

# **BAB 1**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Bahasa kasar adalah ungkapan yang mengandung frasa atau kata-kata kasar dan dikomunikasikan secara lisan atau tertulis kepada lawan bicara (individu atau kelompok), yang berdampak mempercepat terjadinya konflik sosial jika disertai dengan ujaran kebencian (Ibrohim & Budi, 2019). Sedangkan menurut Tuarob & Mitranont (2017), Bahasa kasar adalah ekspresi yang berisi kata atau frasa kasar / kotor, baik lisan maupun tulisan.

Bahasa Kasar sangat mudah disebarkan melalui berbagai media. Bahasa kasar disebarkan dalam berbagai bentuk mulai dari teks hingga suara. Bahasa kasar melalui lisan atau suara tidak pantas didengarkan oleh siapapun dan dimanapun karena Bahasa kasar tidak untuk ditiru dan dipraktikan. Bahasa kasar lisan atau suara dapat dengan mudah diterima dan didengarkan. Anak kecil dapat dengan mudah menerima suara tersebut dan dipraktikan. Hal tersebut sangat berbahaya jika tidak dicegah. Berdasarkan penelitian sebelumnya yang telah dilakukan menjelaskan bahwa konten yang diunggah ke media sosial youtube dapat berisi bahasa kasar (Sukmono, 2020; David, Sondakh, & Harilama, 2017; Rochmatin, 2019).

Solusi berdasarkan permasalahan diatas adalah dibuatnya sistem yang dapat melakukan pemeringkatan konten suara berdasarkan bahasa kasar sesuai dengan tingkatan usia. Sistem pemeringkat konten suara dibuat sesederhana mungkin supaya mempermudah dalam penggunaannya. Format konten suara yang digunakan adalah mp3 dan wav. Untuk membuat sistem ini menggunakan metode pendekatan NLP (Natural Language Processing) karena cocok untuk pengolahan bahasa.

Penerapan NLP (Natural Language Processing) pada kegiatan sehari-hari sudah sangat banyak mulai dari chatbot hingga klasifikasi teks spam pada gmail. Klasifikasi teks digunakan untuk melakukan klasifikasi terkait informasi yang terdapat pada teks. Hal itu yang membuat NLP dapat diterapkan dalam sistem pemeringkat konten suara berdasarkan bahasa kasar.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang disebutkan sebelumnya, masalah utama yang akan dibahas dalam tugas akhir ini yaitu bagaimana membuat sistem yang membantu dalam pemeringkatan konten berdasarkan bahasa kasar dengan melakukan transkripsi dan deteksi Bahasa kasar pada konten suara.

## 1.3. Tujuan Penelitian

Tujuan dari penulisan tugas akhir ini untuk membuat sistem pemeringkatan konten suara berdasarkan umur dengan melakukan transkripsi dan deteksi bahasa kasar berbasis website dengan framework django.

Manfaat dari penulisan tugas akhir ini adalah melakukan pemeringkatan konten suara berdasarkan bahasa kasar yang bertujuan untuk membantu mengklasifikasi konten suara sesuai dengan tingkatan usia.

## 1.4. Batasan Masalah

Bedasarkan masalah maka dalam penelitian tugas akhir ini, peneliti membatasi masalah sebagai berikut :

1. Data penelitian yang digunakan berbahasa Indonesia.
2. Data penelitian yang digunakan terdiri dari 2 jenis yaitu mengandung bahasa kasar dan tidak mengandung bahasa kasar.
3. Bahasa pemrograman yang digunakan adalah Python.
4. Hanya mendeteksi dari transkrip suara.
5. Tahapan pendeteksian kata-kata masih dalam fase morfologi.

## 1.5. State of The Art

| Judul Jurnal   | Pembahasan  |
|--|---|
| <b>Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python</b> | <b><u>Hasil Penelitian:</u></b><br>Jurnal ini membahas tentang membangun sebuah sistem pencatatan teks hasil wawancara dengan menggunakan <i>Google</i> |

|  |  |
|--|--|
| <p><b>Peneliti</b><br/>I Komang Setia Buana</p> <p><b>Lokasi</b><br/>Politeknik Siber Dan Sandi Negara</p> <p><b>Tahun</b><br/>2020</p> <p><b>Nama Jurnal</b><br/>Jurnal Sistem Dan Informatika (JSI)</p>  | <p><i>Speech</i> API berbasis desktop menggunakan Bahasa pemrograman Python yang mampu mencatat informasi dalam audio hasil wawancara dalam bentuk teks secara otomatis.</p> <p><b><u>Alasan Menjadi Tinjauan Penelitian:</u></b><br/>Jurnal ini menjadi acuan dalam penggunaan <i>Google Speech</i> API pada tahap transkripsi teks.</p>  |
| <p><b>Implementasi klasifikasi usia film televisi pada penguatan pendidikan karakter berbasis partisipasi masyarakat</b></p> <p><b>Peneliti</b><br/>Fetrimen Fetrimen, Rhini Fatmasari</p> <p><b>Lokasi</b><br/>Universitas Muhammadiyah Prof. DR. HAMKA, Universitas Terbuka</p> <p><b>Tahun</b><br/>2021</p> <p><b>Nama Jurnal</b><br/>Jurnal Konseling dan Pendidikan</p> | <p><b><u>Hasil Penelitian:</u></b><br/>Jurnal ini membahas tentang klasifikasi usia penonton pada tayangan film televisi menggunakan metode kuantitatif deskriptif dimana penentuan klasifikasi didasarkan pada pengamatan dan penilaian tema, gambar, adegan, suara dan teks terjemahan pada film. Hasil akhir yang dilihat dari klasifikasi usia semua umur, tiga belas, tujuh belas dan dua puluh satu tahun merupakan penggolongan usia yang paling baik.</p> <p><b><u>Alasan Menjadi Tinjauan Penelitian:</u></b><br/>Jurnal ini memiliki kasus yang sama yaitu penelitian terhadap klasifikasi usia serta dengan melihat hasil pemberian klasifikasi usia pada film televisi yang akan dipertunjukkan pada khalayak umum bagian dari upaya melindungi masyarakat dari pengaruh negatif, menjadikan</p> |

|  |   |
|--|---|
|  | langkah awal dalam pengembangan sistem ini.   |
| <p><b>Pendeteksian Bahasa Kasar (Abusive Language) dan Ujaran Kebencian (Hate Speech) Dari Komentar di Jejaring Sosial</b></p> <p><b>Peneliti</b><br/>Luh Putu Ary Sri Tjahyanti</p> <p><b>Lokasi</b><br/>Universitas Panji Sakti</p> <p><b>Tahun</b><br/>2020</p> <p><b>Nama Jurnal</b><br/>DAIWI WIDYA Jurnal Pendidikan</p> | <p><b><u>Hasil Penelitian:</u></b></p> <p>Jurnal ini membahas tentang deteksi Bahasa kasar dan ujaran kebencian pada komentar jejaring sosial menggunakan berbagai API dan library pada python untuk melakukan klasifikasi teks multi-label dimana hasil akhir yang dilihat dari klasifikasi tiga label (bahasa tidak kasar, bahasa kasar tapi bukan ujaran kebencian, dan ujaran Kebencian) lebih sulit daripada hanya mengklasifikasikan apakah teks itu bukan bahasa kasar atau bahasa kasar.</p> <p><b><u>Alasan Menjadi Tinjauan Penelitian:</u></b></p> <p>Jurnal ini memiliki kasus yang sama yaitu penelitian terhadap deteksi bahasa kasar serta dengan melihat hasil pendeteksian bahasa kasar pada komentar sosial media yang menggunakan klasifikasi menjadi tiga label, menjadikan langkah awal dalam proses pendeteksian bahasa kasar pada pengembangan sistem ini.</p> |

## 1.6. Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang sistem yang akan dibangun. Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut :

**BAB I PENDAHULUAN**

Bab ini berisi penjelasan secara singkat mengenai latar belakang, perumusan masalah, tujuan masalah, batasan masalah, *state of the art*, dan sistematika penulisan.

**BAB II LANDASAN TEORI**

Bab ini berisi landasan teori yang terdiri dari teori dasar yang dibutuhkan untuk membangun tugas akhir.

**BAB III ANALISIS DAN PERANCANGAN**

Bab ini memuat tentang analisis dan perancangan mengenai perancangan sistem, perancangan fungsi, perancangan basis data, dan perancangan tampilan.

**BAB IV IMPLEMENTASI DAN PENGUJIAN**

Bab ini membahas mengenai implementasi kode, implementasi program, dan pengujian sistem

**BAB V KESIMPULAN DAN SARAN**

Bab ini memuat kesimpulan dan saran dari pembahasan dan pengujian pada sistem.