

BAB II

LANDASAN TEORI

2.1. Algoritma

Algoritma adalah kumpulan perintah untuk menyelesaikan suatu masalah dimana masalah di selesaikan di tuntut secara sistematis, terstruktur dan logis (Maulana, 2017). Algoritma juga merupakan penjelasan atau deskripsi langkah-langkah dari penyelesaian masalah yang tersusun secara sistematis (Fathoni, 2015). Dalam pemecahan masalah, ada kriteria tertentu tentang kondisi awal yang harus dipenuhi sebelum menjalankan algoritma. Algoritma akan selalu berakhir untuk semua kondisi awal yang memenuhi kriteria. Mulai dari nilai awal, kemudian sekumpulan perintah saat dijalankan akan memproses kondisi yang telah ditetapkan untuk menghasilkan keluaran dan kemudian tentukan kondisi akhir. Algoritma juga merupakan prosedur komputasi yang menerima sekumpulan nilai sebagai masukan untuk ke urutan langkah kalkulasi tertentu yang nantinya akan dikonversikan menjadi satu set nilai yang disebut *output*. Algoritma digunakan sebagai alat untuk memecahkan masalah komputasi yang mendefinisikan hubungan antara *input* dan *output*. Algoritma juga menjelaskan prosedur khusus untuk mendapatkan hubungan antara *input* dan *output* (Sartika & Indra, 2017).

2.2. Natural Language Processing

Natural Language Processing adalah cabang ilmu komputer dan linguistik yang meneliti interaksi antar komputer dan bahasa manusia (bahasa alami). Riset NLP mencakup segmentasi suara antara lain segmentasi tuturan (*speech segmentation*), segmentasi teks (*text segmentation*), penandaan kelas kata (*partof-speech*), serta pengawataksaan makna (*word sense disambiguitas*) (Hoppe et al., 2018). meskipun konten penelitian dapat mencakup teks dan ucapan, pemrosesan tuturan (*speech processing*) telah berkembang menjadi wilayah penelitian tersendiri. sejarah NLP dimulai pada 1950-an, meskipun ada penelitian NLP pada waktu itu sebelum-sebelumnya (Hermawan, 2011). Natural Language Processing (NLP) juga merupakan area penelitian dan aplikasi yang membahas metode komputer untuk memahami dan menggunakan teks bahasa manusia untuk berbagai hal yang berguna. Pembentukan sistem NLP sendiri biasanya dimulai pada level kata pertama, kemudian diikuti oleh tingkat kalimat untuk memahami seluruh konteksnya (Goldberg, 2016).

Proses penyajian bahasa dalam bentuk yang lebih memungkinkan untuk sebuah perhitungan yang dilakukan oleh seorang ahli bernama Chomsky pada tahun 1957. Dalam bidang pemrosesan bahasa alami, hal ini tidak mudah karena ada beberapa alasan yang menghambat pemrosesan bahasa alami yaitu kesulitan dalam mengolah bahasa alami adalah memilih arti yang benar dari kata yang memiliki makna ganda seperti kata "mungkin" yang dapat berarti "racun" atau "mungkin" tergantung kalimatnya. Fenomena ini terjadi saat menentukan jenis kata (*part of speech*) seperti kata "lanjutan" yang bisa berfungsi sebagai kata kerja aktif di masa lalu, kata kerja

pasif, atau kata sifat. Selain itu, jumlah kosakata dalam bahasa alami sangat banyak dan terus bertambah. NLP biasanya digunakan untuk aplikasi bahasa tertulis, namun saat ini, aplikasi bahasa lisan juga dikembangkan. Dalam bahasa lisan, orang sering membentuk tuturan yang tidak sesuai dengan aturan yang berlaku. Salah satu aplikasi pengolah bahasa alami adalah sistem penerjemahan bahasa alami (*natural language translation*), seperti sistem terjemahan dari bahasa Inggris ke bahasa Indonesia. Penerjemah tidak hanya dapat menerjemahkan kata demi kata, tetapi juga menerjemahkan bahasa asli ke bahasa target dengan makna yang dapat dimengerti (Desiani, Primartha, Arhami, & Orsalan, 2019).

2.3. Python

Python adalah bahasa pemrograman dinamis yang umum digunakan untuk pengembangan aplikasi di berbagai bidang. Python pada dasarnya memiliki gaya yang mirip dengan pseudocode, satu-satunya perbedaannya adalah Python bisa dijalankan di komputer dan menampilkan hasilnya. Bahasa Python itu mudah dan dapat dimengerti serta tidak perlu dikompilasi, Jadi kita bisa menjalankan *script* Python seperti yang kita lakukan dan simpan tanpa langkah tambahan berikutnya. Python dibuat pada tahun 1990 oleh Guido van Rossum dari Belanda dan dinamai menurut acara TV favorit Guido Monty Python, Flying Circus. Van Rossum mengembangkan Python sebagai hobi, dan kemudian Python menjadi bahasa pemrograman karena bahasa tersebut banyak digunakan dalam industri dan pendidikan sederhana, ringkas, sintaks yang intuitif, dan memiliki perpustakaan yang luas. Python adalah pemrograman berorientasi objek (OOP). Data Python

adalah objek yang terbuat dari *class*. Pemrograman berorientasi objek adalah alat pengembangan perangkat lunak yang kuat dan dapat digunakan kembali. Saat ini, tim relawan besar sedang mengembangkan serta memelihara Python dan akan tersedia secara gratis dari Python Software Foundation. Python mendukung banyak paradigma pemrograman, tetapi sebenarnya tidak terbatas seperti di bidang pemrograman berorientasi objek, pemrograman imperatif dan pemrograman fungsional. Salah satu fitur yang tersedia di Python adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis seperti bahasa pemrograman dinamis lainnya, Python biasanya digunakan sebagai bahasa *script*, meskipun dalam praktiknya penggunaan bahasa ini secara lebih luas mencakup konteks umum penggunaan yang tidak dijalankan menggunakan bahasa *script*. Python bisa digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat dijalankan berbagai platform sistem operasi, saat ini kode Python dapat dijalankan platform sistem operasi yang berbeda (Chollet, 2018).

2.4. Machine Learning

Di era perkembangan pesat teknologi kecerdasan buatan, sebuah cabang kecerdasan buatan muncul dan menarik perhatian besar para peneliti. Salah satunya adalah *Machine Learning*. *Machine Learning* dapat mempelajari teori suatu pembelajaran mesin atau komputer sehingga komputer bisa belajar dari data. *Machine Learning* melibatkan berbagai disiplin ilmu misalnya statistika dalam matematika, ilmu komputer bahkan neurologi. Istilah *Machine Learning* pertama ditentukan oleh Arthur Samuel pada tahun 1959 (Tu, 2019). *Machine Learning* juga

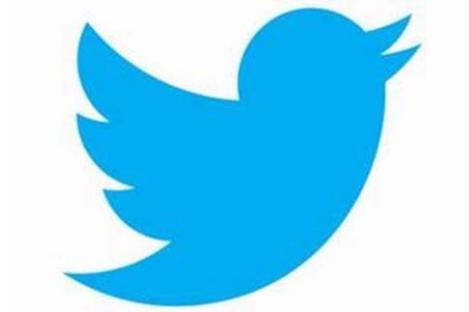
dapat menentukan peningkatan kinerja atau dapat membuat prediksi yang akurat. Menurut J. Putra, *Machine Learning* (ML) adalah inferensi data melalui metode matematika. Jadi pada dasarnya *Machine Learning* adalah pembuatan model matematika yang menggambarkan pola data. *Machine Learning* adalah salah satu dari banyak kecerdasan buatan dalam tahap pemecahan masalah untuk data dan pembelajaran, ada dua jenis *Machine Learning* yang dapat digunakan yaitu klasifikasi dan prediksi. Fitur yang menonjol dari *Machine Learning* itu adalah proses pelatihan. *Machine Learning* sangat membutuhkan data tersebut untuk dijadikan data latih. *Machine Learning* mengacu pada sebuah metode pembuatan komputer yang mempunyai kemampuan untuk belajar dan bekerja secara otomatis. Proses *Machine Learning* dilakukan melalui algoritma tertentu sehingga pekerjaan yang dipesan ke komputer dapat diselesaikan secara otomatis (Hairani, 2018).

2.5. *Natural Language Understanding (NLU)*

Natural Language Understanding (NLU) merupakan salah satu cabang dari *artificial intelligence* (AI) yang menggunakan perangkat lunak komputer untuk memahami masukan yang dibuat dalam bentuk kalimat dalam format teks atau ucapan. NLU secara langsung memungkinkan interaksi manusia-komputer. Pemahaman NLU tentang bahasa manusia alami memungkinkan komputer untuk memahami perintah tanpa sintaks bahasa komputer formal dan komputer untuk berkomunikasi kembali ke manusia dalam bahasa mereka sendiri. Bidang NLU adalah bagian penting dan menantang dari pemrosesan bahasa alami (NLP). Meskipun keduanya memahami bahasa manusia, NLU bertugas untuk

berkomunikasi dengan individu yang tidak terlatih dan memahami maksud mereka, yang berarti bahwa NLU lebih dari sekadar memahami kata-kata dan menafsirkan makna. NLU bahkan diprogram dengan kemampuan untuk memahami makna meskipun ada kesalahan manusia yang umum seperti kesalahan pengucapan atau transposisi huruf atau kata. NLU menggunakan algoritma untuk mereduksi ucapan manusia menjadi ontologi terstruktur. Natural Language Understanding (NLU) is defined by Gartner as “the comprehension by computers of the structure and meaning of human language (e.g., English, Spanish, Japanese), allowing users to interact with the computer using natural sentences”. In other words, NLU is Artificial Intelligence that uses computer software to interpret text and any type of unstructured data. NLU can digest a text, translate it into computer language and produce an output in a language that humans can understand. Natural Language Understanding (NLU) juga didefinisikan oleh Gartner sebagai "pemahaman dari komputer tentang struktur dan arti bahasa manusia (misalnya Bahasa Inggris, Spanyol, Jepang), yang memungkinkan pengguna untuk berinteraksi dengan komputer menggunakan kalimat alami". Dengan kata lain, NLU adalah Artificial Intelligence yang menggunakan perangkat lunak komputer untuk menafsirkan teks dan semua jenis data tidak terstruktur. NLU dapat mencerna suatu teks, menerjemahkannya ke dalam bahasa komputer dan menghasilkan keluaran dalam bahasa yang dapat dipahami manusia (Gartner, 2015).

2.6. *Twitter*



Gambar 2. 1 Logo Twitter

Sumber: (“Twitter bird logo”, 2016)

Gambar 2.1 *Twitter* merupakan situs berita dan jejaring sosial online tempat orang berkomunikasi dalam pesan singkat yang disebut tweet. Tweeting memposting pesan singkat untuk siapa saja yang mengikuti anda di *Twitter*, dengan harapan kata – kata anda bermanfaat dan menarik bagi seseorang di audiens anda. Deskripsi lain tentang *Twitter* dan tweeting mungkin *microblogging* (Shelar & Huang, 2018). *Twitter* didirikan dan diresmikan pada tahun 2006 tepatnya pada bulan maret. Didirikan oleh Jack Dorsey, *Twitter* sudah cukup di kenal di dunia, bahkan pada tahun 2014 *Twitter* sudah menjadi salah satu dari 5 besar situs yang paling sering dikunjungi oleh banyak orang (Basri, 2017). *Twitter* sendiri memiliki beberapa fitur, seperti Tweet atau kicauan yang merupakan fitur utama di *Twitter* untuk mengirim dan melihat kicauan setiap pengguna *Twitter*. *Following* merupakan fitur untuk mengikuti pengguna lain pada *Twitter*. *Followers*, fitur yang digunakan untuk melihat siapa yang mengikuti anda pada media sosial *Twitter*. *Biography* merupakan fitur yang digunakan untuk mengetahui pesan akun *Twitter*

anda pada profil. Profil merupakan salah satu fitur utama pada *Twitter*, fitur ini digunakan untuk melihat avatar *Twitter*, biografi *Twitter*, dan lainnya. Pada *Twitter* juga terdapat fitur *top trending* yang digunakan untuk memudahkan penggunanya untuk melihat tweet yang paling populer dan paling sering di-tweet oleh para pengguna *Twitter* lainnya (Basri, 2017).

2.7. Sifat Dasar Manusia

Selama tahun 1970-an, psikolog Paul Eckman mengidentifikasi enam emosi dasar yang menurutnya dialami secara universal di semua budaya manusia. Emosi yang dia identifikasi adalah kebahagiaan, kesedihan, jijik, ketakutan, kaget dan kemarahan. Dia kemudian memperluas daftar emosi dasarnya untuk memasukkan hal-hal seperti kesombongan, rasa malu, malu, dan kegembiraan. Psikolog Robert Plutchik mengemukakan "roda emosi" yang bekerja seperti roda warna. Emosi dapat digabungkan untuk membentuk perasaan yang berbeda, seperti halnya warna dapat dicampur untuk menciptakan corak lain. Menurut teori ini, emosi yang lebih mendasar bertindak seperti blok bangunan. Emosi yang lebih kompleks, terkadang bercampur, adalah campuran dari emosi yang lebih mendasar ini. Misalnya, emosi dasar seperti kegembiraan dan kepercayaan dapat digabungkan untuk menciptakan cinta (Cherry, 2017).

2.8. Tweepy

Tweepy adalah *library* dari bahasa pemrograman python untuk berinteraksi dengan akun *Twitter*, dengan menggunakan *Twitter application program interface* (API). Interaksi tersebut menggunakan metode RESTful API. Penggunaan *library* tersebut memerlukan ijin autentifikasi dari pemilik akun twitter dengan mendapatkannya melalui dengan membuat aplikasi twitter di *dev.twitter.com* sehingga memiliki *consumer key*, *consumer secret*, *access token* dan *access token secret*. Dimana nantinya digunakan pada *library* tweepy untuk akses pada akun twitter tersebut. 4 kombinasi angka huruf tersebut dimasukkan pada *OAuth tweepy*. Kemudian, *access token* digunakan untuk memasukkan API key dan *access token secret* API secret key dimana digunakan pada potongan *code* terminal untuk mendapatkan *tweeter bearer*.

2.9. Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

Nama Penulis	Identitas Jurnal	Permasalahan	Metode	Hasil Penelitian	Yang Diadopsi
Nurhayati, Busman, Rayi Pradono Iswara	<p>“Pengembangan Algoritma <i>Unsupervised Learning Technique</i> Pada <i>Big Data Analysis</i> Di Media Sosial Sebagai Media Promosi Online Bagi Masyarakat”</p> <p>Jurnal Teknik Informatika Vol 12 No. 1, April 2019</p>	<p>Machine Learning (ML) yang seringkali digunakan adalah untuk mengklasifikasikan satu permasalahan menjadi beberapa kelompok. Dalam kehidupan sehari-hari, obyek dapat diidentifikasi dengan mudah oleh manusia, namun belum tentu dapat dijelaskan secara spesifik.</p>	<p>Metode yang digunakan adalah <i>supervised learning</i> dan <i>unsupervised learning</i>.</p>	<p>Hasil penelitian ini berupa aplikasi <i>software</i> serta tabel dan grafik yang mampu menggabungkan data yang didapat dari media sosial dan kuesioner secara <i>offline</i>. Hasil analisa pengolahan data tersebut dapat digunakan sebagai <i>DSS (Decision Support System)</i> oleh masyarakat dalam mengambil keputusan pengembangan pemasaran</p>	<p>Yang diadopsi dari jurnal ini adalah cara pengolahan datanya yang diambil dalam dari sosial media.</p>

				produksinya selanjutnya.	
Agnes Theresia Damanik, Masayu Leylia Khodra	<p>“Prediksi Kepribadian <i>Big 5</i> Pengguna <i>Twitter</i> dengan <i>Support Vector Regression</i> “</p> <p>Jurnal Cybermatika Vol. 3 No. 1 Juni 2015</p>	Sampai saat ini belum ada penelitian yang memprediksi kepribadian <i>Big 5</i> untuk pengguna <i>Twitter</i> .	Metode yang digunakan adalah dengan metode <i>Support Vector Regression (SVR)</i>	<p>Hasil penelitian ini berupa model yang menggunakan gabungan fitur perilaku sosial dan linguistik bigram. Model tersebut menghasilkan nilai <i>Mean Absolute Error (MAE)</i> terkecil 0.2739, sedangkan nilai <i>Pearson correlation</i> terbesar 0.8559 didapatkan untuk kepribadian <i>Extraversion</i>.</p>	Yang diadopsi dari jurnal ini adalah Teknik pengambilan datanya dari <i>Twitter</i> .

<p>Roji Ellandi, Erwin Budi Setiawan S.Si., M.T, Dr. Fida Nirmala Nugraha, s. Psi., M. Psi.</p>	<p>“Prediksi kepribadian <i>Big Five</i> dengan <i>Term-Frequency Inverse Document Frequency</i> Menggunakan Metode <i>K-Nearest Neighbor</i> pada <i>Twitter</i> “ <i>e-Proceeding of Engineering: Vol.6, No.2 Agustus 2019 Page 9955</i></p>	<p>Tujuan dari penelitian ini adalah mengetahui kepribadian pengguna <i>Twitter</i> yang telah diperiksa dan ditandai</p>	<p>Metode yang digunakan adalah metode <i>KNN dan TF-IDF</i> yang diterapkan pada analisis <i>Big Five</i></p>	<p>Hasil dari penelitian ini adalah dapat memperoleh nilai akurasi 60,97% dengan pembobotan melalui tahap <i>TF-IDF (Term-Frequency Invert Document Frequency)</i>.</p>	<p>Yang diadopsi dari jurnal ini adalah Teknik pengambilan data dan Teknik data cleansing</p>
<p>Sethunya R Joseph, Hlomani Hlomani, Keletso Letsholo, Freeson Kaniwa, Kutlwano Sedimo</p>	<p>“<i>Natural Language Processing: A Review</i>” <i>International Journal of Research in Engineering and Applied Sciences: VOLUME 6, ISSUE 3 (March 2016) (ISSN 2249-3905)</i></p>	<p>Untuk membantu mengembangkan alat dan teknik yang tepat untuk membuat sistem memahami dan memanipulasi bahasa alami untuk melakukan berbagai tugas</p>	<p>Metode yang digunakan adalah <i>NLP Linguistik</i></p>	<p>Hasil dari penelitian ini adalah mendapatkan teknologi, seperti pencocokan string, pencarian kata kunci, pencarian glosarium di masa</p>	<p>Yang diadopsi dari jurnal ini adalah Teknik dalam NLP (<i>Natural Language Processing</i>)</p>

				<p>lalu, untuk lebih maju dalam mencari teknologi seperti pemeriksa tata bahasa, pencarian konsep dan ekstraksi acara</p>	
<p>Doaa Mohey El-Din</p>	<p><i>“Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis”</i> <i>(IJACSA)</i> <i>International Journal of Advanced Computer Science and Applications,</i> <i>Vol. 7, No. 1, 2016</i></p>	<p>Untuk mengetahui cara mengevaluasi sentimen polaritas dan skor secara otomatis dengan menggunakan kata-kata berat</p>	<p>Metode yang digunakan adalah <i>Bag-of-Words</i></p>	<p>Hasil dari penelitian ini menghasilkan perbandingan antara model <i>Bag-of-Words</i> dan kata-kata tambahan yang telah diusulkan dan diuji dampaknya pada tantangan analisis sentimen dan ketepatan.</p>	<p>Yang diadopsi dari jurnal ini adalah cara penggunaan metode <i>Bag-of-Words</i></p>

<p>Abdur Rasool, Ran Tao, Kamyab Marjan, Tayyab Naveed</p>	<p>“Twitter Sentiment Analysis: A Case Study for Apparel Brands” <i>IOP Conf. Series: Journal of Physics: Conf. Series 1176 (2019) 022015</i> <i>IOP Publishing</i> <i>doi:10.1088/1742-6596/1176/2/022015</i></p>	<p>Perusahaan membutuhkan umpan balik tentang layanan untuk meningkatkan kualitas dan kuantitas. Namun, karena jumlah data yang banyak, sulit untuk mendeteksi opini konsumen.</p>	<p>Metode yang digunakan adalah <i>the machine learning techniques, Naïve Bayes, and lexicon dictionary</i></p>	<p>Hasil dari penelitian ini menemukan opini publik tentang dua merek pakaian internasional teratas dan membandingkan sikap positif dan negatif pengguna umum tentang setiap merek</p>	<p>Yang diadopsi dari jurnal ini adalah konsep kerangka pikir dalam analisis sentiment pada jurnal ini</p>
--	---	--	---	--	--

Pada tabel 2.1 yaitu penelitian terdahulu merupakan upaya peneliti untuk mencari perbandingan dan untuk menemukan inspirasi baru untuk penelitian selanjutnya. Kajian terdahulu juga membantu penelitian dalam memposisikan penelitian serta menunjukkan orsinalitas suatu penelitian. Total jurnal yang dijadikan bahan untuk inspirasi baru dan membantu penelitian adalah 6 jurnal yaitu 3 jurnal Nasional dan 3 jurnal Internasional.