ANALISIS SENTIMEN MASYARAKAT TERHADAP DUGAAN KONTROVERSI PONDOK PESANTREN AL-ZAYTUN MENGGUNAKAN NAÏVE BAYES

Derley Akbar Baturaja, Didi Juardi, Agung Susilo Yuda Irawan
Program Studi Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Karawang Jawa Barat
1810631170088@student.unsika.ac.id

ABSTRAK

Pesantren tempat dimana orang memperdalam ilmu agama, akan tetapi terkadang ajaran yang di ajarkan kepada orang kadang di salah artikan atau sering melenceng dari ajaran agama yang umum. Seperti kasus yang sempat viral yaitu Pondok Pesantren Al-Zaytun dengan segala problematika nya tujuan dari penelitian ini adalah untuk menganalisis sentimen yang dirasakan masyarakat Indonesia terhadap kontroversi yang melibatkan Pondok Pesantren Al-Zaytun, menggunakan Instagram sebagai sumber datanya. Data komentar dikumpulkan melalui Instagram API, kemudian dibersihkan dan diproses, menghasilkan 1.010 data kemudian digunakan dalam analisis. Metode yang diterapkan adalah analisis sentimen menggunakan algoritma klasifikasi Naïve Bayes dengan pendekatan *Knowledge Discovery in Database* (KDD). Pendekatan ini meliputi tahap Data Selection, Pre-Processing, Transformation, Data Mining, dan Evaluation.Selain itu, dilakukan juga pembobotan sentimen dengan menggunakan kamus Lexicon dan kamus *Negative Words* untuk mengklasifikasikan data ke dalam kategori positif, negatif, dan netral. Hasil dari penelitian ini menunjukkan bahwa model dengan pembagian data skenario 90:10 memiliki kualitas analisis sentimen yang baik, dengan tingkat akurasi mencapai 80% dan nilai *Area Under the Curve* (AUC) sebesar 0.82. Beberapa kata yang umum ditemukan dataset meliputi "pondok", "pesantren", "Al-Zaytun", dan "sesat". Penelitian ini memberikan wawasan mengenai pandangan masyarakat, diharapkan dapat digunakan sebagai referensi bagi pihak berwenang dalam memahami persepsi publik terkait kasus tersebut.

Kata Kunci: *Analisis sentimen, Naïve Bayes, Al-Zaytun, Instagram, Skor sentimen*

1. PENDAHULUAN

Analisis sentimen adalah sebuah proses otomatis yang melibatkan pemahaman, ekstraksi, dan pemrosesan data teks dengan tujuan menghasilkan informasi terkait sentimen yang tersembunyi dalam suatu pernyataan. Keseluruhan tujuan dari analisis sentimen adalah untuk mengenali pandangan atau sikap yang dinyatakan oleh pembicara atau penulis terkait suatu topik atau kecenderungan polaritas secara keseluruhan dalam teks tersebut. Sikap ini bisa berupa penilaian, evaluasi, ungkapan emosi dari penulis saat menulis, atau dampak emosional dari tulisan penulis pada audiens atau pembaca.

Di Indonesia, bidang informatika berkembang dengan pesat sejalan dengan kebutuhan akan teknologi informasi yang canggih dan inovatif. Kesadaran akan pentingnya penerapan solusi informatika semakin meningkat di kalangan masyarakat, industri, dan pemerintah. Dalam konteks ini, kajian dan penelitian yang mendalam dalam bidang informatika menjadi semakin relevan untuk menghadapi tantangan dan meraih peluang yang muncul dalam era digital saat ini. Saat ini, masyarakat memiliki kemudahan dalam mengakses informasi melalui berbagai platform media sosial. Platform ini meliputi WhatsApp, Facebook, YouTube, Instagram, Twitter, dan platform lainnya.

Dalam beberapa tahun terakhir, platform media sosial Instagram telah menjadi salah satu yang paling diminati untuk berbagi gambar dan video serta berinteraksi dengan pengguna lainnya. Fitur unik dari Instagram, seperti kemampuan berbagi konten visual, penggunaan tagar, dan opsi interaktif seperti stories dan live streaming, telah menjadikannya platform yang menarik bagi berbagai kalangan. Sebagai salah satu media sosial dengan jumlah pengguna yang sangat besar, Instagram memiliki potensi yang signifikan untuk memahami pandangan masyarakat terhadap beragam topik, peristiwa, merek, produk, atau kejadian tertentu.

2. TINJAUAN PUSTAKA

2.1. Media Sosial

Media sosial merupakan suatu platform di internet yang memungkinkan pengguna dari berbagai belahan dunia untuk berinteraksi, berfungsi sebagai wadah untuk menghubungkan individu yang berjauhan agar dapat merasa lebih dekat satu sama lain. Layanan media sosial memfasilitasi komunikasi antar pengguna tentang berbagai topik, menyiratkan bahwa media sosial adalah sumber data yang signifikan dan menjadi tempat di mana pengguna dengan mudah dapat mengungkapkan sentimen atau opini mereka. Dalam lingkup media sosial, terdapat tiga makna yang terungkap, yaitu pengenalan (cognition), komunikasi (communication), dan kolaborasi (cooperation). Saat ini, media sosial telah mencapai basis pengguna yang luas dan digunakan oleh individu di seluruh dunia. Keadaan ini memberikan potensi bagi media sosial untuk menjadi salah satu teknologi yang memainkan peran penting dalam era saat ini [1].

2.2. *Instagram*

Instagram telah mengalami pertumbuhan yang signifikan dan menjadi salah satu platform media sosial dengan jumlah pengguna terbesar di seluruh dunia. Platform ini sangat populer dan fokus pada berbagi foto dan video. Melalui unggahan gambar atau

video di Instagram, individu dapat berkomunikasi dan mengungkapkan pandangan mereka tanpa batasan. Dengan sejarah perkembangan selama lebih dari sepuluh tahun, Instagram saat ini memiliki lebih dari 1,32 miliar pengguna per Januari 2023. Di Indonesia sendiri, jumlah pengguna Instagram mencapai 89,15 juta. Oleh karena itu, kondisinya sangat memungkinkan untuk melakukan penelitian dalam analisis sentimen, karena adanya banyak pengguna aktif di Indonesia [2].

2.3. Pondok Pesantren

Pondok pesantren merupakan institusi pendidikan tradisional dalam Islam di Indonesia yang memiliki fokus pada penguasaan ilmu agama Islam dan pembentukan moral serta akhlak para siswa. Di lingkungan pondok pesantren, santri (siswa) tinggal bersama selama proses belajar. Biasanya, kepemimpinan pondok pesantren dipegang oleh seorang Syekh (Ulama) yang bertanggung jawab atas pendidikan dan perkembangan spiritual santri. Materi kurikulum di pondok pesantren mencakup pelajaran agama seperti Al-Qur'an, hadis, tafsir, fiqih, dan bahasa Arab untuk pemahaman teks-teks agama. Walaupun terdapat pelajaran umum seperti matematika, ilmu pengetahuan alam, dan bahasa Indonesia, pendekatan utama tetap berfokus pada pendidikan agama dan pembangunan spiritualitas [3].

2.4. *Data Mining*

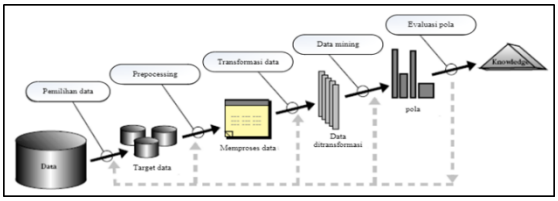
Data Mining merupakan proses pengelompokan data, di mana data diorganisir (dikelompokkan) ke dalam kategori atau kelompok yang telah ditetapkan sebelumnya. Secara alternatif, data mining bisa diartikan sebagai langkah-langkah pengeksplorasian data yang berfungsi sebagai alat untuk menganalisis informasi secara lebih tepat dan efektif. Tujuan utama dari data mining adalah untuk mengenali pola dan pengetahuan baru yang berguna dalam pengambilan keputusan dan perencanaan strategis di berbagai bidang seperti bisnis, ilmu pengetahuan, kedokteran, dan lainnya [4].

2.5. Analisis Sentimen

Analisis sentimen adalah cabang dalam ilmu komputer yang fokus pada bidang komputasi linguistik, pemrosesan bahasa alami, dan penambangan teks. Tujuan utama dari analisis sentimen adalah untuk memeriksa dan menganalisis emosi, penilaian, sikap, pandangan, serta sentimen yang dinyatakan oleh seseorang terhadap pembicara atau penulis terkait berbagai hal, seperti produk, layanan, organisasi, individu, tokoh publik, topik, acara, atau kegiatan tertentu. Oleh karena itu, diperlukan pendekatan analisis sentimen untuk mengenali dan mengkelompokkan opini pembaca menjadi positif dan negatif. Analisis sentimen juga bermanfaat dalam memperoleh wawasan mengenai sikap, opini, dan emosi yang tertanam dalam teks. Dengan bantuan analisis sentimen, data teks dapat diolah sehingga bisa mengenali sentimen yang ada, seperti apakah opini tersebut bersifat positif, negatif, atau netral terhadap suatu isu atau topik tertentu. Informasi yang dihasilkan dari analisis sentimen dapat memberikan pandangan berharga bagi para pengambil keputusan, peneliti, dan pelaku bisnis untuk memahami respons dan persepsi masyarakat terhadap suatu hal [5].

2.6. KDD *(Knowledge Discovery in Database)*

Knowledge Discovery in Database (KDD) adalah proses yang rumit dan terstruktur untuk menemukan, mengambil, serta menganalisis informasi berharga, pola, pengetahuan, dan wawasan baru dari data yang luas dan kompleks dalam suatu basis data. Ada 6 tahapan dalam proses KDD [6] seperti pada Gambar 1.



Gambar 1. Tahapan KDD

2.7. *Instagram* API

Instagram API (Application Programming Interface) merujuk pada sekumpulan petunjuk dan aturan yang memungkinkan para pengembang dan programmer berinteraksi dengan platform Instagram untuk mengakses data dan fungsi yang ada dalam aplikasi Instagram. Untuk mendapatkan akses ke API ini, pengguna diharuskan untuk mengajukan akun pengembang atau developer. Melalui Instagram API, individu dapat menggabungkan fitur-fitur Instagram ke dalam aplikasi atau situs web yang mereka kembangkan [7]

2.8. *Crawling Data*

Crawling adalah proses otomatis yang bertujuan untuk mengumpulkan data dalam jumlah besar dari berbagai halaman web. Teknik ini digunakan oleh mesin pencari dan aplikasi lainnya untuk mengambil informasi dari internet. Bot ini mengikuti tautan (link) dari satu halaman ke halaman lainnya dan mengambil data dari masing-masing halaman yang dikunjungi.
Pada Instagram, crawling data merujuk pada proses akses terhadap informasi pengguna Instagram dan postingan dari server Instagram. Data ini diambil melalui penggunaan Instagram API [8].

2.9. Kamus *Lexicon*

Lexicon merupakan sekumpulan kata yang ada dalam ingatan seseorang dan digunakan dalam proses berkomunikasi. Kamus Lexicon yang diterapkan dalam penelitian ini adalah kamus Lexicon dan *Negative Word* yang menggunakan bahasa Indonesia dan telah digunakan dalam penelitian sebelumnya oleh Koto dengan jumlah 3.609 kata positif dan 6.609 kata negatif. Kamus Lexicon dalam penelitian ini digunakan untuk menganalisis komentar-komentar masyarakat berdasarkan klasifikasi positif, negatif, atau netral. Lexicon memiliki peran yang sangat penting dalam analisis sentimen karena memungkinkan penyaringan kata-kata pendek seperti "tdk," "mkn," "gk," dan sejenisnya [9].

2.10. *Naïve Bayes Classifier*

Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi yang mengintegrasikan prinsip-prinsip statistik dan teori probabilitas. Teknik ini digunakan untuk menangani situasi dalam pembelajaran berbimbing *supervised learning*. Pendekatan klasifikasi *Naïve Bayes* menggunakan probabilitas bersyarat untuk setiap kelas yang mungkin, yang diambil dari data matriks, dan metode ini bersumber dari Teorema Bayes [10]. Formula dasar dari algoritma ini dirangkai seperti berikut ini:

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

Keterangan :

X : Sampel data belum diketahui labelnya
H : Hipotesis bahwa x merupakan data label
P(H) : Peluang dari hipotesa H
P(X) : Peluang data sampel
P(X|H) : Peluang data sampel apabila hipotesis benar
P(H|X) : Peluang hipotesis berdasarkan keadaan sampel

Ketika X adalah bukti, H adalah hipotesis, P (H|X) adalah probabilitas bahwa hipotesis H benar untuk bukti X, atau dengan kata lain P (H|X) merupakan probabilitas posterior H dengan syarat X. P(X|H) adalah probabilitas posterior X dengan syarat H. P (H) adalah probabilitas prior hipotesis H, dan P(X) adalah probabilitas prior bukti X.

2.11. Text Mining

Text mining merujuk pada proses penemuan dan ekstraksi informasi, pola, pengetahuan, atau wawasan yang awalnya tidak diketahui dari kumpulan teks atau dokumen. Teknik-teknik dari bidang pemrosesan bahasa *alami (Natural Language Processing/NLP)*, pembelajaran mesin *(machine learning)*, dan statistik diterapkan dalam text mining untuk mengidentifikasi hubungan dan pola dalam teks yang tidak terstruktur, sehingga informasi yang berharga dapat diambil dari data teks tersebut [11].

3. METODE PENELITIAN

Metodologi penelitian yang diterapkan dalam studi ini ialah knowledge discovery in Database (KDD). Proses tahapan metodologi ini dengan mengadopsi metode KDD melibatkan langkah-langkah seperti pemilihan data, pra-pemrosesan data, transformasi data, penambangan data, serta evaluasi dan interpretasi pengetahuan. Adapun tahapan KDD yang digunakan yaitu:

3.1. *Data Selection*

Pada tahap awal ini, data-data pada database akan diseleksi kemudian disimpan didalam berkas yang terpisah.

3.2. *Pre-processing/Cleaning*

Tahapan berikutnya adalah melakukan pembersihan terhadap data-data yang sudah diperoleh sebelum dilakukan proses data mining.

3.3. *Transformation*

Tahapan ini merupakan suatu proses dimana akan dilakukannya transformasi atau perubahan bentuk data terhadap data-data yang sudah di bersihkan sehingga nantinya data tersebut dapat dilanjutkan ke dalam proses data mining.

3.4. *Data Mining*

Di dalam tahapan ini akan dilakukan proses pencarian pola atau informasi dengan menggunakan metode atau teknik tertentu terhadap data-data.

3.5. *Interpretation/Evaluation*

Tahap ini merupakan tahap terakhir yang dimana akan dilakukannya proses evaluasi terhadap hasil dari proses data mining yang sudah dilakukan.

4. HASIL DAN PEMBAHASAN

Penelitian ini mengkaji analisis sentimen masyarakat terhadap dugaan kontroversi dengan menggunakan algoritma Naïve Bayes. Metodologi yang diterapkan adalah *Knowledge Discovery in Database* (KDD), yang mencakup tahapan Data Selection, Pre-Processing, Transformation, Data Mining, dan Evaluation. Pengumpulan data dilakukan melalui teknik Crawling Data dari Instagram API menggunakan bahasa pemrograman Python. Jumlah komentar yang berhasil dikumpulkan mencapai 1.483 komentar dalam rentang waktu 1 bulan, yaitu bulan Juli.Langkah berikutnya setelah mengumpulkan data komentar adalah melanjutkan menuju tahap Pre-Processing dengan tujuan menghilangkan noise dan karakter yang tidak relevan dari data dalam proses klasifikasi. Setelah proses pembersihan selesai, data yang tersisa adalah 1.010 data komentar. Hasil dari pembobotan sentimen ini menghasilkan data dengan 64 komentar positif, 400 komentar negatif, dan 546 komentar netral dari total 1.010 data dalam dataset.
Hasil penelitian yang telah dilakukan yaitu mengetahui sentimen komentar para pengguna Instagram di Indonesia terhadap dugaan kontroversi Pondok Pesantren Al-Zaytun. Data komentar Instagram tersebut kemudian dibagi ke dalam tiga kategori, yaitu positif, negatif, dan netral, dengan menggunakan algoritma klasifikasi Naïve Bayes Classifier. Hasil klasifikasi dan perhitungan skor sentimen ini nantinya dapat dijadikan sebagai referensi data informasi mengenai opini masyarakat terkait dugaan tersebut di masa yang akan datang.

	userId	createdAt	text
0	18377019274023941	2023-07-15T06:21:03.000Z	Ini proses pengusutannya kok lama yaah ,,, kok...
1	17991648760945848	2023-07-15T12:53:14.000Z	❤❤❤❤❤❤❤❤❤❤❤❤❤❤❤
2	18011805094695580	2023-07-16T00:43:09.000Z	Ajaran ODGJ
3	18226703467227754	2023-07-16T05:19:12.000Z	Knp ngk mati aj yg nyiptakan zaitun
4	17976827678380463	2023-07-16T05:37:52.000Z	Abu Jahal versi Indonesia
...	...	...	...
1478	17974285265385796	2023-07-12T15:57:12.000Z	@mia.musa14 Wah, ada lagi nih pengikut fanatik...
1479	18029918299548318	2023-07-12T16:02:36.000Z	@edokurnia1 😊
1480	18291245143139389	2023-07-13T06:34:31.000Z	iya pesantren mmek ngajarin sesad tolol udah t...
1481	18290035900188458	2023-07-13T06:34:56.000Z	@wildaneka19 wkwkwk, iya dosa nya di bayar pke...
1482	18252831817167077	2023-07-26T10:04:58.000Z	@febriant1212 caper bgt hahah ,kyk udh dapet t...

1483 rows × 3 columns

Gambar 2. *Dataset* Penelitian Hasil *Data Selection*

4.1. *Data Selection*

Hasil dari tahap ini adalah kumpulan data yang akan digunakan dalam penelitian. Teknik yang digunakan adalah pengambilan data menggunakan crawling dari Instagram API.

4.2. *Pre Processing*

Setelah proses pengumpulan data, ditemukan bahwa jenis data yang akan digunakan masih belum ideal untuk menjalani proses data mining. Oleh karena itu, diperlukan tahapan pre-processing untuk menghilangkan kata-kata atau karakter-karakter yang tidak relevan.

4.3. *Cleaning*

Tabel 1. Hasil dari Tahap Cleaning

Komentar sebelum dilakukan cleaning	Komentar sesudah dilakukan cleaning
baca dah biar ga bod*k banget https://youtu.be/va_klOxPTxs Takbir, selamatkan agama Islam #AlZaytunSesat	baca dah biar ga bodk banget Takbir, selamatkan agama Islam
Sok baik pas di kamera doang, spesialis ngangong ngangong ðŸ'ðŸ'ðŸ'ðŸ' #ZaitunTitisanDajjal	Sok baik pas di kamera doang, spesialis ngangong ngangong

Proses cleaning bertujuan untuk penghapusan URL, angka, dan simbol-simbol yang tidak relevan dalam proses klasifikasi. Berikut ini adalah contoh hasil dari penerapan tahap cleaning, dapat dilihat pada Tabel 1.

Setelah proses *cleaning* dilakukan, komentar telah dibersihkan sehingga dapat dilakukannya proses klasifikasi.

4.4. *Case Folding*

Pada tahap ini, dilakukan perubahan terhadap semua huruf kapital dalam data menjadi huruf kecil. Tahap ini dilakukan untuk memastikan bahwa seluruh data yang akan diproses memiliki karakter yang seragam. Hal ini bertujuan agar proses *pre-processing* selanjutnya dapat berjalan lebih efektif dan menghasilkan hasil *pre-processing* yang optimal. Berikut contoh dari tahap *case folding* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Tahapan *Case Folding*

Komentar sebelum dilakukan case folding	Komentar sesudah dilakukan case folding
Muridnya Kan Bisa Dipindahkan Karena Ajaran Sesat , Kenapa Masih Proses Trus , Nunggu Apa , Trus Itu Kan Ajaran Sesat Klo Trlalu Lama Apa Gak Tambah Membahayakan Siswa Sm Siswinya , Generasi Penerus Bangsa Loh	muridnya kan bisa dipindahkan karena ajaran sesat , kenapa masih proses trus , nunggu apa , trus itu kan ajaran sesat klo trlalu lama apa gak tambah membahayakan siswa sm siswinya , generasi penerus bangsa loh
FPI Yg Baik Bntu Bencana Gempa,,Di Bubarin, Al Zaitun Yg Sesah Bikin Aqidah Rusak Masa Ga Di Bubarin, Berati Apa Pemerintah Nye Pd Ga Bener 2019-2024 Maka Nya Indonesia Ancur Byk Utang	fpi yg baik bntu bencana gempa,,di bubarin, al zaitun yg sesah bikin aqidah rusak masa ga di bubarin, berati apa pemerintah nye pd ga bener 2019-2024 maka nya indonesia ancur byk utang
Pake Akal Pikiran Klo Mau Pilih2 Masukin Ke Pesantren. Modelan Pesantren Begini Aliran Nya Bukan Islam Sesungguh Nya, Malah Pada Pake Jas, Kek Di Gereja. Agak Laen Emang	pake akal pikiran klo mau pilih2 masukin ke pesantren. modelan pesantren begini aliran nya bukan islam sesungguh nya, malah pada pake jas, kek di gereja. agak laen emang

Setelah tahap *case folding* dilakukan, terlihat bahwa semua huruf dalam dataset menjadi huruf kecil, termasuk huruf kapital.

4.5. *Tokenizing*

Setelah data telah dibersihkan dan diubah menjadi huruf kecil, langkah berikutnya adalah tahap *tokenizing*. *Tokenizing* melibatkan pemecahan kalimat dalam data komentar menjadi kata-kata individual. Tahap ini akan mempermudah langkah transformasi, karena data tidak akan diproses dalam bentuk kalimat utuh, melainkan sebagai kumpulan kata-kata dari kalimat tersebut. Hasil dari tahap tokenizing dapat dilihat pada Tabel 3.

Tabel 3. Hasil Tahapan *Tokenizing*

Komentar sebelum dilakukan tokenizing	Komentar sebelum dilakukan tokenizing
Lambat amat jelas2 menyesatkan masih diberi kelonggaran, bubarkan saja	'lambat', 'amat', 'jelas', 'menyesatkan', 'masih', 'diberi', 'kelonggaran', 'bubarkan', 'saja'
Heran masih ada orang tua yg masukin anaknya di pesantren sesat gini	'heran', 'masih', 'ada', 'orang', 'tua', 'yg', 'masukin', 'anaknya', 'di', 'pesantren', 'sesat', 'gini'
Pondok apaan kaya gini??, Bukannya bikin Sholeh/ Sholehah, malah sesat	'pondok', 'apaan', 'kaya', 'gini', 'bukannya', 'bikin', 'sholeh', 'sholehah', 'malah', 'sesat'

Pada Tabel 3 memperlihatkan hasil dari tahap *tokenizing* di mana data komentar yang semula berbentuk kalimat berhasil dipecah menjadi kata-kata secara individu. Setiap kata dipisahkan dengan menggunakan tanda kutip satu yang ditempatkan di awal dan akhir setiap kata.

4.6. *Stopword Removal*

Pada tahap ini, digunakan dokumen yang berisi daftar kata-kata *stopword* untuk membantu dalam penyaringan kata-kata yang tidak dibutuhkan dalam proses klasifikasi. Hasil dari tahap *stopword removal* yang dibantu oleh dokumen daftar *stopword* terlihat dalam Tabel 4.

Tabel 4. Hasil Tahapan

Komentar sebelum dilakukan stopword removal	Komentar sesudah dilakukan stopword removal
'loh', 'yg', 'mlesetin', 'kan', 'panji', 'gumilang', 'kok', 'marahnya', 'ke', 'saya', 'gimana', 'kau', 'ni'	'loh', 'mlesetin', 'panji', 'gumilang', 'marah', 'gimana', 'kau', 'ni'
'ingat', 'ajaran', 'nabi', 'muhammad', 'tidk', 'seperti', 'itu', 'ajaran', 'nabi', 'selalu', 'benar'	'ajar', 'nabi', 'muhammad', 'tidk', 'ajar', 'nabi'

Pada contoh di atas dijelaskan bahwa ada beberapa kata yang dihilangkan seperti kata itu, ingat, dan kata seperti.

4.7. *Stemming*

Langkah terakhir dari tahap *pre-processing* adalah tahap *stemming*, yang bertujuan untuk merubah kata-kata dalam dokumen menjadi bentuk kata dasar. Tahap *stemming* ini dilakukan menggunakan package kata dasar. Perbedaan antara sebelum dan setelah tahap stemming dapat dilihat pada Tabel 5.

Tabel 5. Hasil Proses *stemming*

Kata sebelum stemming	Kata sesudah stemming
Kelonggaran	Longgar
Sesungguhnya	Sungguh
Menyesatkan	Sesat
Membubarkan	Bubar
Membahayakan	Bahaya
Ditambahkan	Tambah

Pada Tabel 5 terdapat contoh kata-kata yang sudah melalui tahap *stemming*, dan berikut adalah hasil tahap *stemming* dalam *dataset*. Dapat dilihat pada Tabel 6.

Tabel 6. Hasil tahapan *stemming*

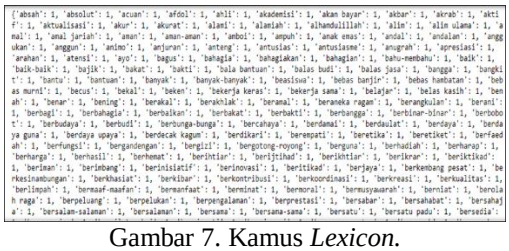
Komentar sebelum dilakukan stemming	Komentar sesudah dilakukan stemming
'lambat', 'amat', 'jelas', 'menyesatkan', 'masih', 'diberi', 'kelonggaran', 'bubarkan', 'saja'	'lambat', 'amat', 'jelas', 'sesat', 'masih', 'beri', 'longgar', 'bubar', 'saja'
'heran', 'masih', 'ada', 'orang', 'tua', 'yg', 'masukin', 'anaknya', 'di', 'pesantren', 'sesat', 'gini'	'heran', 'masih', 'ada', 'orang', 'tua', 'masuk', 'anak', 'pesantren', 'sesat', 'gini'

Jika dilihat dari Tabel 6, kata 'menyesatkan' dihapus imbuhan yang ada pada kata tersebut. Melalui tahap stemming kata 'menyesatkan' menjadi 'sesat', kata tersebut kembali kepada kata dasarnya. Setelah melalui proses *pre-processing,* dataset berkurang menjadi 1.010 data.

4.8. *Transformating Data*

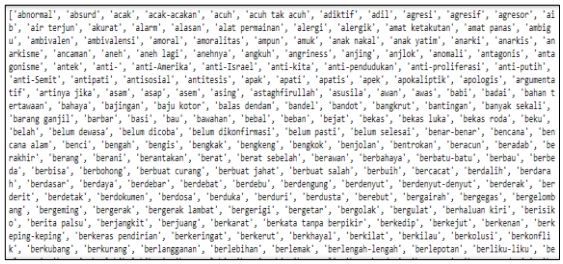
Tahap selanjutnya setelah melakukan tahap *pre-processing* merupakan langkah dalam proses *data transformation.* Sebelum memasuki tahap ini, terlebih dahulu dilakukan penilaian sentimen pada dataset untuk melakukan pembobotan memberikan label pada data. Pada tahap ini, skor sentimen dihitung dengan menghitung jumlah kata-kata yang Mengacu pada kata-kata yang memiliki aspek *positif* dan kata-kata yang memiliki aspek negatif.

Hasil dari penambahan kamus Lexicon dapat dilihat pada Gambar 7.



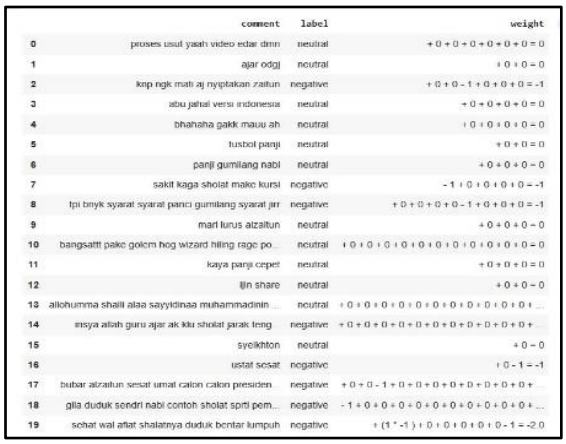
Gambar 7. Kamus *Lexicon.*

Gambar 7 pada bagian sebelumnya mengilustrasikan penerapan *kamus Lexicon* yang telah diperluas dengan memberikan penilaian 1 untuk kata-kata yang memiliki *polaritas positif* dan nilai -1 untuk kata-kata dengan *polaritas negatif* kamus Kata-Kata *Negatif* terlihat pada bagian berikutnya dalam Gambar 8 di bawah ini.



Gambar 8. Kamus *Negative Words*

Setelah penambahan Langkah berikutnya adalah mengkomputasi skor sentimen dengan menganalisis polaritas dalam setiap kalimat dalam dataset komentar. Efek dari proses pembobotan sentimen menggunakan *kamus Lexicon* dan *Negative Words* dapat diamati pada hasilnya. dalam Gambar 9 berikut.



	comment	label	weight
0	proses usuf yaah video edar dmn	neutral	+ 0 + 0 + 0 + 0 + 0 + 0 = 0
1	ajar odgj	neutral	+ 0 + 0 = 0
2	knp ngk mati aj nyiptakan zaitun	negative	+ 0 + 0 - 1 + 0 + 0 + 0 = -1
3	abu jahal versi indonesia	neutral	+ 0 + 0 + 0 + 0 = 0
4	bhahaha gakk mauu ah	neutral	+ 0 + 0 + 0 + 0 = 0
5	husbol panji	neutral	+ 0 + 0 = 0
6	panji gumilang nabi	neutral	+ 0 + 0 + 0 = 0
7	sakit kaga sholat make kursi	negative	- 1 + 0 + 0 + 0 + 0 = -1
8	tpi bnyk syarat syarat panci gumilang syarat jin	negative	+ 0 + 0 + 0 + 0 - 1 + 0 + 0 + 0 = -1
9	mari lurus alzaitun	neutral	+ 0 + 0 + 0 = 0
10	bangsattt pake golem hog wizard hiling rage po...	neutral	+ 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0
11	kaya panji cepet	neutral	+ 0 + 0 + 0 = 0
12	ijin share	neutral	+ 0 + 0 = 0
13	allohumma shalli alaa sayyidinaa muhammadinin ...	neutral	+ 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + ...
14	insya allah guru ajar ak klu sholat jarak teng...	negative	+ 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 +
15	syelkhton	neutral	+ 0 = 0
16	ustat sesat	negative	+ 0 - 1 = -1
17	bubar alzaitun sesat umat calon calon presiden ...	negative	+ 0 + 0 - 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 +
18	gila duduk sendri nabi contoh sholat sprti pem...	negative	- 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + ...
19	sehat wal afiat shalatnya duduk bentar lumpuh	negative	+ (1 * -1) + 0 + 0 + 0 + 0 + 0 - 1 = -2.0

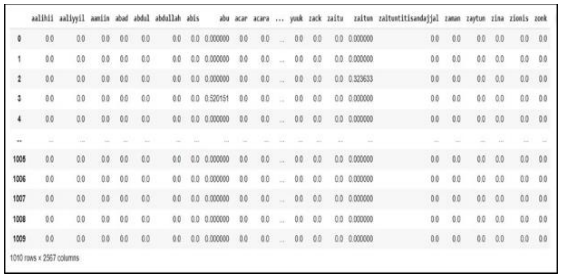
Gambar 9. Hasil Pembobotan Sentimen

Pada Gambar 9 terlihat bahwa perhitungan skor dilakukan dengan memberikan nilai pada setiap kata dalam kalimat dan mengakumulasikan totalnya. Jika total skor sentimen lebih dari 0, maka masuk ke dalam kelas positif. Jika skor sentimen kurang dari 0, kalimat masuk ke dalam kelas negatif. Dan jika skor sentimen tidak termasuk apabila nilai skor sentimen berada dalam *rentang positif* atau *negatif*, kalimat tersebut akan dikategorikan sebagai kelas netral. Jumlah data yang telah diberi label *positif*, *negatif*, dan *netral* setelah proses pembobotan sentimen tersebut dapat diobservasi dalam Tabel 7 berikut.

Tabel 7. Label dan Jumlah Data

Label/Kelas	Jumlah
Positif	64
Negatif	400
Netral	546
Total	**1.010**

Dengan menggunakan kamus *Lexicon dan Negative Words*, hasil dari pembobotan sentimen berhasil mengelompokan total 1.010 data ke dalam tiga kelas. Terdapat 64 data yang masuk ke dalam kelas positif, 400 data masuk ke dalam kelas negatif, dan 546 data masuk ke dalam kelas netral. Setelah melakukan pembobotan sentimen, langkah selanjutnya adalah melakukan pembobotan kata dengan menerapkan metode TF-IDF *(Term Frequency-Inverse Document Frequency)* Berikut adalah hasil dari TF-IDF pada Gambar 10.



	aalihii	aaliyyil	aamiin	abad	abdul	abdullah	abis	...	yuuk	zack	zaitu	zaitun	zaitunitisunadjjal	zaman	zaytun	zina	zionis	zonk
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1005	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1006	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1007	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1009	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1010 rows × 2567 columns

Gambar 10. Hasil Pembobotsn TF-IDF

Pada Gambar 10 terlihat bahwa setiap kata/term dalam korpus diurutkan secara abjad dan kemudian dihitung probabilitas kemunculan kata tersebut dalam dokumen, sebagai langkah persiapan untuk tahap selanjutnya.

4.9. *Data Mining*

Setelah tahap *transformation* selesai, langkah selanjutnya adalah Tahap ini melibatkan ekstraksi pengetahuan dari data yang telah terkumpul. Proses ini melibatkan klasifikasi data dengan menggunakan model dari algoritma *Naïve Bayes* yang dikenal dengan sebutan *Multinomial Naïve Bayes*. Proses klasifikasi ini diterapkan dengan memanfaatkan perpustakaan *(library)* sklearn. naive_bayes pada bahasa pemrograman Python. *Implementasi algoritma Multinomial Naïve Bayes* ini secara lebih rinci dapat ditemukan pada gambar 11 berikut ini.

```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB().fit(X_train_df, y_train)
prediction_mi = model.predict(X_test_df)
prediction_proba_mi = model.predict(X_test_df)
```

Gambar 11. *Library Multional Bayes*

Dalam penelitian ini, dilakukan pengolahan data menggunakan algoritma *Naïve Bayes* dengan 4 skenario pembagian data, yaitu 90:10, 80:20, 70:30, dan 60:40. Hasil analisis sentimen terhadap dugaan kontroversi di Ponpes Al-Zaytun *menggunakan Naïve Bayes Classifier* dievaluasi dengan *Confusion Matrix*. Berikut adalah pembagian antara data training dan data testing Tabel 8 skenario pengujian data training dan testing.

Tabel 8. Skenario Pengujian *Data Training* dan *Testing*.

Persentase Data		Jumlah Data	
Training	**Testing**	**Training**	**Testing**
90%	10%	909	101
80%	20%	808	202
70%	30%	707	303
60%	40%	606	404
Total		**1.010**	

4.10. *Evaluation*

Setelah proses pembuatan model selesai, model yang telah dibuat dari setiap skenario akan diuji. Hasil pengujian model meliputi Skor akurasi, presisi, recall, dan nilai *F-measure (f1-score)* dihitung untuk mengevaluasi hasil klasifikasi.

1. Skenario 1. (90% : 10%)

Gambar di bawah ini merupakan hasil klasifikasi dengan menggunakan 90% data training dan 10% data *testing*.

```
Multinomial NB Accuracy  :  0.7920792079207921
Multinomial NB Precision :  0.5292929292929293
Multinomial NB Recall    :  0.5655172413793104
Multinomial NB F-Measure :  0.546236559139785
```

Gambar 12. Hasil Skenario 1.

Dari Gambar 12 hasil Penerapan klasifikasi menggunakan algoritma *Naïve Bayes* dilakukan dengan membagi data dalam perbandingan 90:10 untuk tujuan evaluasi. Dalam pengujian ini, 90% data digunakan untuk pelatihan model, sementara 10% sisanya digunakan untuk menguji kinerja model. menunjukkan nilai *accuracy* sebesar 80%, *precision* 53%, *recall* 56%, dan *f-measure* 54%.

2. Skenario 2 (80% : 20%)

Gambar di bawah ini merupakan hasil klasifikasi dengan menggunakan 80% data training dan 20% data testing.

```
Multinomial NB Accuracy  :  0.7425742574257426
Multinomial NB Precision :  0.5010974539069358
Multinomial NB Recall    :  0.5204351204351204
Multinomial NB F-Measure :  0.5070364548244628
```

Gambar 13. Hasil Skenario 2.

Dari Gambar 13, hasil klasifikasi menggunakan *Naïve Bayes* pada perbandingan 80:20 menunjukkan nilai *accuracy* sebesar 74%, precision 50%, recall 52%, dan *f-measure* 50%.

3. Skenario 3 (70% : 30%)

Gambar di bawah ini merupakan hasil klasifikasi dengan menggunakan 70% data *training* dan 30% data *testing*.

```
Multinomial NB Accuracy  :  0.7326732673267327
Multinomial NB Precision :  0.4984560046150192
Multinomial NB Recall    :  0.5063685636856369
Multinomial NB F-Measure :  0.4950514895642806
```

Gambar 14. Hasil Skenario 3

Dari Gambar 14, hasil klasifikasi menggunakan *Naïve Bayes* pada perbandingan 70:30 menunjukkan nilai *accuracy* sebesar 73%, *precision* 49%, *recall* 50%, dan *f-measure* 49%.

4. Skenario 4 (60% : 40%)

Gambar di bawah ini merupakan hasil klasifikasi dengan menggunakan 60% data *training* dan 40% data *testing*.

```
Multinomial NB Accuracy  :  0.7202970297029703
Multinomial NB Precision :  0.4855072463768116
Multinomial NB Recall    :  0.5005850005850006
Multinomial NB F-Measure :  0.4879945313167564
```

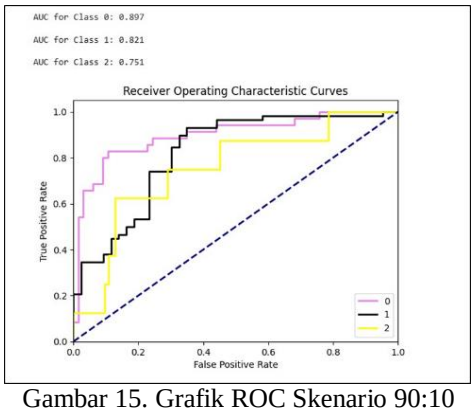
Gambar 14. Hasil Skenario 4

Dari Gambar 4.10, hasil klasifikasi menggunakan *Naïve* Bayes pada perbandingan 60:40 menunjukkan nilai *accuracy* sebesar 72%, *precision* 48%, *recall* 50%, dan *skor f-measure* mencapai 48%. Nilai AUC untuk semua model Multinomial Naïve Bayes dapat diakses pada bagian yang relevan. dalam tabel 9 berikut.

Tabel 9. Hasil Nilai AUC Setiap Skenario

Skenario	AUC	Kualitas
90:10	0.82	*Good Classification*
80:20	0.74	*Fair Classification*
70:30	0.76	*Fair Classification*
60:40	0.75	*Fair Classification*

Dari tabel di atas, hampir semua model dalam skenario menghasilkan nilai AUC yang memiliki kualitas klasifikasi yang cukup baik. Nilai AUC terendah terdapat pada model skenario 80:20 dengan nilai 0,74, sementara nilai AUC tertinggi terdapat pada model skenario 90:10 dengan nilai 0,82 dan memiliki kualitas klasifikasi yang baik. Berikut ini adalah grafik ROC dari model dengan nilai AUC tertinggi, yaitu pada model skenario 90:10



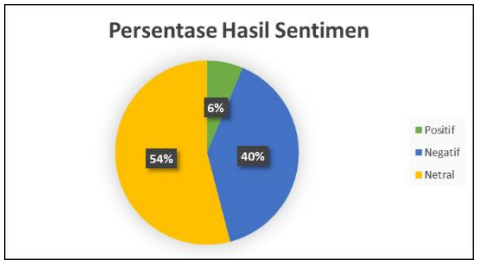
Gambar 15. Grafik ROC Skenario 90:10

Gambar 15 menampilkan Dengan mempertimbangkan nilai-nilai kelas tersebut, nilai skor ROC AUC pada kelas model positif yang berwarna kuning sebesar 0.751, sedangkan netral dan berwarna hitam sebesar 0.821 sedangkan untuk negatif yang berwarna merah muda sebesar 0.897. Dengan angka tersebut maka skor ROC AUC adalah

$$\text{ROC AUC } Score = \frac{0.751+0.821+0.897}{3} = 0.82$$

Dapat disimpulkan bahwa pada model skenario 90:10, nilai AUC yang tertinggi adalah 0,80 merupakan kualitas *Good Classification.*

Pada Gambar 16 di berikut ini merupakan persentase sentimen masyarakat yang terdapat pada komentar Instagram terhadap kontroversi Ponpes Al-Zaytun, komentar netral mempunyai persentase 54%, komentar negatif 40%, dan komentar positif hanya 6%.



Gambar 16. Persentase Hasil Sentimen

.

Langkah berikutnya setelah *Split Data* adalah menerapkan pembobotan kata menggunakan metode *TF-IDF*. Hasil dari pengujian dan perbandingan dari keempat model *Multinomial Naïve Bayes* dapat ditemukan ditemukan dalam Tabel 10 berikut.

Tabel 10. Perbandingan Hasil *Evaluasi Model Naïve Bayes*

Model	*Accuracy*	*Precision*	*Recall*	*F-Measure*	AUC	Kualitas
90:10	80%	53%	56%	54%	0.82	*Good Classification*
80:20	74%	50%	52%	50%	0.74	*Fair Classification*
70:30	72%	49%	50%	49%	0.76	*Fair Classification*
60:40	72%	48%	50%	48%	0.75	*Fair Classification*

Tabel 10 menggambarkan perbandingan antara model-model skenario yang telah diuji. Dalam pengujian nilai accuracy, model dengan accuracy terendah adalah skenario 70:30 dan 60:40 dengan nilai masing-masing 72%, sementara accuracy tertinggi terdapat pada skenario 90:10 dengan 82%. Pada nilai precision, skenario 60:40 memiliki nilai terendah yaitu 48%, sedangkan nilai tertinggi ada pada skenario 90:10 dengan 53%. Nilai recall paling tinggi ditemukan pada skenario 90:10 dengan 56%, sementara nilai terendah ada pada skenario 70:30 dan 60:40 dengan nilai masing-masing 50%. Untuk F-Measure, skenario 90:10 memiliki nilai tertinggi yaitu 54%, sementara nilai terendah ada pada skenario 60:40 dengan 48%. Pada nilai AUC, skenario 80:20 memiliki nilai terendah dan skenario 90:10 memiliki nilai tertinggi bahwa skenario 90:10 memiliki kualitas *Good Classification.*

Gambar 17 *Wordcloud* Komentar Kata Kunci "al-zaytun"

Pada Gambar 17, menampilkan *wordcloud* yang merupakan kumpulan kata-kata yang sering muncul dalam dataset. Kata-kata yang paling sering muncul antara lain adalah "sesat", "aja", "gak". dan "Allah". Semakin besar ukuran kata dalam wordcloud, semakin tinggi frekuensi penggunaan kata tersebut. Ini menunjukkan bahwa kata-kata tersebut sering digunakan oleh masyarakat dalam komentar Instagram pada postingan Al-Zaytun.

DAFTAR PUSTAKA

[1] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.

[2] A. Novantirani, M. K. Sabariah, dan V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeeding Eng.*, vol. 2, no. 1, hal. 1–7, 2015.

[3] B. Liu, *Web data mining: exploring hyperlinks, contents, and usage data*, vol. 1. Springer, 2011.

[4] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," vol. 4, no. 1, hal. 156–161, 2018.

[5] W. Gata dan P. P. Purnomo, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," *Format*, vol. 6, no. 1, hal. 1–13, 2017.

[6] B. Dash, D. Mishra, A. Rath, dan M. Acharya, "A hybridized K-means clustering approach for high dimensional dataset," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 2, hal. 59–66, 2010, doi: 10.4314/ijest.v2i2.59139.

[7] F. Febriansyah, N. R, A. I. Purnamasari, O. Nurdiawan, dan S. Anwar, "Pengenalan Teknologi Android Game Edukasi Belajar Aksara Sunda untuk Meningkatkan Pengetahuan," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, hal. 336, Des 2021, doi: 10.30865/jurikom.v8i6.3676.

[8] F. C. Ningrum, D. Suherman, S. Aryanti, H. A. Prasetya, dan A. Saifudin, "Pengujian Black Box pada Aplikasi Sistem Seleksi Sales Terbaik Menggunakan Teknik Equivalence Partitions," vol. 4, no. 4, 2019.

[9] F. Darwis *et al.*, "SISTEM ANALISIS SENTIMEN PUBLIK TENTANG OPINI PEMILIHAN KEPALA DAERAH JAWA TIMUR 2018 PADA DOKUMEN TWITTER MENGGUNAKAN NAIVE BAYES," 2018.

[10] D. Suprapti, M. Kamisutara, dan P. Artaya, "Analisa Pengujian Informasi Penjualan menggunakan Metode White Box Seminar Nasional Ilmu Terapan (SNITER) 2017-Universitas Widya Kartika."

[11] A. Sentimen dan F. Media, "ANALISIS SENTIMEN REVIEW CUSTOMER TERHADAP PRODUK INDIHOME DAN FIRST MEDIA MENGGUNAKAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK REVIEW ANALYSIS SENTIMENT CUSTOMER PRODUCT INDIHOME AND FIRST MEDIA USING CONVOLUTIONAL NEURAL NETWORK," vol. 8, no. 5, hal. 9049–9061, 2021.