

BAB 2 LANDASAN KEPUSTAKAAN

Pada bagian ini dibahas secara detail tentang kajian pustaka yang berisi informasi penelitian sebelumnya. Selain itu juga dibahas dasar teori yang digunakan dalam penelitian ini secara detail.

2.1 Kajian Pustaka

Berdasarkan topik penelitian ini maka penulis mencari dan mendalami beberapa penelitian yang relevan dengan penelitian ini, yaitu *klasifikasi karakter* pada *twitter* untuk mengetahui karakter calon karyawan menggunakan metode KNN. Penelitian lain dilakukan perbandingan metode untuk melakukan klasifikasi dokumen *corpus* menjadi beberapa bagian klas. Metode yang diujikan antara lain *support vector machine* (SVM), KNN dan *Naïve Bayes*. Hasil pengujian menunjukkan bahwa metode KNN memberikan hasil yang lebih baik dibandingkan dengan SVM dan *Naïve Bayes*, dimana nilai akurasi yang diperoleh sebesar 95.9 %.

Penelitian Qiu *et al* (2012), melakukan studi pada *tweet* dari *twitter* pengguna. Berdasarkan hasil penelitian tersebut didapatkan bahwa *tweet* yang berisi bahasa alami sehari-hari memiliki hubungan yang sangat erat dengan karakter. Dimana *tweet* merupakan ekspresi yang bersifat pribadi dari pengguna. Analisis teks berbahasa Indonesia juga dilakukan untuk mengklasifikasikan *tweet* dari *twitter* menjadi kelompok positif dan kelompok negatif (Putranti dan Winarko, 2014). Klasifikasi dilakukan menggunakan metode *Maximum Entropy* dan SVM. Penelitian sejenis yang menggunakan Bahasa Indonesia juga pernah dilakukan, akan tetapi pengelompokan dilakukan untuk mengetahui karakter pengguna, dimana dibagi menjadi 4 kelompok yaitu Guardia, Artisan, Idealist, dan Rasional (Sarwani dan Mahmudy, 2015). Proses klasifikasi untuk mengetahui karakter kepribadian seseorang menggunakan metode *Naïve Bayes*. Tabel 2.1 menunjukkan penjelasan tentang penelitian terdahulu yang relevan dengan penelitian ini.

Tabel 2. 1. Daftar penelitian terdahulu yang relevan

No	Judul	Objek	Metode	Parameter	Hasil Penelitian
1	<i>Sentiment Analysis of Twitter Data</i>	<i>Twitter</i>	<i>Tree Kernel</i>	Kata dari <i>tweet</i>	Hasil klasifikasi seseorang berdasarkan <i>tweet</i> , kelompok positif, negatif, dan neutral.
2	<i>An Efficient Text Classification using KNN and Naïve Bayesian</i>	<i>Corpus</i> dokumen	SVM, KNN, <i>Naïve Bayesian</i>	Kata dari dokumen <i>corpus</i>	Hasil klasifikasi dokumen dikelompokkan

					menjadi klas <i>science, business, sports, health, education, travel, dan movies.</i>
3	<i>You are what you tweet: Personality expression and perception on Twitter</i>	<i>Twitter</i>	<i>Linguistic Analysis</i>	Kata dari <i>tweet</i>	Terdapat hubungan antara Bahasa alami yang ditulis di <i>tweet</i> dengan karakter seseorang.
4	Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan <i>Maximum Entropy</i> dan <i>Support Vector Machine</i>	<i>Twitter</i>	<i>Maximum Entropy</i> dan <i>Support Vector Machine</i>	Kata dari <i>tweet</i>	Hasil klasifikasi seseorang berdasarkan <i>tweet</i> , kelompok positif dan negative
5	Analisis Twitter untuk Mengetahui Karakter Seseorang Menggunakan Algoritme <i>Naïve Bayes Classifier</i>	<i>Twitter</i>	<i>Naïve Bayes Classifier</i>	Kata dari <i>tweet</i>	Hasil klasifikasi seseorang berdasarkan <i>tweet</i> , kelompok Guardian, Artisan, Idealist, dan Rasional.

2.2 Tes Psikologi

Tes psikologi atau tes kepribadian merupakan suatu metode untuk mengukur ciri-ciri emosi, motivasi, antar-pribadi, dan sikap yang dibedakan dari kemampuan (Anastasi dan Urbina, 1997). Ada berbagai macam tes kepribadian sampai ratusan macamnya. Salah satu metode tes kepribadian yang paling akurat adalah indikator MBTI (Sarwani dan Mahmudy, 2015). Dimana terdapat 16 indikator yang kemudian dikelompokkan menjadi 4 kelompok kepribadian kepribadian besar. 16 indikator tersebut antara lain ISTJ, ESTJ, ISFJ, ESFJ, ESTP, ISTP, ESFP, ISFP, ENTJ, INTJ, ENTP, INTP, ENFJ, INFJ, ENFP, dan INFP. Sedangkan 4 kelompok karakter kepribadian antara lain guardian, artisan, idealist, dan rasional.

Setiap kelompok karater kepribadian memiliki ciri masing-masing yang ditunjukkan pada Tabel 2.2.

Tabel 2. 2. Ciri-ciri karakter kepribadian tiap kelompok

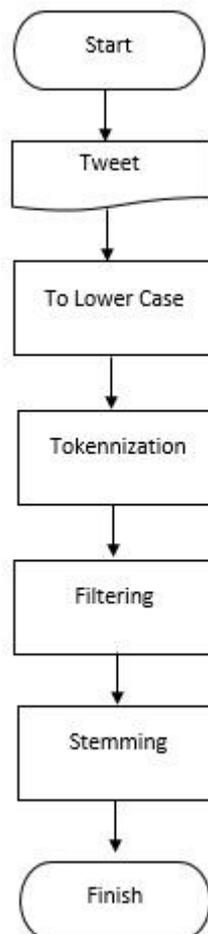
Kelompok	Ciri-ciri	Indikator
Guardian	<ul style="list-style-type: none"> - Taat aturan - Kurang menyenangkan perubahan - Bertanggung jawab - Susah menyesuaikan diri - Berorientasi pada masa kini - Lebih menyenangkan teori - Tidak imajinatif - Tidak senang dengan hal baru - Bagus dalam memimpin - Teliti dan detail 	<ul style="list-style-type: none"> - ESTJ - ISTJ - ESFJ - ISFJ
Artisan	<ul style="list-style-type: none"> - Senang mengambil resiko - Bertindak sebelum berfikir - Mengikuti naluri - Tidak menyukai aturan - Tidak menyukai jadwal yang teratur - Tidak berkomitmen - Efisien - Berani 	<ul style="list-style-type: none"> - ESTP - ISTP - ESFP - ISFP

	<ul style="list-style-type: none"> - Banyak akal - Tidak tergesa-gesa - Suka berimprovisasi 	
Idealist	<ul style="list-style-type: none"> - Mengambil keputusan berdasarkan nilai pribadi - Suka memotivasi orang lain - Jujur - Selalu mencari arti kehidupan - Berintegritas - Baik dalam berkomunikasi - Pintar dalam bernegosiasi - Pengamat yang tajam - Mudah berempati - Karismatik - Mau menerima gagasan baru - Kurang tegas - Rela berkorban - Terlalu emosional 	<ul style="list-style-type: none"> - ENFJ - INFJ - ENFP - INFP
Rasional	<ul style="list-style-type: none"> - Rasa ingin tahu besar - Pintar - Mandiri - Tidak memihak manapun - Melihat sesuatu dari banyak sisi - Melihat masalah secara fleksibel karena melihat dari berbagai hal - Berinovasi secara luwes - Unggul membuat strategi - Penuh percaya diri - Tangkas - Imajinatif - Terlalu rumit dipahami orang 	<ul style="list-style-type: none"> - ENTJ - INTJ - ENTP - INTP

	<p>lain</p> <ul style="list-style-type: none">- Mengabaikan detail tertentu- Skeptis- Menantang aturan jika tidak menyukainya- Kompetitif- Terkadang arogan- Suka menarik diri- Asyik dengan dunianya sendiri	
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

2.3 Text Mining

Text mining merupakan suatu model yang digunakan untuk melakukan pemrosesan teks (Wiess *et al*, 2005). Terdapat beberapa proses dalam *text mining* yaitu *text processing* untuk mendapatkan bobot untuk setiap kata penting yang mengandung informasi. Dalam *text processing* terdapat 2 proses yaitu *text processing* dan *featur selection*. *Text processing* merupakan proses untuk mendapatkan kata baku dan nilai bobot tiap kata. Beberapa tahapan *text processing* antara lain *case folding*, *Tokennization* (Wiess *et al*, 2005). *Case folding* merupakan proses penyamaan class dalam sebuah dokumen. *Tokennization* merupakan proses memecah kalimat. Sedangkan pada proses *feature selection* terdapat beberapa tahapan yaitu *filtering*, *stemming*. Proses pertama adalah *filtering*, yaitu proses menghilangkan kata yang sama dengan kata yang berada pada *stopword*. *Stopword* kata umum yang tidak memiliki informasi penting (Dragut *et al*, 2009), semisal kata “di”, “dia”, “kamu”, “antara”, dan seterusnya. Lalu dilakukan proses *stemming*, yaitu proses merubah kata menjadi kata dasar. Semakin banyak kamus yang digunakan maka semakin presisi hasil yang diberikan. Gambar 2.1. menunjukan alur kerja proses *text processing*.



Gambar 2. 1. Alur proses *text processing*

2.4 K-Nearest Neighbor

K-Nearest Neighbor (KNN) adalah salah satu metode yang populer untuk melakukan klasifikasi teks (Sreemathy dan Balamurungan, 2012). KNN merupakan metode yang menerapkan teknik *lazy learning* dalam proses klasifikasi. Yaitu dengan melakukan proses pembelajaran dari data latih untuk menentukan kelompok k objek. Sehingga dalam menentukan hasil klasifikasi KNN melihat jarak terdekat dari objek dengan masing-masing kelompok. Jarak tersebut diperoleh dari hasil kedekatan antara data masukan dengan data yang berada dalam kelompok berdasarkan nilai sejumlah fitur yang ada. Tahapan untuk menjalankan proses KNN sebagai berikut:

1. Menentukan jumlah tetangga k .
2. Menghitung jarak objek dengan masing-masing data kelompok. Perhitungan jarak menggunakan rumus *euclidian distance* (Sreemathy dan Balamurungan, 2012) yang ditunjukkan pada persamaan 2.1

$$D(x, y) = \sqrt{\sum (x_i - y_i)^2} \quad (2.1)$$

Dimana D adalah jarak, x dan y adalah data latih dan data uji.

3. Lalu didapatkan hasil pengklasifikasian.

Setelah semua data dilakukan pengklasifikasian menggunakan KNN langkah selanjutnya adalah melakukan validasi untuk mencocokkan dengan data pakar dan menghitung nilai akurasi.

2.5 Akurasi

Akurasi adalah proses menghitung nilai keakuratan sistem. Tujuan penghitungan akurasi untuk mengetahui seberapa besar prosentase kebenaran dari hasil klasifikasi karakter kepribadian calon karyawan. Nilai akurasi diperoleh dari perbandingan hasil klasifikasi benar dengan data uji. Persamaan 2.2 menunjukkan formulasi untuk menghitung nilai akurasi.

$$\text{Akurasi} = \frac{\text{Jumlah Data Benar}}{\text{Jumlah Data Uji}} \times 100\% \quad (2.2)$$

