

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 atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

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

BAB III

METODOLOGI PENELITIAN

Pada bab ini akan dijelaskan tahap-tahap yang dilakukan dalam melakukan penelitian. Tahapan penelitian berguna agar pelaksanaan penelitian dapat berjalan dengan baik dan sistematis serta memenuhi tujuan yang diinginkan. Susunan tahapan penelitian dijelaskan pada gambar 3.1 di bawah ini :



Gambar 3.1 Tahapan Metodologi Penelitian

3.1. Perumusan Masalah

Perumusan masalah merupakan tahapan awal dalam metodologi penelitian. Yang dilakukan pada tahapan ini adalah mencari dan mempelajari masalah dalam penelitian, lalu melakukan pencarian solusi untuk memecahkan masalah tersebut. Pada tahapan ini ruang lingkup dan latar belakang dari topik penelitian juga ditentukan. Perumusan masalah dalam penelitian sudah ditentukan yaitu bagaimana menerapkan algoritma *K-Nearest Neighbors* untuk

mengklasifikasikan kepribadian seseorang dengan teori kepribadian *Big Five Personality* melalui status Twitter.

3.2. Studi Pustaka

Tahapan ini merupakan tahapan lanjutan dari tahapan pertama. Setelah perumusan masalah selesai dilakukan pada tahap ini dilakukan pengumpulan data dan informasi yang berhubungan dengan penelitian melalui berbagai media. Studi pustaka bertujuan untuk menemukan teori-teori yang berhubungan dengan penelitian melalui jurnal, media *online*, buku atau penelitian lain berhubungan dengan penelitian yang akan dilakukan.

3.3. Pengumpulan Data

Tahapan ini merupakan tahapan dimana data dikumpulkan yang kemudian akan diproses sebelum diuji dengan algoritma *K-Nearest Neighbor*. Adapun sumber data dan jumlah data adalah sebagai berikut :

1. Data yang digunakan adalah data sekunder yang telah digunakan dalam penelitian Rinaldi Syafrianto tahun 2018.
2. Data merupakan kumpulan *tweet* dari 15 akun Twitter dari orang yang telah mengisi angket tes psikologi berdasarkan teori kepribadian *Big Five Personality* yang telah disebar oleh peneliti sebelumnya dimana hasil dari tes kepribadian tersebut ditentukan oleh psikolog terkait.
3. Jumlah *tweet* yang digunakan adalah sebanyak 1500 *tweet*, dimana setiap akun memiliki masing-masing 100 *tweet*.

3.4. Analisa

Pada tahapan ini gambaran tentang penelitian yang akan dilakukan telah diketahui. Analisa dilakukan terhadap data-data yang telah diperoleh yang akan di proses menggunakan Algoritma *K-Nearest Neighbors*. Adapun rincian analisisnya sebagai berikut:

1. *Text Mining*

Tahapan ini menjelaskan langkah yang dilakukan dalam memproses data yang akan digunakan dalam melakukan pengujian, adapun tahapan praprocess yang dilakukan dalam penelitian ini sebagai berikut:

a. *Case Folding*

Case Folding ini merupakan proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Jadi semua huruf besar akan diubah menjadi huruf kecil dan juga semua karakter selain huruf a sampai z seperti tanda baca akan dihapus.

b. *Tokenizing*

Tahap *tokenizing/parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada proses ini setiap kalimat akan dipotong atau di pisahkan menjadi kata per kata.

c. *Filtering*

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing*. Proses *filtering* dapat menggunakan *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Jadi pada proses ini kata-kata yang merupakan kata sambung atau kata-kata yang tidak deskriptif akan di hapus atau dibuang.

2. *Term Weighting* Dengan *TF-IDF*

Term weighting adalah proses dimana tiap kata yang muncul pada tiap dokumen akan diberi nilai bobot berdasarkan persentase jumlahnya.

3. Tahapan Perhitungan *K-Nearest Neighbor*.

Pada tahapan ini dilakukan perhitungan manual dengan menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan data yang akan digunakan pada tahap pengujian menggunakan *tools*.

3.5. Pengujian

Tahapan ini merupakan tahapan pengujian yang dilakukan setelah melakukan analisa terhadap data yang digunakan. Tahap pengujian diperlukan sebagai ukuran bahwa algoritma yang digunakan dapat digunakan sesuai dengan tujuan. Pada pengujian ini ada beberapa hal yang dijadikan acuan, yaitu :

1. Pengujian algoritma *K-Nearest Neighbor* dengan menggunakan *tools Spyder*
2. Pengujian kinerja sistem dengan menghitung tingkat akurasi.

3.6. Kesimpulan Dan Saran

Tahap ini berisikan tentang kesimpulan dari penelitian yang telah dilakukan serta hasil yang didapatkan. Tahap ini juga berisikan hal yang disimpulkan dan disarankan penulis bagi pembaca untuk melakukan pengembangan terhadap penelitian ini kedepannya.



- 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 atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.