BAB II
LANDASAN TEORI

2.1. Penelitian Terkait

Penelitian ini tidak terlepas dari penelitian sebelumnya yang berfungsi sebagai pembanding terhadap penelitian yang akan dilaksanakan dan juga sebagai sumber inspirasi yang dapat membantu pelaksanaan penelitian. Berdasarkan hal tersebut, penelitian terkait yang peneliti gunakan sebagaimana tabel 2.1. berikut ini.

Tabel 2. 1 Penelitian Yang Relevan

Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
T Mustaqim , K Umam and M A Muslim [9]	*Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm*	2019	Hasil pengujian yang dibangun menggunakan *tools* rapidminer menunjukkan tingkat akurasi algoritma KNN sebesar 79,45%, tertinggi jika dibandingkan dengan algoritma *classifier* lainnya seperti *decision trees, naïve Bayes* dan *random forest.* Proses analisis sentimen hampir bisa berjalan secara otomatis tanpa sentuhan manusia karena sudah ada pelabelan otomatis menggunakan *Vader.* Pengujian analisis sentimen terkait respon	Peneliti merinci proses penggunaan model untuk menjelaskan kemampuan dari model KNN dalam proses kerja gabungan antara deteksi polaritas leksikon VADER dan *K-Nearest Neighbors* dapat bekerja dengan baik pada analisis data mentah Twitter	Peneliti tidak menunjukkan hasil perbandingan dengan metode klasifikasi lain seperti *decision trees* dan *naïve Bayes* tetapi dapat menyatakan bahwa hasil kedua model klasifikasi ini hasilnya lebih rendah dari KNN.

Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
			pemerintah terhadap kebakaran hutan dapat dianalisis menggunakan algoritma KNN dan deteksi polaritas leksikon Vader dapat dilakukan dengan baik.		
Harun Sujadi, Sandi Fajar, Cecep Roni [10]	Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan *Metode Naive Bayes Classifier* dan *Support Vector Machine*	2022	Setelah dilakukan perbandingan, diperoleh hasil bahwa metode *Support Vector Machine* terbukti memiliki nilai akurasi yang lebih tinggi daripada metode *Naïve Bayes Classifier*	Proses pengklasifikasian sentimen covid 19 dengan metode *Naïve Bayes Classifier dan Support Vector Machine* dapat menghasilkan algoritma yang lebih baik yaitu *Naïve Bayes Classifier*	Hasil tidak dilakukan optimasi misalnya menggunakan *PSO*
Aprilia Wandani, Fauziah, Andrianingsih [2]	Sentimen Analisis Pengguna Twitter pada *Event Flash Sale* menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes	2021	Hasil penelitian menunjukkan bahwa algoritma K-NN dan *Random Forest* memiliki kinerja yang kurang baik dalam data sampel kecil. Hal ini berbeda dengan algoritma *Naive Bayes* yang memiliki akurasi lebih stabil pada data sampel besar atau kecil.	Peneliti melakukan perbandingan model dalam melakukan penelitian dengan Algoritma K-NN, *Random Forest,* dan *Naive Bayes*	Peneliti tidak menggunakan Algoritma PSO sebagai salah satu algoritma optimasi yang dapat digunakan untuk pengambilan keputusan.
Erina Undamayanti,	Analisis Sentimen	2022	Diperoleh hasil analisis berupa	Penelitian masih update dengan	Penelitian ini menggunakan

Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
Teguh Iman Hermanto, Ismi Kania-wulan [11]	Menggunakan Metode *Naive Bayes* Berbasis *Particle Swarm* Terhadap Pelaksanaan Program Merdeka Belajar Kampus Merdeka		sentimen positif sebesar 61.92% yang berarti bahwa program MBKM ini dapat diterima dengan baik oleh mahasiswa sebagai masyarakat pengguna Twitter meskipun terdapat 38.08% sentimen negatif yang muncul.	perkembangan saat ini. Pengembang kebijakan MBKM dapat menjadikan penelitian ini sebagai rujukan pengembangan MBKM terutama oleh tim POKJA Kemendikbud dengan memperhatikan hasil penelitian bahwa MBKM ini bisa memberikan manfaat dan pengalaman yang tidak didapatkan mahasiswa di lingkungan kampus. Program ini juga memberikan pengaruh terhadap akreditasi kampus.	metode klasifikasi *Naive Bayes* berbasis PSO, tidak menggunakan algoritma lain sebagai pembanding hasil penelitian
Angelina Puput Giovani, Ardiansyah, Tuti Haryanti, Laela Kurniawati, Windu Gata [8]	Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi	2020	Penelitian ini membandingkan metode NB, SVM, K-NN tanpa menggunakan *feature selection* dengan metode NB, SVM, K-NN yang menggunakan *feature selection*. Penelitian ini juga	Penelitian menggunakan berbagai algoritma klasifikasi sehingga dapat diperoleh perbandingan dari masing-masing algoritma yang dapat menghasilkan akurasi paling	Penelitian hanya melihat perbandingan hasil pengujian dari masing-masing algoritma, tidak menguraikan kategori hasil analisis sentimen pengguna

Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
			membandingkan nilai *Area Under Curve* (AUC) dari metode-metode tersebut untuk mengetahui algoritma yang paling optimal. Hasilnya adalah algoritma PSO berbasis SVM merupakan aplikasi yang terbaik dalam model yang diteliti dengan nilai akurasi sebesar 78,55% dan AUC sebesar 0,853.	besar.	twitter, apakah positif, negatif, atau netral.
Sudianto, Puspa Wahyuningtias, Hapsari Warih Utami, Uli Ahda Raihan, Hasna Nur Hanifah, Yehezkiel Nicholas Adanson [12]	*Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Vennya Escape From Quarantine)*	2022	Metode yang digunakan yaitu *Random forest* dan SVM. Terhadap analisis sentimen Twitter terhadap kaburnya Selebgram Rachel Vennya, algoritma *Random Forest* memperoleh hasil yang lebih baik dibandingkan dengan algoritma SVM.	Data yang besar sesuai jika menggunakan *Random forest* sehingga hasilnya juga bisa membedakan algoritma yang lebih unggul antara *Random forest* dengan SVM. Selain itu langkah-langkah yang telah dilalui dengan baik sehingga menghasilkan hasil penelitian yang akurat.	Peneliti tidak memberikan rekomendasi terhadap penelitian selanjutnya
Rikip Ginanjar, Rosalina, Aldo	Aplikasi Pemantauan Media Sosial untuk Analisa	2020	Penelitian ini mengimplementasikan metode *Lexicon-Based*	Penelitian ini menghasilkan aplikasi android yang dapat	Aplikasi masih perlu dikembangkan lebih lanjut

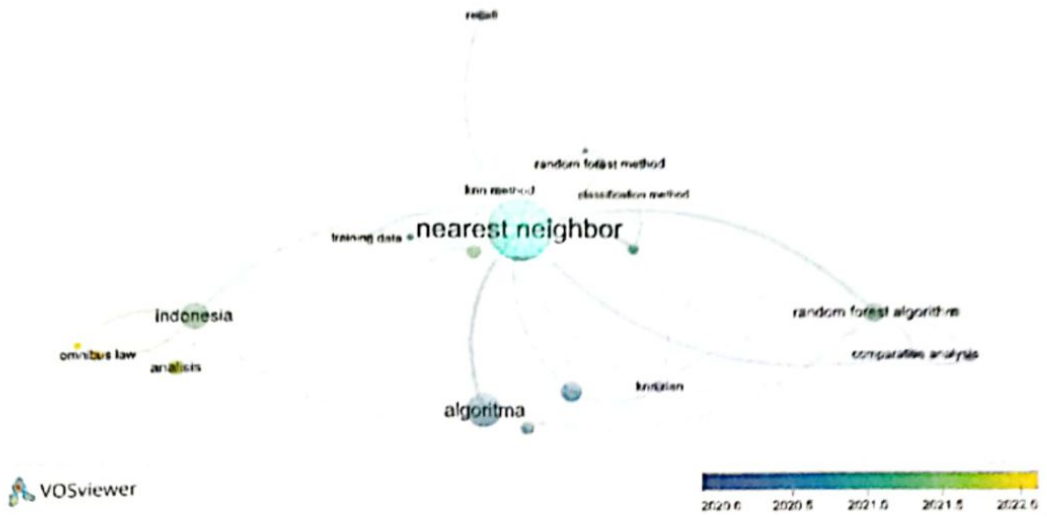
Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
Wijaya [13]	Merek		untuk mengklasifikasikan tweet. Hasilnya aplikasi yang dikembangkan dapat menentukan sentimen positif dan negatif dari setiap tweet.	menganalisis tweet, membuat daftar brand sesuai kategorinya dan divisualisasikan dalam bentuk *bar chart* sesuai dengan nilai rating tertinggi. Untuk melakukan klasifikasi teks penelitian ini mengimplementasikan AFINN Lexicon sehingga teks dapat dikelompokkan dalam kategori positif dan negatif, serta membuat grafik persentase sentimen setiap bulan dari data tweet terbaru.	sehingga dalam penelitian ini belum optimal karena sumber data yang dijangkau baru pada twitter saja.
Louis Madaerdo Sotarjua, Dian Budhi Santoso [14]	Perbandingan Algoritma KNN, *Decision Tree*, dan *Random Forest* Pada Data *Imbalanced Class* Untuk Klasifikasi Promosi Karyawan	2022	Berdasarkan hasil performa model klasifikasi, model KNN memiliki hasil performa yang terbaik nilai metriks evaluasinya. Dapat disimpulkan bahwa dengan model KNN, *Decision Tree,* dan *Random Forest* maka model KNN merupakan	Dataset yang digunakan pada penelitian ini merupakan data yang mengalami *imbalanced class*, sehingga untuk menyeimbangkan kelas data diterapkan metode SMOTE. Selain itu hasil performa model klasifikasi yang dilakukan tidak ditemukan *overfitting*	Peneliti tidak memberikan rekomendasi terhadap penelitian selanjutnya sehingga penelitian terkesan sudah final.

Peneliti	Judul Penelitian	Tahun	Metode dan Hasil Penelitian	Kelebihan	Kelemahan
			model klasifikasi yang lebih baik digunakan pada penelitian ini.	maupun *underfitting*, sehingga model dapat berperforma baik pada *training* maupun *testing*.	

Berdasarkan hasil penelitian sebagaimana diuraikan pada tabel 2.1. di atas, dapat disintesakan bahwa terlepas dari banyaknya penelitian empiris tentang teknik *data mining* dan media sosial, hanya sedikit penelitian yang membandingkan teknik penambangan data dalam hal akurasi, kinerja, dan kesesuaian. Misalnya, telah diamati bahwa keakuratan teknik pembelajaran mesin tertentu dihitung dengan cara yang berbeda, sehingga sulit untuk menemukan jawaban tentang kesesuaian teknik penambangan data, terutama terkait topik penelitian analisis sentimen terhadap PERPPU Tentang Cipta Kerja.

Sepanjang yang peneliti ketahui belum ada penelitian dengan topik optimalisasi akurasi klasifikasi menggunakan *machine learning* untuk menganalisis sentimen pengguna media sosial terhadap PERPPU Tentang Cipta Kerja menggunakan Algoritma KNN dan *Random Forest* berbasis *Particle Swarm Optimization (PSO)*. Dengan demikian, penelitian ini dapat dikatakan memiliki kebaruan (*novelty*) yang dapat dipergunakan untuk pengembangan ilmu pengetahuan dan sebagai salah satu dasar pertimbangan bagi pemangku kepentingan. Hal ini juga dibuktikan dengan hasil analisis bibliografi menggunakan aplikasi *Publish of Perish* dan *Vos Viewer* menggunakan data base dari *google schoolar, semantic scholar*, dan lainnya. Dari hasil analisis yang

dilakukan didapatkan *research gap* dengan penelitian yang dilakukan. Lebih jelasnya ditampilkan dalam gambar 2.1. berikut ini.



Gambar 2. 1. *Network Visualization*

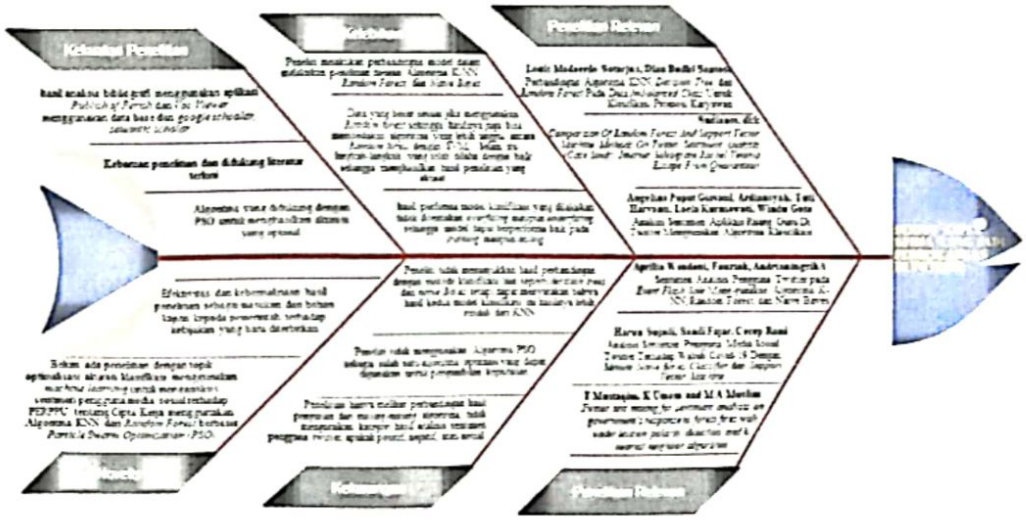
Berdasarkan gambar 2.1. di atas terlihat bahwa penelitian dengan kata kunci KNN, *Random Forest*, dan RUU Cipta Kerja belum banyak dilakukan. Hal ini terlihat dari hasil data bahwa sepanjang tahun 2020 sampai tahun 2022 terlihat kecil ukuran dan warna simpul. Ukuran simpul menunjukkan tingkat pentingnya simpul tersebut dalam jaringan, sedangkan warna simpul menunjukkan kelompok atau komunitas yang terbentuk dalam jaringan.

Selain itu, warna yang diberikan pada node dalam gambar menunjukkan adanya kelompok atau asosiasi yang terbentuk dalam jaringan. Node dengan warna yang sama menunjukkan hubungan atau karakteristik serupa. Warna-warna ini membantu mengidentifikasi pola atau hubungan yang mungkin tidak langsung terlihat dari data mentah.

Hasil visualisasi ini secara keseluruhan menunjukkan bahwa penelitian dengan kata kunci KNN, *Random Forest*, dan RUU Cipta Kerja belum

mendapatkan perhatian yang signifikan dalam rentang waktu yang disebutkan. Hal ini dapat diartikan bahwa ada potensi untuk lebih menggali dan mendalami topik ini dalam konteks penelitian lebih lanjut, mengingat minimnya literatur atau kajian yang telah ada sejauh ini.

Untuk memperjelas kekuatan penelitian ini dan *novelty* yang dihasilkan dibandingkan dengan penelitian-penelitian sebelumnya, penulis gambarkan dalam diagram *fishbone* yang ditampilkan pada gambar 2.2. berikut ini.



Gambar 2. 2. *Fishbone* Penelitian Terkait

2.2. Landasan Teori

2.2.1. *Machine Learning*

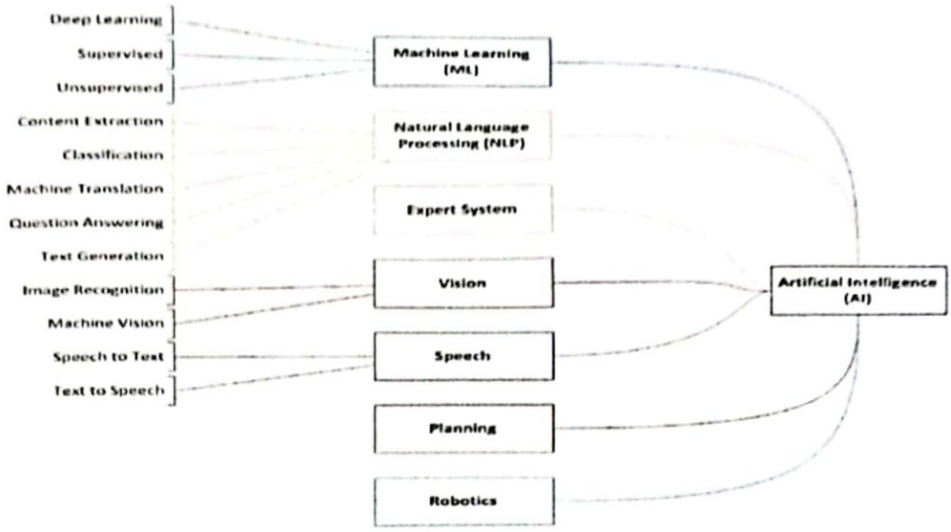
Perkembangan teknologi kecerdasan buatan atau *artificial intelligence* (AI) berkembang pesat. Kecerdasan buatan terdiri dari beberapa bidang, salah satunya adalah pembelajaran mesin. Pembelajaran mesin adalah cabang kecerdasan buatan

(AI) dan ilmu komputer yang berfokus pada penggunaan data dan algoritma untuk meniru pembelajaran manusia dan lebih meningkatkan akurasinya [15]. Algoritma pembelajaran mesin memungkinkan sistem secara otomatis belajar dan meningkatkan diri tanpa perlu pemrograman eksplisit dan berfokus pada pengembangan program komputer yang dapat mengakses data dan menggunakannya untuk belajar mandiri.

Proses pembelajaran mesin dimulai dengan mengamati data seperti pengalaman langsung dan instruksi. Tujuannya adalah untuk menemukan dan mempelajari pola data untuk menginformasikan keputusan masa depan yang lebih baik. Komputer/mesin secara otomatis mempelajari dan mengadaptasi tindakan yang tepat berdasarkan itu tanpa campur tangan/bantuan manusia [16]. Dengan kata lain, pembelajaran mesin memiliki kemampuan untuk mengumpulkan data sendiri, bukan perintah manusia. Data yang dihasilkan diperiksa oleh pembelajaran mesin untuk melakukan tugas tertentu. Tugas yang dapat dilakukan pembelajaran mesin sangat bervariasi tergantung pada apa yang dipelajarinya. Model pembelajaran mesin dapat memberikan akurasi tinggi sambil mempertahankan kecepatan tinggi [17].

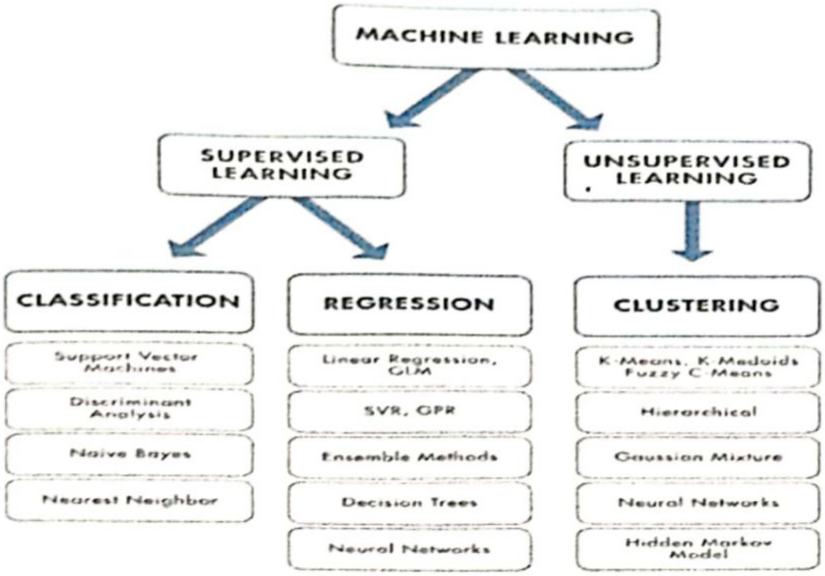
Kecerdasan buatan terapan dapat secara luas dibagi menjadi tujuh bidang: pembelajaran mesin (*machine learning)*, pemrosesan bahasa alami (*natural language processing)*, sistem pakar (*expert system)*, visi (*vision)*, bahasa (*speech)*, perencanaan (*planning)*, dan robotika (*robotics*). Percabangan ini bertujuan untuk mempersempit ruang lingkup pada saat pengembangan atau belajar AI, karena

pada dasarnya kecerdasan buatan memiliki ruang lingkup yang sangat luas [18]. Lebih jelasnya sebagaimana gambar 2.3 berikut [18].



Gambar 2. 3. Bidang Ilmu AI

Simulasi model *machine learning* sangat terkait dengan *Computational Statistics* yang tujuan utamanya adalah untuk fokus membuat prediksi melalui komputer. Terdapat banyak algoritma pembelajaran mesin yang diawasi (*supervised learning*) yang melatih model pada data masukan dan keluaran yang diketahui sehingga dapat memprediksi keluaran dimasa mendatang, dan tidak diawasi (*unsupervised learning*) yang menemukan pola tersembunyi atau struktur intrinsik dalam data masukan, dan masing-masing menggunakan pendekatan pembelajaran yang berbeda [19]. Selain itu, tidak ada metode terbaik atau satu ukuran cocok untuk semua. Pemilihan algoritma juga bergantung pada ukuran dan jenis data yang dikerjakan. Teknik *machine learning* ini dapat dijelaskan dalam gambar 2.4. berikut ini.



Gambar 2. 4 Tekhnik *machine learning* [20]

2.2.2. *Data Mining*

Pada bidang statistik, *data mining* dan *machine learning* semuanya berperan dalam memahami data, mendeskripsikan karakteristik kumpulan data, dan menemukan hubungan dan pola dalam data tersebut untuk membangun model [15]. *Data mining* adalah teknik penambangan yaitu proses mengidentifikasi pola dan tren data untuk mendapatkan informasi yang berguna dari kumpulan data yang sangat besar untuk membuat penilaian atau keputusan. Banyak teknik *data mining* telah dikembangkan dan digunakan dalam *data mining*. Termasuk asosiasi, klasifikasi, pengelompokan, pohon keputusan, prediksi, jaringan saraf, dan banyak lagi. Setiap teknik memiliki aturan dan metode yang menentukan jenis masalah yang dipecahkannya [4].

Data mining dapat diterjemahkan sebagai proses menemukan struktur yang menarik dalam data. Struktur dapat mengambil banyak bentuk, seperti

sekumpulan aturan, grafik atau jaringan, pohon, atau satu atau lebih persamaan. Struktur ini bisa menjadi bagian dari dasbor visual yang kompleks, atau bisa sesederhana daftar kandidat pemilu dan nomor terkait yang mewakili sentimen pemilih berdasarkan *feed Twitter* [21]. Proses menemukan struktur data yang ada membutuhkan Teknik yang disebut teknik *data mining*. Teknik *data mining* adalah proses identifikasi pola dan tren data untuk mendapatkan informasi yang berguna dalam kumpulan data yang sangat besar sehingga pengguna dapat menilai atau memutuskan [7].

Data mining terdiri dari penggalian informasi dan penemuan pengetahuan tanpa asumsi eksplisit, yaitu tanpa penelitian dan desain sebelumnya, informasi yang diperoleh harus memiliki tiga karakteristik: tidak diketahui sebelumnya, efektif, dan dapat ditindaklanjuti [7] [21]. Setiap bidang dibagi menjadi tiga kategori yang merujuk pada jenis pembelajaran mesin atau tugas *data mining* dan mengejar tujuan: deskriptif (misalnya mengidentifikasi pola yang tidak diketahui), prediktif (misalnya perkiraan berdasarkan pengetahuan yang tersedia), dan preskriptif (misalnya optimalisasi berdasarkan *machine learning* yang dikendalikan pengambilan keputusan).

Berdasarkan pengertian di atas, dapat disarikan bahwa *data mining* adalah proses otomatis dari jumlah data yang sangat besar, bertujuan untuk mendapatkan koneksi dan pola yang membawa informasi baru yang bernilai. *Data mining* juga merupakan proses pendukung keputusan yang mencari pola informasi dalam data. Penambangan data menggunakan satu atau lebih algoritma untuk mengidentifikasi tren dan pola yang menarik dalam data. Pengetahuan yang diperoleh dari sesi

penambangan data adalah model data umum. Tujuan utamanya adalah menerapkan apa yang ditemukan pada situasi baru.

2.2.3. Pengelompokkan *Data Mining*

Banyak teknik *data mining* yang telah dikembangkan dan digunakan dalam penambangan data. Setiap teknik memiliki tugas dan metodenya sendiri berdasarkan jenis masalah yang dipecahkannya. Tugas-tugas yang dikerjakan *data mining* dibagi menjadi beberapa kelompok [7] [15] yaitu:

a. Asosiasi (*Association*)

   Asosiasi adalah salah satu penambangan data yang terkenal teknik, menemukan pola berdasarkan hubungan antara variabel dalam transaksi yang sama. Asosiasi juga dikenal sebagai teknik relasi karena menggunakan hubungan antara item dan menemukan seringnya kemunculan item berbeda yang muncul dengan frekuensi tertinggi dalam kumpulan data [7]. Misalnya meneliti jumlah siswa dari sekolah tertentu yang diharapkan memberikan respons positif terhadap kualitas layanan pendidikan yang diberikan.

b. Klasifikasi (*classification*)

   Teknik klasifikasi digunakan untuk mengklasifikasikan suatu koleksi data ke dalam kelompok atau kelas yang berbeda untuk mendapatkan prediksi dan analisis yang akurat dalam kumpulan data yang sangat besar [7]. Klasifikasi memiliki target variabel kategorikal. Misalnya, kelompok pendapatan dapat dibagi menjadi tiga kategori: pendapatan tinggi, sedang, dan rendah.

c. Pengelompokkan *(Clustering)*

Clustering adalah pengelompokan *record*, pengamatan atau observasi untuk membentuk kelas objek yang serupa. *Cluster* adalah kumpulan *record* yang mirip satu sama lain dan berbeda dari *record* di *cluster* lain. Proses *clustering* melibatkan analisis satu atau lebih atribut untuk mengidentifikasi data yang mirip satu sama lain untuk memahami perbedaan dan persamaan antara kumpulan data [7].

d. Pohon keputusan (*Decision tree*)

Teknik pohon keputusan dapat diterapkan sebagai bagian kriteria seleksi. Selain itu, untuk membantu penggunaan dan pemilihan data tertentu dalam keseluruhan struktur [7].

e. Prediksi (*Prediction*)

Prediksi termasuk analisis tren, klasifikasi, pencocokan pola, dan hubungan. Prediksi dibuat dengan menganalisis peristiwa atau contoh masa lalu [7]. Prediksi mirip dengan klasifikasi dan estimasi, kecuali bahwa membuat prediksi nilai hasil masa depan. Contoh: Perkiraan curah hujan untuk wilayah tertentu.

f. Jaringan saraf Tiruan (*Neural Network*)

Neural Network adalah teknik penting yang banyak digunakan pada tahap awal teknologi *data mining*. Jaringan syaraf tiruan terbentuk dari komunitas kecerdasan buatan [7].

2.2.4. Tahapan Proses *Data Mining*

Data mining merupakan salah satu dari rangkaian *Knowledge Discovery in Databases* (KDD) yang berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Ungkapan KDD diciptakan pada tahun 1989 untuk menekankan bahwa pengetahuan dapat diturunkan dari data-*driven discovery* dan sering digunakan secara bergantian dengan *data mining* [21]. Proses penemuan pengetahuan meliputi langkah-langkah berikut [22].

a. Penemuan Data (*data discovery*): tahap ini adalah tahap pengumpulan data yang mencakup deteksi, identifikasi, dan karakterisasi data yang tersedia.
b. Pembersihan data dan Pembersihan (*data cleaning and cleaning*): kebisingan dan data yang tidak penting dihapus pada tahap ini, termasuk data yang bertentangan dan data yang tidak konsisten.
c. Integrasi data (*data integration*): pada tahap ini, data serupa dan terkait dikumpulkan dari beberapa sumber data dan digabungkan menjadi satu.
d. Pemilihan data (*data selection*): pada tahap ini, data yang sesuai diidentifikasi dan diambil dari kumpulan data.
e. Transformasi data (*data transformation*): pada tahap ini, data diubah menjadi bentuk khusus yang sesuai untuk prosedur pencarian dan pengambilan melalui umpan pencapaian atau operasi pengelompokan.

f. Penambangan data (*data mining*): pada tahap ini, penggunaan metode cerdas yang diterapkan untuk mengekstraksi pola data dan ekstraksi model yang berguna.

g. Evaluasi pola (*pattern evaluation*): pada tahap inilah pola yang sangat penting yang mewakili basis pengetahuan untuk penggunaan beberapa metrik penting diidentifikasi.

h. Presentasi pengetahuan (*knowledge presentation*): tahap ini adalah tahap terakhir dari penemuan pengetahuan dalam basis data, dan ini adalah tahap yang dilihat pengguna. Tahap dasar ini menggunakan metode visual untuk membantu pengguna memahami dan menginterpretasikan hasil ekstraksi data.

2.2.5. *Text Mining*

Pengolahan data yang banyak biasanya disebut *data mining* dan dalam ranah pengolahan data mentah teks disebut penambangan teks *(Text mining)* [9]. *Text mining* juga dikenal sebagai penambangan data teks, dirancang untuk memperoleh pengetahuan implisit yang tersembunyi di dalam yang tidak terstruktur teks [23]. *Text mining* merupakan proses ekstraksi berita dari data asal yang belum terstruktur. Data yang belum terstruktur diolah menggunakan teknik serta metode tertentu sehingga menjadi berita yang bermanfaat bagi pengguna [24].

Text mining adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, yang sebelumnya tidak diketahui dari informasi yang secara otomatis diambil dari berbagai sumber data tekstual. Proses

penambangan teks mirip dengan data klasik pengolahan. Pengambilan informasi dimaksudkan untuk memperoleh teks yang diinginkan, mirip dengan pengumpulan data. Ekstraksi informasi digunakan untuk mengekstrak informasi yang telah ditentukan sebelumnya, yaitu pemrosesan awal dari data yang dikumpulkan [23].

Tujuan dari tahap pemilihan fitur adalah mengurangi dimensi dari koleksi teks. Tujuan lainnya yaitu agar proses klasifikasi menjadi lebih efektif dan akurat dengan menghilangkan kata-kata yang dianggap tidak penting dan tidak menggambarkan isi dokumen yang diambil dari data twitter. Tindakan yang dilakukan peneliti pada tahap ini adalah menghilangkan kata *stopword* (*remove stopwords*) dan menghilangkan kata (*stemming*) yang berimbuhan.

2.2.6. Pra-Pemrosesan Data (*Preprocessing*)

Preprocessing merupakan proses normalisasi istilah yang berasal dari kalimat. Tahap ini dilakukan untuk mendapatkan data penelitian yang baik dan fitur yang diekstraksi tersinkronisasi dengan fitur yang diinginkan sehingga mempermudah pengolahan data. Pengumpulan data opini dari media sosial Twitter tidak boleh identik dengan kata baku, kata kamus atau bahasa daerah yang digunakan atau dihilangkan [24]. Teks-teks dikembalikan ke teks alami dengan melakukan eliminasi ekspresi tipikal agar dapat meminimalisir noise pada tahap selanjutnya sehingga perlu dilakukan normalisasi.

Teks yang diproses dalam proses *text mining* biasanya memiliki karakteristik seperti ukuran yang besar, *noise* pada data dan struktur teks yang

kurang baik. Caranya dengan terlebih dahulu menentukan fitur yang mewakili setiap kata untuk setiap fitur dalam dokumen. Sebelum menentukan fitur yang representatif, diperlukan langkah *preprocessing* yang biasanya dilakukan selama *text mining* dokumen [9] [25]. Adapun langkah-langkahnya sebagai berikut.

a. *Data cleaning*

Data cleaning merupakan proses membersihkan data yang dikombinasikan dengan *regex* untuk mendeteksi karakter yang tidak berguna dan langsung dihapus dari data utama untuk meningkatkan kualitas *dataset* [9]. Setiap data tweet dari media sosial Twitter biasanya mengandung banyak kata dan karakter yang tidak berguna untuk proses analisis data. Oleh karena itu, tahap ini adalah membersihkan data dari elemen yang tidak diinginkan seperti emoticon, link, username, dan hashtag.

b. *Tokenisasi*

Proses ini adalah proses pembuatan token, yaitu pembagian teks menjadi kata-kata. Proses tokenisasi membantu memudahkan analisis sentimen pada teks. Analisis kualitatif dan validasi empiris sangat penting dalam menentukan skor polaritas, karena memastikan bahwa hasil akhir memenuhi standar kualitas dan validitas yang diterima secara universal [9].

c. *Normalisasi / Case folding*

Teks yang diperoleh dari media sosial Twitter memiliki huruf besar atau kecil yang beragam. Supaya tidak mengganggu proses analisis data perlu diubah menjadi huruf kapital atau kecil semua. Normalisasi dilakukan untuk menghilangkan variasi dalam data. Contoh dari normalisasi adalah mengubah

semua teks menjadi huruf kecil dan menghilangkan tanda baca. Dengan *case folding*, misalnya huruf besar dan huruf kecil dalam kata sama-sama diubah menjadi huruf kecil, sehingga hasil pencarian tidak terpengaruh oleh perbedaan ukuran huruf. *Case folding* juga membantu mengurangi jumlah data yang harus dikenali oleh algoritma, sehingga dapat mempercepat proses analisis dan meminimalisir penggunaan memori [9].

d. *Remove Stopwords*

Stop word adalah kata yang tidak memiliki arti dalam analisis sentimen. Menghapus *stopwords* sangat berguna seperti preposisi, konjungsi, kata sifat, kata slank, kata ganti dan sebagainya. Kata-kata seperti ini biasanya muncul bersamaan dengan kata utama sehingga tidak unik, tidak memiliki arti tertentu, dan tidak berkontribusi. Daftar kata yang tidak berkontribusi terlalu banyak pada teks analitik disebut *stopword* atau *stoplist* [9].

e. *Stemming / Lemmatization*

Stemming dan *lemmatization* adalah proses untuk mengubah kata-kata dalam teks menjadi bentuk dasarnya. *Stemming* merupakan proses menghilangkan akhiran kata, sedangkan *lemmatization* merupakan proses mengembalikan kata-kata ke bentuk dasarnya dengan menggunakan kamus. *Stemming* merupakan proses menghilangkan atribut tambahan dari kata seperti menghapus "mem" dan "-kan" dari "making" menjadi "making". Pada analisis sentimen kata tidak baku mempengaruhi perhitungan analisis data [9].

2.2.7. Model TF-IDF

Model TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode pembobotan kata yang digunakan dalam analisis teks untuk mengukur tingkat pentingnya kata dalam dokumen atau korpus teks. Model ini bekerja dengan mengalikan frekuensi kemunculan kata dalam sebuah dokumen (*term frequency*) dengan nilai invers dari frekuensi kemunculan kata tersebut dalam seluruh dokumen (*inverse document frequency*). Tujuannya adalah untuk memberikan bobot yang lebih tinggi pada kata-kata yang penting dalam dokumen dan menghilangkan kata-kata yang umum atau tidak relevan.

Model TF-IDF mengukur tingkat pentingnya suatu kata dengan melihat seberapa sering kata tersebut muncul dalam dokumen dan seberapa umum kata tersebut dalam seluruh dokumen. Kata-kata yang muncul lebih sering dalam sebuah dokumen akan diberi bobot yang lebih tinggi, sementara kata-kata yang muncul lebih umum dalam seluruh dokumen akan diberi bobot yang lebih rendah. Model TF-IDF dapat digunakan untuk memproses setiap dokumen dalam korpus teks, termasuk dokumen dalam analisis sentimen di media sosial. Model TF-IDF dalam sentimen analisis dapat membantu mengidentifikasi kata-kata penting dalam dokumen dan membantu mengelompokkan dokumen ke dalam kelompok-kelompok yang berbeda berdasarkan pola kata-kata yang muncul di dalamnya.

Optimalisasi fitur perlu dilakukan ke data pelatihan dengan mengurangi dimensi data pelatihan sambil mempertahankan model tinggi ketepatan. Fitur optimasi yang digunakan adalah ekstraksi dari TF-IDF fitur. Ekstraksi fitur dengan TF-IDF merupakan salah satu proses dari teknik ekstraksi fitur dengan

proses nilai untuk setiap kata dalam data pelatihan. Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, itu dihitung. Nilai TF-IDF tergantung pada frekuensi kemunculan kata dalam dokumen [25].

Kata-kata yang tersusun dalam masing-masing tweet akan diberikan bobot pada tahapan ini. Caranya dengan mengalikan nilai *Term Frequency (TF)* dengan *Inverse Document Frequency (IDF)* [26]. Pengukuran bobot dilakukan dengan mencari kemunculan masing-masing kata yang terdapat dalam tweet pada *feature list* positif, negatif, dan netral. Tweet tersebut akan digunakan sebagai data latih dan juga data uji.

Menurut Salton [8], nilai IDF didapatkan dengan persamaan 1 sebagai berikut.

TF-IDF(w, d, D) = TF(w, d) x IDF(w, D) ..... (1)

Dimana:

TF(w, d) adalah nilai frekuensi kata (*term frequency*) w dalam dokumen d. Rumusnya dapat dinyatakan sebagaimana persamaan 2 berikut.

$$TF(w,d) = \left(\frac{\text{jumlah kemunculan kata w dalam dokumen d}}{\textit{jumlah kata dalam dokumen d}}\right) \text{ ..... (2)}$$

Sedangkan IDF(w, D) adalah nilai *inverse document frequency* dari kata w dalam koleksi dokumen D. Rumusnya dapat dinyatakan sebagaimana persamaan 3 berikut.

$$IDF\ (w,D) = \log 10\ \left(\frac{\mathrm{N}}{\mathrm{n\,w}}\right) \text{ .....(3)}$$

Dimana:

N adalah jumlah dokumen dalam koleksi dokumen D, dan n_w adalah jumlah dokumen dalam koleksi dokumen D yang mengandung kata w.

Nilai IDF semakin tinggi jika kata w semakin jarang muncul dalam koleksi dokumen, dan semakin rendah jika kata w muncul dalam banyak dokumen. Jadi TF-IDF(w, d, D) adalah nilai TF-IDF dari kata w dalam dokumen d dalam koleksi dokumen D.

2.2.8. Analisis Sentimen

Analisis sentimen Twitter semakin mendapat perhatian dalam beberapa tahun terakhir [27]. Analisis sentimen merupakan salah satu proses menganalisis kumpulan teks korpus yang bertujuan untuk menganalisis polaritas emosi baik emosi negatif, emosi netral maupun emosi positif [9]. Analisis sentimen adalah bidang penting yang memungkinkan kami memahami sikap umum pengguna tentang topik tertentu [28]. Analisis sentimen adalah teknik yang secara otomatis mengenali, mengekstrak, dan mengeksekusi informasi tekstual untuk menemukan informasi emosional dari ekspresi pikiran [24]. Analisis sentimen adalah jenis pemrosesan bahasa alami untuk melacak sentimen orang terhadap produk atau topik tertentu [29].

Analisis sentimen atau disebut juga *opinion mining* merupakan bidang yang menganalisis opini, perasaan, peringkat, penilaian, sikap, dan sentimen orang tentang entitas seperti produk, layanan, organisasi, individu, peristiwa, masalah, dan atributnya area penelitian [10]. Dalam penelitian ini analisis sentimen digunakan untuk menemukan informasi berharga yang dibutuhkan dari data yang tidak terstruktur. Tujuannya supaya dapat diketahui sentimen pengguna twitter

terhadap PERPPU Tentang Cipta Kerja. Berdasarkan uraian tersebut dapat disintesakan bahwa analisis sentimen merupakan proses menentukan opini atau sentimen seseorang yang diungkapkan dalam bentuk tekstual dan dapat diklasifikasikan sebagai positif atau negatif.

2.2.9. Teknik Klasifikasi

Teknik klasifikasi dijelaskan sebagai sebuah model dalam *data mining* dimana *classifier* diatur untuk memprediksi *categorical label*. Adapun algoritma klasifikasi yang akan digunakan dalam penelitian ini sebagai berikut.

a. ***K-Nearest Neighbors (KNN)***

K-Nearest Neighbor yang disingkat KNN merupakan algoritma klasifikasi yang prinsip dasarnya adalah perhitungan jarak terdekat [9] [8]. KNN merupakan salah satu algoritma paling sederhana yang digunakan untuk memecahkan masalah klasifikasi. KNN adalah algoritma yang mengklasifikasikan objek baru berdasarkan atribut dan sampel pelatihan [30].

Selain itu, J. Alzubi menegaskan bahwa KNN merupakan metode non-parametrik yang digunakan untuk klasifikasi dan regresi. Diberikan N vektor pelatihan, algoritma KNN mengidentifikasi k-tetangga terdekat dari vektor fitur yang tidak diketahui yang kelasnya akan diidentifikasi [19]. Algoritma KNN merupakan penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada

label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes [29]. Rumusnya sebagaimana persamaan 4 berikut.

$$d(x_{1,}, x_2) = \sqrt{(x_{1\,1} - x_{2\,1})^2 + \cdots + (x_{1p} - x_{2p})^2} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

Dimana:

$d(x_1,x_2)$: jarak antara variabel x_1 dan x_2

x : variabel

p : jumlah dimensi variabel

Cara menghitung jarak *euclidean* antara vektor tweet baru dengan vektor setiap tweet dalam dataset menggunakan persamaan 5 sebagai berikut.

$$d(x,y) = \sqrt{(\sum(x_i - y_i)^2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

di mana x dan y adalah vektor tweet baru dan vektor tweet dalam dataset, dan xi dan yi adalah elemen ke-i dari masing-masing vektor.

b. *Random Forest*

Random forest merupakan algoritma yang digunakan untuk pengklasifikasian *dataset* dalam jumlah besar dan menggabungkan *tree* dengan melakukan training *dataset* yang dimiliki. *Random forest* merupakan metode pembelajaran *ensemble* yang digunakan dalam klasifikasi dan regresi. Metode ini menggunakan pendekatan *bagging* untuk membuat sekumpulan pohon keputusan dengan subset data acak. *Output* dari semua pohon keputusan di hutan acak digabungkan untuk membuat pohon keputusan akhir. Terdapat dua tahap dalam algoritma *Random Forest*, satu adalah membuat hutan acak, dan yang lainnya membuat prediksi dari pengklasifikasi hutan acak yang dibuat pada tahap pertama [19].

c. ***Particle Swarm Optimization (PSO)***

Penemu PSO adalah James Kennedy dan Russ Eberhart pada tahun 1995. Inspirasi PSO yaitu tingkah laku sosial kawanan burung yang terbang berduyun-duyun *(bird flocking)* tanpa bertabrakan atau gerombolan ikan yang berenang berkelompok (*fish schooling*) bergerak cepat tanpa saling bertabrakan meskipun jarak mereka begitu dekat satu sama lainnya.

PSO adalah salah satu teknik komputasi evolusioner dengan populasinya didasarkan pada algoritma pencarian dan dimulai dengan populasi acak yang disebut partikel [31]. Terdapat beberapa cara dalam teknik PSO untuk melakukan optimasi, diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menyeleksi atribut (*attribute selection*) dan *feature selection* [8].

2.2.10. Evaluasi dan Validasi Model

Confusion matrix adalah metode evaluasi yang digunakan untuk mengevaluasi performa dari suatu model klasifikasi. Dalam suatu *confusion matrix*, data test dibandingkan dengan hasil prediksi dari model dan dicatat sebagai *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN) [32]. Setiap baris menunjukkan hasil aktual dari data test, sementara setiap kolom menunjukkan hasil prediksi dari model.

TP menunjukkan banyaknya data test yang benar-benar positif dan juga diprediksi sebagai positif oleh model. FP menunjukkan banyaknya data test yang

negatif tetapi diprediksi sebagai positif oleh model. TN menunjukkan banyaknya data test yang benar-benar negatif dan juga diprediksi sebagai negatif oleh model. FN menunjukkan banyaknya data test yang positif tetapi diprediksi sebagai negatif oleh model.

Dengan menggunakan informasi dari *confusion matrix,* beberapa metrik evaluasi seperti *accuracy, precision, recall*, dan F1 *score* dapat ditemukan untuk memahami performa dari suatu model [32]. Contoh positif (negatif) yang diklasifikasikan dengan benar oleh *classifier* disebut *True Positive (true negative)*, contoh positif (negatif) yang salah diklasifikasikan disebut *False Negative (false positive)* [33].

Model yang digunakan dievaluasi menggunakan metode *confusion matrix* sebagai indikasi aturan sifat klasifikasi (diskriminan). *Confusion matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel *matrix*. *Confusion matrix* ini berisi jumlah elemen yang telah dikelompokkan dengan benar atau tidak benar untuk setiap kelas. Tabel *matrix* yang digunakan dalam *mining data* disajikan dalam tabel 2.2. berikut ini.

Tabel 2. 2 *Confusion Matrix*

Correct classification	***Classified as***	
	+	-
+	*True positive*	*False negative*
-	*False positive*	*True negative*

Sumber : Bramer dalam Ibrahim (2017) dalam [33]

Dalam *confusion matrix* akan dihitung *acuracy, precision, recall dan f-measure* yang dirumuskan dengan persamaan sebagai berikut.

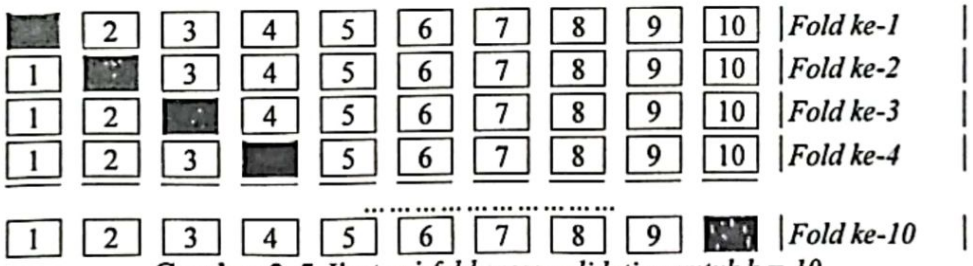
$$Acuracy = \frac{TP+TN}{TP+FN+FP+TN} x\ 100\% \quad \text{..... (6)}$$

$$Precision = \frac{TP}{TP+FP} x\ 100\% \quad \text{..... (7)}$$

$$Recall = \frac{TP}{TP+FN} x\ 100\% \quad \text{..... (8)}$$

$$F - measure = \frac{2 \text{ x precision x } recall}{precision+recall} x\ 100\% \quad \text{..... (9)}$$

Untuk melakukan validasi model, pada penelitian ini penulis memakai metode 10-*fold cross-validation.* Witten menjelaskan bahwa model validasi umum menggunakan 10 *fold cross validation* untuk data *learning* dan pengujian. Dengan kata lain, data *training* dibagi menjadi 10 bagian yang sama dan dilakukan proses *learning* sebanyak 10 kali. Bagian dari *dataset* untuk menguji 9 bagian yang tersisa dan digunakan dalam proses *learning.* Kemudian menghitung *mean* dan deviasi dari 10 hasil uji beda. Validasi silang 10 *fold cross validation* telah menjadi metode standar dan metode praktis validasi *state-of-the-art* [33]. Ilustrasi 10 *fold cross validation* sebagaimana gambar 2.5. berikut.



Gambar 2. 5. Ilustrasi *fold cross validation untuk k = 10*

Keterangan:

	Data Testing
	Data Training

Nilai Kappa juga menjadi salah satu bahan analisis peneliti untuk melihat algoritma mana yang mampu memberikan kinerja terbaik. Alat statistik yang paling banyak digunakan untuk menentukan kesepakatan adalah statistik kappa. Statistik kappa adalah ukuran yang dikoreksi secara kebetulan; yaitu, berusaha

untuk menilai kesepakatan di luar apa yang terjadi secara kebetulan [34]. *Performance* nilai kappa [35] dapat diklasifikasikan menjadi lima kelompok sebagaimana tabel 2.3. berikut.

Tabel 2. 3. Klasifikasi nilai kappa

No	Nilai Kappa	Klasifikasi
1	0.81 – 1.00	*Very Good*
2	0.61 – 0.80	*Good*
3	0.41 – 0.60	*Moderate*
4	0.21 – 0.40	*Fair*
5	0.00 – 0.20	*Poor*

2.3. Tinjauan Objek Penelitian

2.3.1. Media Sosial Twitter

Twitter merupakan salah satu media sosial yang memberikan penggunanya mengirim dan membaca pesan yang berbasis teks [10]. Twitter adalah salah satu media sosial yang paling populer di kalangan pengguna internet karena sederhana dan mudah digunakan. Pengguna bebas untuk mengeluarkan opini atau pendapatnya [8]. Twitter memiliki data terbanyak dibandingkan media sosial lainnya, sehingga membantu dalam hal akurasi penerapan metode analisis sentimen [36].

Salah satu fitur dalam twitter adalah *hashtag*. *Hashtag* digunakan untuk menunjukkan relevansi tweet dengan topik tertentu. *Hashtag* yang dibuat menggunakan karakter # diikuti dengan nama topik (#topik) telah muncul dari kebutuhan untuk memberi label informasi pada pesan yang diposting. Tag dihasilkan oleh pengguna secara spontan dan dapat digunakan untuk memperoleh

semua tweet dengan tagar yang sama. Tagar yang muncul disejumlah besar tweet dianggap sebagai topik yang sedang tren [37].

Informasi yang terdapat dalam tweet dapat digunakan untuk pengambilan informasi dan analisis. Komponen dalam twitter meliputi a) *User name* dan identifikasi pengguna memberikan informasi tentang siapa yang mengirimkan tweet, serta bisa memberikan *insight* tentang latar belakang dan afiliasi dari pengguna; b) *Hashtags* membantu dalam mengaitkan topik-topik tertentu dan mempermudah untuk mengidentifikasi tweet yang membahas topik yang sama; c) *Time Stamp* membantu dalam memahami waktu kapan tweet dikirim dan memungkinkan untuk melakukan analisis waktu; d) *Replies* membantu dalam memahami bagaimana interaksi antar pengguna dan bagaimana informasi tersebut diterima oleh komunitas; e) *Tweet text* adalah bagian yang paling penting karena mengandung informasi utama yang ingin disampaikan oleh pengguna, termasuk opini dan pandangan pengguna, dan f) *Retweet* membantu dalam memahami bagaimana informasi tersebut menyebar dan seberapa banyak pengguna yang membagikan tweet tersebut [2].

2.3.2. PERPPU Nomor 2 Tahun 2022 Tentang Cipta Kerja

Peraturan Pemerintah Pengganti Undang-Undang Republik Indonesia Nomor 2 tahun 2022 Tentang Cipta Kerja merupakan penyempurnaan dari Undang-Undang Nomor 11 Tahun 2020 tentang Cipta Kerja. Pada awalnya, UU Cipta Kerja merupakan pintu jalan untuk meningkatkan realisasi investasi di tahun 2022 [38] yang disusun melalui mekanisme *omnibus law*. Terobosan *Omnibus Law*

memungkinkan 80 Undang-Undang dan lebih dari 1.200 pasal direvisi dengan adanya Undang-Undang Cipta Kerja yang mengatur multi sektor [38]. Menurut pemerintah, reformasi Perppu harus mendorong investasi, membuka lapangan kerja lebih luas, meningkatkan efisiensi tenaga kerja dan mengurangi pengangguran [39].

Omnibus law berasal dari kata Latin yaitu *omnibus* yang berarti semua dan *Law* yang berarti hukum. Jadi *omnibus law* adalah undang-undang yang mengatur berbagai macam materi berbeda dalam satu undang–undang untuk semua orang. Intinya, *omnibus law* merupakan konsep yang disusun pemerintah Indonesia untuk mengatasi beragam permasalahan regulasi, dan memerlukan formula baru untuk mengatasi [40] regulasi dan aturan yang tumpang tindih sehingga tidak memakan waktu yang lama dan tidak mengeluarkan biaya secara berlebihan [41].

PERPPU Cipta Kerja adalah peraturan yang bertujuan untuk mempercepat proyek strategis nasional, termasuk mempromosikan, melindungi dan memberdayakan koperasi, usaha mikro dan UKM, meningkatkan ekosistem investasi, serta melindungi dan meningkatkan kesejahteraan pekerja. PERPPU Cipta Kerja berisi upaya pemerintah untuk menciptakan lapangan kerja dan mengatasi masalah pengangguran. PERPPU ini diterbitkan untuk menggantikan UU Nomor 11 Tahun 2020 tentang Cipta Kerja yang dinyatakan inkonstitusional bersyarat oleh Mahkamah Konstitusi (MK). PERPPU Cipta Kerja terbit satu tahun setelah putusan Mahkamah Konstitusi (MK) yang mengabulkan gugatan terhadap UU Cipta Kerja itu. Menurut MK UU Cipta Kerja bertentangan dengan konstitusi, sehingga UU Cipta Kerja dianggap inkonstitusional. Dalam dua tahun,

MK memberi amanat kepada DPR dan pemerintah agar memperbaiki UU itu. Setahun berlalu, pemerintah menerbitkan PERPPU, bukan perbaikan UU yang dilakukan.

2.3.3. Tweet tentang PERPPU Cipta Kerja

Sejak diterbitkan PERPPU Cipta Kerja, ramai terjadi perbincangan pro dan kontra terhadap kebijakan ini di media sosial Twitter. Kalangan masyarakat yang menolak penerbitan PERPPU beranggapan bahwa PERPPU Cipta Kerja merupakan bentuk pembangkangan, pengkhianatan, atau kudeta terhadap konstitusi RI karena dianggap terdapat banyak pasal yang merugikan tenaga kerja. Misalnya tweet dengan ID: 1612482615197720000 yang mentweet: *Ketika kalangan pengusaha menyambut baik adanya PERPPU cipta kerja itu wajar tapi ketika ada kalangan buruh yg setuju ada PERPPU cipta kerja itu nama nya kurang ajar #batalkanPERPPUciptakerja #batalkanPERPPUciptakerja.*

Bagi yang pro menganggap bahwa PERPPU Cipta Kerja dapat mengatasi berbagai macam persoalan yang muncul dalam dunia kerja dan lebih baik dari UU Cipta Kerja yang dibatalkan MK. Misalnya tweet dari ID 1612719691142750000 yang mengatakan: *PERPPU ciptaker lebih baik daripada UU Cipta Kerja. #JokowiPresidenku #PembangunanUntukRakyat #PresidenJokowiHebat #JokowiMembangunNegeri #PERPPUCiptaKerjaDukungPertumbuhanEkonomi #JokowiBapakInfrastruktur #JokowiMajuBersamaIndonesia.*