

Implementasi Algoritma Support Vector Machine untuk Analisis Sentimen LGBT di Indonesia

Mustakim*, Muhammad Ridwan, Nanda Try Luchia

Prodi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Puzzle Research Data Technology, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Jl. H.R Soebrantas No 155 KM.15 Simpang Baru Panam Pekanbaru 28293, Riau, Indonesia

E-mail: ^{1,*}mustakim@uin-suska.ac.id, ²muhammad1ridwan7@gmail.com, ³12050320445@students.uin-suska.ac.id

Email Penulis Korespondensi: mustakim@uin-suska.ac.id

Submitted: 99/99/9999; Accepted: 99/99/9999; Published: 99/99/9999

Abstrak—Kasus LGBT mulai muncul secara terbuka di Indonesia terjadi di tahun 2016. Kasus ini banyak mendapat perbincangan pada tahun tersebut hingga sampai saat ini karena banyaknya dari masyarakat yang berkomentar setuju dan tidak setuju terhadap aksi, kegiatan, serta adanya nya gender LGBT tersebut di Indonesia. Sentimen dari komentar masyarakat tersebut banyak merujuk berbagai aspek kehidupan sehingga menghasilkan opini masyarakat yang bersifat positif, negatif dan netral. Melihat hal ini, perlu dilakukan klasifikasi serta analisis dari sentimen tweet untuk melihat kecenderungan dari masing-masing opini masyarakat. Analisis dan klasifikasi dilakukan dengan teknik pengolahan data text mining menggunakan algoritma Support Vector Machine (SVM). Proses klasifikasi dilakukan secara 3 tahap dengan pembagian data 90%:10%, 80%:20% dan 70%:30% dengan menggunakan 3 kernel yaitu linier, polinomial dan Radial Basic Function (RBF). Hasil klasifikasi yang diperoleh dari ke tiga kernel tersebut menunjukkan kecenderungan pandangan masyarakat terhadap kasus LGBT adalah negatif dan netral yang di tunjukkan dengan akurasi tertinggi pada kernel linear dan RBF. Percobaan SVM menghasilkan akurasi 74% pada kernel linier dengan percobaan data 90%:10% dan 74% dan pada kernel RBF dengan C=100 gamma=0,01. Pengelompokan data sentimen tweet ini menghasilkan analisis kecenderungan tidak mendukung atau tidak setuju pada gender LGBT karena tidak sesuai dengan landasan yang telah ditetapkan di negara Indonesia yang lebih mengutamakan aspek keagamaan dari pada aspek yang lainnya.

Kata Kunci: Analisis Sentimen, LGBT, Support Vector Machine, Text Mining, Twitter

Abstract—LGBT cases began to appear openly in Indonesia in 2016. This case has received a lot of discussion in that year until now because of the number of people who commented agreeing and disagreeing with actions, activities, and the existence of the LGBT gender in Indonesia. The sentiments from the community's comments refer to various aspects of life so as to produce community opinions that are positive, negative and neutral. Seeing this, it is necessary to perform a classification and analysis of tweet sentiments to see the tendency of each community's opinion. Analysis and classification is done with text mining data processing techniques using the Support Vector Machine (SVM) algorithm. The classification process is done in 3 stages with the division of data 90%:10%, 80%:20% and 70%:30% using 3 kernels namely linear, polynomial and Radial Basic Function (RBF). The classification results obtained from the three kernels show that the tendency of society's view of LGBT cases is negative and neutral which is shown with the highest accuracy on the linear and RBF kernels. The SVM experiment produced an accuracy of 74% on the linear kernel with 90%:10% and 74% data experiments and on the RBF kernel with C=100 gamma=0.01. The grouping of this tweet sentiment data resulted in an analysis of the tendency not to support or disagree with the LGBT gender because it is not in accordance with the established basis in the country of Indonesia which prioritizes religious aspects over other aspects.

Keywords: Sentiment Analysis; LGBT; Support Vector Machine; Text Mining; Twitter

1. PENDAHULUAN

Lesbian, Guy, Bisexual and Transgender (LGBT) adalah suatu isu kejadian peristiwa yang menjadi dipembicaraan di masyarakat Indonesia melalui media sosial karena kaum ini telah berani muncul secara terbuka dengan sejumlah kegiatan mereka seperti kegiatan pengakuan terhadap mereka berupa promosi dan iklan yang mencuat di media sosial [1]. Perusahaan bernama Swan Statistic Consultan yang bergerak dibidang data analitic dan survey metodologi telah melakukan survey dan mengatakan bahwa penyakit HIV/AIDS di indonesia di tahun 2016 mencapai 28.640 jiwa. Jumlah ini terus meningkat dari beberapa tahun setelahnya dengan penyumbang angka terbanyak berasal dari kaum LGBT, hal ini sampaikan langsung oleh CEO perusahaan tersebut yaitu Hasan Azhari. Tidak hanya sampai di tahun tersebut, tahun 2019 juga menunjukkan jumlah LGBT terus meningkat yang dibuktikan di beberapa provinsi di Indonesia seperti Bali melalui data dinas kesehatannya tercatat 21.018 mengidap penyakit HIV/ AIDS, dan wilayah Sumatera Barat yang mencapai 18.000 orang [2]. Karena kasus ini sudah berani secara terbuka bermunculan hal ini membuat masyarakat hingga ormas islam melakukan berbagai penolakan salah satunya MUI yang telah membentuk fatwa bahwa aktifitas LGBT diharamkan dan bertentangan dengan Pancasila dalam sila pertama dan UUD 45 pasal 29 ayat 1 [3], namun dikeadaan yang sebenarnya, gerakan LGBT masih tetap ada, hal ini disebabkan oleh RUU KUHP pasal 292 yang membuat kaum LGBT masih tetap menjalankan gerakannya [4].

LGBT adalah sebuah singkatan yang memiliki arti lesbian yaitu perempuan yang menyukai dan mencintai perempuan, guy yaitu laki-laki yang mencintai dan menyukai laki-laki, bisexual yaitu mencintai dan menyukai dengan cara menjalin hubungan asmara dengan laki-laki dan juga perempuan, transgender yaitu ketidaksesuaian identitas gender saat dilahirkan [16]. Gerakan tersebut menimbulkan banyak komentar seperti yang mendukung atau sebaliknya di media sosial. Dengan kata lain, media sosial sekarang banyak dijadikan sumber informasi utama sekaligus tempat memberikan komentar terhadap suatu topik dari kejadian yang muncul dan sedang terjadi atau sering juga disebut dengan trending

informasi, termasuk informasi mengenai LGBT. Media Sosial juga digunakan sebagai media meluapkan komentar berupa pendapat yang berbentuk sikap emosional yang ditunjukkan baik itu berbentuk lisan atau tulisan adalah bentuk defenisi sentimen [5]. Hal ini yang membuat munculnya penelitian ini dengan tujuan untuk melihat bagaimana dan seberapa banyak pandangan masyarakat Indonesia terhadap gerakan LGBT dan sejenisnya. Dalam penelitian ini hal-hal yang ingin dibuktikan yaitu jumlah dari tweet yang mendukung, menolak atau memilih untuk acuh dari tweet masyarakat Indonesia terhadap gerakan LGBT melalui media sosial Twitter dengan menggunakan teknik text mining.

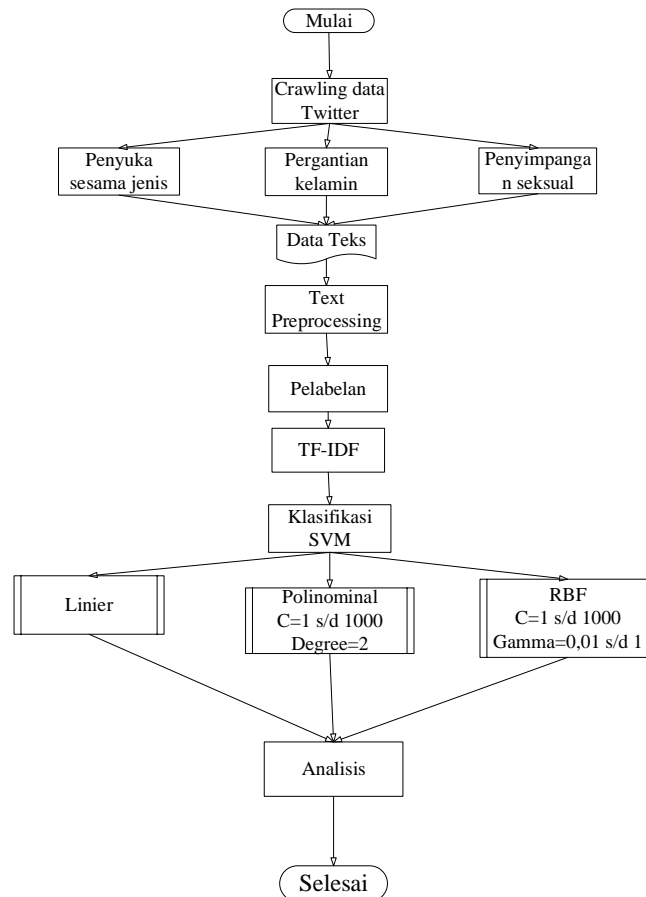
Text mining yaitu proses penambangan data berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah untuk mencari kata-kata yang dapat mewakili isi dari dokumen, sehingga dapat dilakukan analisis keterhubungan antar dokumen [6]. Penelitian terkait text mining sosial media twitter dilakukan oleh Namugera dkk tentang penentuan sentimen terhadap media cetak dan non cetak di negara Uganda pada media sosial twitter dengan melihat polaritas positif dan negatif pada media-media tersebut [28]. Hasil dari penelitian ini menjelaskan bahwa media non cetak lebih baik dari pada media cetak namun begitu, untuk perolehan informasi yang baik, pembaca media noncetak juga harus memperhatikan media cetak sebagai sikap tindak lanjut pengetahuan yang terbatas dari media cetak [28]. Penelitian lain terkait dengan twitter juga pernah dilakukan oleh Watequlis tentang analisis sentimen feedback masyarakat terhadap kebijakan pemerintah dengan menggunakan algoritma Naive Bayes yang bertujuan untuk mengetahui tanggapan positif, negatif dari publik terhadap kebijakan pemerintah dan hasil dari penelitian ini menunjukkan feedback masyarakat terhadap pemerintah cenderung negatif hal ini dibuktikan dengan akurasi 84% negatif, 47% positif [29]. Penelitian lainnya juga pernah dilakukan oleh Effendy dkk tentang menganalisis pendapat masyarakat terhadap penggunaan transportasi di kota-kota besar Indonesia dengan menggunakan algoritma SVM, hasil penelitian tersebut menunjukkan 78% masyarakat berpendapat layak untuk dijadikan sebagai sarana transportasi dari pada memakai kendaraan pribadi [30]. Selain itu beberapa kasus yang pernah menjadi trending di Indonesia telah diteliti oleh Ghulam tentang topik sentimen analisis dalam pemilihan gubernur Jawa Timur dengan tujuan untuk menganalisis pendapat masyarakat terhadap kandidat calon gubernur Khofifah dan Gusipul serta algoritma yang gunakan yaitu Naive Bayes (NBC) yang menghasilkan akurasi sebesar 77% untuk Khofifah dan 76% untuk Gusipul [26], di tahun yang sama juga terdapat pembahasan yang trending yaitu tentang prediksi pemilihan kepala negara atau presiden Republik Indonesia yang dilakukan oleh Budi Harto dan Meiliana dengan tujuan menganalisis dan memprediksi pendapat masyarakat pada kandidat yang akan menjadi presiden Indonesia antara Jokowi dan Prabowo melalui Twitter [27]. Hasil dari penelitian tersebut menunjukkan bahwa analisis sentimen dari tagar terpopuler positif terbanyak ditujukan kepada Jokowi sebesar 40 tagar sedangkan tagar yang ditujukan untuk kandidat Prabowo hanya memiliki 16 tagar populer saja [27].

Terkait dengan text mining, terdapat algoritma yang sering digunakan dalam melakukan klasifikasi terhadap text classification salah satunya adalah Support Vector Machine (SVM), yang merupakan algoritma yang bisa digunakan untuk melakukan berbagai masalah klasifikasi [22], proses klasifikasi ini sudah diimplementasikan diberbagai bidang seperti analisis resiko kredit, diagnosa medis, kategorisasi text, ekstraksi informasi dan sebagainya [23]. Penelitian yang berkaitan dengan SVM terhadap kasus text mining pernah dilakukan oleh Ahmad dkk terkait dengan keefektifan analisis sentimen tweet pada produk apel dan mobil tanpa pengemudi dalam mendeteksi polaritas dari text secara textual dengan hasil kinerja akurasi yang dihasilkan sebesar 59,1% untuk mobil tanpa pengemudi dan 71,2% untuk produk apel [21]. SVM dianggap lebih baik daripada algoritma KNN karena proses algoritma ini sangat kompleks sehingga hasil akurasi klasifikasi algoritma ini tidak bergantung pada ukuran dari sebuah data set tetapi hasil akurasi dihasilkan bergantung kepada siklus saat training yang dihasilkan oleh parameter dari kernel dan algoritma sangat cocok digunakan untuk kasus text mining [20]. Oleh karena itu, dalam penelitian ini akan membuktikan akurasi terbaik SVM pada kasus klasifikasi tweet LGBT di Indonesia yang terdiri dari kelas negatif, positif atau netral.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

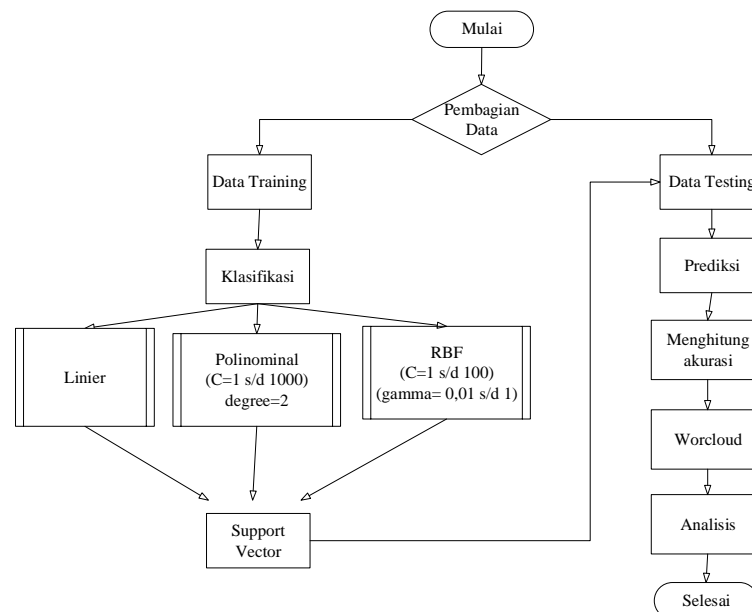
Adapun langkah-langkah yang ditempuh ini dapat dilihat pada gambar 2.



Gambar 1. Metodologi Penelitian

Gambar 1 merupakan alur penelitian yang dimulai dari pengolahan data teks hingga proses analisis data. Penelitian ini berawal dari pengambilan data teks pada media sosial twitter dengan menggunakan teknik penambangan data yang disebut dengan *crawling*. Proses ini dilakukan dengan memanfaatkan bahasa pemrograman python untuk menambang data tersebut. Pengambilan data ini pada twitter dilakukan dengan cara memasukkan kata kunci (*key*), atau #dari kalimat teks yang ingin didapatkan.

Proses terpenting dari text mining adalah preprocessing, ini dilakukan secara berurutan yang diawali dengan proses Case Folding, Cleaning, Filtering hingga Stemming [18]. Selanjutnya merupakan proses TF-IDF. Proses ini berjalan dengan cara memisah masing kalimat menjadi beberapa potongan kata dan memberikan bobot pada kata tersebut sesuai dengan banyaknya frekuensi besar atau tingginya dari sebuah kata tersebut muncul pada setiap dokumen guna untuk melihat tingkat kepentingan kata yang di cari dalam sebuah dokumen [25]. Proses klasifikasi data teks dapat dilaksanakan. Klasifikasi data teks LGBT ini menggunakan algoritma SVM dengan menggunakan 3 kernelnya yaitu linier, polinomial dan RBF. Proses ini dapat dilihat pada alur flowchart pada gambar 2.



Gambar 2. Alur Proses Text Mining dengan SVM

Dalam tahapan klasifikasi ini data teks akan dikelompokkan berdasarkan label sentimen dari masing-masing kalimat dengan cara menentukan hyperplane (pemisah) yang maksimum dari data teks tersebut sehingga data tersebut akan dipisahkan oleh daerah yang disebut dengan margin. Selain itu proses klasifikasi ini juga berpengaruh terhadap kedekatan data *testing* dengan data *support vector* yang ditentukan dari masing-masing label.

2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Proses pembobotan kata ini berguna untuk memberikan nilai pada kata berdasarkan frekuensi kemunculan kata tersebut [11]. Tahapan TF-IDF ini akan dilakukan jika proses stemming dan stopwords removal telah dilakukan [12]. Dalam skema arsitektur arus informasi, adanya sebuah proses yang berkaitan dengan pembobotan kata atau term baik secara lokal maupun global yang hanya berpedoman pada frekuensi munculnya term dalam suatu dokumen, dan juga tidak melihat adanya frekuensi timbulnya term tersebut di dalam dokumen lainnya [13]. Pembobotan lokal hanya berpedoman pada frekuensi munculnya term dalam suatu dokumen dan tidak melihat frekuensi kemunculan term tersebut di dalam dokumen lainnya, sedangkan pembobotan global digunakan supaya sebuah kata tersebut memiliki tekanan yang dapat mengakibatkan perbedaan yang diukur menurut penyebarannya di seluruh dokumen [13]

2.3 Support Vector Machine (SVM)

Support vector machine (SVM) adalah suatu algoritma yang handal dalam menyelesaikan masalah klasifikasi data. Permasalahan SVM dipecahkan dengan menyelesaikan persamaan Lagrangian yang merupakan bentuk dual dari SVM melalui quadratic programming [14]. SVM pertama kali dikenalkan oleh Vapnik pada tahun 1992 pada saat dipresentasikan di Annual Workshop on Computational Learning Theory [15]. Prinsip dasar SVM adalah linier classifier, yaitu kasus klasifikasi yang secara linier dapat dipisahkan. Misalkan diberikan himpunan $X = \{x_1, x_2, x_3, \dots, x_n\}$, dinyatakan sebagai kelas positif jika $f(x) \geq 0$ dan yang lainnya termasuk ke dalam kelas negatif. Untuk penyelesaian masalah klasifikasi dalam SVM terbagi menjadi 2 bagian yaitu *separable data* dan *non separable*. Untuk *non separable* data proses klasifikasi ditambah dengan menggunakan fungsi kernel sebagai cara untuk menentukan hyperplane (fungsi pemisah) dalam mengklasifikasikan data [24].

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x_j) + b \quad (1)$$

dimana nilai $K(x_i, x_j)$ dapat dirubah sesuai dengan formasi kernelnya seperti pada kernel RBF dimana nilai $K(x_i, x_j) =$

$$\exp \left(-\frac{\|x - x'\|^2}{2\sigma} \right) \quad (1)$$

3. HASIL DAN PEMBAHASAN

Data yang diperlukan pada penelitian ini yaitu tweet dengan kata kunci #penyuka sesama jenis, #pergantian kelamin dan #penyimpangan seksual. Pengambilan kata kunci ini dirujuk berdasarkan trending topik yang sering dibahas sebanyak 10.000 tweet diambil dengan teknik crawling. Hasil data tersebut merupakan gabungan data dari masing-masing kata kunci yang telah di crawling sebelumnya dan belum memiliki kategori label yang jelas dari data tersebut. Data yang di *crawling* tersebut belum memiliki label (kelas) atau masih dalam jenis data *unsupervised*. Sehingga untuk diproses menggunakan *supervised data* diperlukan cara untuk melakukan pelabelan. Pada penelitian ini pelabelan data *tweet*

dilakukan secara manual dengan melibatkan pakar Bahasa Indonesia, sebagai sampel data teks yang telah diberikan label dapat dilihat pada tabel 1.

Tabel 1. Sampel Dataset yang telah memiliki label

No	Tweet	Sentimen (label)
1	dituding saammy simorangkir penyuka sesama jenis okan cornelius angkat bicara	negatif
2	tsania marwa tidak mau bicarakan soal kelainan seksual atalarik	netral
...
3826	zoopilia adalah penyimpangan seksual dimana pelaku senang bercinta dengan binatang	negatif

Masing-masing data dari kata kunci yang berbeda dicari pembobotan TF-IDF nya. TF-IDF dihitung dengan ketentuan mengeliminasi *term* dengan maksimal proporsi dokumen frekuensi melebihi 90% dan minimal dokumen frekuensi kurang dari 3%. Pada perhitungan TF-IDF menggunakan Bahasa pemrograman Python dengan menggunakan modul *scikit-learn* yaitu *TfidfVectorizer*. Hasil proses TF-IDF dapat dilihat pada tabel 2.

Tabel 2. TF-IDF dengan Data yang sudah dibersihkan

No	Term					
	anak	bilang	ganggu	gay	...	suka
1	0,00	0,00	0,00	0,00	...	0,00
2	0,00	0,00	0,00	0,35	...	0,00
...
3442	0,00	0,00	0,00	0,00	...	0,00

3.1 Algoritma Support Vector Machine (SVM)

Berdasarkan tahap *preprocessing*, seleksi fitur, dan pembobotan yang telah dilakukan sebelumnya, tahapan selanjutnya adalah proses klasifikasi algoritma SVM dengan menggunakan modul *scikit-learn* pada *python* yaitu SVC. Proses klasifikasi pada SVM terbagi menjadi 3 kali percobaan dengan menggunakan kernel *linier*, *polinomial* dan *RBF* pada data yang telah dibagi tersebut 90:10, 80:20, 70:30.

Hasil proses klasifikasi algoritma SVM dilakukan pada data yang diberikan label secara manual dengan menggunakan data uji sebanyak 10% yaitu 384 data, 20%=766 data dan 30%=1149 data pada masing kernel. Akurasi yang dihasilkan akan bergantung kepada sebuah kernel yang akan menghasilkan nilai yang berbeda, namun akurasi pada data penelitian terdapat beberapa yang memiliki persamaa. Hasil klasifikasi data pada kernel linear, polynomial dan RBF terbaik dapat dilihat pada tabel 3, 4 dan 5.

Tabel 3. Hasil klasifikasi SVM kernel linier

Percobaan		Linier						Akurasi
Training	Testing	Positif	True Negatif	Netral	Positif	False Negatif	Netral	
90%	10%	0	162	124	59	3	35	0.74
80%	20%	0	249	197	129	123	67	0.58
70%	30%	0	365	301	190	181	111	0.58

Hasil dari tabel diatas menunjukkan akurasi yang tertinggi di peroleh dari pengujian 10% data testing yaitu 74%. Selanjutnya kernel polynomial terbaik dapat dilihat pada tabel 4.

Tabel 4. Hasil klasifikasi SVM kernel polinomial data 90/10 (Terbaik).

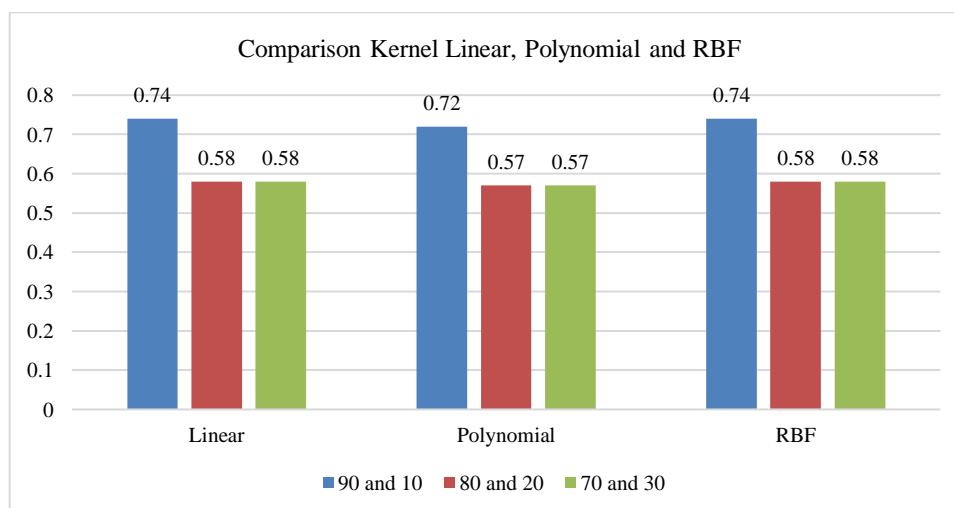
C	Degree	True			False			Akurasi
		Positif	Negatif	Netral	Positif	Negatif	Netral	
1	2	0	127	0	59	0	197	0.33
10	2	0	125	82	59	2	155	0.54
100	2	0	124	146	59	3	51	0.7
1000	2	1	121	153	58	5	43	0.72

Hasil percobaan di atas dengan data uji sebanyak 10% yaitu jumlah data sebanyak 384 tweet pada kernel polinomial dengan C=1000 dan degree=2, hasil akurasi yang didapat sebesar 72%. Sedangkan RBF, kernel terbaik dapat dilihat pada tabel 5.

Tabel 5. Hasil klasifikasi SVM kernel RBF dan 90/10

C	gamma	True			False			Akurasi
		Positif	Negatif	Netral	Positif	Negatif	Netral	
1	0.01	0	120	104	59	7	93	0.58
10	0.01	0	125	152	59	2	45	0.72
100	0.01	0	162	124	59	3	35	0.74
1000	0.01	0	123	154	59	4	43	0.72
1	0.1	0	125	148	59	2	49	0.71
10	0.1	0	123	154	59	4	43	0.72
100	0.1	0	122	153	59	5	44	0.71
1000	0.1	2	118	154	57	9	43	0.71
1	1	0	123	150	59	4	47	0.71
10	1	4	118	155	55	9	42	0.71
100	1	3	112	152	56	15	45	0.69
1000	1	4	109	152	55	18	45	0.69

Hasil akurasi tertinggi yang diperoleh dari ketiga kernel yaitu 74% pada pengujian dengan C=1000 gamma=0,01 pada kernel Linear. Dari beberapa percobaan tersebut terdapat penurunan dan kenaikan yang dihasilkan pada kernel dengan masing-masing paramater, sehingga untuk melihat konsistensya sebuah akurasi yang di hasilkan oleh algoritma SVM ini dapat di lihat pada pada Gambar 3.



Gambar 3. Perbandingan akurasi Kernel Linear, Polynomial and RBF

Hasil analisis pada penelitian ini ditentukan dengan melihat pandangan masyarakat indonesia terhadap LGBT ini berdasarkan akurasi klasifikasi yang telah dilakukan pengolahan data tweet pada masing-masing percobaan. Berdasarkan hal tersebut dapat dilihat bahwa pengakuan gender yang dilakukan oleh kelompok LGBT selama ini mendapatkan penolakan dari masyarakat indonesia. Hal ini dibuktikan dengan hasil akurasi pengolahan data klasifikasi yang menunjukkan 74% atau 0,74 sentimen yang mengarah kepada sentimen negatif. Namun, tidak hanya disitu saja penolakan dari masyarakat dalam sentimen negatif tersebut disampaikan berupa kata yang berdasarkan atas asas keagamaan dan asas biologis yang dapat dilihat dari kata perwakilan yang berfrekuensi besar yaitu simpang perilaku, simpang seksual, dosa dan juga didukung dengan sebagian kata-kata berfrekuensi rendah yang juga mengarahkan kepada penolakan. Disamping itu, juga terdapat banyak sentimen yang memberikan tanggapan atau respon yang tidak memihak kepada kelompok LGBT ini, yang dapat ditunjukan pada kata seperti biar, semoga sembuh, dan sembuh kelamin yang mewakili dari sentimen netral.tonton berdasarkan kategori sentimen. Penjelasan dan hasil dari wordcloud pada setiap siaran televisi dapat dilihat dibawah ini.

4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa kelompok LGBT harus mengakui bahwa kehadiran mereka tidak diakui oleh masyarakat di Indonesia sebagai seksual yang legal karena masyarakat berpandangan bahwa LGBT tersebut merupakan sesuatu perbuatan yang melanggar syari'at keagamaan dan aspek biologis khususnya agama islam. Hasil ini ditunjukan dengan akurasi klasifikasi sentimen tertinggi melalui perbandingan kernel linier, polinomial dan RBF yang diperoleh oleh sentimen negatif sebesar 74% dan 0,74. Serta kernel yang terbaik dalam mengklasifikasikan data adalah

kernel linier dan RBF yang menghasilkan akurasi yang sama. Dari proses pembagian data pada kasus ini, semakin kecil data latih yang digunakan, maka semakin besar pula akurasi yang dihasilkan.

REFERENCES

- [1] Dahar, R. K. A. H. (2016). Lgbt di indonesia : “Perspektif hukum islam, ham, psikologi dan pendekatan masalah”. Al-hakam.
- [2] Panji, B. (2019). “Wilayah yang dihuni kaum LGBT”. <https://wartakota.tribunnews.com/2019/02/24/ini-wilayah-yang-terbanyak-dihuni-kaum-lgbt-se-indonesia-ada-18.000-orang>.
- [3] Ernawati, T. (2017). “LGBT dalam perspektif islam”. Humanika, 6 (1).
- [4] Saputra, A. (2019). “Kuhp saat ini : Gelandangan di hukum 3 bulan penjara, lgbt tidak di pidana”. <https://news.detik.com/berita/d-47177154>.
- [5] Oktasari. (2015). “Text mining dalam analisis sentimen asuransi menggunakan metode naive bayes classifier”. Skripsi, Universitas Jenderal Ahmad Yani.
- [6] Yulian, E. (2018). “Text mining dengan k-means clustering pada tema lgbt dalam arsip tweet masyarakat kota bandung”. Matematika (Mantik), 04 (01).
- [7] Ahmed, K. , dan Jesmin, T. (2014). “Comparative analysis of data mining clasificatiion algoritms in 2 type diabetes prediction data uisng approach”. Internasional Journal of Sciene and Engineering.
- [8] Prasetyo, E. (2014). “Data Mining mengolah data menjadi informasi menggunakan matlab”. CV, Andi
- [9] Twiiter. (2016). “Twitter for bussines study”. <https://bussiness.twitter.com/id/basic>
- [10] Efendi, Z dan Mustakim. (2017). “Text mining clasification sebagai rekomendasi dosen pembimbing tugas akhir program studi sistem informasi”. Jurnal SNTIKI (hal 138-143).
- [11] Estu, dan Winda, N. (2017). “Analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial twitter menggunakan metode knn dan jumlah pembobotan retweet”. Jurna; Pengembangan Teknologi Informasi dan Komunikasi, 1(15).
- [12] Okfalisa, dan Harahap, A.H. (2016). “Implementasi metode *term frequency-invers document frequency*”. Sains dan Teknologi, 13(2), 152-159.
- [13] Wijaya, H. (2016). “Penerapan metode tf-idf dan cosine similarity dalam aplikasi kitab undang-undang hukum dagang. Skripsi UNES.
- [14] Rahamatil, R. F. (2017). “Penerapan teknik data mining dalam metode support vector machine untuk memprediksi siswa yang berpeluang drop out (studi kasus SMKN 1 sutera)”. Jaringan sistem informasi robotik, 1(01).
- [15] Nugroho, Witarto, dan Dwi. H. (2003). “Support Vector Machine-teori dan aplikasi dalam bio informatika”. Kuliah umum ilmu komputer.com
- [16] Hayadi, F. (2017). “Persepsi masyarakat tentang fenomena lgbt dikab. Bengkulu selatan tahun 2016”. Jurnal ilmiah ilmu sience kesehatan.
- [17] Hartanto. (2017). “ Text mining dan sentimen analisis twitter pada gerakan lgbt”. Jurnal Psikologi Ilmiah, 9 (1).
- [18] Okfalisa dan Harahap. (2016). “Implementasi metode terms frequency-inverse document frequency”. Sains, Teknologi dan Industri, 13(2), 151-159.
- [19] Tala. (2003). “Pembelajaran stemming pada bahasa indonesia”. *Institute for Logic, Language and Computation University van Amsterdam The Netherland*.
- [20] Raikwal dan Saxena. (2012). “Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data Set”. *International Journal of Computer Applications (0975 – 8887).Volume 50-No14*.
- [21] Ahmad, Munir., Aftab, Shahib., dan Ali, Iftikhar. (2017). . “Sentiment Analysis of Tweets using SVM”. *International Journal of Computer Applications (0975 – 8887). Volume 177 – No.5*
- [22] Demidova L, Nikulchev E, dan Sokolova Yu. (2016). “Big Data Classification Using The SVM Classifiers With The Modified Particle Swarm Optimization And The SVM Ensembles”. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7(5), p. 294-312.
- [23] Demidova, L., Nikulchev, E., Sokolova, Yu., N, Stepanov., dan N, Tyart. (2017). “Intellectual approaches to improvement of the classification decisions quality on the base of the SVM classifiers”. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7(5), p. 294-312.
- [24] Jadav, Bhumika M dan Vaghela, Vimalkumar B. (2016). “Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis”. *International Journal of Computer Applications (0975 – 8887). Volume 146 – No.13*.
- [25] Qaiser, Shahzad., dan Ali, Ramsa. (2018). “Text Mining: Use of TF-IDF to Examine the Relevance
- [26] of Words to Documents”. *International Journal of Computer Applications (0975 – 8887). Volume 181 – No.1*.
- [27] Asrifi, Ghulam Buntoro. (2018). “Sentiments analysis for prediction the governor of east java 2018 in twitter”. *International Journal Artificial Intelligent and Informatics*. Vol. 1, No. 2, 2018, pp. 82-90.
- [28] Budiharto, Widodo dan Meiliana. (2018). “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis”. *Journal of Big Data*.
- [29] Namugera, Frank., Wesonga, Ronald., dan Jehopio, Peter. (2019). “Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda”. *Journal Computational Sosial Network*.
- [30] Watequlis, Yan S., dan Puspitasari, Dwi. (2017). “Twitter Data Mining For Sentiment Analysis On Peoples Feedback Against Government Public Policy”. *International Journal of Science and Technology*, 3(1), 110 – 122
- [31] Effendy, Veronikha., Novantirani, Anita., dan Kania, Mira Sabariah. (2016). “Sentiment Analysis on Twitter about the Use of City Public Transportation Using Support Vector Machine Method”. *Internasional Journal on ICT*, Vol 2, June 2016, pp. 57 – 66.