



Klasifikasi sentimen terhadap larangan pernikahan beda agama menggunakan metode *Naive Bayes Classifier*

Muhammad Rizki Syafapri^{*1}, Elin Haerani², Iwan Iskandar³, Liza Afriyanti⁴

Email: ¹12050110483@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³iwan.iskandar@uin-suska.ac.id,

⁴liza.afriyanti@uin-suska.ac.id

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Diterima: 18 Februari 2024 | Direvisi: - | Disetujui: 27 April 2024
©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Pernikahan beda agama masih menjadi isu kontroversial bagi masyarakat Indonesia yang multikultural. Aturan Undang-Undang Nomor 1 Tahun 1974 yang melarang pernikahan beda agama memicu berbagai tanggapan di masyarakat. Pada tahun 2023, Mahkamah Agung (MA) mengeluarkan putusan yang melarang pengadilan agama mencatatkan pernikahan beda agama. Hal ini semakin memperkuat kontroversi dan memicu berbagai reaksi dari masyarakat. Platform seperti Instagram menjadi wadah bagi masyarakat untuk mengekspresikan berbagai sentimen mereka terhadap isu ini, mulai dari dukungan, penolakan, hingga pertanyaan dan keraguan. Penelitian ini mengklasifikasikan 1000 komentar Instagram yang dikumpulkan dari lima akun media sosial berita. Komentar tersebut dilabeli secara manual oleh seorang pakar yang berprofesi sebagai dosen bahasa Indonesia, sehingga terbagi menjadi 500 komentar positif dan 500 komentar negatif. Setelah melalui proses *text preprocessing* dan pembobotan TF-IDF, metode *Naive Bayes Classifier* berhasil mencapai tingkat akurasi tertinggi sebesar 76% dengan menggunakan 10% data uji dari dataset untuk mengklasifikasikan sentimen masyarakat terhadap larangan pernikahan beda agama pada komentar Instagram.

Kata kunci: pernikahan beda agama, sentimen, Instagram, naive bayes

Sentiment classification of interfaith marriage ban using Naive Bayes Classifier method

Abstract

Interfaith marriage is still a controversial issue in multicultural Indonesian society. The provisions of Law Number 1 of 1974 which prohibit interfaith marriages have triggered various responses in society. In 2023, the Supreme Court (MA) decided to prohibit religious courts from registering interfaith marriages. This further strengthened the controversy and sparked various reactions from the public. Platforms like Instagram have become a forum for people to express their various sentiments regarding this issue, ranging from support, rejection, to questions and doubts. This research classifies 1000 Instagram comments collected from five news social media accounts. These comments were labeled manually by an expert who works as an Indonesian language lecturer, so they were divided into 500 positive comments and 500 negative comments. After going through the text preprocessing process and TF-IDF weighting, the Naive Bayes Classifier method succeeded in achieving the highest level of accuracy of 76% by using 10% of test data from the dataset to classify public sentiment towards the ban on interfaith marriage in Instagram comments.

Keywords: interfaith marriage, sentiment, Instagram, naive bayes

1. PENDAHULUAN

Kehadiran Undang-Undang Nomor 1 Tahun 1974 tentang perkawinan memiliki keterbatasan dalam memberikan perlindungan hukum yang menyeluruh terhadap berbagai aspek pernikahan, terutama dalam hal pernikahan beda agama di Indonesia. Indonesia merupakan sebuah negara multikultural yang kaya akan keragaman budaya, bahasa, suku bangsa, dan agama [1]. Meskipun suatu pernikahan diakui berdasarkan norma agama, penting untuk diingat bahwa hak asasi manusia menggarisbawahi hak dasar individu untuk mengekspresikan kebebasan dalam menjalani perkawinan dan memilih agama sebagai hak dasar yang tidak dapat dicampuri oleh siapapun [2].

Pernikahan beda agama terjadi ketika dua orang yang berbeda agama menikah sambil mempertahankan agama masing-masing [3]. Masyarakat Indonesia yang beragam sudah tidak asing dengan pernikahan beda agama. Pernikahan semacam ini sudah ada sejak lama dan telah terjadi di berbagai lapisan sosial. Meskipun demikian, perlu diakui bahwa isu pernikahan beda agama selalu menjadi perdebatan di kalangan masyarakat [4].

Larangan pernikahan beda agama kemudian menimbulkan banyak opini dan tanggapan dari kalangan masyarakat di media sosial. Media sosial memungkinkan penerima informasi untuk memberikan tanggapan, sehingga distribusi informasi tidak hanya berlangsung secara satu arah, tetapi juga dua arah. Hal ini berbeda dengan media lain seperti televisi atau radio yang hanya memungkinkan pengguna untuk menerima informasi tanpa dapat memberikan tanggapan. Tanggapan tersebut dapat berupa pesan berbasis teks melalui kolom komentar. Hampir seluruh media sosial menyediakan kolom komentar agar para pengguna dapat menuliskan tanggapannya terkait informasi yang dipublikasikan [5].

Tanggapan dan opini mengenai larangan pernikahan beda agama memunculkan adanya sentimen positif dan negatif. Sentimen masyarakat terhadap larangan pernikahan beda agama dapat diamati melalui media sosial yang banyak digunakan seperti Instagram. Instagram memiliki fitur komentar yang memungkinkan pengguna untuk memberikan tanggapan, opini, atau masukan terhadap postingan. Jumlah karakter pada komentar Instagram dibatasi hingga maksimal 2200 karakter, berbeda jauh dengan Twitter yang hanya bisa memuat 280 karakter [6].

Komentar di Instagram merupakan sumber informasi berharga yang informatif dan sangat membantu [7]. Hal ini dikarenakan komentar di Instagram merupakan data teks yang dapat dianalisis untuk mengetahui pendapat dan perasaan masyarakat terhadap larangan pernikahan beda agama yang ramai diperbincangkan. Proses ini biasanya dimulai dengan mengumpulkan data yang nantinya akan diproses menggunakan pendekatan *machine learning* [8].

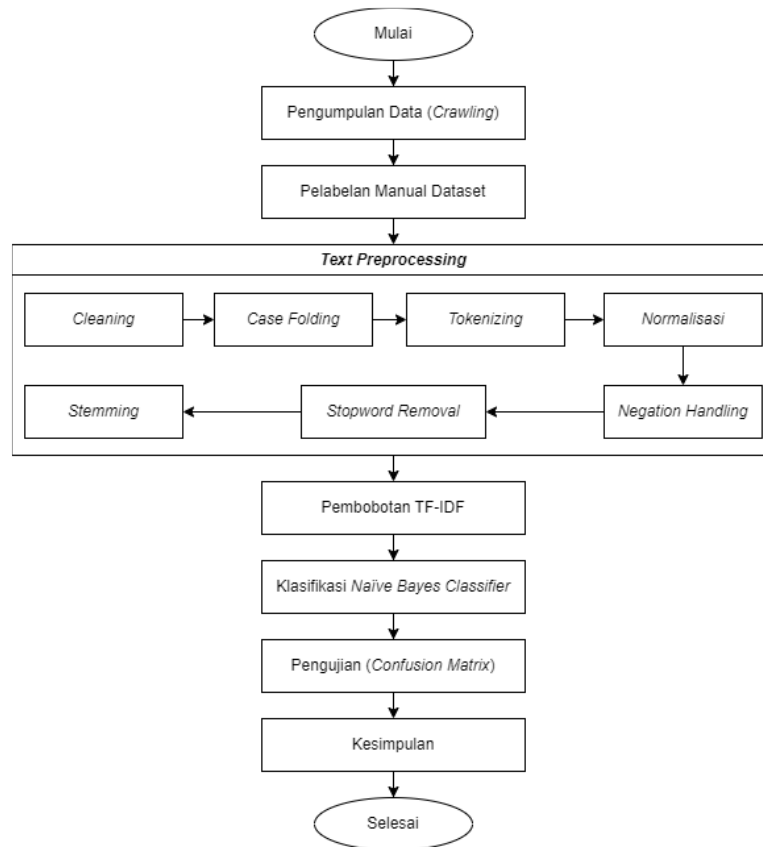
Salah satu pendekatan *machine learning* yang dapat digunakan adalah *Naïve Bayes Classifier*. *Naïve Bayes Classifier* merupakan salah satu metode untuk mengklasifikasikan data dengan probabilitas sederhana yang mengaplikasikan teorema bayes dengan karakter independen yang tinggi [9]. Klasifikasi sentimen positif maupun negatif dengan metode *Naïve Bayes Classifier* menunjukkan hasil yang cukup baik [10].

Beberapa penelitian sebelumnya menunjukkan bahwa *Naïve Bayes Classifier* efektif untuk klasifikasi sentimen dalam berbagai kasus, seperti penelitian dengan kasus terkait komentar netizen terhadap pernyataan pejabat publik pada akun media sosial Instagram dengan data sebanyak 3072 komentar berlabel positif, 1119 komentar berlabel netral, dan 53 komentar berlabel negatif. Dari hasil penelitian tersebut, diperoleh akurasi sebesar 94,12% dengan RapidMiner menggunakan metode *cross-validation* [11]. Selanjutnya, penelitian dengan kasus terkait kenaikan harga bahan pokok menggunakan data sebanyak 2070 tweet dan menghasilkan tingkat akurasi sebesar 94,38%, presisi 59,67%, *recall* 67,93%, dan *f-measure* 62,32% [12]. Selanjutnya, penelitian dengan kasus terkait penilaian masyarakat berdasarkan komentar YouTube terhadap *childfree*, diperoleh sebanyak 2506 data dan menghasilkan persentase akurasi sebesar 97% [13].

Berdasarkan permasalahan yang diuraikan, penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat berdasarkan komentar di Instagram terhadap larangan pernikahan beda agama dengan menggunakan *Naïve Bayes Classifier*. Penelitian ini akan memberikan nilai akurasi untuk klasifikasi sentimen masyarakat terhadap larangan pernikahan beda agama melalui pengujian yang dilakukan.

2. METODE PENELITIAN

Penelitian ini terbagi menjadi enam tahap, dimulai dari pengumpulan data, pelabelan manual dataset, *text preprocessing*, pembobotan kata, klasifikasi *Naïve Bayes Classifier*, hingga pengujian. Gambar 1 menjabarkan alur tahapan penelitian ini menggunakan *flowchart*.



Gambar 1. Flowchart Tahap Penelitian

2.1. Pengumpulan Data

Pengumpulan data meliputi peninjauan terhadap kategori data yang digunakan untuk klasifikasi dan bagaimana mengumpulkan data tersebut. Data yang dikumpulkan dibagi menjadi komentar berlabel positif dan komentar berlabel negatif. Data yang digunakan merupakan komentar dari Instagram yang didapatkan dengan menggunakan alat PhantomBuster. Data sebanyak 2502 komentar berbahasa Indonesia pada postingan Instagram tentang MA resmi melarang semua pengadilan catatkan pernikahan beda agama dikumpulkan pada penelitian ini yang masih bersifat mentah (*raw*) dan belum diketahui labelnya.

2.2. Pelabelan Manual Dataset

Tahap selanjutnya adalah memberi label secara manual terhadap seluruh data dikumpulkan sebelumnya. Label yang diberikan adalah positif dan negatif. Label positif menunjukkan komentar yang setuju atau mendukung kebijakan atau aturan, termasuk apresiasi, penyambutan baik, dan masukan konstruktif. Label negatif menunjukkan komentar yang menentang, menolak, atau mengkritik kebijakan atau aturan dengan cara tidak konstruktif, serta komentar yang berpotensi menimbulkan perpecahan, seperti SARA dan provokasi. Pelabelan manual dataset dilakukan dengan membaca komentar satu per satu dan memberikan label positif atau negatif oleh seorang pakar yang berprofesi sebagai dosen bahasa Indonesia.

2.3. Text Preprocessing

Text preprocessing adalah tahap penting dalam klasifikasi teks yang bertujuan untuk membersihkan dan menyederhanakan teks yang awalnya tidak terstruktur menjadi terstruktur. Proses mengubah data mentah menjadi data yang bersih dan siap untuk digunakan dalam proses klasifikasi disebut sebagai *preprocessing* data [14]. Fungsi dari *text preprocessing* adalah untuk mengurangi *noise* pada dataset [15]. Tahap *text preprocessing* dilakukan melalui tujuh tahapan.

1. *Cleaning*, merupakan tahapan membersihkan kata-kata yang tidak penting, seperti simbol, tanda baca, *username*, URL, *hashtag*, dan emoji.
2. *Case Folding*, merupakan tahapan menghilangkan perbedaan antara huruf besar dan kecil, sehingga seluruh teks diubah menjadi huruf kecil (*lowercase*), serta menyeragamkan format teks untuk memudahkan proses analisis.
3. *Tokenizing*, merupakan tahapan memisahkan kata-kata dalam teks berdasarkan spasi yang ditemukan menjadi token.
4. *Normalisasi*, merupakan tahapan mengubah kata yang semula dalam bentuk tidak baku atau salah ejaan menjadi kata yang baku sesuai dengan kaidah bahasa Indonesia. Salah satu teknik normalisasi yang umum digunakan adalah dengan kamus normalisasi teks untuk kata sehari-hari bahasa Indonesia.
5. *Negation Handling*, merupakan tahapan mengidentifikasi dan menangani negasi dalam teks untuk interpretasi yang akurat. Teknik yang umum digunakan adalah kamus antonim, di mana kata-kata dan antonimnya dicatat untuk mengganti kata yang terkena negasi dengan makna berlawanan.

6. *Stopword Removal*, merupakan tahapan menghapus kata-kata dalam teks yang dianggap tidak relevan. Kata-kata yang dihapus ini disebut sebagai *stopword*.
7. *Stemming*, merupakan tahapan mengubah kata-kata dalam teks menjadi bentuk dasarnya (*stem*) dengan memotong imbuhan yang terdapat pada kata. Imbuhan ini dapat berupa awalan (prefiks), sisipan, akhiran (sufiks), atau kombinasi awalan dan akhiran.

2.4. Pembobotan TF-IDF

Untuk mengetahui nilai frekuensi setiap token yang ada dalam dokumen dapat menggunakan salah satu metode yaitu pembobotan TF-IDF. Frekuensi suatu kata dari suatu istilah dalam dokumen dapat dihitung dengan menggunakan *Term Frequency* (TF). Untuk menentukan seberapa sering suatu istilah digunakan dalam suatu dokumen dapat menggunakan *Inverse Document Frequency* (IDF) [16]. Untuk mengubah data teks menjadi bentuk vektor, penelitian ini menggunakan *library TfidfVectorizer*. Proses ini sangat penting dalam transformasi data teks yang tidak terstruktur menjadi bentuk numerik yang dapat diolah oleh algoritma *machine learning*. *TfidfVectorizer* merupakan sebuah fungsi dalam pustaka *Scikit-learn* untuk ekstraksi fitur teks, memanfaatkan metode TF-IDF dalam memberi bobot pada setiap kata di dalam teks [17]. Bobot ini menunjukkan betapa pentingnya kata tersebut dalam teks. Berikut rumus untuk menghitung TF-IDF.

$$TF = \frac{\text{jumlah kemunculan term dalam dokumen}}{\text{total jumlah kata dalam dokumen}} \quad (1)$$

$$IDF = \log \left(\frac{\text{jumlah dokumen dalam korpus}}{\text{jumlah dokumen yang mengandung term}} \right) \quad (2)$$

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right) \quad (3)$$

Keterangan:

$w_{i,j}$: bobot term i terhadap dokumen j

$tf_{i,j}$: jumlah kemunculan term i dalam dokumen j

N : total jumlah dokumen

df_i : jumlah dokumen yang terdapat term i

2.5. Klasifikasi Naive Bayes Classifier

Naive Bayes Classifier adalah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data tertentu [18]. Teknik ini didasari oleh Teorema Bayes dan dilatih dengan data *training* untuk menghasilkan klasifikasi yang akurat dan memperkirakan probabilitas dari setiap kategori yang terdapat pada ciri dokumen yang diuji [19]. Persamaan (4) menunjukkan rumus metode klasifikasi *Naive Bayes Classifier*.

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)} \quad (4)$$

Persamaan (4) menunjukkan bahwa c merupakan kelas atau label, dan x adalah vektor fitur yang dipakai untuk memprediksi kelas c . $P(c|x)$ menunjukkan *posterior probability* suatu kelas yang diberikan fitur [20]. $P(x|c)$ adalah *likelihood*, yaitu *probability* fitur x jika diketahui kelasnya c . $P(c)$ adalah *prior probability* kelas, mewakili keyakinan awal tentang distribusi kelas. Lalu, $P(x)$ adalah *prior probability* fitur, menggambarkan seberapa umum atau jarang fitur x muncul dalam dataset. Dalam klasifikasi, persamaan ini digunakan untuk memperkirakan probabilitas kelas yang diberikan fitur-fitur yang diamati.

2.6. Pengujian

Data dibagi menjadi data latih (*training*) dan data uji (*testing*) sebelum dilakukan pengujian. Metode *Naive Bayes Classifier* kemudian dilatih pada data latih untuk membangun model klasifikasi [21]. Model ini kemudian diuji pada data uji untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* [22]. Evaluasi performa klasifikasi ini dilakukan dengan menggunakan *confusion matrix*, yang ditampilkan pada Tabel 1 [23].

Tabel 1. Pengujian dengan *Confusion Matrix*

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

Hasil pengujian akan menentukan pembagian data mana yang memiliki tingkat akurasi terbaik. Berikut adalah rumus untuk menghitung nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (8)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data penelitian ini dikumpulkan dengan menggunakan alat PhantomBuster. Sebanyak 2502 komentar dikumpulkan dari 5 akun Instagram, yaitu @folkative, @kompascom, @narasinewsroom, @indozone.id, dan @detikcom. Komentar-komentar tersebut dikumpulkan dari postingan yang membahas tentang putusan Mahkamah Agung (MA) terkait larangan pernikahan beda agama. Data disimpan dalam format file CSV, baik pada proses pengumpulan data maupun proses selanjutnya.

3.2. Pelabelan Manual Dataset

Proses pelabelan data dilakukan secara manual dan divalidasi oleh seorang pakar linguistik yang berprofesi sebagai dosen bahasa Indonesia. Dari 2502 komentar yang dianalisis, terpilih 1000 komentar yang siap digunakan untuk proses klasifikasi sentimen, dengan 500 komentar dinilai sebagai sentimen positif dan 500 komentar lainnya dinilai sebagai sentimen negatif. Gambar 2 menunjukkan tampilan *load* data hasil pelabelan dan analisis.

	username	comment	kelas
0	leonhart_8902	Pdhal scara dogmatis ada agama2 yg bolehin bed...	negatif
1	membummi.22	Kalo apa apa hukum dari agama islam doang. Men...	negatif
2	idabaguswidya	saat Agama di atas HAM memang jadi serba ribet...	negatif
3	made.deo	Daripada ngurisin larangan nikah beda agama, m...	negatif
4	rgn1rst	Tapi di Katolik boleh nikah dispensasi beda ag...	negatif
...	...	...	...
995	allaika.rizqia130116gmail.7	Setujuuu....Ketetapan Allah pastilah yg terbai...	positif
996	nirwanfilez	Alhamdulillah.. agama itu sakral jadi jangan a...	positif
997	riaa.es	Sudah aturan dari Tuhan.. jgn d langgar..	positif
998	_mayyraa80_	@rosdiana9541 dlm agama islam pernikahan beda ...	positif
999	mas.danar_dono	Kalau mau nikah ya silahkan agamanya disamakan...	positif

1000 rows x 3 columns

Gambar 2. Hasil Pelabelan Manual

3.3. Text Preprocessing

Sebelum diklasifikasikan menggunakan metode *machine learning*, data teks dibersihkan dan disusun melalui proses *text preprocessing*. Proses ini dijalankan dengan memanfaatkan *library* Python seperti NLTK dan PySastrawi. Tabel 2 menunjukkan langkah-langkah dan contoh hasil *text preprocessing*.

Tabel 2. Text Preprocessing

Proses	Hasil
Data Mentah	Pirtinyiinnyiiii Cinta Kasih Dan Sayang itu datang dari manaaa??? Klo ternyata takdir membawa pada yang berbeda agama, apakah harus di 'matikan'? Apakah agama tidak mengenal TAKDIR?
Cleaning	Pirtinyiinnyiiii Cinta Kasih Dan Sayang itu datang dari manaaa Klo ternyata takdir membawa pada yang berbeda agama apakah harus di matikan Apakah agama tidak mengenal TAKDIR
Case Folding	pirtinyiinnyiiii cinta kasih dan sayang itu datang dari manaaa klo ternyata takdir membawa pada yang berbeda agama apakah harus di matikan apakah agama tidak mengenal takdir
Tokenizing	['pirtinyiinnyiiii', 'cinta', 'kasih', 'dan', 'sayang', 'itu', 'datang', 'dari', 'manaaa', 'klo', 'ternyata', 'takdir', 'membawa', 'pada', 'yang', 'berbeda', 'agama', 'apakah', 'harus', 'di', 'matikan', 'apakah', 'agama', 'tidak', 'mengenal', 'takdir']
Normalisasi	['pertanyaannya', 'cinta', 'kasih', 'dan', 'sayang', 'itu', 'datang', 'dari', 'mana', 'kalau', 'ternyata', 'takdir', 'membawa', 'pada', 'yang', 'berbeda', 'agama', 'apakah', 'harus', 'di', 'matikan', 'apakah', 'agama', 'tidak', 'mengenal', 'takdir']
Negation Handling	['pertanyaannya', 'cinta', 'kasih', 'dan', 'sayang', 'itu', 'datang', 'dari', 'mana', 'kalau', 'ternyata', 'takdir', 'membawa', 'pada', 'yang', 'berbeda', 'agama', 'apakah', 'harus', 'di', 'matikan', 'apakah', 'agama', 'melupakan', 'takdir']
Stopword Removal	['pertanyaannya', 'cinta', 'kasih', 'sayang', 'takdir', 'membawa', 'berbeda', 'agama', 'matikan', 'agama', 'melupakan', 'takdir']
Stemming	['tanya', 'cinta', 'kasih', 'sayang', 'takdir', 'bawa', 'beda', 'agama', 'mati', 'agama', 'lupa', 'takdir']

Seluruh tahapan *text preprocessing* diimplementasikan sesudah pengumpulan dan pelabelan manual dataset. Tahap terakhir dari proses ini adalah *stemming*. Gambar 3 menunjukkan tampilan Google Colab untuk hasil *text preprocessing* pada seluruh dataset yang digunakan.

	comment	kelas	stemming
0	Pdhal scara dogmatis ada agama2 yg bolehin bed...	negatif	['dogmatis', 'agama', 'boleh', 'beda', 'agama']...
1	Kalo apa apa hukum dari agama islam doang. Men...	negatif	['hukum', 'agama', 'islam', 'mending', 'revisi']...
2	saat Agama di atas HAM memang jadi serba ribet...	negatif	['agama', 'ham', 'serba', 'ribet', 'tau', 'tuh']...
3	Daripada ngurisin larangan nikah beda agama, m...	negatif	['urus', 'larang', 'nikah', 'beda', 'agama', '']...
4	Tapi di Katolik boleh nikah dispensasi beda ag...	negatif	['katolik', 'nikah', 'dispensasi', 'beda', 'ag']...
...	...	...	...
995	Setujuuu....Ketetapan Allah pastilah yg terbai...	positif	['tuju', 'tetap', 'allah', 'baik']...
996	Alhamdulillah.. agama itu sakral jadi jangan a...	positif	['alhamdulillah', 'agama', 'sakral', 'abu', 'a']...
997	Sudah aturan dari Tuhan.. jgn d langgar..	positif	['atur', 'tuhan', 'langgar']...
998	@rosdiana9541 dlm agama islam pernikahan beda ...	positif	['agama', 'islam', 'nikah', 'beda', 'agama', '']...
999	Kalau mau nikah ya silahkan agamanya disamakan...	positif	['nikah', 'sila', 'agama', 'sama', 'atur', 'ub']...

1000 rows x 3 columns

Gambar 3. Hasil *Text Preprocessing* Seluruh Dataset

Proses *text preprocessing* menghasilkan data teks yang dikelompokkan berdasarkan frekuensi kata. Kata-kata dengan frekuensi tertinggi divisualisasikan dalam bentuk *wordcloud* positif dan negatif untuk memahami topik dan sentimen utama dalam data teks.



Gambar 4. *Wordcloud* Komentar Positif dan Negatif

3.4. Pembobotan TF-IDF

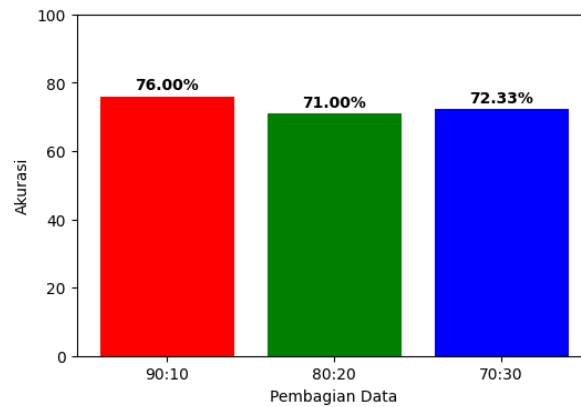
Untuk proses ini dilakukan dengan pustaka Python yang dikembangkan oleh Scikit-learn yaitu TfidfVectorizer. Gambar 5 menunjukkan hasil pembobotan TF-IDF untuk setiap kata dalam data teks. Dalam data teks, kata-kata dengan bobot TF-IDF yang tinggi dianggap penting.

[illegible]

Gambar 5. *Data Frame* Pembobotan TF-IDF

3.5. Klasifikasi Naive Bayes Classifier

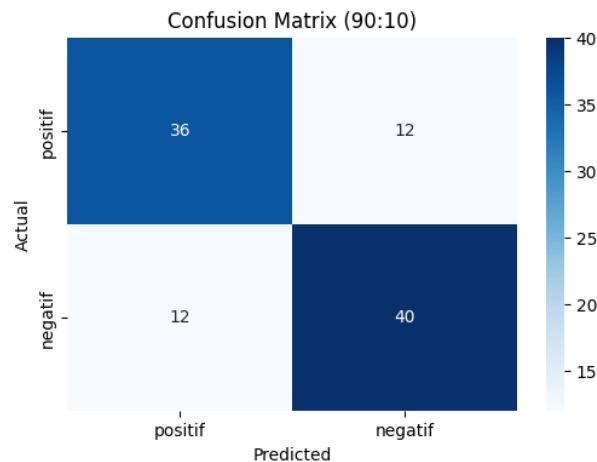
Tiga percobaan klasifikasi Naive Bayes dilakukan dengan skema pembagian data *training* dan data *testing* yang berbeda. Pada pembagian data pertama diperoleh akurasi klasifikasi tertinggi sebesar 76% menggunakan data *training* sebesar 90% dan data *testing* sebesar 10%.



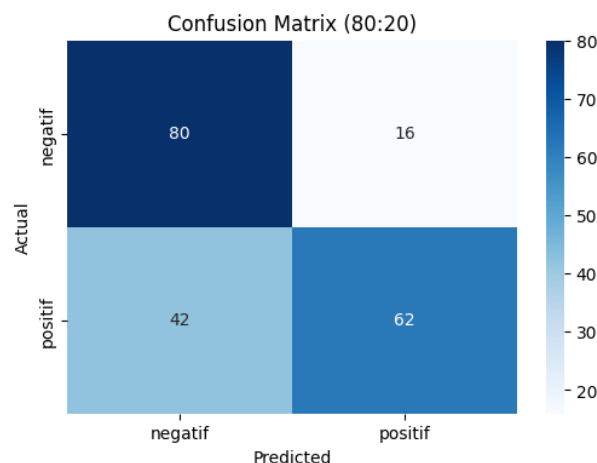
Gambar 6. Klasifikasi Pada Setiap Percobaan

3.6. Pengujian

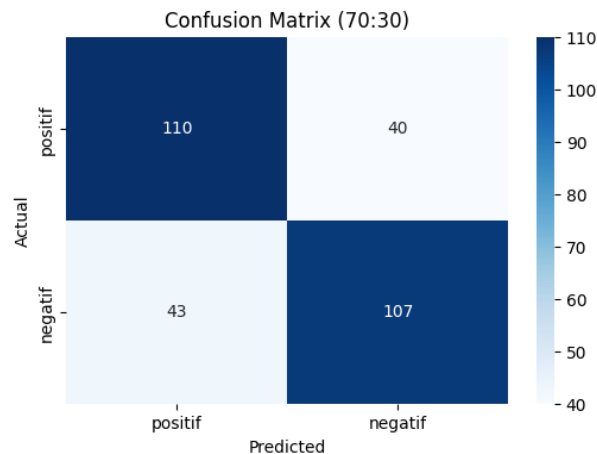
Untuk mengevaluasi kinerja algoritma yang dikembangkan, penelitian ini melakukan pengujian pada 1000 data yang dibagi menjadi data uji dan data latih. Pembagian data dilakukan dengan tiga skenario berbeda. Skenario pertama menggunakan data uji sebesar 10% dan data latih sebesar 90%. Skenario kedua menggunakan data uji sebesar 20% dan data latih sebesar 80%. Sedangkan, skenario ketiga menggunakan data uji sebesar 30% dan data latih sebesar 70%. *Confusion matrix* membandingkan prediksi algoritma (*predicted*) dengan label sebenarnya (*actual*) dari dataset untuk mengevaluasi kinerja algoritma. Gambar 6, 7, dan 8 menampilkan *Confusion matrix* untuk setiap pembagian data.



Gambar 7. *Confusion Matrix* Pada Pembagian Data 90%:10%



Gambar 8. *Confusion Matrix* Pada Pembagian Data 80%:20%

Gambar 9. *Confusion Matrix* Pada Pembagian Data 70%:30%

Tabel 3 menunjukkan hasil pengujian menggunakan *confusion matrix* untuk setiap pembagian data. Hasil ini diperoleh dengan menghitung elemen-elemen dalam *confusion matrix* (Gambar 7, 8, dan 9) dan menggunakan rumus yang telah dijelaskan sebelumnya.

Tabel 3. Pengujian dengan *Confusion Matrix*

<i>Split Data</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
90%:10%	76,00%	76,92%	76,92%	76,92%
80%:20%	71,00%	79,49%	59,62%	68,13%
70%:30%	72,33%	72,79%	71,33%	72,05%

Berdasarkan hasil klasifikasi dan pengujian dengan skenario pembagian data uji sebesar 10%, menghasilkan kinerja algoritma yang paling optimal. Skenario ini menghasilkan nilai *accuracy* 76%, *precision* 76,92%, *recall* 76,92%, dan *f1-score* 76,92%.

4. KESIMPULAN

Penelitian ini menggunakan dataset yang terdiri dari 1000 komentar Instagram, dengan 500 komentar positif dan 500 komentar negatif. Komentar tersebut dikumpulkan dari 5 postingan Instagram akun media sosial berita yang membahas tentang larangan pernikahan beda agama. Komentar tersebut dilabeli secara manual oleh seorang pakar yang berprofesi sebagai dosen bahasa Indonesia. Proses klasifikasi dilakukan dengan metode *Naïve Bayes Classifier* setelah melalui tahapan *text preprocessing* dan pembobotan TF-IDF. Hasil pengujian menggunakan *confusion matrix* menunjukkan bahwa metode tersebut berhasil mencapai kinerja yang baik untuk mengklasifikasikan sentimen masyarakat terhadap larangan pernikahan beda agama pada komentar Instagram, dengan nilai akurasi mencapai 76% dan *f1-score* sebesar 76,92%. Hasil ini diperoleh dengan menggunakan data uji sebesar 10%.

DAFTAR PUSTAKA

- [1] F. P. Samodra, "Sebutkan Keragaman yang Ada di Indonesia, dari Suku Hingga Agama." Accessed: Feb. 10, 2024. [Online]. Available: <https://www.liputan6.com/hot/read/5275820/sebutkan-keragaman-yang-ada-di-indonesia-dari-suku-hingga-agama>
- [2] M. W. Sekarbuana, I. A. P. Widiawati, and I. W. Arthanaya, "Perkawinan Beda Agama dalam Perspektif Hak Asasi Manusia di Indonesia," *J. Prefer. Huk.*, vol. 2, no. 1, pp. 16–21, Mar. 2021, doi: 10.22225/jph.2.1.3044.16-21.
- [3] A. Fuadi and D. A. Sy, "PERNIKAHAN BEDA AGAMA PERSPEKTIF HUKUM ISLAM DAN HUKUM POSITIF DI INDONESIA," *J. Hadratul Madaniyah*, vol. 7, no. 2, pp. 1–14, Dec. 2020, doi: 10.33084/jhm.v7i2.1986.
- [4] A. Amri, "Perkawinan Beda Agama Menurut Hukum Positif dan Hukum Islam," *Media Syari'ah Wahana Kaji. Huk. Islam dan Pranata Sos.*, vol. 22, no. 1, pp. 48–64, May 2020, doi: 10.22373/jms.v22i1.6719.
- [5] K. A. Nugraha, "Analisis Sentimen Berbasis Emoticon pada Komentar Instagram Bahasa Indonesia Menggunakan Naïve Bayes," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 3, pp. 715–721, Dec. 2021, doi: 10.28932/jutisi.v7i3.4094.
- [6] A. Muhaddisi, B. N. Prastowo, and D. U. Kusumaning Putri, "Sentiment Analysis With Sarcasm Detection On Politician's Instagram," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 4, pp. 349–358, Oct. 2021, doi: 10.22146/ijccs.66375.
- [7] R. L. Musyarofah, E. Utami, and S. Raharjo, "Analisis Komentar Potensial pada Social Commerce Instagram Menggunakan TF-IDF," *J. Eksplorasi Inform.*, vol. 9, no. 2, pp. 130–139, Mar. 2020, doi: 10.30864/eksplorasi.v9i2.360.
- [8] F. A. Wenando, R. Hayami, Soni, A. Fitria, and D. F. Shifana, "Sentimen Analisis Masyarakat terhadap Kasus Penembakan Brigadir J Menggunakan Algoritma Naïve Bayes Classifier," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 2, pp. 484–490, Sep. 2023, doi: 10.37859/coscitech.v4i2.5686.
- [9] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform.*, vol. 5, no. 2, pp. 697–711, 2021, doi: 10.30645/j-sakti.v5i2.369.
- [10] N. P. G. Naraswati, D. C. Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, and R. Nooraeni, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sist. J. Sist. Inf.*, vol. 10, no. 1, pp. 222–238, Jan. 2021, doi: 10.32520/stmsi.v10i1.1179.
- [11] A. Anggara, S. Widiono, A. T. Hidayat, and Sutarman, "Analysis of Netizen Comments Sentiment on Public Official Statements on Instagram Social

- Media Accounts,” *Int. J. Adv. Data Inf. Syst.*, vol. 3, no. 2, pp. 87–97, Nov. 2022, doi: 10.25008/ijadis.v3i2.1244.
- [12] M. Muslimin and V. Lusiana, “Analisis Sentimen Terhadap Kenaikan Harga Bahan Pokok Menggunakan Metode Naive Bayes Classifier,” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1200–1209, 2023, doi: 10.30865/mib.v7i3.6418.
- [13] F. Syofiani, S. Alam, and M. I. S. S., “Analisis Sentimen Penilaian Masyarakat Terhadap Childfree Berdasarkan Komentar di Youtube Menggunakan Algoritma Naive Bayes,” *J. Teknol. Inform. dan Komput.*, vol. 9, no. 2, pp. 688–703, Sep. 2023, doi: 10.37012/JTIK.V9I2.1661.
- [14] S. W. Ritonga, Yusra, M. Fikry, and E. P. Cynthia, “Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naive Bayes Classifier,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 134–143, Jun. 2023, doi: 10.47065/bits.v5i1.3535.
- [15] P. Pandunata, C. K. Ananta, and Y. Nurdiansyah, “Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naive Bayes Classifier,” *INFORMAL Informatics J.*, vol. 7, no. 2, pp. 146–156, Aug. 2022, doi: 10.19184/isj.v7i2.33928.
- [16] Z. Firmansyah and N. F. Puspitasari, “Analisis Sentimen Masyarakat Terhadap Vaksinasi COVID-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes,” *J. Tek. Inform.*, vol. 14, no. 2, pp. 171–178, Oct. 2021, doi: 10.15408/JTI.V14I2.24024.
- [17] A. Prananda, E. Haerani, M. Fikry, and F. Yanto, “Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine dalam Analisis Sentimen Terhadap Pemilihan Presiden 2024,” *Krea-TIF J. Tek. Inform.*, vol. 11, no. 2, pp. 84–94, Dec. 2023, doi: 10.32832/KREA-TIF.V11I2.15364.
- [18] A. Y. P. Yusuf and R. Sari, “Implementasi Algoritma Naive Bayes untuk Klasifikasi Pemahaman Program MBKM bagi Mahasiswa,” *J. Inform. Inf. Secur.*, vol. 3, no. 2, pp. 171–180, 2022, Accessed: Mar. 05, 2024. [Online]. Available: <https://ejurnal.ubharajaya.ac.id/index.php/jiforty/article/view/1713>
- [19] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, “Analisis Kinerja Algoritma Naive Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.
- [20] S. H. A. Samsudin, N. M. Sabri, N. Isa, and U. F. M. Bahrin, “Sentiment Analysis on Acceptance of New Normal in COVID-19 Pandemic using Naive Bayes Algorithm,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 581–588, Dec. 2022, doi: 10.14569/IJACSA.2022.0130968.
- [21] S. Saepudin, S. Widiastuti, and C. Irawan, “Analisis Sentimen Ulasan Platform Media Sosial Menggunakan Algoritma Naive Bayes Classifier,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 2, pp. 236–243, Jul. 2023, doi: 10.32736/sisfokom.v12i2.1650.
- [22] M. D. Nurmalasari, Kusri, and Sudarmawan, “Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 52–59, Jul. 2021, doi: 10.31603/komtika.v5i1.5140.
- [23] M. Yusran, S. Siswanto, and A. Islamiyati, “Comparison of Multinomial Naive Bayes and Bernoulli Naive Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking,” *Sist. J. Sist. Inf.*, vol. 13, no. 1, pp. 96–106, Jan. 2024, doi: 10.32520/STMSI.V13I1.3187.