

PROPOSAL SKRIPSI
ISLAMOFOBIA DI INDONESIA: *INSIGHTS* DAN MODEL
KLASIFIKASI DARI MEDIA ELEKTRONIK
MENGGUNAKAN *SUPPORT VECTOR MACHINE*



Disusun oleh :
Izzuddin
11190940000052

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH
JAKARTA
2022

PROPOSAL SKRIPSI

Islamofobia di Indonesia: *Insights* dan Model Klasifikasi dari Media Elektronik Menggunakan *Support Vector Machine*

Diajukan oleh :

Izzuddin

11190940000052

Usulan Dosen Pembimbing

Dosen Pembimbing I

Bapak Taufik Sutanto M.Sc.Tech., Ph.D

NIP 197905302006041002

Bapak Taufik merupakan dosen Matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta. Keahlian beliau adalah Komputasi Statistik. Penelitian pada tahun 2020 berjudul, “*Analisa Historis Karakteristik Percakapan Media Sosial Masyarakat Indonesia dalam Membahas Topik Agama Menggunakan Topic Modelling*” menginspirasi penulis untuk turut meneliti perkembangan topik mengenai agama, terutama terkait Islam.

Dosen Pembimbing II

Ibu Madona Yunita Wijaya M.Sc.

NIP 198506242019032007

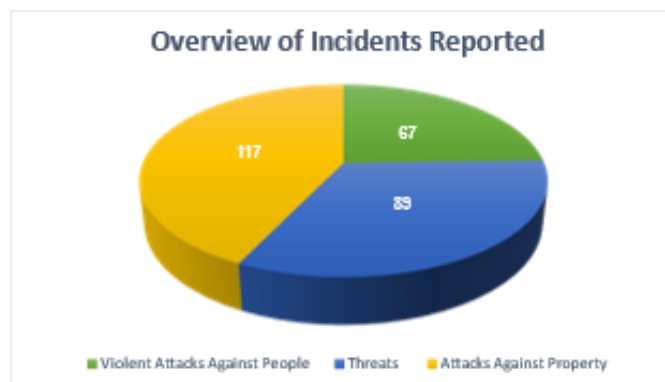
Ibu Madona merupakan dosen Matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta. Keahlian beliau adalah Statistik dan Data Sains. Beberapa penelitian terbaru beliau, yaitu “*Relative Importance Analysis for Psychological Research*” (2021), “*Energy Saving Potential Prediction and Anomaly Detection in College Buildings*” (2020), dan “*World Gold Price Forecast using APARCH, EGARCH and TGARCH Model*” (2020).

BAB I. LATAR BELAKANG

A. *Perumusan Masalah*

Setelah Perang Dunia II, negara-negara Eropa “mengimpor” tenaga kerja dari luar (negara yang mayoritas penduduk dan budayanya berpedoman pada ajaran Islam, seperti Aljazair, Maroko, India, dan Turki) untuk membangun kembali negara mereka yang hancur setelah Perang Dunia II. Para tenaga kerja yang kebanyakan merupakan tenaga kerja asing (TKA) dan keluarganya berselisih dalam kehidupan sosial di negara tempat mereka bekerja, hal ini sering menimbulkan konflik di masyarakat yang berujung pada kerusuhan dan kekerasan. Terkadang kejadian ini menimbulkan stigma negatif terhadap Islam yang lambat laun berkembang menjadi ketakutan terhadap Islam atau dikenal dengan Islamofobia [1]. Namun, baru semakin menjadi perbincangan setelah serangan teroris di Menara Kembar World Trade Center di pusat kota New York City pada 11 September 2001. Sejak 9/11, serangan ini dilakukan secara intensif melalui tak terhitung banyaknya media, artikel, surat kabar, jurnal, film, dan film dokumenter [2].

Hingga lebih dari dua dekade terakhir, masih banyak terjadi kasus-kasus Islamofobia di dunia. Mengutip dari hatecrime.osce.org, terdapat total 333 kejadian kebencian Islam yang terjadi di 23 negara di tahun 2020 yang dirincikan pada Gambar 1 [3].



Gambar 1. Detail 333 Kasus yang Terjadi di 23 Negara (2020).

Melalui keterangan lebih lanjut sebuah laporan berjudul “*European Islamophobia Report 2020*”, Islamofobia mulai berpindah ke lingkup yang lebih privasi dan merambat ke sosial media [4]. Kejadian terbaru di tahun 2022 melibatkan sebuah kasus kebencian terhadap Islam yang menghina Nabi Muhammad melalui media sosial di India oleh pemimpin pemuda dari Partai Bharatiya Janata Party (BJP). Selanjutnya, kejadian ini memicu protes dari negara-negara Muslim, seperti Qatar, Arab Saudi, UEA, Oman, Indonesia, Malaysia, Pakistan, Iran, dan Afghanistan untuk menuntut permintaan maaf [5].

Dari beberapa kasus yang terkuak dan menjadi sorotan, kebanyakan dari mereka berasal dari negara dengan populasi mayoritas non-muslim. Lalu, bagaimana kondisi Islamofobia di negara dengan mayoritas muslim terbesar? Apakah hal-hal tersebut juga terjadi dan mengapa bisa terjadi?

Indonesia merupakan negara mayoritas Muslim dengan jumlah penganutnya sebesar 231,06 juta penduduk atau setara 86,7% dari total populasi [6]. Meski begitu, ada beberapa kasus Islamofobia yang sering ditemukan dan dilakukan secara terbuka. Misalnya, pada tahun 1963-1964, terdapat hujatan terhadap Allah SWT, Alquran, Rasulullah SAW secara terbuka yang dilakukan oleh Lembaga Kebudayaan Rakyat (LEKRA). Kemudian, terdapat kasus penghinaan terhadap Nabi Muhammad di pers pada zaman Orde Baru. Hingga pada tahun 2020, diterbitkannya sebuah buku yang memuat hinaan serta pelecehan secara ekstrem terhadap Nabi Muhammad SAW di beberapa halaman [7].

Indonesia sendiri tidak membatasi adanya kebebasan berekspresi dan berpendapat, terutama di media elektronik. Kebebasan ini memiliki banyak aspek positif dan konstruktif. Kendati demikian, hal ini juga menebarkan bias atau maksud negatif yang membawa ke arah konflik. Oleh karena itu, diperlukan penelitian lebih lanjut terkait pola-pola yang terjadi dalam wacana-wacana yang tersebar di internet mengenai Islamofobia. Berdasarkan beberapa penelitian sebelumnya, *machine learning*, sebuah metode berbasis komputasi, mampu menangani masalah tersebut secara efektif dan efisien. Davidson et al. menggunakan feature-based model klasifikasi tradisional yang menggabungkan *distributional* TF-IDF dan fitur-fitur linguistik lain menggunakan *Support Vector Machines* (SVM). Mereka menggunakan

tiga label: *hate speech*, *offensive*, dan *neither hate speech nor offensive* [8]. Di lain sisi, Qasim et al berhasil mengklasifikasikan teks menjadi Islamofobia atau bukan menggunakan *deep learning* [9]. Berangkat dari penelitian-penelitian tersebut, peneliti ingin mengungkap lebih jauh tentang konten digital, khususnya di Twitter, untuk meningkatkan kesadaran dan memberikan klarifikasi terkait Islamofobia dan ujaran kebencian Islam yang, secara umum, muncul di Indonesia melalui situs internet dengan menggunakan metode *Support Vector Machine* (SVM).

B. Manfaat dan Tujuan Penelitian

Manfaat dan tujuan penelitian ini adalah sebagai berikut:

1. Mengungkap ujaran-ujaran kebencian terhadap Islam yang terdapat di media elektronik;
2. Mengungkap alasan-alasan di balik penyampaian ujaran kebencian yang ada di media elektronik;
3. Mengklasifikasikan teks menjadi berindikasi Islamofobia dan netral atau tidak berkaitan dengan Islamofobia; dan
4. Membuat model yang bisa digunakan untuk menghasilkan insight yang dapat digunakan untuk mendesain kampanye berbasis data dalam mengatasi dan mencegah Islamofobia di masa depan.

BAB II. TINJAUAN PUSTAKA DAN LANDASAN TEORI

A. *Sentiment Analysis*

Sentiment analysis atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sikap, penilaian, dan emosi seseorang berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu [10]. Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian mengkategorikan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif, atau netral [11]. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

B. *Twitter*

Twitter adalah suatu media elektronik yang merupakan layanan dari *microblog*, yang memiliki pengertian suatu bentuk blog yang membatasi ukuran setiap kirimannya di mana memberikan fasilitas bagi pengguna untuk dapat menuliskan pesan dalam Twitter update hanya berisi 140 karakter. Dengan kemudahan Twitter dan waktu pemrosesan aplikasi yang cepat, informasi yang terdapat di dalamnya dengan mudah menyebar luas [12]. Kemudian, fitur-fitur yang terdapat dalam Twitter adalah:

1. Halaman Utama (*Home*)

Merupakan halaman utama untuk melihat kiriman yang dikirimkan oleh orang-orang yang sudah berteman dengan pengguna atau yang pengguna ikuti (*Following*). Selain itu, pengguna juga bisa melihat kiriman akun pengguna lain selama akun tersebut tidak dikunci saat mengunjungi halaman profil mereka.

2. Profil (*Profile*)

Halaman ini yang bisa dilihat oleh semua orang, untuk pengaturan public, yang berisi data diri serta kiriman yang sudah pernah dituliskan atau pernah dikirim. Kiriman dapat berbentuk teks, gambar, atau video.

3. *Followers*

Disebut juga pengikut, merupakan pengguna lain yang tertarik untuk berteman atau sekedar mengikuti aktivitas pengguna lain. Pengguna lain yang menjadi pengikut akun seseorang, maka Tweets seseorang yang ia ikuti tersebut akan muncul di halaman utamanya.

4. *Following*

Merupakan akun seseorang yang mengikuti akun pengguna lain agar Tweets yang dikirim oleh orang yang diikuti bisa masuk ke dalam halaman utama Twitternya.

5. Penyebutan (*Mention*)

Isitilah penyebutan (*mentions*) merupakan balasan dari percakapan agar sesama pengguna bisa langsung “menandai” dengan mudah orang yang akan diajak bercakap.

6. Tanda Pagar (*Hashtag*)

Simbol “#” atau bisa dibilang hashtag yang ditulis didepan suatu topik tertentu yang bertujuan agar pengguna lain atau akun lain bisa mencari topik yang sama sesuai yang dicari dengan hashtag itu.

7. Topik Viral (*Trending Topic*)

Hal ini bisa dilihat berdasarkan peringkat yang membicarakan topik yang sama dalam waktu bersamaan dan di dalam suatu daerah.

C. Text Mining

Text mining (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda [13]. *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, pengklasteran, ekstraksi informasi, dan pengambilan informasi [14].

Pada dasarnya, proses kerja *text mining* banyak mengadopsi dari penelitian *data mining* namun yang menjadi perbedaan adalah pola yang digunakan oleh *text mining* diambil dari sekumpulan bahasa alami yang tidak terstruktur, sedangkan dalam *data mining* pola yang diambil dari *database* yang terstruktur [15]. Tahap-tahap *text mining* secara umum adalah *text preprocessing* dan seleksi fitur [13].

1. *Text Preprocessing*

Tahap *text preprocessing* adalah tahap awal dari *text mining*. Tahap ini mencakup semua aktivitas dan proses untuk menyiapkan data yang akan digunakan pada kegiatan *text mining* [13]. Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing*, yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan pemisah-pemisah, seperti tanda titik (.), koma (,), spasi, dan karakter angka yang ada pada kata tersebut [14].

2. Seleksi Fitur

Tahapan seleksi fitur bertujuan untuk mengurangi dimensi dari suatu kumpulan teks atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak memiliki arti bermakna sehingga proses pengklasifikasian menjadi lebih efektif dan akurat. Pada tahap ini tindakan yang dilakukan akan menghilangkan *stopwords* dan *stemming* terhadap kata yang berlebihan [13].

D. Vector Space Model

Data yang digunakan pada penelitian adalah berupa data teks di mana termasuk pada data tidak terstruktur. Sedangkan, komputer hanya mampu mengolah data terstruktur yang biasanya berbentuk tabular. Maka dari itu, agar data ini bisa diolah oleh komputer, data perlu dikonversi menjadi angka. Proses inilah yang dimainkan peran oleh VSM yang diperlukan untuk mengubah dokumen menjadi nilai-nilai vector.

E. Support Vector Machine

Support vector machine (SVM) adalah sebuah metode pembelajaran yang menganalisis data dan mengenal pola, digunakan untuk klasifikasi dan analisis regresi. SVM diperkenalkan oleh Vladimir Vapnik, Boser dan Guyon. Kemudian, dipresentasikan di *Annual Workshop on Computational Learning Theory* pada tahun 1992. Ide dasar SVM adalah melakukan transformasi data pembelajaran ke ruang berdimensi lebih tinggi (ruang fitur) dan menemukan sebuah *hyperplane* yang memisahkan data latih menjadi dua kelas yang berbeda dengan margin maksimum [16]. *Hyperplane* adalah bidang datar penentu yang menjadi pemisah antara kelas-kelas pada data di dimensi ke- n dan margin adalah jarak antar *hyperplane* dengan data terdekat (*support vector*) dari masing-masing kelas.

F. Evaluasi Model

Evaluasi model perlu dilakukan untuk mengetahui seberapa baik suatu model dapat mengklasifikasikan suatu kelas. Salah satu cara untuk melakukan hal tersebut adalah dengan menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah tabel berdimensi 2x2 yang menyatakan berapa banyak data uji yang benar dan salah diklasifikasikan. Parameter yang digunakan pada data uji yaitu TP (*true positive*), FN (*false negative*), TN (*true negative*), dan FP (*false positive*) [17].

BAB III. METODOLOGI

Adalah bagian yang menjelaskan bagaimana penelitian akan dijalankan. Jelaskan tentang rencana jalannya penelitian secara detail baik masalah data apa saja yang akan diambil dan alat apa yang akan digunakan.

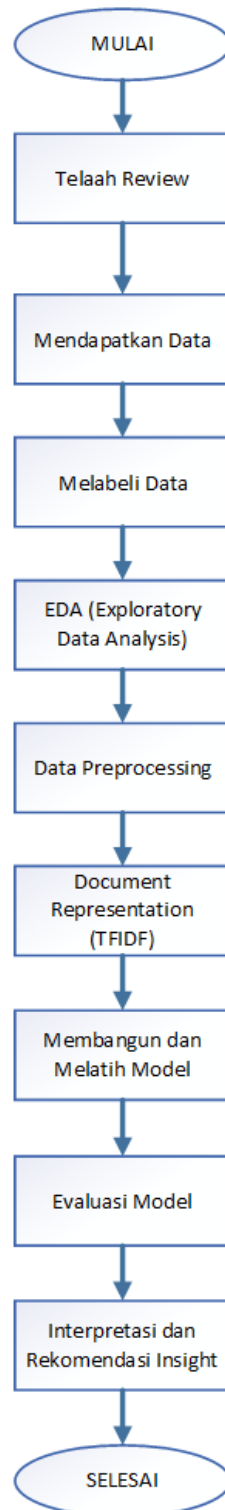
A. Sumber dan tipe data

Data yang akan digunakan dalam penelitian ini merupakan data sekunder yang berupa cuitan-cuitan yang dibagikan melalui Twitter dan judul-judul berita yang didapatkan melalui beberapa situs di internet.

B. Usulan Metode

Penelitian ini akan menggunakan metode *Support Vector Machine* (SVM).

C. Diagram alur penelitian



Gambar 2. Diagram Alur Penelitian.

BAB IV. JADWAL PENELITIAN

Penelitian direncanakan akan dilaksanakan selama enam bulan. Rincian rencana jadwal penelitian dicantumkan dalam Tabel 1.

No.	Keterangan	Bulan					
		1	2	3	4	5	6
1.	Studi pustaka						
2.	Mendapatkan data						
3.	Penulisan 'Pendahuluan'						
4.	Melabeli Data						
5.	Penulisan 'Tinjauan Pustaka'						
6.	EDA						
7.	Penulisan 'Metodologi Penelitian'						
8.	Preprocessing						
9.	Membangun dan Melatih Model						
10.	Evaluasi Model						
11.	Optimasi Model						
12.	Interpretasi dan Rekomendasi						
13.	Penulisan 'Hasil dan Pembahasan'						
14.	Penulisan 'Penutup'						
15.	Penulisan 'Daftar Pustaka' dan 'Lampiran'						

Tabel 1. Jadwal Penelitian.

DAFTAR PUSTAKA

- [1] F. Nugraha, “Latar Belakang Munculnya Islamophobia di Kalangan Masyarakat Eropa,” 2012. .
- [2] K. Dikarma, “Tragedi 9/11 dan Perjuangan Muslim AS Lawan Stereotip,” *Republika*, 2021. .
- [3] OSCE ODIHR, “Anti-Muslim hate crime,” *OSCE ODIHR HATE CRIME REPORTING*, 2020. [Online]. Available: <https://hatecrime.osce.org/index.php/anti-muslim-hate-crime>. [Accessed: 09-Jun-2022].
- [4] S. Sevencan, “COVID-19 pandemic deepens online Islamophobia in Europe,” *AA*, 2021. [Online]. Available: <https://www.aa.com.tr/en/europe/covid-19-pandemic-deepens-online-islamophobia-in-europe/2462001>. [Accessed: 09-Jun-2022].
- [5] “Kembali Komentar Anti-Islam, Polisi India Tangkap Pimpinan Pemuda Partai Berkuasa,” *VoA*, 2022. [Online]. Available: <https://www.voaindonesia.com/a/kembali-komentar-anti-islam-polisi-india-tangkap-pimpinan-pemuda-partai-berkuasa-/6607952.html>. [Accessed: 09-Jun-2022].
- [6] The Royal Islamic Strategic Studies Center, *The Muslim 500 - The World's 500 Most Influential Muslim 2022*. Jordan National Library, 2021.
- [7] N. Nashrullah, “Bukti Islamofobia Juga Terjadi di Indonesia, Sikap Umat?,” *Republika*, 2020. [Online]. Available: <https://www.republika.co.id/berita/qivsyf320/bukti-islamofobia-juga-terjadi-di-indonesia-sikap-umat>. [Accessed: 09-Jun-2022].
- [8] H. Khan, “Language Agnostic Model: Detecting Islamophobic Content on Social Media,” Middle Tennessee State University, 2021.
- [9] Q. Mehmmod, I. Siddiqi, and A. Kaleem, “Islamophobic Hate Speech Detection from Electronic Media using Deep Learning,” no. January, p. 15, 2022.
- [10] H. S. Maning, D. C., P Raghavan., *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.
- [11] A. H. Wang, “Don’t Follow Me : Spam Detection in Twitter.”

- IEEE*, no. Security and Cryptography, 2011.
- [12] N. E. Setyani, "Penggunaan Media Sosial Sebagai Sarana Komunikasi Bagi Komunitas," *J. Komun. Univ. Sebel. Maret*, 2013.
 - [13] J. S. Feldman, R., *The Text Mining Handbook: Advance Approaches in Analyzing Unstructured Data*. New York: Cambride University Press, 2007.
 - [14] J. K. Berry, M W., *Text Mining Applications and Theory*. Sessex: Wiley.
 - [15] J. P. Han, J., M Kamber., *Data Mining : Concepts and Techniques*. Waltham: Morgan Kaufmann Publisher, 2011.
 - [16] W. M. Andrew, "Support Vector Machine.," Carniege, 2011.
 - [17] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier," *Semin. Nas. Apl. Teknol. Inf.*, vol. 20, pp. 5–10, 2014.