

**ANALISIS SENTIMEN UU CIPTA KERJA
MENGUNAKAN METODE *SUPPORT VECTOR
MACHINE* (SVM)**

(Studi Kasus : UU Cipta Kerja)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Ahmad Rohiqim Makhtum

18611118

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2022**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Analisis Sentimen UU Cipta Kerja Menggunakan
Metode *Support Vector Machine* (SVM) dengan
Studi Kasus : UU Cipta Kerja

Nama Mahasiswa : Ahmad Rohiqim Makhtum

NIM : 18611118

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 1 Juli 2022

Ketua Prodi Statistika

Pembimbing


(Dr. Edy Widodo, S.Si., M.Si.)


(Muhammad Muhajir, S.Si., M.Sc.)

**HALAMAN PENGESAHAN
TUGAS AKHIR**

**ANALISIS SENTIMEN UU CIPTA KERJA MENGGUNAKAN METODE
SUPPORT VECTOR MACHINE (SVM)**

(Studi Kasus : UU Cipta Kerja)

Nama Mahasiswa : Ahmad Rohiqim Makhtum

NIM : 18611118

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 12 Juli 2022**

Nama Penguji

Tanda Tangan

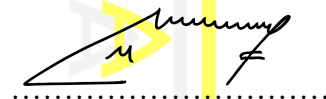
1. Ayundyah Kesumawati, S.Si., M.Si



2. Dina Tri Utari, S.Si., M.Sc.



3. Muhammad Muhajir, S.Si., M.Sc.



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)



<https://science.uii.ac.id/surat-digital/validasi/REG008208>

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Dengan memanjatkan puji syukur kehadiran Allah SWT atas segala rahmat dan inayahnya, sholawat serta salam juga tercurahkan kepada baginda Nabi agung Muhammad SAW yang telah membawa zaman kegelapan menuju zaman yang terang benderang *addinul islam*.

Akhirnya dengan kerendahan hati, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah membantu, membimbing, dan memberikan dorongan sehingga tersusunnya Penelitian Tugas Akhir ini. Terkhusus penulis ucapkan terima kasih kepada :

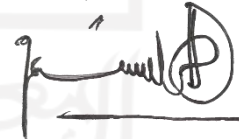
1. Prof. Riyanto, S.Pd., M.Si., Ph.D. yang menjabat Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
2. Dr. Edy Widodo, S.Si., M.Si. yang menjabat sebagai Ketua Jurusan Statistika dan jajaran Program Studi Statistika.
3. Bapak Muhammad Muhajir, S.Si., M.Sc. selaku dosen Pembimbing yang sudah sangat sabar dan pengertian dalam membimbing, mengajari, menjelaskan, dan menyemangati penulis sampai selesainya penelitian ini.
4. Dosen dan staf Program Studi Statistika Universitas Islam Indonesia yang sudah memberikan begitu banyak ilmu pengetahuan kepada penulis.
5. Kedua orang tua yang setiap hari dan setiap malam memberikan doa, dukungan dan motivasi kepada penulis. Kepada kedua adik yang telah memberikan senyuman-senyuman ringan sebagai penyemangat penulis selama ini.
6. Sahabat kontrakan tercinta (Riski Sarwo, Alvin M, Juliana Saputra, Akmal Fauzan, Zahrul Wafi, Baginda Attariq (Rembo), Lutfi Anshori, Wahyu Aji, Takayuki Pratama, Alan, Mikung dan teman-teman lainnya yang tidak mungkin disebutkan semua yang selalu setia menemani, memberikan dukungan dan semangat motivasi dari awal penulis kuliah tahun 2018 sampai di penghujung tahun 2022.

7. Serta terimakasih untuk semua pihak yang telah membantu penulis dalam menyelesaikan penelitian ini yang tidak mungkin penulis sebut satu per satu.

Demikian yang bisa penulis sampaikan. Sekali lagi penulis menyampaikan bahwa penelitian ini masih sangat jauh dari kata sempurna dan masih banyak kekurangan karena keterbatasan pengetahuan, wawasan, serta pengalaman penulis. Akan tetapi penulis berusaha menyajikan penelitian ini dengan sebaik-baiknya. Terakhir, penulis berharap dengan adanya Penelitian Tugas Akhir ini menjadi *washilah* penulis dalam kemanfaatan untuk masyarakat yang lebih luas.

*Wallahulmuwaffiq Ilaa Aqwamitthoriq
Wassalamu 'alaikum Wr.Wb.*

Yogyakarta, 24 Juni 2022



Ahmad Rohiqim Makhtum

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL | i |
| HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR..... | ii |
| HALAMAN PENGESAHAN TUGAS AKHIR | iii |
| KATA PENGANTAR..... | iv |
| DAFTAR ISI | vi |
| DAFTAR TABEL | viii |
| DAFTAR GAMBAR..... | ix |
| DAFTAR LAMPIRAN | x |
| PERNYATAAN | xi |
| INTISARI..... | xii |
| ABSTRACT | xiii |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang Masalah..... | 1 |
| 1.2. Rumusan Masalah | 6 |
| 1.3. Batasan Masalah..... | 6 |
| 1.4. Tujuan Penelitian | 7 |
| 1.5. Manfaat Penelitian | 7 |
| BAB II TINJAUAN PUSTAKA..... | 8 |
| 2.1. Kajian Induktif | 8 |
| BAB III LANDASAN TEORI..... | 13 |
| 3.1. <i>Twitter</i> | 13 |
| 3.2. UU Cipta Kerja (<i>Omnibus law</i>)..... | 14 |
| 3.2.1 Definisi <i>Omnibus law</i> | 14 |
| 3.2.2 Tujuan <i>Omnibus law</i> | 15 |
| 3.2.3 <i>Omnibus law</i> di Indonesia | 16 |
| 3.3. <i>Text Mining</i> | 18 |
| 3.3.1 <i>Text Preprocessing</i> | 18 |
| 3.4. Pembobotan TF-IDF | 19 |
| 3.5. Analisis Sentimen..... | 21 |
| 3.6. Klasifikasi | 22 |
| 3.7. <i>Confusion Matrix</i> | 24 |
| 3.8. <i>Support Vector Machine (SVM)</i> | 25 |
| 3.9. <i>Support Vector Machine Nonlinear</i> dan <i>Kernel Trick</i> | 33 |
| 3.10. <i>Wordcloud</i> | 34 |
| BAB IV METODOLOGI PENELITIAN | 36 |
| 4.1. Populasi Penelitian | 36 |
| 4.2. Tempat dan Waktu Penelitian | 36 |
| 4.3. Variabel penelitian | 36 |
| 4.4. Teknik Sampling | 36 |
| 4.5. Tahapan Penelitian | 37 |
| BAB V HASIL DAN PEMBAHASAN..... | 38 |
| 5.1. <i>Scraping Data</i> | 38 |
| 5.2. <i>Preprocessing Data</i> | 39 |
| 5.2.1 <i>Case Folding</i> | 39 |
| 5.2.2 <i>Cleaning Data</i> | 40 |

| | | |
|-------|--|----|
| 5.2.3 | <i>Tokenizing</i> | 41 |
| 5.2.4 | <i>Filtering</i> | 42 |
| 5.3. | Pelabelan Kelas Sentimen..... | 43 |
| 5.4. | <i>Statistical Term Frequency-Invers Document Frequency (TF-IDF)</i> | 46 |
| 5.5. | Klasifikasi SVM..... | 48 |
| 5.5.1 | Kernel <i>Linear</i> | 50 |
| 5.5.2 | Kernel <i>Polynomial</i> | 50 |
| 5.5.3 | Kernel <i>RBF</i> | 52 |
| 5.5.4 | Kernel <i>Sigmoid</i> | 53 |
| 5.6. | Asosiasi Kata..... | 55 |
| 5.7. | <i>Wordcloud</i> | 56 |
| | BAB VI PENUTUP..... | 57 |
| 6.1. | Kesimpulan..... | 57 |
| 6.2. | Saran..... | 58 |
| | DAFTAR PUSTAKA..... | 59 |
| | LAMPIRAN..... | 65 |



DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Tabel Penelitian Sebelumnya | 10 |
| Tabel 3.1 Tabel <i>Confusion matrix</i> | 24 |
| Tabel 3.2 Tabel Nilai <i>AUC</i> | 25 |
| Tabel 3.3 Contoh Data <i>Linear SVM</i> | 29 |
| Tabel 4.1 Variabel Penelitian | 36 |
| Tabel 5.1 Sampel Data <i>Tweet</i> | 39 |
| Tabel 5.2 Proses <i>Case Folding</i> | 40 |
| Tabel 5.3 Proses <i>Cleaning data</i> | 41 |
| Tabel 5.4 Proses <i>Tokenizing</i> | 42 |
| Tabel 5.5 Proses <i>Filtering</i> | 43 |
| Tabel 5.6 Pelabelan <i>Tweet</i> | 44 |
| Tabel 5.7 Contoh Perhitungan Skor | 45 |
| Tabel 5.8 Klasifikasi Pelabelan <i>Tweet</i> | 45 |
| Tabel 5.9 Perhitungan TF | 46 |
| Tabel 5.10 Perhitungan IDF | 47 |
| Tabel 5.11 TF-IDF pada <i>Omnibus law</i> | 47 |
| Tabel 5.12 Frekuensi Kata | 48 |
| Tabel 5.13 Nilai Akurasi Kernel SVM | 55 |
| Tabel 5.14 Asosiasi Kata | 55 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 3.1 Proses SVM untuk Menemukan <i>Hyperline</i> | 27 |
| Gambar 3.2 <i>Hyperline</i> dan Jarak <i>Margin</i> | 27 |
| Gambar 3.3 Grafik Simulasi Data | 29 |
| Gambar 3.4 SVM <i>Non Linear</i> | 33 |
| Gambar 3.5 Tampilan <i>Wordcloud</i> | 35 |
| Gambar 4.1 Tahapan Penelitian | 37 |
| Gambar 5.1 <i>Install Tweepy</i> | 38 |
| Gambar 5.2 <i>Import Tweepy, Configparser, Pandas</i> | 38 |
| Gambar 5.5 Plot Data Sentimen | 46 |
| Gambar 5.6 <i>Confusion Matrix</i> | 49 |
| Gambar 5.7 Kernel <i>Linear</i> | 50 |
| Gambar 5.8 Kernel <i>Polynomial</i> | 51 |
| Gambar 5.9 <i>Confusion Matrix Kernel Polynomial</i> | 51 |
| Gambar 5.10 Kernel RBF | 52 |
| Gambar 5.11 <i>Confusion Matrix Kernel RBF</i> | 52 |
| Gambar 5.12 Kernel <i>Sigmoid</i> | 53 |
| Gambar 5.13 <i>Confusion Matrix Kernel Sigmoid</i> | 54 |
| Gambar 5.14 <i>Wordcloud</i> Klasifikasi Positif | 56 |
| Gambar 5.15 <i>Wordcloud</i> Klasifikasi Negatif | 56 |

DAFTAR LAMPIRAN

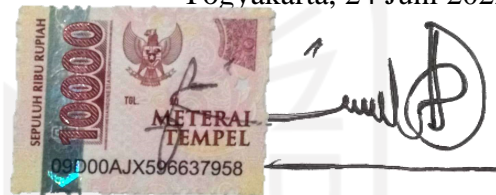
| | |
|------------------|----|
| Lampiran 1 | 65 |
| Lampiran 2 | 66 |
| Lampiran 3 | 69 |
| Lampiran 4 | 72 |
| Lampiran 5 | 75 |



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 24 Juni 2022



Ahmad Rohiqim Makhtum

INTISARI

Analisis Sentimen UU Cipta Kerja Menggunakan Metode *Support Vector Machine* (SVM)

(Studi Kasus : UU Cipta Kerja)

Ahmad Rohiqim Makhtum

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Undang-Undang *Omnibus law* merupakan Undang-Undang baru yang digunakan dalam sistem peraturan di Indonesia, sistem ini biasa disebut dengan Undang-Undang sapu jagat karena menggantikan beberapa norma-norma hukum menjadi satu peraturan. *Omnibus law* memiliki 11 klaster, salah satunya adalah mengenai regulasi ketenagakerjaan. Namun banyak tanggapan yang berbeda-beda dari masyarakat Indonesia mengenai *Omnibus law* khususnya di media sosial *twitter*. Peneliti menggunakan teknik *scraping* dalam pengambilan data cuitan warganet *twitter*. Didapatkan data sebanyak 3067 data yang diambil dari 20 Maret 2022 sampai 20 Mei 2022. Data tersebut selanjutnya diberikan pelabelan (positif, negatif dan netral), pembobotan dan pengklasifikasian menggunakan metode *Support Vector Machine* (SVM) dengan tujuan mengetahui sentimen dari masyarakat terhadap *Omnibus law* apakah cenderung positif atau negatif dan mengetahui akurasi dari metode SVM. Klasifikasi algoritma SVM dengan menggunakan kernel *linear* mendapatkan nilai akurasi tertinggi yakni sebesar 97.05% dan AUC sebesar 0,97. Berdasarkan hasil analisis diketahui bahwa masyarakat lebih banyak menilai dengan tanggapan yang positif terkait *Omnibus law* dibandingkan dengan tanggapan negatif. Tanggapan positif diantaranya mendukung pengusaha muda dengan legislasi yang mudah, sedangkan tanggapan negatif salah satunya tentang gaji buruh yang masih murah walaupun sudah disahkan UU *Omnibus law*.

Kata Kunci : *Omnibus law, Twitter, Analisis Sentimen, Support Vector Machine*

ABSTRACT

Sentiment Analysis of Omnibus Law Using *Support Vector Machine* (SVM) Method

(Case Study : Omnibus Law)

Ahmad Rohiqim Makhtum

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

The Omnibus Law is a new law that is used in the regulatory system in Indonesia, this system is commonly referred to as the Universal Sweeping Law because it replaces several legal norms into one regulation. The Omnibus law has 11 clusters, one of which is regarding labor regulations. However, there are many different responses from the Indonesian people regarding the Omnibus law, especially on Twitter social media. Researchers used scraping techniques in retrieving tweets from Twitter citizens. There were 3067 data collected from March 20, 2022 to May 20, 2022. The data were then labeled (positive, negative and neutral), weighted and classified using the Support Vector Machine (SVM) method with the aim of knowing the sentiments of the public towards the Omnibus law and determine the accuracy of the Support Vector Machine (SVM) method. The classification of the SVM algorithm using a linear kernel gets the highest accuracy value of 97.05%. The public is more concerned with positive responses regarding the Omnibus law compared to negative responses. The positive responses included supporting young entrepreneurs with easy legislation, while the negative responses were one of them regarding the wages of workers who were still cheap despite the existence of the Omnibus Law..

Keywords: *Omnibus law, Twitter, Sentiment Analysis, Support Vector Machine*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pada awal tahun 2020, pemerintah sedang menyiapkan rancangan undang-undang (selanjutnya disingkat RUU) tentang Cipta Kerja dengan menggunakan konsep *Omnibus law*. Pada prosesnya RUU ini disiapkan oleh pemerintah untuk digunakan sebagai suatu strategi dalam usaha membangun perekonomian Indonesia sehingga para investor dapat tertarik untuk menanamkan modalnya di Indonesia. Pemerintah menganggap perlu adanya RUU Cipta Kerja ini dikarenakan tingginya angka pengangguran di Indonesia yang mencapai 7 juta orang, sehingga diharapkan RUU ini mampu menciptakan lapangan kerja. Kini *omnibus law* Cipta Kerja resmi disahkan oleh pemerintah dalam rapat paripurna yang diselenggarakan di Gedung DPR yang berada di Senayan Jakarta pada 5 Oktober 2020 dengan jumlah halaman total 905 (Kemenkumham, 2020). Konsep *Omnibus law* ini merupakan konsep baru yang digunakan dalam sistem peraturan di Indonesia. Sistem ini biasa disebut dengan Undang-Undang sapu jagat karena menggantikan beberapa norma-norma hukum menjadi satu peraturan. Selain itu konsep seperti ini dijadikan misi untuk memotong beberapa norma hukum yang dianggap tidak relevan dengan perkembangan zaman dan dirasa merugikan negara. (Prabowo, 2020)

UU Cipta Kerja memiliki 11 klaster, salah satunya adalah mengenai regulasi ketenagakerjaan. Klaster ini mencakup 3 Undang-Undang yang digabungkan menjadi satu, yaitu Undang-Undang Nomor 13 Tahun 2003 mengenai ketenagakerjaan, Undang-Undang Nomor 40 Tahun 2004 mengenai sistem jaminan sosial, dan Undang-Undang Nomor 24 Tahun 2011 mengenai badan penyelenggara jaminan sosial. Pada klaster terkait ketenagakerjaan, Pemerintah berupaya menyelaraskan 3 undang-undang tersebut dengan harapan mampu memberikan dimensi dan ruang untuk investor melihat peraturan yang sudah disempurnakan tanpa harus khawatir dengan peraturan yang saling tumpang tindih dan mengakibatkan kerugian bagi investor itu sendiri (UU Cipta Kerja).

Pemerintah juga menganggap bahwa UU Cipta Kerja bermanfaat untuk memperbaiki iklim investasi dan mewujudkan suatu kepastian hukum. Adapun tujuan dari pendorongan investasi diantaranya adalah : sebagai upaya untuk mempercepat transformasi ekonomi, sebagai langkah menyelaraskan kebijakan pusat dan daerah, upaya untuk memberikan kemudahan dalam membuka usaha, dan sebagai produk hukum untuk mengatasi problematika tentang regulasi yang masih tumpang tindih. (kemendikbud, 2020) Sesuai yang disampaikan Presiden Indonesia Joko Widodo dalam acara pertemuan tahunan Bank Indonesia di Jakarta pada Kamis 28 November 2019 bahwa “50 tahun waktu yang dibutuhkan jika merevisi UU satu per satu. Melalui *Omnibus law*, penyederhanaan regulasi bisa dipercepat” (kemendikbud, 2019). Namun dalam proses perencanaan RUU ini, bahkan sampai disahkan menjadi UU Cipta Kerja pada tahun 2020 banyak sekali opini-opini yang dikeluarkan oleh masyarakat yang tetap menolak adanya UU Cipta Kerja ini. Banyaknya opini publik beberapa disebabkan karena *deadline* pengerjaannya hanya selama 100 hari oleh Presiden Jokowi dan juga tidak melibatkan pihak-pihak lain dalam pembuatannya. Namun, ada satu hal yang sangat penting dan menjadi masalah utama dalam UU ini yakni persoalan adanya pemotongan pesangon bagi pekerja yang diputus hubungan kerja oleh perusahaan, hilangnya hari cuti melahirkan dan lain sebagainya.

Berdasarkan berita-berita yang beredar selama tahun 2021 sampai 2022 masih begitu banyak aksi demo yang dilakukan oleh masyarakat, buruh maupun mahasiswa yang menuntut adanya pencabutan UU Cipta Kerja. Seperti contoh pada hari Jumat 7 Desember 2021 puluhan ribu buruh yang melibatkan para petani, buruh honorer, buruh migran, pekerja rumah tangga, masyarakat bahkan mahasiswa untuk mencabut UU Cipta Kerja sesuai keputusan Mahkamah Konstitusi yang sudah terlebih dahulu menyatakan inkonstitusional bersyarat dan cacat formil (Amani, 2021). Begitu juga pada awal tahun 2022 tepatnya 7 februari 2022 massa buruh bergerak ke Gedung DPR RI untuk menyelenggarakan unjuk rasa demi menyuarkan kembali tuntutan terhadap penolakan *Omnibus law* UU Cipta Kerja. Bukan hanya di Jakarta, aksi massa buruh juga dilakukan di Semarang, Jepara, Makassar, Aceh, Surabaya, Medan hingga Banjarmasin (CNN, 2022).



Gambar 1.1 Massa Buruh Gabungan Menggelar Aksi Tolak *Omnibus law*

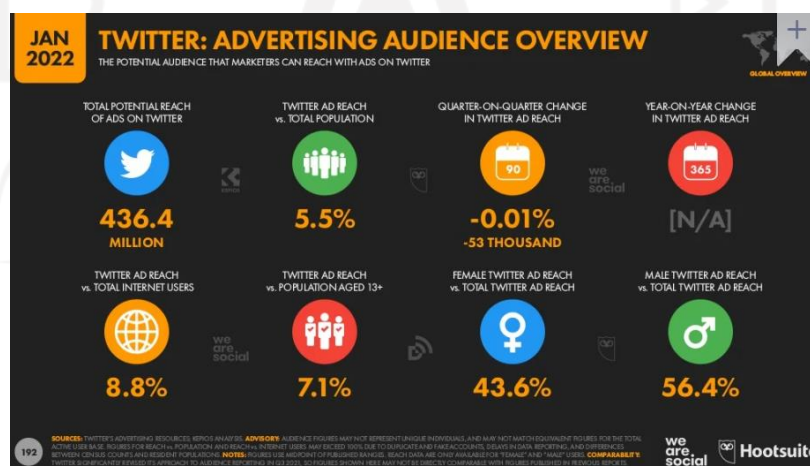
Source : <https://www.liputan6.com/>

Berdasarkan kejadian ini, mengapa begitu banyak pekerja dan masyarakat yang menolak adanya RUU Cipta Kerja ini. Inti masalahnya terletak pada Pasal 89 ayat 45 RUU Cipta Kerja yang merevisi ketentuan pada pasal 156 Undang-Undang Nomor 13 tahun 2003 mengenai Ketenagakerjaan yang mengalami perubahan regulasi perhitungan uang pada ayat 1 pasal 156 Undang-Undang Nomor 13 tahun 2003 yang dirubah dengan redaksi “*Dalam hal terjadi pemutusan hubungan kerja, pengusaha wajib membayar uang pesangon dan/atau uang penghargaan masa kerja*”, selanjutnya pada ayat 3 pasal 156 dilakukan perubahan pada huruf g dan h diganti dengan ketentuan regulasi pada huruf g dengan bunyi “*masa kerja 21 (dua puluh satu) tahun atau lebih, 8 (delapan) bulan upah*”. Perubahan ini merupakan upaya untuk penghapusan perhitungan uang penghargaan jasa karena telah diatur bahwa masa kerja 24 tahun atau lebih adalah 10 bulan upah yang diterima (Kurniawan, 2020).

Sebenarnya bukan hanya aksi buruh, kasus terbaru yang terdapat di desa Wadas kabupaten Purworejo Jawa Tengah tentang penolakan masyarakat terhadap proyek strategis nasional berupa penambangan juga imbas dari adanya UU Cipta Kerja. Menurut pakar hukum lingkungan dari Pusat Studi Lingkungan Hidup UGM Wahyu Yun Santoso peristiwa di Desa Wadas, Purworejo, Jawa Tengah pada Selasa 8 Februari 2022 merupakan dampak dari proyek strategis nasional sebagai implementasi *Omnibus law* UU Cipta Kerja. Pemerintah seolah-olah mengesampingkan mengenai dampak lingkungan, konsolidasi dengan masyarakat, dan lebih mementingkan dan memfokuskan Program Strategis

Nasional (PSN). Begitu juga dengan keran investasi penambangan secara besar-besaran yang telah ditopang dengan adanya UU Cipta Kerja (Amin, 2022). Kasus terbaru pada hari buruh tahun 2022 yang bertepatan dengan hari raya idul fitri 1443 H para buruh juga melakukan aksi terkait omnibus law. Andi Gani yang sekarang menjabat sebagai Ketua Buruh ASEAN (ATUC), menilai perjuangan tenaga kerja Indonesia masih susah dan lama karena persoalan yang belum terselesaikan, di antaranya permasalahan Omnibus Law dan meminta penghapusan klaster ketenagakerjaan. Para buruh juga menuntut penolakan terhadap UU Cipta Kerja (Anwar, 2022)

Pembicaraan terkait UU Cipta Kerja bukan hanya didunia nyata, didunia media sosial juga ramai diperbincangkan oleh warganet. Mulai dari kekhawatiran warganet terkait kredibilitas UU Cipta Kerja dan implikasi dari penerapannya. Warganet pun banyak yang melontarkan asumsinya atas segala keadaan dari imbas UU Cipta Kerja. Walaupun pengesahan UU ini dilakukan tahun 2020 akan tetapi sampai pada awal tahun 2022 ini isi dan implementasi dari UU Cipta Kerja masih menjadi salah satu topik pembicaraan yang dibahas di *Twitter*. Mengapa perlu mengetahui topik yang dibahas di twitter? Karena pada kuartal pertama tahun 2022 tercatat terdapat 436 juta orang yang aktif menjadi pengguna jejaring sosial *Twitter*. *Twitter* juga sebagai aplikasi jejaring sosial yang menempati peringkat ke 7 sebagai aplikasi jejaring sosial yang paling diminati diseluruh dunia.



Gambar 1.2 Jumlah Audience Twitter Pada Tahun 2022

(Sumber: <https://www.slideshare.net/>)

Twitter dengan pengguna yang begitu banyak dari berbagai kalangan masyarakat akan memunculkan begitu banyak ragam komentar atau opini pada setiap berita yang ada. Masyarakat banyak beropini baik pro dan kontra mengenai dampak adanya UU Cipta Kerja. Kalangan dunia usaha menyambut baik dan memberikan apresiasi kepada pemerintah dan DPR karena telah ada UU Cipta Kerja, akan tetapi Amnesty International Indonesia menilai bahwa UU Cipta Kerja banyak ketentuan dan sangat tidak progresif, ketentuan dalam UU Cipta Kerja banyak melanggar prinsip non-retrogesi sehingga menyebabkan kemunduran dalam aspek pemenuhan hak-hak masyarakat (Mukaromah, 2020).

Guna membantu menyelesaikan masalah pro kontra mengenai UU Cipta Kerja diperlukannya sebuah riset atau analisis terkait pro kontra tentang UU Cipta Kerja yang didapatkan dari tanggapan masyarakat melalui jejaring sosial *Twitter*. Ulasan atau opini dari masyarakat tentang UU Cipta Kerja dapat dijadikan sebagai alat yang efektif dan efisien untuk menentukan informasi terhadap suatu kejadian dan dijadikan sebagai bahan pertimbangan untuk kebijakan-kebijakan selanjutnya. Metode yang dapat digunakan untuk memperoleh informasi opini masyarakat tentang UU Cipta Kerja secara efektif dan efisien salah satunya adalah dengan metode *Text Mining*. Salah satu teknik analisis dengan metode *Text Mining* adalah analisis sentimen. Analisis sentimen digunakan untuk mendeteksi keluhan, persepsi produk atau kebijakan, dan persepsi tentang merek tertentu. Dengan analisis sentimen pemerintah dapat mengetahui informasi tentang tanggapan dan sikap masyarakat terhadap topik diskusi kontekstual dari seluruh dokumen. (Alwasi'a, 2020). Informasi dan hasil yang didapatkan dari proses analisis sentiment tersebut dapat digunakan sebagai acuan pemerintah dalam melakukan revisi atau perbaikan.

Berdasarkan uraian tersebut, *Sentiment Analysis* atau *opinion mining* adalah sebuah proses memahami, mengekstrak dan mengolah data yang berbentuk kata secara otomatis yang bertujuan mendapatkan informasi dalam suatu kalimat tersebut apakah beropini positif atau negatif. Algoritma yang akan digunakan pada penelitian ini adalah dengan metode *Super Vector Machine* (SVM). Metode klasifikasi *Super Vector Machine* (SVM) dipilih karena memiliki akurasi yang baik dalam klasifikasi (Budianto, 2018), mampu menemukan *hyperline* terbaik

sebagai pemisah (Ubaidillah, 2019), proses learning yang cepat dan metode klasifikasi teks yang paling efektif karena dapat mengelola ruang fitur yang besar dan mempunyai kemampuan generalisasi yang tinggi (Dharmadhikari, 2011).

Pada penelitian sebelumnya seiring perkembangan teknologi *Sentiment Analysis* juga dapat digunakan untuk melihat sentiment masyarakat terhadap Komisi Pemberantasan Korupsi (KPK) di Indonesia, seperti yang dilakukan oleh (Dedi , Eka , & Ferico , 2020) dalam penelitiannya yang berjudul “Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia”. Penelitian ini dilakukan untuk melihat opini masyarakat terhadap kinerja Komisi Pemberantasan Korupsi (KPK). Dari hasil klasifikasi menggunakan metode SVM dibagi menjadi tiga kelas, yaitu sentiment positif terhadap KPK sebanyak 8%, sentimen negatif terhadap KPK sebanyak 77%, dan sentimen netral terhadap KPK sebanyak 15%. Maka pada penelitian ini peneliti akan menggunakan metode *Super Vector Machine* (SVM) untuk mengetahui tanggapan dan opini dari masyarakat melalui jejaring sosial *Twitter* tentang UU Cipta Kerja di Indonesia, yang harapannya dari penelitian ini dapat menyajikan klasifikasi analisis sentimen yang baik dan tepat sehingga dapat memberikan informasi yang berguna sekaligus dapat dijadikan bahan rekomendasi bagi pemerintah untuk mengatasi permasalahan dampak dari pengesahannya UU Cipta Kerja di Indonesia.

1.2. Rumusan Masalah

Berdasarkan uraian dari latar belakang, rumusan masalah yang akan dikaji dalam penelitian ini adalah :

1. Bagaimana analisis sentimen UU Cipta Kerja pada media sosial *Twitter* menggunakan metode *Support Vector Machine* (SVM)?
2. Bagaimana tingkat akurasi dari metode *Support Vector Machine* (SVM)?

1.3. Batasan Masalah

Batasan masalah yang peneliti gunakan meliputi :

1. Penelitian ini menggunakan data hasil *Scraping Twitter* pada bulan Maret-Mei 2022

2. Metode yang peneliti gunakan adalah *Support Vector Machine* (SVM).
3. Aplikasi yang peneliti gunakan adalah *Python* dan *RStudio*.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Mengetahui sentimen dari masyarakat terhadap UU Cipta Kerja pada media sosial Twitter menggunakan metode *Support Vector Machine* (SVM).
2. Mengetahui tingkat akurasi yang didapatkan menggunakan metode *Support Vector Machinen* (SVM) dengan beberapa kernel.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan gambaran umum mengenai opini masyarakat terhadap UU Cipta Kerja berdasarkan *tweet* pada media sosial Twitter.
2. Memberikan gambaran tingkat akurasi dari metode *Support Vector Machine* untuk dijadikan acuan pengembangan penelitian berikutnya.
3. Hasil dari analisis sentimen dapat digunakan sebagai bahan rekomendasi dan evaluasi pemerintah terhadap Undang-Undang yang telah ada yang berguna untuk pengembangan Undang-Undang yang lebih baik lagi.

BAB II

TINJAUAN PUSTAKA

2.1. Kajian Induktif

Penelitian terdahulu menjadi acuan penelitian ini untuk mengetahui hubungan metode atau studi kasus dengan penelitian terdahulu dan menghindari duplikasi dengan penelitian terdahulu. Berikut ini adalah studi dari beberapa penelitian tentang analisis sentimen.

Penelitian yang pertama dilakukan oleh (Harijianto, 2019) dengan metode *Multinomial Naïve Bayes* dari hasil cuitan *twitter*. Hasil dari penelitiannya menghasilkan nilai akurasi dengan tidak k-Fold Cross Validation sebesar 72.941%, jika menggunakan k-Fold Cross Validation dengan 3 Fold mendapatkan nilai akurasi 71.601%, 5-Fold sebesar 70.72% dan 10-Fold sebesar 70.68%.

Penelitian ke dua dilakukan oleh (Passi, 2020) dengan membandingkan metode Support Vector Machine, Naive Bayes, K-Nearest Neighbor dan Random Forest untuk mengklasifikasikan ulasan Piala Dunia 2014 menjadi tiga kelas, yaitu positif, netral, dan negatif. Data diambil dari Twitter pada bulan Juni-Juli 2014 dan mendapat 2 juta *tweet*. Hasil dari penelitiannya ini metode Naïve Bayes memiliki performa terbaik, dikarenakan mempunyai nilai akurasi dan nilai AUC tertinggi jika dibandingkan dengan metode lain.

Penelitian ke tiga dilakukan oleh (Aini, 2021) tentang prediksi kelulusan mahasiswa dengan menggunakan metode SMOTE dan SVM. Klasifikasi dengan menggunakan metode SMOTE yang digabungkan dengan metode Support Vector Machine (SVM) menghasilkan tingkat nilai akurasi sebesar 77%, sensitifitas 80%, spesifisitas 74%. Adapun tingkat akurasi tanpa Support Vector Machine (SVM) mendapatkan nilai akurasi 74%, spesifisitas 84% dan sensitifitas 52%.

Penelitian ke empat oleh (Zalyhaty, 2021) tentang tanggapan masyarakat terhadap Vaksin Covid-19 menggunakan metode Support Vector Machine menghasilkan kesimpulan dari 283 tanggapan masyarakat, mayoritas masyarakat bersentimen positif bahwa vaksin mampu melindungi dari virus covid-19, dan sentiment negative masyarakat yang menolak dan ragu terhadap vaksin karena

khawatir efek samping dari vaksin tersebut. Adapun nilai akurasi dari penelitian ini sebesar 82.47%.

Penelitian ke lima dilakukan oleh (Hakim S. N., 2021) mengenai ulasan *myIndiHome* pada tanggal 1 Novemer – 15 Desember 2020 didapatkan 2.539 ulasan pada situs Google Play dengan menggunakan *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC). Nilai akurasi dengan menggunakan metode SVM lebih besar yaitu sebesar 87,54% jika dibandingkan dengan metode NBC dengan akurasi sebesar 84,69%, yang berarti bahwa algoritma SVM memiliki kinerja lebih baik dalam menghasilkan klasifikasi data ulasan *myIndiHome* jika dibandingkan dengan menggunakan metode NBC.

Penelitian ke enam dilakukan oleh (Santoso, 2021) tentang analisis sentimen pengguna Twitter topik #BPJS rasarentenir, menggunakan sampel data sebanyak 200 dengan metode Support Vectore Machine (SVM) telah berhasil dilakukan dan didapatkan nilai akurasi sebesar 94%. Penelitian ini menghasilkan model yang dapat mengklasifikasikan dan memprediksi respon masyarakat terhadap BPJS, apakah respon positif atau negatif. Dan dalam penelitian ini, analisis sentimen publik tentang #BPJSrasarentenir lebih menarah kepada sentiment negatif.

Penelitian selanjutnya ke tujuh dari (Rezki, 2021) yang menggunakan metode K-Nearest Neighbor dalam melakukan analisis klasifikasi emosi pada media sosial twitter. Didapatkan nilai akurasi sebesar 91,04%, *precision* sebesar 89,36%, nilai *recall* sebesar 87,92%, dan *f-1 score* sebesar 88,16% dengan nilai $k = 17$ dan *threshold* 5. Terdapat nilai eror atau hasil kesalahan menurut sistem sebanyak 8,96% atau 43 data, dan kelas yang diprediksi salah oleh sistem adalah kelas berlabel kepercayaan sebanyak 14 data. Hal ini dikarenakan banyak fitur yang berpotongan dan memperkirakan fitur kata dengan kelas selain kelas percaya.

Penelitian ke delapan dilakukan oleh (Aliffia , Intan, & Budi, 2021) terkait analisis sentiment pelaksanaan PILKADA pada saat pandemi dengan metode Support Vectore Machine dengan jumlah data scraping *tweet* dari aplikasi Twitter sebanyak 496 data. Hasil dari klasifikasi data *tweet* dengan menggunakan metode *Support Vector Machine* beserta kernel *linear* mendapatkan tingkat nilai akurasi yang cukup besar yaitu sebesar 92%, dengan nilai *precision* 22%, dan *recall* 67%.

Jumlah kelas negatif lebih banyak daripada kelas positif, diketahui mayoritas masyarakat tidak setuju dengan Pilkada 2020 karena adanya pandemi.

Berdasarkan kajian induktif tentang penelitian-penelitian sebelumnya yang peneliti jadikan sebagai acuan referensi baik secara metode maupun studi kasus penelitian ini. Berikut merupakan hasil perbandingan penelitian-penelitian terdahulu:

Tabel 2.1 Tabel Penelitian Sebelumnya

| Tahun | Nama | Judul | Hasil Penelitian |
|-------|--------------------------|--|---|
| 2019 | Servasius Dwi Harijianto | Analisis Sentimen pada twitter menggunakan Multinomial Naive Bayes | Hasil dari penelitiannya menghasilkan nilai akurasi dengan tidak k-Fold Cross Validation sebesar 72.941%, jika menggunakan k-Fold Cross Validation dengan 3 Fold mendapatkan nilai akurasi 71.601%, 5-Fold sebesar 70.72% dan 10-Fold sebesar 70.68% |
| 2020 | Patel & Passi | Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning | Hasil dari penelitian Patel dan Passi menunjukkan bahwa metode Naïve Bayes memiliki performa terbaik, karena memiliki nilai akurasi dan AUC tertinggi jika dibandingkan dengan metode lain |
| 2021 | Nadiatul Aini | Prediksi Kelulusan Mahasiswa dengan Menggunakan Metode Smote dan Support Vector Manchine (SVM) | Klasifikasi dengan menggunakan metode SMOTE yang digabungkan dengan metode Support Vector Manchine (SVM) menghasilkan tingkat nilai akurasi sebesar 77%, sensitifitas 80%, spesifisitas 74%. Adapun tingkat akurasi tanpa Support Vector Manchine (SVM) mendapatkan nilai akurasi 74%, spesifisitas 84% dan sensitifitas 52%. |
| 2021 | Layla Qodary Zalyhaty | Analisis Sentimen Tanggapan Masyarakat terhadap Vaksin Covid-19 Menggunakan | Didapatkan hasil mayoritas masyarakat bersentimen positif bahwa vaksin mampu melindungi dari virus covid-19, dan |

| | | | |
|------|--|--|---|
| | | Algoritma SVM | sentiment negative masyarakat yang menolak dan ragu terhadap vaksin karena khawatir efek samping dari vaksin tersebut. Adapun nilai akurasi dari penelitian ini sebesar 82.47% |
| 2021 | Sulton Nur Hakim | Analisis Sentimen Persepsi Pengguna Myindihome Menggunakan Metode <i>Support Vector Machine</i> (SVM) Dan Naïve Bayes Classifier (NBC) | Nilai akurasi dengan menggunakan metode SVM lebih besar yaitu sebesar 87,54% jika dibandingkan dengan metode NBC dengan akurasi sebesar 84,69%, yang berarti bahwa algoritma SVM memiliki kinerja lebih baik dari pada NBC |
| 2021 | Gading Teguh Santoso | Analisis Sentimen pada <i>Tweet</i> dengan Tagar #BPJSrasarentenir Menggunakan Metode <i>Support Vectore Machine</i> (SVM) | Metode <i>Support Vectore Machine</i> (SVM) telah berhasil dilakukan dan didapatkan nilai akurasi sebesar 94%. Dan dalam penelitian ini, analisis sentimen publik tentang #BPJSrasarentenir lebih menarah kepada sentiment negatif |
| 2021 | Alia Sri Rezki | Klasifikasi Emosi Pada Twitter Dengan Metode K-Nearest Neighbor (KNN) | Hasil penelitiannya didapatkan nilai akurasi sebesar 91,04%, <i>precision</i> sebesar 89,36%, nilai <i>recall</i> sebesar 87,92%, dan <i>f-1 score</i> sebesar 88,16% dengan nilai $k = 17$ dan <i>threshold 5</i> |
| 2021 | Aliffia Kulsumarwati, Intan Purnamasari & Budi Arif Dermawan | Penerapan SVM Dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi | Dengan metode <i>Support Vector Machine</i> kernel <i>linear</i> didapatkan nilai akurasi yang cukup besar yaitu sebesar 92%, dengan nilai <i>precision</i> 22%, dan <i>recall</i> 67%. Jumlah kelas negatif lebih banyak daripada kelas positif. |

Dari beberapa penelitian yang telah dilakukan terdapat perbedaan dengan penelitian yang dilakukan pada saat ini salah satunya adalah perbedaan studi kasus dan perbedaan beberapa metode yang digunakan pada penelitian terdahulu.



BAB III

LANDASAN TEORI

3.1. Twitter

Twitter termasuk dalam layanan media sosial *microblog*, sehingga *user* dapat membuat suatu *long message* terbatas dan bisa opsional yang juga bisa ditambahkan multimedia (Nugraheni & Yuni, 2017). *Twitter* merupakan layanan yang memungkinkan user dalam melakukan pengiriman dan pembacaan pesan tidak melebihi dari 280 karakter yang biasa disebut sebagai *tweet*. *Twitter* sebelumnya hanya mampu melakukan *tweet* 140 karakter, akan tetapi pada 7 November 2017 bertambah menjadi 280 karakter. *Twitter* didirikan oleh Jack Dorsey di San Francisco, Amerika Serikat pada 21 Maret 2006 (D'monte, 2009).

Menurut (Tamburion, 2015) terdapat beberapa istilah-istilah yang digunakan pada *Twitter*, yaitu:

- a. *Tweet* : sebuah informasi yang tidak lebih dari 280 karakter yang berisi berita terbaru yang berkaitan dengan akun *user*.
- b. *Reply tweet / response tweet* (RT) ialah komentar serta balasan atas *tweet*.
- c. *Retweet* adalah membagikan informasi yang telah di bagikan.
- d. *Follow* adalah mengikuti akun serta informasi yang disampaikan oleh pengguna nya.
- e. *Follower* adalah mengikuti akun pengguna lainnya.
- f. *Hastags* : atau biasa dikenali dengan tanda pagar (#) merupakan tanda yang menandai kata kunci pada topik diskusi dan informasi yang dibagikan.
- g. *Mention* : merupakan cara untuk menyebutkan nama akun pengguna lain untuk diajak berkomunikasi dengan cara ditambahkan awalan “@”.
- h. *Direct Message* (DM) : merupakan fasilitas kirim dan balas pesan yang digunakan oleh pengguna secara *private* terhadap pihak yang diikuti.

- i. *Timeline* : merupakan *tweet* terbaru dari pengguna atau pemilik *Twitter* yang diikuti pengguna akun, termasuk *tweet* yang dibuat oleh pengguna akun.
- j. *Trending topics* : merupakan tema atau isu yang sedang ramai diperbincangkan di *Twitter*

Twitter API (Application Programming Interface) merupakan layanan yang biasa digunakan dalam pengolahan data ketika sebuah software dibangun. *Twitter API* memiliki satu yang dapat memproses tugas yang diminta, sehingga dengan memanggil fungsi perangkat lunak, dapat memanggil fungsi perangkat lunak yang dibangun. *Twitter API* dibangun dengan arsitektur REST (*Representational State Transfer*), memungkinkan untuk digunakan dengan berbagai format data seperti XML dan JSON. *Twitter API* terdiri dari *Twitter Streaming API* dan *Twitter Search API*. Perbedaannya adalah bahwa *Twitter Search API* berfokus pada fungsi pencarian sebelumnya, sedangkan *Twitter Streaming API* berfokus pada fungsi pencarian yang akan datang. (Nugraheni & Yuni, 2017).

3.2. UU Cipta Kerja (*Omnibus law*)

3.2.1 Definisi *Omnibus law*

Menurut (Raharjo S. , 1981) *omnibus law* berasal dari dua kata yakni *omnibus* dan *law*. Definisi *omnibus* berasal dari bahasa latin yang artinya “untuk semuanya”. Adapun definisi *law* berarti “hukum”. Sedangkan kata *omnibus* jika digabungkan dengan kata *law* maka akan membentuk sebuah arti kata baru yakni hukum untuk semua. (Gamer, 2004) dalam kamus *Black’s Law Dictionary* mencetuskan konsep *omnibus law* :

The Omnibus Bill is a single bill that contains many different things, usu. designed in such a way as to compel the executive to accept all unrelated minor provisions or to veto major provisions. An omnibus bill is a bill that deals with proposals related to a specific subject, such as an omnibus Judiciary Bill that covers all proposals for new judges or a crime bill that deals with new crimes and crime control.

Menurut (Santo, 2019) *omnibus law* adalah suatu cara atau konsep pembuatan peraturan yang menggabungkan beberapa regulasi dengan substansi regulasi yang berbeda menjadi satu regulasi besar yang berfungsi sebagai satu

payung hukum. Ketika peraturan itu disahkan, akibatnya beberapa peraturan hasil penggabungan itu dicabut dan selanjutnya dinyatakan tidak berlaku, baik sebagian maupun keseluruhan.

Sedangkan *omnibus law* menurut (Busroh, 2017) merupakan konsep untuk merancang undang-undang yang isinya mengubah dan/atau mencabut beberapa regulasi dalam beberapa undang-undang. Konsep ini beberapa berkembang di wilayah negara *common law* dengan menggunakan sistem hukum *anglo saxon*. Negara-negara tersebut seperti Amerika Serikat, Kanada, Belgia dan Inggris. Tawaran untuk mengimplementasikan konsep ini biasanya dibuat dikarenakan regulasi yang terlalu gemuk dan aturan yang tumpang tindih. Jika masalah regulasi diselesaikan dengan seperti biasa seperti merevisi undang-undang, hal tersebut akan memakan waktu lebih lama dan menghabiskan terlalu banyak anggaran.

3.2.2 Tujuan *Omnibus law*

Busroh menambahkan bahwa setiap sesuatu ketentuan diterapkan dalam sebuah bidang maka pasti memiliki tujuan. Seperti konsep *omnibus law* dalam membentuk suatu regulasi, adanya tujuan dari konsep tersebut dapat dikatakan sebagai kelebihan termasuk dalam konsep *omnibus law*. Diantara tujuannya adalah :

1. Mengatasi konflik regulasi perundang-undangan dengan cepat, efektif dan efisien;
2. Menyelaraskan regulasi pemerintah baik tingkat pusat maupun daerah sebagai langkah penunjang iklim investasi;
3. Mampu memotong rantai birokrasi yang bertele-tele;
4. Pengurusan perizinan lebih terstruktur, efisien dan efektif;
5. Kepastian dan perlindungan hukum untuk pengambil kebijakan lebih terjamin;
6. Hubungan koordinasi antar instansi terkait lebih meningkat dikarenakan telah diatur dalam *omnibus regulation* yang terpadu.

Penyusunan regulasi dengan konsep *omnibus law* akan sudah diterapkan dan dijalankan di lapangan jika tidak ada partisipasi *stakeholder* dan koordinasi yang bagus antar instansi terkait. Tentu saja substansi peraturan perundang-undangan

yang baru harus dibuat semudah dan seefisien mungkin. Maka dari itu konsep *omnibus law* juga tidak bisa lepas dari beberapa kelemahan diantaranya :

1. Kemungkinan membuka peluang lebih banyak ditolak pada saat paripurna atau lebih dikenal sebagai *judicial review* terhadap kebijakan *omnibus regulation* yang diterbitkan;
2. Proses pembentukan peraturan perundang-undangan yang tidak melibatkan legislatif membuat seakan-akan legislatif merasa “dikebiri”.
3. Orientasi kebijakan hukum pemerintah yang dapat diubah-ubah sesuai kehendak rezim yang memerintah akan banyak mempengaruhi stabilitas sistem hukum nasional.

3.2.3 Omnibus law di Indonesia

Awal tahun 2020, pemerintah sedang menyiapkan RUU tentang Cipta Kerja dengan mengadopsi konsep *Omnibus law*. Pada prosesnya RUU ini disiapkan oleh pemerintah untuk digunakan sebagai suatu strategi dalam upaya membangun perekonomian Indonesia sehingga para investor tertarik untuk menanamkan modalnya di Indonesia. Pemerintah menganggap perlu adanya RUU Cipta Kerja ini dikarenakan tingginya angka pengangguran di Indonesia yang menyentuh angka 7 juta orang, sehingga harapan pemerintah RUU ini mampu menciptakan banyak lapangan kerja. Kini *omnibus law* Cipta Kerja resmi disahkan oleh pemerintah dalam rapat paripurna yang diselenggarakan di Gedung DPR yang berada di Senayan Jakarta pada 5 Oktober 2020 dengan jumlah halaman total 905 (Kemenkumham, 2020). Konsep *Omnibus law* ini merupakan konsep baru dalam sistem peraturan di Indonesia. Sistem ini biasa disebut dengan Undang-Undang sapu jagat karena menggantikan beberapa norma-norma hukum menjadi satu peraturan. Selain itu konsep seperti ini dijadikan misi untuk memotong beberapa norma hukum yang dianggap tidak relevan dengan perkembangan zaman dan dirasa merugikan negara. (Prabowo, 2020)

Dalam *booklet* UU Cipta Kerja dipaparkan bahwa dengan menggunakan Teknik *omnibus law* sekitar delapan puluh undang-undang dan lebih dari seribu dua ratus pasal mampu direvisi sekaligus hanya dengan satu UU *multisector* yakni UU Cipta Kerja dengan 11 klaster yaitu :

1. Persyaratan investasi

2. Ketenagakerjaan
3. Penyederhanaan perizinan
4. Kemudahan berusaha
5. Administrasi pemerintahan
6. Pengenaan sanksi
7. Dukungan riset dan inovasi
8. Investasi dan Proyek Pemerintah
9. Pengadaan lahan
10. Kemudahan, Pemberdayaan, dan Perlindungan UMKM
11. Kawasan ekonomi

Selain daripada itu, pemerintah menilai bahwa UU Cipta Kerja mempunyai beberapa manfaat diantaranya :

1. Memperbaiki iklim investasi dan mewujudkan kepastian hukum.
2. Menciptakan banyak lapangan kerja yakni tiga juta setiap tahunnya dibandingkan dengan tanpa *omnibus law* yang hanya dua sampai dua setengah juta pertahun.
3. Penyederhanaan perizinan berusaha.
4. Kemudahan perizinan dasar. Menyederhanaan perizinan dasar terkait tiga sektor yakni izin lokasi, lingkungan dan bangunan Gedung.
5. Persyaratan investasi dipermudah. Karena regulasi persyaratan berinvestasi menjadi lebih sederhana.
6. Percepatan birokrasi perizinan usaha. UU Cipta Kerja membuat proses berusaha menjadi lebih cepat dan mudah.
7. Pengembangan bisnis halal. Pemerintah meyakini UU Cipta Kerja akan mendongkrak pengembangan produk halal termasuk didalamnya kemudahan akses dan birokrasi bagi usaha mikro dan kecil (UMK).
8. Perlindungan dan jaminan bagi pekerja. Salahsatunya dengan program jaminan kehilangan pekerjaan yang didalamnya memuat tidak bertambahnya beban iuran pekerja atau pengusaha.
9. Percepatan pengadaan tanah. UU Cipta Kerja mempermudah pengadaan tanah untuk pembangunan dengan tujuan untuk kemaslahatan umum dan investasi melalui Bank Tanah.

10. Dukungan BUMN untuk riset dan inovasi. Dalam hal ini pemerintah bisa menugaskan Badan Usaha Milik Negara (BUMN) untuk melakukan pengembangan riset dan inovasi di pusat atau di berbagai daerah.

Akan tetapi terlepas dari manfaat UU Cipta Kerja menurut pemerintah, (Fuady, 2013) memaparkan UU Cipta Kerja yang saat ini menggunakan konsep *omnibus law* sebagai pengubah sebagian Undang-undang Ketenagakerjaan sedikit banyak akan berdampak kepada perubahan tatanan sosial. Berubahnya hukum yang mampu mempengaruhi perubahan sosial harus selaras dengan fungsi hukum sebagai sarana perubahan sosial atau sarana rekayasa masyarakat. Dengan demikian UU Cipta Kerja sebagai sumber hukum harus mampu menjawab tantangan ketika dihadapkan pada persoalan ketenagakerjaan di masa mendatang demi perubahan ke arah yang lebih baik bagi seluruh *stakeholder*, termasuk perlindungan tenaga kerja.

3.3. Text Mining

Text Mining merupakan sebuah teknologi yang dapat digunakan untuk menganalisis data teks baik berupa teks semi struktur ataupun tidak struktur (Han, Kamber, & Pei, 2012). Sedangkan menurut (Maria, 2014) *Text Mining* merupakan proses penggalian data dalam bentuk teks dimana sumber datanya biasanya diperoleh dari dokumen dan bertujuan untuk menemukan kata-kata yang dapat mewakili isi dari dokumen, sehingga hubungan antar dokumen dapat dianalisis.

Dalam suatu proses *Text Mining*, dokumen yang akan digunakan harus dilakukan pemrosesan terlebih dahulu, Adapun langkah-langkah dalam melakukan *Text Mining* adalah sebagai berikut ini:

3.3.1 Text Preprocessing

Text Preprocessing adalah tahap awal yang dilakukan untuk mempersiapkan teks yang terstruktur agar dapat diproses sesuai kebutuhan. *Preprocessing* umumnya digunakan untuk mengubah informasi dari setiap sumber data ke dalam bentuk atau format standar berupa nilai numerik yang dapat digunakan sebagai sumber data yang nantinya dapat diolah dan diproses dengan berbagai metode (Feldman & Sangar, 2006).

Pre-processing mempunyai beberapa tahapan antara lain:

a. Cleaning Data

Merupakan pembersihan kata untuk menghilangkan atribut-atribut atau tanda baca yang tidak terdapat hubungan dengan data yang bertujuan untuk mengurangi *noise* seperti simbol (!@#%\$%^&*():{ }.,?~/[]), *URL*, *hashtag*, *emoticon* dan *mention*.

b. Case Folding

Case folding merupakan langkah perubahan untuk mengkonversi karakter huruf “a-z” diubah menjadi huruf kecil. *Case folding* juga menghilangkan bagian yang dianggap tidak valid atau menghilangkan karakter seperti angka dan lain-lain.

c. Tokenizing

Tokenizing merupakan proses pemotongan suatu kalimat berdasarkan kata perkata yang menyusunnya. Tahapan ini juga dapat menghilangkan beberapa karakter dalam kalimat yang dianggap sebagai tanda baca.

d. Filtering

Filtering merupakan tahapan dengan menggunakan *stopword removal* yakni daftar kata yang akan dihapus pada dokumen seperti “dan”, “di”. Proses *filtering* juga terkadang terdapat penghapusan terhadap spasi yang berlebihan.

3.4. Pembobotan TF-IDF

Menurut (Kamal, 2021) data yang telah melewati tahap *preprocessing* harus berupa angka-angka, sehingga data tersebut dapat dihitung dan diolah. Data diubah menjadi *vector* kemudian diberikan bobot dan nilai setiap kata dan selanjutnya diolah menggunakan prediksi algoritma.

Pembobotan kata dapat menggunakan algoritma *Term-Frequency Times Inverse Document Frequency* (TF-IDF). Algoritma ini dinilai mempunyai nilai keakurasian dan efisiensi. Metode ini mencari nilai *Term Frequency* (TF) dan *Invers Document Frequency* (IDF) pada masing-masing kata pada tiap dokumen. Secara sederhana, algoritma ini dipakai untuk mencari kata yang sering muncul atau *term* dalam suatu dokumen.

Term Frequency (TF) merupakan kemunculan frekuensi suatu kata atau *term* dalam suatu dokumen. Semakin frekuensi kata yang muncul besar (TF tinggi) maka semakin memberikan nilai kemiripan kata yang semakin besar.

Ada beberapa macam *Term Frequency* (TF) diantaranya:

b. TF murni (*raw TF*)

Nilai TF murni bergantung pada jumlah munculnya kata dalam dokumen. Seperti contoh jika muncul lima (5) kali maka kata tersebut juga bernilai lima (5).

c. TF biner (*binary TF*)

TF biner hanya menilai ada atau tidaknya kata dalam dokumen. Jika terdapat kata maka diberi nilai satu (1) jika tidak ada katanya maka diberi nilai nol (0).

d. TF logaritmik

TF logaritmik dipakai untuk mengatasi dokumen yang memiliki sedikit *term* dalam *query* akan tetapi mempunyai frekuensi tinggi.

$$tf = 1 + \log(tf) \quad (3.1)$$

e. TF normalisasi

Tf normalisasi menghitung perbandingan antara jumlah frekuensi kata dengan jumlah semua kata pada suatu dokumen.

$$tf = \frac{\text{Jumlah frekuensi kata terpilih}}{\text{Jumlah kata}} \quad (3.2)$$

Sedangkan *Inverse Document Frequency* (IDF) adalah metode yang digunakan untuk menghitung berapa kata yang diperlukan dalam sebuah dokumen. Metode *Inverse Document Frequency* (IDF) dipakai untuk memboboti penelusuran informasi dalam dunia *Text Mining*. Rumus *Inverse Document Frequency* (IDF) sebagai berikut :

$$idf_j = \log \left(\frac{D}{df_j} \right) + 1 \quad (3.3)$$

Keterangan :

D : Jumlah dari keseluruhan dokumen

df_j : Jumlah dokumen yang memiliki *term* t

Penggabungan perhitungan TF murni dan rumus IDF dengan mengalikan keduanya menghasilkan rumus TF-IDF sebagai berikut:

$$w_{ij} = tf_{ij} \times idf_j \quad (3.4)$$

$$w_{ij} = tf_{ij} \times \left(\log \left(\frac{D}{df_j} \right) + 1 \right) \quad (3.5)$$

Keterangan :

w_j : Bobot term t_j terhadap dokumen d_j

tf_{ij} : Jumlah munculnya *term* t_j dalam dokumen d_j

D : Jumlah keseluruhan dokumen

df_j : Jumlah dokumen yang memiliki term t_j

3.5. Analisis Sentimen

Analisis Sentimen merupakan suatu metode yang digunakan untuk menilai produk atau layanan untuk menentukan hasil opini atau perasaan terhadap suatu produk atau layanan tersebut (Haddi, Liu, & Shi, 2013). Adapun menurut (Medhat, 2014) analisis sentimen merupakan sebuah studi komputasi tentang pendapat, sifat, dan emosi seseorang terhadap suatu objek. Perlu diketahui bahwa objek yang dimaksud berupa produk, aktivitas, layanan atau bahkan topik. Secara umum, analisis sentimen digunakan untuk membantu dalam pengumpulan informasi yang nantinya akan digunakan untuk memilah nilai positif, negatif dan netral terhadap sebuah topik atau produk tertentu. Produk atau topik yang mendapatkan nilai sentimen positif yang akan direkomendasikan kepada pihak tertentu. Biasanya orang mengungkapkan pendapat mereka tentang produk atau layanan melalui *blog*, *twitter*, atau *platform* serupa lainnya.

Menurut (Pradhan, 2016) terdapat beberapa tantangan dalam melakukan analisis sentimen, salah satunya adalah kalimat yang memiliki nilai positif dalam satu situasi akan menjadi negatif dalam situasi lain, pengguna atau publik tidak selalu memberikan pendapat yang sama, pada *platform twitter* yang memiliki karakter terbatas biasanya komputer akan sulit untuk mengekstrak opini pengguna atau publik. Adapun menurut (Dehhaf, 2012) fungsi analisis sentimen ialah mengelompokkan teks, selanjutnya menentukan pendapat dari teks yang diekspresikan ke dalam kalimat yang dianalisis apakah teks tersebut bersifat positif, negatif, atau bahkan netral. Analisis sentimen dapat digunakan dalam proses pelacakan suatu produk, merek, jasa, dan target pasar konsumen. Selain itu,

analisis sentimen dapat digunakan untuk menilai keunggulan atau kelemahan suatu produk berdasarkan keluhan atau persepsi konsumen.

Menurut (Walaa, Ahmed , & Hoda, 2014) analisis sentiment dibagi menjadi 3 kategori : sentiment positif, sentiment negatif dan sentiment netral yang biasanya berupa pendapat, tanggapan atau opini terhadap suatu subjek tertentu baik menggunakan media sosial ataupun media yang lain. Terdapat 3 klasifikasi dalam analisis sentiment :

1. Level dokumen

Level analisis sentimen dokumen merupakan klasifikasi dari suatu dokumen opini yang menunjukkan bahwa dokumen opini tersebut menunjukkan apakah opini positif, negative ataupun netral dan menganggap bahwa isi dari keseluruhan dokumen tersebut berisi opini dari satu topik saja.

2. Level kalimat

Analisis sentimen kalimat dapat mengklasifikasikan sentimen opini dalam setiap kalimat. Pada tataran kalimat, hal ini menentukan apakah kalimat tersebut mengungkapkan sentimen positif, negatif, atau netral.

3. Level aspek

Pada tingkat aspek, analisis sentimen mengklasifikasikan kelas berdasarkan karakteristik tertentu. Pembuat opini dapat berbeda pada aspek yang berbeda dari entitas yang sama, seperti "kualitas suara ponsel ini buruk, tetapi baterai dapat bertahan lama."

3.6. Klasifikasi

Klasifikasi merupakan proses membentuk model atau fungsi yang dapat membedakan dan menggambarkan kelas data. Klasifikasi juga bertujuan untuk memprediksi kelas dari suatu objek yang kelasnya tidak diketahui. Sebelum melakukan proses klasifikasi, pertama-tama adalah membangun model klasifikasi dari sekumpulan kelas data yang sudah ada sebelumnya (*training data*), menggunakan model untuk klasifikasi pengujian data (prediksi) dan terakhir mengukur keakuratan model. Klasifikasi dapat digunakan dalam berbagai aplikasi seperti diagnostik seperti medis, aplikasi kredit perbankan, selective marketing, dan analisis sentimen. Klasifikasi dapat disajikan dalam berbagai model klasifikasi seperti *Naive Bayes Classifier (NBC)*, *Decision Trees*, *K Nearest-*

Neighbourhood classifier (KNN) dan *support vectore machine (SVM)*. (Santoso, 2021)

Terdapat tahapan-tahapan dalam proses klasifikasi. Salah satunya yaitu *learned model*, tahapan ini merupakan tahapan yang di dalamnya terdapat pengembangan model menggunakan hasil analisis *record database* dari beberapa serangkaian kelas. Setiap *record* diasumsikan memiliki kelas yang telah ditentukan yang didasarkan pada atribut kelas label, maka dari itu klasifikasi termasuk dalam *supervised learning*. Tahap ini disebut juga dengan tahapan belajar atau latihan. Algoritma dalam proses klasifikasi akan membangun sebuah model klasifikasi dengan mempelajari dan menganalisis data *training*. Tahapan proses *training* juga dapat dilihat sebagai tahap pembentukan fungsi atau pemetaan $y = f(x)$, di mana y adalah kelas yang diprediksi dan x adalah *record* yang ingin diprediksi kelasnya. (Hakim S. N., 2021)

Baik buruknya hasil klasifikasi dapat dilihat dan dianalisis berdasarkan beberapa hal berikut :

12. *Accuracy*

Accuracy merupakan jumlah dari proporsi dengan nilai prediksi benar. Tingkat akurasi dapat digunakan untuk mengukur tingkat ketepatan antara nilai actual dengan nilai prediksi (Hakim S. N., 2021).

13. *Precision*

Precision dapat diasumsikan adalah cara yang dapat dipakai untuk melihat seberapa banyak item yang “tepat” antara informasi yang diminta dengan jawaban yang diberikan sistem (Hakim S. N., 2021).

14. *Recall*

Recall merupakan proporsi jumlah dokumen teks yang cocok di bawah kendali diantara semua dokumen teks yang relevan dalam koleksi. Mengingat *recall* dipakai sebagai ukuran keberhasilan sistem dalam *reinventing* informasi (Manning, Raghavan, & Schutze, 2009).

15. *F-Measure*

F-Measure adalah kata-kata yang harmonis yang dikeluarkan oleh nilai *recall* dan *precision* sehingga mampu memberikan suatu penilaian yang lebih

seimbang. *F-Measure* juga dapat dipakai pengukuran kinerja sistem secara keseluruhan dalam klasifikasi (Manning, Raghavan, & Schutze, 2009).

3.7. *Confusion Matrix*

Confusion matrix adalah metode untuk menghitung akurasi dalam konsep data mining. Tabel yang berisi jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan salah dapat digunakan untuk menggambarkan *confusion matrix*. (Rahman, Darmawidjadja, & Alams, 2017).

Menurut (Sasongko, 2016) *Confusion matrix* merupakan sebuah alat ukur berbentuk matrix dengan menggunakan algoritma tertentu yang dapat digunakan untuk mendapatkan suatu jumlah ketepatan klasifikasi terhadap kelas yang dipakai. Adapun tabel *Confusion matrix* dapat dilihat pada **Tabel 3.1**.

Tabel 3.1 Tabel *Confusion matrix*

| <i>Confusion matrix</i> | | <i>Nilai Aktual</i> | |
|-------------------------|-------------------------|----------------------------|----------------------------|
| | | <i>Aktual Positif</i> | <i>Aktual Negatif</i> |
| <i>Nilai Prediksi</i> | <i>Prediksi Positif</i> | <i>True Positive (TP)</i> | <i>False Positive (FP)</i> |
| | <i>Prediksi Negatif</i> | <i>False Negative (FN)</i> | <i>True Negative (TN)</i> |

Penjelasan :

1. *True Positive (TP)* : adalah nilai yang dihasilkan dari prediksi pada klasifikasi positif dan nilai sebenarnya positif.
2. *True Negative (TN)* : adalah nilai yang dihasilkan dari prediksi pada klasifikasi negatif padahal nilai sebenarnya positif.
3. *False Positive (FP)* : adalah nilai yang dihasilkan dari prediksi pada klasifikasi negatif padahal nilai sebenarnya positif.
4. *False Negative (FN)* : adalah nilai yang dihasilkan dari prediksi pada klasifikasi negatif dan nilai sebenarnya negatif.

Dari penjelasan tabel dan deskripsi, rumus *recall*, *precision*, akurasi, *Spesificity*, *FPR* dan *AUC* dapat dicari dengan :

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3.6)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (3.7)$$

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3.8)$$

$$Spesificity = \frac{TN}{TN+FP} \times 100 \quad (3.9)$$

$$False Positive Rate (FPR) = 1 - spesificity \quad (3.10)$$

$$Area Under Curve = \frac{1+Recall-FPR}{2} \quad (3.11)$$

AUC (area under curve) merupakan luas area di bawah kurva. Luas area dari *AUC* selalu berada dalam interval nilai 0 hingga 1. Nilai *AUC* didapatkan berdasarkan gabungan luas trapesium titik-titik (*sensitivity* dan *specificity*). Secara umum *AUC (area under curve)* dipakai dalam pengukuran performa menggunakan perkiraan peluang dari hasil sampel yang telah dipilih secara acak dari suatu populasi positif dan negatif.

Standar tabel yang digunakan untuk mengklasifikasikan nilai berdasarkan *AUC (area under curve)* dapat dilihat pada Tabel 3.1 (Sasongko, 2016)

Tabel 3.2 Tabel Nilai *AUC*

| Nilai <i>AUC</i> | Keterangan Klasifikasi |
|------------------|------------------------|
| 0.90 - 1.00 | Sangat Baik |
| 0.80 - 0.90 | Baik |
| 0.70 - 0.80 | Cukup |
| 0.60 - 0.70 | Buruk |
| 0.50 - 0.60 | Salah |

3.8. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) merupakan metode yang dirumuskan oleh Boser, Guyon, Vapnik, dan pertama kali dideklarasikan pada tahun 1992 di *Workshop on Computational Learning Theory*. Konsep dasar dari *SVM* sebenarnya merupakan penggabungan dari teori-teori komputasi yang telah ada

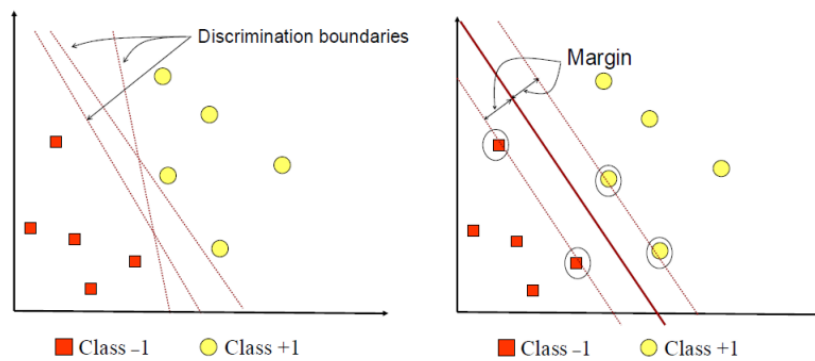
beberapa dekade sebelumnya, seperti halnya *margin hyperplane* (Duda & Hart, 1973).

Menurut (Ariyani, 2016) SVM merupakan salah satu metode yang paling berpengaruh dan ampuh untuk memecahkan klasifikasi. Konsep SVM dapat digambarkan secara sederhana sebagai sebuah usaha mencari *hyperplane* terbaik untuk pemisah antara dua kelas pada ruang input.

Support Vector Machine (SVM) merupakan suatu metode terbaik yang dapat digunakan dalam mengatasi permasalahan klasifikasi. Konsep dari SVM bermula dari permasalahan dua kelas klasifikasi sehingga membutuhkan *training* set negatif atau positif (Pratama, Wihandika, & Ratnawati, 2018).

Sedangkan menurut (Fiska, 2017) SVM merupakan metode yang masih berkaitan dengan *data mining* karena metode SVM merupakan klasifikasi dengan prinsip pengklasifikasi linier. Permasalahan diselesaikan dengan metode SVM dengan persamaan *lagrangian* yang merupakan bentuk ganda dari SVM melalui pemrograman kuadrat. SVM digagas sebagai alternatif untuk SVM standar karena terbukti lebih efisien daripada SVM tradisional dalam pemrosesan data berskala besar.

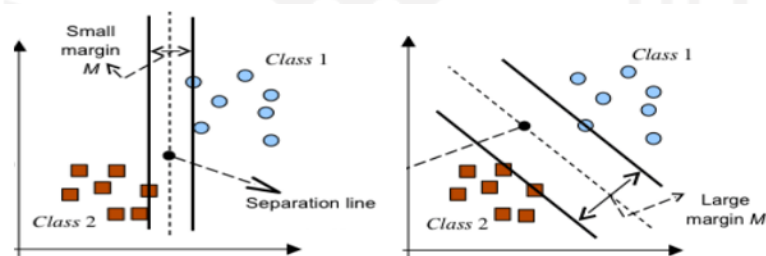
Secara umum, (Hakim S. N., 2021) dalam penelitiannya menjelaskan bahwa proses metode SVM berjalan dapat diilustrasikan seperti pada Gambar 3.1. Gambar sebelah kanan menggambarkan batas diskriminasi dengan margin terbesar yang mungkin. Margin adalah jarak antara dua kelas data yang paling dekat dengan garis *hyperplane*. Ketika *hyperplane* digunakan untuk memisahkan dua kelas data yang berbeda, *hyperplane* dengan margin terbaik akan memberikan generalisasi yang berguna untuk meningkatkan hasil klasifikasi. Sedangkan gambar di sebelah kiri menggambarkan beberapa skenario yang dapat terjadi dalam batas-batas diskriminatif dalam proses SVM untuk mengatur dataset



Gambar 3.1 Proses SVM untuk Menemukan *Hyperline*

Sumber: (Handayani et al, 2020)

Pada gambar 3.1 kategori +1 ditunjukkan oleh warna kuning dan kategori -1 ditunjukkan warna merah. Adapun klasifikasi dapat dilakukan dengan menemukan ketepatan garis (*hyperplane*) yang memisahkan kedua kelompok (antara kuning dan merah). Pada Gambar 3.1 sebelah kiri mengindikasikan terdapat beberapa batas perbedaan alternatif (*discrimination boundaries*). Pada gambar 3.1 bagian (b) menunjukkan *hyperplane* yang optimal dimana garis merah tebal terletak di tengah antara dua kelompok (kuning dan merah), Adapun titik persegi merah dan bulatan kuning pada koordinat dinamakan support vector.



Gambar 3.2 *Hyperline* dan Jarak *Margin*

Gambar 3.2 menjelaskan dua kondisi dari *hyperplane* dan juga jarak margin-nya. Kedua bentuk dapat mengklasifikasikan grup data dengan benar. Hasil yang lebih akurat dalam mengklasifikasikan data ditandai *hyperplane* dan jarak margin yang lebih lebar dibanding *hyperplane* dengan jarak margin yang lebih kecil. Terkadang menggunakan proses *Maximum Marginal Hyperplane* (MMH) untuk mencari *hyperplane* jarak margin terbesar dalam SVM. *Hyperplane* mempunyai formula :

$$w \cdot x_i + b = 0 \quad (3.12)$$

Dengan keterangan :

w : vector bobot

x_i : data ke-i

b : nilai bias

Nilai b jika digunakan sebagai bobot tambahan , w_0 , maka formulanya dapat menggunakan :

$$w_0 + w \cdot x_i = 0 \quad (3.13)$$

Titik berada di atas hyperplane atau titik berada di bawah *hyperplane* maka menggunakan formula :

$$w_0 + w \cdot x_i > 0 \quad (3.14)$$

$$w_0 + w \cdot x_i < 0 \quad (3.15)$$

Svm merupakan metode mengklasifikasikan data menjadi dua kelas. Dimana y_i merupakan kelas data ke-I, didapatkan formula :

$$H_1: w_0 + w \cdot x_i \geq 1 \text{ untuk } y_i = +1 \quad (3.16)$$

$$H_2: w_0 + w \cdot x_i \geq -1 \text{ untuk } y_i = -1 \quad (3.17)$$

(Han, Kamber, & Pei, 2012)

Adapun *Maximum Marginal Hyperplane* (MMH) dapat diukur dengan memaksimalkan jarak hyperplane dengan titik terdekat yakni $\frac{1}{||w||}$. kejadian tersebut bisa disebut masalah Quadratic Programming (QP), dengan memperhatikan constraint pada formula 3.19, Adapun titik minimum dicari dengan formula 3.18

$$\min_w \tau(w) = \frac{1}{2} ||w|| \quad (3.18)$$

$$y_i (x_i \cdot w + b) - 1 \geq 0 \quad (3.19)$$

Permasalahan tersebut dapat dibantu dengan Teknik *Lagrange Multiplier* seperti formula :

$$L(w,b,a) = \frac{1}{2} ||w||^2 - \sum_{i=1}^l (a_i ((x_i \cdot w + b) - 1)) \quad (3.20)$$

dengan $i=1,2,\dots,1$

a_i ialah Lagrange multiplier memiliki nilai 0 atau positif ($a_i \geq 0$). Nilai optimal dapat dihitung dengan meminimalkan nilai L terhadap w dan b , dan tidak memaksimalkan L untuk a_i . Menggunakan nilai $L=0$ untuk menggambarkan sifat pada titik optimal formula 3.20 dapat dimodifikasi dengan memaksimalkan masalah yang hanya terdapat a_i .

$$\sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j x_i x_j \quad (3.21)$$

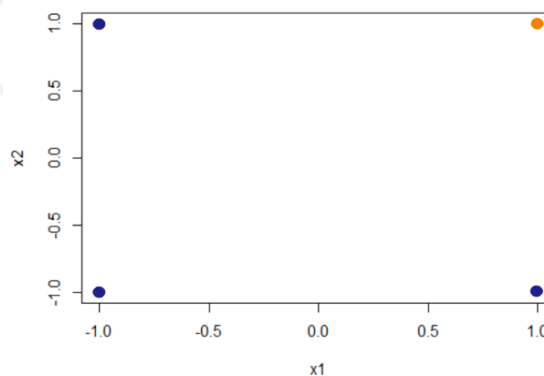
dimana $a_i \geq 0$ ($i = 1, 2, \dots, l$) $\sum_{i=1}^l a_i y_i = 0$

Hasil dari perhitungan nanti akan banyak data bernilai positif, dan data-data yang nantinya berkorelasi dengan a_i dinamakan dengan *support vector* (Nugroho, Witarto, & Handoko, 2013). Selanjutnya peneliti akan menjelaskan simulasi perhitungan *linear SVM*.

Tabel 3.3 Contoh Data *Linear SVM*

| X1 | X2 | Kelas (y) |
|----|----|-----------|
| 1 | 1 | 1 |
| 1 | -1 | -1 |
| -1 | 1 | -1 |
| -1 | -1 | -1 |

Pada Tabel 3.3 merupakan contoh data dimana x_1 dan x_2 merupakan titik data, sedangkan kelas (y) merupakan kelas yang bernilai kelas positif jika bernilai (1) dan kelas negatif jika bernilai (-1). Langkah selanjutnya adalah membuat grafik dari data tersebut.



Gambar 3.3 Grafik Simulasi Data

Pada gambar 3.3 titik berwarna biru merupakan data kelas negatif, dan titik orange merupakan kelas positif. Selanjutnya menentukan titik minimum margin menggunakan rumus.

$$\frac{1}{2} \|W\|^2 = \frac{1}{2} (W_1^2 + W_2^2) \quad (3.22)$$

didapatkan persamaan *hyperplane* nya :

$$(w_1 + w_2 + b) \geq 1 \text{ untuk } y_1 = 1, x_1=1, x_2=1 \quad (3.23)$$

$$(-w_1 + w_2 - b) \geq 1 \text{ untuk } y_2 = -1, x_1=1, x_2=-1 \quad (3.24)$$

$$(w_1 - w_2 - b) \geq 1 \text{ untuk } y_3 = -1, x_1=-1, x_2=1 \quad (3.25)$$

$$(w_1 + w_2 - b) \geq 1 \text{ untuk } y_4 = -1, x_1=-1, x_2=-1 \quad (3.26)$$

Selanjutnya mencari nilai w_1 , w_2 dan b dengan menjumlahkan persamaan-persamaan yang sudah terbentuk sebelumnya.

- a. Menjumlahkan persamaan 1 dan 2

$$\begin{aligned} (w_1 + w_2 + b) &\geq 1 \\ (-w_1 + w_2 - b) &\geq 1 \\ \hline 2w_2 &= 2 \\ w_2 &= 1 \end{aligned} +$$

- b. Menjumlahkan persamaan 1 dan 3

$$\begin{aligned} (w_1 + w_2 + b) &\geq 1 \\ (w_1 - w_2 - b) &\geq 1 \\ \hline 2w_1 &= 2 \\ w_1 &= 1 \end{aligned} +$$

- c. Menjumlahkan persamaan 2 dan 3

$$\begin{aligned} (-w_1 + w_2 - b) &\geq 1 \\ (w_1 - w_2 - b) &\geq 1 \\ \hline -2b &= 2 \\ b &= -1 \end{aligned} +$$

Setelah didapatkan hasil nilai w_1 , w_2 dan b maka hasil dari persamaan *hyperplane* seperti berikut ini.

$$f(x) = 1x_1 + 1x_2 - 1$$

Dari hasil persamaan *hyperplane* dapat menjelaskan bahwa garis *hyperplane* sudah dapat memisahkan dua buah kelas. Selanjutnya menurut (Anto, Arief, &

Handoko, 2003) metode *Support Vector Machine* (SVM) mempunyai beberapa karakteristik diantaranya:

1. Prinsip dari metode *Support Vector Machine* (SVM) adalah *linear classifier*.
2. Prinsip kerja metode *Support Vector Machine* (SVM) pada dasarnya hanya mampu mengklasifikasikan dua kelas.
3. *Pattern recognition* dilakukan dengan mentransformasikan data dalam ruang input ke ruang berdimensi lebih tinggi, kemudian optimisasi dilakukan pada ruang vektor baru. Ini membedakan SVM dari solusi pengenalan pola tipikal, yang melakukan optimasi parameter pada ruang hasil transformasi dengan dimensi ruang input yang lebih rendah.
4. Mengaplikasikan strategi *Structural Risk Minimization* (SRM)

Selain itu metode *Support Vector Machine* (SVM) juga salah satu metode yang mempunyai kelebihan dan kekurangan dalam menjalankannya. Dalam sebuah penelitian, kelebihan dan kekurangan metode harus diperhatikan. Setelah itu metode yang tepat digunakan dengan memperhatikan dari karakteristik data yang akan diteliti. Setiap metode penelitian mempunyai kelebihan dan kekurangan, begitu juga dengan metode SVM. Berikut ini kelebihan dari metode SVM dibandingkan dengan metode lain:

1. Generalisasi

Generalisasi diartikan sebagai kemampuan suatu metode dalam menyelesaikan permasalahan klasifikasi suatu pattern dengan tidak memasukkan data yang digunakan dalam fase pembelajaran metode tersebut. Menurut *Vapnik generalization error* dipengaruhi dua faktor yaitu error kepada *training set* dan dipengaruhi data oleh dimensi VC (*Vapnik-Chervokinensis*). Dalam hal ini SVM selain meminimalisir eror yang lebih kecil pada *training set*, juga meminimalisir faktor yang kedua, atau dalam hal ini biasa disebut *Structural Risk Minimization* (SRM).

2. Landasan Teori

Support Vector Machine (SVM) yang merupakan metode berbasis statistic, mempunyai landasan teori yang mampu dan mudah dianalisa dengan detail dan jelas, dan tidak mempunyai sifat *black box*.

3. *Feasibility*

Proses penentuan support vector yang dapat dirumuskan dalam QP problem menjadikan metode *Support Vector Machine* (SVM) tergolong relatif mudah diimplementasikan. Maka andai saja ada library untuk menyelesaikan QP problem, maka dengan sendirinya metode SVM mudah diimplementasikan. Selain itu juga dapat diselesaikan dengan menggunakan metode sekuensial.

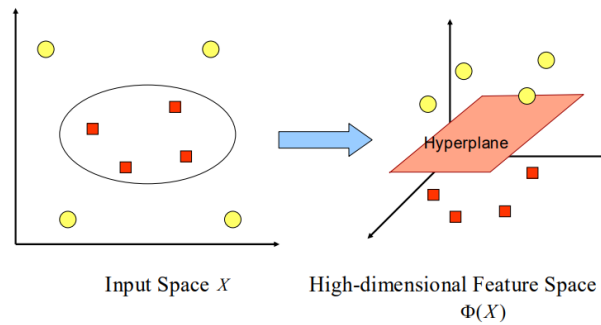
4. *Curse of dimensionality*

Curse of dimensionality diartikan sebagai masalah yang dihadapi oleh metode pengenalan pola di parameter estimasi (misalnya jumlah neuron tersembunyi di jaringan saraf). Karena banyaknya sampel data yang relatif kecil dibandingkan dengan dimensi ruang vektor data alhasil semakin tinggi dimensi ruang vektor informasi diproses, membawa konsekuensi jumlah data yang diperlukan dalam proses pembelajaran. Bahkan terkadang sering terjadi, data yang diproses dengan jumlah terbatas, dan untuk mengumpulkan lebih banyak data tidak memungkinkan dilakukan karena kesulitan teknis dan kendala dalam pembiayaan. Di dalam kondisi ini, jika metode dipaksa harus bekerja pada data dalam jumlah yang relatif lebih kecil dibandingkan dengan dimensinya, maka hal ini akan menjadikan proses estimasi parameter metode tersebut menjadi sangat sulit.

Metode *Support Vector Machine* (SVM) juga mempunyai kekurangan diantaranya :

1. Sulit jika digunakan dalam masalah dengan skala besar. Skala besar disini diartikan sebagai besarnya sampel data yang dianalisis.
2. Metode SVM secara teoritik hanya mampu dalam mengatasi problem klasifikasi dua kelas. Walaupun sebenarnya baru-baru ini telah dimodifikasi dengan menyelesaikan masalah lebih dari dua kelas seperti strategi Tree Structure, namun strategi terbaru dari pengembangan SVM dapat dikatakan masih *multiclass-problem*.

3.9. Support Vector Machine Nonlinear dan Kernel Trick



Gambar 3.4 SVM Non Linear

Sumber : (Nugroho, Witarto, & Handoko , 2003)

Pada kasus data yang sering ditemui di dunia nyata yang berupa teks, sangat jarang ditemukan data linear separable karena hampir pasti bersifat non linear. Untuk mengatasi masalah data non linear, SVM dapat dimodifikasi dengan menggunakan trik yakni kernel. Langkah pertama data dipetakan ke ruang vector lebih tinggi. Pada dimensi ini hyperplane untuk pemisah antar kelas bisa dikonstruksikan. Data pelatihan muncul karena adanya perkalian antara $\phi(X_i) \cdot \phi(X_j)$ dengan $\phi(X)$ merupakan pemetaan dari fungsi *nonlinear* yang dipakai dalam mentransformasikan data pelatihan.

$$K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j) \quad (3.27)$$

Setelah melakukan pemetaan pada dimensi baru, selanjutnya dapat dicari *hyperplane* dengan margin maksimal. Berikut merupakan rumus dari beberapa kernel yang sering digunakan :

1. *Kernel Linear*

$$K(X_i, X_j) = (X_i \cdot X_j) \quad (3.28)$$

2. *Kernel Polynomial*

$$K(X_i, X_j) = (X_i \cdot X_j + 1)^d \quad (3.29)$$

3. *Kernel Sigmoid*

$$K(X_i, X_j) = \tanh(y(X_i \cdot X_j) + c) \quad (3.30)$$

4. *Kernel Gaussian Radial Basis Function (RBF)*

$$K(X_i, X_j) = \exp(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2) \quad (3.31)$$

Keterangan :

$K(X_i, X_j)$: fungsi kernel

$X_i \cdot X_j$: *dot product vector* untuk X_i dan X_j

c : koefisien

d : derajat untuk *polynomial*

y : parameter *gamma*

Tidak terdapat syarat khusus untuk penggunaan linear yang mana dan tingkat akurasi paling tinggi, karena biasanya kernel yang dipilih tidak menghasilkan tingkat akurasi yang berbeda signifikan (Han, Kamber, & Pei, 2012).

3.10. Wordcloud

Menurut (Qeis, 2015) *wordcloud* merupakan sebuah sistem yang dapat menghasilkan visualisasi dari setiap kata-kata dalam sebuah kalimat dengan memberikan sebuah penekanan pada frekuensi kemunculan kata terkait dalam wacana yang tertulis. Penggunaan *wordcloud* dapat memudahkan pembaca atau peneliti dikarenakan mampu memberikan gambaran dengan cepat mengenai isi dari suatu kalimat

(Pradana, 2020) menambahkan bahwa *wordcloud* populer dalam dunia *Text Mining* dikarenakan mudah difahami, dengan menggunakan *wordcloud* gambaran frekuensi kata dapat divisualisasikan dengan cepat dalam bentuk yang menarik namun informatif. Ukuran text pada gambar di *wordcloud* tergantung dengan frekuensi data, semakin banyak frekuensi kata yang digunakan, maka semakin besar ukuran kata yang divisualisasikan di *wordcloud*, begitu juga sebaliknya, semakin sedikit frekuensi kata yang digunakan maka semakin kecil ukuran kata yang divisualisasikan di *wordcloud*. Tampilan *wordcloud* dapat dilihat pada gambar 3.5.

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi Penelitian

Populasi dari data penelitian yang peneliti gunakan merupakan seluruh cuitan atau komentar pengguna media sosial twitter terkait isu atau masalah pro dan kontra UU Cipta Kerja yang telah disahkan oleh DPR RI. Sedangkan peneliti hanya mengambil dan menggunakan sampel sebanyak 3064 data baik *tweet* atau *retweet* terkait isu pro kontra UU Cipta Kerja dengan menggunakan *key word* “omnibus law” dan “cipta kerja”.

4.2. Tempat dan Waktu Penelitian

Peneliti melakukan pengambilan data twitter dengan menggunakan metode *scraping* data mulai tanggal 20 Maret 2022 sampai 20 Mei 2022. Demo buruh sempat menjadi *trending twitter* pada *may day* 2022 yang salah satu tuntutan adalah penolakan *omnibus law*.

4.3. Variabel penelitian

Variabel pada penelitian ini ditampilkan pada Tabel 4.1.

Tabel 4.1 Variabel Penelitian

| Variabel | Definisi Operasional | Skala |
|-------------------|--|----------|
| <i>Text/Tweet</i> | Cuitan atau postingan di <i>Twitter</i> . | Nominal |
| <i>Score</i> | Nilai yang diprediksi oleh mesin pada teks berdasarkan kamus kata positif dan negatif. | Interval |
| Klasifikasi | Kategori sentimen yang berupa teks dari hasil nilai <i>score</i> | Nominal |

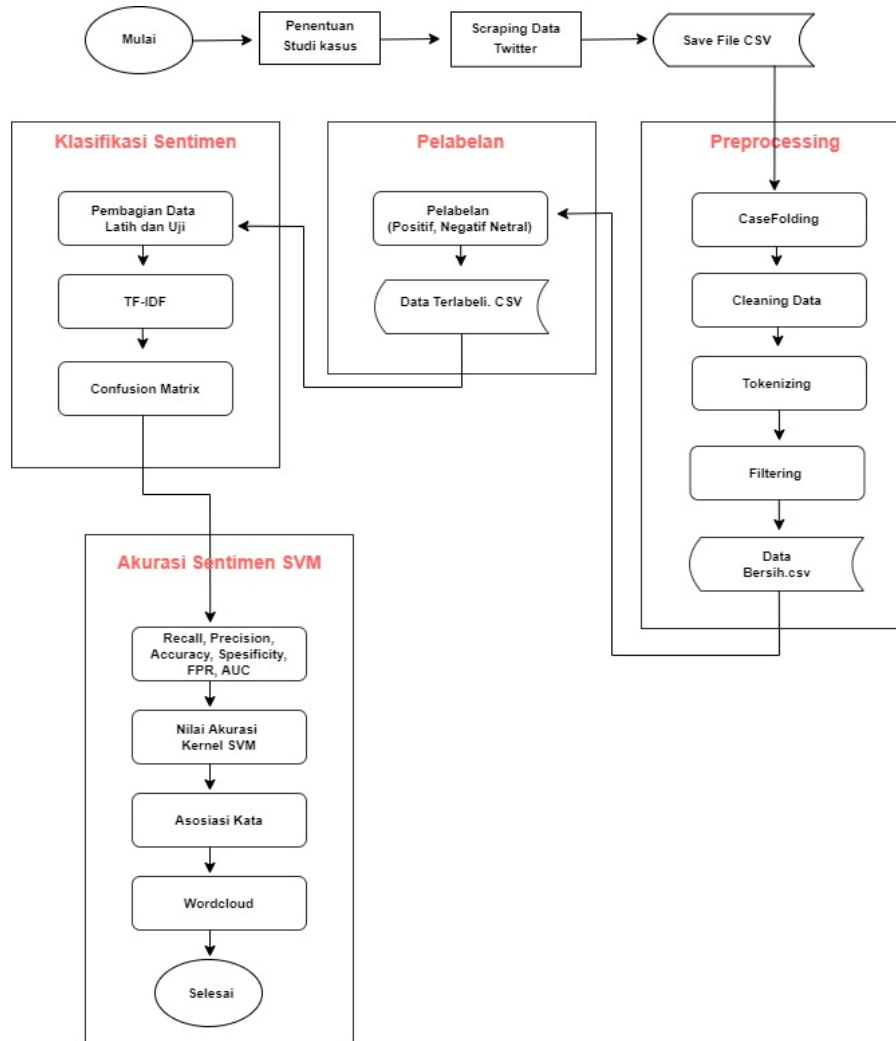
4.4. Teknik Sampling

Peneliti menggunakan data sekunder dalam melakukan pengumpulan data dengan teknik *scraping*. Teknik ini sangat berguna untuk pengambilan data yang berjumlah sangat banyak atau *big data*. Adapun teknik *sampling* dalam proses ini adalah *random sampling* atau biasa disebut Teknik pengambilan sampel secara

acak tanpa memperhatikan strata dalam populasi tersebut dan setiap unit sampel mempunyai peluang terambil yang sama.

4.5. Tahapan Penelitian

Tahapan dalam melakukan analisis data dalam penelitian ini dapat digambarkan pada Gambar 4.1.



Gambar 4.1 Tahapan Penelitian

BAB V

HASIL DAN PEMBAHASAN

5.1. *Scraping Data*

Proses pengumpulan *dataset* yang peneliti gunakan untuk analisis sentimen ini melalui proses *scraping* data dari *twitter*. Data yang peneliti kumpulkan adalah data dari proses *scraping* dari tanggal 20 Maret 2022 sampai 20 Mei 2022. Proses tahapan *scraping* ini peneliti menggunakan *library* dari *python* yaitu *tweepy*. *Library tweepy* merupakan *library* di *python* yang dapat dipakai untuk mengakses atau menyambungkan data *twitter* melalui akses API (*Application Programming Interface*). *Library* ini biasa dipakai oleh para pengembang aplikasi untuk menyempurnakan atau merancang aplikasi yang membutuhkan sinkronisasi data dari *twitter*. Gambar 5.1 merupakan scrip atau perintah instalasi *library python tweepy*.

```
[ ] pip install tweepy
```

Gambar 5.1 *Install Tweepy*

Setelah berhasil *install library tweepy*, sebelum masuk ke proses *scraping* data dengan akses API, terlebih dahulu peneliti melakukan *import* dari *library* yang sudah terinstal. Peneliti melakukan *import tweepy*, *configparser* dan *import pandas as pd* yang terdapat dalam Gambar 5.2.

```
import tweepy
import configparser
import pandas as pd
```

Gambar 5.2 *Import Tweepy, Configparser, Pandas*

Proses *scraping* yang peneliti gunakan mempunyai tiga bagian pemrosesan. Pertama adalah proses *read configs*, *authentication* dan *create dataframe*. *Read configs* merupakan proses aktivasi akses API dengan mengisi API key, API key secret, access token dan acces token secret. *Authentication* merupakan proses pengambilan *tweet* yang didalamnya memuat *key word tweet* yang akan diambil, peneliti menggunakan key word “Omnibus Law” dan “Cipta Kerja. Hasil *scraping* ini selanjutnya membutuhkan tahapan *preprocessing* agar teks *tweet* menjadi bersih dan dapat digunakan untuk proses ke tahap selanjutnya.

5.2. Preprocessing Data

Data cuitan pada *twitter* merupakan data yang berupa teks yang belum terstruktur, hal tersebut terjadi karena data masih terdapat banyak *noise*, sehingga belum dapat dilakukan tahap klasifikasi, perlu adanya proses agar data menjadi lebih terstruktur atau biasa dikenal sebagai *preprocessing* data. Tahapan *preprocessing* digunakan untuk membersihkan data teks, terdapat beberapa tahapan diantaranya, yaitu *case folding*, *tokenizing*, *cleaning* dan *filtering*. Tabel 5.1 merupakan contoh dari tampilan empat data asli yang peneliti ambil dari hasil data *tweet* sebelum dilakukan proses *preprocessing* data.

Tabel 5.1 Sampel Data *Tweet*

| No | User | <i>Tweet</i> |
|----|-----------------|--|
| 1 | EdwardCullen347 | Undang-Undang Cipta Kerja memperkuat, mempercepat dan mempermudah Hilirisasi Riset. #PercepatanEkonomiMaju ##DukungEtlePresisi |
| 2 | refiihsanp | Minyak, Toak Masjid , Omicron, Autobots, Megatron, 3 periode ialah contoh PENGALIHAN ISU dari pada JHT BPJS, Wadas, Omnibuslaw &€ https://t.co/LWldYUyzUA |
| 3 | Fir_bonc | RT @Nissa_4yu: Masih Inget RUU cipta kerja yg bikin syok mayoritas rakyat bawah Indonesia? |
| 4 | Ardiesuhardi1 | YANG DIPERLUKAN BANGSA KITA ITU: 1. Pawang utang negara 2. Pawang rampok BUMN 3. Pawang maling Bansos 4. Pawang <i>omnibus law</i> |

5.2.1 Case Folding

Tahapan *case folding* merupakan langkah perubahan untuk mengkonversi karakter huruf “a-z” kapital diubah menjadi huruf kecil. Contoh hasil dari pemrosesan *tweet* dengan *case folding* ditunjukkan pada Tabel 5.2. Karakter yang berwarna merupakan karakter yang berubah karena melalui proses *case folding*.

Tabel 5.2 Proses *Case Folding*

| No | <i>Tweet</i> | <i>Tweet Setelah Proses Case Folding</i> |
|----|---|---|
| 1 | Undang-Undang Cipta Kerja memperkuat, mempercepat dan mempermudah Hilirisasi Riset. #PercepatanEkonomiMaju ##DukungEtlePresisi | undang-undang cipta kerja memperkuat, mempercepat dan mempermudah hilirisasi riset. #percepatanekonomimaju ##dukungetlepresisi |
| 2 | Minyak, Toak Masjid , Omicron, Autobots, Megatron, 3 periode ialah contoh PENGALIHAN ISU dari pada JHT BPJS, Wadas, Omnibuslaw â€ https://t.co/LWldYUyzUA | minyak, toak masjid , omicron, autobots, megatron, 3 periode ialah contoh pengalihan isu dari pada jht bpjs, wadas, omnibuslaw â€ https://t.co/lwldyuyzua |
| 3 | RT @Nissa_4yu: Masih Inget RUU cipta kerja yg bikin syok mayoritas rakyat bawah Indonesia? | rt @nissa_4yu: masih inget ruu cipta kerja yg bikin syok mayoritas rakyat bawah indonesia? |
| 4 | YANG DIPERLUKAN BANGSA KITA ITU: 1. Pawang utang negara 2. Pawang rampok BUMN 3. Pawang maling Bansos 4. Pawang omnibus law | yang diperlukan bangsa kita itu: 1. pawang utang negara 2. pawang rampok bumn 3. pawang maling bansos 4. pawang omnibus law |

5.2.2 *Cleaning Data*

Pada tahapan ini dilakukan pemrosesan kata untuk mengurangi *noise* seperti penghapusan tanda baca, *HTML/url*, *username/mention*, *emoticon*, angka, *hashtag*, dan lain sebagainya. Hasil dari proses *cleaning* data ditunjukkan Tabel 5.3.

Tabel 5.3 Proses *Cleaning* data

| No | <i>Tweet Setelah Proses Case Folding</i> | <i>Tweet Setelah Proses Cleaning</i> |
|----|---|--|
| 1 | undang-undang cipta kerja memperkuat, mempercepat dan mempermudah hilirisasi riset. #percepatanekonomimaju ##dukungetlepresisi | undang undang cipta kerja memperkuat, mempercepat dan mempermudah hilirisasi riset. |
| 2 | minyak, toak masjid , omicron, autobots, megatron, 3 periode ialah contoh pengalihan isu dari pada jht bpjs, wadas, omnibuslaw â€ https://t.co/lwldyuyzua | minyak, toak masjid , omicron, autobots, megatron, periode ialah contoh pengalihan isu dari pada jht bpjs, wadas, omnibuslaw |
| 3 | rt @nissa_4yu: masih inget ruu cipta kerja yg bikin syok mayoritas rakyat bawah indonesia? | masih inget ruu cipta kerja bikin syok mayoritas rakyat bawah indonesia |
| 4 | yang diperlukan bangsa kita itu: 1. pawang utang negara 2. pawang rampok bumh 3. pawang maling bansos 4. pawang omnibus law | yang diperlukan bangsa kita itu pawang utang negara pawang rampok bumh pawang maling bansos pawang omnibus law |

5.2.3 Tokenizing

Setelah melakukan proses *cleaning* data, selanjutnya adalah proses *tokenizing* atau proses pemecahan dari suatu kalimat atau teks menjadi sebuah kata per kata atau biasa dikenal sebagai “token”. *Tokenizing* berguna untuk mempermudah dalam menghitung frekuensi banyaknya kata yang muncul dalam data tersebut. Hasil dari proses *tokenizing* terdapat pada Tabel 5.4.

Tabel 5.4 Proses *Tokenizing*

| No | <i>Tweet Setelah Proses Cleaning</i> | <i>Tweet Setelah Proses Tokenizing</i> |
|----|---|---|
| 1 | undang undang cipta kerja memperkuat, mempercepat dan mempermudah hilirisasi riset. | 'undang undang', 'cipta', 'kerja', 'memperkuat', 'mempercepat', 'dan', 'mempermudah', 'hilirisasi', 'riset' |
| 2 | minyak, toak, masjid , omicron, autobots, megatron, periode ialah contoh pengalihan isu dari pada jht bpjs, wadas, omnibuslaw | 'minyak', 'toak', 'masjid' , 'omicron', 'autobots', 'megatron', 'periode', 'ialah', 'contoh', 'pengalihan', 'isu', 'dari', 'pada', 'jht', 'bpjs', 'wadas', 'omnibuslaw' |
| 3 | masih inget ruu cipta kerja bikin syok mayoritas rakyat bawah indonesia | 'masih', 'inget', 'ruu', 'cipta', 'kerja', 'yg', 'bikin', 'syok', 'mayoritas', 'rakyat', 'bawah', 'indonesia' |
| 4 | yang diperlukan bangsa kita itu pawang utang negara pawang rampok bumh pawang maling bansos pawang <i>omnibus law</i> | 'yang', 'diperlukan', 'bangsa', 'kita', 'itu', 'pawang', 'utang', 'negara', 'pawang', 'rampok', 'bumh', 'pawang', 'maling', 'bansos', 'pawang', 'omnibus', 'law' |

5.2.4 *Filtering*

Filtering merupakan tahapan yang dilakukan untuk menyaring dan menghapus kata yang tidak memiliki makna sesuai dalam kamus *stopwords*, termasuk juga penghapusan kata-kata penghubung. Hasil dari *filtering* terlihat pada Tabel 5.5.

Tabel 5.5 Proses *Filtering*

| No | <i>Tweet Setelah Proses Tokenizing</i> | <i>Tweet Setelah Proses Filtering</i> |
|----|--|---|
| 1 | 'undang undang', 'cipta', 'kerja', 'memperkuat', 'mempercepat', 'dan', 'mempermudah', 'hilirisasi', 'riset' | 'undang undang', 'cipta', 'kerja', 'memperkuat', 'mempercepat', 'mempermudah', 'hilirisasi', 'riset' |
| 2 | 'minyak', 'toak', 'masjid', 'omicron', 'autobots', 'megatron', 'periode', 'ialah', 'contoh', 'pengalihan', 'isu', 'dari', 'pada', 'jht', 'bpjs', 'wadas', 'omnibuslaw' | 'minyak', 'masjid', 'omicron', 'autobots', 'megatron', 'periode', 'pengalihan', 'isu', 'jht', 'bpjs', 'wadas', 'omnibuslaw' |
| 3 | 'masih', 'inget', 'ruu', 'cipta', 'kerja', 'yg', 'bikin', 'syok', 'mayoritas', 'rakyat', 'bawah', 'indonesia' | 'ruu', 'cipta', 'kerja', 'mayoritas', 'rakyat', 'bawah', 'indonesia' |
| 4 | 'yang', 'diperlukan', 'bangsa', 'kita', 'itu', 'pawang', 'utang', 'negara', 'pawang', 'rampok', 'bumn', 'pawang', 'maling', 'bansos', 'pawang', 'omnibus', 'law' | 'diperlukan', 'bangsa', 'kita', 'pawang', 'utang', 'negara', 'pawang', 'rampok', 'bumn', 'pawang', 'maling', 'bansos', 'pawang', 'omnibus', 'law' |

Setelah proses *filtering* selesai, maka tahapan *preprocessing* data telah selesai dilakukan. Data setelah *preprocessing* menjadi data yang lebih struktur dan mempermudah dalam proses klasifikasi.

5.3. Pelabelan Kelas Sentimen

Teknik yang penulis gunakan dalam proses pelabelan sentimen adalah teknik *hybrid approach*. Teknik ini merupakan gabungan antara pendekatan *lexicon* dan *machine learning*. *Lexicon* merupakan pendekatan dengan cara melabeli kelas sentimen, selanjutnya data diproses dengan pendekatan *machine learning* untuk menguji kinerja model yang didapatkan.

Teknik *hybrid approach* ini biasa digunakan untuk data-data yang sebelumnya belum memiliki pelabelan dengan sentimen apapun, sehingga dalam hal ini penulis menggunakan teknik tersebut. Penulis juga ingin mengetahui kinerja dari model yang didapatkan dengan metode SVM pada proses *machine*

learning approach. Penggabungan pendekatan tersebut penulis berharap dapat meningkatkan kinerja klasifikasi sentimen.

Pada tahapan ini penulis menggunakan *tools RStudio* dengan melakukan pelabelan kelas sentimen menjadi tiga kelas. Yaitu kelas sentimen positif, kelas sentimen negatif dan kelas sentimen netral. Pelabelan sentimen dengan pembobotan *lexicon* mempertimbangkan nilai sentimen secara otomatis. Cara kerja perhitungan nilai sentimen tersebut ialah dengan cara pencarian jumlah kata dengan sentimen positif dan jumlah kata dengan sentimen negatif. Kemudian nilai dari sentimen positif yang didapatkan dikurangi dengan nilai dari sentimen negatif untuk mendapatkan skor sentimen. Beberapa contoh *tweet* yang telah diberikan skor dan pelabelan sentimen ditunjukkan pada Tabel 5.6.

Tabel 5.6 Pelabelan *Tweet*

| No | <i>Tweet</i> | Skor | Sentimen |
|----|---|------|----------|
| 1 | uu cipta kerja akan memperkuat, mempercepat dan mempermudah hilirisasi riset. | 1 | Positif |
| 2 | minyak, toak, masjid , omicron, autobots, megatron, periode ialah contoh pengalihan isu dari pada jht bpjs, wadas, omnibuslaw | -2 | Negatif |
| 3 | masih inget ruu cipta kerja bikin syok mayoritas rakyat bawah indonesia | -1 | Negatif |
| 4 | yang diperlukan bangsa kita itu pawang utang negara pawang rampok bumh pawang maling bansos pawang <i>omnibus law</i> | 0 | Netral |

Tabel 5.6 merupakan gambaran dari perhitungan dengan menggunakan metode *lexicon* untuk menentukan nilai dari skor sentimen dan hasil klasifikasi sentimen. Adapun untuk tabel dari perhitungan skor sentimen lebih detail dapat dilihat pada tabel 5.7.

Tabel 5.7 Contoh Perhitungan Skor

| <i>Tweet</i> | Kata Positif | Kata Negatif |
|---|---------------------|---------------------|
| dampak luar biasa uu cipta kerja tingkatkan iklim usaha dan investasi di indonesia tetapi belum banyaknya tercipta lapangan pekerjaan | tingkat | dampak |
| | | belum |
| | 1 | 2 |

Pada tabel 5.7 dapat dijelaskan dengan perhitungan skor sentimennya sebagai berikut :

$$Skor = \text{Jumlah Kata Positif} - \text{Jumlah Kata Negatif} \quad (5.1)$$

$$Skor = 1-2 \quad (5.2)$$

$$Skor = -1 \quad (5.3)$$

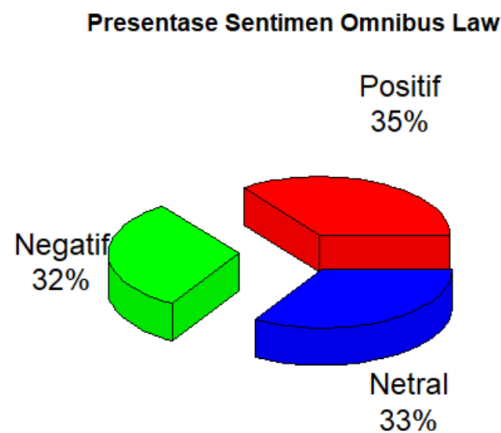
Dari *tweet* yang terdapat pada Tabel 5.7 mendapatkan nilai skor sebesar -1 dengan 1 kata positif dan 2 kata negatif. Karena nilai -1 kurang dari 0, maka *tweet* tersebut dapat dikategorikan kedalam klasifikasi sentimen negatif. Selanjutnya hasil dari pelabelan sentimen keseluruhan data ditunjukkan pada Tabel 5.8.

Tabel 5.8 Klasifikasi Pelabelan *Tweet*

| Jumlah Data | | |
|--------------------|----------------|---------------|
| Positif | Negatif | Netral |
| 1067 | 971 | 1026 |

Dari hasil Tabel 5.8 didapatkan jumlah dari hasil pelabelan dari data *tweet* dengan pelabelan sentimen positif sebanyak 1067 *tweet*, pelabelan sentimen negatif sebanyak 971 *tweet*, dan sisanya sentimen netral sebanyak 1026 *tweet*. Dari hasil pelabelan sentimen tersebut menunjukkan bahwa sentimen positif terhadap UU Cipta Kerja lebih dominan jika dibandingkan dengan sentimen yang lain.

Secara keseluruhan hasil dari klasifikasi sentimen positif, negatif dan netral divisualisasikan pada Gambar 5.5. Dikarenakan data yang peneliti gunakan berjumlah ribuan maka peneliti menjustifikasi bahwa data yang peneliti gunakan merupakan data non linear, yang selanjutnya dapat dipisahkan dengan menggunakan fungsi kernel.



Gambar 5.3 Plot Data Sentimen

5.4. *Statistical Term Frequency-Invers Document Frequency (TF-IDF)*

TF-IDF merupakan tahapan sebelum melakukan klasifikasi untuk memberikan pembobotan suatu kata. Besar kecilnya TF-IDF dipengaruhi oleh banyaknya kemunculan frekuensi kata dalam suatu dokumen.

Kata kunci : uu cipta kerja omnibuslaw

Dokumen 1 (D1) : dampak omnibus iklim usaha investasi indonesia tercipta lapangan pekerjaan jutaan buruh masyarakat

Dokumen 2 (D2) : tolong jokowi omnibus dibatalkan menyusahkan buruh

Dokumen 3 (D3) : kawal omnibuslaw

Nilai TF dapat dilakukan dengan membagi antara jumlah kata yang dicari dalam sebuah dokumen dibagi dengan jumlah kata keseluruhan yang terdapat dalam sebuah dokumen

Tabel 5.9 Perhitungan TF

| Kata | Dokumen | | | | |
|---------|-------------|-----|-----|-----|------|
| | 1 | 2 | ... | ... | 3064 |
| Omnibus | $1/12=0.08$ | ... | ... | ... | 0 |
| Buruh | $1/12=0.08$ | ... | ... | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Demo | 0 | ... | ... | ... | 0 |

| | | | | | |
|---------------|----|----------|-----|-----|---|
| Jokowi | 0 | 1/6=0.16 | ... | ... | 0 |
| Jumlah | 12 | ... | ... | ... | 0 |

Berdasarkan **Tabel 5.9** diketahui bahwa salah satu sampel yakni kata “omnibus” pada dokumen 1 berjumlah 1 dan jumlah keseluruhan kata pada dokumen 1 berjumlah 12 kata. Sehingga didapatkan nilai TF untuk kata “Omnibus” yaitu $1/12 = 0.08$. berlaku untuk kata-kata yang lain. Setelah mencari nilai TF, selanjutnya mencari nilai IDF seperti pada Tabel 5.10.

Tabel 5.10 Perhitungan IDF

| Kata | Dokumen | | | | Jumlah | IDF = $\log\left(\frac{N}{df}\right)+1$ |
|---------|---------|---|-----|------|--------|--|
| | 1 | 2 | ... | 3064 | | |
| Omnibus | 1 | 1 | ... | 0 | 1044 | $\text{Log}\left(\frac{3064}{1044}\right) = 1.467$ |
| Buruh | 1 | 1 | ... | 1 | 684 | $\text{Log}\left(\frac{3064}{684}\right) = 1.651$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Demo | 0 | 0 | ... | 0 | 362 | $\text{Log}\left(\frac{3064}{177}\right) = 1.927$ |
| Jokowi | 0 | 1 | ... | 0 | 155 | $\text{Log}\left(\frac{3064}{155}\right) = 2.295$ |

Berdasarkan **Tabel 5.10**, nilai IDF didapatkan dari hasil log jumlah keseluruhan data dibagi jumlah kata keseluruhan dalam data set kemudian ditambah 1. Kata “Omnibus” pada seluruh dokumen adalah sebesar 0.467 dan seterusnya untuk kata yang lain. Setelah mendapatkan hasil perhitungan IDF dari masing-masing kata, selanjutnya adalah mencari nilai TF-IDF, dengan melakukan perkalian nilai TF dan nilai IDF. Perhitungan TF-IDF dapat dilihat pada **Tabel 5.11**.

Tabel 5.11 TF-IDF pada *Omnibus law*

| Kata | Dokumen (TF) | | | | | IDF | TF-IDF | | | | |
|---------|--------------|------|----|-----|------|-------|--------|-------|-----|-----|------|
| | 1 | 2 | 50 | ... | 3064 | | 1 | 2 | 100 | ... | 3064 |
| Omnibus | 0.08 | ... | 0 | ... | 0 | 1.467 | 0.117 | ... | 0 | ... | 0 |
| Buruh | 0.08 | ... | 0 | ... | 0 | 1.651 | 0.132 | ... | 0 | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Demo | 0 | 0.16 | 0 | ... | 0 | 1.927 | ... | 0.308 | 0 | ... | 0 |
| Jokowi | 0 | ... | 0 | ... | 0 | 2.295 | 0 | ... | 0 | ... | 0 |

TF-IDF menggambarkan bahwa semakin besar nilai TF-IDF dari sebuah kata maka semakin besar hubungan kata tersebut pada sebuah dokumen.

Selanjutnya peneliti akan menampilkan jumlah per kata dari keseluruhan data *tweet* yang peneliti gunakan. Jumlah kata dalam seluruh dokumen *tweet* berjumlah 33578 kata dan setiap kata mempunyai frekuensi jumlah yang berbeda-beda. Dapat dilihat selain kata yang peneliti gunakan dalam *keyword scraping* di *twitter*, terdapat kata-kata yang mempunyai jumlah dan frekuensi terbanyak yang terdapat dalam keseluruhan *tweet* pada Tabel 5.12

Tabel 5.12 Frekuensi Kata

| Word | Frekuensi | Word | Frekuensi |
|-----------|-----------|-----------|-----------|
| Cipta | 1247 | Jokowi | 156 |
| Kerja | 1175 | Day | 144 |
| Omnibus | 1044 | Dpr | 142 |
| Law | 832 | rezim | 134 |
| Buruh | 684 | May | 128 |
| UU | 296 | Kpk | 124 |
| tolak | 248 | Tuntutan | 122 |
| Indonesia | 190 | Investasi | 109 |
| upah | 186 | murah | 106 |

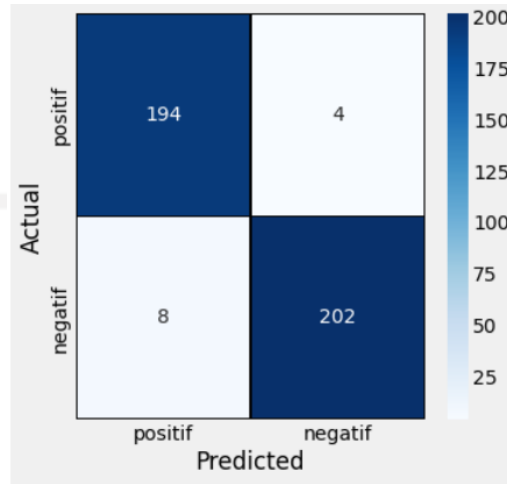
5.5. Klasifikasi SVM

Dalam melakukan klasifikasi, penenili terlebih dahulu menentukan data latih dan data uji. Data latih (*training* data) merupakan data yang dipakai dalam melatih suatu algoritma untuk menentukan model yang sesuai, Adapun data uji (*testing* data) digunakan dalam pengujian dari nilai akurasi model yang didapatkan pada tahapan *testing*. Pembagian dari data latih dan uji dengan total data yang peneliti pakai adalah 2038 data yang terdiri dari data positif dan data negatif. Dalam pembagian data ini peneliti menggunakan proporsi 80% untuk data latih (data *training*) dan 20% untuk data uji (data *testing*).

$$\begin{aligned} \text{Data latih} &= 80\% \times 2038 \\ &= 1630 \end{aligned}$$

$$\begin{aligned} \text{Data uji} &= 20\% \times 2038 \\ &= 408 \end{aligned}$$

Hasil dari data latih 1630 dan data uji sebesar 408, selanjutnya dilakukan proses pengujian nilai *Confusion Matrix* dengan menggunakan hasil dari data uji atau data *testing* untuk kernel linear dengan menggunakan nilai $C=1$.



Gambar 5.4 *Confusion Matrix*

Gambar 5.6 *confucion matrix* menjelaskan bahwa dari data aktual yang bernilai sentimen positif didapatkan klasifikasi yang sesuai dengan data uji tersebut sebesar 194. Dalam artian data yang diklasifikasikan oleh model atau mesin dengan tepat sebesar 194 data. Akan tetapi mesin mendeteksi bahwa terdapat 4 data sentien positif yang diklasifikasikan atau terprediksi ke dalam sentimen negatif. Adapun untuk klasifikasi sentimen negatif, data prediksi yang diklasifikasian sesuai dengan data aktual sebesar 202 data, akan tetapi mesin mendeteksi terdapat 8 data sentimen negatif yang terprediksi oleh mesin ke dalam sentimen positif.

Setelah menentukan nilai *confusion matrix* selanjutnya peneliti mencari nilai untuk *recall*, *precision*, *accuracy*, *specificity*, *False Positive Rate (FPR)*, dan *Area Under Curve (AUC)*.

$$Recall = \frac{194}{194+8} = 0,96 = 96\%$$

$$Precision = \frac{194}{194+4} = 0,98 = 98\%$$

$$Accuracy = \frac{194+202}{194+4+8+202} = 0,97 = 97\%$$

$$Spesificity = \frac{202}{202+4} = 0,98 = 98\%$$

$$FPR = 1 - 0,98 = 0,02$$

$$AUC = \frac{1+0,96-0,02}{2} = 0,97$$

Berdasarkan perhitungan diatas didapatkan bahwa *recall* 96% yang artinya rata-rata dari data klasifikasi sebanyak 96% mampu memprediksi sentimen dengan relevan. Nilai *precision*-nya adalah 98%, yang berarti rata-rata 98% data klasifikasi tergolong sentimen relevan. Klasifikasi tersebut benar, berdasarkan nilai *recall* dan *precision* yang tinggi. *Spesificity*, FPR, dan AUC juga menghasilkan nilai yang cukup tinggi. Nilai AUC sebesar 0,97 yang menunjukkan bahwa nilai tersebut sangat layak atau klasifikasinya sangat baik.

Selanjutnya metode yang digunakan dalam mengklasifikasian sentimen pada penelitian adalah dengan menggunakan algoritma *Support Vector Machine* (SVM). Klasifikasi metode SVM mempunyai beberapa *kernel* yakni *Linier*, *polynomial*, *Sigmoid* dan RBF. Semua kernel praktikan coba untuk mengetahui hasil akurasi terbaik pada metode SVM.

5.5.1 Kernel Linear

Kernel yang pertama adalah *linear*, beberapa akurasi kernel linear dengan menggunakan beberapa nilai C yang berbeda terdapat pada Gambar 5.7.

```
#SVM LINIER
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='linear', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test, svm.predict(x_test))))

Akurasi untuk C = 0.01: 0.8504901960784313
Akurasi untuk C = 0.05: 0.9240196078431373
Akurasi untuk C = 0.25: 0.9656862745098039
Akurasi untuk C = 0.5: 0.9681372549019608
Akurasi untuk C = 1: 0.9705882352941176
```

Gambar 5.5 Kernel Linear

Gambar 5.7 merupakan akurasi kernel *linear* untuk hasil C=0.01 sebesar 85.04%, 0.05 sebesar 92.40%, 0.25 sebesar 96.56%, 0.5 sebesar 96.81% dan untuk C=1 sebesar 97.05%.

5.5.2 Kernel Polynomial

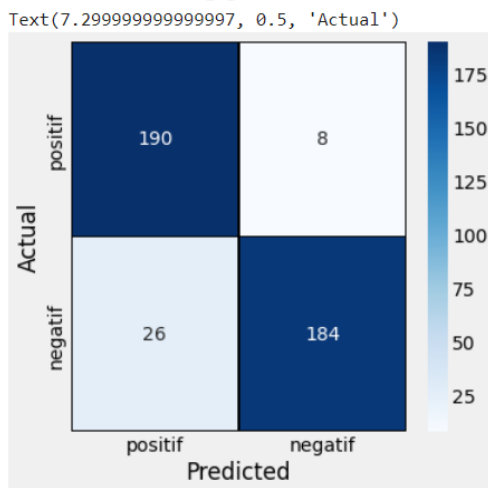
Kernel yang kedua adalah *polynomial*, beberapa akurasi kernel *polynomial* dengan menggunakan beberapa nilai C yang berbeda terdapat pada Gambar 5.8.

```
#SVM POLINOMIAL
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='poly', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test, svm.predict(x_test))))
```

↳ Akurasi untuk C = 0.01: 0.5147058823529411
 Akurasi untuk C = 0.05: 0.9117647058823529
 Akurasi untuk C = 0.25: 0.9019607843137255
 Akurasi untuk C = 0.5: 0.9117647058823529
 Akurasi untuk C = 1: 0.9166666666666666

Gambar 5.6 Kernel *Polynomial*

Gambar 5.8 merupakan akurasi kernel *polynomial* untuk hasil C=0.01 sebesar 51.47%, 0.05 sebesar 91.17%, 0.25 sebesar 90.19%, 0.5 sebesar 91.17% dan untuk C=1 sebesar 91.66%. Nilai prediksi *confusion matrix* kernel *polynomial* dengan C=1 dapat dilihat pada **Gambar 5.9**.



Gambar 5.7 *Confusion Matrix* Kernel *Polynomial*

Gambar 5.9. data positif yang diklasifikasikan oleh mesin atau model dengan tepat sebesar 190 data. Akan tetapi mesin mendeteksi bahwa terdapat 8 data sentien positif yang diklasifikasikan atau terprediksi ke dalam sentimen negatif. Adapun untuk klasifikasi sentimen negatif, data prediksi yang diklasifikasian sesuai dengan data aktual sebesar 184 data, akan tetapi mesin mendeteksi terdapat 26 data sentimen negatif yang terprediksi oleh mesin ke dalam sentimen positif. Setelah menentukan nilai *confusion matrix* selanjutnya peneliti mencari nilai untuk *recall*, *precision*, *accuracy*, *specificity*, *False Positive Rate (FPR)*, dan *Area Under Curve (AUC)* untuk kernel *polynomial*.

$$Recall = \frac{190}{190+26} = 0,88 = 88\%$$

$$Precision = \frac{190}{190+8} = 0,96 = 96\%$$

$$Accuracy = \frac{190+184}{190+8+26+184} = 0,92 = 92\%$$

$$Spesificity = \frac{184}{184+8} = 0,96 = 96\%$$

$$FPR = 1 - 0,96 = 0,04$$

$$AUC = \frac{1+0,88-0,04}{2} = 0,92$$

5.5.3 Kernel RBF

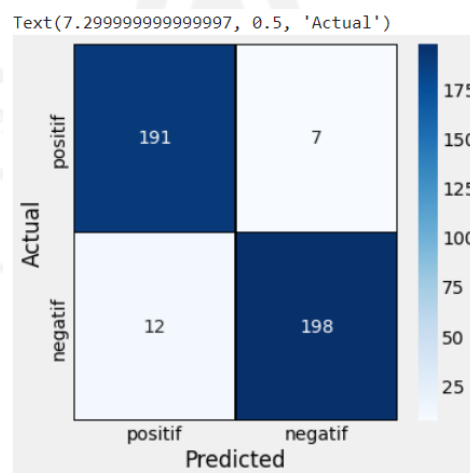
Kernel yang ketiga adalah RBF, beberapa akurasi kernel RBF dengan menggunakan beberapa nilai C yang berbeda terdapat pada Gambar 5.10.

```
#SVM RBF
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='rbf', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test, svm.predict(x_test))))

Akurasi untuk C = 0.01: 0.6029411764705882
Akurasi untuk C = 0.05: 0.8455882352941176
Akurasi untuk C = 0.25: 0.9044117647058824
Akurasi untuk C = 0.5: 0.9313725490196079
Akurasi untuk C = 1: 0.9534313725490197
```

Gambar 5.8 Kernel RBF

Gambar 5.10 merupakan akurasi kernel RBF dengan nilai gammanya 'scale' sebesar $\frac{1}{(n_{feature} * X.var())}$ dan didapatkan hasil C=0.01 sebesar 60.29%, 0.05 sebesar 84.55%, 0.25 sebesar 90.44%, 0.5 sebesar 93.13% dan untuk C=1 sebesar 95.34%. Nilai prediksi *confusion matrix* kernel RBF dengan C=1 dapat dilihat pada **Gambar 5.9**.



Gambar 5.9 Confusion Matrix Kernel RBF

Gambar 5.11. data positif yang diklasifikasikan oleh mesin atau model dengan tepat sebesar 191 data. Akan tetapi mesin mendeteksi bahwa terdapat 7 data sentimen positif yang diklasifikasikan atau terprediksi ke dalam sentimen negatif. Adapun untuk klasifikasi sentimen negatif, data prediksi yang diklasifikasikan sesuai dengan data aktual sebesar 198 data, akan tetapi mesin mendeteksi terdapat 12 data sentimen negatif yang terprediksi oleh mesin ke dalam sentimen positif. Setelah menentukan nilai *confusion matrix* selanjutnya peneliti mencari nilai untuk *recall*, *precision*, *accuracy*, *specificity*, *False Positive Rate* (FPR), dan *Area Under Curve* (AUC) untuk kernel *RBF*.

$$Recall = \frac{191}{191+12} = 0,94 = 94\%$$

$$Precision = \frac{191}{191+7} = 0,96 = 96\%$$

$$Accuracy = \frac{191+198}{191+7+12+198} = 0,95 = 95\%$$

$$Spesificity = \frac{198}{198+7} = 0,96 = 96\%$$

$$FPR = 1 - 0,96 = 0,04$$

$$AUC = \frac{1+0,94-0,04}{2} = 0,95$$

5.5.4 Kernel Sigmoid

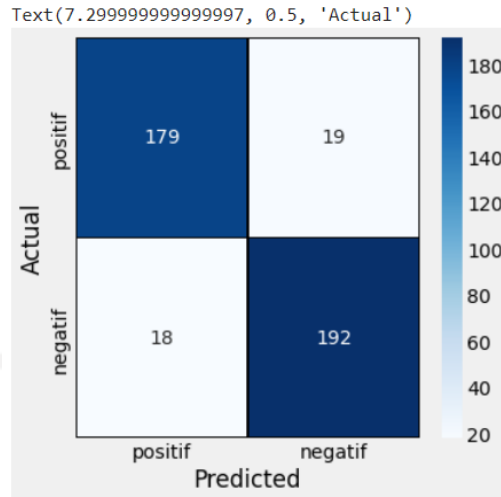
Kernel yang kedua adalah sigmoid, beberapa akurasi kernel sigmoid dengan menggunakan beberapa nilai C yang berbeda terdapat pada Gambar 5.12.

```
#SVM SIGMOID
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='sigmoid', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test, svm.predict(x_test))))

Akurasi untuk C = 0.01: 0.8333333333333334
Akurasi untuk C = 0.05: 0.8382352941176471
Akurasi untuk C = 0.25: 0.8578431372549019
Akurasi untuk C = 0.5: 0.9019607843137255
Akurasi untuk C = 1: 0.9093137254901961
```

Gambar 5.10 Kernel Sigmoid

Gambar 5.12 merupakan akurasi kernel *sigmoid* untuk hasil C=0.01 sebesar 83.33%, 0.05 sebesar 83.82%, 0.25 sebesar 85.78%, 0.5 sebesar 90.01% dan untuk C=1 sebesar 90.93%. Nilai prediksi *confusion matrix* kernel *Sigmoid* dengan C=1 dapat dilihat pada Gambar 5.13.



Gambar 5.11 *Confusion Matrix* Kernel Sigmoid

Gambar 5.13. data positif yang diklasifikasikan oleh mesin atau model dengan tepat sebesar 179 data. Akan tetapi mesin mendeteksi bahwa terdapat 19 data sentien positif yang diklasifikasikan atau terprediksi ke dalam sentimen negatif. Adapun untuk klasifikasi sentimen negatif, data prediksi yang diklasifikasikan sesuai dengan data aktual sebesar 192 data, akan tetapi mesin mendeteksi terdapat 18 data sentimen negatif yang terprediksi oleh mesin ke dalam sentimen positif. Setelah menentukan nilai *confusion matrix* selanjutnya peneliti mencari nilai untuk *recall*, *precision*, *accuracy*, *specificity*, *False Positive Rate* (FPR), dan *Area Under Curve* (AUC) untuk kernel *sigmoid*.

$$Recall = \frac{179}{179+18} = 0,91 = 91\%$$

$$Precision = \frac{179}{179+19} = 0,90 = 90\%$$

$$Accuracy = \frac{179+192}{179+192+18+19} = 0,91 = 91\%$$

$$Spesificity = \frac{192}{192+19} = 0,91 = 91\%$$

$$FPR = 1 - 0,91 = 0,09$$

$$AUC = \frac{1+0,91-0,09}{2} = 0,91$$

Perbandingan dari kinerja keempat kernel SVM yang praktikan gunakan untuk keseluruhan data, akan praktikan pilih dari masing-masing kernel dengan menggunakan besar $C=1$ karena nilai tersebut merupakan nilai tertinggi dibandingkan dengan menggunakan nilai C yang lain. Hasil perbandingan klasifikasi dari keempat kernel ditampilkan pada tabel 5.13.

Tabel 5.13 Nilai Akurasi Kernel SVM

| Kernel | Nilai C | Akurasi | AUC |
|-------------------|----------------|----------------|------------|
| <i>Linear</i> | 1 | 97.05% | 0.97 |
| <i>Polynomial</i> | 1 | 91.66% | 0.92 |
| <i>RBF</i> | 1 | 95.34% | 0.95 |
| <i>Sigmoid</i> | 1 | 90.93% | 0.91 |

Berdasarkan hasil dari tabel 5.13 didapatkan metode kernel *linear* dengan menggunakan algoritma *Support Vector Machine* mendapatkan hasil nilai akurasi terbaik jika dibandingkan dengan kernel-kernel lain. Kernel *linear* dengan $C=1$ mendapatkan nilai akurasi sebesar 97.05% dan nilai AUC sebesar 0.97.

5.6. Asosiasi Kata

Melihat hubungan antara keterkaitan antar kata satu dengan kata yang lainnya dapat dilakukan dengan cara asosiasi kata. Menurut (Sarwono, 2006) hasil dari asosiasi menunjukkan kekuatan antar dua kata yang memiliki hubungan dengan rentang nilai 0 s/d 1. Beberapa nilai dalam hubungan asosiasi adalah sebagai berikut:

- 0 : Tidak ada hubungan antara dua variabel
- >0 - 0,25 : Hubungan lemah
- >0,25 - 0,5 : Hubungan cukup
- >0,5 – 0,75 : Hubungan kuat
- 1 : Hubungan sangat kuat

Peneliti menggunakan beberapa kata dan melihat asosiasinya seperti kata “pengusaha” dan “buruh” pada tabel 5.14.

Tabel 5.14 Asosiasi Kata

| Data | Keyword | Asosiasi Kata | Nilai |
|----------------|------------------|----------------------|--------------|
| Positif | Pengusaha | Dukung | 0.70 |
| | | Muda | 0.65 |
| | | Dinilai | 0.60 |
| Negatif | Buruh | Murah | 0.43 |
| | | Kirim | 0.43 |
| | | Petisi | 0.43 |

6. BAB VI PENUTUP

6.1. Kesimpulan

Berdasarkan hasil dari uraian pembahasan dan rumusan masalah yang telah disebutkan sebelumnya, peneliti dapatkan kesimpulan untuk menjawab beberapa rumusan masalah tersebut, yakni:

1. Hasil *sentiment analysis* diperoleh klasifikasi sentimen positif terhadap *Omnibus law* sebanyak 1067 *tweet*, kemudian klasifikasi sentimen negatif sebanyak 971 *tweet*, dan klasifikasi netral sebanyak 1026 *tweet*. Dari hasil sentimen positif dan negatif menunjukkan bahwa sentimen positif masyarakat didominasi oleh pengusaha yang mendapatkan kemudahan dengan adanya *omnibus law*, sentimen negatif masyarakat didominasi buruh yang menganggap gaji buruh masih terbilang murah walaupun sudah adanya undang-undang *omnibus law*.
2. Berdasarkan perhitungan metode *Support Vector Machine* (SVM) didapatkan nilai akurasi terbaik sebesar 97.05% yakni kernel *linear* dengan $C=1$ dan AUC sebesar 0.97, yang berarti tingkat akurasi kernel *linear* SVM sudah sangat baik.

6.2. Saran

1. Diharapkan pada penelitian selanjutnya mampu memperbanyak waktu dan kuantitas data *tweet* yang digunakan.
2. Peneliti selanjutnya dapat membandingkan dengan metode turunan dari *Support Vector Machine* seperti *Reduced Support Vector Machine* (RSVM).
3. Bagi pemerintah, diharapkan hasil dari penelitian ini terutama ulasan/cuitan positif dan negatif terhadap *Omnibus Law* dapat dijadikan sebagai bahan pertimbangan dan bahan evaluasi untuk perundang-undangan yang lebih baik lagi di masa yang akan datang.



DAFTAR PUSTAKA

- Aini, N. (2021). *Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Smote Dan Support Vector Machine (SVM)*. Mataram: Program Studi Ilmu Komputer Fakultas Teknik Dan Desain Universitas Bumigora.
- Aliffia , K., Intan, P., & Budi, A. D. (2021). Penerapan SVM Dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi. *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, 101-109.
- Alwasi'a. (2020). *Analisis Sentimen Pada Review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detik.com pada Google Play 2019) [Skripsi]*. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Amani, N. K. (2021, Desember 3). *100 Ribu Buruh Bakal Demo Besar-besaran 7 Desember 2021*. Diambil kembali dari Liputan6.com: <https://www.liputan6.com/bisnis/read/4727440/100-ribu-buruh-bakal-demo-besar-besaran-7-desember-2021>
- Amin, I. (2022, Februari 10). *Pakar Hukum Lingkungan UGM Sebut Konflik Wadas Dampak Omnibus Law*. Diambil kembali dari Tirto.id: <https://tirto.id/pakar-hukum-lingkungan-ugm-sebut-konflik-wadas-dampak-omnibus-law-goNz>
- Andini, P. D. (2014). Penerapan Analisis Pohon Regresi pada Data Perlindungan Sosial.
- Anto, S. N., Arief, B. W., & Handoko, D. (2003). Support Vector Machine (Teori dan Aplikasinya dalam Bioinformatika). *Kuliah Umum IlmuKomputer.Com*.
- Anwar, M. C. (2022, Mei 03). *Sederet Tuntutan Buruh di Aksi May Day yang Digelar 12 Mei 2022*. Diambil kembali dari Kompas.com: <https://money.kompas.com/read/2022/05/03/054100326/sederet-tuntutan-buruh-di-aksi-may-day-yang-digelar-12-mei-2022?page=all>
- Ariyani, L. (2016). Kajian Penerapan Model C45, Support Vector Machine (SVM), dan Neural Network Dalam Prediksi Kenaikan Kelas. *Faktor Exacta*, 72–86.

- Budianto, M. A. (2018). Perbandingan K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) Dalam Pengenalan Karakteristik Plat Kendaraan Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik Kejuruan*, 1-9.
- Busroh, F. F. (2017, Agustus). Konseptualitas Omnibus Law dalam Menyelesaikan Permasalahan Regulasi Pertanahan. *Arena Hukum*, 10, 241.
- CNN. (2022, Februari 07). *Buruh Bergerak Demo di DPR, Serukan Kembali Penolakan UU Cipta Kerja*. Diambil kembali dari [cnnindonesia.com: https://www.cnnindonesia.com/nasional/20220207104337-20-755901/buruh-bergerak-demo-di-dpr-serukan-kembali-penolakan-uu-cipta-kerja](https://www.cnnindonesia.com/nasional/20220207104337-20-755901/buruh-bergerak-demo-di-dpr-serukan-kembali-penolakan-uu-cipta-kerja)
- Dedi , D., Eka , S. P., & Ferico , O. P. (2020, November). Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic*, 7, 1-11.
- Dehhaf. (2012, Februari 10). *Sentiment Analysis for Business, Finance, and Social Media Showcased at May 8, New York Symposium*. Diambil kembali dari Customerthink: https://customerthink.com/sentiment_analysis_for_business_finance_and_social_media_showcased_at_may_8_new_york_symposium/
- Dharmadhikari, I. M. (2011). Empirical Studies on Machine Learning Based Text Classification Algorithms. *Advanced Computing: An International Journal (ACIJ)*, 161-169.
- D'monte, L. (2009). *Swine Flu's tweet Tweet Causes Online Flutter*. Business-standard.com.
- Duda & Hart. (1973). *PATTERN CLASSIFICATION AND SCENE ANALYSIS*. New York: John Wiley & Sons.
- Feldman, R., & Sangar, J. (2006). *The Text Mining Handbook*.
- Fiska, R. R. (2017, Juni 1). Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera). *SATIN – Sains dan Teknologi Informasi*, 3.

- Fuady, M. (2013). Teori-Teori Besar (Grand Theory) Dalam Hukum. *Kencana*, 249.
- Gamer, B. A. (2004). *Black's Law Dictionary*. Minnesota: West Publishing Co.
- Haddi, E., Liu, X., & Shi, Y. (2013). *The Role of Text Pre-processing in Sentiment Analysis*. *Procedia Computer Science* 26-32.
- Hakim, S. N. (2021). *Analisis Sentimen Persepsi Pengguna Myindihome Menggunakan Metode Support Vector Machine (SVM) Dan Naïve Bayes Classifier (NBC)*. Yogyakarta: Program Studi Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. Waltham: Morgan Kaufmann.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques Third Edition*. . Waltham USA: Morgan Kaufmann.
- Handayani et al. (2020, November). Klasifikasi Kualitas Udara Dengan Metode Support Vector. 3.
- Harijianto, S. D. (2019). *Analisis Sentimen pada twitter menggunakan Multinomial Naive Bayes*. Yogyakarta: Fakultas Sains dan teknologi universitas Sanata Dharma.
- Kamal, W. W. (2021). *Analisis Sentimen Ulasan Produk Skincare Menggunakan Metode Support Vector Machine (Studi Kasus: Forum Female Daily)*. Yogyakarta: Program Studi Informatika Program Sarjana Universitas Islam Indonesia.
- kemendikbud. (2019, November 28). *Booklet-UU-Cipta-Kerja*. Diambil kembali dari <https://dikti.kemdikbud.go.id/wp-content/uploads/2020/10/Booklet-UU-Cipta-Kerja.pdf>
- kemendikbud, D. (2020, Oktober). *Omnibus Law*. Diambil kembali dari dikti.kemdikbud: <https://dikti.kemdikbud.go.id/wp-content/uploads/2020/10/Booklet-UU-Cipta-Kerja.pdf>
- Kemenkumham, B. H. (2020, Februari 18). *RUU Omnibus Law Cipta Lapangan Kerja untuk Tingkatkan Pertumbuhan Ekonomi*. Diambil kembali dari Kemenkumham: <https://www.kemenkumham.go.id/berita/ruu-omnibus-law-cipta-lapangan-kerja-untuk-tingkatkan-pertumbuhan-ekonomi>

- Kurniawan, F. (2020). Problematika Pembentukan Ruu Cipta Kerja Dengan Konsep Omnibus Law Pada Klaster Ketenagakerjaan Pasal 89 Angka 45 Tentang Pemberian Pesangon Kepada Pekerja Yang Di Phk. *Jurnal Panorama Hukum*.
- Manning, C., Raghavan, P., & Schutze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University.
- Maria, R. (2014). Pembangunan Aplikasi Pencairan Dokumen Menggunakan *Text Mining* Berbasis Web. *Jurnal Universitas Atma Jaya Yogyakarta*.
- Medhat. (2014). Sentiment analysis algorithms and applications A survey. *Ain Shams Engineering Journal*, 10.1016/j.asej.2014.04.011.
- Mukaromah, V. F. (2020, Oktober 07). *Ini Pro Kontra yang Muncul Setelah Omnibus Law UU Cipta Kerja Disahkan*. Diambil kembali dari Kompas.com:
<https://www.kompas.com/tren/read/2020/10/07/143101665/ini-pro-kontra-yang-muncul-setelah-omnibus-law-uu-cipta-kerja-disahkan?page=all>
- Nugraheni, & Yuni. (2017, Juni). Social Media Habit Remaja Surabaya. *Jurnal Komunikatif*, 1.
- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2013). Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika. *IlmuKomputer.Com*.
- Passi, P. d. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. *IoT*, 218-239.
- Prabowo, A. S. (2020, April). Politik Hukum Omnibus Law. *Jurnal Pamator*, Volume 13 No 1. Diambil kembali dari
<https://journal.trunojoyo.ac.id/pamator/article/view/6991>
- Pradana, M. G. (2020, Maret 25). Penggunaan Fitur Wordcloud dan Document Term Matrix dalam *Text Mining*. *Jurnal Ilmiah Informatika (JIF)*, 2615-1049.
- Pradhan. (2016). *Journal of Pharmacognosy and Phytochemistry*. Golden Heart of the Nature : Piper betle L. 6 Vol. 1 pp. 147-167.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. 1704–1708.

- Qeis, M. I. (2015). Aplikasi Wordcloud Sebagai Alat Bantu Analisis Wacana. *Jakarta : Program Studi Desain Komunikasi Visual Fakultas Bahasa dan Seni Universitas Indraprasta PGRI.*
- Raharjo, S. (1981). *Hukum Masyarakat & Pembangunan.* Bandung : Alumni.
- Rahman, Darmawidjadja, & Alams. (2017, Januari). Klasifikasi untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization neural Network (RBNN). *Jurnal Informatika, 11 No 1.*
- Rezki, A. S. (2021). *Klasifikasi Emosi Pada Twitter Dengan Metode K-Nearest Neighbor (KNN).* Pekanbaru: Fakultas Sains Dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Sahputra, I. (2021). *Klasifikasi Sentimen Terhadap Komentar Twitter Menggunakan Algoritma Support Vector Machine Dan Naïve Bayes.* Pekanbaru: Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
- Santo, P. A. (2019, Oktober). *Memahami Gagasan Omnibus Law.* Diambil kembali dari [business-law.binus.ac.id:](https://business-law.binus.ac.id/) <https://business-law.binus.ac.id/2019/10/03/memahami-gagasan-omnibus-law/>
- Santoso, G. T. (2021). *Analisis Sentimen pada Tweet dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM).* Pekanbaru: Fakultas Teknik Universitas Islam Riau .
- Sarwono, J. (2006). *Metode Penelitian Kuantitatif dan Kualitatif.* Yogyakarta: Graha Ilmu.
- Sasongko, T. B. (2016, Agustus 2). Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA). *Jurnal Teknik Informatika dan Sistem Informasi, 2,* 2443-2229.
- Tamburian, D. (2015, Juli). Interpretasi Tagar #Savehajilulung Di Kalangan Netizen Pengguna. *Jurnal Komunikasi, 7,* 81-97.
- Ubaidillah, M. &. (2019). Klasifikasi Gelombang Otot Lengan pada Robot Manipulator Menggunakan Support Vector Machine. *REKAYASA, 91-97.*
- Walaa, M., Ahmed , H., & Hoda, K. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 1093-1113.*

Zalyhaty, L. Q. (2021). *Analisis Sentimen Tanggapan Masyarakat terhadap Vaksin Covid-19 Menggunakan Algoritma SVM*. Surabaya: Universitas Dinamika.



LAMPIRAN

Lampiran 1 Scraping Data Twitter

```
pip install tweepy

import re
import pandas as pd
import numpy as np
import seaborn as sns
import csv
import string
import nltk
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')

import tweepy
from tweepy import OAuthHandler
import pandas as pd

api_key = "I0E7xGaPsVIIyg0sq8tfw"
api_key_secret = "frk7rjhTVHEuRErf4V2h93xZ6eSAr2myy9gH4RaU"
access_token = "228245421-
FPyC4SFufgkHDmHy7gTpEqm36mbFepYmkQ2p54xf"
access_token_secret = "cYc2xsbeKiTAlb8RuQ9btHZxERGNofrlazKBPTn
Q"

auth = tweepy.OAuthHandler(api_key, api_key_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

api.update_status

hasilSearch = api.search(q="omnibus law", lang="id", count=100
, tweet_mode="extended", lat="indonesia" )
hasilSearch

print("show the 100 recent tweets: \n")
i =1
for tweet in hasilSearch [0:200]:
    print(str(i)+' ) '+tweet.full_text + '\n')
    i = i +1
df=pd.DataFrame([tweet.full_text for tweet in hasilSearch], co
lumnns=['Tweets'])
df
df.to_csv('tweets.csv')
```

Lampiran 2 Preprocessing Data

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('/content/Data Tweet Skirpsi fix.csv', sep=';
', encoding="ISO-8859-1")
df.tail()
# melihat jumlah baris dan kolom
df.shape
#2. Proses Filtering/Preprocessing
import numpy as np
import pandas as pd
import re
import string
import nltk

from nltk.sentiment.vader import SentimentIntensityAnalyzer
#-----Case Folding-----#
df['text'] = df['Tweet'].str.lower()
print('Case Folding Result : \n')
print(df['text'])
print('\n\n\n')

#-----Cleaning Data-----#
from tqdm import tqdm
preprocessed_tweets = []
# tqdm is for printing the status bar
for sentence in tqdm(df['Tweet']):
    sentence = re.sub(r'https?:\/\/\S+|www\.\S+', r'', str (sent
ance)) # remove URLs
    sentence = re.sub(r'<.*?>', r'', str (sentence)) # remove
HTML
    sentence = re.sub(r'\d+', '', str (sentence)).strip() # re
move number
    sentence = re.sub(r"^\w\s\d]", "", str (sentence)) # remov
e pnctuations
    sentence = re.sub(r'@\w+', '', str (sentence)) # remove men
tions
    sentence = re.sub(r'#\w+', '', str (sentence)) # remove has
h
    sentence = re.sub(r"\s+", " ", str (sentence)).strip() # re
move space
    sentence = re.sub("\S*\d\S*\Ã", "", str (sentence)).strip(
)
```

```

    sentence = re.sub('[^A-Za-
z]+', ' ', str (sentence)) # let only alphabet

    preprocessed_tweets.append(sentence.strip())
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
df.Tweet
#-----Tokenizing-----#
import string
import re

from nltk.tokenize import word_tokenize
x = df.iloc[0]
print(nltk.word_tokenize(x['Tweet']))
def identify_tokens(row):
    message = row['Tweet']
    tokens = nltk.word_tokenize(message)
    token_words = [w for w in tokens if w.isalpha()]
    return token_words

df['Tweet'] = df.apply(identify_tokens, axis = 1)
df.Tweet
#-----stemming-----#
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
stemming = PorterStemmer()
def stem_list(row):
    text = row['Tweet']
    stem = [stemming.stem(word) for word in text]
    return (stem)

df['Tweet'] = df.apply(stem_list, axis=1)
df.Tweet
#-----Filtering-----#
#stopword
from nltk.corpus import stopwords

stops = set(stopwords.words('indonesian'))
def remove_stops(row):
    message = row['Tweet']
    stopwords = [w for w in message if not w in stops]
    return stopwords

df['Tweet'] = df.apply(remove_stops, axis = 1)
df['Tweet'] = df['Tweet'].str.join(" ")

```

```
df['Tweet']  
#-----Save Data-----#  
df.to_csv('Data Hasil Preprocessing.csv')
```



Lampiran 3 Pelabelan Sentimen (RStudio)

```
dataset <- read.csv(file.choose(), header = TRUE, sep = ";")
View(dataset)
# define "tolower error handling" function
try.error = function(x)
{
  # create missing value
  y = NA
  # tryCatch error
  try_error = tryCatch(tolower(x), error=function(e) e)
  # if not an error
  if (!inherits(try_error, "error"))
    y = tolower(x)
  # result
  return(y)
}
# lower case using try.error with sapply
deskripsi=dataset$text
deskripsi = sapply(deskripsi, try.error)
# remove NAs in deskripsi
deskripsi = deskripsi[!is.na(deskripsi)]
names(deskripsi) = NULL
length(deskripsi)
getwd()
opinion.lexicon.pos = scan("positif.txt", what = "character", comment.char = ";")
opinion.lexicon.neg = scan("negatif.txt", what = "character", comment.char = ";")
head(opinion.lexicon.pos)
head(opinion.lexicon.neg)
#membuat fungsi score.sentiment(), yang bisa menghitung hasil sentimen mentah
berdasarkan algoritma pencocokan sederhana:
getSentimentScore = function(deskripsi, opinion.lexicon.pos, opinion.lexicon.neg,
.progress = "none")
```

```

{
  require(plyr)
  require(stringr)
  scores = laply(deskripsi, function(sentence, opinion.lexicon.pos,
opinion.lexicon.neg) {
    #remove digit, punctuation, dan special/control character:
    sentence = gsub("[[:cntrl:]]", "", gsub("[[:punct:]]", "", gsub("\\d+", "",
sentence)))
    #convert semua teks menjadi lowercase:
    sentence = tolower(sentence)
    #pisahkan setiap kalimat menggunakan spasi (space delimiter):
    words = unlist(str_split(sentence, "\\s+"))
    #lakukan boolean match dari setiap kata-kata menggunakan pos
&amp;&amp;&amp; neg opinion-lexicon:
    pos.matches = !is.na(match(words, opinion.lexicon.pos))
    neg.matches = !is.na(match(words, opinion.lexicon.neg))
    #score sentimen = total positive sentiment - total negative:
    score = sum(pos.matches) - sum(neg.matches)
    return(score)
  }, opinion.lexicon.pos, opinion.lexicon.neg, .progress=.progress)

  #return data frame berisi kalimat beserta sentimennya:
  return(data.frame(text = deskripsi, score = scores))
}
#terapkan ke data tweet yang telah kita bersihkan:
dataResult = getSentimentScore(deskripsi, opinion.lexicon.pos,
opinion.lexicon.neg)
#export to csv:
write.csv(dataResult, file = "dataResult2.csv")
hasil <- read.csv(file.choose(), header = TRUE, sep = ";")
hasil$klasifikasi<- ifelse(hasil$score<0, "Negatif", ifelse(hasil$score>0, "Positif",
"netral"))

```

```
hasil$klasifikasi
#Tukar Row
data <- hasil[c(1,2,3)]
write.csv(data, file = "data ter labeli2.csv")
#Memisahkan twit
data.pos <- hasil[hasil$score>0,]
write.csv(data.pos, file = "data-pos.csv")
data.neg <- hasil[hasil$score<0,]
write.csv(data.neg, file = "data-neg.csv")
data.net <- hasil[hasil$score==0,]
data=rbind(data.neg,data.pos,data.net)
write.csv(data, file = "data.csv")
table(data$klasifikasi)
```



Lampiran 4 Klasifikasi SVM

```
#Memanggil Data
df_data = pd.read_csv('/content/Data Pelabelan Fix.csv', sep='
;', encoding="ISO-8859-1")
df_data
import re
import pandas as pd
import numpy as np
import seaborn as sns
from textblob import TextBlob
import csv
import string
import nltk
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
import seaborn as sns
from PIL import Image
from wordcloud import WordCloud, ImageColorGenerator

#-----TF-IDF-----#
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer()
tfidf = vectorizer.fit_transform(df_data['text'])
vectorizer.get_feature_names_out()
tfidf.toarray()
tfidf

from sklearn.model_selection import train_test_split
x = df_data['text']
y = df_data['klasifikasi']
x_train, x_test, y_train, y_test, = train_test_split(x, y, tes
t_size = 0.2, random_state=0)
x_train
x_test
x_train.shape, x_test.shape

# vectorisasi
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer()
vectorizer.fit(x_train)
x_train = vectorizer.transform(x_train)
x_test = vectorizer.transform(x_test)

print(x_train.shape)
print(x_test.shape)
```



```

x_train.toarray()

#SVM LINIER
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='linear', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test,
svm.predict(x_test))))

#SVM POLINOMIAL
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='poly', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test,
svm.predict(x_test))))

#SVM RBF
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='rbf', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test,
svm.predict(x_test))))

#SVM SIGMOID
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
for c in [0.01,0.05,0.25,0.5,1]:
    svm = SVC(kernel='sigmoid', C=c)
    svm.fit(x_train, y_train)
    print('Akurasi untuk C = %s: %s' %(c, accuracy_score(y_test,
svm.predict(x_test))))

print('Final Model SVM Akurasi: %s' %accuracy_score(y_test, sv
m.predict(x_test)))
# evaluasi model svm
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
y_predd = svm.predict(x_test)

```

```

print('Akurasi of SVM classifier on test set: {:.4f}'.format(s
vm.score(x_test, y_test)))
print(classification_report(y_test, y_predd))

import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
cm=confusion_matrix(y_test, y_predd)
plt.figure(figsize = (5,5))
sns.heatmap(cm,cmap= "Blues",
            linecolor = 'black',
            linewidth = 1,
            annot = True,
            fmt='',
            xticklabels = ['positif','negatif'],
            yticklabels = ['positif','negatif'])
plt.xlabel("Predicted")
plt.ylabel("Actual")

wr_data = pd.read_csv('/content/data untuk wordcloud.csv', sep
=';', encoding="ISO-8859-1")
wr_data
#menampilkan wordcloud
from wordcloud import WordCloud
import matplotlib.pyplot as plt
#Positif
train_kdr = wr_data[wr_data['klasifikasi'] == 'Positif']
all_text_pst = ' '.join(word for word in train_kdr['text'])
wordcloud = WordCloud(colormap = 'Blues', width = 1000, mode =
'RGBA', background_color = 'white').generate(all_text_pst)
plt.figure(figsize = (20,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.margins(x=0, y=0)
plt.show
#Negatif
train_kdr = wr_data[wr_data['klasifikasi'] == 'Negatif']
all_text_pst = ' '.join(word for word in train_kdr['text'])
wordcloud = WordCloud(colormap = 'Reds', width = 1000, mode =
'RGBA', background_color = 'white').generate(all_text_pst)
plt.figure(figsize = (20,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.margins(x=0, y=0)
plt.show

```

Lampiran 5 Asosiasi Kata (RStudio)

```
library(tm)
library(SnowballC)
library(wordcloud)
library(RColorBrewer)
library(stringr)

dokumen <- readLines("C://Users/LENOVO/Documents/data/data frekuensi kata
fix.csv")
dokumen
dokumen <- VCorpus(VectorSource(dokumen))
dokumen
str(dokumen)
dokumenDTM<-DocumentTermMatrix(dokumen,control=list(tolower=TRUE,
removeNumbers=TRUE,
stopwords=TRUE,
removePunctuation=TRUE,
stemming=TRUE))

dokumenDTM
inspect(dokumenDTM)
dokumenDTM$dimnames$Terms
dokumenfreq<-findFreqTerms(dokumenDTM,3)
dokumenfreq
dokumenfreq<-findFreqTerms(dokumenDTM,3)
dokumenfreq
dokkudtm <- TermDocumentMatrix(dokumen)
dokkudtm
em <- as.matrix(dokkudtm)
em
ve <- sort(rowSums(em),decreasing=TRUE)
ve
de <- data.frame(word = names(ve),freq=ve)
head(de, 15)

wordcloud(words = de$word, freq = de$freq, min.freq = 1,
max.words=50, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))

vee<-as.list(findAssocs(dokkudtm, terms =c("buruh", "demo", "jokowi"), corlimit
= c(0.15,0.15,0.15,0.15,0.15,0.15)))
vee
```