

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Tahapan Penelitian**

##### **3.1.1 Studi Pustaka**

Algoritma KNN, analisis sentimen, *preprocessing* teks, pembobotan kata, dan ide-ide lain yang berkaitan dengan penelitian yang sedang dilakukan dimasukkan oleh penulis dalam bagian tinjauan pustaka ini. Untuk mendukung proses penelitian ini, penulis mencari sumber literatur untuk digunakan sebagai acuan atau referensi. Buku, artikel, dan jurnal yang berhubungan dengan *data sains* merupakan literatur yang digunakan.

##### **3.1.2 Pengumpulan Data**

Sebuah teknik dan proses yang disebut pengumpulan data harus dilakukan untuk mempelajari apa yang perlu dilakukan untuk memasukkan algoritma K-Nearest Neighbor ke dalam proses klasifikasi. Untuk membangun sistem pada titik ini dalam proses pengumpulan data, banyak tugas yang harus diselesaikan, termasuk penggunaan teknik scrapping. Teknik ini melibatkan ekstraksi data dari situs web, database, dan sumber lain untuk mengumpulkan informasi yang relevan yang dibutuhkan untuk menerapkan algoritma KNN. Data yang dikumpulkan kemudian akan diproses.

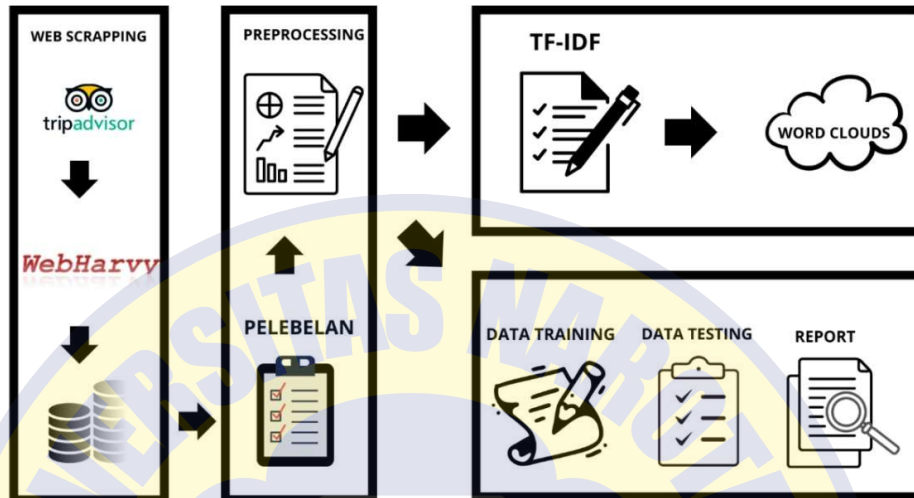
### **3.1.3 Perancangan Sistem**

Pada tahap ini, sistem akan dibuat dengan memulai proses pembuatan test kit untuk mengevaluasi algoritma K-Nearest Neighbor untuk melakukan klasifikasi ulasan dan memastikan akurasi sistem. Proses ini sangat penting untuk menentukan kinerja sistem dan melihat apakah sistem dapat bekerja dengan baik dan memenuhi harapan. Test kit yang dibuat akan digunakan untuk menguji sistem dan mengukur seberapa baik sistem dapat melakukan klasifikasi ulasan dan menentukan akurasi sistem. Hasil yang diperoleh dari tahap ini akan menentukan apakah sistem siap untuk diterapkan pada konteks nyata atau perlu diterapkan perbaikan-perbaikan tertentu.

### **3.1.4 Evaluasi dan Analisis Hasil Penelitian**

Hasil yang dihasilkan dari sistem yang dibuat diperiksa selama tahap pengujian. Ini sebagai jawaban atas rumusan masalah yang telah diterangkan pada bab sebelumnya. Melalui tahap pengujian ini, sistem dievaluasi dan dicari tahu apakah sistem tersebut dapat menjawab masalah yang telah ditentukan. Tahap ini sangat penting karena memungkinkan penentuan keberhasilan atau kegagalan sistem dalam menyelesaikan masalah yang ada. Evaluasi yang dilakukan melalui tahap pengujian ini akan menentukan apakah sistem memenuhi standar yang diinginkan atau perlu diterapkan perbaikan-perbaikan tertentu.

### 3.2 Gambaran Umum Proses Sistem



Gambar 3.1 Gambaran umum sistem

Berdasarkan Gambar 3.1, Penelitian ini dimulai dengan proses pengumpulan data melalui teknik *scrapping* komentar dari *websitesite* Tripadvisor, dengan objek restoran ayam bakar Primarasa di Surabaya. Data yang diperoleh kemudian diolah melalui tahap pelabelan yang dilakukan secara manual. Setelah itu, tahap *preprocessing* dilakukan untuk membersihkan data, termasuk proses *casefolding*, *cleaning*, *tokenization*, dan *removal stopword*. Selanjutnya, setiap kata dari dokumen dihitung dengan menggunakan metode TF-IDF dan divisualisasikan menggunakan *Word Cloud*. Data yang sudah melalui tahap *preprocessing* kemudian digunakan dalam tahap pelatihan dan pengujian dengan menggunakan metode *K-Nearest Neighbors* dan akan diuji menggunakan metode *percentage split* dan *cross validation*.

### 3.2.1. Data Scrapping

Pada penelitian ini digunakan teknik scrapping dengan menggunakan webharvy untuk mengambil data opini pesan dan kesan konsumen Restoran Ayam Bakar Primarasa Surabaya yang ada di Tripadvisor. Data yang digunakan merupakan data dengan format .csv. Data yang diambil merupakan data komentar, dimana orang-orang menulis kritik maupun saran terhadap restoran Ayam Bakar Primarasa Surabaya.

### 3.2.2 Pelabelan

Setelah melakukan proses *scrapping*, data yang didapat dilakukan pelabelan secara manual dengan diberikan label positif dan label negatif. Label positif digunakan komentar yang tidak mengandung kalimat yang merugikan restoran tersebut. Sedangkan label negatif digunakan untuk komentar yang merugikan restoran Ayam Bakar Primarasa Surabaya.

### 3.2.3 Pre-processing

Preprocessing merupakan langkah awal dalam proses klasifikasi berbasis algoritma K-Nearest Neighbor. Data yang akan digunakan dalam proses TF-IDF pada tahap ini merupakan data mentah dari kumpulan teks ulasan, dimana setiap kata akan diberi bobot agar dapat diolah pada tahap selanjutnya. Tahapan pra-pemrosesan adalah sebagai berikut:

Tahap pertama *pre-processing*, disebut *case folding*, berupaya menempatkan kata-kata yang diperoleh ke dalam format yang sama. Istilah diubah

menjadi huruf kecil atau huruf kecil pada saat ini. Tata cara pelaksanaan case folding adalah sebagai berikut:

Mesin akan mengubah semua kalimat menjadi huruf kecil pada saat ini setelah mengumpulkan data. Berikut tahapan *case olding*:

1. Dokumen teks akan dibaca satu baris per satu baris.
2. Semua huruf besar pada dokumen akan diubah menjadi huruf kecil.
3. Saat dokumen disimpan, semua huruf akan dikonversi menjadi huruf kecil.

Saat mengumpulkan data mentah dari Tripadvisor dalam bentuk ulasan, data tersebut dapat diterapkan dalam langkah-langkah yang disebutkan sebelumnya dalam situasi klasifikasi ini. Informasi ini akan digunakan sebagai sampel dari tahap pra-pemrosesan dari proses pertama.

Untuk menghapus karakter seperti tanda baca, tahap sistem yang dikenal sebagai tokenisasi akan membagi teks menjadi fragmen-fragmen kecil menjadi satu kata yang dikenal sebagai token. Proses penerapan tokenisasi adalah sebagai berikut:

1. Pemecahan Teks: Dalam tahap ini, teks dibagi menjadi fragmen-fragmen kecil yang mewakili satu entitas semantik atau kata.
2. Identifikasi Token: Dalam tahap ini, setiap fragmen teks dikenali sebagai satu token yang terpisah dan mewakili satu entitas semantik.
3. Filtrasi Token: Dalam tahap ini, beberapa token mungkin harus dihapus atau difilter karena tidak memiliki makna atau informasi yang relevan, seperti tanda baca, angka, dan simbol.

4. Penyimpanan Token: Dalam tahap ini, token-token yang sudah difilter disimpan untuk digunakan dalam proses selanjutnya.

Setiap ulasan diproses melalui tahapan menghapus kata-kata yang tidak signifikan setelah melalui tahap memisahkan kata-kata dalam sebuah teks menjadi token. Proses ini dikenal sebagai stop word, dimana kata-kata yang tidak memiliki makna yang penting akan dihilangkan dari hasil akhir. Ini bertujuan untuk memperjelas dan menyederhanakan teks, memfokuskan pada informasi yang penting saja. Berikut tahapan dari *stopword removal* :

1. Pemisahan token: Tahap ini melibatkan memisahkan kata-kata dalam teks menjadi token individu.
2. Perbandingan dengan kamus stopword: Setelah token dibentuk, maka setiap token akan dibandingkan dengan kamus stopword.
3. Penghapusan token: Token yang teridentifikasi sebagai kata yang tercantum dalam kamus stopword akan dihapus.
4. Penyimpanan token yang tersisa: Token yang tidak tercantum dalam kamus stopword akan disimpan dan digunakan sebagai bagian dari teks yang akan dianalisis.
5. Proses selesai: Proses stopword removal selesai setelah semua token dalam teks selesai dibandingkan dan diolah sesuai dengan tahap-tahap sebelumnya.

### 3.2.4 Pembobotan TF-IDF

Tahap pembobotan menggunakan tf-idf muncul setelah *pre-processing*. Pada tahap ini akan menentukan bobot komentar yang dihitung dari setiap kata dan dikalikan dengan idf. Berikut adalah implementasi dari metode Tf-Idf :

3.2.4.1 Menentukan nilai Tf: Langkah ini melibatkan penentuan frekuensi tiap kata dalam sebuah dokumen.

3.2.4.2 Menghitung bobot Tf x Idf: Setelah nilai Tf ditentukan, maka bobot Tf akan dikalikan dengan nilai Idf (Inverse Document Frequency) untuk menentukan bobot akhir dari setiap kata.

3.2.4.3 Mengulangi proses untuk setiap data: Langkah-langkah 1 dan 2 akan diulang untuk setiap data dalam corpus yang akan dianalisis. Proses ini akan memberikan bobot akhir untuk setiap kata dalam setiap dokumen dalam corpus.

### 3.2.5 Data Training

Setelah melalui proses *pre-processing*, data selanjutnya akan diolah menggunakan algoritma K nearest neighbor. Algoritma ini akan menggunakan seluruh data dari 1124 komentar sebagai data latih untuk membentuk model.

Dalam proses pelatihan, data akan dianalisis dan diklasifikasikan berdasarkan fitur-fitur yang diambil dari setiap komentar. Algoritma akan mencari kemiripan antar data dan menemukan komentar terdekat untuk setiap data baru.

Dengan menggunakan seluruh data sebagai data latih, diharapkan model yang dihasilkan akan memiliki akurasi yang baik dan mampu melakukan klasifikasi dengan benar pada data baru.

Dengan demikian, pelatihan menggunakan algoritma K nearest neighbor akan membantu membuat model yang handal dan dapat menentukan klasifikasi suatu komentar dengan baik.

### **3.2.6 Data *Testing***

Setelah melalui tahap pelatihan, data selanjutnya akan dilakukan pengujian untuk menentukan akurasi dari model yang digunakan. Proses ini penting dilakukan untuk memastikan bahwa model yang dihasilkan dapat melakukan klasifikasi dengan benar. Untuk memastikan hasil yang akurat, data pengujian akan dilakukan dengan menggunakan dua metode yaitu percentage split dan cross validation.

Metode percentage split akan membagi data menjadi dua bagian yaitu data latih dan data pengujian dengan mempertimbangkan persentase tertentu.

Sedangkan cross validation akan membagi data menjadi beberapa bagian dan melakukan pelatihan dan pengujian secara bergantian pada setiap bagian. Dengan metode ini, diharapkan dapat memastikan bahwa model yang dihasilkan tidak hanya mampu melakukan klasifikasi pada data tertentu saja, tetapi juga mampu melakukan klasifikasi pada semua data.

Hasil dari kedua metode tersebut akan dibandingkan dan digunakan untuk menentukan hasil akurasi terbaik. Dengan melakukan pengujian ini, diharapkan model yang dihasilkan mampu menentukan klasifikasi dengan akurasi yang tinggi.