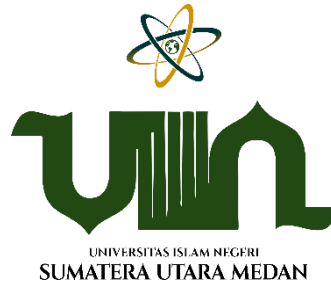


**LAPORAN PENELITIAN KOLABORATIF FAKULTAS
PROGRAM STUDI ILMU KOMPUTER**



**ANALISIS SENTIMEN TERHADAP ISU ISLAMOPHOBIA
MENGUNAKAN ALGORITMA SVM**

Disusun :

Oleh :

AIDIL HALIM LUBIS, M.Kom

ALI DARTA, MA

SILVA UKTHI FILLA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUMATERA UTARA

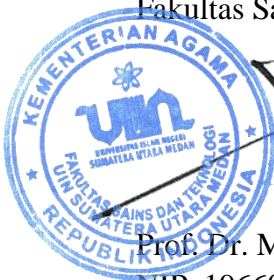
MEDAN

2022

LEMBAR PENGESAHAN

1. a. Judul Penelitian : **Analisis Sentimen Terhadap Isu Islamophobia Menggunakan Algoritma SVM**
- b. Kluster Penelitian : Penelitian Program Studi
- c. Bidang Keilmuan : Ilmu Komputer
- d. Kategori : Kelompok
2. Peneliti 1 : Aidil Halim Lubis, M.Kom (Ketua)
NIP Peneliti 1 : 198805272019031010
3. Peneliti 2 : Ali Darta, M.A (Anggota)
NIP Peneliti 2 : 198310102019031007
5. Unit Kerja : Fakultas Sains dan Teknologi – UIN SU
6. Waktu Penelitian : Medan
7. Waktu Penelitian : 3 bulan, tahun 2021
7. Biaya Penelitian : Rp. 20.028.200,-
(Dua Puluh Juta Dua Puluh Delapan Rupiah)

Disahkan oleh
Dekan
Fakultas Sains dan Teknologi



Prof. Dr. MHD. Syahnan, M.A
NIP. 196608101999031002

Medan, 05 September 2022
Peneliti,

Aidil Halim Lubis, M.Kom
NIP. 198805272019031010

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Aidil Halim Lubis, M.Kom
Jabatan : Dosen / Asisten Ahli / Gol. III.b
Unit Kerja : Fakultas Sains dan Teknologi
 : UIN Sumatera Utara Medan
Alamat : Jalan Lapangan Golf Desa Durian Jangak

Dengan ini menyatakan bahwa :

1. Judul penelitian “Analisis Sentimen Terhadap Isu Islamophobia Menggunakan Algoritma SVM” merupakan karya orisinal saya.
2. Jika dikemudian hari ditemukan fakta bahwa judul, hasil atau bagian dari laporan penelitian saya merupakan karya orang lain dan/atau plagiasi, maka saya akan bertanggung jawab untuk mengembalikan 100% dana hibah penelitian yang telah saya terima dan siap mendapatkan sanksi sesuai ketentuan yang berlaku.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Medan, 05 September 2022

Yang Menyatakan,



Aidil Halim Lubis, M.Kom
NIP. 198805272019031010

ANALISIS SENTIMENT TERHADAP ISU ISLAMOPHOBIA MENGUNAKAN ALGORITMA SVM

ABSTRAK

Analisis sentimen adalah proses untuk memahami dan mengelompokkan suatu kata yang menjadi beberapa kelas. Analisis sentimen juga disebut opinion mining yang dapat diartikan menggali opini dan emosi dari data uji. Sentimen akan diklasifikasikan menjadi beberapa kelas positif, negatif dan netral. Islam merupakan agama yang muncul berabad-abad tahun yang lalu. Ajaran islam datang bertujuan untuk membangun dan menciptakan manusia dalam kedamaian dan berserah diri terhadap penciptanya yaitu Allah SWT. Pandangan konstruktivisme terhadap islam memunculkan fenomena islamophobia yang merupakan hasil konstruksi sejak lama yang menggambarkan sisi kebalikan dari ajaran islam. Saat ini isu islamophobia semakin berkembang menciptakan pandangan – pandangan beragam. Isu tersebut berkembang pada platform media sosial yang menciptakan opini yang menarik untuk diklasifikasikan. Analisis yang dilakukan menggunakan algoritma SVM dan dataset yang bersumber dari tweet pada platform media sosial twitter dengan jumlah 1000 data menghasilkan tingkat akurasi sebesar 99.99% setelah dilakukan pengujian. Hasil ini menunjukkan algoritma SVM cukup baik untuk melakukan analisis sentimen dimana tingkat error yang dihasilkan menggunakan MSE sebesar 0.010 dan RMSE sebesar 0.099.

Kata kunci : *analisis sentimen, SVM, MSE, RMSE, Islamophobia.*

ABSTRACT

Sentiment analysis is the process of understanding and classifying a word into several classes. Sentiment analysis is also called opinion mining, which can be interpreted as exploring opinions and emotions from test data. Sentiments will be classified into positive, negative, and neutral classes. Islam is a religion that emerged centuries ago. Islamic teachings come with the aim of building and creating humans in peace and surrender to their creator, namely Allah SWT. The constructivism view of Islam raises the phenomenon of Islamophobia, which is the result of a long construction that describes the opposite side of Islamic teachings. Currently, the issue of Islamophobia is growing, creating diverse views. The issue develops on social media platforms that create interesting opinions to classify. The analysis was carried out using the SVM algorithm and a dataset sourced from tweets on the Twitter social media platform, with a total of 1000 tweets, resulting in an accuracy rate of 99.99% after testing. These results show the SVM algorithm is good enough to perform sentiment analysis where the error rate generated using MSE is 0.010 and RMSE is 0.099.

Keywords : *sentiment analysis, SVM, MSE, RMSE, Islamophobia*

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillahirobbil Alamin, segala puji syukur bagi Allah SWT atas segala rahmat dan karunia-Nya, sehingga saya dan tim penelitian dapat menyelesaikan penelitian ini dengan judul “**Analisis Sentimen Terhadap Isu Islamophobia Menggunakan Algoritma SVM**”. Sholawat dan salam senantiasa dipanjatkan kepada Nabi Besar Muhammad SAW beserta para sahabat – sahabat beliau yang telah membawa manusia dan seisi alam dari kegelapan ke cahaya sehingga kita menjadi manusia beriman, berilmu dan tetap beramal shaleh agar menjadi manusia yang berakhlak mulia.

Penulisan laporan ini bertujuan untuk melengkapi persyaratan luaran penelitian. Laporan ini juga diharapkan dapat menambah wawasan ilmu pengetahuan, khususnya dibidang ilmu komputer dalam ranah nilai – nilai Islam yang terpadu dalam proses pembelajaran dalam lingkungan Universitas Islam Negeri Sumatera Utara.

Dalam penulisan ini, penulis menyadari bahwa semua yang tertuang dalam laporan ini jauh dari kata sempurna. Oleh karena itu penulis mengharapkan masukan berupa kritik dan saran demi evaluasi untuk memperbaiki kekurangan dalam penelitian ini agar nanti dapat menjadi sempurna. Terima kasih kepada anggota peneliti dan tim penelitian yang sudah fokus dan berkontribusi dalam membantu penelitian ini dan pastinya bantuan dari semua sangat berperan dari awal hingga laporan ini selesai. Dan kepada semua pihak yang tidak bisa disebutkan satu persatu, ribuan terima kasih kami aturkan atas bantuan yang telah diberikan selama ini kepada kami. Akhir kata penulis berharap semoga laporan penelitian ini dapat bermanfaat.

Medan, 05 September 2022
Penulis,

Aidil Halim Lubis, M.Kom
NIP. 198805272019031010

DAFTAR ISI

ABSTRAK	iii
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Permasalahan	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
BAB II LANDASAN TEORI	4
2.1. Text Mining	4
2.2. Machine Learning	5
2.3. Sentimen Analisis	8
2.4. Pre-Processing	10
2.5. Feature Selection.....	12
2.6. Dataset	14
2.7. Support Vector Machine (SVM)	14
2.8. K-Fold Cross Validation	18
2.9. Mean Square Error (MSE)	18
2.10. RMSE (Root Mean Square Error)	19
2.11. Mean Absolute Error (MAE)	20
2.12. Orange Data Mining	21
2.13. Flowchart	22

BAB III METODE PENELITIAN	25
3.1. Perencanaan	25
3.2. Analisa Kebutuhan.....	27
3.3. Pemodelan dan Perancangan.....	27
3.4. Hasil Prediksi Algoritma.....	28
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Pembahasan	29
4.2. Analisis Data.....	29
BAB V PENUTUP.....	35
5.1. Kesimpulan	35
5.2. Penutup	35
DAFTAR PUSTAKA	36

DAFTAR TABEL

Tabel 1. Simbol Pada Flowchart	23
Tabel 2. Pre-Processing	31
Tabel 3. Hasil Evaluasi Model	33
Tabel 4. Hasil Prediksi Model SVM	34

DAFTAR GAMBAR

Gambar 1. Support Vector Machine.....	15
Gambar 2. Pengaruh parameter C pada SVM.....	17
Gambar 3. Visualisasi kelengkungan akibat parameter γ	17
Gambar 4. Pengaruh parameter γ pada SVM	18
Gambar 5. Orange Data Mining	21
Gambar 6. Flowchart Sistem	26
Gambar 7. Proses Pengolahan Data dengan Orange Data Mining	29
Gambar 8. Preprocess Text dengan Orange.....	30
Gambar 9. Hasil Pembobotan TF-IDF	33
Gambar 10. Visualisasi sebaran data.....	34

BAB I

PENDAHULUAN

1.1. Latar Belakang

Islam merupakan agama yang muncul berabad-abad tahun yang lalu. Agama ini muncul pertama kali di wilayah Arab, yaitu 610 M yang ditandai dengan terimanya wahyu Al-Quran yang pertama di Makkah oleh Muhammad SAW. Ajaran islam yang datang ke dunia intinya bertujuan untuk membangun dan menciptakan manusia dalam kedamaian dan sikap berserah diri terhadap penciptanya yaitu Allah SWT. Secara umum seorang yang beragama islam harus mampu untuk menyelamatkan diri sendiri dan juga orang lain agar tercipta kedamaian dimuka bumi ini. Banyak tantangan muncul untuk menciptakan kedamaian dimuka bumi, antara lain saat ini beredar isu fenomena islamophobia. Pandangan konstruktivisme mengenai fenomena islamophobia yang merupakan hasil konstruksi sejak lama bahwa islam itu kejam, keji dan tidak manusiawi. Isu islamophobia menjadi masalah yang menakutkan, karena islamophobia umat muslim mengalami pelecehan dan ketidakadilan dan orang-orang yang membenci islam justru menghukum orang yang tidak bersalah alias tidak ada kausal dari masalah (Zulian, 2020). Pada kasus yang muncul belakang ini mengenai dideportasinya seorang tokoh agama islam oleh sebuah Negara di Asia. Kasus tersebut membuat polemik dan perdebatan dimedia sosial dan menciptakan polarisasi pro dan kontra terhadap kasus tersebut. Media sosial adalah suatu tempat berbagi informasi yang cepat dan mudah. Melalui media sosial mereka bebas berekspresi dengan berasumsi bahwa pesan dapat dilihat dan didengar semua orang (Dewi & Delliana, 2020). Salah satu media sosial yang banyak digunakan adalah Twitter. Twitter merupakan media yang memiliki karakteristik yang unik yaitu

memiliki simbol dan aturan khusus, twitter membatasi jumlah teks yang dipost atau dalam istilah di twitter adalah tweet maksimal sebanyak 140 karakter (Zhang et al., 2011). Dengan media sosial pengguna bisa mengungkapkan ekspresi ataupun opini, baik berupa ungkapan positif, negatif dan netral. Ekpresi yang dihasilkan dari media sosial tersebut akan diproses menggunakan algoritma sebagai pendukung dalam menganalisis yang menghasilkan presentase tweet positif, negative dan netral (Permatasari et al., 2021). Pada penelitian ini algoritma yang digunakan adalah Support Vektor Machine (SVM) untuk proses klasifikasi.

1.2. Rumusan Permasalahan

Berdasarkan latar belakang yang telah dijabarkan diatas, maka dapat dirumuskan permasalahan penelitian sebagai berikut :

1. Bagaimana menerapkan algoritma SVM untuk melakukan klasifikasi datasheet komentar dari tweet pada media sosial twitter?
2. Bagaimana mempresentasikan data yang diambil dari twitter ?

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Dataset diambil dari tweet yang berasal dari media sosial twitter.
2. Jumlah dataset sebanyak 200 tweet.
3. Algoritma yang digunakan adalah SVM untuk klasifikasikan dataset.
4. Klasifikasi akan dikelompokkan menjadi 3 yaitu positif, negative dan netral.

5. Penelitian ini menggunakan aplikasi Orange versi 3.30.1 sebagai pengolahan data.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Mengetahui model analisis sentiment tentang isu islamophobia pada media sosial twitter menggunakan algoritma SVM
2. Mengetahui performance dari algoritma SVM pada dataset komentar twitter.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Membantu memahami cara kerja klasifikasi menggunakan algoritma SVM.
2. Membantu memahami dalam pengambilan keputusan berdasarkan klasifikasi komentar yang bersumber dari tweet dari media sosial twitter menggunakan algoritma SVM.

BAB II

LANDASAN TEORI

2.1. Text Mining

Text Mining merupakan suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tool analisis yang dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian atau pengelompokkan dan menganalisa unstructured text dalam jumlah besar. Solusi yang dapat diberikan oleh text mining, mengadopsi dan mengembangkan banyak teknik dari bidang lain seperti Data Mining, Information Retrieval, Statistik dan Matematik, Machine Learning, Linguistic, Natural Language Processing (NLP) dan Visualization (Herdhianto, 2020).

Text mining didefinisikan secara luas sebagai proses pengetahuan intensif dimana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Dalam cara yang sejalan dengan data mining, text mining berupaya untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik (Fieldman dan Sanger, 2007). Penelitian dibidang text mining menangani masalah yang berkaitan dengan representasi teks, klasifikasi, pengelompokkan, ekstraksi atau pencarian informasi dan pemodelan pola.

Terdapat banyak penerapan dalam melakukan text mining, salah satunya adalah klasifikasi teks. Klasifikasi teks merupakan proses pengelompokkan sebuah dokumen kedalam kategori atau kelas tertentu yang telah ditentukan. Tugas utama dari klasifikasi teks adalah untuk menentukan kategori dari masing-masing dokumen berdasarkan karakteristik masing-masing kelas (Sahputra, 2020). Adapun tahapan dalam klasifikasi teks, antara lain sebagai berikut :

1. Document Collecting

Document Collecting merupakan tahap pengumpulan data.

2. Pre-processing

Tahap ini merupakan tahap dimana data dinormalisasi agar data yang digunakan memiliki standar yang konsisten dan terstruktur sehingga data siap untuk diolah dan dilanjutkan ke tahap selanjutnya.

3. Ekstraksi fitur

Tahap ini merupakan tahap mengubah bentuk ciri data agar dapat diproses ke tahap selanjutnya. Dalam text mining, ciri data merupakan kumpulan kata-kata, maka kata-kata ini harus diubah terlebih dahulu menjadi angka yang dapat digunakan untuk proses klasifikasi selanjutnya.

4. Klasifikasi

Tahap ini merupakan tahapan untuk menentukan kategori setiap dokumen berdasarkan karakteristik dari masing-masing kelas terhadap dokumen.

5. Validasi

Tahap ini adalah mengukur akurasi dari klasifikasi yang dihasilkan.

2.2. Machine Learning

Machine learning pertama kali diperkenalkan oleh Arthur Samuel, seorang ahli komputer yang berasal dari Amerika Serikat pada tahun 1959. Pada saat itu, Samuel masih bekerja disalah satu perusahaan teknologi yaitu IBM. Samuel mengartikan machine learning merupakan suatu sistem yang dapat belajar tanpa harus diprogram kembali serta dapat memprediksi data dan mengambil keputusan dengan menggunakan beberapa algoritma.

Machine learning merupakan bagian dari kecerdasan buatan untuk mengembangkan suatu sistem yang diprogram oleh manusia kemudian diajarkan ke sistem agar sistem tersebut dapat belajar sendiri dan manusia tidak perlu memprogram sistem berulang kali. Untuk mengolah data menggunakan machine learning diperlukan adanya data learning dan data testing. Data learning yaitu data yang digunakan sistem untuk belajar dan data testing untuk menguji kemampuan sistem yang telah dilakukan pembelajaran dari data learning. Terdapat beberapa pendekatan utama dalam proses machine learning yaitu, pendekatan supervised learning dan unsupervised learning.

2.2.1. Supervised Learning

Supervised learning adalah sebuah pendekatan dimana sudah terdapat data latih dan terdapat variabel yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan (klasifikasi) suatu data ke data yang sudah ada. Algoritma ini mengamati sekumpulan pasangan masukan – keluaran dalam jumlah yang cukup besar dan kemudian menghasilkan sebuah model yang mampu memetakan masukan yang baru menjadi keluaran yang tepat.

Terdapat empat isu yang harus dipertimbangkan dalam menggunakan teknik supervised learning yaitu perlunya memutuskan kategori yang akan digunakan untuk mengklasifikasikan kasus. Kedua, dibutuhkan satu set pelatihan untuk masing-masing kategori. Ketiga, perlu menentukan fitur dari setiap kategori. Biasanya lebih baik untuk menghasilkan fitur sebanyak mungkin karena sebagian besar algoritma akan dapat fokus hanya pada fitur yang relevan. Terakhir, perlu memutuskan algoritma yang digunakan untuk kategorisasi tersebut. Pada kelompok supervised learning ini ada banyak

algoritma yang dapat digunakan sebagai metode untuk mengklasifikasikan suatu data. Dalam hal ini salah satunya adalah algoritma SVM termasuk kelompok supervised learning.

2.2.2. Unsupervised Learning

Dalam pendekatan unsupervised learning, metode ini diterapkan tanpa adanya latihan (training) dan tanpa adanya guru (teacher). Guru yang dimaksud adalah label dari data. Misalkan ada sekelompok pengamatan atau data tanpa ada label (output) tertentu, maka dalam unsupervised learning harus mengelompokkan data tersebut kedalam beberapa kelas yang dikehendaki. Ini terutama dilakukan karena data yang tidak memiliki label. Label menandai kemana data akan dikelompokkan. Untuk melakukan tugas (task) ini bisa diterapkan metode unsupervised learning.

Unsupervised learning adalah teknik yang terlebih dahulu menciptakan sebuah sentimen tanpa data latih, dan kemudian menentukan orientasi sentimen dari unit teks melalui beberapa fungsi berdasarkan positif dan negatif. Menentukan sentimen dengan pendekatan unsupervised learning adalah melalui kata-kata atau frase dengan polarisasi sentimen, juga disebut sebagai orientasi semantik.

2.2.3. Semi-Supervised Learning

Semi-supervised Learning adalah tipe learning dimana kita mempunyai data masukan (input data) dalam jumlah besar dan hanya beberapa dari data tersebut yang dilabeli. Permasalahan ini berada diantara supervised learning dan unsupervised learning.

2.2.4. Reinforcement Learning

Reinforcement learning terjadi ketika kita menyajikan algoritma dengan contoh yang kekurangan label, tetapi kita dapat menyertakan contoh dengan feedback positif atau negatif bergantung pada solusi yang ditawarkan oleh algoritma tersebut. Dengan reinforcement learning, mesin dan software dapat menentukan tingkah laku ideal terhadap sebuah konteks yang spesifik secara otomatis, dengan tujuan untuk memaksimalkan performanya. Umpan balik (feedback) ‘simple reward’ dibutuhkan untuk mesin mempelajari tingkah lakunya, hal ini disebut reinforcement signal.

2.3. Sentimen Analisis

Analisis sentimen adalah proses untuk memahami dan mengelompokkan suatu kata yang dibagi menjadi beberapa kelas. Analisis sentimen juga disebut opinion mining yang dapat diartikan menggali opini dan emosi dari data uji. Tujuan dari analisis sentimen adalah untuk mengekstrak atribut dan komponen dari beberapa komentar yang ada di media sosial sehingga dapat menentukan beberapa kelas positif, negatif dan netral (Permatasari et al., 2021).

Menurut (Liu, 2011) sentiment analysis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistic dan text mining yang memiliki tujuan untuk menganalisis pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu ataupun kegiatan tertentu. Analisis sentimen sangat berguna sebagai pemrosesan penganalisis komentar seperti pendapat kemudian diproses menjadi sesuatu yang lebih bermakna (Palucoumputer, 2015) seperti diolah menjadi bentuk

rating. Dengan analisis sentimen inilah dapat mencoba mengaplikasikan komentar kedalam system rating.

Kecenderungan penelitian tentang analisis sentimen berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negative. Pendapat mewakili hamper semua aktivitas manusia, karena pendapat dapat mempengaruhi terhadap perilaku seseorang. Analisis sentimen pada suatu kalimat menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu (Pang, dkk. 2008).

2.3.1. Sentimen Analisis pada Twitter

Secara definisi analisis sentimen twitter merupakan bagian dari pendapat pada media sosial twitter. Pesan atau sering disebut juga dengan tweet pada kenyataannya lebih mudah untuk dianalisis karena penulisan yang dibatasi dibanding dengan forum diskusi pada media sosial lainnya. Hal ini berbeda pada forum diskusi yang lebih sulit, dikarenakan pengguna dapat mendiskusikan apapun dan berinteraksi satu sama lain. Kalimat sering kali memuat pendapat tunggal, meskipun tidak bersifat mutlak bahwa setiap kalimat berisi pendapat tunggal. Dalam kasus lain terdapat kalimat dengan pendapat lebih dari satu pada suatu kalimat namun hanya sebagian kecil (Liu, 2012).

pada umumnya sentimen analisis merupakan tahapan klasifikasi. Namun tahapan klasifikasi sentimen pada Twitter (tidak terstruktur) sedikit lebih sulit dibandingkan dengan klasifikasi dokumen terstruktur. Dalam kasus ini, analisis sentimen Twitter yang merupakan gambaran dari kalimat, langkah pertama (Liu, 2012) adalah untuk mengklasifikasikan apakah kalimat

mengungkapkan pendapat atau tidak. Langkah kedua adalah mengklasifikasikan kalimat-kalimat pendapat menjadi positif dan negative. Menurut (Barbosa, dkk. 2010 dikutip Liu, 2012) terdapat beberapa fitur yang dapat digunakan, diantaranya *retweets*, *hashtag*, link, kata-kata huruf, *emoticon*, tanda seru dan tanda tanya.

Ada beberapa hal yang menjadi alasan Twitter digunakan sebagai analisis sentiment (pak, dkk. 2010) yaitu :

1. Twitter adalah *microblogging* yang digunakan oleh orang-orang yang berbeda untuk mengungkapkan pendapat mereka tentang topik yang berbeda, sehingga merupakan sumber yang baik dalam menemukan pendapat orang lain.
2. Twitter memuat berbagai pesan teks setiap harinya dalam jumlah yang besar.
3. Pengguna Twitter bervariasi dari berbagai kalangan. Oleh karena itu, dapat menemukan pesan pengguna dari berbagai sosial dan kepentingan kelompok.

2.4. Pre-Processing

Sebelum melakukan proses text mining, maka perlu dilakukan pre-processing terlebih dahulu. Pre-processing ini seperti layaknya melakukan pembersihan data, membuang duplikasi, memperbaiki typo, memeriksa data yang tidak konsisten dan lain sebagainya. Tahap pre-processing merupakan tahapan dimana dilakukan seleksi data agar data yang akan digunakan menjadi lebih terstruktur dan sesuai dengan format yang ada (Aditama et al, 2020).

Pre-processing ini sangat penting untuk dilakukan dalam pengolahan data analisis sentimen dari media sosial, dengan tujuan antara lain untuk

menyeragamkan bentuk kata, menghilangkan noise, mengurangi volume kata dan memadatkan kata menjadi beberapa token yang penting. Terdapat beberapa tahapan yang dilakukan dalam pre-processing, antara lain yaitu :

1. *Case folding* merupakan tahap mengubah semua huruf menjadi bentuk huruf kecil (lower case). Hal ini dilakukan untuk menyeragamkan bentuk format kata pada data teks agar mudah untuk diproses pada tahapan selanjutnya.
2. *Filtering* merupakan tahap membuang noise, menghapus karakter-karakter, emoticon dan tanda baca seperti ” @, . ! “. Angka, situs web, link url atau lain sebagainya untuk mempermudah tahapan selanjutnya.
3. *Tokenizing* merupakan pemisahan rangkaian term, dimana tiap deretan kata didalam kalimat, paragraph atau halaman dipisahkan menjadi token atau potongan kata tunggal atau termed word.
4. *Normalisasi* merupakan tahap pemeriksaan ulang untuk mengidentifikasi kata atau huruf yang berlebihan pada setiap token agar diganti sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).
5. *Stopword Removal* merupakan proses untuk menghapus kata-kata yang sering muncul pada suatu dokumen. Kata-kata tersebut merupakan kata yang tidak begitu penting karena terlalu banyak.
6. *Stemming* merupakan proses mengubah kata menjadi kata dasar dari seluruh kata tanpa menghilangkan makna dari setiap kata tersebut. Setiap kata yang memiliki imbuhan diawal dan diakhir kata akan dihilangkan. Contoh seperti kata “perancangan” menjadi “rancang”. Adapun algoritma stemming sama seperti yang diterapkan pada algoritma Nazief dan Andriani, yaitu :

- a. Cari kata pada daftar kamus. Jika ada kata pada daftar kamus, maka kata tersebut merupakan kata dasar.
- b. Jika kata seperti langkah pertama tidak ada pada kamus, periksa imbuhan pada kata belakang (sufiks) seperti kata “lah” dan “kah”. Jika ada hilangkan sufiks.
- c. Periksa kata ganti milik seperti “-ku”, “-mu”, dan “nya” pada dokumen. Jika terdapat kata tersebut maka dihilangkan.
- d. Memeriksa akhiran (“-i”, “-an”), (“-lah”, “-kah”) dan kata ganti milik (“-ku”, “-mu”, “-nya”). Jika ada ditemukan, maka akhiran dihilangkan.
- e. Periksa awalan dokumen seperti kata “di-“, “se-“, “ke-“, “te-“, “be-“, “pe-“, dan “me-“. Bila ditemukan, kata awalan dihapus. Adanya kemungkinan kata multi-prefiks diperlukan berulang kali pemeriksaan pada dokumen.
- f. Setelah semua tahapan telah selesai dilakukan dan berhasil, kata dasar yang ditemukan akan dikembalikan oleh algoritma.

2.5. Feature Selection

Tahapan ini berperan dalam menentukan *term* atau kata kunci yang menjadi ciri dari suatu dokumen yang dapat membedakan dokumen satu dengan lainnya dalam satu *corpus*. Dalam text mining, *feature selection* merupakan tahapan paling penting yang memiliki peranan sangat signifikan dalam akurasi text analytic, dikarenakan feature selection merupakan suatu proses yang digunakan untuk menghilangkan atau menghapus *features* yang tidak relevan yang terdapat dalam satu *dataset*. Terdapat empat pendekatan yang paling banyak digunakan dalam proses *features selection* diantaranya

yaitu *Document Frequency (IDF)*, *Term Frequency (TF)*, *Inverse Document Frequency (IDF)* serta *Term Frequency/Inverse Document Frequency(TF/IDF)*.

a. *Document Frequency (IDF)*

DF memiliki prinsip kerja membuang term umum yang terdapat didalam dokumen yang ada pada suatu corpus dokumen teks. Hal tersebut menjadikan term yang tersisa dalam suatu dokumen merupakan term yang memiliki tingkat overlapping rendah dibanding dengan term yang terdapat pada dokumen lain dalam suatu corpus.

b. *Term Frequency (TF)*

Pendekatan TF berbeda dengan DF, dalam pendekatan ini tidak mengindahkan term yang terkandung dalam suatu dokumen yang berbeda. Metode TF secara sederhana hanya menghitung kemunculan dari term yang terdapat dalam suatu dokumen. Term dengan frekuensi kemunculan yang tinggi akan menjadi tanda dari dokumen dimana term tersebut berada.

c. *Inverse Document Frequency (IDF)*

Pendekatan IDF hampir sama dengan TF, yakni menghitung frekuensi kemunculan suatu term. Namun, jika TF menghitung kemunculan suatu term hanya di satu dokumen teks, maka IDF menghitung kemunculan suatu term di keseluruhan corpus dokumen.

d. *Term Frequency/Inverse Document Frequency (TF/IDF)*

TF/IDF adalah gabungan dari pendekatan TF dan IDF yang mana proses perhitungan mengambil rasio antara nilai TF dan nilai IDF.

2.6. Dataset

Dataset atau set data merupakan kumpulan instansi sebuah informasi yang isinya mengandung atribut. Set data dapat menyimpan informasi seperti catatan medis penyakit, catatan asuransi, perpustakaan makro, dan berbagai macam jenis data lainnya yang dibutuhkan oleh sistem itu sendiri. Kumpulan data ini yang kemudian di proses ke dalam algoritma mesin pembelajaran untuk melatih model yang dinamakan dengan *training dataset*. Namun, jika terdapat nilai yang hilang dalam suatu kumpulan data dapat menyebabkan masalah selama fase *training* dan klasifikasi. Beberapa kemungkinan alasan yang menyebabkan hilangnya data adalah data yang ada tidak relevan dengan entri, penghapusan data karena penyimpangan, dan adanya kerusakan data. Masalah adanya data yang hilang dapat diatasi dengan pendekatan seperti penambang data mengabaikan data yang dihilangkan, menukar seluruh nilai yang dihilangkan dengan global atau dengan rata-rata fitur kelas yang diberikan, mengamati sampel secara manual dengan menghilangkan nilai dan memasukkan nilai yang layak atau mungkin. Set data penelitian ini diambil dari API twitter berupa postingan atau data tweet.

2.7. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma klasifikasi dalam bidang pembelajaran mesin. Ide utama dari SVM adalah untuk menemukan hyperlane terbaik yang memisahkan 2 daerah keputusan dengan baik. Secara umum, SVM bekerja dengan mengikuti langkah-langkah berikut:

1. Fitur-fitur dari data dipetakan kedalam ruang dimensi tinggi menggunakan fungsi kernel (linear, polynomial, sigmoid, atau radial basis function). Artinya, dengan menggunakan fungsi kernel,

akan dihasilkan fitur-fitur baru yang akan digunakan untuk mengklasifikasikan (tidak lagi menggunakan fitur-fitur lama).

Masing-masing kernel dirumuskan sebagai berikut:

a. Kernel Linear

$$K(x_i, x_j) = x_i^T x_j \quad (1)$$

b. Kernel Polynomial

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (2)$$

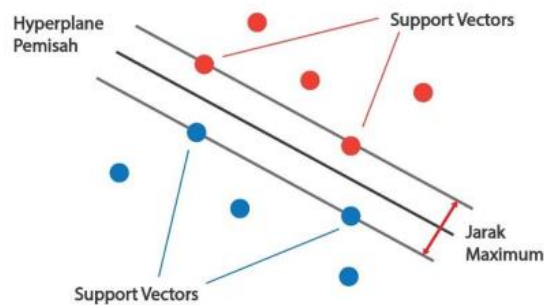
c. Kernel Sigmoid

$$K(x_i, x_j) = \tanh(x_i^T x_j + r) \quad (3)$$

d. Kernel Radial Basis Function

$$K(x_i, x_j) = \exp(-\gamma \|x_i^T - x_j\|^2), \gamma > 0 \quad (4)$$

2. Mencari hyperlane terbaik yang memisahkan data yang telah dipetakan dalam ruang dimensi tinggi. Hyperlane terbaik dapat diperoleh dengan memaksimalkan jarak hyperlane dengan titik data terdekat (Support Vectors).



Gambar 1. Support Vector Machine

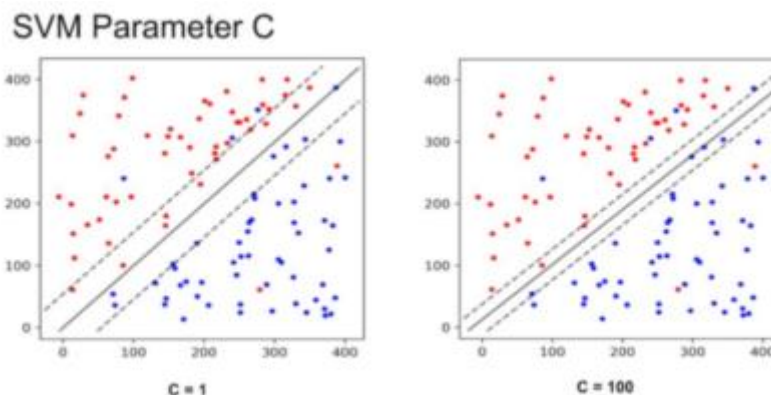
Contoh penggunaan kernel, adalah sebagai berikut. Misalkan digunakan fungsi kernel Radial Basis Function dengan data $x = x_1, x_2, \dots, x_n$. Fitur-fitur hasil kernel RBF dimisalkan sebagai $f = f_1, f_2, \dots, f_m$, dimana fitur-fitur tersebut

digunakan untuk menentukan hasil klasifikasi sebagai berikut (w adalah bobot, f adalah fitur).

$$\text{If : } w_1f_1 + w_2f_2 + \dots + w_n f_m > 0, \text{ then :class} = 1, \text{ else: class} = 0$$

Nilai f dapat dicari dengan $f = K(x_i x_j) = \exp(-\gamma \|x_j^T - x_i\|^2)$. Untuk sementara, substitusikan x_j dengan suatu titik random l , sehingga persamaan akan diubah sementara menjadi $f_1 = K(x_i x_j) = \exp(-\gamma \|x_j^T - x_i\|^2)$. Dengan menggunakan fungsi kernel RBF dan memahami bahwa $\|x_i^T - l\|^2$ adalah jarak antara x_i dengan titik l , maka dapat disimpulkan bahwa ketika suatu titik data x_i berada sangat dekat dengan titik l , maka nilai f akan mendekati 1, sebaliknya jika x_i semakin jauh dengan titik l , maka nilai f akan mendekati 0.

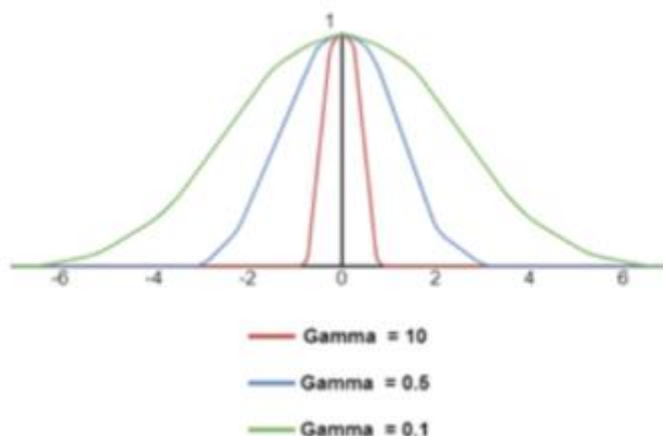
Terdapat parameter-parameter yang dapat diubah untuk menghasilkan hyperlane pemisah terbaik, diantaranya adalah parameter C dan parameter γ . Parameter regulasi C digunakan SVM untuk menemukan jarak terbaik dari hyperlane. Nilai C yang terbaik harus dicari agar SVM terjadi keseimbangan antara jarak semaksimal dan kesalahan klasifikasi seminimal mungkin. Nilai C yang tinggi menyebabkan bahwa model akan menerima lebih banyak penalti ketika model gagal mengklasifikasikan data dengan tepat. Sebaliknya nilai C yang rendah menyebabkan bahwa model akan mentolerir data yang salah diklasifikasikan. Nilai C yang rendah biasanya baik digunakan pada data yang memiliki banyak noise, sebaliknya nilai C yang tinggi biasanya baik digunakan pada data yang memiliki sedikit noise.



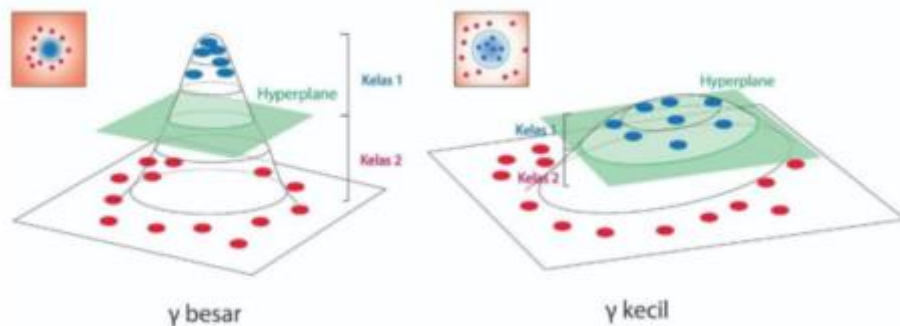
Gambar 2. Pengaruh parameter C pada SVM

Sumber : <https://learnopencv.com/svm-using-scikit-learn-in-python/>

Selain parameter C, pada kernel Radial Basis Function terdapat parameter γ yang besar akan menghasilkan kelengkungan yang tajam pada ruang dimensi tinggi. Sebaliknya, nilai γ yang kecil akan menghasilkan ruang dimensi tinggi yang lebih landai. Pasangan nilai parameter C dan γ harus ditemukan agar SVM dapat menggeneralisasi data dengan baik.



Gambar 3. Visualisasi kelengkungan akibat parameter γ



Gambar 4. Pengaruh parameter γ pada SVM

2.8. K-Fold Cross Validation

K-fold adalah salah satu metode dari cross validation. Konsep dari *k fold cross validation* tidak hanya membuat beberapa sampel data uji berulang kali saja, akan tetapi juga membagi dataset kedalam bagian terpisah dengan ukuran yang sama. Model klasifikasi dilatih oleh subset data latih kemudian divalidasi oleh subset validasi (data uji) sebanyak k . misalkan dengan memilih data latih secara acak sebanyak 10 kali untuk digunakan dalam membangun model klasifikasi untuk digunakan, pada data uji. Penggunaan cara ini dapat diperoleh rata – rata nilai kinerja dari 10 kali test error. Prosedur ini merupakan cara yang standar untuk memvalidasi model klasifikasi. Penggunaan k-fold cross validation mampu mengurangi waktu komputasi yang mungkin terjadi karena adanya proses iterasi dengan tetap menjaga keakuratan estimasi model (Mustika et al., 2021).

2.9. Mean Square Error (MSE)

Didalam membuat model estimasi tidak akan lepas dari *Mean Square Error* (MSE) dan dua komponennya yaitu bias dan standard error atau varians.

Kriteria error yang digunakan dalam regresi non parametrik ini bukan lagi *Least Square Error* melainkan dengan mencari nilai Mean Square Error (MSE) yang terkecil. Untuk menentukan jenis uji yang paling mendekati kebenaran dilakukan dengan mengukur *error* (kesalahan). Untuk mengukur *error* biasanya digunakan *Mean Square Error*. Penguji yang menghasilkan *error* terkecil adalah uji yang dipilih. Mean Square Error (MSE) adalah kuadrat dari rata-rata kesalahan.

$$MSE = \frac{\sum(Y_i - \hat{Y}_i)^2}{n} \quad (5)$$

Keterangan :

y_i = Data sebenarnya

\hat{y}_i = Nilai prediksi dari variabel Y

n = Banyaknya observasi

2.10. RMSE (Root Mean Square Error)

Kriteria yang digunakan untuk mengukur kebaikan model setelah diperoleh suatu model adalah *root mean square error* (RMSE) (Wahyuningsih, 2012). RMSE merupakan alat seleksi model berdasarkan pada *error* hasil estimasi. Error yang ada menunjukkan seberapa besar perbedaan hasil estimasi dengan nilai yang akan estimasi. Nilai ini akan digunakan untuk menentukan model mana yang terbaik. Definisi RMSE dapat ditulis sebagai berikut.

$$RMSE = \sqrt{\frac{\sum_i^n 1(y_i - \hat{y}_i)^2}{n}} \quad (6)$$

n = Jumlah sampel

y_i = Nilai aktual

Y_i = Nilai prediksi

2.11. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Nilai MAE menunjukkan rata – rata kesalahan (*error*) absolut antara hasil peramalan/prediksi dengan nilai riil (Subagyo, 1986). Secara umum persamaan MAE dijelaskan sebagai berikut,

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_1 - y_i| \quad (7)$$

Dimana :

f_i : nilai hasil peramalan,

y_i : nilai sebenarnya

n : jumlah data

berdasarkan persamaan diatas, MAE menghitung rata – rata *error* dengan memberikan bobot yang sama untuk seluruh data ($i = 1 \dots n$) secara intuitif. Untuk evaluasi model peramalan, MAE lebih intuitif dalam memberikan rata – rata *error* dari keseluruhan data. Dalam kasus ini pemilihan MAE menjadi tepat karena seluruh data diberikan bobot yang sama.

2.12. Orange Data Mining

Data mining merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data. Data mining hadir sebagai bidang ilmu yang berusaha mengolah data, saat ini semakin banyak dihasilkan dan tersimpan menjadi suatu pengetahuan yang bermanfaat serta berharga. Data mining pada umumnya diimplementasikan menggunakan teknik-teknik seperti dengan menggunakan bahasa pemrograman tertentu atau bahasa pemrograman lainnya.

Orange atau lebih dikenal dengan Orange Data Mining merupakan perangkat lunak open source yang dapat melakukan proses data mining atau data analytic melalui koden visual programming. Keunggulan orange data mining terlihat pada kemampuan mengolah data sumber dari berbagai tipe seperti csv atau lainnya. Pada orange data mining terdapat berbagai fitur dan algoritma yang bisa digunakan sesuai dengan kebutuhan seperti KNN, Naïve bayes, K-Means Clustering dan lainnya.



Gambar 5. Orange Data Mining

2.13. Flowchart

Flowchart merupakan gambar atau bagan yang memperlihatkan urutan dan hubungan antar proses beserta instruksinya. Gambaran ini dinyatakan dengan simbol. Dengan demikian setiap simbol menggambarkan proses tertentu. Sedangkan hubungan antar proses digambarkan dengan garis penghubung.

Flowchart ini merupakan langkah awal pembuatan program. Dengan adanya *flowchart* urutan poses kegiatan menjadi lebih jelas. Jika ada penambahan proses maka dapat dilakukan lebih mudah. Setelah *flowchart* selesai disusun, selanjutnya pemrogram (*programmer*) menerjemahkannya ke bentuk program dengan bahasa pemrograman.

Terdapat 2 jenis *flowchart* yaitu sebagai berikut :

1. Sistem *Flowchart*

Sistem *flowchart* merupakan diagram alir yang menggambarkan suatu sistem peralatan komputer yang digunakan dalam proses pengolahan data serta hubungan antar peralatan tersebut. Sistem *flowchart* tidak digunakan untuk menggambarkan urutan langkah untuk memecahkan masalah , tetapi hanya untuk menggambarkan prosedur dalam sistem yang dibentuk.

2. Program *Flowchart*

Merupakan bagan alir yang menggambarkan urutan logika dari suatu prosedur pemecahan masalah. Untuk menggambarkan *flowchart* program telah tersedia simbol-simbol standart. Berikut ini adalah gambar dari simbol-simbol standart yang digunakan pada *flowchart* program pada tabel 2.1 berikut ini : (Rifka R, 2017)

Tabel 1. Simbol Pada Flowchart

Sumber: (Rifka R, 2017)





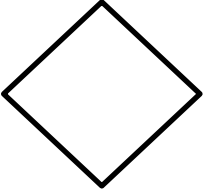
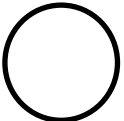

No	Simbol	Nama	Keterangan
1		<i>Terminator</i>	digunakan pada awal dan akhir aliran proses.
2		<i>Process</i>	Fungsinya adalah proses integrasi pengolahan data.
3		<i>Sub Process</i>	Fungsinya untuk menggambarkan sub program yang bisa berupa prosedur atau fungsi.
4		<i>Input and Output</i>	Simbol yang digunakan sebagai sumber data untuk diproses.
5		<i>Decision</i>	Simbol yang digunakan untuk keputusan data dengan relasi yang sedang diproses.
6		<i>On Page Connector</i>	Simbol yang digunakan untuk menghubungkan

			diagram alur dengan halaman lain.
7		<i>Flow</i>	Sebagai arah untuk tujuan program aliran.

BAB III

METODE PENELITIAN

Kerangka kerja dalam sebuah penelitian diperlukan untuk mencapai hasil yang optimal, karena hal itu maka peneliti akan membuat kerangka kerja menggunakan algoritma *Support Vector Machine* dimana pada algoritma ini menjadi suatu bagian dari pembelajaran mesin yang diterapkan untuk mempelajari *text mining* ataupun sekumpulan data yang jumlahnya sangat besar.

Adapun kerangka kerja dalam penelitian ini antara lain ialah:

3.1. Perencanaan

a. Ruang Lingkup Masalah

Masalah pokok dalam penelitian ini adalah menganalisis sentimen masyarakat terhadap isu islamophobia yang berkembang berdasarkan topik-topik yang muncul pada aplikasi media sosial twitter menggunakan algoritma *support vector machine* (SVM). Dataset yang digunakan adalah postingan atau tweet yang diambil melalui media sosial twitter tersebut.

b. Tujuan Penelitian

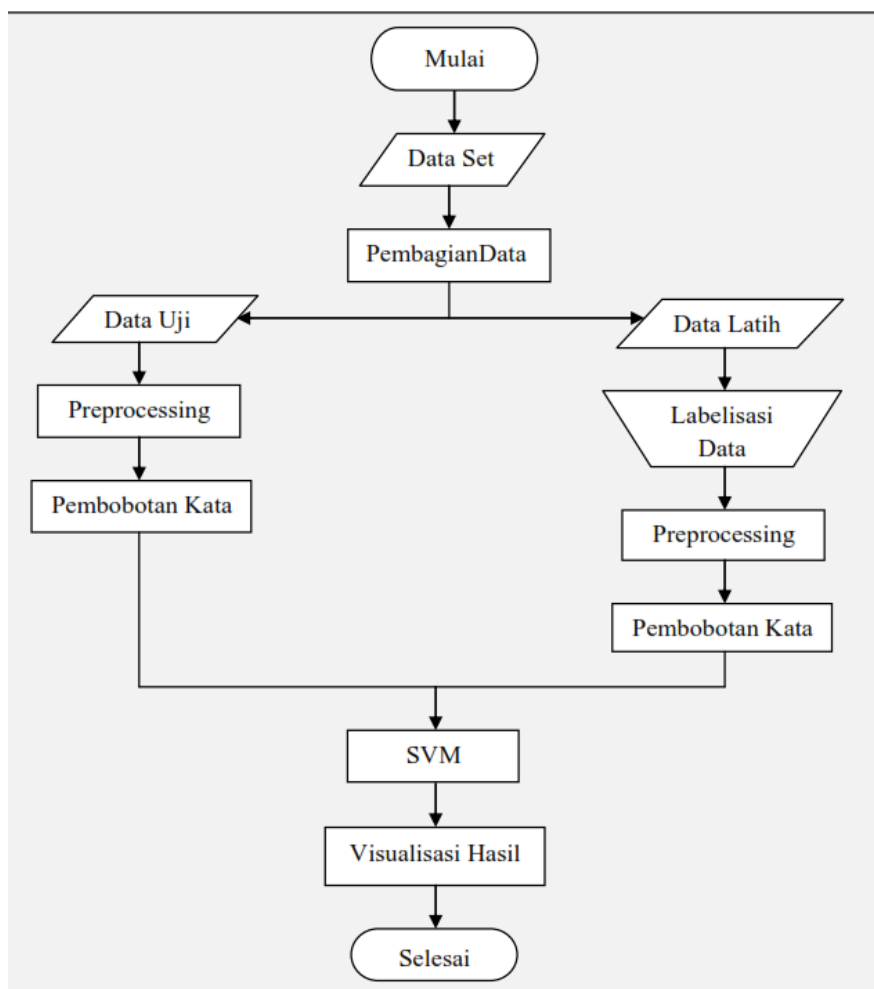
Tujuan dari pelaksanaan penelitian ini adalah agar mengetahui pola sentimen masyarakat terhadap isu islamophobia berupa sentimen positif, negatif atau netral yang bersumber dari tweet pada aplikasi media sosial twitter. Data tweet ini akan dianalisis menggunakan algoritma SVM.

c. Manfaat Penelitian

Manfaat penelitian ini memberikan model analisis sentimen dimana akan mengklasifikasikan beberapa jenis sentimen yaitu positif, negatif, dan netral dari hasil pengolahan data tersebut menggunakan algoritma SVM terhadap isu islamophobia yang ada pada media sosial twitter.

d. Flowchart rancangan

Pada penelitian ini seluruh tahapan-tahapan yang dilakukan dapat digambarkan melalui diagram alur seperti berikut.



Gambar 6. Flowchart Sistem

3.2. Analisa Kebutuhan

Pada tahap analisa kebutuhan sistem adalah dengan menganalisa kebutuhan informasi fungsi sistem yaitu berupa proses – proses yang terjadi terhadap sistem dan kebutuhan non fungsi yaitu kebutuhan yang harus dipenuhi untuk perilaku sistem. Kebutuhan perilaku berkaitan dengan kebutuhan operasional dan kinerja sistem yang akan dijalankan atau digunakan. Untuk mendapatkan kebutuhan tersebut diperlukan teknik pengumpulan kebutuhan dengan studi literatur dan analisis data.

3.3. Pemodelan dan Perancangan

Adapun pada tahapan ini metode penelitian yang digunakan dalam penelitian ini sebagai berikut :

a. Text Preprocessing

Tahap ini merupakan proses yang berfungsi untuk membersihkan teks sebelum dilakukan pengolahan lebih lanjut. Data teks diperoleh biasanya tidak terstruktur serta terdapat banyak noise seperti tanda baca, imbuhan, angka, karakter-karakter khusus dan lainnya.

b. Case Folding

Tahap ini dilakukan untuk mengkonversi keseluruhan teks menjadi satu bentuk standar. Case folding biasanya akan mengkonversi keseluruhan teks menjadi lowercase.

c. Cleaning

Tahap ini dilakukan pembersihan data, seperti penghapusan karakter-karakter selain yang ditentukan seperti tanda baca, karakter selain alphabet dari A-Z, menghapus url (link), menghapus hashtag, menghapus tanda username.

d. Tokenizing

Tahap tokenizing merupakan tahapan yang dilakukan untuk melakukan pemotongan string input berdasarkan kata penyusun. Secara garis besar, proses tokenizing akan memecah teks sekumpulan karakter yang terdapat dalam teks ke dalam satuan kata, sehingga memudahkan untuk membedakan karakter tertentu yang dapat diperlukan sebagai pemisah kata atau bukan.

e. Filtering

Tahap ini dilakukan pengambilan kata-kata penting hasil dari proses tokenizing. Proses filtering menggunakan algoritma stoplist (membuang kata kurang penting) atau wordlist (menyimpan kata penting).

e. Stemming

Stemming tidak hanya digunakan untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, namun juga merupakan tahapan yang dilakukan untuk mencari bentuk kata dasar tiap kata dari hasil proses filtering.

3.4. Hasil Prediksi Algoritma

Pada tahapan ini akan memaparkan hasil serta membahas analisis sentimen terhadap isu islamophobia menggunakan dataset tweet dari aplikasi media sosial twitter yang diambil melalui API twitter. Dataset akan dianalisis menggunakan algoritma *support vector machine* (SVM). Pada tahap ini akan dilakukan pengujian berdasarkan dataset yang telah dilatih untuk melakukan pengujian terhadap data uji. Hasil pengujian data uji akan menghasilkan prediksi berupa komentar yang mengandung sentimen positif, negatif dan netral. Data latih yang diperoleh dari crawling menggunakan API Twitter berjumlah 1000 data, dan data tersebut sudah dilakukan langkah awal yaitu preprocessing untuk mendapatkan data yang sudah benar-benar dapat diolah oleh menggunakan algoritma SVM.

BAB IV

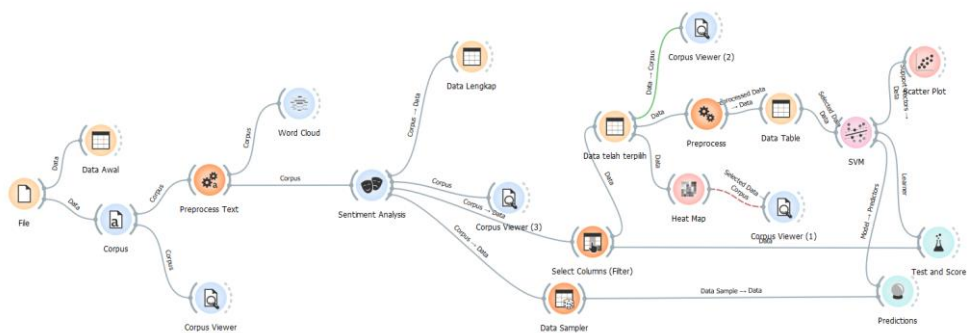
HASIL DAN PEMBAHASAN

4.1. Pembahasan

Pada penelitian ini akan diimplementasikan algoritma *support vector machine* (SVM) untuk menganalisis dataset tweet yang bersumber dari aplikasi media sosial twitter. Dataset tweet terkait isu “islamophobia” yang diperoleh bulan Juni 2022 dengan cakupan yang lokasi berada di Indonesia.

4.2. Analisis Data

Pada tahap seluruh proses dari awal hingga akhir dapat dilihat melalui visual programming menggunakan orange data mining. Pada proses awal yang dilakukan adalah pre-processing text dan pada proses akhir adalah melakukan analisis menggunakan algoritma SVM dan diperoleh hasil prediksi serta score dari algoritma SVM tersebut. Proses tersebut dapat dilihat pada gambar berikut.

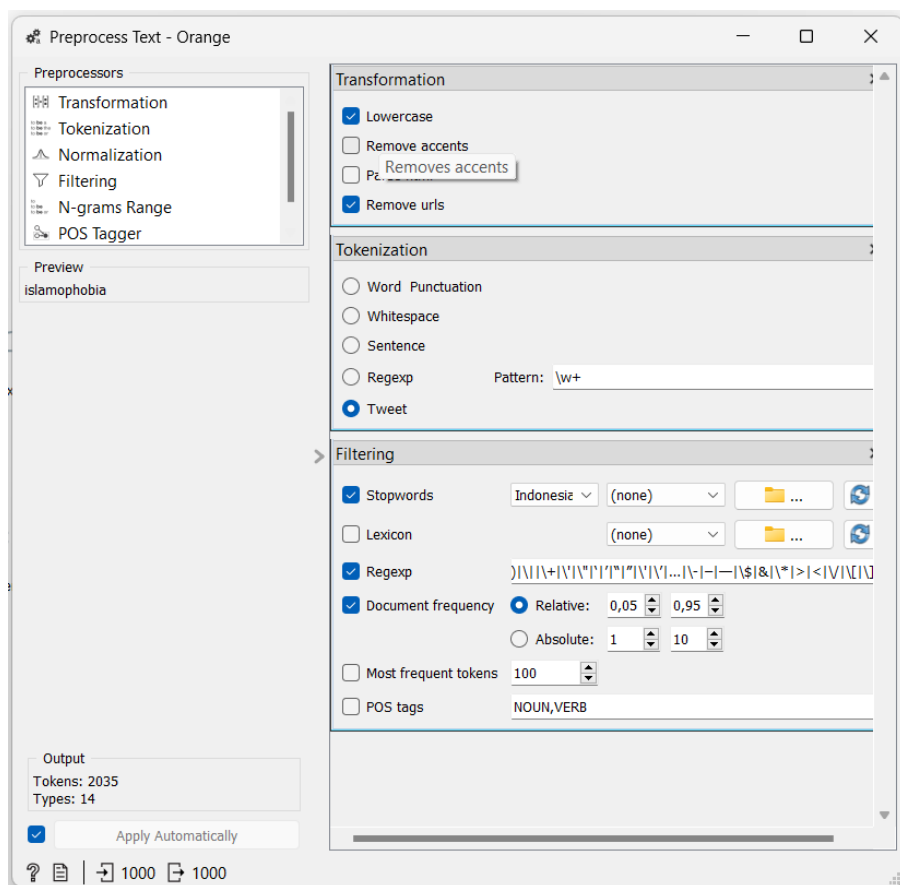


Gambar 7. Proses Pengolahan Data dengan Orange Data Mining

4.2.1 Pre-Processing

Pada tahap ini dataset tweet yang telah diperoleh akan diolah berdasarkan tahapan – tahapan pre-processing teks. Dibagian ini seluruh

dataset tweet dilakukan pre-processing yaitu membuat seluruh kata menjadi bentuk lowercase. Kemudian dilanjutkan ketahap berikutnya cleaning data berupa menghilangkan tanda baca, url, dan hastag.



Gambar 8. Preprocess Text dengan Orange

Dataset yang diperoleh sebanyak 1000 tweet, setelah dilakukan proses pre-processing maka data siap untuk dilakukan analisis selanjutnya. Adapun beberapa contoh dataset yang telah dilakukan pre-processing terlihat pada tabel berikut.

Tabel 2. Pre-Processing

Text Awal	Hasil Pre-Processing	Label
#Islamophobia https://t.co/RTIFDvTd5W	islamophobia	Netral
Mayjen (Purn) Deddy Budiman: Era Oligarki Neo-Komunisme Propaganda Islamophobia Terstruktur, Sistematis, dan Masif! https://t.co/U2PlaJAuEX	mayjen purn deddy budiman era oligarki neo komunisme propaganda islamophobia terstruktur sistematis dan masif	Netral
@Dennysiregar7 Proporsional dalam menilai dong den. Islamophobia amat makhluk pemakan bangkai satu ini.	proporsional dalam menilai dong den. islamophobia amat makhluk pemakan bangkai satu ini	Netral
@OposisiCerdas ISLAMOPHOBIA LAGI NGARANG BEBAS...😊😊😊	oposisicerdas islamophobia lagi ngarang bebas	Netral
Islamophobia https://t.co/tTg0a9aOka	islamophobia	Netral
@geloraco Kok Kemensos yg selewengkan dana Bansos	geloraco kok kemensos yg selewengkan dana bansos tidak	Netral

tidak dibubarkan, Islamophobia ?	dibubarkan islamophobia	
yg urgent itu UU ANTI ISLAMOPHOBIA & ANTI KORUPSI	yg urgent itu uu anti islamophobia anti korupsi	Negatif
@yaniarsim Dikit2 islamophobia, padahal sih cuma ini orang aja sebenarnya males nyari fakta	yaniarsim dikit islamophobia padahal sih cuma ini orang aja sebenarnya males nyari fakta	Netral
@HisyamMochtar Mereka adalah Islamophobia group	hisyammochtar mereka adalah islamophobia group	Netral
slamophobia suka cari gara2. Ntar kalo dibunuh, Islam dibilang teroris.	Islamophobia suka cari gara ntar kalo dibunuh islam dibilang teroris caper banget sih insecure kah sama agamanya sendiri	Negatif

4.2.2. Pembobotan Kata TF-IDF

Metode TF-IDF adalah metode pemberian bobot untuk setiap token pada sebuah dokumen. Nilai bobot diperoleh dari hasil perkalian frekuensi kemunculan token pada suatu dokumen (*term frequency*) dan frekuensi token

pada seluruh dokumen (*inverse document frequency*). Pada bagian proses yang dihasilkan oleh TF-IDF seperti pada gambar berikut berikut :

Word	TF-IDF
singapura	0.347
islamophobia	0.112
yq	0.099
@uas_abdulso...	0.071
konser	0.065
singapura,	0.057
indonesia	0.054
2022	0.051

Gambar 9. Hasil Pembobotan TF-IDF

4.2.3. Cross Validation

Metode yang paling biasa digunakan untuk evaluasi kinerja prediktif dari model, model ini menggeneralisasikan data independen, proses pemisahan data training dan data testing oleh fungsi K-Fold. K-Fold yang dipakai dalam penelitian ini adalah $K = 5$.

4.3.4. Hasil Evaluasi dan Prediksi

Dari hasil evaluasi menggunakan algoritma SVM didapat hasil seperti berikut:

Tabel 3. Hasil Evaluasi Model

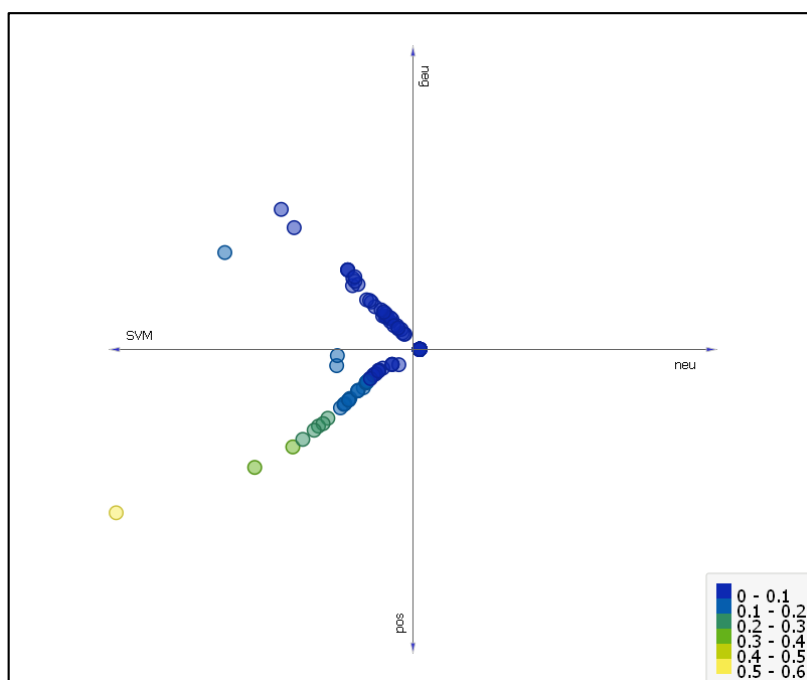
Model	MSE	RMSE	MAE
SVM	0.010	0.100	0.098

Dari hasil evaluasi model yang ditunjukkan oleh tabel diatas, maka didapat evaluasi sebesar 99,99% dan untuk akurasi prediksi model menggunakan algoritma SVM didapat sekitar 99,99%. Ini menunjukkan model telah berjalan dengan baik setelah dilakukan pengujian.

Tabel 4. Hasil Prediksi Model SVM

Model	MSE	RMSE	MAE
SVM	0.010	0.099	0.098

Hasil visualisasi sebaran dataset yang bersumber dari tweet yang telah dilakukan evaluasi menggunakan algoritma SVM terlihat pada gambar berikut.



Gambar 10. Visualisasi sebaran data

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan penelitian dan pembahasan yang telah dilakukan, hasil analisis sentimen terhadap isu islamophobia menggunakan algoritma SVM seperti berikut :

1. Hasil analisis mendapatkan nilai akurasi sebesar 99.99%. ini menunjukkan algoritma SVM bekerja secara optimal dalam melakukan analisis terhadap dataset tweet yang digunakan selama penelitian. Dan algoritma *support vector machine* (SVM) mampu melakukan klasifikasi terhadap dataset tersebut dan mengelompokkan menjadi beberapa sentimen, baik itu sentimen positi, sentimen negatif dan netral.
2. Hasil nilai evaluasi dan prediksi dapat dibuktikan dengan berdasarkan nilai *Mean Square Error* (MSE) yaitu 0,010 dan nilai *Root Mean Square Error* (RMSE) yaitu 0,099. Untuk nilai *Mean Absolute Error* (MAE) berada pada angka 0,098.

5.2. Penutup

Dengan selesainya penelitian ini dilaksanakan hingga tahap akhir dan laporan yang dihasilkan, maka:

1. Model analisis sentimen terhadap isu islamophobia terdiri dari sentimen positif, negatif dan netral.
2. Dataset untuk penelitian ini selanjutnya perlu untuk selalu diperbaharui sesuai dengan perkembangan yang terarah agar algoritma dapat terus belajar (*learner*).

DAFTAR PUSTAKA

- Aditama, M. I., Pratama, R. I., Wiwaha, K. H., & Rakhmawati, N. A. (2020). Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19. *Journal Information Engineering and Educational Technology*, 90-92.
- Dewi, A. P., & Delliana, S. (2020). Self Disclosure Generasi Z Di Twitter. *Ekspresi Dan Persepsi: Jurnal Ilmu Komunikasi*, 3(1), 62. <https://doi.org/10.33822/jep.v3i1.1526>
- Herdhianto, A. (2020). Sentiment Analysis Menggunakan Naive Bayes Classifier (NBC) pada Tweet Tentang Zakat. Skripsi: Universitas Islam Negeri Syarif Hidayatullah.
- Liu, (2011) Sentiment Naliysis and Opinion Mining., Bing Liu Departmen Of Computer Science University III inoise at Chicago.
- Liu, B. (2012). Sentiment Analysis and Subjectivity. Synthesis Lectures on Human Language Technologies. USA: editor: Graeme Hirst Morgan & Claypool Publishers.
- Mustika, Ardilla, Y., & Manuhutu, A. (2021). *Data Mining dan Aplikasinya*(Rismawati (ed.)). Widiana Bhakti Persada Bandung.
- Palucomputer (2015). Pengertian Sejarah Dan Pendiri Facebook,<http://www.palucomputer.com/2015/06/pengertian-sejarah-dan-pendiri-facebook-terlengkap.html>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* , 1-2.

- Permatasari, P. A., Linawati, L., & Jasa, L. (2021). Survei Tentang Analisis Sentimen Pada Media Sosial. *Majalah Ilmiah Teknologi Elektro*, 20(2), 177. <https://doi.org/10.24843/mite.2021.v20i02.p01>
- Pisner, D. A., & Schnyer, D. M. (2019). Support vector machine. In *Machine Learning: Methods and Applications to Brain Disorders*. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
- Rifka R, N. (2017). *Step by Step Lancar Mambuat SOP*. Yogyakarta: Huta Publisher.
- Santosa, B. (2007). *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu : Yogyakarta.
- Sahputra, N. K. (2020). Mengklasifikasi Cyberbullying dari Media Sosial Twitter Menggunakan Algoritma Klasifikasi K-Nearest Neighbor dan Pendekatan Lexicon. Yogyakarta: Universitas Sanata Dharma.
- Subagyo, Pangestu, 1986, *Forecasting Konsep dan Aplikasi*, Yogyakarta, BPPE UGM
- Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). Combining lexicon-based and learning-based methods for twitter sentiment analysis. *HP Laboratories Technical Report*.
- Zulian, I. (2020). Analisis Pengaruh Islamophobia Terhadap Kebijakan Luar Negeri Amerika Serikat Di Pemerintahan Donald Trump. *Jurnal PIR : Power in International Relations*, 3(2), 140. <https://doi.org/10.22303/pir.3.2.2019.140-155>

LAMPIRAN 1. DATASET

Text	Label
@SarakinArt2 @ZukiAzuka @fyrishere islamophobia ðŸ’ž	Netral
@tvOneNews Kok nggak dibubarkanâ€œ ACT dibekukan Kemensos, islamophobia nich yeeee	Netral
@DMrsI_ @mengslo Kok ada ya manusia begini, mending deportasi aja ke korut bgst bgt	Netral
Kahmi Eropa Raya: Islamophobia Masih Menjadi Isu Utama di Beberapa Negara https://t.co/qkwxoc8jUD	Netral
Sesama maling dana umat harus saling melindungi, bukan begitu @musniumar @UAS_AbdulSomad @AdiHidayatLcMa @Pak_JK Taliban, ISIS, FPI, @PKSejahtera butuh dana dong. https://t.co/hkQPGKPs4 https://t.co/oEi6mnnb0P	Netral
also kalo kalian butuh referensi jurnal atau ebook mengenai islamophobia dan kehidupan muslim di prancis, bisa dm aku email kalian yaa nanti aku kirimin ðŸ’ž»»	Netral
Islamophobia suka cari gara2. Ntar kalo dibunuh, Islam dibilang teroris. Caper banget sih.. Insecure kah sama agamanya sendiri? ðŸ’ž† https://t.co/fLR9jvxiPF	Netral

<p>@Malukutoleransi @jaya2045 Gk ada islamophobia di sini. Yg ada kaum islamophilia yg mengklaim islam dizalimi trs membangun persepsi islam dng akal jahil mereka dan kefasikan</p>	<p>Netral</p>
<p>@HusinShihab @Hilmi28 Kelihatan sekali gerombolan islamophobia pada nongol dari gotd~,</p>	<p>Netral</p>
<p>@AzulZulian Saya pikir itu masih wacana beli Kpl selam usul UAS,Politik Islamophobia ðŸ’ª</p>	<p>Netral</p>
<p>@DwiHeriMulyant1 @PutraWi54049130 @Pakde_yan @PadmavatiPrati1 Ini Ngomongin soal Islamophobia , Wukuf Arofah , Perbedaan Tgl Iedul Adha ataw Pemerintahan Jokowi sihh..??</p>	<p>Netral</p>
<p>yg urgent itu UU ANTI ISLAMOPHOBIA &</p>	<p>Netral</p>
<p>Islamophobia https://t.co/plqSywHBQi</p>	<p>Netral</p>
<p>@kangdede78 @triajaaa Tomat kau bukan ustad, setan lebih jago dari kau, tpi bukan ustad sama dgn kau setan...@UAS_AbdulSomad @UAS_AbdulSomad @aagym @budimandjatmiko @gpansor @MUIPusat @DakwahPBNU @gusmusgusmu @GusBahaOnline @GUSBAHAOFFICIA1</p>	<p>Negatif</p>
<p>@Diah_Mario_ Sll saja standar Ganda diterapkan dinegara yg terpapar islamophobia.</p>	<p>Netral</p>
<p>@Dennysiregar7 Islamophobia katanya bang. Perasaan kita islam aman dan</p>	<p>Netral</p>

<p>tentram aja tu ibadah. Mereka yg buat isu, mereka sendiri yg panas ðŸ˜¸, ðŸ˜¸, ðŸ˜¸,</p>	
<p>@kangngabei @focus_or_die @detikcom Kalo si abu janda cs diproses hukumnya Baru gua percaya ni rezim menentang islamophobia. https://t.co/6u3WMgOk0f</p>	Netral
<p>Mayjen (Purn) Deddy S Budiman: Era Oligarki Neo Komunisme Propaganda Islamophobia Terstruktur, Sistematis dan Masif https://t.co/IRJJKx0t4K</p>	Netral
<p>@kangngabei @focus_or_die @detikcom Hahaha lu berani ga bilang agama si leopard itu agama teroris kek yg dilakukam si abu janda? ðŸ˜¸, ðŸ˜¸, ðŸ˜¸, kagak kan. Kelen emang senangnya menghina islam doang. Lu bagin dari Islamophobia itu. Bacot doang lu</p>	Netral
<p>@i_noside @Muhammad_Saewad Ternyata kadrun itu islamophobia ya bang karena menjatuhkan jati diri agama islam agar islam tetap di benci dengan cara jual agama, ngaku jadi ustad, cabul, dll https://t.co/Row8tgjP3d</p>	Netral
<p>@kangngabei @focus_or_die @detikcom Di amrik di eropa gada larangan un beribadah bagi muslim. Tp banyak muslim disana mengalami serangan islamophobia baik secara fisik maupun mental. Sama kek disini. Bukan sekali dua kali ust/kiyai yg diserang saat di mas</p>	Netral

<p>@ChusnulCh__ Anda dan gerombolan islamophobia koar2 nyinyirin pak Anies gimana ? Lempar batu sembunyi tangan...</p>	<p>Netral</p>
<p>@CNNIndonesia ACT memang ada penyimpangan tetapi tuduhan terorisme, kriminalisasi pembekuan dana dan lain-lainnya adalah modus rejim Islamophobia sudutkan umat sekaligus rampok dana umat.. Betul2 terlalu!</p>	<p>Netral</p>
<p>@_ekokuntadhi berita hoax yang di share oleh Ganjaris Islamophobia</p> <p>parah nih Ganjaris islamophobia seneng banget kalau nemu berita beginian</p>	<p>Netral</p>
<p>@__AnakKolong Kaum Islamophilia dodolan Islamophobia di negeri mayoritas muslim terbesar</p>	<p>Negatif</p>
<p>Islamophobia dgn provokasi.</p> <p>"Mobil pelaku dikejar dan ditabrak sehingga terbalik oleh 2 muslimah"</p> <p>https://t.co/5A513a3YE4</p>	<p>Netral</p>
<p>Waspadai Kelompok Radikal Pencipta Islamophobia. Maju Bersama IKN</p> <p>https://t.co/ynjMr3uPNN</p>	<p>Netral</p>
<p>@heatnotburn234</p> <p>@MNW_MNW_MNW Dapet donasi brp tu mon? Sudah dipake kah utk keliling dunia melawan wabah Islamophobia... Eehh btw, maksudmu dunia nyata atau keliling dunia maya jagad sosmed aja??</p>	<p>Netral</p>

<p>Maju Bersama IKN Waspadai Kelompok Radikal Pencipta Islamophobia https://t.co/QUWxJkPyBd</p>	<p>Netral</p>
<p>Paling lantang teriak "Islamophobia", eh dirinya malah phobia terhadap semua agama ðŸ¥²</p>	<p>Netral</p>
<p>@jaja9888 @gustratoo @mojokdotco nah ini. apalagi yg suka koar2 islamophobia. Di negara mayoritas muslim terbesar di dunia. Logis? memang minderan aja.</p>	<p>Netral</p>
<p>@2609The @AndaloKifli @hdsambodo @muchlis_ar Hei....saya termasuk yg tdk setuju dgn bom bunuh diri...meski hadiahnya menggiurkan,saya tdk setuju...dan yg membuat doktrin tsb,jauh dr ajaran Islam...kamu sudah termakan propaganda islamophobia!!! Tdk ada aya</p>	<p>Netral</p>
<p>@andhikatsu @AmanillahH Ya yang numpang numpang, semua ideologi kan gini, yg bikin Islamophobia makin kenceng juga isis sama taliban toh ðŸŸ~,</p>	<p>Netral</p>
<p>@kangdede78 Mereka2 ini yg membuat islamophobia. Kelakuan daj*al, pakai manusia berkedok agama.</p>	<p>Netral</p>
<p>@bogypratama25 @detikcom kasian sih polisinya. dilepas dibilang pengecut, ditangkap paksa tar dibilang kriminalisasi ulama dan santri, islamophobia dll sama â€˜gerombolanâ€™ mereka</p>	<p>Netral</p>

<p>@mgreantara @herricahyadi @pukismatcha Dakwah islam my ass, kalau menjual agama demi nilep duit sih tepat, pasang stempel islam sebagai facade kalau diserang mainkan kartu kalau pemerintah rakus duit dan tentunya islamophobiaðŸ˜,ðŸ˜,</p>	<p>Netral</p>
<p>Waspada Kelompok Radikal Pencipta Islamophobia!! Maju Bersama IKN https://t.co/z3eq2y1Do8</p>	<p>Netral</p>
<p>Sterilisasi Doktrin Radikal di Mimbar Agama Adalah Sterilisasi Akar Islamophobia Maju Bersama IKN https://t.co/bxFq8Usi8q</p>	<p>Netral</p>
<p>@UyokBack CECEB-CEBI memang di- design untuk menyerang yang berbau islam. Mereka islamophobia bayaran.</p>	<p>Netral</p>
<p>@korantempo mendadak kadrun pakai istilah "UMAT" lalu sok perhatian dengan kebocoran "DANA UMAT". Media #Islamophobia biasanya punya agenda yang justru menyerang ke- islamannya. sepatutnya umat Islam berhati-hati. Sudah 2 hari terus jadi headline. https://</p>	<p>Netral</p>
<p>Bagaimana bisa seorang muslim menjadi pengikut akun yang beraroma Islamophobia?</p> <p>Keilmuannya? Padahal bahkan ilmu seorang Islamophobia hanya digunakan untuk tipu daya.</p> <p>Akunnya bercentang Biru?</p>	<p>Netral</p>

<p>Preeetâ—ï Kebanggaan yang sesat.</p>	
<p>Bila suatu negeri dipimpin oleh seorang berideologi komunis maka tindakan gagal yg dilakukan pemimpinnya tsbt akan ditutupi dg isu2 Sperti: Covid. Terorisme, radikalisme/Islamophobia, dll. Untuk menghindari bullyang rakyat. @KSPgoid @luhut_binsar @jokowi</p>	<p>Netral</p>
<p>@qid_23 Sakai gila. Dia taktau ke kat US behavior mcm ni org buat out of islamophobia. Depa tu dah sah2 bukan islam. Yang dia ni dah kenapa?</p>	<p>Netral</p>
<p>@MNW_MNW_MNW Mereka yang ngotot itu coba ditelisik lebih detail..paling ga mereka bukan donatur dan cenderung termasuk gerombolan islamophobia</p>	<p>Netral</p>
<p>Udah bawaan orok otak islamophobia https://t.co/taIp72qId9</p>	<p>Netral</p>
<p>@dnayza1212 @AT_AbdillahToha Ada aroma kodok islamophobia</p>	<p>Netral</p>
<p>Wajah monyong kayak Somad @UAS_AbdulSomad ini hatinya buruk.suka bohong.suka adu domba. suka berkhianat https://t.co/nFGEmHb66C</p>	<p>Negatif</p>
<p>@DefrizonZ @Gus_Ariffin @MUIPusat Maaf saya tegaskan Islamophobia itu gak ada di Indonesia, apa yg ditakutkan sama Islam orang di Indonesia itu 80% beragama Islam. Kalo di negara muslim mungkin itu bener. Justru yg di Indonesia itu Kristenphobia. Saya Musl</p>	<p>Netral</p>

@DefrizonZ @Gus_Arifin @MUIPusat Gimana ceritanya islamophobia? Masa kaum muslim Indonesia islamophobia? Masa jeruk makan jeruk?	Netral
Kalo Islamophobia ada mah gak bakalan ada yg berani minta sumbangan bangun masjid di jalanan đŸŽ https://t.co/qhboHp4Yix	Netral
ISLAMOPHOBIA sebenarnya berawal dari orang2 yg ngaku2 Islam. https://t.co/UZA5XC41IB	Netral
Nama aslinya Robi Nur Huda.. Gara2 zionis yg islamophobia namanya diubah menjadi Robin Hood... Biar berbau kebarat baratan... Itulah kejahatan kaum kafir... Takbirrr. .. Kumpulkan donasi melawan islamophobia, ibadah kalian, demi kemewahan kami. Takbirrr Ttd ngustad Kadrun. https://t.co/remljg9NaY	Netral
@yaniarsim Dikit2 islamophobia, padahal sih cuma ini orang aja sebenarnya males nyari fakta	Netral
@bbbekti pasti ini ngomongin WNI Singapura kan?	Netral
Mau jadi wasit gak usah jauh jauh damaikan dl kericuhan di NKRI krn mulut" usil buzzerp dn para pejabat islamophobia https://t.co/q6HiJzgerz	Netral
@tanyakanrl Quora sekarang udah kayak Twitter, banyak buzzer islamophobia. Isinya anak-anak edgy. Males maen Quora, tulisan gw diserang mulu ama buzzer tolol	Netral