

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Aplikasi *streaming online* merupakan tempat penyiaran yang menggunakan jaringan internet. Dengan beberapa konten yang sering dimunculkan seperti *gaming, music, coding*, dan lain-lain. Kolom komentar pada aplikasi *streaming online* disediakan agar para penonton dapat berinteraksi atau memberikan ekspresi dengan *streamers*. Fakta tersebut didasari oleh pengamatan mengenai *bullying* (perundungan) terhadap *live streaming gamers* wanita menurut UU ITE 2008 yang dilakukan oleh Zendika Dayongki S, dkk. Dengan menggunakan data dari 5 *gamer* wanita pada sosial media yang melakukan *live streaming game* seperti defiarkaw, RINA ClauPaw, Audrey FF, Chole Berry Gaming, Cici Pertiwi. Akun-akun yang disebutkan dipilih berdasarkan minimal *subscriber* 100.000, dan *streamer* mengunggah konten paling terbaru minimal di bulan september 2019. Menghasilkan sebanyak 80% perundungan terjadi kepada *gamers* wanita saat melakukan *live streaming*. Terakhir, perundungan secara verbal meliputi menggoda, memanggil nama secara tidak pantas, mengejek, menghina, dan mengancam [1]. Maka dari itu, kata hinaan termasuk kedalam perundungan secara verbal.

Ada beberapa penelitian yang sudah membahas mengenai pendeteksian terhadap komentar di media sosial. Penelitian pertama yang dilakukan oleh Antonius Rachmat, Yuan Lukito untuk deteksi komentar *spam* bahasa Indonesia pada Instagram menggunakan algoritma *Naive Bayes* mendapatkan nilai akurasi sebesar 77,25%, dengan data yang diolah sebanyak 17.007 [2]. Selanjutnya yang dilakukan oleh Naomi Chatrina Siregar, dkk untuk mengetahui apakah komentar tersebut bersifat negatif atau positif. Metode yang digunakan merupakan *Naive Bayes Classifier*, hasil akurasi sebesar 68% [3]. Berdasarkan hal tersebut

diperlukan penelitian lanjut untuk meningkatkan akurasi dan dapat melakukan proses terhadap jumlah data yang kecil.

Terdapat sebuah cara untuk melakukan peningkatan akurasi dengan menggunakan *Local Mean K-Nearest Neighbors* (LMKNN) dan *Distance Weight K-Nearest Neighbors* (DWKNN). Penelitian ini dilakukan oleh Khairul Umam Syaliman, dkk yang menguji pada 4 dataset [4] menyimpulkan bahwa metode yang diusulkan lebih baik daripada *K-Nearest Neighbor* asli dengan kenaikan sebesar 2,492%. Penelitian selanjutnya yang melakukan klasifikasi komentar spam pada instagram berbasis REST Services menggunakan DWKNN, menghasilkan akurasi tertinggi sebesar 90% untuk  $k = 2$  [5]. Terakhir, *Local Mean-Based K-Nearest Neighbor* terbukti dapat memperbaiki kinerja metode *K-Nearest Neighbor*, terutama untuk jumlah data yang kecil [6].

Berdasarkan paparan yang dipaparkan sebelumnya, pada penelitian ini akan mencoba menganalisis dan mengolah suatu teks apakah teks tersebut merupakan sebuah kata hinaan atau bukan.

Penelitian ini dilakukan untuk mendeteksi komentar dan menguji performa algoritma *Local Mean-Based K-Nearest Neighbor* untuk mendeteksi komentar yang mengandung unsur kata hinaan.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah Seberapa besar akurasi dari metode *Local Mean-Based K-Nearest Neighbor* untuk mendeteksi komentar yang mengandung unsur kata hinaan?

## **1.3 Maksud dan Tujuan**

Maksud dari penelitian yang akan dilakukan adalah membangun aplikasi pengolahan teks untuk dapat mendeteksi kata hinaan dengan menggunakan metode *Local Mean-Based K-Nearest Neighbor*. Memiliki tujuan untuk mampu mengenali komentar yang memiliki unsur kata hinaan dan mengetahui besaran akurasi dari *Local Mean-Based K-Nearest Neighbor*.

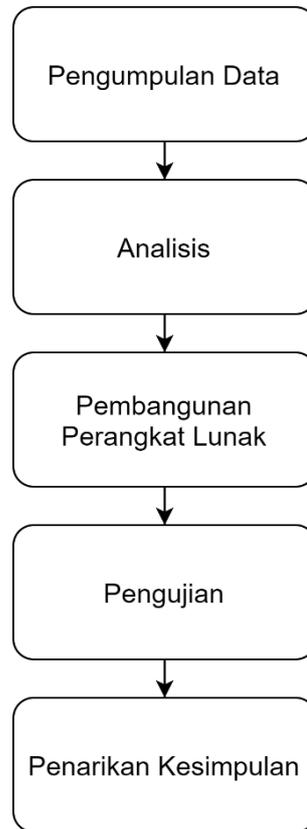
#### 1.4 Batasan Masalah

Batasan masalah dimaksudkan untuk membatasi ruang lingkup pekerjaan yang dilaksanakan. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data komentar diambil dari aplikasi Youtube berdasarkan isi konten *video* berupa *game* dengan channel The lazy Monday, seputar dota, Tony Astan, dan FansV Gaming.
2. Kata hinaan yang akan ditangani hanya satu kata.
3. Mengambil data menggunakan metode *Web Scraping* secara otomatis dengan membuat program di aplikasi Google Apps Script.
4. Kelas klasifikasi kata hinaan hanya berupa hinaan, dan bukan hinaan.
5. Pada tahap ekstraksi fitur akan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).
6. Komentar yang hanya memiliki satu karakter akan dianggap sebagai bukan hinaan.
7. Yang dideteksi hanya tulisan saja atau teks tidak dengan *video* nya.
8. Evaluasi model menggunakan K-Fold Cross Validation dengan  $K = 5$  [7].
9. Metode analisis perancangan yang digunakan adalah analisis dan perancangan perangkat lunak berorientasi objek.

#### 1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam aplikasi ini adalah metode deskriptif. Metode deskriptif akan memberikan gambaran yang jelas tentang hal-hal yang diperlukan. Dalam penelitian ini telah terbagi menjadi 5 tahapan, Pengumpulan Data, Analisis, Pembangunan Perangkat Lunak, Pengujian, Penarikan Kesimpulan. Langkah-langkah yang harus dilakukan selama penelitian ini dapat dilihat pada gambar 1.1.



Gambar 1.1 Alur penelitian yang dilakukan

### 1.5.1 Pengumpulan Data

Pada tahapan ini akan melakukan pengumpulan data dengan cara mengambil data yang terdapat pada video di aplikasi Youtube, mengacu pada beberapa *channel game*. Pertama, The Lazy Monday dengan video yang berjudul “Gw CINTA Mobile Legends bang bang” diunggah tanggal 14 Mei 2021. Kedua, seputar dota dengan video yang berjudul “windah basudara psarkas handal” diunggah tanggal 8 Juni 2021. Ketiga, Tony Astan dengan video yang berjudul “PUBG DAN FREE FIRE GAME SAMPAH” diunggah tanggal 3 Juli 2020. Keempat, FansV Gaming dengan video yang berjudul “Cheat Point Blank Indonesia Auto Headshot,Teleport Flat,Esp Hack,Aim Bullet || Terbaru Mei-Juni 2020” diunggah tanggal 23 Mei 2020. Setelah itu, diambil data tersebut dari API *comment* Youtube yang mengacu pada ID video dengan menggunakan Apps Script berformat .gs, dan disimpan pada suatu file berformat csv/xlsx.

### 1.5.2 Analisis

Pada tahapan ini akan dilakukan analisis masalah, analisis sistem, analisis data mentah dari hasil pengumpulan data, analisis *preprocessing*, analisis ekstraksi fitur TF-IDF, dan analisis terhadap metode *Local Mean-Based K-Nearest Neighbor* dalam pendeteksian komentar yang mengandung kata hinaan.

### 1.5.3 Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang akan digunakan dalam penelitian ini adalah metode *Linear Sequential* yang meliputi proses-proses sebagai berikut:

1. *Analysis*

Menganalisis hal-hal yang berkaitan dengan apa yang dibutuhkan dalam pembuatan dan pengembangan perangkat lunak.

2. *Design*

Melanjutkan dari tahapan analisis untuk dapat dirancang menggunakan *Unified Modelling Language (UML)* yang termasuk metode pemrograman berorientasi objek. Juga perancangan antarmuka yang akan menjadi tampilan pada sistem.

3. *Coding*

Tahapan ini memasuki pengkodean perangkat lunak. Pengkodean menggunakan bahasa tertentu yang mengacu pada perancangan sistem sebelumnya.

4. *Testing*

Melakukan pengecekan terhadap perangkat lunak yang sudah dibangun. Pengecekan mengacu kepada perancangan sistem, agar perangkat lunak sesuai dengan rancangan yang sudah ditentukan sebelumnya dan dapat digunakan dengan baik tanpa ada kesalahan.

#### **1.5.4 Pengujian**

Tahapan penelitian ini akan melakukan pengujian terhadap aplikasi yang telah dibangun. Agar mengetahui besaran akurasi dari pengklasifikasian kata hinaan yang terdapat pada suatu teks dengan menggunakan *K-Fold Cross Validation*.

#### **1.5.5 Penarikan Kesimpulan**

Tahapan terakhir penelitian ini akan melakukan penarikan kesimpulan dari hasil pengujian yang telah diketahui. Menyimpulkan hasil akurasi dari klasifikasi kata hinaan menggunakan metode *Local Mean-Based K-Nearest Neighbor*.

#### **1.6 Sistematika Penulisan**

Sistematika penulisan laporan disusun agar dapat memberikan gambaran secara umum mengenai apa yang dilakukan terhadap penelitian ini, untuk mendeteksi kata hinaan menggunakan metode *Local Mean-Based K-Nearest Neighbor*. Sistematika penulisan penelitian sebagai berikut.

### **BAB 1 PENDAHULUAN**

Bab ini membahas mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan.

### **BAB 2 TINJAUAN PUSTAKA**

Bab ini membahas mengenai buku, jurnal dan sumber yang digunakan untuk menganalisis masalah dan teori yang dipakai dalam mengolah data penelitian yaitu teori mengenai metode *K-Nearest Neighbor*, *Local Mean-Based K-Nearest Neighbor*, teori mengenai pembangunan sistem, teori *Text Preprocessing*, TF-IDF, kata hinaan.

### **BAB 3 ANALISIS DAN PERANCANGAN SISTEM**

Bab ini menguraikan penjelasan analisis kebutuhan yang akan dibangun mulai dari analisis data masukan, analisis metode, dan analisis kebutuhan

perangkat lunak. Selain dari analisis kebutuhan, bab ini akan membahas tentang perancangan sistem.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bab ini akan menampilkan implementasi hasil analisis dan perancangan yang telah dilakukan sebelumnya serta pengujian mengenai akurasi dari metode *Local Mean-Based K-Nearest Neighbor*.

#### **BAB 5 KESIMPULAN DAN PEMBAHASAN**

Bab ini berisi tentang kesimpulan dari hasil pembangunan sistem yang telah diimplementasikan disertai juga dengan saran yang diharapkan dapat berguna untuk tahap penelitian lanjutan.

