

IMPLEMENTASI DAN PERBANDINGAN ALGORITMA KNN DAN NAIVE BAYES PADA ANALISIS SENTIMEN KOMENTAR YOUTUBE ISU PENGUNDURAN KETUA PBNU

Mohammad Nailul Fahmi¹, Zaehol Fatah², Syarif Aminul Khoiri³

Universitas Ibrahimy

fhmm808@gmail.com

ABSTRAK

Perkembangan media sosial menjadikan platform digital sebagai ruang publik utama untuk menyampaikan opini, termasuk mengenai isu strategis keagamaan dan organisasi seperti pengunduran Ketua Umum PBNU yang memicu beragam reaksi di kolom komentar YouTube. Besarnya volume komentar yang tidak terstruktur membuat analisis sentimen manual menjadi tidak efektif, sehingga dibutuhkan pendekatan komputasi berbasis “*Machine Learning*”. Penelitian ini bertujuan mengimplementasikan serta membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam mengklasifikasikan sentimen warganet. Menggunakan pendekatan kuantitatif berkerangka “*Knowledge Discovery in Database*” (KDD), data dikumpulkan melalui teknik “web scraping” pada video YouTube terkait. Tahapan “*Preprocessing*” mencakup “*Case Folding*”, “*Cleaning*”, “*Tokenizing*”, “*Stopword Removal*”, dan “*Stemming*”, yang dilanjutkan dengan pembobotan “*TF-IDF*”. Evaluasi model dilakukan menggunakan “*Confusion Matrix*” berdasarkan metrik akurasi, “*Precision*”, “*Recall*”, serta “*F1-score*”. Hasil riset ini diharapkan mampu mengidentifikasi algoritma dengan performa optimal, sekaligus memberikan kontribusi nyata dalam pengembangan analisis opini publik berbasis media sosial pada ranah isu keagamaan dan keorganisasian.

Kata Kunci: Analisis Sentimen, K-Nearest Neighbor, Naive Bayes, YouTube, PBNU, Machine Learning

ABSTRACT

The growth of social media has turned digital platforms into public spaces for expressing opinions, including on religious and organizational issues. One issue that has drawn public attention is the resignation of the chairman of the Nahdlatul Ulama Central Board (PBNU), which triggered various public reactions in YouTube comment sections. The large number of unstructured comments makes manual analysis of public perception less effective, so a machine-learning-based sentiment analysis approach is needed. This study aims to implement and compare the performance of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms in classifying the sentiment of YouTube comments related to the PBNU chairman resignation issue. The data were obtained through web scraping of YouTube videos discussing the issue, followed by preprocessing stages consisting of case folding, cleaning, tokenizing, stopwords removal, and stemming. The data were then transformed using TF-IDF weighting before being processed with the KNN and Naive Bayes algorithms. A quantitative approach within the Knowledge Discovery in Database (KDD) framework was used. Model evaluation was carried out using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results are expected to show which algorithm performs best in sentiment classification and to contribute to the development of social-media-based public opinion analysis on religious and organizational issues.

Keywords: Sentiment Analysis, K-Nearest Neighbor, Naive Bayes, YouTube, PBNU, Machine Learning

Pendahuluan

Nahdlatul Ulama (NU) merupakan salah satu organisasi islam yang cukup besar di dunia dengan anggotanya yang tersebar di hampir seluruh Indonesia dan beberapa negara lain melalui PCNU-nya. Nahdlatul Ulama memiliki arti kebangkitan para ulama. Istilah “kebangkitan” itu sendiri pada dasarnya mengandung arti yang lebih aktif jika dibandingkan dengan kata “perkumpulan” atau “perhimpunan”. Seperti kita ketahui, para ulama merupakan panutan umat dimana umat akan mengikutinya. Oleh karena itu, dengan kepemimpinan para ulama, diharapkan arah kebangkitan dan kejayaan umat islam serta kaum muslimin akan lebih terlihat jelas dan nyata. Nahdlatul Ulama disingkat NU, yang artinya kebangkitan ulama. Sebuah organisasi yang didirikan oleh para ulama pada tanggal 31 Januari 1926/16 Rajab 1344 H di kampung Kertopaten Surabaya, dengan Kiai Muhammad Hasyim Asy'ari sebagai Rais Akbar pertamanya [1].

Organisasi kemasyarakatan merupakan bagian penting dalam kehidupan sosial-politik di Indonesia. Salah satu organisasi terbesar yang memiliki pengaruh signifikan adalah Nahdlatul Ulama (NU), dengan struktur kepemimpinan tertinggi yang berada di bawah Pengurus Besar Nahdlatul Ulama (PBNU). Ketua umum merupakan jabatan tertinggi di jajaran tanfidziyah Pengurus Besar Nahdlatul Ulama (PBNU). Di dalam Ensiklopedia NU disebutkan bahwa ketua umum bertugas melaksanakan seluruh kebijakan dan program yang telah digariskan oleh jajaran syuriah PBNU [2].

Sosok ketua umum PBNU yang diangkat dalam penelitian ini adalah Yahya Cholil Staquf yang akrab disapa “Gus Yahya”. Yang saat ini menjabat ketua umum PBNU periode 2021-2026. Yahya Cholil Staquf dipilih melalui proses Mukhtar ke-34 NU di Lampung pada Desember 2021, mengungguli calon petahana. Sebagai Ketua Umum, Gus Yahya menduduki posisi sentral dalam struktur kepengurusan harian (tanfidziyah) PBNU [3]. Sejak kepemimpinannya, Gus Yahya telah tampil aktif, baik dalam urusan internal organisasi maupun dalam dialog

keagamaan, nasional, dan internasional. Karena posisinya yang strategis memimpin PBNU, organisasi islam terbesar di Indonesia, setiap keputusan, tindakan, dan pernyataannya kerap menjadi sorotan publik dan pemicu opini luas di masyarakat. Hal ini menjadikannya sebagai tokoh penting dalam memetakan persepsi publik terhadap isu-isu keagamaan, sosial, dan kebangsaan [4]. Isu pengunduran ketua PBNU yang diangkat dalam penelitian ini menjadi salah satu topik yang ramai diperbincangkan di berbagai kanal media sosial. Peristiwa ini memunculkan berbagai opini mulai dari dukungan, kritik, spekulasi, hingga analisis dari berbagai pihak. Reaksi masyarakat tersebut semakin meluas seiring dapat diamati melalui banyaknya komentar dan diskusi yang muncul pada konten YouTube yang membahas isu pengunduran Ketua PBNU.

Di era digital modern ini, Indonesia mengalami pergeseran dalam pola konsumsi informasi masyarakatnya. Meskipun sebagian besar orang masih menggunakan koran dan majalah sebagai sumber berita, semakin banyak individu yang beralih ke platform digital untuk memperoleh informasi terkini. Kemajuan teknologi, khususnya akses mudah ke internet dan perkembangan media sosial, telah menjadi katalisator utama perubahan ini. Masyarakat Indonesia kini lebih cenderung mendapatkan berita melalui situs web berita, aplikasi berita, dan platform media sosial [5]. YouTube merupakan sebuah media sosial yang memiliki konten berbentuk video, dimana videonya bertujuan sebagai hiburan, pengetahuan dan lain sebagainya. Pengguna YouTube juga dapat membuat, melihat, serta berbagi video secara gratis. Di YouTube juga disediakan fitur komentar sebagai penyalur respons *viewers* terhadap konten video yang ditontonnya. Komentar tersebut bisa bersifat positif atau negatif. Di sisi lain situs YouTube juga dapat memberikan umpan balik berupa *like*, *dislike*, dan komentar terhadap video yang ditayangkan di media YouTube [5].

Analisis sentimen merupakan teknik yang digunakan untuk mengklasifikasikan opini atau emosi yang terdapat dalam teks menjadi kategori seperti positif, negatif, atau netral. Dalam konteks komentar YouTube tentang isu



SENABISTEKES
Seminar Nasional Bisnis Teknologi dan Kesehatan

pengunduran ketua PBNU, analisis sentimen dapat membantu memahami bagaimana masyarakat memandang isu tersebut. Permasalahan utamanya adalah bagaimana mengekstraksi ribuan komentar menjadi informasi sentimen yang terstruktur. Tanpa pendekatan algoritmik yang tepat, persepsi publik hanya dapat dibaca secara sepintas, padahal data tersebut memiliki nilai informasi yang tinggi untuk memahami dinamika opini masyarakat terkait isu sosial-keagamaan. Berdasarkan kondisi tersebut, pemilihan topik ini dilatarbelakangi oleh semakin masifnya penggunaan media sosial sebagai ruang publik dalam menyampaikan opini, serta kebutuhan akan pendekatan analitis untuk memahami persepsi masyarakat secara objektif. Penelitian ini bertujuan untuk mengolah komentar-komentar tersebut menggunakan dua algoritma yang berbeda dan membandingkan performanya dalam mengklasifikasikan sentimen.

Melalui perbandingan kedua algoritma, penelitian ini diharapkan dapat memberikan gambaran algoritma mana yang lebih optimal digunakan dalam analisis sentimen komentar YouTube, sekaligus memberikan kontribusi ilmiah dalam pengembangan analisis opini publik di era digital. Dengan demikian, hasil penelitian ini dapat menjadi rujukan bagi studi-studi lanjutan yang berkaitan dengan dinamika opini publik, isu keagamaan, serta pemanfaatan *machine learning* dalam analisis data media sosial.

Materi dan Metode

Metode penelitian adalah pendekatan sistematis yang digunakan untuk mengumpulkan, mengolah, dan menganalisis data guna mencapai tujuan penelitian secara ilmiah. Pada penelitian ini, metode yang diterapkan bertujuan untuk membangun sebuah model klasifikasi sentimen yang akurat dan mampu menggambarkan persepsi publik terhadap isu pengunduran ketua PBNU berdasarkan komentar video YouTube [6].

1. Jenis Penelitian

Penelitian ini tergolong dalam jenis penelitian kuantitatif dengan pendekatan *data mining*, karena berfokus pada mengolah data

komentar publik dalam volume besar untuk mengklasifikasikan sentimen secara sistematis. Pendekatan ini juga bersifat deskriptif kuantitatif, menggambarkan kecenderungan sentimen publik terhadap pengunduran ketua PBNU. Selain itu, penelitian ini merupakan eksperimen komputasi, dimana algoritma *K-Nearest Neighbor* dan *Naive Bayes* diimplementasikan dengan bahasa pemrograman *Python* dan *framework Streamlit* sebagai pustaka *Python* yang *open source* untuk memudahkan membuat dan berbagi aplikasi web yang sesuai kebutuhan untuk *machine learning* dan *data science* [7].

2. Sampel dan Populasi

Populasi dalam penelitian ini adalah seluruh komentar yang terdapat pada video YouTube yang membahas isu pengunduran ketua Pengurus Besar Nahdlatul Ulama (PBNU), khususnya yang diunggah pada periode November hingga Desember 2025. Video yang dijadikan sumber data adalah video dengan judul “Gejolak di Tubuh PBNU, Desak Gus Yahya Mundur | tvOne” yang dipublikasikan pada tanggal 22 November 2025 melalui kanal YouTube tvOneNews. Populasi ini bersifat dinamis karena jumlah komentar dapat terus bertambah seiring dengan interaksi pengguna terhadap video tersebut.

Sampel penelitian berupa komentar yang berhasil dikumpulkan melalui proses *scraping* dan telah melalui tahap seleksi data. Jumlah data yang digunakan dalam penelitian sebanyak 657 komentar yang terdiri dari 344 komentar negatif, 295 komentar netral, dan 18 komentar positif. Teknik pengambilan sampel menggunakan *purposive sampling*, yaitu memilih komentar yang relevan dengan konteks penelitian dan memiliki keterkaitan langsung dengan isu pengunduran Ketua PBNU.

3. Teknik Pengumpulan Data

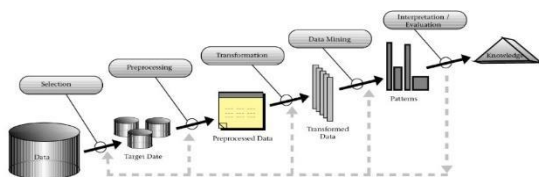
Teknik pengumpulan data dilakukan menggunakan metode *web scraping* terhadap komentar video YouTube terkait isu pengunduran Ketua PBNU. Proses *scraping* dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan pustaka *YouTube*

Comment API untuk mengekstraksi komentar secara otomatis dan efisien. Data yang berhasil dikumpulkan kemudian disimpan dalam format CSV untuk mempermudah proses pengolahan dan analisis data.

Selain pengumpulan data primer melalui *scraping* komentar YouTube, penelitian ini juga menggunakan teknik studi pustaka. Studi pustaka dilakukan dengan mengkaji berbagai sumber seperti jurnal ilmiah, dan referensi terkait analisis sentimen, klasifikasi teks, algoritma KNN, algoritma *Naive Bayes*, serta metode *data mining*. Kajian pustaka digunakan sebagai landasan teoritis dalam pengembangan sistem dan analisis penelitian.

4. Metode Pengembangan Sistem

Penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)* sebagai pendekatan dalam proses analisis data. Metode KDD digunakan untuk menemukan pola dan pengetahuan dari data dalam jumlah besar melalui beberapa tahapan sistematis, yaitu pemilihan data, *preprocessing*, transformasi data, implementasi algoritma, dan evaluasi model [8].



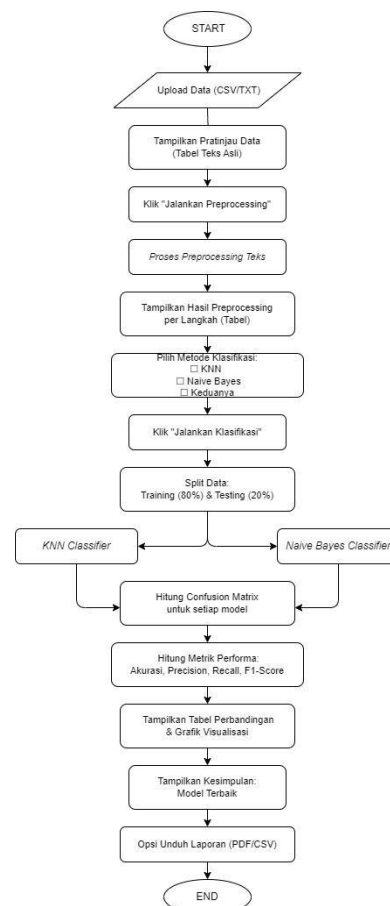
Gambar 1. Knowledge Discovery in Database (KDD)

Tahapan awal dimulai dengan pemilihan dataset berupa komentar YouTube terkait isu pengunduran Ketua PBNU. Data yang diperoleh kemudian dibersihkan melalui tahap *preprocessing* yang meliputi *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*. Selanjutnya data ditransformasikan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Tahap berikutnya adalah implementasi algoritma klasifikasi menggunakan *K-Nearest Neighbor (KNN)* dan *Naive Bayes*. Dataset dibagi menjadi data

training dan *testing* untuk mengukur performa model. Evaluasi dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

5. Perancangan Desain Sistem

Perancangan desain sistem menggambarkan alur proses analisis sentimen secara sistematis mulai dari input data hingga menghasilkan evaluasi model klasifikasi. Sistem yang dibangun terdiri dari beberapa tahapan utama, yaitu *upload data*, *preprocessing* teks, pelabelan sentimen, klasifikasi, evaluasi model, dan visualisasi hasil.



Gambar 2. Flowchart Sistem

Tahap awal dimulai dari proses *upload* data komentar YouTube ke dalam sistem. Setelah data berhasil dimasukkan, sistem menampilkan pratinjau data untuk memastikan data telah terbaca dengan benar. Selanjutnya pengguna menjalankan proses *preprocessing* yang terdiri dari *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*.

Hasil *preprocessing* menunjukkan bahwa seluruh data berhasil diproses dengan baik sebanyak 657 data komentar. Setelah *preprocessing* selesai, data dilakukan pelabelan sentimen ke dalam kategori positif, negatif, dan netral. Distribusi data sentimen menunjukkan dominasi sentimen negatif sebanyak 344 komentar (52,36%), sentimen netral sebanyak 295 komentar (44,90%), dan sentimen positif sebanyak 18 komentar (2,74%).

Tahap berikutnya adalah pembobotan kata menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Pada tahap klasifikasi, sistem membagi dataset menjadi data *training* dan *testing*. Selanjutnya algoritma KNN dan *Naive Bayes* diterapkan untuk melakukan klasifikasi sentimen.

Berdasarkan hasil pengujian *confusion matrix*, algoritma *Naive Bayes* mampu mengklasifikasikan sentimen negatif dengan benar sebanyak 62 data, sentimen netral sebanyak 34 data, dan sentimen positif sebanyak 2 data. Sementara itu, algoritma KNN mampu mengklasifikasikan sentimen negatif sebanyak 54 data, sentimen netral sebanyak 39 data, dan sentimen positif sebanyak 2 data.

Hasil evaluasi model menunjukkan bahwa algoritma *Naive Bayes* memperoleh nilai *accuracy* sebesar 74,24%, *precision* sebesar 79,59%, *recall* sebesar 74,24%, dan *F1-score* sebesar 74,79%. Sedangkan algoritma KNN memperoleh nilai *accuracy* sebesar 71,97%, *precision* sebesar 76,74%, *recall* sebesar 71,97%, dan *F1-score* sebesar 73,88%.

Selain itu, dilakukan pengujian *cross validation* untuk mengetahui tingkat konsistensi performa model. Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* memperoleh nilai *cross validation* sebesar 71,54%, sedangkan algoritma KNN memperoleh nilai sebesar 76,86%.

Tahap akhir sistem menampilkan hasil evaluasi dalam bentuk tabel, *confusion matrix*, dan grafik visualisasi sehingga memudahkan pengguna dalam membandingkan performa

kedua algoritma. Sistem juga menyediakan fitur unduh laporan hasil analisis dalam format CSV sebagai dokumentasi penelitian.

Hasil dan Pembahasan

Pada bagian ini dijelaskan seluruh hasil penelitian yang telah dilakukan, meliputi proses pengumpulan data, *preprocessing* teks, pelabelan sentimen, implementasi algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes*, serta evaluasi kinerja kedua algoritma. Penyajian hasil mengikuti kaidah analisis univariat dan bivariat sesuai dengan jenis penelitian kuantitatif eksperimen komputasi.

Analisis univariat bertujuan untuk menggambarkan karakteristik masing-masing variabel secara deskriptif tanpa melakukan hubungan antar variabel. Dalam penelitian ini, analisis univariat menyajikan gambaran umum data yang meliputi jumlah data, distribusi sentimen, serta hasil *preprocessing* teks.

Karakteristik data penelitian diperoleh dari komentar pengguna YouTube yang membahas isu pengunduran Ketua PBNU Yahya Cholil Staquf. Setelah dilakukan proses *scraping* dan seleksi data, diperoleh sebanyak 657 komentar yang digunakan dalam penelitian. Data kemudian dikelompokkan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral.

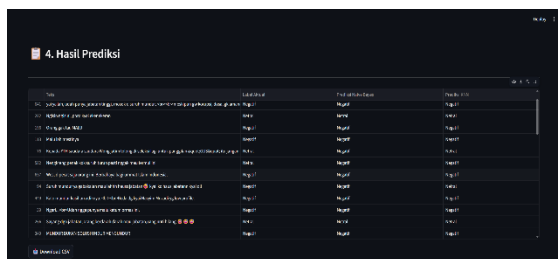
Tabel 1. Distribusi Kategori Sentimen

Kategori Sentimen	N	%
Positif	18	2.74
Negatif	344	52.36
Netral	295	44.90
Total	657	100

Berdasarkan Tabel 1 diketahui bahwa kategori sentimen negatif mendominasi data penelitian dengan jumlah 344 komentar (52,36%), diikuti sentimen netral sebanyak 295 komentar (44,90%), sedangkan sentimen positif hanya sebanyak 18 komentar (2,74%). Hasil tersebut menunjukkan bahwa mayoritas opini publik pada komentar YouTube terkait isu pengunduran Ketua PBNU cenderung bernada negatif.

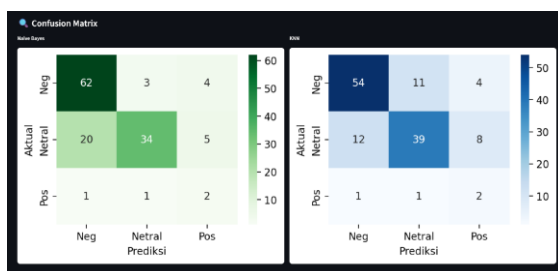
Tahapan *preprocessing* dilakukan untuk membersihkan data teks sebelum proses klasifikasi. Tahapan *preprocessing* meliputi *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*. Seluruh proses *preprocessing* berhasil dilakukan terhadap 657 data komentar sehingga data siap digunakan pada tahap klasifikasi sentimen.

Tahap *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil. Selanjutnya tokenisasi digunakan untuk memecah kalimat menjadi token kata. Tahap *stopword removal* dilakukan untuk menghapus kata-kata umum yang tidak memiliki makna penting, sedangkan *stemming* digunakan untuk mengubah kata berimbuhan menjadi kata dasar.



Gambar 3. Hasil Prediksi Algoritma

Penelitian ini menggunakan dua algoritma klasifikasi, yaitu *Naive Bayes* dan *K-Nearest Neighbor* (KNN). Kedua algoritma diterapkan pada data komentar YouTube yang telah melalui tahap *preprocessing* dan pelabelan sentimen. Hasil klasifikasi menunjukkan bahwa kedua algoritma mampu mengelompokkan komentar ke dalam kategori positif, negatif, dan netral. Sebagian besar komentar berhasil diprediksi sesuai dengan label aktual.



	Actual Neg	Actual Netral	Actual Pos
Predicted Neg	62	20	1
Predicted Netral	3	34	1
Predicted Pos	4	5	2

Gambar 4. Hasil Confusion Matrix

Berdasarkan hasil *confusion matrix*, algoritma *Naive Bayes* mampu mengklasifikasikan sentimen negatif dengan

benar sebanyak 62 data, sentimen netral sebanyak 34 data, dan sentimen positif sebanyak 2 data. Sementara itu, algoritma KNN mampu mengklasifikasikan sentimen negatif sebanyak 54 data, sentimen netral sebanyak 39 data, dan sentimen positif sebanyak 2 data.

Hasil *confusion matrix* menunjukkan bahwa kedua algoritma masih mengalami kesulitan dalam mengenali sentimen positif karena jumlah data positif yang sangat sedikit dibandingkan kategori lainnya. Ketidakseimbangan data memengaruhi kemampuan model dalam melakukan klasifikasi secara optimal.

Tabel 2. Perbandingan Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	74.24%	79.59%	74.24%	74.79%
KNN	71.97%	76.74%	71.97%	73.88%

Berdasarkan Tabel 2 diketahui bahwa algoritma *Naive Bayes* memiliki performa lebih baik dibandingkan KNN pada seluruh parameter evaluasi. *Naive Bayes* memperoleh nilai *accuracy* sebesar 74,24%, *precision* sebesar 79,59%, *recall* sebesar 74,24%, dan *F1-score* sebesar 74,79%. Sedangkan algoritma KNN memperoleh nilai *accuracy* sebesar 71,97%, *precision* sebesar 76,74%, *recall* sebesar 71,97%, dan *F1-score* sebesar 73,88%.

Tabel 3. Hasil Cross Validation

Model	Cross Validation Score
Naive Bayes	71.54%
KNN	76.86%

Berdasarkan Tabel 3 diketahui bahwa algoritma KNN memperoleh nilai *cross validation* sebesar 76,86%, sedangkan *Naive Bayes* memperoleh nilai sebesar 71,54%. Hasil tersebut menunjukkan bahwa model KNN



memiliki tingkat kestabilan yang cukup baik pada proses validasi silang. Namun secara keseluruhan, algoritma *Naive Bayes* tetap menunjukkan performa klasifikasi yang lebih unggul berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

Pada subbab ini dijelaskan teori-teori pendukung yang menjadi landasan dalam penelitian perbandingan algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes* untuk klasifikasi sentimen isu pengunduran ketua PBNU. Pembahasan teori disusun secara spesifik berdasarkan variabel dan komponen penelitian, meliputi analisis sentimen, algoritma KNN, algoritma *Naive Bayes*, klasifikasi teks, serta teknik evaluasi kinerja model.

Analisis sentimen merupakan teknik yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi dari suatu data teks ke dalam kategori tertentu, seperti positif, negatif, dan netral [9]. Dalam penelitian ini, analisis sentimen digunakan untuk mengetahui persepsi masyarakat terhadap isu pengunduran Ketua PBNU melalui komentar pada platform YouTube. Hasil penelitian menunjukkan bahwa mayoritas komentar pengguna YouTube memiliki kecenderungan sentimen negatif sebesar 52,36%, diikuti sentimen netral sebesar 44,90%, dan sentimen positif sebesar 2,74%. Dominasi sentimen negatif menunjukkan bahwa isu pengunduran Ketua PBNU memunculkan banyak kritik, ketidaksetujuan, dan opini bernada penolakan dari masyarakat pengguna media sosial. Sementara itu, rendahnya jumlah sentimen positif menunjukkan bahwa dukungan terhadap isu tersebut relatif sedikit dibandingkan sentimen negatif dan netral. Temuan ini memperlihatkan bahwa media sosial, khususnya YouTube, dapat menjadi sumber data yang efektif dalam menggambarkan opini publik secara *real-time*. Hal ini sejalan dengan teori analisis sentimen yang menyatakan bahwa komentar pada media sosial dapat digunakan untuk mengukur persepsi masyarakat terhadap suatu isu, tokoh, kebijakan, maupun fenomena sosial tertentu [10]. Selain itu, YouTube memiliki karakteristik sebagai media interaktif yang

memungkinkan pengguna menyampaikan opini secara langsung melalui kolom komentar [11]. Penelitian ini juga memiliki keterkaitan dengan penelitian terdahulu berjudul “Perbandingan Algoritma *Naive Bayes* dan *K-Nearest Neighbor* (KNN) Pada Video YouTube Mengenai *Global Warming*” [12]. Penelitian tersebut menunjukkan bahwa komentar YouTube dapat digunakan sebagai sumber data analisis sentimen untuk mengetahui persepsi masyarakat terhadap suatu isu publik. Kesamaan tersebut memperkuat bahwa platform YouTube merupakan media yang relevan dalam penelitian analisis sentimen berbasis opini publik.

K-Nearest Neighbor (KNN) merupakan algoritma *supervised learning* yang bekerja berdasarkan prinsip kedekatan jarak antar data [13]. Pada penelitian ini, algoritma KNN digunakan untuk mengklasifikasikan komentar YouTube ke dalam kategori sentimen positif, negatif, dan netral berdasarkan kemiripan antar data teks yang telah melalui proses pembobotan TF-IDF. Hasil evaluasi menunjukkan bahwa algoritma KNN memperoleh nilai *accuracy* sebesar 71,97%, *precision* sebesar 76,74%, *recall* sebesar 71,97%, dan *F1-score* sebesar 73,88%. Nilai tersebut menunjukkan bahwa algoritma KNN mampu melakukan klasifikasi sentimen dengan cukup baik, meskipun performanya masih berada di bawah algoritma *Naive Bayes*. Secara teori, KNN bekerja dengan menghitung jarak antar data menggunakan pendekatan *Euclidean Distance*, kemudian menentukan kelas berdasarkan mayoritas tetangga terdekat [13]. Pada penelitian ini, performa KNN dipengaruhi oleh distribusi data yang tidak seimbang, terutama karena jumlah sentimen positif jauh lebih sedikit dibandingkan sentimen negatif dan netral. Ketidakseimbangan data tersebut menyebabkan algoritma KNN mengalami kesulitan dalam menentukan tetangga terdekat secara optimal pada kelas minoritas.

Hasil penelitian ini memiliki kesamaan dengan penelitian “Analisis Perbandingan Algoritma *Naive Bayes* dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio di Google *Play Store*” [14]. Pada



SENABISTEKES

Seminar Nasional Bisnis Teknologi dan Kesehatan

penelitian tersebut, algoritma KNN memperoleh performa yang cukup baik dengan tingkat akurasi sebesar 74,92%. Penelitian tersebut menyatakan bahwa KNN efektif dalam mengenali pola kemiripan antar data teks. Hal ini menunjukkan bahwa algoritma KNN tetap memiliki kemampuan yang baik dalam klasifikasi teks, terutama pada dataset dengan pola data yang stabil dan seimbang. Selain itu, hasil *cross validation* pada penelitian ini menunjukkan bahwa algoritma KNN memperoleh nilai sebesar 76,86%, lebih tinggi dibandingkan *Naive Bayes*. Nilai tersebut menunjukkan bahwa model KNN memiliki tingkat kestabilan yang cukup baik ketika diuji menggunakan data yang berbeda.

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan *Teorema Bayes* dalam menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu [15]. Algoritma ini banyak digunakan pada *text mining* dan analisis sentimen karena memiliki kemampuan yang baik dalam mengenali pola kata pada data teks. Berdasarkan hasil penelitian, algoritma *Naive Bayes* memperoleh nilai *accuracy* sebesar 74,24%, *precision* sebesar 79,59%, *recall* sebesar 74,24%, dan *F1-score* sebesar 74,79%. Hasil tersebut menunjukkan bahwa *Naive Bayes* memiliki performa yang lebih baik dibandingkan algoritma KNN dalam proses klasifikasi sentimen komentar YouTube terkait isu pengunduran Ketua PBNU. Keunggulan *Naive Bayes* pada penelitian ini disebabkan karena algoritma tersebut menggunakan pendekatan probabilistik yang mampu menghitung peluang kemunculan kata pada masing-masing kategori sentimen. Pada data teks yang memiliki jumlah fitur sangat banyak, pendekatan probabilistik *Naive Bayes* lebih efektif dibandingkan pendekatan berbasis jarak seperti KNN. Selain itu, *Naive Bayes* juga lebih stabil dalam menangani data teks yang memiliki tingkat ambiguitas tinggi.

Hasil penelitian ini sejalan dengan penelitian terdahulu “Perbandingan Algoritma *Naive Bayes* dan *K-Nearest Neighbor* (KNN) Pada Video YouTube Mengenai *Global Warming*” [12]. Penelitian tersebut menunjukkan bahwa *Naive Bayes* memperoleh

tingkat akurasi sebesar 79,17%, lebih tinggi dibandingkan KNN sebesar 76,06%. Penelitian tersebut menyimpulkan bahwa *Naive Bayes* lebih efektif dalam mengklasifikasikan opini publik pada media sosial.

Selain itu, penelitian “*Comparative Study of SVM, KNN, and Naive Bayes for Sentiment Analysis of Religious Application Reviews*” [16] juga menunjukkan bahwa *Naive Bayes* memperoleh akurasi sebesar 90,27%, lebih tinggi dibandingkan KNN sebesar 82,55% pada dataset bertema keagamaan. Temuan tersebut memperkuat hasil penelitian ini bahwa algoritma *Naive Bayes* memiliki performa yang baik dalam klasifikasi data teks bertema sosial dan keagamaan.

Dengan demikian, berdasarkan hasil evaluasi model dan dukungan penelitian terdahulu, algoritma *Naive Bayes* dinilai lebih efektif dalam melakukan klasifikasi sentimen komentar YouTube terkait isu pengunduran Ketua PBNU.

YouTube merupakan salah satu platform media sosial yang menyediakan ruang interaksi publik melalui fitur komentar [11]. Dalam penelitian ini, YouTube digunakan sebagai sumber data utama karena mampu merepresentasikan opini masyarakat secara langsung terhadap isu pengunduran Ketua PBNU. Komentar-komentar yang terdapat pada video YouTube mencerminkan berbagai bentuk opini, baik berupa dukungan, kritik, maupun tanggapan netral terhadap isu yang sedang dibahas. Oleh karena itu, YouTube menjadi media yang relevan dalam penelitian analisis sentimen karena menyediakan data tekstual dalam jumlah besar yang dapat dianalisis menggunakan teknik *machine learning*. Selain itu, penggunaan YouTube sebagai sumber data juga didukung oleh penelitian terdahulu [17] yang memanfaatkan komentar YouTube untuk mengukur persepsi masyarakat terhadap isu *global warming*. Kesamaan tersebut menunjukkan bahwa platform YouTube memiliki potensi besar sebagai media pengumpulan data dalam penelitian berbasis opini publik.

Penelitian ini diimplementasikan menggunakan bahasa pemrograman *Python* karena memiliki banyak pustaka yang



mendukung pengolahan data dan *machine learning*. *Python* dipilih karena sintaksnya sederhana, fleksibel, serta memiliki pustaka seperti *Pandas*, *NLTK*, *Sastrawi*, dan *Scikit-learn* yang mempermudah proses *preprocessing* teks, pembobotan TF-IDF, hingga klasifikasi sentimen [9]. Pada tahap *preprocessing*, pustaka *NLTK* dan *Sastrawi* digunakan untuk melakukan *tokenisasi*, *stopword removal*, dan *stemming* pada data komentar berbahasa Indonesia. Selanjutnya, pustaka *Scikit-learn* digunakan dalam proses implementasi algoritma KNN dan *Naive Bayes* serta evaluasi model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Selain *Python*, penelitian ini juga memanfaatkan *framework Streamlit* sebagai media implementasi sistem berbasis web. *Streamlit* digunakan untuk membangun antarmuka sistem analisis sentimen yang interaktif sehingga pengguna dapat melakukan *upload* data, *preprocessing*, klasifikasi, hingga visualisasi hasil secara langsung melalui web [7].

Penggunaan *Streamlit* memberikan kemudahan dalam proses implementasi sistem karena mampu menyajikan model *machine learning* dalam bentuk aplikasi yang praktis dan mudah digunakan. Dengan adanya sistem berbasis web tersebut, proses analisis sentimen menjadi lebih informatif, interaktif, dan efisien dalam membantu pengguna memahami hasil klasifikasi sentimen secara *real-time*.

Kesimpulan

Kesimpulan penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) dan *Naive Bayes* berhasil diimplementasikan untuk melakukan analisis sentimen terhadap komentar YouTube terkait isu pengunduran Ketua PBNU. Proses analisis dilakukan melalui tahapan *web scraping*, *preprocessing* teks, pembobotan TF-IDF, hingga klasifikasi sentimen menggunakan kedua algoritma tersebut. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* memiliki performa yang lebih baik dibandingkan KNN dalam mengklasifikasikan sentimen komentar YouTube berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Keunggulan *Naive Bayes* menunjukkan bahwa pendekatan

probabilistik lebih efektif dalam menangani data teks dengan jumlah fitur yang besar dan tingkat ambiguitas yang tinggi. Penelitian ini juga menunjukkan bahwa mayoritas komentar masyarakat pada isu pengunduran ketua PBNU cenderung bernada negatif, sehingga media sosial seperti YouTube dapat digunakan sebagai sumber data yang efektif untuk menggambarkan persepsi publik secara *real-time*. Selain itu, implementasi sistem menggunakan *Python* dan *framework Streamlit* mampu mendukung proses analisis sentimen secara lebih interaktif, praktis, dan mudah digunakan, sehingga penelitian ini dapat memberikan kontribusi dalam pengembangan penerapan *machine learning* pada analisis opini publik berbasis media sosial, khususnya pada isu organisasi dan keagamaan di Indonesia.

Daftar Pustaka

- [1] A. Faujiyanto dan B. Ibrahim, "Dengan Masyarakat Daerah Transmigrasi di Kabupaten Kuantan Singingi (1981-2019)," *Fikri: Jurnal Kajian Agama, Sosial dan Budaya*, 2021.
- [2] L. A. Triono, "Ketua Umum PBNU, Wewenang dan Tugasnya," *NU Online*, 2021.
- [3] *NU Online*, "Susunan Lengkap Kepengurusan PBNU 2022-2027," *NU Online*, 2022.
- [4] *NU Online*, "Ketum PBNU Harap Konsolidasi Global bagi Kemanusiaan Dapat Terwujud," *NU Online*, 2025.
- [5] P. Studi, T. Informatika, F. I. Komputer, U. M. Riau, D. Muallafah, R. Gunawan, dan D. M. Suratno, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM," vol. 13, no. 1, pp. 72–80, 2023.
- [6] B. A. B. Iii dan A. M. Penelitian, "Metodologi Penelitian," 2010.
- [7] T. K. Amarya, R. Firliana, dan A. Ristyawan, "Aplikasi Deteksi Dini



SENABISTEKES
Seminar Nasional Bisnis Teknologi dan Kesehatan

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan

Universitas Muhammadiyah Muara Bungo

Volume 3, No 1, 2026

e-ISSN 3089-3763

- Penyakit Stroke Menggunakan Streamlit,” vol. 9, pp. 453–462, 2025.
- [8] F. Zafira, B. Irawan, dan A. Bahtiar, “Penerapan Data Mining untuk Estimasi Stok Barang dengan Metode K-Means Clustering,” vol. 8, no. 1, pp. 156–161, 2024.
- [9] A. P. Candra, “Analisis Data Menggunakan Python: Memperkenalkan Pandas dan NumPy,” vol. 3, no. 1, pp. 11–16, 2025.
- [10] P. A. Permatasari dan L. Jasa, “Survei Tentang Analisis Sentimen Pada Media Sosial,” vol. 20, no. 2, 2021.
- [11] A. Vira dan E. Reynata, “Penerapan YouTube sebagai Media Baru dalam Komunikasi,” 2003.
- [12] A. Himawan dan C. A. Sugianto, “Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbor (KNN) pada Video YouTube Mengenai Global Warming,” vol. 2, pp. 98–104, 2024.
- [13] I. H. Kusuma dan N. Cahyono, “Analisis Sentimen Masyarakat terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” vol. 8, no. 3, pp. 302–307, 2023.
- [14] D. Pratmanto, A. Widayanto, Y. M. Kristania, dan R. Wijianto, “Analisis Perbandingan Algoritma Naive Bayes dan KNN untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio di Google Play Store,” vol. 4, no. 2, pp. 119–124, 2024.
- [15] D. Darwis, N. Siskawati, dan Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” vol. 15, no. 1, pp. 131–145, t.t.
- [16] H. Aprilianti, H. Mustofa, K. Umam, dan M. R. Handayani, “Comparative Study of SVM, KNN, and Naive Bayes for Sentiment Analysis of Religious Application Reviews,” vol. 9, no. 3, pp. 920–927, 2025.
- [17] A. Safira, F. N. Hasan, K. J. Timur, dan N. B. Classifier, “Paylater Menggunakan Metode Naive Bayes,” vol. 5, no. 1, pp. 59–70, 1978.