



BAB II

LANDASAN TEORI

© Hak cipta dimiliki IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

A. Tinjauan Pustaka

1. Informasi

Pengertian Informasi menurut Patricia M. Wallace (2019:7) adalah

“data or facts that are assembled and analyzed to add meaning and usefulness.”

Data atau fakta yang dikumpulkan menjadi satu, serta dirangkai dan dianalisis untuk menambah sebuah makna dan kegunaan.

Pengertian Informasi menurut Kenneth C.Laudon dan Jane P. Laudon (2019:16)

adalah

“By information we mean data that have been shaped into a form that is meaningful and useful to human beings”

Yang dimaksud dengan informasi adalah data yang telah diproses sedemikian rupa sehingga bermanfaat dan dapat dimengerti oleh manusia.

2. Data

Data adalah sebuah informasi yang didapatkan berdasarkan observasi, wawancara dan komunikasi yang akan dibutuhkan untuk kebutuhan tertentu, data bisa bermacam-macam variasi contohnya seperti : data sampel, data observasi, data primer dan data sekunder. Data dibagi menjadi 4 kategori, yaitu : sifat, sumber data, cara memperolehnya dan waktu pengumpulannya.

Menurut Syafrizal Helmi Situmorang (2014:2) data memiliki 2 sifat :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



C a. Kualitatif

Data yang tidak berbentuk angka, dan mempunyai ciri tidak bisa dilakukan operasi matematika seperti penambahan, pengurangan, perkalian dan pembagian. Contohnya adalah Kuesioner pertanyaan, kualitas pelayanan sebuah dealer mobil.

b. Kuantitatif

Data yang berbentuk angka yang bisa dilakukan dengan operasi matematika. Contohnya seperti pendapatan per kapita, jumlah orang mempunyai mobil.

Menurut Syafrizal Helmi Situmorang (2014:3) sumber data ada 2 macam :

a. Internal

Data dari dalam suatu organisasi yang menggambarkan keadaan organisasi tersebut. Contohnya seperti jumlah karyawan, jumlah pendapatan.

b. Eksternal

Data dari luar suatu organisasi yang dapat menggambarkan faktor-faktor yang mempengaruhi hasil kerja suatu organisasi. Contohnya seperti daya beli masyarakat dalam mempengaruhi hasil penjualan suatu perusahaan.

Menurut Syafrizal Helmi Situmorang (2014:3) cara memperoleh ada 2 macam :

a. Primer

Data yang dikumpulkan sendiri oleh perorangan/organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan dapat berupa *interview* dan observasi.\

Hak Cipta Dilindungi Undang-Undang

C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



© b. Sekunder

Data yang dikumpulkan dari studi-studi sebelumnya atau diterbitkan oleh instansi lain yang dapat berupa data dokumentasi dan arsip-arsip yang disimpan dalam penyimpanan di *website*.

Menurut Syafrizal Helmi Situmorang (2014:3) waktu pengumpulan ada 2 macam :

a. Cross Section

Data yang dikumpulkan pada suatu waktu tertentu untuk menggambarkan keadaan dan kegiatan pada waktu tersebut. Contohnya seperti data penelitian yang menggunakan kuesioner.

b. Time Series

Data yang dikumpulkan dari waktu ke waktu untuk melihat perkembangan kejadian/kegiatan selama periode tersebut. Misalnya, perkembangan uang beredar, harga 9 macam bahan pokok, penduduk.

3. Data Mining

Pengertian *Data Mining* Menurut Pan Ning Tan (2019:24) adalah

“Data mining is the process of automatically discovering useful information in large data repositories”

Data mining adalah sebuah proses otomatis untuk menemukan informasi yang bermanfaat didalam penyimpanan data yang besar.

Data mining merupakan bagian dari *Knowledge Discovery In Databases* (KDD) yang prosesnya mengkonversi data mentah menjadi informasi berguna, proses yang dilakukan dari data *preprocessing* sampai *postprocessing* hasil dari *data mining*. Berikut gambar yang menjelaskan proses tersebut.

Hak Cipta Dilindungi Undang-Undang

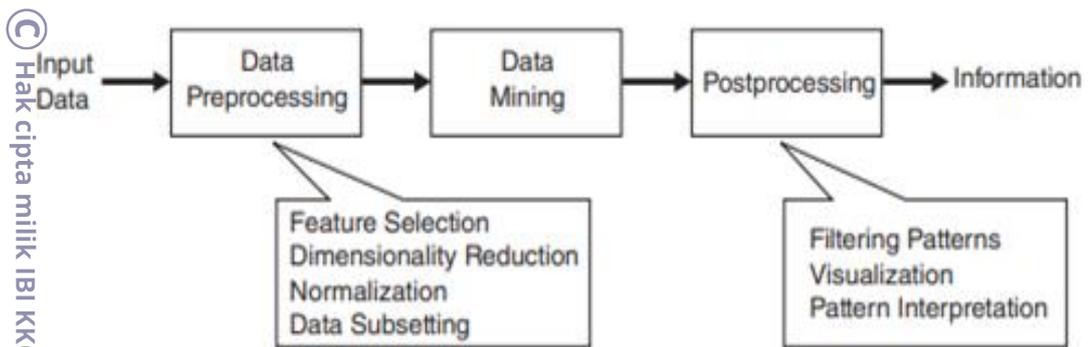
© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 2. 1
Proses Data Mining

Sumber : Pang Ning-Tan (2019:7)

Input data dapat disimpan kedalam beberapa format (Tabel Relasional, *Spreadsheets*) yang lalu disimpan kedalam penyimpanan data atau bisa didistribusi di beberapa area. Tujuan dari *Preprocessing* ialah untuk mengubah data mentah menjadi format yang digunakan untuk analisis, tahapan dari data *Preprocessing* seperti menggabungkan data dari beberapa sumber, membersihkan data untuk menghapus observasi duplikat, *noise*, menyeleksi catatan dan fitur yang relevan dalam data mining. Setelah proses *data mining* dilakukan maka selanjutnya proses *Postprocessing* memiliki fungsi untuk memastikan bahwa hasil valid dan berguna yang akan diintegrasikan pada pengambilan keputusan.

Tugas dari *data mining* dibagi menjadi 2 kategori :

a. Predictive Tasks

Tujuan dari tugas tersebut untuk memprediksi nilai dari atribut tertentu berdasarkan nilai atribut lainnya. Atribut yang akan diprediksi biasa disebut dengan variabel tidak bebas atau target sedangkan atribut yang dilakukan untuk prediksi disebut dengan variabel bebas atau bersifat penjelasan. (Pan Ning Tan 2019:29).

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

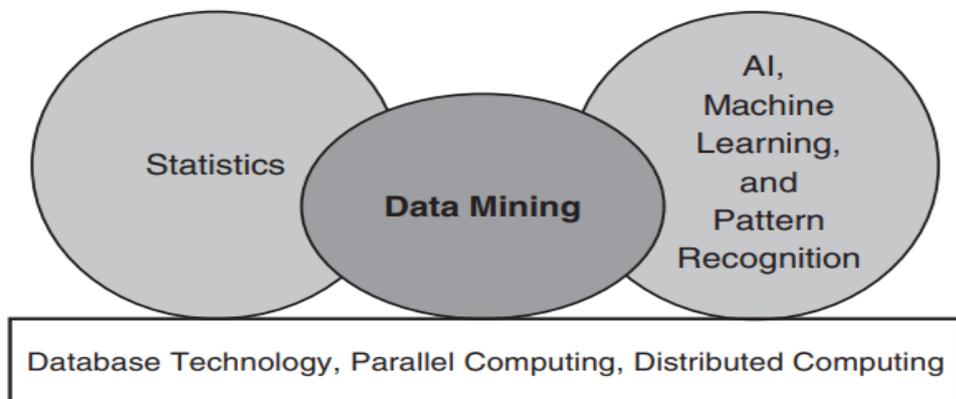
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



b. Descriptive Tasks

Tujuan dari tugas tersebut untuk mendapatkan pola (korelasi, tren, kluster, lintasan dan anomali) yang merangkum hubungan yang mendasari data. Untuk tugas deskriptif bersifat eksploratif dan sering membutuhkan teknik *postprocessing* untuk memvalidasi dan menjelaskan hasilnya. (Pan Ning Tan 2019:29).

Data mining membantu dalam bidang lainnya seperti statistik dengan pengujian hipotesis dan estimasi sampel, untuk *artificial intelligence* seperti algoritma pencarian, teknik modeling, *machine learning* dan pengenalan pola. Berikut gambar yang menunjukkan relasi antar bidang dari *data mining*.



Gambar 2. 2
Relasi Bidang *Data Mining*
Sumber : Pang Ning-Tan (2019:7)

Menurut Pan Ning Tan (2019:29) *data mining* memiliki beberapa metode pengolahan :

a. Predictive

Teknik prediksi digunakan apabila suatu nilai memiliki atribut yang berbeda, contohnya Algoritma Linear Regression, Neural Network dan lain-lain.

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

b. Association

Teknik Asosiasi digunakan untuk hubungan antar data, contohnya Algoritma Apriori.

c. Clustering

Teknik klustering digunakan untuk pengelompokan data dalam suatu kelompok tertentu, contohnya Algoritma K-Means, K-Medoids, Self Organization Map (SOM), C-Means.

d. Classification

Teknik klasifikasi mengelompokkan data yang mempunyai variabel tertentu berbeda dengan clustering yang tidak memiliki variabel yang dependen, contohnya Algoritma ID3, Naive Bayes dan K Nearest Neighbour.

4. Text Mining

Pengertian *Text Mining* menurut Ramesh Sharda, Dursun Dellen et al (2018:278)

adalah

“Text mining (also known as text data mining or knowledge discovery in textual databases) is the semiautomated process of extracting patterns (useful information and knowledge) from large amounts of unstructured data sources.”

Text mining (juga dikenal sebagai penambangan data teks atau penemuan pengetahuan dalam database teks) adalah proses semi otomatis ekstraksi pola (informasi dan pengetahuan yang berguna) dari sumber data yang tidak terstruktur dalam jumlah besar.

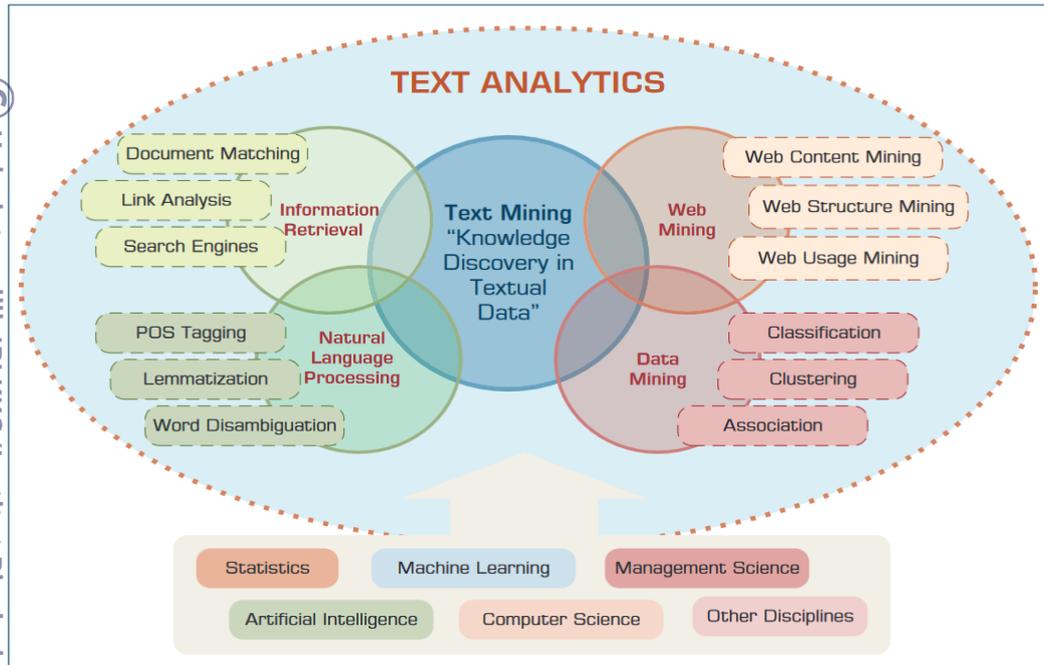
Berikut ilustrasi yang bisa menggambarkan fondasi dari sebuah *Text Mining*.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 2. 3
Fondasi Dan Relasi Area *Text Mining*
Sumber : Ramesh Sharda, Dursun et al (2018:277)

Text mining sama dengan *data mining* dalam hal tujuan dan menggunakan proses yang sama, namun pada *text mining* input prosesnya adalah kumpulan *file* data yang tidak terstruktur (atau kurang terstruktur) seperti dokumen Word, *file* PDF, cuplikan teks, *file* XML, dan sebagainya. Pada dasarnya, *text mining* dapat dianggap sebagai proses (dengan dua langkah utama) yang dimulai dengan memberikan struktur pada sumber data berbasis teks dan diikuti dengan mengekstrak informasi dan pengetahuan yang relevan dari data teks terstruktur ini menggunakan teknik dan alat *data mining*.

Menurut Ramesh Sharda, Dursun et al (2018:282) juga menjelaskan pengertian dari *Natural Language Processing* atau NLP adalah

“Natural language processing (NLP) is an important component of text mining and is a subfield of artificial intelligence and computational linguistics. It studies the problem of “understanding” the natural human language, with the view of converting depictions of human language (such as textual documents) into more formal representations (in the form of numeric and symbolic data) that are easier for computer programs to manipulate.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Pemrosesan bahasa alami (*Natural Language Processing* atau NLP) adalah komponen penting dalam *text mining* dan merupakan sub-bidang dari kecerdasan buatan dan linguistik komputasi. NLP mempelajari masalah "pemahaman" bahasa manusia alami, dengan tujuan mengkonversi teks-teks dalam bahasa manusia (seperti dokumen teks) menjadi representasi yang lebih formal (dalam bentuk data numerik dan simbolik) yang lebih mudah dimanipulasi oleh program komputer.

Tujuan dari NLP adalah untuk mencapai pemrosesan bahasa alami yang lebih canggih, yang melampaui manipulasi teks berdasarkan sintaksis atau perhitungan kata-kata, dan mempertimbangkan konteks, konstruksi semantik, dan peraturan tata bahasa.

5. Analisis Sentimen

Pengertian Analisis Sentimen menurut Bing Liu (2020:1) adalah

“Sentiment analysis, also called opinion mining, is the field of study that analyzes people’s opinions, sentiments, appraisals, attitudes, and emotions toward entities and their attributes expressed in written text.”

Analisis sentimen, juga disebut penambangan opini, adalah bidang studi yang menganalisis opini, sentimen, penilaian, sikap, dan emosi orang terhadap entitas dan atributnya yang diungkapkan dalam teks tertulis.

Entitas-entitas tersebut dapat berupa produk, layanan, organisasi, individu, acara, isu, atau topik. Bidang ini merepresentasikan ruang masalah yang besar.

Menurut Bing Liu (2020:1) ruang lingkup analisis sentimen

“Many related names and slightly different tasks – for example, sentiment analysis, opinion mining, opinion analysis, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining – are now all under the umbrella of sentiment analysis.”

Banyak nama terkait dan tugas yang sedikit berbeda, misalnya, analisis sentimen, penambangan opini, analisis opini, ekstraksi opini, penambangan sentimen, analisis



melaksanakan pemilu, yakni KPU, Bawaslu, serta DKPP. Kedudukan ketiga lembaga tersebut diperkuat dan diperjelas tugas dan fungsinya serta disesuaikan dengan perkembangan kebutuhan hukum dalam penyelenggaraan pemilu. Penguatan kelembagaan dimaksudkan untuk dapat menciptakan pemilu yang lancar, sistematis dan demokratis. Secara umum Undang-Undang ini mengatur mengenai penyelenggara pemilu, pelaksanaan pemilu, pelanggaran pemilu dan sengketa pemilu, serta tinda pindana pemilu.

C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

- a. Berdasarkan Peraturan Pemerintah Pengganti Undang-Undang (Perpu) No. 1 Tahun 2022 tentang perubahan atas Undang-Undang Nomor 7 Tahun 2017 tentang Pemilihan Umum bahwa, Sebagai implikasi dari pembentukan Provinsi Papua Selatan, Provinsi Papua Tengah, dan Provinsi Papua Pegunungan yang merupakan pemekaran dari Provinsi Papua serta pembentukan Provinsi Papua Barat Daya yang merupakan pemekaran dari Provinsi Papua Barat perlu kebijakan dan langkah luar biasa untuk mengantisipasi dampak pembentukan daerah baru tersebut terhadap penyelenggaraan tahapan pemilihan umum tahun 2024 agar tetap terlaksana sesuai dengan jadwal dan tahapan sehingga menciptakan stabilitas politik dalam negeri.

7. Twitter

Twitter merupakan layanan bagi teman, keluarga, dan rekan kerja untuk berkomunikasi dan tetap terhubung melalui pesan singkat. Melalui Twitter, orang dapat memposting tweet berisi teks, tautan, foto, dan video, yang kemudian akan *diposting* di profil, dikirimkan ke pengikut (*follower*) dan dapat dicari melalui fitur pencarian di Twitter.

Twitter memiliki 2 fitur dengan nama slogan unik yang khas :

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



a. Tweet

Tweet adalah pesan yang dapat mencakup foto, video, tautan, dan teks yang diposting di Twitter. Pengguna dapat memposting pembaruan di profil dengan menekan tombol tweet.

b. Retweet

Retweet adalah ketika memposting ulang sebuah tweet, dengan fitur retweet di Twitter, pengguna lain dapat dengan mudah berbagi tweet tersebut dengan semua pengikut (*follower*). Pengguna dapat melakukan Retweet pada tweet yang dibuat sendiri atau milik orang lain.

Ada kalanya orang mengetik "RT" pada awal Tweet untuk menandakan bahwa mereka sedang memposting ulang konten milik orang lain. Ini bukan sebuah perintah atau fitur resmi dari Twitter, namun hanya menunjukkan bahwa mereka mengutip tweet orang lain.

8. Machine Learning

Pengertian *Machine Learning* menurut Andriy Burkov (2019:3) adalah

“Machine learning is a subfield of computer science that is concerned with building algorithms which, to be useful, rely on a collection of examples of some phenomenon. These examples can come from nature, be handcrafted by humans or generated by another algorithm.”

Machine learning adalah cabang ilmu komputer yang berhubungan dengan membangun algoritma yang, untuk dapat berguna, mengandalkan kumpulan contoh fenomena tertentu. Contoh-contoh ini dapat berasal dari alam, dibuat oleh manusia, atau dihasilkan oleh algoritma lain.

Machine learning juga dapat didefinisikan sebagai proses pemecahan masalah praktis dengan cara mengumpulkan dataset dan membangun model statistik secara algoritmis

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



berdasarkan dataset tersebut. Model statistik tersebut diasumsikan digunakan untuk memecahkan masalah praktis tersebut.

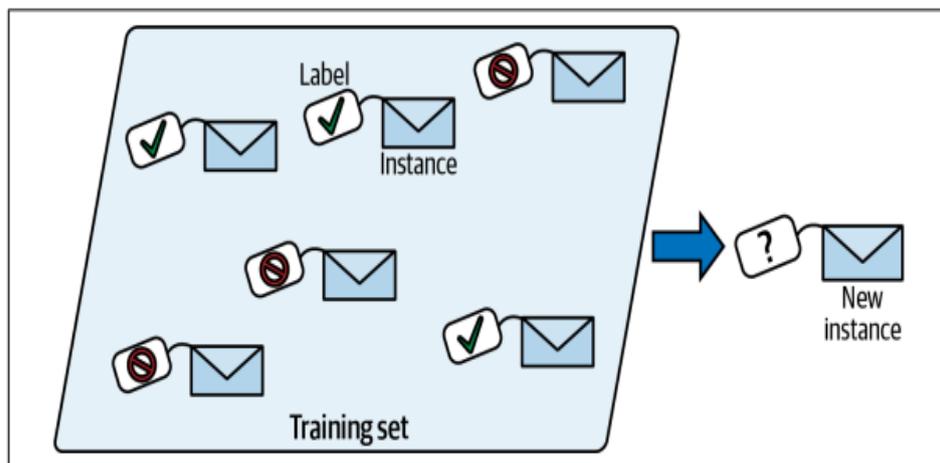
Pengertian *Machine Learning* menurut Aurelien Geron (2023:4) adalah
 “*Machine learning is the science (and art) of programming computers so they can learn from data.*”

Machine learning adalah ilmu (dan seni) dalam memprogram komputer agar dapat belajar dari data.

Menurut Aurelien Geron (2023:10) ada beberapa kategori dari *Machine learning*, beberapa kategori tersebut dapat dibagi, seperti berikut :

a. Supervised learning

Dalam *Supervised learning*, set pelatihan yang disediakan untuk algoritma mencakup solusi yang diinginkan, yang disebut label. Berikut ilustrasi yang dapat menjelaskan *Supervised learning*.



Gambar 2. 4
Contoh Supervised learning
 Sumber : Aurelien Geron (2023:10)

Tugas *Supervised learning* yang umum adalah klasifikasi. Filter spam adalah contoh yang baik untuk hal ini: ia dilatih dengan banyak email contoh beserta

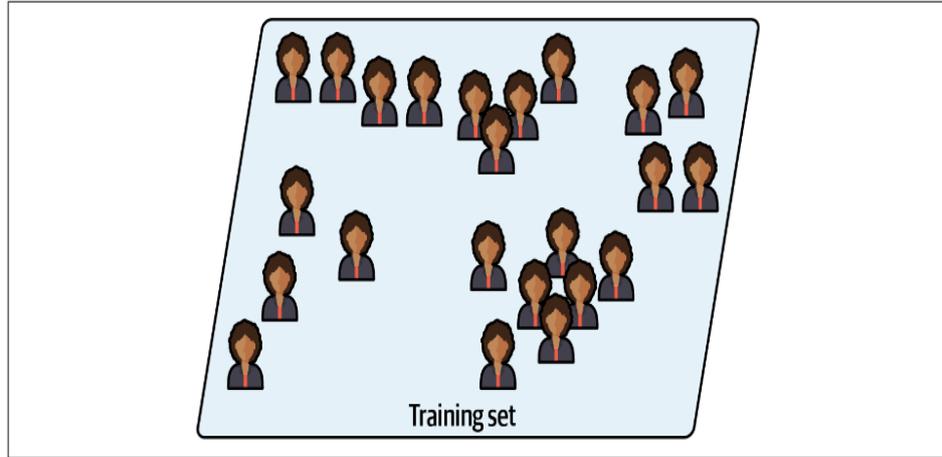
Hak cipta milik IBIKKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang
 Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

kelasnya (spam atau ham), dan harus belajar cara mengklasifikasikan email baru.

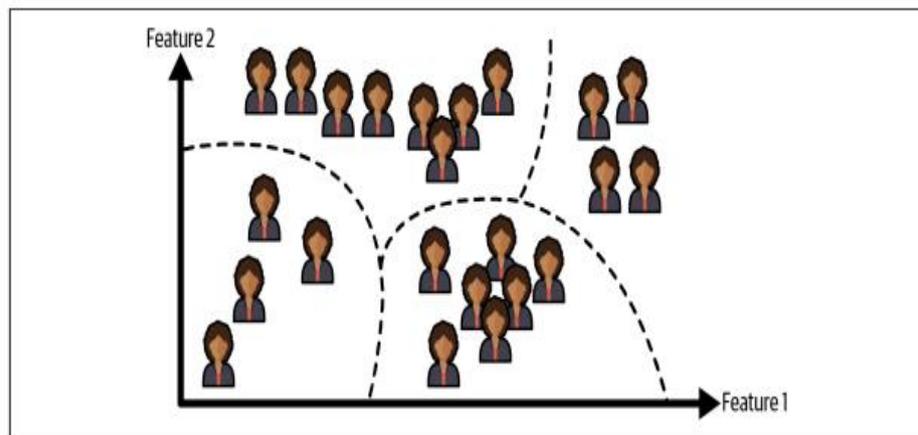
b. Unsupervised learning



Gambar 2. 5
Ilustrasi Unsupervised learning
 Sumber : Aurelien Geron (2023:12)

Dalam *Unsupervised learning*, seperti yang dapat dilihat pada gambar, data pelatihan tidak diberi label. Sistem mencoba belajar tanpa guru.

Berikut contoh yang dapat menjelaskan secara lengkap *Unsupervised learning*.



Gambar 2. 6
Contoh Unsupervised learning
 Sumber : Aurelien Geron (2023:12)

Dalam contoh tersebut, dijelaskan bagaimana penggunaan algoritma klustering dalam *Unsupervised learning* dapat membantu pemilik blog untuk memahami

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



kelompok pengunjung yang berbeda. Algoritma klustering dapat digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau kategori yang mirip berdasarkan pada pola yang ditemukan dalam data. Dalam contoh ini, pemilik blog ingin mengetahui lebih banyak tentang pengunjung blognya dengan melakukan clustering pada data pengunjung yang dimilikinya.

Tanpa harus memberi tahu algoritma kelompok mana pengunjung blog itu dapat dimasukkan, algoritma dapat menemukan sendiri pola yang terdapat pada data. Sebagai contoh, algoritma mungkin dapat mengelompokkan 40% pengunjung sebagai remaja yang suka buku komik dan sering membaca blog setelah sekolah, serta 20% pengunjung sebagai orang dewasa yang menikmati fiksi ilmiah dan mengunjungi blog selama akhir pekan. Dengan menggunakan algoritma pengelompokan hierarkis, kelompok-kelompok tersebut dapat dibagi lagi menjadi kelompok yang lebih kecil. Dengan mengetahui kelompok-kelompok pengunjung yang berbeda, pemilik blog dapat mengembangkan strategi pemasaran yang lebih baik.

c. *Semi-supervised learning*

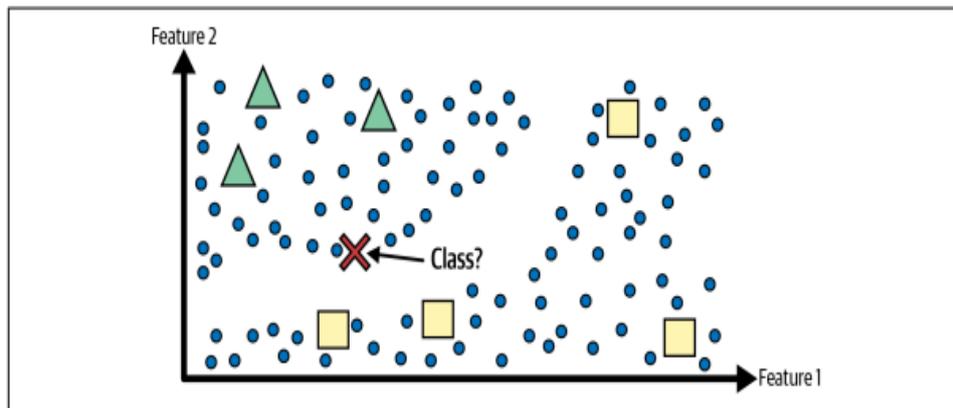
Karena memberi label pada data biasanya memakan waktu dan biaya yang cukup besar, maka seringkali Anda akan memiliki banyak contoh data yang tidak memiliki label, dan hanya sedikit contoh data yang sudah diberi label. Beberapa algoritma dapat menangani data yang sebagian telah diberi label, hal ini disebut *Semi-supervised learning*

Berikut contoh yang dapat menjelaskan Semi-supervised learning.

© Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang menggunakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 2. 7
Contoh *Semi-supervised learning*
Sumber : Aurelien Geron (2023:14)

Semi-supervised learning dengan dua kelas (segitiga dan persegi): contoh-contoh tidak berlabel (lingkaran) membantu mengklasifikasikan contoh baru (tanda silang) kedalam kelas segitiga daripada kelas persegi, meskipun contoh baru tersebut lebih dekat ke contoh persegi yang sudah diberi label.

d. ***Self-supervised learning***

Pendekatan lain dalam *machine learning* melibatkan pembuatan kumpulan data yang sepenuhnya diberi label dari kumpulan data yang sepenuhnya tidak diberi label. Sekali lagi, setelah seluruh kumpulan data diberi label, dapat menggunakan algoritma *Supervised learning* apapun. Pendekatan ini disebut *self-supervised learning*.

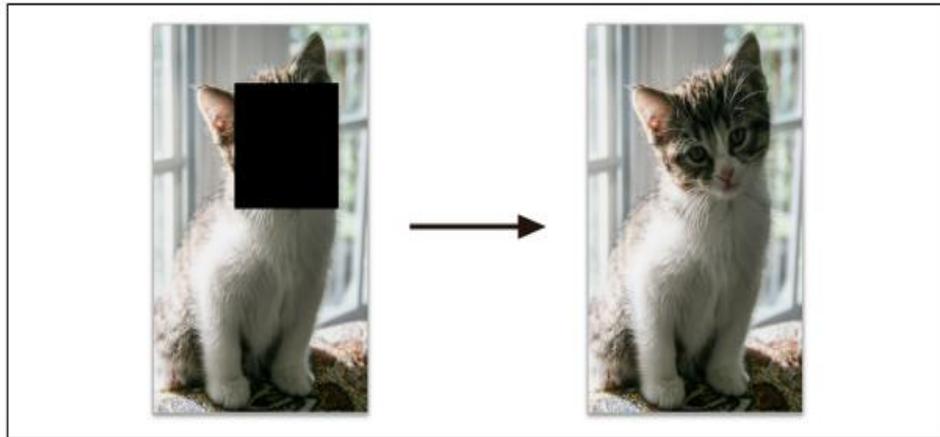
Contoh dan gambar berikut untuk menjelaskan *Self-supervised learning*, jika Anda memiliki dataset besar gambar yang tidak berlabel, Anda dapat secara acak menutupi sebagian kecil gambar dan kemudian melatih model untuk mengembalikan gambar asli. Selama pelatihan, gambar yang ditutupi digunakan sebagai input untuk model, dan gambar asli digunakan sebagai label.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Gambar 2. 8

Self-supervised learning

Sumber : Aurelien Geron (2023:15)

9. *Application Programming Interface (API)*

Pengertian *Application Programming Interface (API)* menurut Glenn Block, pablo

Cibraro et al (2014:23) adalah

“Web API is a programmatic interface to a system that is accessed via standard HTTP methods and headers. A Web API can be accessed by a variety of HTTP clients, including browsers and mobile devices.”

Web API adalah antarmuka pemrogramman untuk sistem yang diakses melalui metode dan header HTTP standar. Sebuah Web API dapat diakses oleh berbagai klien HTTP, termasuk browser dan perangkat *mobile*.

Pengertian API menurut JJ Geewax (2020:3) adalah

“An API defines the way in which computer systems interact. And since an exceptionally small number of systems live in a vacuum, it should come as no surprise that APIs are everywhere”

API adalah cara di mana sistem komputer berinteraksi, dan karena hampir tidak ada sistem yang beroperasi secara mandiri, maka tidak mengherankan jika API tersebar di mana-mana.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



10. Python

Ⓒ Pengertian Python menurut Al Sweigart (2019:28) adalah

“Python is a programming language (with syntax rules for writing what is considered valid Python code) and the Python interpreter software that reads source code (written in the Python language) and performs its instructions.”

Python adalah sebuah bahasa pemrograman dengan aturan *syntax* tertentu untuk menulis kode Python yang valid, dan sebuah perangkat lunak *interpreter* Python yang membaca kode sumber yang ditulis dalam bahasa Python dan menjalankan instruksinya.

Pengertian Python menurut William Wizner (2020:35) adalah

“Python is an object-oriented and interpretive computer program language. Its syntax is simple and contains a set of standard libraries with complete functions, which can easily accomplish many common tasks.”

Python adalah bahasa pemrograman komputer yang bersifat objek-berorientasi dan interpretatif. Sintaksnya sederhana dan terdapat kumpulan pustaka standar yang lengkap dengan fungsi-fungsi bawaan, yang memungkinkan untuk menyelesaikan berbagai tugas umum dengan mudah.

Menurut Eric Matthes (2019:38) alasan mengapa bahasa pemrograman Python digunakan oleh penggunanya adalah

“Python is an incredibly efficient language: your programs will do more in fewer lines of code than many other languages would require. Python’s syntax will also help you write “clean” code. Your code will be easy to read, easy to debug, and easy to extend and build upon compared to other languages.” (Laden, p. xxxvi) alasan mengapa menggunakan python”

Python adalah bahasa pemrograman yang sangat efisien: program Anda akan melakukan lebih banyak dalam jumlah baris kode yang lebih sedikit dibandingkan dengan banyak bahasa pemrograman lainnya. *Syntax* Python juga akan membantu Anda menulis kode yang "bersih". Kode Anda akan mudah dibaca, mudah diperbaiki,

Hak Cipta Dilindungi Undang-Undang

Ⓒ Hak Cipta dimiliki IBI IKKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Ⓒ Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



dan mudah diperluas dan dibangun dibandingkan dengan bahasa pemrograman lainnya.

1.1 Orange Data Mining

Orange Data Mining adalah perangkat lunak *open-source* yang digunakan untuk analisis data dan pembelajaran mesin (*Machine Learning*). Perangkat lunak ini menyediakan antarmuka *visual* yang intuitif untuk mengintegrasikan dan menganalisis data dengan berbagai metode, termasuk klustering, klasifikasi, regresi, dan visualisasi data. Orange Data Mining dirancang untuk mempermudah pengguna untuk memahami dan menerapkan algoritma pembelajaran mesin dan statistik dalam konteks analisis data.

Beberapa fitur yang mampu dilakukan pada Orange Data Mining :

a. *Interactive Data Visualization*

Visualisasi data untuk mengungkap pola data yang tersembunyi, memberikan intuisi dalam prosedur analisis data, atau mendukung komunikasi antara Data Scientists dan Domain Expert. Ada berbagai jenis *widget* visualisasi, termasuk scatter plot, box plot, histogram, serta visualisasi khusus model seperti dendrogram, silhouette plot, dan tree visualizations. Selain itu, banyak visualisasi lainnya tersedia dalam *add-ons*, seperti visualisasi jaringan, *word clouds*, peta geografis, dan lainnya.

b. *Visual Programming*

Orange merupakan alat yang sangat baik untuk penambangan data, baik bagi pemula maupun data scientist ahli. Dengan antarmuka pengguna yang dimilikinya, pengguna dapat lebih fokus pada analisis data daripada



menghabiskan waktu untuk *coding* yang melelahkan, sehingga memudahkan dalam membuat konstruksi *pipeline* analisis data yang kompleks.



Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

c. *Add-ons Extend Functionality*

Dapat memanfaatkan berbagai *add-on* (berupa *widget*) yang tersedia di dalam Orange untuk mengekstrak data dari sumber data eksternal, melakukan pemrosesan bahasa alami dan penambangan teks, menganalisis jaringan, mengekstrak *itemset* yang sering muncul, serta melakukan penambangan aturan asosiasi.

12. Penelitian

Menurut Umesh Kumar dan D.P. Kothari (2022:1) penelitian adalah

“Research is a process through which new knowledge is discovered. Research helps us to organize this new information into a coherent body, a set of related ideas that explain events that have occurred, and predict events that may happen.”

Penelitian adalah sebuah proses yang digunakan untuk menemukan pengetahuan baru. Dalam riset, kita dapat mengorganisir informasi baru ini ke dalam satu kesatuan yang teratur dan terkait satu sama lain, berupa kumpulan gagasan yang menjelaskan peristiwa yang terjadi dan memperkirakan peristiwa yang mungkin terjadi di masa depan.

13. Penelitian Kuantitatif

Menurut Umesh Kumar dan D.P. Kothari (2022:6) penelitian kuantitatif adalah

“Quantitative research is variables based whereas qualitative research is attributes based. Quantitative research is based on measurement or quantification of the phenomenon under study. That is, it is data based and hence more objective and popular.”

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Penelitian kuantitatif didasarkan pada variabel, sedangkan penelitian kualitatif didasarkan pada atribut. Penelitian kuantitatif melibatkan pengukuran atau kuantifikasi dari fenomena yang diteliti, sehingga memiliki dasar data dan lebih objektif, sehingga lebih banyak diminati.

B. Penelitian Terdahulu

1. Opini Masyarakat Twitter Terhadap Kandidat Bakal Calon Presiden

Republik Indonesia Tahun 2024

Penelitian yang dilakukan oleh Artha Dini Akmal, Lip Permana, Hidayatul Fajri dan Yuliarti menghasilkan kesimpulan yang dapat diambil :

Opini masyarakat twitter terhadap kandidat Bakal calon presiden republik indonesia tahun 2024 banyak menyebutkan nama Ganjar Pranowo dan Prabowo Subianto. Dari hasil pengolahan data yang diambil dari data Twitter mulai dari tanggal 23 Desember 2022 sampai tanggal 29 Desember 2022, Ganjar Pranowo merupakan kandidat calon Presiden yang sering disebut sebanyak 5836 kali. Dengan begitu dapat disimpulkan dari rentang waktu 23 sampai 29 Desember 2022 ini, masyarakat Twitter lebih banyak memberikan perhatian kepada Ganjar pranowo sebagai kandidat calon Presiden Indonesia.

2. Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law

Menggunakan Orange Data Mining

Penelitian yang dilakukan oleh Ratu Aghnia Raffaidy Wiguna dan Andri Irfan Rifai menghasilkan kesimpulan yang dapat diambil :

- Beberapa trending topik yang menjadi fokus pembicaraan mengenai omnibus law di twitter adalah mengenai buruh maju Bersama omnibus law, Bersama



pulih ekonomi bangsa, UU Cipta Kerja, waktunya milenial berbisnis, omnibus sejahtera, dimana yang sekiranya menjadi hal penting bagi pemerintah untuk diperhatikan keluhan rakyat.

- b. Metode analisis menggunakan tweet profiler dapat mengetahui mood atau emosi para pengguna twitter dengan trending topik yang sedang terjadi di Indonesia khususnya mengenai UU Cipta Kerja Omnibus Law.
- c. Dengan melakukan analisis *clustering* distribution kita dapat mengetahui klasifikasi pengguna twitter dengan visualisasi emosi yang telah diinput ke dalam setiap corpus dalam orange data mining.

3. Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

3. Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis

Penelitian yang dilakukan oleh Septian Fendyputra Pratama, Ricsa Andean, Aryo Nugroho menghasilkan kesimpulan yang dapat diambil :

Berdasarkan penelitian yang dilakukan tentang analisis sentimen percakapan di Twitter pada saat debat pertama calon presiden di Indonesia berlangsung pada tanggal 17 Januari 2018 yang menggunakan data dari Twitter pada jam 18.30 hingga 23.59 dapat disimpulkan bahwa, sentimen yang di tunjukan publik dunia maya pada saat debat berlangsung cenderung bersentimen positif.

Berdasarkan hasil pada tabel dan barplot pada pemrogramman R, hasil persentase, baik tweet yang menggunakan hashtag #JokowiAminMenangDebat maupun hashtag #PrabowoIndonesiaMenang menunjukkan cenderung mengandung sentimen positif lebih banyak dari sentimen negatif dan netral pada tiap tweet. Untuk sentimen negatif pada kedua hashtag, jumlahnya cukup banyak pada setiap tweet yang diposting saat debat berlangsung, walaupun memang tak sebanyak sentimen positif.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



Hasil sentimen netral pada kedua hashtag ini, bisa dikatakan cukup banyak dan bisa disimpulkan juga bahwa, terdapat pengguna Twitter yang tidak terlalu berorientasi pada salah satu calon presiden dan wakil presiden tertentu dan cenderung menjadi pemilih yang belum menentukan pilihan mereka saat pemilihan umum nanti yang menjadikan debat sebagai cara meyakinkan mereka untuk kepada siapa mereka akan menentukan pilihannya dan memposting tweet mereka di Twitter saat debat berlangsung.

4. Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter

Penelitian yang dilakukan oleh Cahyo Prianto, Nisa Hanum Harani, Indra Firmansyah menghasilkan kesimpulan yang dapat diambil :

- a. Media sosial twitter dapat digunakan untuk menganalisis kecenderungan sebuah opini netizen baik itu untuk analisis sebuah produk atau tokoh yang dianggap penting dalam suatu bidang.
- b. Kandidat 01 Joko Widodo diketahui lebih banyak menjadi opini massa sehingga mendapatkan respon sentimen yang besar dari pada kandidat 02 Prabowo Subianto. Baik itu sentimen positif maupun sentiment negatif.
- c. Menggunakan data bulan maret sebelum pemilu 2019, terlihat bahwa baik menggunakan NRC lexicon maupun Bing lexicon, kandidat 01 Joko Widodo mendapat respon lebih banyak dengan sentimen yang lebih positif dibandingkan dengan kandidat 02 Prabowo Subianto. Untuk kandidat 01 Joko Widodo analisis NRC lexicon memberikan nilai sebesar 249 dan bing lexicon sebesar 267 dengan nilai rata-rata 0,11, sedangkan untuk kandidat 02 Prabowo Subianto analisis NRC lexicon memberikan nilai sebesar 195 dan bing lexicon sebesar 204 dengan nilai rata-rata 0,085.

Hak Cipta Dilindungi Undang-Undang

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.



- d. Dengan menggunakan data bulan april setelah pemilu 2019, terlihat adanya pergeseran opini terhadap dua kandidat calon presiden. Kandidat 01 Joko Widodo tetap mendapatkan respon yang banyak dari masyarakat netizen tetapi nilai sentimennya bergeser ke arah lebih negatif dibandingkan dengan kandidat 02 Prabowo Subianto. Untuk kandidat 01 Joko Widodo analisis NRC lexicon memberikan nilai sebesar 17 dan bing lexicon sebesar -273 dengan nilai rata-rata -0,246, sedangkan untuk kandidat 02 Prabowo Subianto analisis NRC lexicon memberikan nilai sebesar 238 dan bing lexicon sebesar -73 dengan nilai rata-rata -0.02430939.

5. Sentimen Analisis Kegiatan Trading Pada Aplikasi Twitter dengan Algoritma SVM, KNN Dan Random Forrest

Penelitian yang dilakukan oleh Neng Resti Wardani, Sudin Saepudin, Cecep Warman menghasilkan kesimpulan yang dapat diambil :

- Berdasarkan hasil analisis yang dilakukan peneliti bahwa respon atau komentar masyarakat terhadap kegiatan trading menuai respon netral karna kegiatan trading hanya beberapa situs ilegal yang mengatasnamakan trading dan itupun sudah dilaporkan kepada pihak yang berwajib.
- Dari ke 3 metode yang digunakan tersebut maka metode yang paling unggul adalah metode KNN karena metode tersebut menggunakan probabilitas dan statistik.
- Tool yang digunakan untuk analisis sentimen yaitu aplikasi orange 3, diharapkan kedepannya mudah-mudahan ada peneliti yang menggunakan tools lain untuk analisisnya agar lebih baik dan lengkap tools nya.
- Perbandingan Metode Klasifikasi hasil perbandingan antara beberapa metode klasifikasi, metode yang menghasilkan akurasi paling tinggi adalah dengan

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

- Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
- Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

metode KNN. Selanjutnya terdapat metode Random forest dan Metode SVM dengan masing-masing akurasi 0,994 dan 0.992.

C Hak cipta milik IBI KKG (Institut Bisnis dan Informatika Kwik Kian Gie)

Hak Cipta Dilindungi Undang-Undang

Institut Bisnis dan Informatika Kwik Kian Gie

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik dan tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IBIKKG.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IBIKKG.

