan kendaraan sepeda motor besar-besaran di polsek kenjeran
43	indonesia dijajah belanda selama 350 tahun, benarkah?
44	dari anaknya owner *_behaestex*_ pabrik sarung atlas,rubat dll : ...
45	: kesalahan penulisan :fitsa hats di bap kepolisian
46	seorang gadis palestina ditembak oleh 10 laki-laki warga israel

NO	JUDUL BERITA HOAX
47	akhirnya 3 bank besar indonesia berhutang besar ke cina, apa selanjutnya ?
48	kata-kata terakhir steve jobs
49	jokowi akan dibaptis ulang dan mengganti nama menjadi hebartus
50	maruf amin ngaku kalo grombolan mereka adalah kodok
51	foto kondisi tenggorokan pasien positif covid-19
52	saddam hussein mengatakan amerika mengancam irak dengan virus corona
53	perkembangan chloroquine untuk obat covid-19
54	selamat datang tka china sebarkan virus corona #tergagapcorona
55	deddy corbuzier berikan mobil untuk shinta nuriyah
56	foto mesut ozil menginjak bendera tionkok
57	foto ahok dipeluk dan dicium dua wanita, alasan alexis dipertahankan
58	menhan: wajib militer ampuh menangkal bangkitnya pki
59	mobile phone burning while charging using powerbank in emporium mall
60	gerindra: pak prabowo sudah mengirim utusan untuk kepulangan habib rizieq shihab
61	demontran mengencingi polisi
62	ambulans dki tidak membawa batu dan bensin
63	unggah foto kejadian tahun 2016 orang utan mati terpanggang, warga aceh minta maaf
64	mahasiswa unsri palembang meninggal dunia dalam aksi hari ini
65	habib bahar smith di gebukin napi didalam sel
66	maruf amin sebut memilih jokowi dengan dirinya sama saja sudah mengantongi tiket surga
67	selebaran waspada penculikan anak berlogo polda jabar
68	seorang anggota komite nasional papua barat tewas dibantai tni/polri
69	pelaku mutilasi anak berkeliaran di pelalawan riau
70	bmkg menilai pemberitaan aljazeera yang mengatakan polusi udara di jakarta tidak sehat jelang asian games adalah berlebihan

NO	JUDUL BERITA HOAX
71	indonesian government still hasn't declared a national disaster. incredible. pemerintah indonesia masih belum mendeklarasikan bencana nasional. luar biasa.
72	mobil terkait letusan gunung sinabung disebut bangkai esemka
73	rs wates pasang spanduk, umumkan bpjs menunggak belasan miliar
74	mui mengeluarkan daftar bumbu dan makanan yang mengandung babi
75	badai tropis cempaka mengarah ke yogyakarta
76	cacing hidup di sarden living worm in sardine
77	presiden jokowi temui perwakilan partai komunis china
78	manajer kampanye ditahan facebook dan twitter tak mengira dampak luas manipulasi konten
79	uus: daripada nonton film g30s/pki mendingan nonton film miyabi, lebih bermanfaat!!
80	ajakan melakukan kekerasan kalau dilarang takbir keliling
81	tangkapan layar pengakuan relawan suntik vaksin covid-19 yang menyebut lupa punya hutang tapi plonga-plongo
82	kecewa terhadap pemerintah, puluhan warga jalan bakri bantaeng tutup jalan raya
83	photo of prime minister shinzo abe beside president of the philippines duterte at a national museum in japan
84	video berdoa untuk italia
85	selain jack ma, habib rizieq galang dana rp 500 m lawan corona
86	covid-19 sudah sampai purwakarta
87	kejadian galaxy mall 3 surabaya
88	penjelasan bmkg soal isu cuaca panas extreme yang melanda indonesia
89	sempat heboh dugaan penculikan, siswa gorontalo ternyata dijemput ibu kandung
90	bayi kembar tiga turis jerman hasil hubungannya dengan anjing
91	salah satu pulau di palu telah di jual ke negara china
92	video emak-emak labrak gudang kpu jombang jatim
93	terjadi penyeludupan dan penyimpanan surat suara beserta berkas di kecamatan bengkong, batam
94	caleg ppp jadi korban kecelakaan helikopter di tasikmalaya dengan membawa uang pecahan rp. 100 ribu sebanyak satu karung

NO	JUDUL BERITA HOAX
95	saling menyerang menggunakan klaim mirip
96	luhut : jokowi shalat jumat sampai 4 rakaat, prabowo kalah jauh
97	hari ini terjadi di malaka-malaysia
98	tulisan ridwan kamil : niat busuk kubu sebelah sejak 2014
99	air selokan di malang berwarna hijau toska
100	terduga penculikan anak ditangkap di aceh
101	maunya rezim jokowi adzan tak perlu teriak teriak
102	ada pemukulan sopir taksi online oleh sopir taksi bandara
103	kemendikbud berencana akan melaksanakan kembali unbk matematika pada 17 april mendatang
104	soal bocornya nik dan nomor kk sekarang dilempar menjadi tanggungjawab provider
105	tolak gugatan, hakim tegaskan nonpribumi tak boleh punya tanah di yogyakarta
106	jangan gunakan ponsel sambil mengisi pulsa meteran, sangat berbahaya!
107	katulampa djebol siap2 ya guys!
108	pesawat ufo di nepal/colorado us
109	micin bikin bodoh
110	banser nu jatim siap jadi relawan ke myanmar
111	: kapolda dki sebut penyebab penyerangan novel baswedan karena bisnis hijab online sang istri
112	anis : saya takut dia memberi racun kepada warga
113	obama memegang bokong melania trump
114	pendeta anies baswedan sedang berkhotbah membawakan renungan tentang kiamat air bah
115	video dengan narasi sudah tertangkap satu orang yang membakar gambar habib rizieq
116	photo barack obama, anthony fauci and melinda gates visit the wuhan china laboratory coronavirus factory in 2015
117	foto naik pesawat? bebaaassss indonesia terserah lu ajah
118	perkembangan informasi empat ton telur bansos pemprov jabar membusuk di gudang penyimpanan bulog garut

NO	JUDUL BERITA HOAX
119	bukit tinggi segera lockdown
120	foto saya lebih hebat dari ibu susi tenggelamkanjakarta
121	mungkin kali ciliwung mw bilang @aniesbaswedan im coming..!!
122	16 juli 2020 jakarta putih. triliunan manusia membludak memenuhi dinding gedung dpr dan tiang monas
123	bandara internasional sepungan di kalimantan timur telah dijual ke pihak asing
124	wawan gangguan jiwa karena kecanduan main game
125	kepala kantor kemenag gresik, markus firdaus jawab tudingan yang menyebut dirinya non muslim
126	teuku yazhid pastikan dirinya bukan pria berserban hijau yang ancam jokowi
127	video said aqil siradj dukung prabowo
128	tni-polri pose dua jari dukung prabowo-sandiaga
129	siswi smp 3 puyung yang terjatuh saat melakukan aksi jumping motor, meninggal dunia
130	sekte kristen meniru islam
131	pt aheb nyatakan torpedo tak mengandung zat benzo yang sebabkan 56 pelajar smp 18 pekanbaru menyayat tangannya
132	pesan berantai berisikan polisi minta uang tebusan rp 300 juta
133	simpan hp dibawah jok sepeda motor saat berdering langsung meledak
134	jepang adalah negara pertama yang dikunjungi mahathir
135	densus 88 mengamankan terduga teroris di bali
136	foto wanita iran yang menonton piala dunia tanpa hijab menjadi viral
137	menteri bumn rini soemarno melarang pegawainya berhijab
138	kader pdi perjuangan membuat puisi yang terkesan mendeskreditkan islam
139	polisi bantah akun @cak_luth milik tersangka muslim cyber army
140	pamflet khofifah mengenai kks proyek pesta para tikus
141	megawati bikin jokowi tidak nyaman usai penyebutan petugas partai
142	<i>HOAX</i> , registrasi kartu prabayar paling lambat 31 oktober 2017

NO	JUDUL BERITA HOAX
143	daun berpaku di madiun
144	teori flat earth bukanlah <i>HOAX</i>

Lampiran II Data Uji Berita Valid

NO	JUDUL BERITA VALID	KATEGORI
1	Zuraida Ungkap Sakit Hatinya ke Hakim Jamaluddin: Suami Saya Selingkuh	news
2	Zuraida Tuding Jamaluddin Ganggu Banyak Wanita: Anak, Adik-Instruktur Senam	news
3	Zuraida Tepuk-tepuk Anak yang Tidur Bareng Jamaluddin Saat Dibekap Bed Cover	news
4	Zuraida Sempat Tidur Sekasur Bareng Jasad Jamaluddin Usai Pembunuhan	news
5	Zuraida Modali Rp 2 Juta, Eksekutor Hakim Jamaluddin Beli HP Sekali Pakai	news
6	Zuraida Minta Cerai ke Hakim Jamaluddin: Dia Sakiti Saya dengan Perempuannya	news
7	Zuraida Khawatir Skenario Bunuh Jamaluddin Terbongkar: Ini Membahayakan Saya	news
8	Zuraida Hanum-Jefri Ternyata Ingin Menikah Usai Bunuh Hakim Jamaluddin	news
9	Zuraida Hanum Ternyata Janjikan Eksekutor Jamaluddin Umrah dan Duit	news
10	Zuraida Hanum Siapkan Rp 2 Juta Rencanakan Pembunuhan Hakim Jamaluddin	news
11	Zuraida Hanum dkk Didakwa Pembunuhan Berencana Terhadap Hakim Jamaluddin	news
12	Zuraida Hanum Dituntut Penjara Seumur Hidup	news
13	Zuraida Hadapi Sidang Tuntutan Kasus Pembunuhan Hakim Jamaluddin Hari Ini	news
14	Zuraida Akan Rekonstruksi 'Bed Cover' Pembunuhan Hakim Jamaluddin di 3 Lokasi	news
15	Zulkifli Nilai PAN Rugi Jika Oposisi, PKS: Justru Kalau Pro Pemerintah Sesak	news
16	Zulkifli Hasan: Kongres PAN Digelar 12 Februari di Sultra	news
17	Zulkifli Hasan Penuhi Panggilan KPK Jadi Saksi Kasus Suap Alih Fungsi Lahan	news
18	Zulkifli Hasan Mangkir dari Panggilan KPK di Kasus Suap Alih Fungsi Hutan	news
19	Zulkifli Hasan ke Kader PAN: Kongres Sudah Usai, Tak Ada Waktu Bertengkar	news
20	Zulkifli Hasan Dipanggil KPK terkait Kasus Suap Alih Fungsi Hutan Besok	news
21	Zulkifli Hasan Cetak Rekor Jadi Ketum PAN 2 Periode	news
22	Zulkifli Hasan Bertemu Jokowi di Istana, Bahas Apa?	news
23	Zulkifli Hasan Berharap Jokowi Tabah atas Berpulangnya Ibunda	news

NO	JUDUL BERITA VALID	KATEGORI
24	Zulkifli Hasan Bakal Dicecar KPK soal Pengajuan Alih Fungsi Hutan Riau	news
25	Zulkifli Beri Sinyal Masuk Koalisi Jokowi, NasDem Bicara Kekecewaan	news
26	Zulkifli Beri Sinyal Gabung Koalisi Pemerintah, Dradjad: Kader PAN Harus Ikut	news
27	Zulkifli Beri Kursi ke Hatta, Masih Pikir-pikir untuk Besannya	news
28	Zulhas-Hatta Rajasa Bersatu, PAN Targetkan Jadi 3 Besar di 2024	news
29	Zulhas: Kompetitor Saya Mulfachri dan Asman, Bukan Amien Rais!	news
30	Zulhas Undang Surya Paloh Hadiri Pelantikan DPP PAN 25 Maret Ini	news
31	Zulhas Undang Prabowo ke Pelantikan Pengurus Baru PAN-Bahas Corona	news
32	Zulhas Terpilih Lagi Jadi Ketum, PAN Akan Merapat ke Jokowi?	news
33	Zulhas Temui Prabowo di Kertanegara Perkenalkan Pengurus PAN Baru	news
34	Zulhas Temui Jokowi di Istana, Loyalis: Sahabat Baik Lama Tak Bersua	news
35	Zulhas Sesalkan Ribut di Kongres PAN: Komputer Diambil, OC Tak Bisa Kerja	news
36	Zulhas Sapa Sandiaga di Kongres PAN: Mudah-mudahan Wapres yang Tertunda	news
37	Zulhas Pimpin Perolehan Suara Sementara Pemilihan Ketum PAN	news
38	Zulhas Ngaku Rugi Masuk Oposisi, PDIP Minta PAN Berembuk	news
39	Zulhas Minta Pengurus Masjid Jaga Disiplin Jemaah di New Normal	news
40	Zulhas Ketum PAN Lagi, Hatta Rajasa Jadi Ketua MPP	news
41	Zulhas Carter 2 Pesawat Lion Air Bawa 392 Pendukungnya ke Kongres PAN	news
42	Zulhas Bertemu Jokowi, Golkar: Bagus Jika Niatnya Dukung Pemerintah	news
43	Zuhur Tetap Ada Meski Salat Jumat Ditiadakan di Masjid Istiqlal, Saf Tak Rapat	news
44	Zona Merah Corona Meluas, DPRD Medan Minta Sanksi Perwal Karantina Ditegakkan	news
45	Zindar Gugat ke PTUN Jakarta Tuntut 5 Anggota BPK Dicopot	news
46	Zakat Fitrah, Arti dan Definisi Menurut 4 Mahdzab	news
47	Zakat Fitrah dan Zakat Mal, Apa Bedanya?	news

NO	JUDUL BERITA VALID	KATEGORI
48	Zakat di Bansos Corona Dianggap SARA, Baznas Babel: Tak Ada Diskriminasi	news
49	Zakat Care Line ACT, Mudahkan Berzakat Tanpa Keluar Rumah	news
50	Zainudin Hasan Dibui 12 Tahun, Harley hingga Tanah Senilai Rp 27 M Dirampas	news
51	Yusuf Wonda, Putra Papua yang Viral Makan Nasi-Tahu Lulus Jadi Prajurit TNI	news
52	Yusuf Mansur Ajak Warga Jalani New Normal dan Berdamai dengan Kesultanan	news
53	Yusril: Utamakan Nyawa Rakyat Hadapi Corona, Ekonomi Nomor Dua	news
54	Yusril: Pandemi Corona Situasi Genting, Jokowi Dapat Terbitkan Perppu	news
55	Yusril: Landasan Hukum PSBB Serba Tanggung karena Tanpa Sanksi Pidana	news
56	Yusril Usul 'Koalisi Parpol' soal Ambang Batas DPR, Pimpinan Komisi II Heran	news
57	Yusril Sarankan Pemerintah Tangani Corona Seperti Tanggap Darurat Bencana	news
58	Yusril Nilai Masalah Masa Jabatan Bupati Talaud Selesai: Tinggal Dilantik	news
59	Yurianto: Edukasi Corona Harus Dilakukan Bersama, Bukan Hanya Pemerintah	news
60	Yuri: Penanganan Pandemi Corona Dilakukan Terpimpin Melalui Gugus Tugas	news
61	Yuri Ungkap Cerita Petugas Hadapi Virus Harus Berjam-jam Tak ke Kamar Kecil	news
62	Yuri soal Kasus di Bogor: Tak Ada Kematian Corona Seperti Serangan Jantung	news
63	Yuri Soal Diskriminasi ke WN Jepang: Yang Melakukan Masyarakat, Jadi Nggak Enak	news
64	Yuri Jawab Pemprov Bali soal WNA Corona Meninggal: Dokternya Sudah Tahu	news
65	Yunahar Ilyas Meninggal Dunia, MUI: Beliau Sosok Rendah Hati-Bersahaja	news
66	Yunahar Ilyas Dikenang Sebagai Sosok Ulama Bijak yang Gigih Berdakwah	news
67	Yuliana Makin Fasih Teknologi karena WFH	news
68	Yuk, Nonton Konser Sabyan di Aplikasi Umma	news
69	Yuk, Kunjungi Pameran UMKM di BNI Kreasi Nusantara Craft Online	news
70	Yuk! Ikutan Sharing Time detiknetwork Soal Jurnalisme Digital 2020	news
71	Yuk Simak, Jadwal Belajar dari Rumah TVRI Hari Ini Rabu 6 Mei 2020	news

NO	JUDUL BERITA VALID	KATEGORI
72	Yuk Sahur! Ini Jadwal Imsak dan Buka Puasa di Jakarta 23 Mei	news
73	Yuk Perkaya Skill Kamu Lewat Online Learning Duta Bangsa	news
74	Yuk Nonton! Kultum Ustaz Hanan Attaki: Keajaiban Niat	news
75	Yuk Intip Jadwal Imsak dan Buka Puasa di DKIÂ Jakarta Hari Ini!	news
76	Yuk Belajar Jadi Orang yang Rida Bareng Ustaz Hanan Attaki	news
77	Yuk Bantu #AntarkanSemangat Ramadhan Lewat Postingan Instagram	news
78	YouTuber Aleh Aleh Khas Medan Terjerat Kasus Ujaran Kebencian	news
79	YOSS Temui Gubernur Sulsel, Hendak Serahkan Stadion Mattoanging	news
80	Yorrys Kritik Mahfud yang Sebut Data Veronica Koman soal Papua 'Sampah'	news
81	YLBHI: Yasonna Laoly Buang Badan dengan Copot Dirjen Imigrasi	news
82	YLBHI: Larangan Merekam Persidangan akan Memperparah Mafia Peradilan	news
83	YLBHI Sebut Pola-pola Tindakan Teror dalam Ruang Demokrasi di 2020	news
84	YLBHI Kritik Mekanisme Penangkapan Aktivis Ravio Patra oleh Polisi	news
85	YLBHI Desak Kenaikan Iuran BPJS Dibatalkan karena Langkah Putusan MA	news
86	Yes! Kini Tukang Ikan Hias Bisa Dapat Sertifikasi Pakai Aplikasi Ini	news
87	Yenti Garnasih Nilai Kasus Wahyu Setiawan Adalah Penipuan, Ini Alasannya	news
88	Yenny Wahid soal Tepuk 'No Kafir': Apa Sih yang Mau Diajarin?	news
89	Yayasan Pendidikan VIVERE Salurkan APD Medis ke Puskesmas dan Klinik	news
90	Yayasan Cerita Kendala Rawat 3 Anak Difabel Positif Corona di Asrama	news
91	Yatim Piatu, 3 Anak Difabel Positif Corona di Jaktim Jalani Isolasi Mandiri	news
92	Yasonna: Paspor Jemaah Haji 80% Selesai, tapi Belum Ada Keputusan dari Saudi	news
93	Yasonna: Di Tanjung Priok Lahir Kriminal dari Kemiskinan, Menteng Tidak	news
94	Yasonna: 5 Petugas Bandara Soetta Positif Corona, 3 Sembuh	news
95	Yasonna, Sri Mulyani dan Jaksa Agung Hadiri Sidang Perppu Corona di MK	news

NO	JUDUL BERITA VALID	KATEGORI
96	Yasonna Ungkap Pegawainya Diperiksa Terkait Corona, Ada 1 yang Meninggal	news
97	Yasonna Teken Permen Penghentian Sementara Bebas Visa bagi China	news
98	Yasonna Tegaskan Pengesahan Perppu Corona Jadi UU Sesuai Prosedur	news
99	Yasonna soal Perppu Corona: Tak Ada yang Kebal Hukum, Koruptor Tetap Ditindak	news
100	Yasonna Siapkan Hukuman Berat untuk Napi Berulah: Mereka Akan Menyesal	news
101	Yasonna Siap Mundur Terkait Harun Masiku, NasDem Minta KPK Turun Tangan	news
102	Yasonna Sebut Residivisme RI Rendah, Komisi III: Bukan soal Perbandingan	news
103	Yasonna Sebut Angkasa Pura II Tolak Beri Rekaman CCTV Harun Masiku	news
104	Yasonna Persilakan HRS Masuk RI, Balik Disindir FPI	news
105	Yasonna Persilakan Habib Rizieq Masuk RI, FPI: Lempar Batu Sembunyi Tangan	news
106	Yasonna Ogah Bicara Kasus Harun Masiku di DPR: Saya Menkum HAM Sekarang	news
107	Yasonna Minta Pemda Segera Daftarkan Rempah Kekayaan Geografis ke Kemenkumham	news
108	Yasonna Laoly Siap Mundur dari Menteri Jika Ronny Sompie Tak Bersalah	news
109	Yasonna Laoly Minta Maaf ke Warga Tanjung Priok soal Pidato 'Kriminal'	news
110	Yasonna Laoly Diadukan terkait Harun Masiku, KPK Telaah Laporan	news
111	Yasonna Laoly Bicara Pembebasan Napi Koruptor hingga BaasyirÂ	news
112	Yasonna ke Habib Rizieq: Sistem Free, Anytime Bisa Masuk RI	news
113	Yasonna Jelaskan Alasan Pemerintah Ambil Opsi PSBB untuk Tangani Corona	news
114	Yasonna Disebut Tak Bohong soal 'Delay' Masiku, PDIP Bicara 'Darah Campuran'	news
115	Yasonna Digugat ke PN Surakarta Terkait Asimilasi Napi yang Kembali Berulah	news
116	Yasonna Copot Ronny Sompie, Komisi III Pilih Beri Atensi ke TPF Kemenkum HAM	news
117	Yasonna Bela Kebijakan Asimilasi Saat Marak Kejahatan di Tengah Pandemi	news
118	Yasonna Bantah Akan Bebaskan Koruptor Lansia Gegara Corona	news
119	Yasonna Bakal Didemo soal 'Kriminal di Priok', Polisi Turunkan Pengamanan	news

NO	JUDUL BERITA VALID	KATEGORI
120	Yasonna Baca Komentar Medsoc tentang Dirinya: Bahasanya Kasar, Ampun Deh	news
121	Yasonna Angkat Tangan Ngaku Tak Kenal Harun Masiku, Ini Kata KPK	news
122	Yang Perlu Kamu Tahu soal 2 WNI Tertular Virus Corona di Depok	news
123	Yang Perlu Diketahui soal Virus Hanta di Tengah Pandemi Corona	news
124	Yang Perlu Diketahui soal Sanksi Larangan Mudik yang Berlaku Hari Ini	news
125	Yang Perlu Diketahui soal Heboh Pelarungan ABK WNI oleh Kapal China	news
126	Yang Perlu Diketahui dari PSBB Jakarta yang Berlaku Hari Ini	news
127	Yang Nekat Sholat di Masjid saat Pandemi, Lihat Rasulullah SAW Ketika Hijrah	news
128	Yang Mentereng di Survei Indo Barometer: Dari Ahok hingga Prabowo	news
129	Yang Beda dari Makassar, Iring-iringan Pengantar Jenazah Kerap Masuk Tol	news
130	Yakin Tak Terlibat Jiwasraya, Moeldoko Minta BAP Diperiksa	news
131	Yakin Dipinang PDIP Jadi Cawalkot Solo, Gibran: Pemimpin Harus Optimistis	news
132	Yakin Bisa Lewati Badai Corona, Jokowi Tekankan Desa-Pusat Harus 1 Visi	news
133	Yakin Anies Jujur, PD Minta Dugaan Manipulasi Rekomendasi TACB Diusut	news
134	Yacht Ditemukan Terombang-ambing di Aceh Utara, Diduga Milik WNA	news
135	Yacht dari Australia Hendak Menuju Jakarta, Dikemudikan Pelaut 70 Tahun	news
136	Yacht Asal Australia Dijarah di Perairan Sumsel, Dievakuasi di Lampung	news
137	Xanana Gusmao Yakin RI Bantu Timor Leste soal Karantina Warganya dari China	news
138	Wusss! Koleksi Nurhadi Eks PNS Tajir Melintir DPO KPK: MINI Cooper-Jaguar	news
139	Wujudkan Kota Cerdas, Gubernur Sulsel Jajaki Pembangunan MRT di Makassar	news
140	Wuhan Masih Terkunci, Mahasiswa RI Belum Bisa Kembali	news
141	WP KPK Minta Jaksa Agung Tunda Penarikan 2 Jaksa	news
142	WP KPK Minta Aturan soal Penarikan Pegawai ke Institusi Asal Diperjelas	news
143	WP KPK Lapor Dewas, Minta Pengembalian Kompol Rosa Dihentikan	news

NO	JUDUL BERITA VALID	KATEGORI
144	Wow! Gojek Bagikan 200.000 Sembako untuk Mitra Driver	news
145	Wow! Biaya Parkir BMW 4 Tahun di Bandara Ngurah Rai Bali Rp 70 Juta	news
146	Wow! Alat Ini Bisa Hemat Tagihan Listrik hingga 15%	news
147	Wow! Ada Pesta Bakso di Arena Salju Trans Snow World Bekasi	news
148	World Press Freedom Day, Dewan Pers Ingatkan soal Empati Terkait Corona	news
149	Woods Bagikan 75 Ribu Masker Atasi COVID-19 di Jabodetabek	news
150	WO di Depok yang Kabur di Hari-H Tipu Puluhan Pengantin, Kerugian Rp 2,5 M	news
151	WNI yang Ingin Pulang, KJRI Houston Imbau Urus Surat Keterangan Perjalanan	news
152	WNI yang Dievakuasi dari Wuhan China Jadi 241 Orang, Ini Kata Kemlu	news
153	WNI yang Dievakuasi dari China Akan Jalani Observasi Standar Protokol WHO	news
154	WNI yang Curi Tas Louis Vuitton Rp 105 Juta di Melbourne Disidang 2 Oktober	news
155	WNI Sandera Abu Sayyaf Dibebaskan Militer Filipina, Panglima: Terima Kasih	news
156	WNI Pulang Kampung Usai Observasi dari Wuhan yang Terkurung	news
157	WNI Positif Virus Corona, Kemlu Cek ke KBRI Singapura	news
158	WNI Positif Virus Corona Tinggal di Perumahan Studio Alam Depok	news
159	WNI Positif Corona Diisolasi di Gedung Terpisah di RSPI	news
160	WNI Positif Corona di Singapura, Pemerintah Perketat Akses Bandara-Pelabuhan	news
161	WNI Positif Corona di Singapura Jadi 40 Orang: 7 Sembuh, 2 Meninggal	news
162	WNI Positif Corona di Singapura Jadi 37 Orang	news
163	WNI Positif Corona di Luar Negeri Jadi 994: 551 Sembuh, 56 Meninggal	news
164	WNI Positif Corona di Luar Negeri Jadi 773: 50% Sudah Sembuh, 43 Meninggal	news
165	WNI Positif Corona di Luar Negeri Jadi 681, Terbaru di Rusia	news
166	WNI Positif Corona di Luar Negeri Jadi 470 Orang: 97 Sembuh, 19 Meninggal	news
167	WNI Positif Corona di Luar Negeri Jadi 406 Orang: 95 Sembuh, 19 Meninggal	news

NO	JUDUL BERITA VALID	KATEGORI
168	WNI Positif Corona di Luar Negeri Jadi 210 Orang: 33 Sembuh, 6 Meninggal	news
169	WNI Positif Corona di Luar Negeri Jadi 1.005 Per 6 Juni, Terbaru di Ethiopia	news
170	WNI Positif Corona di Luar Negeri 786 Kasus: 386 Sembuh, 42 Meninggal	news
171	WNI Positif Corona di Bali PNS yang Kembali dari Jakarta	news
172	WNI Positif Corona Batuk 10 Hari Tak Sembuh-Sembuh	news
173	WNI Positif Corona Ada di 33 Negara, Menlu Pastikan Dapat Layanan Kesehatan	news
174	WNI Pasien Corona di Kuwait Meninggal, Ini Sebaran Kasus WNI COVID-19 di LN	news
175	WNI Pasien 07 yang Positif Corona Pulang dari AS dan Transit di Jepang	news
176	WNI Meninggal di Singapura Akibat Corona, Kemlu Siap Bantu Urus Jenazah	news
177	WNI Kru Diamond Princess Tak Mau Dievakuasi Via Laut, KSP: Dipertimbangkan	news
178	WNI Kru Diamond Princess Dievakuasi Pakai Kapal atau Pesawat? Ini Kata Jokowi	news
179	WNI Kena Virus Corona, Menkes: Biar Singapura yang Periksa	news
180	WNI Kena Virus Corona, Jokowi: Ditangani Singapura dan Didampingi KBRI	news
181	WNI Kena Virus Corona di Singapura, Menko PMK: Jika Sudah Sembuh Dipulangkan	news
182	WNI Kena Virus Corona di Singapura, Ma'ruf: Ditangani Serius Kedutaan	news
183	WNI Kena Corona-WN Jepang Bertemu di Paloma, Manajemen: Semua Karyawan Sehat	news
184	WNI Kena Corona ke RS Mitra Keluarga Depok, 70-an Tenaga Medis Dirumahkan	news
185	WNI Disandera Lagi	news
186	WNI Dinyatakan Positif Corona di Singapura, Sempat Dirawat di Jakarta	news
187	WNI Dievakuasi dari Wuhan China, BNPB: Akan Jalani Observasi 2 Minggu	news
188	WNI di Natuna Selesai Observasi Sabtu, Akan Dibawa Pesawat TNI AU ke Jakarta	news
189	WNI di Jepang Belum Dievakuasi, Pemerintah Masih Negosiasi	news
190	WNI di Hubei China Dijemput Kurang dari 24 Jam Lagi	news
191	WNI di Hubei China Akan Dijemput Pakai Batik Air	news

NO	JUDUL BERITA VALID	KATEGORI
192	WNI di Australia Positif Corona, Menlu: Statusnya Permanent Resident	news
193	WNI dari Wuhan Tiba di Bandara Hang Nadim Batam	news
194	WNI dari Wuhan Mulai Diobservasi, TNI-Polri Gencarkan Patroli di Natuna	news
195	WNI dari Wuhan Diobservasi, Ada Warga Natuna yang 'Pindah' Pulau	news
196	WNI dari China Akan Dikarantina Usai Evakuasi dari Wabah Virus Corona	news
197	WNI Dari 3 Negara Lonjakan Corona Tak Ditolak Balik ke RI, tapi...	news
198	WNI Asal Aceh Ceritakan Kondisi Observasi di Hanggar Raden Sadjad Natuna	news
199	WNI ABK World Dream Segera Dievakuasi, Bagaimana yang di Diamond Princess?	news
200	WNI ABK World Dream Mulai Dievakuasi, Dipercirakan Tiba di Sebaru Jumat	news
201	WNI ABK World Dream Diobservasi, 3 RS Disiapkan Jadi Rujukan Suspect Corona	news
202	WNI ABK World Dream Akan Diobservasi di Pulau Sebaru Kepulauan Seribu	news
203	WNI ABK Diamond Princess: 68 Dievakuasi, 8 Tertahan karena Corona	news
204	WNI ABK Diamond Princess yang Sembuh Corona di Jepang Dipulangkan Malam Ini	news
205	WNI ABK Diamond Princess Tiba di Tanjung Priok, Dipulangkan ke 11 Provinsi	news
206	WNI ABK Diamond Princess Sembuh Corona akan Diobservasi di Bapelkes Cikarang	news
207	WNI ABK Diamond Princess Merasa Dibunuh Pelan-pelan, Menkes: Itu Persepsi	news
208	WNI ABK Diamond Princess Juga Akan Diobservasi di Pulau Sebaru	news
209	WNI ABK Diamond Princess Dievakuasi ke Kertajati, Lanjut Pulau Sebaru Via Laut	news
210	WNA Tewas di Kolam Hotel Gorontalo, Evakuasi Pakai Standar Pasien Corona	news
211	WNA Finlandia yang Hilang Saat Snorkeling Ditemukan Tewas	news
212	WNA Asal Irlandia di Sumbar Positif Corona, Wagub: Dia Berpindah-pindah Hotel	news
213	WNA Asal China Dijambret di Palembang, Duit Rp 75 Juta Raib	news
214	WN Singapura Positif Virus Corona Kunjungi Batam, 4 WNI Dikarantina	news
215	WN Singapura Positif Corona 3 Hari di Batam, Begini Jejaknya	news

NO	JUDUL BERITA VALID	KATEGORI
216	WN Singapura Meninggal di Batam, Sempat Diduga Corona Namun Negatif	news
217	WN Prancis Natalhie Ditolak Masuk ke Makassar	news
218	WN Lithuania Tewas Usai Main Kano di Bali, Dievakuasi Petugas Ber-APD	news
219	WN Korsel di Solo Bunuh Diri karena Merasa Corona, Ini Kata Istana	news
220	WN Jepang di Bali Ditemukan Tewas Gantung Diri di Kamar	news
221	WN Finlandia Hilang Saat Snorkeling di Gili Meno Lombok	news
222	WN Filipina yang Tewas di Kapal di Laut Sulsel Diautopsi Malam Ini	news
223	WN Filipina Tewas di Kapal di Laut Sulsel, Polisi: Tak Terjangkit Corona	news
224	WN Australia Ditemukan Tewas di Hotel Bali, Polisi Tunggu Hasil Tes COVID-19	news
225	WN Ausie Ditemukan Tewas di Kosan, Dievakuasi Petugas Ber-APD Lengkap	news
226	WN Amerika Ditemukan Tewas di Vila Bali, Dievakuasi Petugas Ber-hazmat	news
227	WN Aljazair Buron Pembobolan ATM di Gianyar Bali Ditangkap	news
228	WN Afrika yang Terseret Arus Pantai Berawa Bali Ditemukan Tewas	news
229	WN Afrika Terseret Arus Pantai Berawa Bali	news
230	Wisma Milik Kemenristek Disulap Jadi Rumah Singgah Tenaga Medis di Tangsel	news
231	Wisma Atlet Rawat 1.018 Pasien, 700 Terkonfirmasi Positif COVID-19	news
232	Wisma Atlet Kemayoran Jadi RS Darurat Corona, PPNI Ingatkan Kesiapan Relawan	news
233	Wisma Atlet Diberdayakan untuk Tangani Pasien Corona Ringan	news
234	Wishnutama-BNPB Rapat Koordinasi Bahas Mitigasi Bencana di Daerah Wisata	news
235	Wishnutama Minta Hotel Tak Pecat Karyawan Selama Masa Pandemi Corona	news
236	Wishnutama & Kang Emil Minta Bogor Kembangkan Festival Cap Go Meh	news
237	Wisatawan Terseret di Pantai Sawarna Lebak, 1 Orang Tewas-1 Hilang	news
238	Wisatawan Berkurang di Taman Nasional Bunaken	news
239	Wisata Lesu Imbas Corona, Banda Aceh Giatkan Promosi	news

NO	JUDUL BERITA VALID	KATEGORI
240	Wiranto soal Wantimpres Belum Setor LHKPN: Kami Baru Sebulan	news
241	Wiranto Resmikan Museum Bang Yos di Cibubur	news
242	Wiranto Lapor Kerja Wantimpres ke Jokowi, Tak Ingin Beri Nasihat Duplikasi	news
243	Wiranto Bikin Video Klip, Nyanyikan Lagu Minta Warga Tak Mudik	news
244	Wilson Pamit, Matt dan BKSDA Masih Berupaya Selamatkan Buaya Berkalung Ban	news
245	Willy Ngaku Sudah 4 Kali Order Fiktif Ojol Kongkalikong dengan Toko	news
246	Wilayahnya Zona Merah Corona, Walkot Prabumulih Siapkan RS-Bagi Sembako	news
247	Wilayah Penyangga Jakarta Terapkan PSBB, Check Point Diperluas	news
248	Wilayah Indonesia Masih Kebal Corona	news
249	WHO: Wabah Corona Meningkat Tiba-tiba di Dunia, Sangat Memprihatinkan	news
250	WHO Tetapkan Pandemi Corona, Ini Saran NasDem ke Pemerintah	news
251	WHO Surati Jokowi Minta Segera Umumkan Darurat Nasional Virus Corona	news
252	WHO Perluas Definisi Kematian COVID-19, Tak Cuma yang Terkonfirmasi	news
253	WHO Paparkan 4 Perbedaan Corona dan Flu, Penanganannya Berbeda	news
254	WHO Nyatakan Darurat Global Virus Corona, Istana: Kami Serius Tangani Ini	news
255	WHO Minta Umumkan Status Darurat Corona, KSP: Belum Diputuskan, Tetap Waspada	news
256	WHO Minta Jokowi Umumkan Darurat Nasional Corona, Istana: Sudah Bentuk Gugus Tugas	news

2- 220605210006 - Deteksi Berita Hoax Dari Media Online Indonesia Menggunakan Algoritma Naive Bayes dan Support Vector Machine

by elytaanam@gmail.com Nur

Submission date: 18-Dec-2023 11:39AM (UTC+0700)

Submission ID: 2261524907

File name: Bab_I_VII_Nur_Elyta_Deteksi_Berita_Hoax.pdf (1.85M)

Word count: 14204

Character count: 79920

2- 220605210006 - Deteksi Berita Hoax Dari Media Online Indonesia Menggunakan Algoritma Naive Bayes dan Support Vector Machine

ORIGINALITY REPORT



PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	2%
2	tafsir.learn-quran.co Internet Source	1%
3	ejournal.undip.ac.id Internet Source	1%
4	repository.ub.ac.id Internet Source	1%
5	openlibrarypublications.telkomuniversity.ac.id Internet Source	1%
6	akhirat.net Internet Source	1%
7	e-jurnal.uajy.ac.id Internet Source	1%
8	dspace.uii.ac.id Internet Source	1%

Submitted to Universitas Brawijaya

9

<1 %

polynoe.lib.uniwa.gr

Internet Source

<1 %

gfd.turnbackhoax.id

Internet Source

<1 %

Submitted to UIN Sunan Ampel Surabaya

Student Paper

<1 %

ejournal.ikado.ac.id

Internet Source

<1 %

www.kompasiana.com

Internet Source

<1 %

repository.its.ac.id

Internet Source

<1 %

lib.unnes.ac.id

Internet Source

<1 %

repository.uir.ac.id

Internet Source

<1 %

blog.csdn.net

Internet Source

<1 %

jurnal.univbinainsan.ac.id

Internet Source

<1 %

Ahmad Kurniadi. "Implementasi Convolutional Neural Network Untuk Klasifikasi Varietas

<1 %

Pada Citra Daun Sawi Menggunakan Keras",
DoubleClick: Journal of Computer and
Information Technology, 2020

Publication

21 Submitted to Kolej Universiti Islam Sultan Azlan Shah <1 %

Student Paper

22 medium.com <1 %

Internet Source

23 سعد عبدالعزيز عواد العنزي. "تفسير آيات الأولياء والصالحين من غير الأنبياء والمرسلين في القرآن الكريم", مجلة الدراسات العربية, 2023 <1 %

Publication

24 Submitted to Management Development Institute Of Singapore <1 %

Student Paper

25 Mohammad Amin Amani, Francesco Marinello. "A Deep Learning-Based Model to Reduce Costs and Increase Productivity in the Case of Small Datasets: A Case Study in Cotton Cultivation", Agriculture, 2022 <1 %

Publication

26 Submitted to Imam Malik College for Sharia and Law <1 %

Student Paper

27 ijadis.org <1 %

Internet Source

28	archive.org Internet Source	<1 %
29	www.slideshare.net Internet Source	<1 %
30	id.123dok.com Internet Source	<1 %
31	Sri Winiarti, Desy Widayanti, Ulaya Ahdiani, Taufiq Ismail. "Klasifikasi Jenis Buku Berdasarkan Cover dan Judul Buku Menggunakan Metode Support Vector Machine dan Cosine Similarity", Sainteks, 2022 Publication	<1 %
32	Submitted to Universitas Jambi Student Paper	<1 %
33	ejournal.up45.ac.id Internet Source	<1 %
34	rcastoragev2.blob.core.windows.net Internet Source	<1 %
35	repository.unja.ac.id Internet Source	<1 %
36	www.idntimes.com Internet Source	<1 %
37	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %

- 38 ejurnal.seminar-id.com <1 %
Internet Source
- 39 rahmadya.com <1 %
Internet Source
- 40 Dhana Aulia Ayu Kurniawan, Ema Utami, Hanif Al Fatta. "ANALISIS SENTIMEN PADA OPINI PENGGUNA APLIKASI QASIR MENGGUNAKAN SUPPORT VECTOR MACHINE DAN RANDOM FOREST", TEKNIMEDIA: Teknologi Informasi dan Multimedia, 2023 <1 %
Publication
- 41 Submitted to Politeknik Negeri Jember <1 %
Student Paper
- 42 Virginia Khoirunnisa, Sri Lestari. "IMPLEMENTASI KLASIFIKASI KEHAMILAN BERESIKO DENGAN METODE NAIVE BAYES PADA PUSKESMAS KELURAHAN MALAKA JAYA", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023 <1 %
Publication
- 43 beritafakta.id <1 %
Internet Source
- 44 infosulawesi.com <1 %
Internet Source
- 45 repo.uinsatu.ac.id <1 %
Internet Source

<1 %

46 Submitted to Universitas Muria Kudus <1 %
Student Paper

47 doaj.org <1 %
Internet Source

48 repositori.iain-bone.ac.id <1 %
Internet Source

49 t.me <1 %
Internet Source

50 www.researchgate.net <1 %
Internet Source

51 Monica Widiasri, Ellysa Tjandra, Lisa Maria Chandra. "Peningkatan Kinerja Pencarian Dokumen Tugas Akhir Menggunakan Porter Stemmer Bahasa Indonesia dan Fungsi Peringkat Okapi BM25", Teknika, 2017 <1 %
Publication

52 Rudy Asrianto, Melda Herwinanda. "Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine", Jurnal CoSciTech (Computer Science and Information Technology), 2022 <1 %
Publication

- 53 Zahra Azhari, Lusiana Efrizoni, Wirta Agustin, Rini Yanti. "Opinion Mining menggunakan Algoritma Deep Learning untuk Menganalisis Penggunaan Aplikasi Jamsostek Mobile", Indonesian Journal of Computer Science, 2023
Publication
-
- 54 adoc.pub <1 %
Internet Source
-
- 55 ejurnal.stmik-budidarma.ac.id <1 %
Internet Source
-
- 56 journal.ipm2kpe.or.id <1 %
Internet Source
-
- 57 jurnal.iaii.or.id <1 %
Internet Source
-
- 58 library.um.ac.id <1 %
Internet Source
-
- 59 repo.itera.ac.id <1 %
Internet Source
-
- 60 vdocuments.mx <1 %
Internet Source
-
- 61 www.droidriky.my.id <1 %
Internet Source
-
- 62 www.idxchannel.com <1 %
Internet Source

63	docplayer.info Internet Source	<1 %
64	ejournal-s1.undip.ac.id Internet Source	<1 %
65	iqra.republika.co.id Internet Source	<1 %
66	pt.scribd.com Internet Source	<1 %
67	pythonformachinelearning.wordpress.com Internet Source	<1 %
68	repository.uin-suska.ac.id Internet Source	<1 %
69	repository.uinjkt.ac.id Internet Source	<1 %
70	text-id.123dok.com Internet Source	<1 %
71	www.kominfo.go.id Internet Source	<1 %
72	www.scribd.com Internet Source	<1 %
73	zadoco.site Internet Source	<1 %
74	publikasi.dinus.ac.id Internet Source	<1 %

- 75 towardsdatascience.com **<1 %**
Internet Source
-
- 76 Arief Setyo Nugroho, Rusydi Umar, Abdul Fadlil. "KLASIFIKASI BOTOL PLASTIK MENGGUNAKAN MULTICLASS SUPPORT VECTOR MACHINE", Jurnal Khatulistiwa Informatika, 2021 **<1 %**
Publication
-
- 77 Indra Griha Tofik Isa. "Aplikasi Asesmen Calon Debitur menggunakan Naive Bayes di Koperasi Mitra Sejahtera SMK Negeri 1 Kota Sukabumi", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2021 **<1 %**
Publication
-
- 78 Luban Abdi Susanto. "PEMILIHAN HYPERPARAMETER PADA ALEXNET CNN UNTUK KLASIFIKASI CITRA PENYAKIT KEDELAI", INDEXIA, 2023 **<1 %**
Publication
-
- 79 Nur Muhammad Farhan, Bayu Setiaji. "Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter Mengenai Persoalan Perpu Cipta Kerja", Indonesian Journal of Computer Science, 2023 **<1 %**
Publication
-
- 80 pt.slideshare.net **<1 %**
Internet Source

Exclude quotes On

Exclude bibliography Off

Exclude matches Off