



**ANALISIS SENTIMEN MENGENAI OPINI PUBLIK TERHADAP
KEBIJAKAN UU ITE PADA MEDIA SOSIAL TWITTER DENGAN
MELAKUKAN KOMPARASI ALGORITMA KLASIFIKASI NAÏVE
BAYES, K-NEAREST NEIGHBOR(K-NN), SUPPORT VECTOR
MACHINE(SVM)**

TUGAS AKHIR

**Abdurrahman Alfath
41518010145**

UNIVERSITAS
MERCU BUANA

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



**ANALISIS SENTIMEN MENGENAI OPINI PUBLIK TERHADAP
KEBIJAKAN UU ITE PADA MEDIA SOSIAL TWITTER DENGAN
MELAKUKAN KOMPARASI ALGORITMA KLASIFIKASI NAÏVE
BAYES, K-NEAREST NEIGHBOR(K-NN), SUPPORT VECTOR
MACHINE(SVM)**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

U N Abdurrahman Alfath

41518010145

MERCU BUANA

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM : 41518010145

Nama : Abdurrahman Alfath

Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 01 Juli 2022



Abdurrahman Alfath

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Abdurrahman Alfath
NIM : 41518010145
Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN) Support Vector Machine(SVM)

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

MERCU BUANA

Jakarta, 01 Juli 2022



Abdurrahman Alfath

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Abdurrahman Alfath
 NIM : 41518010145
 Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naive Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Menyatakan bahwa :

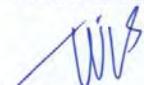
1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal : Jurnal Informatika : Jurnal Pengembangan IT		
	ISSN :		
	Link Jurnal : https://ejournal.poltektegal.ac.id/index.php/informatika		
	Link File Jurnal Jika Sudah di Publish :		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
 Dosen Pembimbing TA


 Dr. Harwikarya, MT

Jakarta, 01 Juli 2022


 Abdurrahman Alfath



LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010145
Nama : Abdurrahman Alfath
Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022



(Sabar Rudiarto, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010145
Nama : Abdurrahman Alfath
Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022



(Afiyati, S.Si, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM : Abdurrahman Alfath
Nama : 41518010145
Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022


(Wawan Gurawan, S.Kom, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

NIM : 41518010145
Nama : Abdurrahman Alfath
Judul Tugas Akhir : Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2022

Menyetujui,


(Dr. Harwikarya, MT)
Dosen Pembimbing

Mengetahui,


(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika


(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Alhamdulillah Puji syukur selalu kita panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia nya mengantarkan untuk menyelesaikan Tugas Akhir yang berjudul “Analisis Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter” sesuai dengan ketentuan yang sudah di tetapkan. Laporan Tugas Akhir ini di buat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa dalam pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kepada Orang tua. Ayah dan Ibu, yang selalu mendoakan dan mendukung bahwa saya bisa menyelesaikan kuliah dengan baik, serta dengan membiayai kuliah dari awal sampai akhir.
2. Ibu Saruni Dwiasnati, ST, MM, M.Kom selaku Dosen Pembimbing Akademik yang membantu proses perkuliahan dari awal kuliah sampai akhir.
3. Bapak Dr. Harwikarya, MT selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan yang baik dan meluangkan waktu untuk membantu saya menyelesaikan Tugas Akhir dari awal sampai akhir.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama perkuliahan berlangsung. Memberikan Kesempatan untuk selalu belajar, berkarya, dan berkembang.
5. Seluruh Staff Administrasi dan Tata Usaha yang telah memberikan bantuan pelayanan dan arahan sehingga dapat memberika kemudahan.
6. Teman Seangkatan yang selalu mengingatkan dan memberikan semangat dalam mengerjakan Tugas Akhir.
7. Senior yang selalu memberikan saran dan membantu dalam menyelesaikan Tugas Akhir.
8. Seluruh Pihak yang membantu dalam menyelesaikan proses Tugas Akhir

Akhir kata, penulis berharap hasil dari Tugas Akhir ini dapat bermanfaat, walaupun di dalam nya masih terdapat kekurangan dalam penulisan, eksperimen. Oleh karena itu, kritik dan saran pembaca sangat berharga bagi penulis untuk berkembangnya penulisan. Semoga Tugas Akhir ini dapat di mengerti sehingga dapat memberikan manfaat kepada pembaca.

Jakarta, 01 Juli 2022
Penulis



DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS.....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ...	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI.....	v
LEMBAR PENGESAHAN.....	viii
ABSTRAK.....	ix
ABSTRACT	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
NASKAH JURNAL	1
KERTAS KERJA	10
BAB 1. LITERATUR REVIEW.....	11
BAB 2. SOURCE CODE.....	26
BAB 3. DATASET	42
BAB 4. TAHAPAN EKSPERIMEN.....	45
BAB 5. HASIL SEMUA EKSPERIMEN.....	53
DAFTAR PUSTAKA.....	61
LAMPIRAN DOKUMEN HAKI	64
LAMPIRAN KORESPONDENSI.....	67

NASKAH JURNAL

Analisis Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(K-NN), Suport Vector Machine(SVM)

Abdurrahman Alfath^{1*)}, Harwikarya²,

^{1,2}Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta
^{1,2}Jln. Raya, Meruya Sel., Kec. Kembangan, Jakarta, Daerah Khusus Ibukota Jakarta, 11650, Indonesia
 email: ¹41518010145@student.mercubuana.ac.id, ²harwikarya@gmail.com

Abstract – Technology that is developing rapidly makes us inseparable from this, using technology can facilitate all daily activities. The ITE Law here plays an important role as a filter for information technology users so that they are not arbitrary in using and utilizing technology. This study aims to see public opinion on social media twitter on the policy of the ITE Law. Sentiment analysis here, we can see tweet data about the ITE Law whether the data is positive or negative by determining automatically and manually. This research can also later be used to assist the government in seeing to what extent the public knows the importance of the ITE Law in carrying out activities on social media. From the results of experimental testing using the Naïve Bayes classification algorithm, K-Nearest Neighbor (K-NN), and Support Vector Machine (svm) the best results were found, namely the SVM algorithm using automatic labeling data with an accuracy of 91%, then manual labeling data after conducting experiments, the best results were found, namely the SVM algorithm with an accuracy of 86%. Two data labeling automatic and manual with a percentage split of 80% : 20% the best. This is because the proportion of data 80% and 20% of the training and testing datasets have better results because they provide an evaluation value that is close to balance.

Abstrak – Teknologi yang sedang berkembang pesat membuat kita tidak bisa lepas dari hal tersebut, dengan menggunakan teknologi dapat memudahkan semua segala aktifitas sehari-hari. UU ITE disini sangat berperan penting sebagai filter pengguna teknologi informasi agar tidak semena-mena dalam menggunakan dan memanfaatkan teknologi. Penelitian ini bertujuan untuk ingin melihat opini masyarakat di sosial media twitter terhadap kebijakan UU ITE. Analisis Sentimen disini kita dapat melihat data tweet tentang UU ITE ini apakah data tersebut termasuk positif atau negatif dengan cara menentukan otomatis dan manual. Penelitian ini juga nantinya dapat digunakan membantu pemerintah dalam melihat sejauh mana masyarakat mengetahui pentingnya UU ITE dalam melakukan kegiatan di media sosial. Dari

hasil pengujian eksperimen dengan menggunakan algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor(K-NN), dan Support Vector Machine(svm) di dapati hasil terbaik yaitu algoritma SVM menggunakan data labeling Otomatis dengan akurasi 91%, kemudian di data labeling manual setelah melakukan eksperimen di dapati hasil terbaik yaitu algoritma SVM dengan akurasi 86%. Dua data labeling otomatis dan manual dengan percentage split 80% : 20% yang terbaik. Hal tersebut karena proporsi data 80% dan 20% dari dataset pelatihan dan pengujian memiliki hasil yang lebih baik karena memberikan nilai evaluasi yang mendekati keseimbangan.

Kata Kunci – Analisis Sentimen, UU ITE, Naïve Bayes, K-NN,SVM.

PENDAHULUAN

Teknologi yang sedang berkembang pesat membuat kita tidak bisa lepas dari hal tersebut, dengan menggunakan teknologi dapat memudahkan semua segala aktifitas sehari-hari. Oleh karena itu teknologi sangat membantu dan dibutuhkan oleh masyarakat dalam mencari segala informasi menggunakan sosial media. Sosial media Twitter contohnya yang sedang banyak digunakan dimasa endemic seperti ini, pengguna twitter dapat melihat atau melontarkan opini terkait apa yang sedang di alami atau sedang trending contohnya seperti KPK,UU ITE, dll. Akan tetapi dalam melakukan segala aktifitas menggunakan teknologi kita tidak bisa sembarang menggunakannya karena semua sudah ada aturan atau hukum di dalam UU ITE (Undang-Undang Informasi Transaksi Elektronik) yang sudah di tetapkan. Polemik yang terjadi di twitter terkait UU ITE sedang hangat diperbincangkan oleh masyarakat pengguna twitter, di dalamnya berisikan komentar pro dan kontra terhadap kebijakan UU ITE tersebut. Undang-Undang Nomor 11 Tahun 2008 tentang UU ITE adalah undang-undang pertama di bidang Teknologi Informasi dan Transaksi Elektronik sebagai produk legislasi yang sangat dibutuhkan dan telah menjadi

pionir yang meletakkan dasar pengaturan di bidang pemanfaatan Teknologi Informasi dan Transaksi Elektronik[1]. UU ITE dimana aturan pemerintah dengan pemanfaatan teknologi informasi menjadi lebih baik dalam perdagangan dan pertumbuhan ekonomi nasional untuk mencapai kesejahteraan masyarakat, dan pemerintah harus mendukung pengembangan teknologi informasi melalui infrastruktur peraturan perundang-undangan sehingga pemanfaatan teknologi informasi dilakukan secara aman untuk mencegah penyalahgunaannya dalam kehidupan bermasyarakat, berbangsa, bernegara, hak dan kebebasan.

Melalui penggunaan dan pemanfaatan teknologi informasi tersebut dengan mempertimbangkan pembatasan yang ditetapkan dengan undang-undang dengan maksud semata-mata untuk menjamin atas hak kebebasan orang lain dan untuk memenuhi tuntutan yang adil sesuai moral, nilai-nilai agama, keamanan, dan ketertiban umum dalam suatu bangsa yang demokratis. UU ITE disini sangat berperan penting sebagai filter pengguna teknologi informasi agar tidak semena-mena dalam menggunakan dan memanfaatkan teknologi, namun disini lain dengan adanya pasal-pasal UU ITE tersebut menyebabkan membuat merasa tidak adil. UU ITE kerap menjerat netizen di medsos. Southeast Asian Freedom of Expression Network (SAFE-net) mencatat puluhankasus sejak 2008 hingga 2020. Melalui beberapa penelitian, Cuitan masyarakat seputar UU ITE memiliki tatanan kalimat yang berbeda. Untuk itu diperlukan filterisasi keluhan dengan analisis sentimen. Analisis Sentimen merupakan ilmu untuk menganalisa data teks yang berupa opini, sikap, emosi, penilaian. ilmu ini termasuk bagian data mining. Data berbentuk teks tersebut akan di beri label positif dan negatif tergantung isi dari teks tersebut. Analisis Sentimen disini bertujuan ingin melihat opini masyarakat di sosial media twitter terhadap kebijakan UU ITE dan menerapkan algoritma klasifikasi untuk mendapatkan nilai akurasi terbaik. Manfaat analisis sentimen sangat penting untuk mengetahui sejauh mana data pengaduan masyarakat mengenai revisi undang-undang informasi dan transaksi elektronik digunakan sebagai alat untuk mengetahui pendapat dan reaksi masyarakat[2].

Masalah penelitian dari berbagai sumber dapat di jadikan acuan untuk penelitian ini, seperti “Analisis Sentimen pada pengguna twitter menggunakan metode K-Nearest Neighbor (K-NN)”, dari hasil pengujian diketahui bahwa nilai akurasi terbesar adalah 67,2% [3]. Masih rendahnya akurasi dalam penelitian tersebut, Sama hal nya pada penelitian yang lainnya “Analisis Sentimen Opini Publik Mengenai Covid-19 pada Twitter “Menggunakan Metode Naïve Bayes dan KKN, dari hasil pengujian diketahui metode Naïve Bayes memiliki nilai lebih tinggi, salah satunya adalah tingkat accuracy sebesar 63.21%, sedangkan metode KNN sebesar 58.10% [4]. Pada kedua sumber masalah penelitian diatas masih rendahnya akurasi sentimen, dan masih kurangnya pengujian model algoritma klasifikasi yang sesuai guna mendapatkan akurasi yang lebih baik. Pada penelitian ini guna mendapatkan akurasi yang lebih baik oleh karena itu pada penelitian ini menggunakan metode algoritma Naïve Bayes, KNN, dan SVM untuk menjadikan acuan dalam mencari nilai akurasi algoritma yang lebih tinggi.

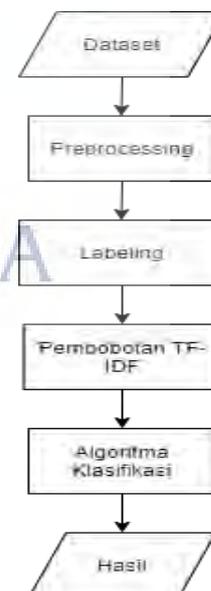
Berdasarkan latar belakang diatas, penelitian ini berfokus untuk mengklasifikasi terhadap opini masyarakat

melalui data tweet yang mengandung nilai positif, dan negatif mengenai UU ITE menggunakan algoritma K-Nearest Neighbor (K-NN), Support Vector Machine(SVM), dan Naïve Bayes, dimana ketiga algoritma tersebut memiliki keunggulan dalam melakukan klasifikasi text sangat cocok dilakukan eksperimen dari ketiga algoritma tersebut,

PENELITIAN YANG TERKAIT

Dalam penelitian sebelumnya mengenai Analisis Sentimen opini publik mengenai sarana dan transportasi mudik tahun 2019 pada twitter menggunakan algoritma naïve bayes, neural network, k-nn dan svm Algoritma K-NN memiliki akurasi yang lebih tinggi, akurasi= 90,76% dan AUC= 0,939; akurasi yang dihasilkan SVM= 89,03% dan AUC= 0,5; akurasi Naïve Bayes= 78,16% dan AUC= 0,567 kemudian yang dihasilkan Neural Network= 52,73% dan AUC=0,0[5]. Kemudian Analisis Sentimen UU ITE Menggunakan Algoritma Support Vector Machine(SVM) Hasil akurasi tertinggi didapatkan dari komposisi data latih 90% dan data uji 10% dengan nilai akurasi 84% dengan rata-rata precision 84%, recall 65%, f1-score 71% tiap kelas sentiment[6]. Berdasarkan pemaparan di atas algoritma klasifikasi yang digunakan mendapatkan akurasi tinggi sehingga sesuai dengan penelitian ini, untuk dapat dijadikan acuan bahwa algoritma tersebut sesuai dengan analisis sentiment dalam penelitian ini.

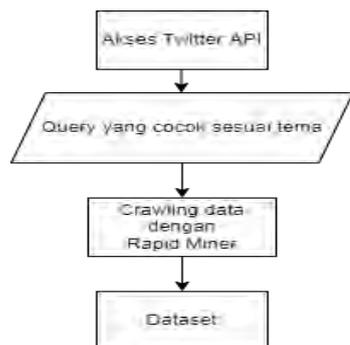
METODE PENELITIAN



Gambar 1. Alur metode penelitian

Pada Penelitian ini menggunakan metode kuantitatif dimana data yang terkumpul diproses melalui API Twitter dengan menggunakan tools RapidMiner. Penelitian ini di bantu dengan tools python dan r untuk melakukan preprocessing dan labeling pada Google Colab. Dalam penelitian ini juga dilakukan eksperimen dengan membandingkan hasil akurasi algoritma klasifikasi Naive Bayes, K-Nearest Neighbor (K-NN), dan Support Vector Machine(SVM) untuk mengetahui algoritma yang memiliki tingkat akurasi terbaik.

A. Dataset



Gambar 2. Perolehan dataset

Metode pengumpulan data ini dilakukan proses menggunakan tools Rapid Miner melalui API Twitter dengan memasukan *keyword* yang sesuai dengan topik. Tahap Selanjutnya crawling data sesuai apa yang sudah diinput melalui *Query* sehingga menjadi dataset. Pada pengumpulan data di peroleh dataset berupa text, yang di ambil dari parameter di atas yaitu UU ITE, pasal uu ite, pencemaran nama baik, berita hoax, judi online, penipuan online, dan ujaran kebencian. Dari parameter tersebut kita ketahui merupakan yang termasuk dalam UU ITE maka sangat cocok untuk dijadikan parameter. Setelah melakukan crawling data dari dari bulan November 20221 sampai bulan Juni 2022 dengan menggunakan parameter keyword diantara ke 7 parameter tersebut mendapatkan dataset berjumlah 2199 data.

Tanggal	Text	Id	Labeling Manual
2022-05-19 07:39:06	Mengancam pakai UU ITE Ancam gunakan kuasa politik (by name dropping jabatan seseorang) Bikin malu sebenarnya. Tapi ini bukan hal baru, meski sudah diprotes karena ngawur! tapi kenapa orang-orang tetap melakukan? Ya karena terus dibiarkan.	152708 641095 134003 2	Negatif
2022-05-21 16:37:09	RT @ddggmmbbkk: Mas tolong hati2 kalo ngomongin MU. Twitnya bisa saya laporkan ke pengacara saya, ini bisa kena pasal UU ITE. Ancaman Hukum...	152794 658919 691878 4	Positif
2022-05-21 16:19:10	RT @denkeropik: Artinya ini saya prediksi akan karet seperti UU ITE dan bergantung ke interpretasi Aparat Penegak Hukum. Kasus narkoba saj...	152794 206266 629734 4	Negatif
2022-05-21 15:33:32	@undvted @thesanolimait @meminumair @ARSIPAJA Bukan ngebela sih. Tapi biar kalian2 paham bahwa spill di medsos itu bukan hal yang baik. Mental2 seperti itu seharusnya dihilangkan. Soalnya sering loh, giliran di lapurin UU ITE eh netizen bergumam Hukum di Negara kita lucu. Padahal sudah sesuai hukum positif.	152793 057969 334272 3	Positif
2022-05-13 21:40:17	ini itu dilarang, ribet lu dah kek uu ite	1525123 7730573 14816	Negatif

Gambar 3. Contoh dataset

	Id	Text	Klasifikasi
0	1527086410951340032	Mengancam pakai UU ITE\nAncam gunakan kuasa po...	Negatif
1	1526101270310035456	Konten spt ini mmg satire, yg menyebarkan tdk ...	Negatif
2	1527952452389322752	@Dennysiregar7 Kena UU ITE gak nih???	Positif
3	1527951001805725698	RT @zucowi: @ciroooo17 Takut uu ite? Pakai uu ...	Negatif
4	1527949752129953792	@ciroooo17 Takut uu ite? Pakai uu tpks otomati...	Negatif
...
2194	1536302748643172352	@tatakujiyati Kalau berita itu Hoax kenapa kal...	Negatif
2195	1536297966302285824	@tatakujiyati Mmhm @detikcom disebut fitnah An...	Negatif
2196	1536297386490081282	@Prboemi_JKT Ga malu emang yaa bkin berita hoax	Negatif
2197	1536296431061184512	@yusuf_dumdum @tatakujiyati Sudah ketawan ngib...	Negatif
2198	1537585678057615361	RT @kurawa: Orang ini harus diproses pidana ka...	Negatif

2199 rows x 4 columns

Gambar 4. Jumlah dataset

B. Preprocessing

Tahapan dimana melakukan pengolahan data di awal sehingga tidak nanti pada saat diolah lebih lanjut tidak adanya data yang bermasalah. Sehingga pada saat di olah data tersebut relevan sesuai digunakan di penelitian ini, Dataset yang digunakan sebanyak 2.199 kemudian di olah, setelah di olah data manjadi 1.168 melalui pemrosesan antara lain:

3.1 Cleansing

Cleansing dalam penelitian ini bertujuan untuk menghapus RT, username, hashtag, angka dan URL pada tweets.

Contoh Tweet : @AndriYS27281093 @rrsofyanti @alisyarief @jokowi @elonmusk Hati" Bacotmu pak kalau gak bisa buktikan pemalsuan ijasah anda berarti hoax Fitnah pencemaran nama baik kena uu ite

Hasil Cleansing : Hati-hati Bacotmu pak kalau gak bisa buktikan pemalsuan ijasah anda berarti hoax Fitnah pencemaran nama baik kena uu ite

3.2 Case folding

Case folding merupakan tahapan awal pada Preprocessing yang bertujuan untuk mengubah setiap bentuk kata menjadi sama. Hal ini dilakukan dengan mengubah kata menjadi lower case atau huruf kecil [7].

Hasil Case Folding : hati-hati bacotmu pak kalau gak bisa buktikan pemalsuan ijasah anda berarti hoax fitnah pencemaran nama baik kena uu ite.

3.3 Tokenization

Tokenisasi adalah proses untuk memotong dokumen menjadi pecahan kecil yang dapat berupa bab, sub-bab, paragraf, kalimat, dan kata (token). Pada proses ini akan menghilangkan whitespace [7].

Hasil Tokenization : hatihati, bacotmu, pak, kalua, gak, bisa, buktikan, pemalsuan, ijazah, anda, berate, hoax, fitnah, pencemaran, nama, baik, kena, uu ite

3.4 Stopword removal

Tahap ini memroses agar kata-kata yang tidak penting atau tidak bermakna dihapus untuk analisis sentimen. Contoh kata-kata tersebut adalah atau, yang, dengan, di, ke, dan tetapi [8].

Hasil Stopword : hatihati, bacotmu, pak, kalua, gak, bisa, buktikan, pemalsuan, ijazah, anda, berate, hoax, fitnah, pencemaran, nama, baik, kena, uu ite,

3.5 Stemming

Tahap stemming adalah tahap mencari root (bentuk dasar) dari tiap kata. Pada tahap ini, dilakukan proses pengembalian berbagai bentukan kata ke dalam suatrepresentasi yang sama [9].

Hasil Stemming : hatihati, bacotmu, pak, kalua, gak, bisa, buktikan, pemalsuan, ijazah, anda, berate, hoax, fitnah, pencemaran, nama, baik, kena, uu ite,

3.6 Remove Duplicate

Tahapan ini digunakan untuk menghapus data yang sama, sehingga tidak adanya data yang sama pada saat di olah ke tahap selanjutnya.

	Id	Text	remove_user
1062	1525824910937190401	@kur0_kum4 @yourclavicula @Lambe48 @Y_ChikaJK.	Abis gw ngescroll tl barusan, kalo em...
1751	146440961643974660	Pengacara... Itu malah makin mojokin dia... Dr...	Pengacara... Itu malah makin mojokin dia... Dr...
81	1526012488524832768	@Prasety72595950 Mungkin itu yang adil bagi po...	Mungkin itu yang adil bagi polri, padahal ada...
1750	1464412864119115786	@schfless aduin aja neder itu gurumu gak sopan b...	aduin aja neder itu gurumu gak sopan bgt aduin...
1585	1464547937040551940	@Innayaputri72 ahli hukum pun tak bs membantu	ahli hukum pun tak bs membantu sebab UU ITE i...
1276	1526049298768883712	Yoi... pastinya dia akan ajukan gugatan Fitn...	Yoi... pastinya dia akan ajukan gugatan Fitn...
1507	1464433562426347520	RT @dediteaumbanu4 Yth pak Kapolri @ListyoSi...	RT : Yth pak Kapolri , Mohon diajak ngopi ak...
1612	1464500978120462336	@Markonah_003 Yuk hidupkan lagi kasus talangsa...	Yuk hidupkan lagi kasus talangsan dengan uu ite
662	1526229164612648960	Yuk sini yg mau gugat pencemaran nama baik Si...	Yuk sini yg mau gugat pencemaran nama baik. Si...
1301	1525843394710601728	@yedi_kumiahadri @Andrian51759679 @nberootuca...	1. Dizaman SBY mencetak rekor Premium naik

remove_http	tweet_cClean	Tweet
Abis gw ngescroll tl barusan, kalo em...	[abis_gw_ngescroll_tl_barusan_kalo_emang]	abis gw ngescroll tl barusan kalo emang foto d...
Pengacara... Itu malah makin mojokin dia... Dr...	[acara_mojokin_dr_no_rek_yg_libat_tn_tau_nama]	acara mojokin dr no rek yg libat ttn tau nama
Mungkin itu yang adil bagi polri, padahal ada...	[adil_polri_uu_ite]	adil polri uu ite
aduin aja neder itu gurumu gak sopan bgt aduin...	[aduin_aja_neder_guru_gak_sopan_bgt_aduin]	aduin aja neder guru gak sopan bgt aduin neder b...
ahli hukum pun tak bs membantu sebab UU ITE i...	[ahli_hukum_bs_bantu_uu_ite_biadab_karna_narasi]	ahli hukum bs bantu uu ite biadab karna narasi
Yoi... pastinya dia akan ajukan gugatan Fitn...	[yoi_pasti_aju_gugat_fitnah_cemar_nama]	yoi pasti aju gugat fitnah cemar nama ngacir...
Yth pak Kapolri , Mohon diajak ngopi ak...	[yth_kapolri_mohon_ajak_ngopi_akun_ya_mngkn]	yth kapolri mohon ajak ngopi akun ya mengkn
Yuk hidupkan lagi kasus talangsan dengan uu ite	[yuk_hidup_talangsan_uu_ite]	yuk hidup talangsan uu ite
Yuk sini yg mau gugat pencemaran nama baik. Si...	[yuk_yg_gugat_cemar_nama_silahkanmn]	yuk yg gugat cemar nama silahkanmn
1. Dizaman SBY mencetak rekor Premium naik	[zaman_sby_cetak_rekor_premium_atas_sorry_zaman]	zaman sby cetak rekor premium atas sorry zaman...

Gambar 5. Tahap preprocessing

	Tweet
120	abis gw ngescroll tl barusan kalo emang foto d...
2189	ada yg lawan berita hoax
81	adil polri uu ite
1864	admin yg salah konten bahasa suka sebar berita...
1778	ah ajg wea makan hoax judul berita
...	...
59	yg uu ite auto nyala gitu aja kaga ngarti
704	yoiiii pasti aju gugat fitnah cemar nama ngacir...
1519	yth kapolri mohon ajak ngopi akun ya mengkri
1234	yuk yg gugat cemar nama silahkanmn
729	zaman sby cetak rekor premium atas sorry zaman...

1168 rows × 1 columns

Gambar 6. Hasil Preprocessing tanpa label

	Klasifikasi	Tweet
120	Positif	abis gw ngescroll tl barusan kalo emang foto d...
2189	Negatif	ada yg lawan berita hoax
81	Positif	adil polri uu ite
1864	Negatif	admin yg salah konten bahasa suka sebar berita...
1778	Negatif	ah ajg wea makan hoax judul berita
...
59	Positif	yg uu ite auto nyala gitu aja kaga ngarti
704	Negatif	yoiiii pasti aju gugat fitnah cemar nama ngacir...
1519	Negatif	yth kapolri mohon ajak ngopi akun ya mengkri
1234	Negatif	yuk yg gugat cemar nama silahkanmn
729	Negatif	zaman sby cetak rekor premium atas sorry zaman...

1168 rows × 2 columns

Gambar 7. Hasil preprocessing dengan label

C. Labeling

Pada tahap ini setelah melakukan Preprocessing selanjutnya melakukan label terhadap data. Labeling sentimen bertujuan untuk mengetahui opini masyarakat terhadap suatu isu topik tersebut. Labeling dibagi menjadi 2 yaitu positif, dan negatif. Labeling otomatis data yang sudah bersih dan hanya memiliki kata yang bermakna diolah dengan bahasa R di google collaboratory. Kemudian data yang sudah bersih dan bermakna dicocokkan ke dalam kamus corpus bahasa Indonesia sentiment negatif dan positif yang disiapkan sebelumnya dalam bentuk csv. Kalimat yang terdeteksi akan diberi nilai, untuk kata positif diberi nilai 1, untuk kata negatif di beri nilai -1, dan jika tidak ditemukan kata tersebut di korpus negatif dan positif maka diberi nilai 0. Selanjutnya dilakukan penghitungan skor dengan menjumlahkan semua nilai dari kata. Jika skor bernilai <0 maka tweet dilabeli negatif dan jika jumlah skor >=0 maka tweet dilabeli positif. Ada berbagai Contoh dari kelas positif yang berisi pujian, saran, masukan, dan cerminan emosi positif seperti puas, senang, dan bahagia. Sedangkan kelas negatif berisi keluhan, kalimat sindiran, kritik, dan cerminan emosi negatif seperti amarah, kesal, dan kecewa. Kemudian Labeling manual dilakukan dengan memberikan label pada setiap tweet yang sesuai dengan kata tersebut apakah memiliki kata positif ataupun negatif pada setiap tweet. Terdapat Label dari otomatis terdapat sentimen positif 422 dan sentimen negatif 746 dan Label dari manual terdapat sentiment positif 483 dan negatif 685. Berikut tabel hasil dari labeling otomatis dan manual:

Positif	Negatif
422	746

Tabel 1. Labeling Otomatis

Positif	Negatif
483	685

Tabel 2. Labeling Manual

score		Tweet	klasifikasi
0	-2	abis gw ngescroll ti barusan kalo emang foto d...	Negatif
1	-1	acara mojokin dr no rek yg libat thn tau nama ...	Negatif
2	1	adil polri uu ite	Positif
3	2	aduin aja nder guru gak sopan bgt aduin nder b...	Positif
4	1	ahli hukum bs bantu uu ite biadab karna narasi...	Positif
5	-1	aja min cabut aja tuh uu ite pikir belah	Negatif
6	0	ajar jerat sakit uu ite pilih	Positif
7	-1	ajar tajwid ketimbang ajar mtk ngga spill oran...	Negatif
8	-2	ajgg cemar nama gue ga porno dude	Negatif
9	-1	ak lapirin km ujar benci	Negatif
10	0	akibat demo sayur takut uu ite hukum rasa pers...	Positif
11	0	aktivis hak asasi manusia ham haris azhar lawa...	Positif
12	0	akun lu anon lacak lu nyebarin aib orang enak ...	Positif
13	-2	akun ya bebas blok hayu aja delete deactivate ...	Negatif
14	2	alas anak ahok damai ayu thalia cemar nama	Positif
15	1	alhamdulillah amanah utk warga harga lembaga p...	Positif
16	2	alhamdulillah chat mesum hrs masuk sp tinggal ...	Positif
17	-2	amat hukum universitas taman siswa Palembang a...	Negatif
18	-6	amber ga saksi sih temennya ga lihat jadi cm t...	Negatif
19	-1	amriani bikin heboh media sosial hantam eks se...	Negatif
20	-1	analogi nya nonton bokep gak kena pasal hukum ...	Negatif
21	-5	ancam bunuh keras masuk ranah pidana lisan tul...	Negatif

Gambar 8. Labeling Otomatis

Klasifikasi		Tweet
0	Positif	abis gw ngescroll ti barusan kalo emang foto d...
1	Negatif	acara mojokin dr no rek yg libat thn tau nama ...
2	Positif	adil polri uu ite
3	Positif	aduin aja nder guru gak sopan bgt aduin nder b...
4	Negatif	ahli hukum bs bantu uu ite biadab karna narasi...
5	Negatif	aja min cabut aja tuh uu ite pikir belah
6	Negatif	ajar jerat sakit uu ite pilih
7	Negatif	ajar tajwid ketimbang ajar mtk ngga spill oran...
8	Negatif	ajgg cemar nama gue ga porno dude
9	Positif	ak lapirin km ujar benci
10	Positif	akibat demo sayur takut uu ite hukum rasa pers...
11	Positif	aktivis hak asasi manusia ham haris azhar lawa...
12	Positif	akun lu anon lacak lu nyebarin aib orang enak ...
13	Positif	akun ya bebas blok hayu aja delete deactivate ...
14	Negatif	alas anak ahok damai ayu thalia cemar nama
15	Positif	alhamdulillah amanah utk warga harga lembaga p...
16	Positif	alhamdulillah chat mesum hrs masuk sp tinggal ...
17	Positif	amat hukum universitas taman siswa Palembang a...
18	Negatif	amber ga saksi sih temennya ga lihat jadi cm t...
19	Negatif	amriani bikin heboh media sosial hantam eks se...
20	Negatif	analogi nya nonton bokep gak kena pasal hukum ...
21	Negatif	ancam bunuh keras masuk ranah pidana lisan tul...

Gambar 9. Labeling Manual

D. Pembobotan TF-IDF

Salah satu fungsi penting yang disediakan oleh tahapan proses ini adalah untuk dapat memilih term atau kata apa saja yang dapat dijadikan wakil penting untuk kumpulan dokumen yang akan dianalisis dengan kata lain kita melakukan pembobotan terhadap setiap term. Metode TF-IDF ini merupakan metode pembobotan dalam bentuk sebuah metode yang merupakan integrasi antar term frequency (TF), dan inverse document frequency (IDF)[10]. Pembobotan kata ini berpengaruh dalam mendapatkan akurasi terbaik.

E. Klasifikasi Algoritma

Klasifikasi Algoritma merupakan salah satu metode dalam penambangan data (data mining) yang digunakan dalam menganalisis sekumpulan data penting, Metode-metode dalam klasifikasi dapat secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan[11]. Sedangkan menurut [12] Klasifikasi adalah tipe analisis data yang dapat membantu orang menentukan kelas label dari sampel yang ingin di

klasifikasi, Berikut algoritma yang di gunakan untuk melakukan eksperimen:

1. Algoritma Naïve Bayes merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi, namun Naive Bayes memiliki kekurangan yaitu sangat sensitive dalam pemilihan fitur sehingga dapat mempengaruhi hasil akurasi [13].
2. Algoritma K-Nearest Neighbor (K-NN) adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari K tetangga terdekatnya (nearest neighbors) dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [14].
3. Support Vector Machine (SVM) adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi [15]. Dalam tujuan SVM adalah untuk menemukan fungsi klasifikasi terbaik untuk membedakan antara anggota dari dua kelas dalam data pelatihan[16].

HASIL DAN PEMBAHASAN

Pada hasil semua eksperimen akan menampilkan hasil yang telah dilakukan eksperimen. Pada penelitian ini melakukan eksperimen dari proses melakukan crawling data, preprocessing, labeling, TF-IDF, modeling sehingga di dapati hasil semua eksperimen dari beberapa algortima klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN), dan Support Vector Machine(SVM) dengan diperoleh hasil yang berbeda-beda. Penelitian ini juga nantinya dapat digunakan membantu pemerintah dalam melihat sejauh mana masyarakat mengetahui pentingnya UU ITE dalam melakukan kegiatan di media sosial, karena pada dasarnya dalam melakukan segala aktifitas di media sosial di pantau dengan UU ITE. Oleh karena ini sangat baik jika masyarakat tau apa saja yang termasuk UU ITE agar tidak adanya kekhawatiran dalam menggunakan media sosial.

Terdapat tiga scenario yang menjadi pembagian presentase data. Pada penelitian ini pembagian persentase dibagi menjadi tiga skenario eksperimen yaitu eksperimen pertama menggunakan pembagian data 70% : 30%, eksperimen kedua menggunakan pembagian data 80% : 20% dan eksperimen yang ketiga menggunakan pembagian data 90% : 10%. Pembagian tersebut merupakan pemisah data training dan testing berdasarkan persentasenya, contoh 90% : 10% yang merupakan data training 90% dan data training 10%. Penelitian ini juga melakukan perbandingan akurasi dengan membandingkan label otomastis dan label manual.

Pada labeling otomatis dengan skenario *percentage split* 70% : 30%, Hasil eksperimen kinerja model dari algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 79%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 78%, dan algortima SVM menghasilkan akurasi sebesar 88%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 70% : 30% dari setiap algoritma :

```
0.7995226730310262
[[174 62]
 [ 22 161]]
```

	precision	recall	f1-score	support
0	0.89	0.74	0.81	236
1	0.72	0.88	0.79	183
accuracy			0.80	419
macro avg	0.80	0.81	0.80	419
weighted avg	0.82	0.80	0.80	419

Gambar 10. Naïve Bayes 70% : 30% Labeling Otomatis

```
0.8285714285714286
[[62 16]
 [ 8 54]]
```

	precision	recall	f1-score	support
0	0.89	0.79	0.84	78
1	0.77	0.87	0.82	62
accuracy			0.83	140
macro avg	0.83	0.83	0.83	140
weighted avg	0.84	0.83	0.83	140

Gambar 14. K-NN 80% : 20% Labeling Otomatis

```
0.7857142857142857
[[61 17]
 [13 49]]
```

	precision	recall	f1-score	support
0	0.82	0.78	0.80	78
1	0.74	0.79	0.77	62
accuracy			0.79	140
macro avg	0.78	0.79	0.78	140
weighted avg	0.79	0.79	0.79	140

Gambar 11. K-NN 70% : 30% Labeling Otomatis

```
0.9139784946236559
[[151 7]
 [ 17 104]]
```

	precision	recall	f1-score	support
0	0.90	0.96	0.93	158
1	0.94	0.86	0.90	121
accuracy			0.91	279
macro avg	0.92	0.91	0.91	279
weighted avg	0.92	0.91	0.91	279

Gambar 15. SVM split 80% : 20% Labeling Otomatis

```
0.8806682577565632
[[212 25]
 [ 25 157]]
```

	precision	recall	f1-score	support
0	0.89	0.89	0.89	237
1	0.86	0.86	0.86	182
accuracy			0.88	419
macro avg	0.88	0.88	0.88	419
weighted avg	0.88	0.88	0.88	419

Gambar 12. SVM 70% : 30% Labeling Otomatis

Pada labeling otomatis dengan skenario *percentage split* 80% : 20%, Hasil eksperimen kinerja model dari algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 82%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 82%, dan algoritma SVM menghasilkan akurasi sebesar 91%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 80% : 20% dari setiap algoritma :

```
0.8243727598566308
[[122 26]
 [ 23 108]]
```

	precision	recall	f1-score	support
0	0.84	0.82	0.83	148
1	0.81	0.82	0.82	131
accuracy			0.82	279
macro avg	0.82	0.82	0.82	279
weighted avg	0.82	0.82	0.82	279

Gambar 13. Naïve Bayes 80% : 20% Labeling Otomatis

Pada labeling otomatis dengan skenario *percentage split* 90% : 10%, Hasil eksperimen kinerja model dari algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 83%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 81%, dan algoritma SVM menghasilkan akurasi sebesar 88%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 90% : 10% dari setiap algoritma :

```
0.8357142857142857
[[62 17]
 [ 6 55]]
```

	precision	recall	f1-score	support
0	0.91	0.78	0.84	79
1	0.76	0.90	0.83	61
accuracy			0.84	140
macro avg	0.84	0.84	0.84	140
weighted avg	0.85	0.84	0.84	140

Gambar 16. Naïve Bayes 90% : 10% Labeling Otomatis

```
0.8142857142857143
[[60 18]
 [ 8 54]]
```

	precision	recall	f1-score	support
0	0.88	0.77	0.82	78
1	0.75	0.87	0.81	62
accuracy			0.81	140
macro avg	0.82	0.82	0.81	140
weighted avg	0.82	0.81	0.81	140

Gambar 17. KNN 90% : 10% Labeling Otomatis

```
0.8857142857142857
[[75 7]
 [ 9 49]]
```

	precision	recall	f1-score	support
0	0.89	0.91	0.90	82
1	0.88	0.84	0.86	58
accuracy			0.89	140
macro avg	0.88	0.88	0.88	140
weighted avg	0.89	0.89	0.89	140

Gambar 18. SVM 90% : 10% Labeling Otomatis

Pada labeling manual dengan skenario *percentage split* 70% : 30%, Hasil eksperimen kinerja model dari algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 74%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 73%, dan algoritma SVM menghasilkan akurasi sebesar 80%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 70% : 30% dari setiap algoritma :

```
0.7463768115942029
[[119 82]
 [ 23 190]]
```

	precision	recall	f1-score	support
0	0.84	0.59	0.69	201
1	0.70	0.89	0.78	213
accuracy			0.75	414
macro avg	0.77	0.74	0.74	414
weighted avg	0.77	0.75	0.74	414

Gambar 19. Naïve Bayes 70% : 30% Labeling Manual

```
0.7391304347826086
[[34 28]
 [ 8 68]]
```

	precision	recall	f1-score	support
0	0.81	0.55	0.65	62
1	0.71	0.89	0.79	76
accuracy			0.74	138
macro avg	0.76	0.72	0.72	138
weighted avg	0.75	0.74	0.73	138

Gambar 20. KNN 70% : 30% Labeling Manual

```
0.8043478260869565
[[176 27]
 [ 54 157]]
```

	precision	recall	f1-score	support
0	0.77	0.87	0.81	203
1	0.85	0.74	0.79	211
accuracy			0.80	414
macro avg	0.81	0.81	0.80	414
weighted avg	0.81	0.80	0.80	414

Gambar 21. SVM 70% : 30% Labeling Manual

Pada labeling manual dengan skenario *percentage split* 80% : 20%, Hasil eksperimen kinerja model dari

algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 76%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 81%, dan algoritma SVM menghasilkan akurasi sebesar 86%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 80% : 20% dari setiap algoritma :

```
0.7608695652173914
[[ 79 54]
 [ 12 131]]
```

	precision	recall	f1-score	support
0	0.87	0.59	0.71	133
1	0.71	0.92	0.80	143
accuracy			0.76	276
macro avg	0.79	0.76	0.75	276
weighted avg	0.79	0.76	0.75	276

Gambar 22. Naïve Bayes 80% : 20% Labeling Manual

```
0.8188405797101449
[[43 19]
 [ 6 70]]
```

	precision	recall	f1-score	support
0	0.88	0.69	0.77	62
1	0.79	0.92	0.85	76
accuracy			0.82	138
macro avg	0.83	0.81	0.81	138
weighted avg	0.83	0.82	0.82	138

Gambar 23. K-NN 80% : 20% Labeling Manual

```
0.8659420289855072
[[113 14]
 [ 23 126]]
```

	precision	recall	f1-score	support
0	0.83	0.89	0.86	127
1	0.90	0.85	0.87	149
accuracy			0.87	276
macro avg	0.87	0.87	0.87	276
weighted avg	0.87	0.87	0.87	276

Gambar 24. SVM 80% : 20% Labeling Manual

Pada labeling manual dengan skenario *percentage split* 90% : 10%, Hasil eksperimen kinerja model dari algoritma klasifikasi Naïve Bayes, K-Nearest Neighbor (K-NN) dengan Best K=3, dan Support Vector Machine(SVM). Dalam melakukan eksperimen di dapati hasil dari algoritma Naïve Bayes menghasilkan akurasi sebesar 77%, kemudian algoritma KNN dengan K=3 menghasilkan akurasi 79%, dan algoritma SVM menghasilkan akurasi sebesar 83%. Dari hasil kinerja yang mendapati hasil akurasi terbaik adalah algoritma SVM. Berikut merupakan hasil *percentage split* 70% : 30% dari setiap algoritma :

- vol. 15, no. 1, pp. 23–28, 2020.
- [5] M. W. Pertiwi, “Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, K-NN dan SVM,” *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 27–32, 2019.
- [6] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021.
- [7] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, “Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 11, pp. 4704–4713, 2018.
- [8] V. K. S. Que, A. Iriani, and H. D. Purnomo, “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
- [9] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [10] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor,” *Kilat*, vol. 8, no. 1, pp. 81–92, 2019, doi: 10.33322/kilat.v8i1.421.
- [11] H. Sulistiani, “Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa,” pp. 300–305, 2018, doi: 10.31227/osf.io/yaavj.
- [12] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [13] R. Wati, “Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 159–164, 2020, doi: 10.33480/jitk.v5i2.1034.
- [14] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [15] M. Kusmira, “Analisis Sentimen Registrasi Ulang Kartu SIM pada Twitter Menggunakan Algoritma SVM dan K-NN,” *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 105–110, 2019.
- [16] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Analisa Sentimen Mengenai Opini Publik Terhadap Kebijakan UU ITE Pada Media Sosial Twitter Dengan Melakukan Komparasi Algoritma Klasifikasi Naïve Bayes, K-Nearest Neighbor(KNN), Support Vector Machine(SVM)”. Kertas kerja ini berisi semua material hasil penelitian Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan mengenai source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.