

ANALISA SENTIMEN UNTUK MENGIDENTIFIKASI KECENDERUNGAN RADIKALISME DENGAN NAÏVE BAYES

Fajar Yulianto^{1*}, Hartarto Junaedi², Suhatati Tjandra³, Amanda Pascarini⁴

^{1,2,3}Magister Teknologi Informasi, Teknik, Institut Sains Dan Teknologi Terpadu

Surabaya, ⁴Magister Psikologi, Universitas 17 Agustus 1945 Surabaya

E-mail: ^{*}1fajar3435@gmail.com

ABSTRACT

Radicalism in Indonesia is still an issue that is often discussed considering that there are still many acts of radicalism in Indonesia. The relatively rapid spread of radicalism requires special handling to stop the spread. This study aims to identify radicalism by grouping it into 3 (three) groups or levels which are later expected to facilitate the search for solutions to stop the spread of radicalism. The analysis system will use sentiment analysis using the method of machine learning, namely Naive Bayes. The data used were collected through surveys to students and religious leaders. The Naive Bayes method will later serve to determine the survey results of each individual based on his group. The total data collected for this study amounted to 250 respondents who have filled out survey questions, the data will be divided into 2 types, there are 165 data used for the training phase and 85 data for testing. After processing the data, the results were obtained using classification report calculations and obtained an accuracy of 85% from 3 radical groups.

Keywords: sentiment analysis; machine learning; naïve bayes; radicalism

ABSTRAK

Radikalisme di Indonesia, hingga saat ini masih menjadi suatu isu yang sering diperbincangkan mengingat masih banyaknya aksi-aksi radikalisme di Indonesia. Penyebaran paham radikalisme yang relative cepat memerlukan adanya penanganan khusus untuk menghentikan penyebaran tersebut. Penelitian ini bertujuan untuk mengidentifikasi radikalisme dengan mengelompokannya menjadi 3 (tiga) kelompok atau tingkatan yang nantinya diharapkan dapat memudahkan pencarian solusi penghentian penyebaran paham radikalisme tersebut. Sistem analisis akan menggunakan analisis sentiment dengan menggunakan metode dari machine learning yaitu naive bayes. Data yang digunakan sendiri dikumpulkan melalui survey kepada santri dan pemuka agama. Metode naive bayes nantinya akan berfungsi untuk menentukan hasil survey tiap individu tersebut berdasarkan kelompoknya. Total data yang terkumpul untuk penelitian ini berjumlah 250 responden yang sudah mengisi pertanyaan survey, data tersebut nantinya akan dibagi 2 jenis, terdapat 165 data digunakan untuk tahap training dan 85 data untuk testing. Setelah melakukan proses olah data hasil yang didapat menggunakan perhitungan classification report dan didapat akurasi sebesar 85% dari 3 kelompok radikal.

Kata Kunci: sentiment analysis; *machine learning*; *naïve bayes*; radikalisme

1. Pendahuluan

Radikalisme merupakan suatu istilah yang berasal dari bahasa latin yakni “*radix*” yang memiliki arti “akar, pengkal, bagian bawah, atau menyeluruh, dan amat keras dalam

menuntut sebuah perubahan. Radikalisme di Indonesia, hingga saat ini masih menjadi suatu isu yang sering diperbincangkan mengingat masih banyaknya aksi-aksi radikalisme di Indonesia. Radikalisme menurut BNPT

merupakan embrio lahirnya terorisme sehingga radikalisme dan terorisme merupakan istilah yang tidak dapat dipisahkan. Radikalisme lebih lanjut, BNPT mendefinisikannya sebagai suatu sikap dimana terdapat suatu keinginan akan perubahan secara total dengan mengubah nilai – nilai secara drastis melalui kekerasan maupun aksi ekstrem. Potensi radikalisme sosial-keagamaan di Indonesia dapat dikatakan cukup tinggi, tercatat pada tahun 2016 sekitar 600.000 muslim Indonesia pernah terlibat tindakan radikal, 11 juta bersedia radikal[1]. Sedangkan terorisme sendiri dalam UU No. 15 Th. 2003 tentang Pemberantasan Tindak Pidana Terorisme dimaknai sebagai aksi yang sesuai dengan tindak kriminal yang tertera dalam PP No. 1 Tahun 2002, terorisme disini dimaknai sebagai perbuatan melawan hukum secara sistematis yang memiliki maksud untuk menghancurkan keutuhan bangsa dan negara.

Radikalisme di Indonesia mulai berkembang setelah berakhirnya masa orde baru. Sejak saat itu Indonesia sering disebut sebagai sarang teroris. Peta konsentrasi jaringan teroris Indonesia menunjukkan hampir di seluruh pulau di Indonesia terdapat kelompok – kelompok teroris. Namun tidak semua yang memiliki sikap dan pemahaman radikal pasti menjadi seorang teroris, banyak faktor yang menyebabkan seseorang akhirnya menjadi seorang teroris

Potensi radikalisme di Indonesia banyak didominasi oleh usia – usia produktif, sebanyak 47,3% berpotensi di usia kisaran 21 – 30 tahun, dan 29,1% di kisaran 31 – 40 tahun, sisanya antara usia 11 – 20 tahun dan lebih dari 40 tahun[2]. Fenomena radikalisme ini makin hari semakin menguat mengikuti arus globalisasi dan didukung dengan

semakin mudahnya kita untuk saling bertukar informasi dalam bentuk media fisik ataupun media sosial. Penguatan radikalisme ini cukup berbahaya mengingat dengan adanya teknologi wilayah persebarannya makin luas dan hampir mencakup seluruh wilayah di Indonesia seperti terlihat dari Gambar 1.



Gambar 1. Peta Persebaran Jaringan Terorisme

Penyebaran paham radikalisme di Indonesia didominasi oleh media teknologi, penyebaran paham ini paling banyak tersebar melalui internet. Tercatat selama tahun 2018 Kementerian Kominfo melakukan pemblokiran terhadap 10.4490 konten dari berbagai media social dan sumber yang dapat diakses melalui internet. Dengan cepatnya penyebaran paham radikalisme di Indonesia perlu adanya penanganan khusus untuk menghentikan penyebaran tersebut. BNPT menyampaikan bahwa penyelesaian terkait radikalisme dan terorisme tidak dapat dilakukan dengan hanya menegakkan dan melakukan penindakan hukum, namun perlu adanya upaya pencegahan dan penindakan sampai ke pusat terkecilnya[3].

Radikalisme sejatinya berkaitan dengan ideology seseorang yang mana dipengaruhi oleh banyak faktor. Adapun faktor tersebut berdasarkan

survey menunjukkan hasil pada Gambar 2.



Gambar 2. Faktor Yang Mempengaruhi Radikalisme

Namun hasil tersebut masih belum spesifik dan kurang lengkap. Natalia menjelaskan dalam penelitiannya bahwa agama di Indonesia memiliki nilai yang cukup sakral dan hal ini secara sadar tidak sadar dapat menguasai kesadaran dan emosi pemeluk agama tersebut hal ini lah yang pada akhirnya dapat memunculkan tindakan radikal apabila penganut salah satu agama tersebut merasa terusik nilai-nilai sakralnya [4]. Lebih lanjut, Fanindy melalui penelitiannya terkait Pergeseran Literasi pada Generasi Milenial Akibat Penyebaran Radikalisme di Media Sosial menunjukkan bahwa adanya media social menjadi salah satu faktor terkuat yang melatarbelakangi munculnya paham radikalisme pada generasi social akibat adanya pergeseran literasi generasi milenial berkaitan dengan ilmu keagamaan, pendidikan, dan hukum agama [5].

Mengingat semakin mudahnya penyebaran radikalisme maka, melalui penelitian ini, peneliti berusaha untuk mengidentifikasi tingkatan radikalisme yang ada di Indonesia dengan pengelompokan radikalisme kedalam 3 kelompok yaitu radikal tingkat rendah, tingkat menengah, tingkat tinggi. Berkaitan dengan metode dalam melakukan pengklasifikasian

radikalisme, Fanindy dalam penelitiannya melihat bahwa penyebaran khususnya melalui media social sebagaimana hasil penelitiannya, cenderung lebih dimungkinkan untuk dianalisis melalui analisis sentimen dengan metode *Naïve Bayes*. Hal ini didukung oleh penelitian terdahulu milik Sundara dkk, bahwa penggunaan metode *Naïve Bayes* berhasil membangun sebuah model klasifikasi tweet dengan akurasi 86% dibandingkan dengan metode yang lainnya [6]. Penelitian terdahulu lainnya juga menunjukkan bahwa penggunaan metode *Naïve Bayes* memberikan kemudahan dalam pemilahan informasi yang didapat sehingga keputusan dapat dilaksanakan dengan cepat dan tepat [7]. Metode *Naïve Bayes* juga banyak diimplementasikan di aplikasi lain [8]–[12]. *Sentiment analysis* adalah area yang berkembang dari Pemrosesan Bahasa Alami dengan penelitian mulai dari klasifikasi tingkat dokumen hingga mempelajari polaritas kata dan frasa [13]–[16]. Data penelitian sebagai dasar pengelompokan sendiri didapatkan melalui kuesioner online yang telah disebar oleh peneliti.

Sehingga dengan beberapa referensi penelitian terdahulu yang telah disebutkan tersebut, penggunaan metode *Naïve Bayes* pada penelitian diharapkan mampu memberikan hasil pengelompokan radikalisme. Proses penelitian ini direncanakan dengan diawali oleh pra-proses dataset, kemudian diekstraksi dari kata sifat dari dataset yang memiliki beberapa makna yang disebut fitur vektor, kemudian memilih daftar vektor fitur dan selanjutnya diterapkan algoritma klasifikasi berbasis machine learning yaitu: Naive Bayes, Entropi maksimum dan SVM bersama dengan Semantik Orientasi berbasis WordNet yang

mengekstrak sinonim dan kesamaan untuk fitur konten. Akhirnya mengukur *performance of classifier in terms of recall, precision and accuracy* dengan mendapatkan hasil akurasi 89.9%. Hasil data yang telah dikumpulkan, kemudian akan diklasifikasikan berdasarkan jenis-jenis atau kelompok kata kunci yang sudah dikelompokkan juga. Pada penelitian ini akan mengelompokkan radikal ke dalam 3 kelompok yaitu radikal tingkat rendah, tingkat menengah, tingkat tinggi.

Penelitian ini berbeda dengan penelitian-penelitian terdahulu dimana penelitian ini menggunakan data yang diperoleh dengan mengumpulkan data menggunakan *google form*. Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi terhadap proses identifikasi radikalisme di Indonesia dan dapat menyempurnakan penelitian-penelitian terdahulu tentang radikalisme.

2. Tinjauan Pustaka

Penelitian yang sudah dilakukan penulis bertujuan untuk menguji hipotesis mereka bahwa radikal Islamis memiliki informasi yang dapat diidentifikasi dan sifat-sifat perilaku

yang dapat digunakan untuk mengidentifikasi motif ideologis mereka secara unik di antara radikal lainnya. Untuk memvalidasi hipotesis ini, dataset PIRUS telah digunakan untuk membangun model klasifikasi untuk radikal Islam di AS. Kinerja empat algoritma *supervised machine learning* yang berbeda (yaitu *deep learning, distributed random forest, gradient boosting* dan *naïve bayes*) dan telah dievaluasi menggunakan platform analisis data besar yang mutakhir, seperti pada Tabel 1. Dua metrik evaluasi utama telah digunakan, yaitu metrik kesalahan generalisasi (yaitu akurasi, *F1-score, precision* dan *recall*) dan waktu pelatihan model. Apalagi efeknya mempertimbangkan masalah klasifikasi sebagai masalah binomial, yaitu, untuk mengklasifikasikan seseorang sebagai radikal Islam atau bukan versus masalah multikelas, yaitu, untuk membedakan radikal Islam di antara radikal lain yang telah dipelajari. Selain itu, efek preprocessing dataset, khususnya, mengubah atribut *datatype* dari numerik ke kategorikal dan sebaliknya telah dibandingkan.

Tabel 1. Hasil evaluasi rata-rata

Model	Class	Attribute Datatype	Accuracy	F1	Precision	Recall	Time (ms)	Time (Scaled)
DL	Binary	Categorical	0.97	0.98	0.98	0.98	27131	0.812999
		Numerical	0.94	0.96	0.95	0.98	20344	0.609642
	Multi	Categorical	0.72	0.83	0.78	0.88	33371	1
		Numerical	0.58	0.72	0.69	0.75	12839	0.384745
DRF	Binary	Categorical	0.97	0.98	0.97	0.99	4008	0.120114
		Numerical	0.96	0.97	0.96	0.99	3494	0.104693
	Multi	Categorical	0.70	0.86	0.92	0.80	2277	0.068227
		Numerical	0.63	0.83	0.92	0.76	1761	0.052765
GBM	Binary	Categorical	0.97	0.98	0.97	0.98	3281	0.098319
		Numerical	0.97	0.98	0.98	0.99	3394	0.101714
	Multi	Categorical	0.74	0.85	0.88	0.82	4269	0.127935
		Numerical	0.75	0.89	0.91	0.86	3428	0.102736
NB	Binary	Categorical	0.95	0.97	0.96	0.99	51	0.001516
		Numerical	0.90	0.94	0.92	0.96	174	0.00522
	Multi	Categorical	0.70	0.80	0.81	0.80	28	0.000839
		Numerical	0.51	0.4	0.70	0.31	53	0.001576

Adapun penelitian tentang *sentiment analysis* dalam mendeteksi tentang *online review* yang telah dilakukan, dijelaskan tentang bagaimana internet telah berkembang pesat dan mengubah cara pandang seseorang saat ini semua bisa terselesaikan lewat *posting blog* yang berbeda, forum diskusi online, *review* produk di situs web, dll, seperti pada Tabel 2.

Tabel 2. Probabilitas Untuk Setiap Tweet Baik Positif Atau Negatif

Tweets	Positi ve	Negati ve
i think i lost my spanish book!!	0.7	0.1
life can be so complicated. i know god has a plan i just have to trust in him	0.6	0.2
i don't know what to do..	0.8	0.1
i did not get cast in but i will keep looking for some dance opportunities. i need some exercise.	0.7	0.1
youaintgotta apologize. i know i'm not ugly	0.5	0.1

Orang-orang biasanya bergantung pada konten buatan pengguna pada produk apa pun datang untuk melakukan tindakan yang diinginkan. Ketika orang ingin membeli produk melalui online, pertama-tama mereka akan mencari ulasan dalam situs web produk tertentu melalui online, sebelum membuat keputusan. Beberapa analisis harus dilakukan pada semua ulasan ini sehingga hasil akhir mengatakan apakah produk bagus untuk dibeli atau tidak. Ada sentimen berbeda teknik analisis yang tersedia dengan banyak aplikasi untuk berbagai domain, seperti dalam bisnis untuk mendapatkan

umpan balik untuk produk dari pelanggan. Pengetahuan dan teknik *machine learning* adalah dua teknik terutama yang digunakan untuk *sentiment analysis*. Sebenarnya, ada berbagai teknik *machine learning* yang digunakan untuk mengklasifikasikan data yaitu, *Naïve Bayes classifier*, *support vector machine*, *decision tree*, *random forest*, *neural networks* dll.

Adapun juga penelitian tentang *sentiment analysis* menggunakan *naïve bayes* yang telah ditingkatkan yang dilakukan oleh *Naïve Bayes* adalah model probabilistik yang sangat sederhana yang cenderung bekerja dengan baik pada teks klasifikasi dan biasanya membutuhkan waktu lebih sedikit untuk melatih jika dibandingkan dengan model seperti mesin vektor dukungan. Kami akan menunjukkan dalam makalah ini bahwa tinggi tingkat akurasi dapat diperoleh dengan menggunakan model *Naïve Bayes*, yang sebanding dengan model terkini dalam klasifikasi sentiment, seperti pada Tabel 3.

Tabel 3. Hasil Uji Coba Dengan Naïve Bayes

Feature Added	Accuracy on test set
Original Naive Bayes algorithm with Laplacian Smoothing	73.77%
Handling negations	82.80%
Bernoulli Naive Bayes	83.66%
Bigrams and trigrams	85.20%
Feature Selection	88.80%

Jumlah fitur yang optimal dipilih dengan menggunakan set validasi 1000 dokumen, plot akurasi v / s. Kemudian akurasi diukur pada seluruh set tes dari 25000 dokumen. Waktu untuk fitur Seleksi adalah sekitar 3 menit.

3. Metode

3.1. Radikalisme

Radikalisme secara umum dipahami sebagai suatu gerakan sosial yang mengarah pada hal-hal yang negatif. Setidaknya persepsi itu yang dikonsepsikan oleh Lukman Hakim, Wakil Kepala LIPI, dalam pengantar buku *Islam dan Radikalisme di Indonesia*. Dari persepsi seperti itu, maka muncul istilah ekstrem, anti barat, anti Amerika, dan teroris. Dari segi bahasa, sebenarnya istilah radikal jauh berbeda dengan teroris. Karena radikal itu sendiri merupakan sebuah proses secara sungguh sungguh untuk melatih keberhasilan atau cita-cita yang dilakukan dengan cara-cara yang positif. Sementara itu, terorisme berasal dari kata teror yang bermakna menakut-nakuti pihak lain. Oleh sebab itu, teror selalu dilakukan dengan cara-cara negatif dan menakutkan pihak lain. Seiring berkembangnya jaman dan semakin banyaknya fenomena-fenomena antara radikal dan terorisme yang berkembang di Indonesia banyak masyarakat yang beranggapan bahwa radikal merupakan teroris yang sedang dididik. Terutama melewati jalur-jalur keagamaan menggunakan aturan-aturan dari agama yang dianut oleh korbannya seolah sedang di cuci otak menggunakan aturan dari agama itu sendiri, jika si korban hanya berfokus ke agamanya saja bisa saja korban akan jatuh kedalam lingkaran terorisme yang tidak berhenti berputar sampai saat ini. Dari penyebaran yang dilakukan melewati agama, paham-paham radikal ini juga semakin mudah menyebar melewati instansi-instansi pendidikan, keluarga, teman sebaya, tetangga dan akan terus menyebar jika orang yang menyebarkan ini memiliki niat yang negatif maka paham radikal ini bisa diubah menjadi paham terorisme. Pada penelitian ini akan membagi jenis

radikalisme menjadi 3 kelompok yaitu radikalisme tingkat rendah, radikalisme tingkat menengah, radikalisme tingkat tinggi.

3.2 *Machine Learning*

Mengeksploitasi teknologi seperti kecerdasan buatan, *machine learning* dan *data mining* guna untuk memerangi terorisme, radikalisme dan ekstremisme kekerasan terutama di media sosial telah menarik perhatian peneliti dalam tujuh belas tahun terakhir, oleh karena itu pada penelitian ini juga akan memanfaatkan teknologi *machine learning* atau sebuah mesin yang menggunakan sebuah model yang dapat untuk memprediksi tentang data masa depan dan dapat memberikan sebuah prediksi atau keputusan yang rasional. Ketidakpastian dalam menentukan keputusan atau prediksi memiliki peran yang mendasar dalam *machine learning*. Data yang diamati bisa konsisten dengan banyaknya model *machine learning* yang tersedia, dan oleh karena itu model mana yang tepat yang akan digunakan, mengingat data tersebut tidak pasti. Begitu juga prediksi, soal data masa depan dan konsekuensi tindakan di masa depan, juga tidak pasti. maka dari itu teori probabilitas merupakan dasar dari *machine learning* karena menyediakan kerangka kerja untuk mencari sesuatu yang tidak pasti. *Machine learning* berupaya untuk terus mengembangkan metode untuk meningkatkan kinerja komputer pada pekerjaan tertentu berdasarkan data yang diamati. Contoh umum dari pekerjaan tersebut misalnya mendeteksi pejalan kaki dalam gambar yang diambil dari kendaraan otonom, mengklasifikasikan ekspresi gen pola dari pasien leukemia menjadi subtype dengan hasil klinis, atau menerjemahkan kalimat bahasa Inggris ke dalam bahasa Indonesia. Namun,

seperti yang akan kita lihat, ruang lingkup tugas dari machine learning bahkan lebih luas daripada klasifikasi pola atau tugas pemetaan ini, dan dapat mencakup optimasi dan pengambilan keputusan, mengompresi data, dan secara otomatis mengekstraksi model yang dapat ditafsirkan dari data. Oleh karena penulis akan menggunakan salah satu metode dari machine learning untuk mengklasifikasikan pola radikal menggunakan Sentiment Analysis atau biasa disebut opinion mining.

3.3. Naïve Bayes Classifier

Klasifikasi *Naïve Bayes* adalah model probabilistik sederhana berdasarkan aturan bayes bersama dengan asumsi independensi yang kuat. Model *naïve bayes* melibatkan asumsi independensi bersyarat yang disederhanakan. Itu diberikan kelas (positif atau negatif), kata-kata secara kondisional independen satu sama lain. Asumsi ini tidak mempengaruhi akurasi dalam klasifikasi teks tetapi membuat algoritma klasifikasi ini berjalan sangat cepat berlaku untuk masalah ini. probabilitas kemungkinan maksimum dari kata yang termasuk dalam kelas tertentu diberikan oleh ungkapan:

$$P(x_i/c) = \frac{\text{Count } x_i \text{ in document of class } c}{\text{Total no of words in documents of class } c} \quad (1)$$

Hitungan frekuensi dari kata-kata disimpan dalam tabel hash selama pelatihan tahap. Berdasarkan *Bayes Rule*, probabilitas dokumen tertentu milik sebuah kelas c_i diberikan oleh [16]

$$P(d) = \frac{P(c_i) * P(c_i)}{P(d)} \quad (2)$$

Jika kita menggunakan asumsi independensi bersyarat penyederhanaan, dan diberikan kelas (positif atau negatif), kata-kata secara kondisional independen satu sama lain. Disebabkan oleh asumsi penyederhanaan model ini disebut sebagai "naive" [16].

$$P(d) = \frac{(\prod P(c_j)) * P(c_j)}{P(d)} \quad (3)$$

Di sini x_i adalah kata-kata individual dari dokumen tersebut. Classifier menampilkan kelas dengan probabilitas posterior maksimum. Kami juga menghapus kata-kata duplikat dari dokumen, mereka tidak menambahkan tambahan informasi; jenis algoritma naïve bayes ini disebut Bernoulli Naïve Bayes. Termasuk hanya kehadiran kata dan bukannya hitungan telah ditemukan memperbaiki kinerja sedikit, ketika ada sejumlah besar contoh pelatihan.

3.4. Word Embedding

Word embedding adalah proses mengubah sebuah teks menjadi angka atau vektor. Sebagian besar arsitektur deep learning dan machine learning membutuhkan angka sebagai input. Hal ini disebabkan karena keduanya tidak dapat melakukan proses analisis pada input data berupa string atau teks. Word embedding memiliki beberapa jenis, seperti Term-Frequency and Inverse Document Frequency (TF-IDF), Word2Vec, Glove, CountVectorizer dan masih banyak lagi.

Pada penelitian ini ada menggunakan *CountVectorizer* sebagai proses dari *word embedding*. Untuk mengetahui sebuah vektor bisa mewakili makna dari sebuah kata adalah dengan mengukur beberapa vektor sebagai perbandingan. Seperti

misalnya mengukur jarak antara vektor dari kata “Japan” dengan “Jepang” atau vektor kata “Jakarta” dengan “Indonesia”, maka akan ditemukan bahwa jarak kedua kata tersebut akan muncul pada angka yang berdekatan. Hal ini disebabkan karena makna kedua kata berdekatan, yakni nama negara dan ibu kotanya. Sehingga nilai vektor yang dihasilkan akan berdekatan juga. Beberapa contoh lain akan ditampilkan pada Tabel 4.

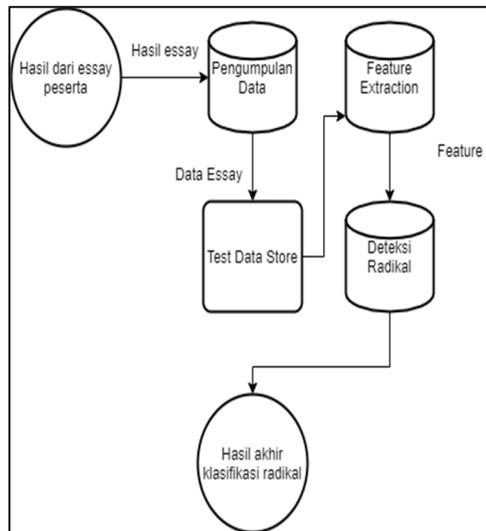
Tabel 4. Contoh Kata Dengan Vektor Berdekatan

Relationship	Example 1	Example 2	Example 3
France - Paris	England: London	Indonesia: Jakarta	Philippine: Manilla
big – bigger	small: larger	cold: colder	quick: quicker
AS – Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Tesla – scientist	Messi: Barca	Mozart: violinist	Vincentso: Casano
Sarkozy – France	Espresso: Italy	Monstadt: Germany	Inazuma: Japan
Iron – Fe	zinc – Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Elon: Musk
Microsoft – Edge	Google: Tensorflow	IBM: Linux	Apple: Iphone
Microsoft – Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Jihad - Agama	Radikal: Teroris	Kilafa: Jihad	Islam: Jihad

3.5. *Sentiment Analysis*

Sentiment analysis atau *opinion mining* adalah studi komputasi atas opini, penilaian, sikap, dan emosi orang terhadap entitas, individu, masalah, peristiwa, topik, dan atributnya. Tugas ini secara teknis menantang dan praktis sangat berguna. Misalnya, pembisnis selalu ingin mengerti opini publik atau konsumen tentang produk atau layanan mereka. Pembeli sendiri juga ingin mengetahui pendapat pengguna yang ada sebelum mereka menggunakan layanan atau membeli produk.

Dengan melesatnya pertumbuhan media social seperti situs *reviews*, forum diskusi, blog, dan jejaring sosial di *web*, individu dan organisasi semakin menggunakan opini publik di media ini untuk pengambilan keputusan mereka. Namun, menemukan dan meninjau sebuah opini di sebuah *web* dan menyaring informasi yang terkandung di dalamnya tetap menjadi tugas yang berat karena perkembangan opini yang beragam. Setiap postingan dalam situs biasanya berisi sejumlah besar tulisan yang berisi pendapat yang tidak selalu mudah diuraikan. Rata-rata pembaca akan mengalami kesulitan dalam mengidentifikasi postingan yang relevan dan akurat, pada umumnya pembaca akan merangkum informasi dan pendapat yang terkandung di dalamnya untuk mencari hasil yang mereka cari. Dengan menggunakan *sentiment analysis* diharapkan dapat mempermudah dalam merangkum dan mendapatkan hasil yang kita inginkan, misalnya hendak akan membeli barang atau sebuah layanan pada tempat tertentu. Pada penelitian ini *sentiment analysis* akan dibagi menjadi 4 tahap dalam menyelesaikan tugasnya. Tahap pertama digunakan untuk mengumpulkan opini opini yang sudah ada sebelumnya, kemudian di tahap ke 2 digunakan untuk membersihkan opini opini yang sudah ada sehingga membentuk sebuah kata kunci, setelah itu ditahap ke 3 setelah berbagai kata kunci didapatkan maka kata kunci tersebut akan dikelompokkan sesuai kelasnya masing-masing, dan di tahap ke 4 digunakan untuk tahap *testing* dari opini yang masuk dan dicocokkan dengan kata kunci yang ada. Untuk lebih memudahkan dalam memahami alur dari proses *sentiment analysis* akan di jelaskan pada diagram Gambar 3.



Gambar 3. Diagram Proses Sentimen Analisis

Pada diagram yang disajikan pada Gambar 3 mulanya peserta akan diminta untuk mengisi sebuah essay berupa pertanyaan-pertanyaan yang berhubungan dengan radikal, setelah peserta mengisi essay tersebut data dari peserta akan ditampung terlebih dahulu, setelah itu akan dilanjutkan pada proses testing dimana data akan mulai diekstrak pada tahap feature extraction menggunakan sebuah kamus atau kumpulan kata kunci yang sudah disiapkan sebelumnya, untuk mendapatkan kata kunci atau dipenelitian ini yaitu menggunakan kumpulan text tentang sudut pandang radikal dari berbagai sudut pandang.

Untuk itu sebelum diolah data akan dilakukan tahap preprocessing seperti berikut:

- **Case Folding**
Yaitu proses penyamaan case dalam sebuah dokumen. Tidak semua dokumen konsisten dalam penggunaan huruf kapital. Oleh karena itu peran case folding dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi bentuk standar (biasanya huruf kecil).

- **Filtering**

Yaitu proses mengambil kata – kata penting dari hasil token. Algoritma yang dapat digunakan dalam filtering yaitu *stopword* dan *wordlist*. *stopword* adalah membuang kata – kata kurang penting atau tidak deskriptif, seperti “yang”, ”dan”, ”di”, ”saya”, ”kamu”, dan lain seterusnya. Sedangkan *wordlist* adalah menyimpan kata – kata penting.

Setelah tahap *preprocessing* selesai dilakukan, maka akan berlanjut ke tahap *feature extraction* untuk menyimpan data-data *training* yang sudah diolah tadi menjadi sebuah kamus baru menggunakan metode *countvectorizer*. Setelah proses *feature extraction* selesai tahap selanjutnya yaitu melakukan tahap *training data* dan hasil dari proses ini baru bisa didapatkan kesimpulan jawaban essay dari responden ini termasuk dalam kelompok radikal tingkat rendah, radikal tingkat menengah atau radikal tingkat tinggi. Untuk lebih detilnya tentang relasi-relasi yang akan ada dalam sistem penulis akan menjelaskan dalam bentuk ERD atau Entity Relationship Diagram pada Gambar 4.

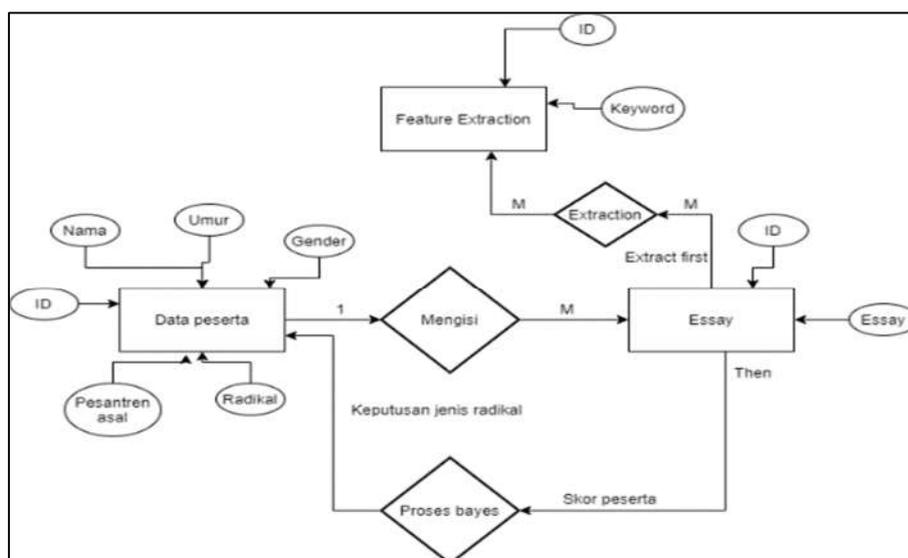
Fungsi dari ERD sendiri dari penggambaran Entity Relationship Diagram (ERD) adalah untuk membuat kita lebih mudah untuk menganalisis pada suatu basis data atau suatu system dengan cara yang cepat dan murah, Fungsi berikutnya dari penggambaran Entity Relationship Diagram (ERD) adalah penulis dapat menguji model yang telah dibuat dan juga bisa megabaikan proses apa yang telah dilakukan hanya dengan menggambar Entity Relationship Diagram (ERD), Fungsi selanjutnya dari penggambaran Entity Relationship Diagram (ERD)

adalah untuk menjelaskan hubungan - hubungan antar data - data dalam basis data berdasarkan objek - objek dasar data yang mempunyai hubungan yang dihubungkan oleh suatu relasi, Fungsi terakhir dari Entity Relationship Diagram (ERD) adalah untuk mendokumentasikan data - data yang ada dengan cara mengidentifikasi setiap entitas dari data - data dan hubungannya pada suatu Entity Relationship Diagram (ERD) itu sendiri. Pada Gambar 4 dijelaskan tentang entity relationship yang akan ada pada sistem, sistem nantinya akan memiliki 3 macam table data peserta, essay, dan feature extraction setiap table akan memiliki entitasnya masing-masing. Untuk relasi tiap tablenya peserta akan mengisi essay relasi yang terjadi disini adalah one to many setelah essay terisi, tahap selanjutnya essay yang sudah terisi akan dilakukan tahap feature extraction relasi yang akan terjadi disini adalah many to many setelah semua essay sudah diisi selanjutnya adalah hasil skor essay akan diproses oleh naïve bayes dan hasil akhir dari perhitungan tersebut

akan dimasukkan kedalam table data peserta lebih tepatnya di entitas radikal.

Pada penelitian ini, peneliti menggunakan essay sebagai bahan ujinya, essay yang akan digunakan disini merupakan hasil jawaban dari beberapa pertanyaan yang telah diisi oleh santri dan beberapa pemuka agama di beberapa pondok pesantren, setelah data terkumpul. Selanjutnya akan dilakukan pelabelan oleh *expert judgement* untuk menentukan kelompok radikal dari setiap responden yang akan digunakan sebagai data training, seperti pada Tabel 5.

Setelah data tersebut selesai dilabeli selanjutnya dilakukan tahap *preprocessing* untuk mendapatkan beberapa kumpulan teks tentang radikal dari berbagai sudut pandang, tahap preprocessing nya meliputi *case folding*, *filtering*, dan *stemming*. Setelah tahap preprocessing selesai tahap selanjutnya yaitu tahap feature extraction pada tahap ini akan menyimpan data-data *training* yang sudah diolah tadi menjadi sebuah kamus baru menggunakan metode *countvectorizer*.

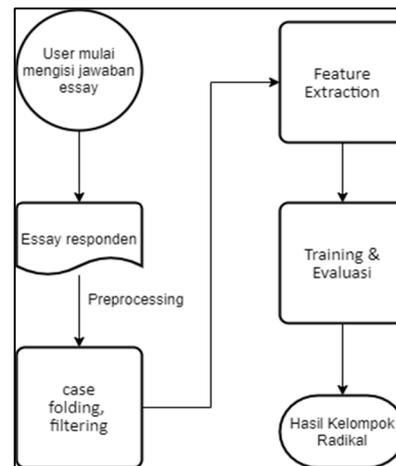


Gambar 4. Entity Relationship Diagram

Tabel 5. Contoh Jawaban Dari Essay Responden

NO	Pertanyaan 1	Pertanyaan 2	Pertanyaan 3	Pertanyaan 4	Kelompok Radikal
1	Jihad suatu pengorbanan yang dilandasi agama	Tidak semua tapi ada oknum tertentu	Karena adanya suatu aliran yg sesat	Tidak semua tapi ada oknum yang menyalahgunakan cadar	Rendah
2	jihad menurut saya adalah cara membela islam saat berperang dengan orang orang yang ingin menjathkan islam	menurut saya jihad dan radikal adalah hal yang berbeda karena jihad hanya untuk islam dan radikal untuk semua agama	radikal merupakan sebuah cara untuk tidak menghargai manusia	iya karena radikal menyebabkan aura negatif masuk kedalam diri	Menengah
3	pengecahan yang dilakukan yaitu dengan membuat prespektif masyarakat tentang jihad adalah sebuah kedongkolan seseorang terhadap agama	menyampaikan bahwa beragama dan bersosial yang baik dapat membuat hidup lebih indah untuk diajalani	yang berhak sesorang itu adalah kafir hanyalah tuhan dan manusia sebagai makhluk hanya bisa meyakinkan orang lain tentang tuhan	jihad yang penuh perjuangan memerangi godaan duniawi yang mengahsut	Menengah
4	Menurut saya hal tersebut tidak benar karena radikalisme bukanlah tujuan dari jihad yang sebenarnya	Radikalisme adalah sebuah ideologi yang menyimpang dari ajaran agama islam	Pembelajaran jihad didalam ponodok dilakukan melalui pembelajaran akhlak	Saya akan memberikan mereka teguran dengan cara yang baik	Rendah
5	Radikalime jalan tercepat dalam mendapatkan kita ingin	Mengetahui agama orang lain itu penting agar kita juga paham	Saya setuju karena tujuan dasar dai agama islam baik	Kita semua memang sama, tetapi soal keyakinan kita semua beda	Tinggi

Setelah proses *feature extraction* selesai tahap selanjutnya yaitu melakukan tahap *training data* dan hasil dari proses ini baru bisa didapatkan kesimpulan jawaban essay dari responden ini termasuk dalam kelompok radikal tingkat rendah, radikal tingkat menengah atau radikal tingkat tinggi. Setelah tahap training selesai maka akan dilanjutkan ke tahap evaluasi model yang telah ditraining menggunakan metode *classification report*. Gambaran lengkapnya dapat dilihat pada Gambar 5.



Gambar 5. Gambaran metode penelitian secara garis besar

4. Hasil dan Pembahasan

Metode *Sentiment Analysis* akan penulis implementasikan kedalam sebuah sistem dengan nama “Sistem Analisa Paham Radikalisme” diharapkan sistem ini nantinya akan dapat membantu dalam menganalisa penyebaran paham radikalisme dalam sebuah pesantren yang ada di Indonesia. Dalam melakukan uji coba pada penelitian ini pertama-tama

dataset yang sudah terkumpul akan dibagi menjadi 2 bagian data 165 untuk data *training* dan 85 untuk *testing* atau bisa, sebelum dataset digunakan untuk melakukan uji coba pada model. Setelah data didapatkan kemudian dilabeli oleh *expert judgement* sebelumnya, kemudian dilakukan tahap preprocessing.

Tabel 6. Hasil preprocessing pada data yang akan ditraining

NO	Pertanyaan 1	Pertanyaan 2	Pertanyaan 3	Pertanyaan 4	Kelompok Radikal
1	jihad pengorbanan dilandasi agama	oknum	aliran yg sesat	oknum menyalahgunakan cadar	Rendah
2	jihad membela islam berperang orang orang menjathkan islam	jihad radikal berbeda jihad islam radikal agama	radikal menghargai manusia	iya karena radikal menyebabkan aura negatif masuk kedalam	Menengah
3	pengecahan prespektif masyarakat jihad kedongkolan agama	beragama bersosial hidup indah diajalani	berhak seseorang kafir tuhan manusia makhluk orang tuhan	jihad penuh perjuangan memerangi godaan duniawi mengahsut	Menengah
4	radikalisme tujuan jihad	radikalisme ideologi menyimpang ajaran agama islam	pembelajaran jihad didlam ponodok pembelajaran akhlak	teguran	Rendah
5	radikalime jalan tercepat ubah	agama orang paham keyakinan anut	setuju tujuan dasar dai agama islam kebaikan	keyakinan perbedaan	Tinggi

Pada tahap ini diterapkan metode seperti *case folding* dan *filtering* yang menghasilkan table seperti pada Tabel 6. Pada tabel ini dijelaskan beberapa kata sambung seperti “sesuatu”, “semua”, “karena” akan dieliminasi menjadikan beberapa kata kunci saja pada setiap jawaban. Setelah tahap ini selesai selanjutnya adalah proses *feature extraction* ini berfungsi sebagai pembuatan kamus baru untuk data yang akan kita ujikan.

['agama', 'agamanya', 'ajaran', 'aksi', 'aktivitas', 'al', 'allah', 'alquran', 'apapun', 'arah', 'arti', 'artian', 'bebas', 'beikhtiyar', 'bela', 'benda', 'bentuk', 'ber', 'berdiri', 'berikhtiar', 'berjihad', 'berjuang', 'berperang', 'bersungguh', 'berusaha', 'bumi', 'caraberjuang', 'dasar', 'daya', 'diajarkan', 'dibela', 'didalam', 'dijalan', 'dikaitkan', 'dilaksanakan', 'dilandasi', 'dimaknai', 'diyakini', 'ejekan', 'fisik', 'fundamental', 'golongan', 'harta', 'hidup', 'hinaan', 'idealisme', 'ikhtiar',

Gambar 6. Hasil tahap *feature extraction*

Pada Gambar 6 dijelaskan bahwa setelah proses *preprocessing* selesai akan dihasilkan sebuah kamus baru yang disini berbentuk kumpulan dari banyak kata, kamus inilah yang nantinya akan digunakan sebagai acuan pada proses *training data*.

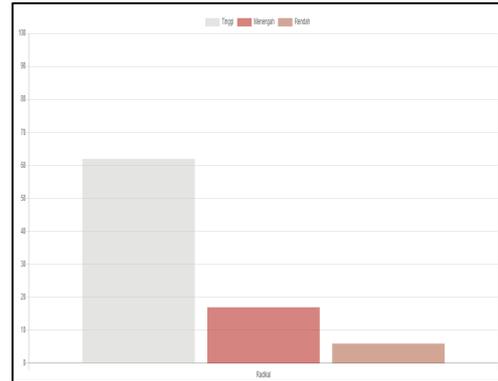
Pada tahap uji coba, uji coba dimulai dengan melakukan proses *training naive bayes*, pertama kita akan membagi data menjadi 2 bagian 165 data responden untuk proses *training* dan 85 data responden untuk tahap *testing*, kemudian hasil dari *training* tadi akan kita evaluasi menggunakan metode *classification report* dengan hasil pada Tabel 7.

Tabel 7. Hasil Classification Report

	Precision	Recall	F1-Score	Support
Rendah	0.93	0.78	0.85	32
Menengah	0.86	0.86	0.86	28
Tinggi	0.77	0.99	0.84	25
Accuracy			0.85	85
Macro avg	0.85	0.85	0.85	85
Weighted avg	0.85	0.85	0.85	85

Dari hasil evaluasi tersebut kita bisa melihat hasil presisi untuk kelompok rendah sebesar 93% sedangkan untuk kelompok menengah sebesar 86% dan untuk kelompok tinggi sebesar 77% dan untuk total akurasi yang didapat dari ke 3 kelompok sebesar 85%. Setelah model terbentuk dari proses *training* dan evaluasi selanjutnya model akan dipasang ke dalam aplikasi yang berbasis *web*.

Hasil yang diperoleh dari perhitungan sistem didapat rendah mendapat total 6, menengah mendapat total 17, dan tinggi dengan total 62, seperti pada Gambar 7.



Gambar 7. Hasil Olah Data Langsung Dari Aplikasi Web

5. Penutup

Setelah dilakukan tahap uji coba dan evaluasi analisa sentimen menggunakan *naive bayes* dapat diperoleh akurasi yang cukup bagus dengan total 85%. Adapun beberapa temuan dalam tahap uji coba seperti beberapa kata yang disingkat-singkat yang akhirnya mempengaruhi sistem dalam melakukan perhitungan yang akan membuat nilai akurasi menurun, kedepannya bisa dikembangkan lagi pada tahap *preprocessing* yaitu melakukan seleksi atau penambahan kamus untuk kata yang disingkat-singkat sehingga menjadi 1 kata yang sesuai.

6. Daftar Pustaka

- [1] T. A. Sundara, S. Ekaputri, and S. Sotar, "Naïve Bayes Classifier untuk Analisis Sentimen Isu Radikalisme," in *Prosiding SISFOTEK*, 2020, pp. 93–98.
- [2] M. Ibrahim, E. Bu'ulolo, and I. Lubis, "Penerapan Algoritma Naive Bayes Classifier Untuk Mendeteksi Tingkat Kredibilitas Hoax News/ Fake News Pada Sosial Media Di Indonesia Berbasis Android (Studi Kasus : Kantor Tribun Medan)," *RESOLUSI : REKAYASA TEKNIK INFORMATIKA DAN*

- INFORMASI*, vol. 1, no. 1, 2020.
- [3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning technique," in *Empirical methods in natural language processing*, 2002, pp. 79–86.
- [4] M. Al-Zewairi and G. Naymat, "Spotting the Islamist Radical within: Religious Extremists Profiling in the United State," in *Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops*, 2017, pp. 162–169.
- [5] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," *Intelligent Data Engineering and Automated Learning IDEAL*, vol. 8206, pp. 194–201, 2013.
- [6] K. Suppala and N. Rao, "Sentiment Analysis Using Naïve Bayes Classifier," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 8, 2019.
- [7] M. S. Muchith, "Radikalisme Dalam Dunia Pendidikan," *ADDIN Media Dialektika Ilmu Islam*, vol. 10, no. 1, 2016.
- [8] A. Hermanto, "Implementasi Text Mining Menggunakan Naive Bayes Untuk Penentuan Kategori Tugas Akhir Mahasiswa Berdasarkan Abstraksinya," *KONVERGENSI*, vol. 12, no. 2, pp. 1–10, 2016.
- [9] J. E. Manurung and E. T. Putri, "Penentuan Minat Bakat Menggunakan Metode Bayes Berbasiss Web," *KONVERGENSI*, vol. 16, no. 2, pp. 80–89, 2020.
- [10] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 2, p. 91, 2019.
- [11] A. W. Syaputri, E. Irwandi, and M. Mustakim, "Naïve Bayes Algorithm for Classification of Student Major's Specialization," *Journal of Intelligent Computing & Health Informatics*, vol. 1, no. 1, p. 17, 2020.
- [12] A. H. Mirza, "Application of Naive Bayes Classifier Algorithm in Determining New Student Admission Promotion Strategies," *Journal of Information Systems and Informatics*, vol. 1, no. 1, pp. 14–28, 2019.
- [13] A. L. Maas, R. Daly, P. T. Pham, D. M. Huang, A. Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," in *Conference: The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.
- [14] N. Normah, "Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews," *SinkrOn*, vol. 3, no. 2, p. 13, 2019.
- [15] S. Khomsah, "Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews," *Jurnal Penelitian Pos dan Informatika*, vol. 10, no. 2, p. 157, 2020.
- [16] L. F. Narulitas, "Analisa Sentimen Pada Tinjauan Buku Dengan Algoritma K-Nearest Neighbour," *KONVERGENSI*, vol. 13, no. 2, pp. 76–81, 2017.