

# ANALISIS SENTIMEN BERITA HOAX MENGGUNAKAN NAÏVE BAYES

Meyti Eka Apriyani<sup>1</sup>, Rasyed Renaldi<sup>2</sup>, Toga Aldila Cinderatama<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknologi Informasi, Politeknik Negeri Malang

<sup>1</sup>meyti.eka@polinema.ac.id, <sup>2</sup>renaldirasyed@gmail.com, <sup>3</sup>toga.aldila@polinema.ac.id

---

## Abstrak

Saat ini WhatsApp merupakan aplikasi yang banyak digunakan oleh masyarakat Indonesia sebagai sarana komunikasi jarak jauh dengan pengguna lain. WhatsApp juga memiliki fitur yang dapat digunakan untuk kepentingan perlindungan dalam mengatasi tindak kejahatan digital, dalam kasus ini kejahatan digital yang dimaksud adalah penyebaran berita hoax presiden Joko Widodo 3 periode. Untuk mengatasi kejahatan digital tersebut maka diperlukan sebuah analisis sentimen agar dapat menyelesaikan tindak kejahatan tersebut. Penelitian ini menggunakan algoritma Naïve Bayes untuk mengklasifikasikan percakapan yang ditemukan dan *User Acceptance Test* (UAT) untuk mengetahui tanggapan responden tentang kemudahan aplikasi desktop python terhadap penggunaan aplikasi tersebut. Hasil yang didapatkan adalah penggunaan aplikasi desktop dengan bahasa pemrograman Python kurang maksimal dalam mendapatkan kalimat bukti percakapan berdasar file gambar yang telah ditentukan, dari 80 kalimat percakapan yang digunakan sebagai bahan penelitian dapat dihasilkan berjumlah 75 kalimat percakapan. Kemudian, pada proses klasifikasi menggunakan Naïve Bayes dengan skenario 50:50 didapatkan akurasi 88%, *precision* 84%, dan *recall* 91,3%, sedangkan hasil dari pengujian UAT memiliki persentase 50% sangat setuju, 40% setuju, dan 7,1% tidak setuju. Dapat disimpulkan aplikasi desktop python berfungsi, sesuai dan layak untuk digunakan, akan tetapi untuk mendapatkan hasil yang maksimal perlu adanya pengembangan lebih lanjut dan disimpulkan juga aplikasi desktop python mudah dan layak digunakan, tetapi untuk mendapatkan hasil yang maksimal perlu adanya pengembangan lebih lanjut.

**Kata kunci** : Naïve Bayes, analisis sentimen, berita hoax, *User Acceptance Test*

---

## 1. Pendahuluan

Pada dasarnya pengguna internet di Indonesia sangat lah besar, diketahui berdasarkan hasil survey Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menyatakan setidaknya ada 210 juta pengguna internet dari 275 juta penduduk di Indonesia periode 2021-2022 (APJII,2022). Hal itu berpengaruh kepada penetrasi pengguna internet di Indonesia yang meningkat sebesar 77%. Dengan dinyatakannya beberapa laporan tersebut membuat banyak perusahaan ternama bersaing untuk menyediakan dan menciptakan sebuah aplikasi yang disertai dengan internet untuk media penghubungnya. Berbagai macam aplikasi yang dimulai dari aplikasi *chatting* maupun aplikasi lainnya, akan tetapi target yang diprioritaskan oleh perusahaan ialah aplikasi *chatting* atau pesan instan tersebut.

Semakin pesatnya perkembangan pengguna internet, hal tersebut menimbulkan kerentanan pada sistem keamanannya terutama pada salah satu aplikasi pesan instan yaitu WhatsApp. Kerentanan tersebut berupa tersebarannya informasi negatif ke beberapa pengguna WhatsApp yang lain, diantara informasi negatif tersebut adalah berita bohong yang mana hal tersebut jelas merugikan bagi pengguna aplikasi WhatsApp yang lain. Berdasarkan hasil riset yang dilakukan oleh *DailySocial.id*, pada konteks ini

WhatsApp memiliki jumlah persentase mencapai 56% untuk penggunaan menyebarkan berita bohong. Selain daripada itu WhatsApp juga merupakan aplikasi pesan instan yang termasuk lintas platform, yang berarti dapat digunakan melalui laptop baik itu dekstop maupun web. Hal ini semakin memungkinkan bahwa WhatsApp dapat dengan mudahnya disadap dan disusupi oleh beberapa oknum tidak bertanggung jawab untuk menyebarkan berita bohong (*hoax*) (Anggraini, *et al*, 2020).

Kemudian berdasarkan sumber *Google Trends* dengan topik “presiden 3 periode” pada tanggal 10-16 April 2022 memiliki minat yang menanjak hal tersebut jika meluas akan berdampak pada berkurangnya kepercayaan masyarakat akan berita politik tersebut dan juga akan berpotensi menimbulkan perpecahan antar golongan dan masyarakat.

Berbagai metode telah diterapkan untuk mendapatkan akurasi terbaik dalam analisis sentimen. Penulis melakukan studi pustaka terhadap beberapa metode yang telah digunakan pada penelitian terdahulu, yaitu algoritma naïve bayes pada analisis sentimen positif dan negatif (Kautsar & Syafrullah, 2022) terhadap isu presiden 3 periode di twitter, analisis sentimen pemilihan presiden tahun 2019 menggunakan K-means untuk klastering data latih

dan Naïve Bayes (Kurniawan & Susanto, 2019), penggunaan UAT menurut (Priyatna, *et al*, 2020) untuk mengetahui hasil yang dibangun pada sistem dapat diterima dan memiliki kemudahan desain dan efisiensi yang baik, dan algoritma Naïve Bayes dan *cosine similarity* (Nugraha & Abdulloh, 2022) untuk klasifikasi hoax berbahasa Indonesia. Penelitian terdahulu menyatakan bahwa algoritma Naïve Bayes memiliki hasil yang lebih baik daripada *cosine similarity* dalam klasifikasi berita hoax. Sehingga berdasarkan pemaparan di atas penulis menggunakan metode Naïve Bayes sebagai algoritma pengklasifikasian sentimen positif, negatif, dan netral pada kasus percakapan berita hoax presiden Joko Widodo 3 periode

Sehingga berdasarkan latar belakang di atas penelitian ini bertujuan melakukan analisis sentimen berita hoax pada aplikasi WhatsApp menggunakan metode Naïve Bayes. Hasil analisis sentimen yang dilakukan, diharapkan dapat membantu permudah pemangku kepentingan Komisi Pemilihan Umum daerah Jawa Timur Kota Malang dalam menyingkap kejahatan penyebaran berita hoax.

## 2. Metode Penelitian

### 2.1 Sumber Data

Data yang digunakan adalah berasal dari skenario dan simulasi, yang mana 2 orang yang melakukan penyebaran berita *hoax* presiden Joko Widodo 3 periode melakukan komunikasi melalui *smartphone* dan menggunakan aplikasi media sosial WhatsApp. Menurut (Wahyuni, *et al*, 2021) berita hoax merupakan sebuah pemberitaan yang kegiatannya berusaha untuk menipu pembaca maupun pendengarnya dalam meyakini suatu informasi tersebut. Informasi yang didapat dapat dijadikan bukti digital, menurut (Riskiyadi, *et al*, 2020) bukti digital adalah setiap informasi yang dapat disimpan maupun disalurkan dalam bentuk digital yang mana pihak yang bersangkutan dapat menjadikannya kasus hukum yang kemudian digunakan untuk pemeriksaan pengadilan. Pada WhatsApp tersebut berisi percakapan yang mengandung unsur dukungan kepada presiden Joko Widodo agar beliau dapat menjabat lagi menjadi 3 periode dan total percakapan tersebut berjumlah 80 percakapan dan percakapan tersebut di *screenshot* sebanyak 6 file dan dikumpulkan pada folder pada laptop yang telah disiapkan sebelumnya.

### 2.2 Ekstraksi Data

Ekstraksi data merupakan proses untuk memperoleh kalimat yang ada pada file gambar yang telah dikumpulkan, kalimat tersebut berisi percakapan yang mengandung unsur berita hoax presiden Joko Widodo 3 periode. Pada prosesnya, ekstraksi data dilakukan menggunakan aplikasi desktop python yang mana menurut (Suraya & Sholeh, 2021) python merupakan salah satu bahasa

pemrograman yang paling direkomendasikan untuk para ilmuwan, terkhusus *big data*, *data mining*, *deep learning data science*, dan *machine learning*. Maka dari itu peneliti untuk memproses ekstraksi menggunakan sistem *Optic Recognition Character* (OCR). Kemudian aplikasi desktop python dapat melakukan ekstraksi sebanyak 6 file text sesuai jumlah file gambar yang dikumpulkan sebelumnya, data tersebut kemudian di simpan pada folder khusus untuk dapat melakukan tahap filter manual dan pelabelan manual.

### 2.3 Filter dan Pelabelan Data

Untuk dapat melakukan proses klasifikasi data, data yang telah diperoleh dilakukan filter dan dilabeli secara manual. Aturan pelabelan mengikuti Kamus Besar Bahasa Indonesia (KBBI). Dalam prosesnya, data atau file gambar yang telah di ekstrak menggunakan aplikasi desktop python menjadi file text dimasukkan ke file excel dan dilakukan filter manual, bertujuan untuk menghilangkan karakter yang tidak perlu, kemudian dilakukan pelabelan manual oleh teman peneliti yang ahli bahasa.

### 2.4 Text Preprocessing

Data yang terkumpul akan dilakukan proses *preprocessing* yang bertujuan untuk menyeragamkan dan mengurangi volume kata yang tidak mempengaruhi klasifikasi (Hafiz Yunas & Fikry, 2018). Proses *preprocessing* memiliki beberapa tahap diantaranya *case folding*, *tokenizing*, *stopword removal*, dan *stemming* menurut (Randhika, *et al*, 2021). Proses *preprocessing* tersebut merupakan pendekatan dari *text mining* yang menurut (Roufia, *et al*, 2018) merupakan definisi dari menambang data berupa teks yang mana datanya bersumber dari data, kemudian bertujuan untuk melakukan pencarian kata yang mewakili isi data tersebut sehingga dapat mewakili keterkaitan antar data.

Tahapan *case folding* adalah suatu teknik yang mengubah seluruh huruf yang ada pada dokumen menjadi huruf kecil (*lower case*) teknik ini disebut normalisasi data. Karakter yang tidak penting seperti tanda baca, angka, dan karakter lain tersebut akan dihilangkan.

Tahapan selanjutnya adalah *tokenizing*, ialah suatu proses yang melakukan pembagian sebuah *string* masukkan menjadi unit terkecil (kata) yang menyusunnya, kemudian *stopword removal* merupakan tahap pada proses pemeriksaan kategori kata yang muncul, apabila kategori kata tersebut termasuk *stop list* maka mengarah pada sekumpulan kata yang tidak memiliki makna tertentu, sedangkan apabila kategori kata tersebut termasuk *word list* maka mengarah pada sekumpulan kata yang memiliki arti sebaliknya. Lalu yang terakhir adalah tahapan *stemming* yang merupakan proses kata dasar dari suatu kata merupakan penentuannya yang menyesuaikan struktur atau tata bahasa yang

digunakan, kata tersebut diantaranya memiliki imbuhan, akhiran, awalan, sisipan, awalan dan akhiran. Berikut contoh kata imbuhan: Memakan = makan; Menyapu = sapu.

### 2.5 Term Frequency

Term frequency merupakan bagian dari pendekatan *text mining* dan analisis teks yang bertujuan untuk melakukan pembobotan kata dan bertujuan untuk mengukur seberapa sering suatu kata tersebut muncul dalam suatu dokumen. (Habib Kusuma, *et al*, 2023). Bobot *tf* dapat dihitung menggunakan Persamaan 1.

$$TF_{td} = (t_i) \tag{1}$$

Berikut merupakan Tabel 1 yang menunjukkan suatu pembobotan setiap kata.

Tabel 1. Pembobotan TF

Term	TF			
	t1	t2	t3	tf
pak	1	0	0	1
presiden	1	0	0	1
jokowi	1	0	0	1
tiga	1	0	0	1
periode	1	0	0	1
ada	0	1	0	1
berita	0	1	0	1
baru	0	1	0	1
kirim	0	1	0	1
mas	0	0	1	1
suka	0	0	1	1
berita	0	0	1	1
politik	0	0	1	1

### 2.6 Naïve Bayes Classification

Naïve Bayes Classification adalah salah satu metode klasifikasi yang berdasar pada teorema Bayes, yang mana metode klasifikasi ini digunakan untuk menentukan kelas suatu item berdasarkan fitur-fiturnya. Kemudian Naïve Bayes juga memiliki akurasi yang tinggi meskipun berupa algoritma sederhana. Menurut Merinda Lestandy, *et al* (2021) metode ini menggunakan perhitungan probabilitas.

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \tag{2}$$

Setiap dokumen selalu digambarkan berupa pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” yang mana atribut  $a_1$  adalah kata pertama, kemudian  $a_2$  merupakan kata kedua dan seterusnya, dan untuk himpunan kategori merupakan  $V$  menurut (Ariyanti, *et al*, 2020). Ketika proses klasifikasi algoritma, algoritma ini mencari nilai probabilitas tertinggi terhadap semua kategori yang dilakukan pengujian ( $V_{map}$ ). Agar dapat mencari nilai probabilitas tertinggi Persamaan 3 dapat digunakan sebagai berikut.

$$V_{MAP} = \underset{v \in V}{\text{arg max}} \left( \frac{P(x_1, x_2, x_3, \dots, x_n | v_j) p(v_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \tag{3}$$

Untuk atribut  $P(x_1, x_2, x_3, \dots, x_n)$  nilai bersifat konstan pada semua kategori ( $V_j$ ) sehingga persamaan dapat ditulis sebagai berikut:

$$V_{MAP} = \underset{v \in V}{\text{arg max}} P(x_1, x_2, x_3, \dots, x_n | v_j) P(v_j) \tag{4}$$

Proses penyederhanaan dapat dilakukan sebagai berikut:

$$V_{MAP} = \underset{v \in V}{\text{arg max}} \prod_{i=1}^n P(x_i | v_j) P(v_j) \tag{5}$$

Keterangan:

$V_j$  = Kategori berita

$P(X_i | V_j)$  = Peluang kemunculan dokumen yang memiliki kategori  $j$

Langkah berikutnya ialah, melakukan perhitungan probabilitas setiap kelas  $j$ , berikut adalah rumusnya:

$$P(V_j) = \frac{|docs_j|}{|data training|} \tag{6}$$

Keterangan:

$P(V_j)$  = Peluang kemunculan dokumen pada kategori  $j$

$|docs_j|$  = Jumlah dokumen pada setiap kategori  $j$

$|data training|$  = Jumlah pada semua kategori dokumen

Dan berikut adalah tahap akhir dalam menentukan perhitungan probabilitas kata  $x_i$  terhadap data testing kepada data uji pada setiap kelas  $j$ , dengan cara:

$$P(X_i | V_j) = \frac{n_k + 1}{n + |kosakata|} \tag{7}$$

Keterangan:

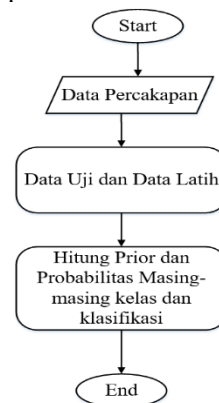
$P(X_i | V_j)$  = Peluang kemunculan  $X_i$  pada kategori  $V_j$

$n_k$  = Jumlah kemunculan frekuensi setiap *term*

$n$  = Jumlah kemunculan frekuensi setiap *term* pada setiap kategori

$|kosakata|$  = Jumlah semua term dari semua kategori

Dalam klasifikasi data percakapan berikut adalah tahapan proses algoritma Naïve Bayes dalam bentuk *flowchart* pada Gambar 1.



Gambar 1. Flowchart Klasifikasi Percakapan Naïve Bayes

### 2.7 Confusion Matrix

*Confusion Matrix* merupakan sebuah tabel yang digunakan untuk menentukan kinerja dari suatu algoritma klasifikasi. *Confusion matrix* merupakan uji klasifikasi jumlah data benar dan jumlah data salah (Normawati, *et al*, 2021). Dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix

	Kelas Prediksi	
	1	0
Nilai sebenarnya	1 TP FN	0 FP TN

Keterangan:

TP (*True Positive*) = Data positif yang nilainya benar  
 TN (*True Negative*) = Data negatif yang nilainya benar

FP (*False Positive*)= Data negatif yang nilainya data positif

FN (*False Negative*)= Data positif yang nilainya data negatif

Adapun beberapa parameter guna untuk menghitung performa klasifikasi yaitu *accuracy*, *precision*, dan *recall*. *Accuracy* adalah perkiraan daripada kinerja rasio dengan benar dari total observasi. Berikut rumus perhitungan *accuracy*.

$$accuracy = \frac{TP+TN}{Total} \tag{8}$$

*Precision* adalah perkiraan yang benar daripada rasio observasi positif dari total observasi positif yang diprediksi. Berikut rumus perhitungan *precision*.

$$precision = \frac{TP}{TP+FP} \tag{9}$$

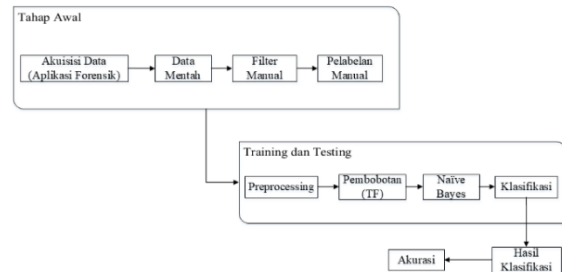
*Recall* adalah jumlah observasi positif yang perkiraannya diamati dengan benar untuk semua kelas sebenarnya, hal tersebut merupakan *sensitivity*. Berikut perhitungan nilai *recall*.

$$recall = \frac{TP}{TP+FN} \tag{10}$$

### 2.8 Deskripsi Sistem

Klasifikasi dilakukan untuk mempermudah peneliti dalam mengkategorikan sebuah teks percakapan yang dijadikan sebagai alat bukti digital sebuah kasus kejahatan digital, sistem klasifikasi dibuat menggunakan *framework* laravel yang mana menurut (Nugroho & Nugroho, 2021) merupakan salah satu *framework* yang dapat memaksimalkan penggunaan bahasa pemrograman PHP untuk pengembangan website. *Framework* ini dibuat untuk mempermudah pengembangan website dengan arsitektur MVC (Model View Controller) yang merupakan arsitektur tersebut merupakan pendekatan perangkat lunak yang memisahkan antara logika dan presentasi dan juga PHP menjadi lebih powerful, cepat, aman, dan sederhana. Sistem memiliki beberapa proses yaitu pengumpulan data teks

percakapan yang diperoleh melalui proses ekstrak sentimen percakapan dengan menggunakan aplikasi desktop Python. Pada Gambar 2 berikut ini menjelaskan bahwa sistem yang dibangun dalam penelitian ini merupakan sistem untuk mengklasifikasikan teks percakapan yang dilakukan oleh peneliti.



Gambar 2. Deskripsi Sistem

Proses selanjutnya yaitu filter manual, setelah itu melakukan proses pelabelan manual yang di kategorikan percakapan positif, negatif, dan netral, adapun syarat data diberikan label positif yaitu apabila terdapat kalimat yang mengandung penyebaran berita hoax presiden Joko Widodo 3 periode. Sedangkan syarat data diberikan label negatif yaitu apabila tidak terdapat kalimat yang mengandung penyebaran berita *hoax* presiden Joko Widodo 3 periode dan syarat data diberikan label netral apabila percakapan tidak membahas tentang berita *hoax* presiden Joko Widodo 3 periode. Langkah selanjutnya yaitu proses *pre-processing* pada dataset, yang mana pada tahap ini meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Setelah melakukan *pre-processing* pada dataset, langkah selanjutnya sistem melakukan proses perhitungan TF yang menghitung probabilitas kata setiap percakapan sehingga dapat menjadi masukan untuk perhitungan ke dalam Naïve Bayes. Nilai TF tersebut dapat digunakan sebagai perhitungan metode dan menghasilkan kategori pada tiap data *training* maupun *testing*. Data *training* dan data *testing* diklasifikasikan setelah mendapatkan nilai masing-masing pada proses sebelumnya dan menjadi acuan untuk hasil klasifikasi pada sistem. Pada sistem ini data *training* dan data *testing* diacak secara *random* oleh sistem dengan perbandingan 50:50.

### 2.9 Pengujian

Mengukur kinerja aplikasi desktop python dan web klasifikasi naïve bayes laravel dalam penelitian ini pada aplikasi desktop python di uji untuk melakukan ekstrak file gambar yang berisi 80 percakapan menjadi file text. File gambar yang diekstrak berjumlah 6 file, kemudian untuk menguji web klasifikasi laravel dalam memperoleh nilai akurasi, *precision*, dan *recall* menggunakan teknik *confusion matrix*. Akurasi adalah pengukuran seberapa akurat algoritma mengklasifikasikan dengan benar. Presisi

menggambarkan rasio prediksi benar positif dibandingkan dengan semua hasil prediksi positif, recall adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Pada penelitian ini rasio data latih dan data uji digunakan secara random dengan perbandingan 50:50. Kemudian untuk menguji fungsional dan kesesuaian hasil serta kelayakan dalam penggunaan aplikasi desktop python menggunakan metode *User Acceptance Test (UAT)* dan responden yang dibutuhkan berjumlah 3 orang.

### 3. Hasil dan Pembahasan

#### 3.1 Hasil Analisis Ekstrak

Untuk hasil ekstraksi yang dilakukan peneliti menggunakan aplikasi desktop python berhasil namun proses ekstraksi sangat bergantung pada kualitas gambar yang digunakan, hal tersebut berpengaruh pada performa aplikasi, dan kalimat yang berhasil di ekstraksi tersebut berjumlah 75 kalimat percakapan dari total 80 kalimat percakapan yang ada pada file gambar tersebut dan 75 kalimat percakapan tersebut telah dilakukan proses filter manual dan pelabelan manual.

#### 3.2 Hasil

Hasil pengujian pada klasifikasi algoritma Naïve Bayes pada kasus penyebaran berita hoax presiden 3 periode dengan menggunakan 50 persen data *training* dan 50 persen data *testing* secara acak mendapat nilai akurasi 88 persen, nilai presisi mendapatkan nilai sebesar 84 persen dan nilai recall mendapatkan nilai sebesar 91,3 persen dan dikarenakan jumlah kelengkapan kalimat pada setiap percakapan yang dilakukan pengujian kemungkinan dapat ditingkatkan dengan melengkapi kalimat pada setiap percakapan. Kelengkapan kalimat ialah yang memperhatikan subjek, predikat, dan objek. Maka dari itu berdasar dari hasil klasifikasi tersebut dapat dilihat pada Gambar 3 berikut.



Gambar 3 Hasil Klasifikasi Data

Berdasarkan pada Gambar 3, diagram batang positif lebih besar daripada diagram batang negatif hal tersebut dikarenakan dari hasil perbandingan analisis manual dan analisis tf yang mana pada analisis manual dan analisis tf positif masing-masing memiliki nilai untuk analisis manual berjumlah 30

sedangkan analisis tf berjumlah 35. Kemudian untuk diagram batang negatif masing-masing memiliki nilai untuk analisis manual negatif berjumlah 10 sedangkan analisis tf berjumlah 7, sedangkan diagram batang netral masing-masing memiliki nilai untuk analisis manual netral berjumlah 35 sedangkan analisis tf berjumlah 33.

Hasil daripada kemampuan Aplikasi Desktop Python untuk melakukan ekstrak file gambar terhadap *smartphone* dianggap kurang maksimal, sebab pada proses ekstrak file gambar terdapat kualitas gambar yang kurang bagus yang dimana hal tersebut berdampak pada hasil daripada proses ekstrak tersebut ketika telah menjadi file text, sehingga kemampuan Aplikasi Desktop Python untuk melakukan ekstrak file bergantung pada kualitas gambar dan berdampak pada kalimat yang ada pada file yang berhasil diekstrak yaitu dari 80 kalimat sentimen yang ada dalam file gambar percakapan tersebut hanya dapat melakukan ekstrak total 75 kalimat sentimen yang ada dalam file gambar percakapan tersebut.

Kemudian algoritma naïve bayes dapat digunakan untuk mengklasifikasikan percakapan kedalam positif, negatif dan netral tentang penyebaran berita *hoax* presiden Joko Widodo 3 periode dan hasilnya adalah positif, dan perentase yang dihasilkan berdasarkan porsi pengujian 50:50 ialah 88% akurasi, 84% *precision*, 91,3% *recall*, dan yang terakhir hasil daripada pengujian eksternal menggunakan metode *User Acceptance Test (UAT)* dengan jumlah responden (*user*) 3 orang menghasilkan nilai persentase sebesar 50% dengan kategori sangat setuju, 40% setuju, dan 7,1% tidak setuju, dari hasil pengujian tersebut dapat disimpulkan bahwa aplikasi desktop python yang telah dibuat berfungsi, sesuai dan layak untuk digunakan. Namun untuk mendapatkan hasil yang maksimal perlu adanya pengembangan lebih lanjut dan semoga penelitian ini dapat dikembangkan untuk kepentingan yang lebih kompleks serta bermanfaat untuk digunakan bagi pemangku kepentingan penelitian yang berkaitan.

### 4. Kesimpulan dan Saran

Kesimpulan dari penelitian diatas bahwa Algoritma Naïve Bayes dapat digunakan untuk mengklasifikasikan percakapan kedalam positif dan negatif tentang penyebaran berita hoax presiden 3 periode dan hasilnya cenderung positif. Hasil uji akurasi daripada klasifikasi ialah tentang kasus penyebaran berita hoax presiden 3 periode yang mana menggunakan porsi pengujian 50:50 sebesar 88% Accuracy, 94% Precision, 90% Recall.

Saran daripada penelitian ini melakukan penelitian pencarian bukti percakapan maupun opini publik lebih lanjut pada aplikasi media sosial yang lain. Kemudian menggunakan jenis *smartphone* yang berbeda dari penelitian ini dan menggunakan versi android yang terbaru atau sistem operasi lain seperti

iOS. Lalu aplikasi dapat dikembangkan dengan memaksimalkan kualitas gambar dengan menambahkan fitur lainnya atau menggunakan sistem lain selain daripada OCR dan juga ditambahkan fitur agar dapat melakukan analisis sentimen suatu percakapan maupun opini public. Kemudian yang terakhir sistem dapat dikembangkan dengan menambahkan *preprocessing* yang lebih banyak sehingga mendapatkan noise yang sedikit.

#### Daftar Pustaka:

- Anggraini, N., Masruroh, S. U., & Tiaraningtias, H. (2020). Analisa Forensik Whatsapp Messenger Pada Smartphone Android. *Jurnal Ilmiah FIFO*, 12(1), 83. <https://doi.org/10.22441/fifo.2020.v12i1.008>
- Ariyanti, D., Iswardani, K., & Rafidah, S. (2020). Klasifikasi Penanganan Keluhan Masyarakat Kota Probolinggo Menggunakan Algoritma Naive Bayes. *J-SAKTI (Jurnal Sains ...*, 4(September), 424–433.
- Asosiasi Penyelenggara Jasa Internet Indonesia. (2022). Survei Penetrasi & Profil Pelaku Pengguna Internet Indonesia. Diakses dari <https://survei.myapjii.id/>.
- Habib Kusuma, I., & Cahyono, N. (2023). *Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor*. 8(3).
- Hafiz Yunas, A., & Fikry, M. (2018). Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine. *Jurnal CoreIT*, 4(2).
- Kautsar, A., & Syafrullah, M. (2022). Implementasi Algoritme Multinomial Naïve Bayes Pada Analisis Sentimen Terhadap Isu Presiden 3 Periode. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia, September 2022*, 675–682.
- Kurniawan, I., & Susanto, A. (2019). Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019. *Eksplora Informatika*, 9(1), 1–10. <https://doi.org/10.30864/eksplora.v9i1.237>
- Merinda Lestandy, Abdurrahim Abdurrahim, & Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Nugraha, A. Y. A., & Abdulloh, F. F. (2022). Optimasi Naive Bayes dan Cosine Similarity Menggunakan Particle Swarm Optimization Pada Klasifikasi Hoax Berbahasa Indonesia. *Jurnal Media Informatika Budidarma*, 6(3), 1444. <https://doi.org/10.30865/mib.v6i3.4170>
- Nugroho, H. E., & Nugroho, A. (2021). Analisis Dan Perancangan E-Commerce Pada Toko Sepatu Dope13Store Menggunakan Framework Laravel. *Information System Journal*, 4(1), 38–44. <https://doi.org/10.24076/infosjournal.2021v4i1.565>
- Priyatna, B., Lia Hananto, A., Nova, M., Studi Sistem Informasi, P., & Buana Perjuangan Karawang, U. (2020). Application of UAT (User Acceptance Test) Evaluation Model in Minggon E-Meeting Software Development. *Systematics*, 2(3), 110–117.
- Randhika, M. N., Young, J. C., Suryadibrata, A., & Mandala, H. (2021). Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online. *Ultimatics : Jurnal Teknik Informatika*, 13(1), 19–25. <https://doi.org/10.31937/ti.v13i1.1921>
- Riskiyadi, M. (2020). *INVESTIGASI FORENSIK TERHADAP BUKTI DIGITAL DALAM MENGUNGKAP CYBERCRIME* (Vol. 3, Issue 2).
- Roufiah, A. (2018). *Text Mining Dengan Metode Naïve Bayes Classifier Untuk Mengklasifikasi Berita Berdasarkan Konten*. 8.
- Suraya, S., & Sholeh, M. (2021). Designing and Implementing a Database for Thesis Data Management by Using the Python Flask Framework. *International Journal of Engineering, Science and Information Technology*, 2(1), 9–14. <https://doi.org/10.52088/ijesty.v2i1.197>
- Wahyuni, W. S., Fithri, B. S., & ... (2021). Sosialisasi Sanksi Penyebaran Berita Bohong (Hoax) berdasarkan Undang-undang Informasi dan Transaksi Elektronik. *Jurnal ABDIMAS ...*, 2(1), 42–45.