

Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes

Habibi Al Rasyid Harpizon^{1*}, Rahmad Kurniawan², Iwan Iskandar³,

Roni Salambue⁴, Elvia Budianita⁵, Fadhilah Syafria⁶

^{1,2,3,5,6} Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. HR. Soebrantas KM. 15, Tuah Karya, Kec. Tampan, Kota Pekanbaru, Riau 28293

^{2,4} Universitas Riau

Jl. HR. Soebrantas, KM. 12,5, Simpang Baru, Kec. Tampan, Kota Pekanbaru, Riau 28293

Corresponding author's e-mail: 11751101289@students.uin-suska.ac.id

Abstrak — Sosial media tidak hanya digunakan oleh masyarakat Indonesia untuk hiburan, tetapi juga sebagai media edukasi. *Youtube* merupakan salah satu media sosial yang terkenal di Indonesia dengan 93,8% pengguna. *Youtube* juga dimanfaatkan sebagai media Dakwah seperti yang dilakukan oleh Ustadz Abdul Somad. Ustadz Abdul Somad merupakan ulama yang berpengaruh di Indonesia. Beliau sering mengunggah video yang membahas berbagai jenis persoalan agama khususnya pada bidang hadist dan fiqh. Pengguna *Youtube* dapat memberikan *feedback* berupa *like*, *dislike* dan komentar terhadap video yang ditayangkan. *Feedback* diperlukan oleh pembuat konten di *Youtube* untuk melihat tanggapan pengguna. Analisa secara manual sulit dilakukan karena jumlah data yang besar. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Ustadz Abdul Somad melalui komentar *youtube* menggunakan algoritma *Naïve Bayes*. Penelitian ini menggunakan 1000 komentar dari 10 video yang ada di *Youtube* mengenai Ustadz Abdul Somad. *Naïve Bayes* merupakan algoritma yang sederhana, namun memiliki akurasi yang tinggi dan dapat digunakan pada data yang sedikit. Berdasarkan hasil penelitian, didapatkan sebanyak 67% berkomentar positif, 27% berkomentar netral dan 6% berkomentar negatif. Berdasarkan pengujian didapatkan akurasi sebesar 87%, presisi 91% dan *recall* 97%. Berdasarkan pengujian tersebut dapat disimpulkan bahwa penelitian ini dapat digunakan untuk hasil sentimen dengan cepat di *Youtube*.

Kata kunci: *Analisis Sentimen, Naïve Bayes, Ustadz Abdul Somad, Youtube*

Abstract — Indonesian people have been used *Youtube* for entertainment and as an education. As Indonesia's most popular social media, *Youtube* has 93.8% users. *YouTube* is also used as a medium of *Da'wah*, like Ustadz Abdul Somad. Ustadz Abdul Somad is an influential Preacher in Indonesia. He often uploads videos that lecture various types of religious issues, especially in the fields of *hadith* and *fiqh*. *YouTube* users can provide *feedback* in the form of *likes*, *dislikes*, and *comments* on videos that are shown. Creators need *feedback* on *YouTube* to see user *feedback*. Manual analysis is complicated because of the large amount of data. Therefore, this study aimed to analyze public sentiment towards Ustadz Abdul Somad through *YouTube* comments using the *Naïve Bayes* algorithm. This study obtained 1000 comments from 10 videos about Ustadz Abdul Somad. *Naïve Bayes* is a simple algorithm with high accuracy and can be used on small data. Based on the results, it was found that 67% commented positively, 27% commented neutrally, and 6% commented negatively. Based on the experimental testing, the accuracy is 87%, precision is 91%, and recall is 97%. Based on these tests, it can be concluded that this research can be used for quick sentiment results on *YouTube*.

Keywords: *Sentiment Analysis, Naïve Bayes, Ustadz Abdul Somad, Youtube*

1. Pendahuluan

Youtube adalah platform media sosial yang menampilkan video dalam durasi tertentu yang dibuat oleh para pembuat video (*creator youtube*) dengan genre yang berbeda-beda. Tingginya antusiasme masyarakat dalam menggunakan *youtube* membuat media sosial ini semakin dimanfaatkan oleh berbagai kalangan. *Youtube* menempati urutan media sosial terpopuler yang ada di Indonesia [1]. Memiliki 170 juta pengguna atau 93,8 persen dari keseluruhan populasi 181,9 juta pengguna internet berusia 16-64 tahun berdasarkan laman berita solopos.com. Disisi lain, situs *youtube* juga dapat memberikan umpan balik (*feedback*) berupa *like*, *dislike*, dan komentar terhadap video yang ditayangkan di media *youtube* [2]. Sehingga penonton dapat berinteraksi serta memberikan pendapat mereka mengenai video tersebut. Tentunya berdasarkan tanggapan penonton, komentar dapat berisi komentar yang dapat membangun serta memotivasi sang kreator atau mungkin dapat berbentuk komentar negatif berupa SARA (Suku, Ras, Agama, dan Antar golongan), terhadap kreator *youtube* sendiri.

Sebagai salah satu platform streaming video terpopuler di dunia, *youtube* menampilkan berbagai video *entertainment* dan juga edukasi. Selain itu *youtube* juga dimanfaatkan sebagai media penyebaran dakwah serta informasi oleh pemuka agama khususnya yang ada di Indonesia pada saat ini, salah satunya ialah Ustadz Abdul Somad. Ustadz Abdul Somad adalah salah satu ulama atau pendakwah yang terkenal di Indonesia. Ustadz

Abdul somad sering membahas berbagai macam persoalan agama, khususnya pada bidang kajian ilmu hadis dan juga ilmu fiqih. Selain itu, Ustadz Abdul Somad juga banyak membahas mengenai nasionalisme dan berbagai masalah terkini yang sedang menjadi pembahasan hangat dikalangan masyarakat [3]. Memiliki jumlah pengikut sebanyak 2 juta *subscriber* per tanggal 3 Mei 2021 di *youtube*, dengan nama channel Ustadz Abdul Somad Official menjadikan channel ini menjadi channel pribadi pendakwah dengan subscriber terbanyak di Indonesia. Menurut news.detik.com, Ustadz Abdul Somad merupakan ulama paling berpengaruh di Indonesia berdasarkan LSI (Lingkaran Survey Indonesia) dan berhasil menduduki urutan pertama dengan perolehan 30,2 persen berdasarkan kategori “didengar imbauannya” dibandingkan dengan 4 ulama lainnya pada tahun 2018. Tentunya dengan tingkat antusiasme yang tinggi, membuat *channel* ini mendapatkan begitu banyak respon dari masyarakat. Hal ini akan berguna bagi Ustadz Abdul Somad untuk mendapatkan informasi yang ada berupa komentar untuk mendapatkan *feedback* dari masyarakat secara langsung tentang materi yang telah disampaikan. Hal ini tentunya akan menyulitkan jika harus menganalisis setiap komentar secara manual, karena komentar berisikan ratusan hingga ribuan tanggapan dari setiap penonton di video tersebut [1].

Salah satu cara untuk melihat respon komentar masyarakat di dalam media *youtube* yaitu dengan analisis sentimen. Analisis sentimen merupakan salah satu cabang dari *Text mining*, dimana pada proses ini dapat menganalisa dokumen atau sumber teks untuk mengelompokkan suatu opini atau tulisan kedalam opini positif, netral maupun negatif secara otomatis [4]. Analisis sentimen sering dijumpai pada media sosial twitter dan Instagram, tetapi *youtube* lebih memiliki peluang yang besar dalam pemanfaatan analisis sentimen. Penggunaan yang terus meningkat dan jumlah *subscriber* yang besar, membuat *youtube* akan lebih berpotensi dibandingkan dengan media sosial lainnya. Minimnya penggunaan bahasa Indonesia dalam analisis sentimen di *youtube*, semakin memperbesar peluang pemanfaatan pada media sosial ini.

Pengolahan data dalam analisis sentimen terdapat dua jenis, yaitu *lexicon-based* dan *machine learning*. Pada *lexicon-based* menggunakan kamus serta perbandingan kata. Sedangkan metode *machine learning* merupakan cara yang lebih modern dan memiliki banyak keunggulan, diantaranya ketepatan yang tinggi dan banyak pilihan pada algoritma yang digunakan. Penelitian analisis sentimen Ustadz Abdul Somad yang dilakukan oleh [5] menghasilkan akurasi sebesar 80,57% dengan menggunakan algoritma SVM. Penelitian yang dilakukan oleh [6] menghasilkan akurasi sebesar 97.33% menggunakan algoritma SVM yang di kombinasikan dengan TF-IDF. Hanya saja algoritma ini terlalu kompleks dan pemrosesan data yang digunakan cenderung membutuhkan waktu yang lebih lama.

Terkait dengan judul penelitian yang akan diusulkan mengenai algoritma *naïve bayes*, terdapat penelitian terkait yaitu yang dilakukan oleh [7], penelitian menghasilkan nilai akurasi sebesar 70% pada algoritma *naïve bayes* sedangkan pada algoritma KNN menghasilkan nilai akurasi sebesar 40%. Hal ini menunjukkan bahwa algoritma *naïve bayes* lebih unggul dibandingkan algoritma KNN dalam melakukan klasifikasi terhadap artikel berbahasa Indonesia. Penelitian lainnya yang dilakukan oleh [8] menghasilkan akurasi sebesar 94% untuk algoritma *naïve bayes* dibandingkan algoritma SVM dengan nilai akurasi sebesar 75.50%. Hal ini menunjukkan bahwa algoritma *naïve bayes* lebih unggul dibandingkan algoritma SVM dalam melakukan klasifikasi di komentar *twitter*.

Berdasarkan latar belakang yang telah dibuat, maka pada penelitian ini akan melakukan analisis sentimen menggunakan algoritma *naïve bayes* dalam klasifikasi komentar *youtube* tentang Ustadz Abdul Somad. Sehingga hasil dari penelitian ini diharapkan dapat melihat seberapa besar akurasi yang dihasilkan oleh algoritma *naïve bayes* dalam mengklasifikasi sentimen pada komentar *youtube* di video Ustadz Abdul Somad ini. Pengolahan data pada penelitian ini menggunakan bahasa pemrograman *python* di *google collab*.

Oleh karena itu, analisis sentimen mengenai Ustadz Abdul Somad menggunakan komentar *youtube* diharapkan dapat memberikan *feedback* serta tanggapan masyarakat mengenai ceramah beliau. Sehingga Ustad Abdul Somad dapat mengoptimalkan kembali ceramah beliau untuk video selanjutnya.

2. Tinjauan Pustaka

2.1. Text Mining

Menurut [9], *text mining* yaitu proses menambang data yang berupa teks, dimana sumber yang dihasilkan biasanya didapatkan dari dokumen dengan tujuan untuk mencari kata-kata yang dapat mewakili isi dari suatu dokumen sehingga dapat dilakukannya. Analisa kethubungan antar dokumen. *Text mining* memiliki tujuan untuk mengekstrak informasi yang berguna dari sumber data. Sehingga sumber data yang digunakan pada *text mining* adalah sekumpulan dokumen atau teks yang memiliki format yang tidak terstruktur melalui identifikasi dan eksplorasi pola yang menarik. *Text mining* juga memiliki tugas khusus yaitu diantaranya pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

Adapun tahapan *Text mining* :

1. Teks, proses pengumpulan data berupa teks (*crawling*) yang nantinya akan digunakan didalam penelitian.

2. Prapemrosesan Teks, prapemrosesan teks yaitu proses pembersihan data yang akan digunakan pada penelitian. Sehingga data yang akan digunakan sudah dalam keadaan bersih dari emoji dan atribut yang tidak diperlukan. Prapemrosesan teks terdapat 5 tahapan, yaitu: *Cleaning, Case Folding, Tokenizing, Stopward Removal dan Stemming*
3. *Text Transformation*, proses melihat teks dokumen yang diwakili oleh kata-kata yang dikandungnya. Pada tahapan ini dilakukannya proses coding pada data yang dipilih.
4. *Feature Selection*, proses penambahan/pemilihan fitur yang diinginkan.
5. *Data Mining*, proses mencari pola atau informasi menarik didalam data yang terpilih dengan menggunakan teknik atau metode tertentu.
6. *Interpretation / Evaluation*, tahapan ini mencakup pada pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Pada tahapan ini juga dilakukan evaluasi terhadap proses yang dijalankan.

2.2. Analisis Sentimen

Sentiment analysis (analisis sentimen) atau yang sering disebut dengan opinion mining (penambangann opini) adalah studi komputasi untuk mengenali serta mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektifitas, penilaian, atau pandangan yang terdapat dalam suatu teks [10].

2.3. Count Vectorizer

Count vectorizer adalah cara untuk mengubah serangkaian string tertentu menjadi representasi frekuensi. Selain itu proses ekstraksi menggunakan *count vectorizer* digunakan untuk menghitung frekuensi kata didalam kalimat atau dokumen yang kemudian akan direpresentasikan ke dalam bentuk *vector* [11].

2.4. Youtube

Tahun 2005 merupakan awal terbentuknya dan lahirnya situs streaming video yaitu *youtube.com*. Perusahaan yang berawal dari 3 orang karyawan ini terinspirasi dari nama nama sebuah kedai pizza dan restoran Jepang di San Maceo, California. Kesuksesan situs ini mulai meningkat sejak setahun dari kelahirannya. Tepatnya pada tahun 2006 situs ini mendapatkan unggahan video baru sebanyak 65.00 dan mencapai 100.000 unggahan hingga Juli 2006. Sebelumnya pada bulan Juni, *youtube* sudah memasuki bidang Kerjasama pemasaran dan periklanan dengan NBC. Hal ini semakin membaik ketika Google membeli saham *youtube* dengan nilai USD 1.65 Juta pada bulan Oktobernya. Hal inilah yang membuat perkembangan serta penggunaan *youtube* semakin pesat di seluruh dunia. Hal ini dibuktikan dengan penghargaan yang didapatkan oleh *youtube* melalui majalah PC world, dan mendapatkan julukan Sembilan dari sepuluh produk terbaik di tahun 2006 [12].

2.5. Ustadz Abdul Somad

Ustadz Abdul Somad lahir di Silo Lama, Kabupaten Asahan, Sumatera Utara pada tanggal 18 Mei 1977. Beliau adalah seorang pendakwah dan ulama di Indonesia yang sering mengulas serta membahas berbagai macam persoalan agama, khususnya ilmu hadist dan ilmu fiqih. Disini lain, ia juga membahas serta mengulas mengenai bagaimana meningkatkan nasionalisme dan berbabagai masalah terbaru yang sedang menjadi pembahasan hangat di lingkungan masyarakat. Ustadz Abdul Somad mulai dikenal oleh publik ketika beliau mampu memberikan penjelasan dengan ilmunya secara lugas dan tepat dalam menyampaikan dakwah yang disiarkan melalui saluran *youtube* nya dan disebarluaskan lagi melalui media sosial *Instagram* dan *facebook* miliknya. Pada saat ini Ustadz Abdul Somad memiliki pengikut (*followers*) di *Instagram* mencapai 5,3 juta dan *follower* di *youtube* mencapai 2.05 juta *subscriber*[3].

2.6. Analisis Sentimen Ustadz Abdul Somad

Perkembangan kepopuleran Ustadz Abdul Somad semakin meningkat seiring berjalannya waktu. Salah satu ustadz yang paham akan materi agama yang ia disampaikan ditambah dengan gaya ceramah yang menarik semakin membuat masyarakat tertarik untuk mendengarkan ceramah beliau. Ustadz Abdul Somad semakin aktif dalam menggunggah ceramahnya di *youtube* maupun media sosial lainnya, membuat antusias masyarakat juga semakin meningkat. Tentunya dengan peningkatan ini akan menimbulkan pro dan kontra dari para penonton yang membuat komentar pada video tersebut.

2.7. Algoritma Naïve Bayes

Naïve bayes classifier merupakan metode pengklasifikasian statistik yang didasarkan pada teorema bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. *Naïve bayes classifier* atau yang sering disebut dengan *Bayesian Classification* terbukti memiliki akurasi dan kecepatan yang tinggi [13]. Alasan lain penggunaan metode *naïve bayes* karena proses pengklasifikasian dengan metode ini memiliki tingkat efektifitas dan mempunyai tingkat akurasi yang tinggi [14]. Algoritma *naïve bayes* menggunakan algoritma yang sederhana, tetapi juga dapat menghasilkan tingkat akurasi yang tinggi [15]. *Naïve bayes* juga memiliki kelebihan diantaranya [16].

Kelebihan *naïve bayes classifier* :

- a) Mudah diimplementasi.
- b) Memberikan hasil yang baik untuk banyak kasus.

2.8. Confusion Matrix

Confusion Matrix bertujuan untuk menganalisa kinerja algoritma dalam melakukan klasifikasi untuk mendeteksi kerusakan [17]. Selain itu *confusion matrix* merupakan sebuah metode yang biasa digunakan untuk perhitungan akurasi, *recall*, *precision*, dan *error* [18]. Pada evaluasi *recall* dilakukan tahap identifikasi aplikasi untuk dilakukan pencarian seluruh item yang signifikan dengan dokumen. *Recall* dapat diartikan sebagai persentase seluruh dokumen yang signifikan dengan *query*. *Precision* dilakukan dengan mengidentifikasi aplikasi untuk mencari peringkat yang sangat signifikan. *Precision* juga dapat diartikan dengan persentase dokumen yang di *retrieve* yang sangat signifikan dengan *query*. Selanjutnya *accuracy* atau perhitungan akurasi dapat diperoleh dari perbandingan kasus yang sudah ditetapkan benar dengan jumlah semua kasus yang ada, sedangkan *error rate* adalah perbandingan kasus yang sudah ditetapkan salah dengan jumlah semua kasus yang ada.

		PREDIKSI		
		Positif	Negatif	Netral
AKTUAL	Positif	TPos	FPosNeg	FPosNet
	Negatif	FNegPos	TNeg	FNegNet
	Netral	FNetPos	FNetNeg	TNet

Tabel 1. *Confusion Matrix*

Keterangan:

- a) TPos : Nilai dari data yan bernilai positif dan juga hasil dari klasifikasi yang bernilai positif.
- b) FPosNeg : Nilai dari data yang bernilai positif dan hasil dari klasifikasi yang bernilai negatif.
- c) FNegPos : Nilai dari data yang bernilai negatif dan hasil dari klasifikasi yang bernilai positif.
- d) TNeg : Nilai dari data yang bernilai negatif dan hasil dari klasifikasi yang bernilai negative

3. Metode Penelitian

Metodologi penelitian adalah tahapan atau proses yang disusun secara sistematis dalam melakukan suatu penelitian yang berguna untuk mencapai target yang diharapkan. Tahapan penelitian yang akan dilakukan pada penelitian ini yaitu sebagai berikut:



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

Pengumpulan data yang dilakukan pada penelitian ini yaitu dengan dilakukannya pengumpulan data dengan dua tahapan, yaitu tahapan *crawling* dan *labelling*. Tahapan *crawling* yaitu proses dilakukannya pengambilan data (komentar) dari platform *youtube*. Proses pengambilan data menggunakan website <https://exportcomments.com/> dan akan didapatkan komentar sebanyak 1000 komentar yang sudah disatukan dalam bentuk excel. Setelah mendapatkan jumlah dataset yang diinginkan yaitu sebanyak 1000 data maka akan dilanjutkan ke tahapan *labelling*. *labelling* (pelabelan) yaitu proses memberikan label data yang sudah didapatkan sebelumnya dengan memberikan label apakah data tersebut bernilai positif, netral dan negatif berdasarkan makna dari satu kalimat dataset tersebut. Proses pelabelan data menggunakan metode *crowd sourcing*. *Crowd sourcing* yaitu sebuah metode yang dilakukan untuk mengumpulkan beberapa orang untuk

menganalisa atau melakukan sesuatu demi mencapai suatu tujuan. Sehingga setelah dikumpulan dataset yang sudah di beri label maka akan dilakukannya proses voting akhir pada label tersebut.

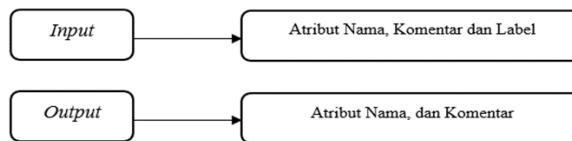
3.2. Prapemrosesan Teks

Prapemrosesan teks atau yang lebih dikenal dengan nama *text preprocessing* yaitu proses membersihkan data sebelum diolah nantinya. Pada tahapan ini terdapat 5 proses diantaranya:



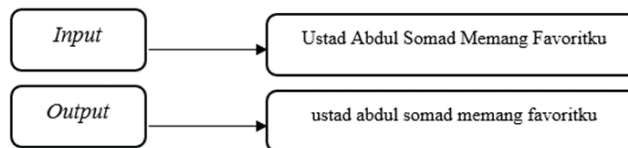
Gambar 2. Tahapan Prapemrosesan Teks/*Preprocessing Teks*

- a) *Cleaning*, pada tahapan ini bertujuan untuk membersihkan data dari data/komentar yang ganda, dan atribut tidak diperlukan yang terdapat pada komentar



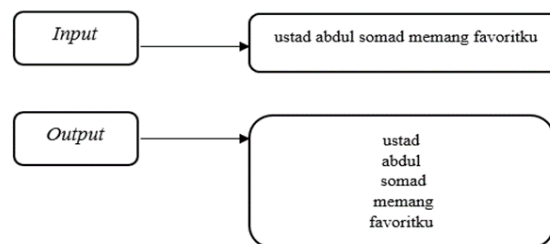
Gambar 3. Tahapan *Cleaning*

- b) *Case Folding*, pada tahapan ini bertujuan untuk menyeragamkan seluruh teks kedalam huruf kecil (*lowercase*).



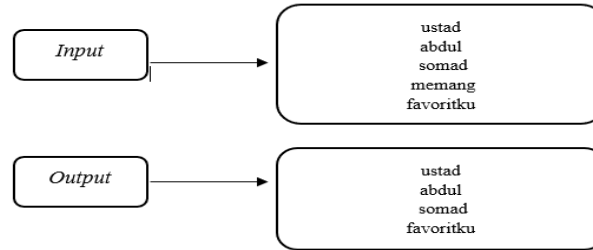
Gambar 4. Tahapan *Case Folding*

- c) *Tokenizing*, yaitu proses dilakukannya pemecahan kata pada kalimat. Pada tahapan ini akan dilakukannya pemecahan dari kalimat di komentar menjadi kata per kata.



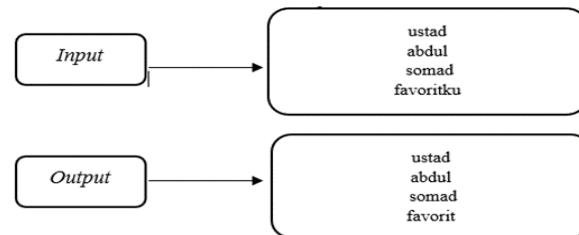
Gambar 5. Tahapan *Tokenizing*

- d) *Stopword Removal*, pada tahapan ini dilakukannya penghilangan kata yang termasuk kedalam kategori *stopward*. Sedangkan *stopward* yaitu kata yang sering muncul namun dianggap tidak memiliki arti yang berarti. Seperti kata *hubung* yang, ini, dengan, dan kata *hubung* lainnya.



Gambar 6. Tahapan *Stopword Removal*

- e) *Stemming*, pada proses ini bertujuan untuk menemukan kata dasar, dengan menghilangkan semua imbuhan yang menyatu pada kata.



Gambar 7. Tahapan *Stemming*

3.3. Klasifikasi

Pada tahapan ini akan dilakukannya proses klasifikasi dengan algoritma yang digunakan. Adapun algoritma yang digunakan adalah algoritma *naïve bayes*. Pada tahap ini data yang sudah melewati proses *crawling* dan *labeling* akan di proses menggunakan algoritma yang dipilih yaitu *naïve bayes*. Dalam tahapan ini mesin akan mengenali pola yang kemudian akan diklasifikasi ke dalam tiga kelas yaitu kelas positif, netral dan kelas negatif.

Naïve bayes memiliki bentuk umum yaitu:

$$P(C|X) = \frac{P(X|C) P(c)}{P(x)} \quad (1)$$

Keterangan:

- X : Data dengan class yang belum diketahui
- C : Hipotesis data yang merupakan suatu class spesifik
- $P(C|X)$: Probabilitas hipotesis C berdasarkan kondisi (*posterior probability*)
- $P(c)$: Probabilitas hipotesis C (*prior probability*)
- $P(x|c)$: Probabilitas berdasarkan kondisi pada hipotesis
- $P(x)$: Probabilitas c

Rumus *naïve bayes* diatas dapat diartikan dengan peluang masuknya sebuah data sampel dari karakteristik tertentu pada kelas C (Posterior). C (Posterior) merupakan peluang munculnya kelas C (sebelum data sampel masuk, atau disebut prior), kemudian dikali dengan peluang munculnya karakteristik sampel dari kelas C (sering disebut likelihood), dan dibagi dengan peluang munculnya karakteristik data sampel secara global (evidence).

$$P(V1|C = c) = \frac{\text{CountTerms}(v1, \text{docsv}(c))}{\text{AllTerms}(\text{docs}(c))} \quad (2)$$

Simbol V1 didalam persamaan diatas melambangkan satu kata tertentu didalam komentar pada *youtube*, sedangkan $\text{CountTerms}(v1, \text{docsv}(c))$ menunjukkan jumlah kemunculan suatu kata berlabel c (“positif”, “netral” atau “negatif”). $\text{AllTerms}(\text{docs}(c))$ menunjukkan jumlah semua kata yang berlabel c yang ada pada dataset.

Tahapan Analisa perhitungan *naïve bayes* pada penelitian ini dilakukan dengan metode klasifikasi yang telah diperoleh dari data latih/*Vocabulary* hasil observasi. Klasifikasi *naïve bayes* perlu dilakukan untuk mengkonversi teks-teks menjadi angka sehingga bisa dibaca oleh algoritma *naïve bayes*. Cara mengubah teks pada komentar ke dalam bentuk angka, yaitu menggunakan *Count Vectorizer*. *Count vectorizer* yaitu cara untuk

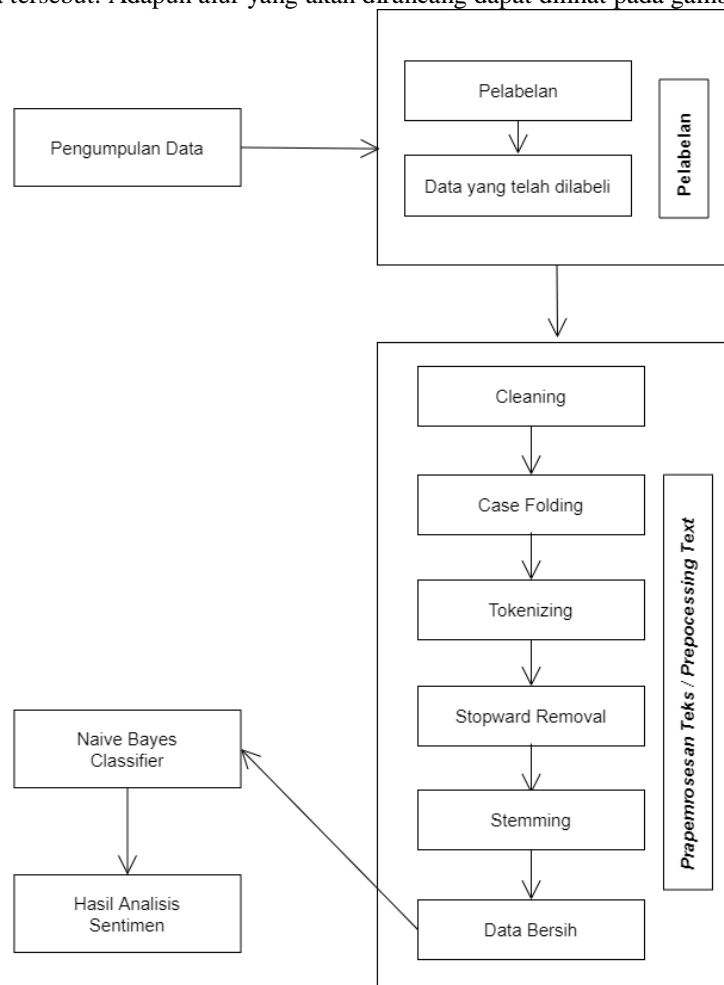
mengubah serangkaian string tertentu menjadi representasi frekuensi. Adapun konversi teks menjadi angka yang dilakukan pada penelitian ini yaitu menggunakan *count vectorizer*.

Tahapan dari perhitungan *naive bayes* adalah :

- Tahapan pertama dalam proses *naive bayes classifier* adalah menghitung probabilitas masing-masing kelas dari keseluruhan data training.
- Proses *testing*. Proses ini berguna untuk mengetahui keakuratan model yang dibangun pada proses training, umumnya digunakan data yang disebut test set untuk memprediksi label. Metode *naive bayes classifier* terdiri dari dua tahapan dalam proses klasifikasi teks, yaitu tahapan pelatihan dan tahapan klasifikasi. Tahap pelatihan dilakukan proses analisis terhadap sample dokumen berupa pemilihan *vocabulary* yaitu kata yang dimungkinkan muncul dalam koleksi dokumen sample yang menjadi representasi dokumen. Langkah selanjutnya adalah menentukan probabilitas bagi tiap kategori berdasarkan sampel dokumen.

3.4. Perancangan

Pada tahapan perancangan ini dilakukannya perancangan yang akan dibangun berdasarkan proses analisa sebelumnya. Adapun perancangan ini dilakukan dengan merancang serta menghasilkan output yang baik dan hasil yang akurat. Sehingga algoritma nantinya dapat dengan mudah dimengerti oleh para pembaca dalam menerapkan algoritma tersebut. Adapun alur yang akan dirancang dapat dilihat pada gambar dibawah ini.



Gambar 8. Alur Analisis Sentimen

3.5. Implementasi dan Pengujian

Implementasi pada penelitian ini menggunakan bahasa pemrograman *Python* yang dijalankan pada media *Google Collab*. Pada proses ini akan dilakukannya proses menguji kinerja, menghitung nilai serta mengidentifikasi klasifikasi algoritma *naive bayes* terhadap mesin yang telah dibangun menggunakan *confusion matrix* dengan menghitung *accuracy, precision, recall* dan *F1-Score* seperti pada persamaan berikut.

$$Accuracy = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

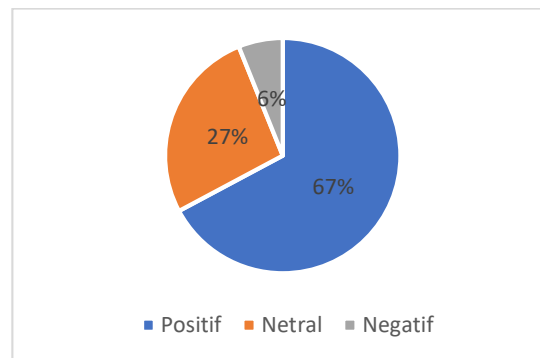
$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F - Measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

4. Hasil dan Pembahasan

Pembersihan data dilakukan pada tahapan *preprocessing text*. Didalam data didapatkan komentar yang ditulis oleh para penonton di video *youtube*. Pada komentar ini terdapat tiga pembagian komentar yaitu komentar positif, netral dan juga komentar negatif. Data terdiri dari 1000 komentar dari 10 video yang ada di youtube mengenai Ustadz Abdul Somad. Pada komentar ini terdapat (67%) atau 672 komentar positif, (27%) atau 267 komentar netral dan (6%) atau 61 komentar negatif seperti yang terlihat pada gambar 9.



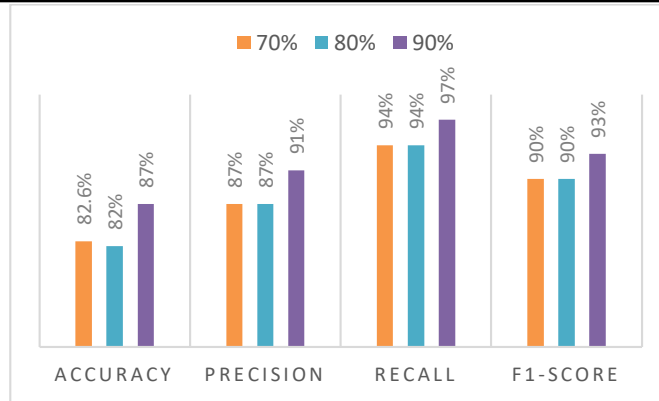
Gambar 9. Perbandingan Sentimen

Pada analisis sentimen ini dilakukan tiga model eksperimen. Eksperimen pertama yaitu menggunakan alokasi data training sebesar 70%. Eksperimen kedua menggunakan alokasi data training sebesar 80% dan yang terakhir menggunakan alokasi data training sebesar 90%. Seperti pada tabel berikut.

No	Alokasi Data (Training)	Accuracy	Precision	Recall	F-1 Score
1	70%	82.6%	87%	94%	90%
2	80%	82%	87%	94%	90%
3	90%	87%	91%	97%	93%

Tabel 2. Hasil Eksperimen

Hasil eksperimen diatas juga dapat dilihat dalam bentuk grafik pada gambar 10.



Gambar 10. Grafik Hasil Eksperimen

Berdasarkan tiga eksperimen diatas maka eksperimen ketiga menghasilkan nilai *accuracy* paling besar yaitu 87% dibandingkan dengan eksperimen lainnya. Hal ini juga didukung dengan perhitungan untuk mencari nilai akurasi berdasarkan rumus 1:

$$Accuracy = \frac{3 + 15 + 69}{3 + 1 + 0 + 3 + 15 + 7 + 0 + 2 + 69} \times 100\%$$

Maka didapatkanlah nilai *accuracy* sebesar 87%

5. Kesimpulan

5.1. Kesimpulan dan Saran

Berdasarkan hasil penelitian ini maka analisis sentimen mengenai Ustadz Abdul Somad menggunakan komentar youtube mendapatkan nilai akurasi sebesar 87% dengan dominan komentar positif. Penelitian ini juga menghasilkan nilai *Precision* sebesar 91%, *Recall* 97% dan nilai *F-Measure* 93%. Hasil ini membuktikan bahwa algoritma *naïve bayes* baik dalam melakukan analisis sentimen pada komentar *youtube* mengenai Ustadz Abdul Somad.

Berdasarkan penelitian ini diharapkan dilakukannya penelitian lainnya mengenai komentar Ustadz Abdul Somad di *youtube* menggunakan metode lain sehingga diharapkan dapat memberikan nilai akurasi yang lebih optimal. Sehingga dengan adanya analisis sentimen ini dapat memberikan informasi mengenai hasil sentimen masyarakat mengenai video Ustadz Abdul Somad yang dapat digunakan sebagai masukan agar selalu memberikan video yang terbaik bagi masyarakat.

5.2. Ucapan Terimakasih

Terima kasih kepada lembaga riset CIDSCI (*Center of Islamic Data Science and Continuous Improvement*) Universitas Islam Negeri Sultan Syarif Kasim Riau yang telah mendukung dan memberikan fasilitas demi tercapainya hasil dari penelitian ini.

Daftar Pustaka

- [1] R. Kurniawan, F. Lestari, abdul somad Batubar, mohd zakree ahmad Nazri, K. Rajab, and R. Munir, "Indonesian Lexicon-Based Sentiment Analysis of Online Religious Lectures Review," 2021.
- [2] C. Dabas, P. Kaur, and N. Gulati, "Analisis Komentar di Video Youtube menggunakan Hadoop," 2021.
- [3] Puspitsari, "STRATEGI DAKWAH USTADZ ABDUL SOMAD DALAM KLARIFIKASI PENOLAKAN DAKWAH MELALUI MEDIA SOSIAL YOUTUBE SKRIPSI," 2018.
- [4] D. A. Muthia, "ANALISIS SENTIMEN PADA REVIEW RESTORAN DENGAN TEKS BAHASA INDONESIA MENGGUNAKAN," vol. 2, no. 2, pp. 39–45, 2017.
- [5] D. Aridarma, R. Sadikin, B. S. Prakoso, and H. S. Utama, "USTADZ ABDUL SOMAD LECTURE SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE ALGORITHM COMPARISON OF," vol. 1, pp. 111–116, 2018.
- [6] J. Pratama, "Klasifikasi sentimen terhadap ustadz abdul somad di media sosial youtube menggunakan metode support vector machine tugas akhir," 2021.
- [7] R. N. Devita *et al.*, "PERBANDINGAN KINERJA METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI ARTIKEL BERBAHASA INDONESIA PERFORMANCE COMPARISON OF NAIVE BAYES AND K-NEAREST NEIGHBOR," vol. 5, no. 4, pp. 427–434, 2018.
- [8] D. Ajeng and L. Marlinda, "Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis

-
- Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter,” *2018 6th Int. Conf. Cyber IT Serv. Manag.*, no. Citsm, pp. 1–6, 2018.
- [9] T. Setiadi, “PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI,” vol. 2, pp. 73–83, 2014.
- [10] O. N. Rahim, “ANALISIS SENTIMEN UNTUK MENGUKUR POPULARITAS TOKOH PUBLIK BERDASAR DATA PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA DATAMINING DENGAN TEKNIK KLASIFIKASI,” vol. VI, no. 2, 2014.
- [11] M. Priandi, “Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor,” no. September, pp. 311–319, 2021.
- [12] E. Chandra, “YOUTUBE, CITRA MEDIA INFORMASI INTERAKTIF ATAU MEDIA PENYAMPAIAN ASPIRASI PRIBADI,” pp. 406–417, 2010.
- [13] H. Muhamad *et al.*, “OPTIMASI NAÏVE BAYES CLASSIFIER DENGAN MENGGUNAKAN PARTICLE,” vol. 4, no. 3, pp. 180–184, 2017.
- [14] M. S. Adhi, M. Z. Naf’an, and E. Usada, “JURNAL RESTI Pengaruh Semantic Expansion pada Naïve Bayes Classifier untuk Analisis Sentimen Tokoh Masyarakat,” vol. 3, no. 2, pp. 141–147, 2019.
- [15] D. A. Putri, D. A. Kristiyanti, and E. Indrayuni, “Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review,” 2020.
- [16] D. S. Pamungkas, N. A. Setiyanto, and E. Dolphina, “ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAIVE BAYES CLASSIFIER TERHADAP KATA KUNCI ‘KURIKULUM 2013,’” vol. 14, no. 4, pp. 299–314, 2015.
- [17] D. Sinaga and C. Jatmoko, “ANALISIS SENTIMEN UNTUK MENGETAHUI KESAN PLAYER GAME MOBILE LEGENDS MENGGUNAKAN NAÏVE BAYES CLASSIFIER,” vol. 0, pp. 540–547, 2020.
- [18] L. K. Wardhani and D. Octaviano, “Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag # 2019gantipresiden,” vol. 9, no. 1, pp. 103–114, 2020.