

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Industri televisi dan radio di Indonesia merupakan bagian penting dalam industri Indonesia dikarenakan memiliki bagian besar dalam penyebaran informasi [1]. Radio sendiri menjadi salah satu media favorit bagi masyarakat Indonesia untuk hiburan dan informasi [2]. Menurut Perkembangan industri televisi dan radio di seluruh Indonesia membuat tingkat kreativitas dan persaingan antar lembaga penyiaran semakin tinggi, sehingga program siaran menjadi tolok ukur keberhasilan meraih keuntungan. Perkembangan persaingan dalam siaran televisi dan radio memunculkan program siaran yang tidak sesuai dengan nilai-nilai yang dianut dan diyakini oleh masyarakat[3].

Siaran televisi sering mengabaikan kualitas untuk mengikuti selera pasar [4]. Salah satu contoh dari mengabaikan kualitas siaran adalah dengan memutar lagu dengan lirik seksual. Lirik musik yang mengandung kata kasar, seksual, serta pemakaian narkoba dapat memberikan dampak negatif bagi kaum remaja [5, 6]. Oleh karena itu KPI membuat aturan Pedoman Perilaku Penyiaran (P3) dan Standar Program Siaran (SPS) bagi lembaga penyiaran untuk melarang *ekspresi seksual* [7]. Peraturan ini mengenai program siaran dilarang berisi lagu dan/atau video klip yang menampilkan judul dan/atau lirik bermuatan seks, cabul, dan/atau mengesankan aktivitas seks. Kebijakan ini membuat siaran lagu dalam radio maupun televisi untuk memilih lagu dengan judul dan lirik tanpa kata seksual dan aman untuk dikonsumsi oleh semua umur[3].

Selain Indonesia, *Recording Industry Association of America (RIAA)* juga telah berupaya untuk memberi tahu orang tua tentang potensi risiko lirik tertentu dan potensi dampaknya terhadap generasi muda [8]. Upaya ini melahirkan *parental advisory label* pada tahun 1990 untuk album dengan lirik yang mengandung seks dan kekerasan [9]. Namun pada kenyataannya, label ini masih secara sukarela dan tidak semua album lagu memiliki *parental advisory label* [10].

Namun dalam melakukan sensor lagu masih terdapat isu seperti banyaknya lagu yang perlu disensor sehingga memerlukan waktu yang cukup lama dalam proses sensor lagu. Siaran dapat memutar lagu sekitar 18–20 lagu dalam 1 jam siaran. Dan dalam satu siaran, dapat memiliki waktu sekitar 18 hingga 24 jam waktu

siaran [11]. Hal ini dapat menimbulkan masalah bagi televisi dan radio karena perlu untuk mengarahkan waktu dan tenaga untuk menentukan lagu yang layak untuk disiarkan ke dalam program siaran. Masalah lainnya dari proses sensor lirik lagu adalah lagu dapat memiliki label yang berbeda dari lembaga yang memberi label. Hal ini dapat menimbulkan masalah dalam lagu yang dianggap memiliki lirik seksual dapat lolos dan ditayangkan oleh program siaran [12].

Terdapat dua metode utama untuk membuat alat klasifikasi konten seksual dalam lirik lagu. Pertama adalah dengan membandingkan lirik lagu dengan kamus bahasa seksual. Metode ini memiliki kekurangan dimana diperlukan kamus bahasa seksual yang perlu untuk diperbarui secara berkala untuk membandingkan lirik lagu dengan kamus bahasa seksual. Diperlukan juga pakar dalam menentukan kata yang masuk ke dalam kamus bahasa seksual agar tidak terjadi kesalahan dalam klasifikasi konten seksual [13]. Cara lainnya adalah dengan menggunakan *machine learning*. *Machine learning* dapat digunakan dalam proses klasifikasi konten seksual dalam lirik lagu dengan melakukan *analisis sentimen* terhadap lirik lagu dan menentukan lirik lagu mengandung kata seksual atau tidak [14].

Salah satu model *machine learning* yang dapat digunakan adalah *Extra Trees Classifier (ETC)*. ETC bekerja dengan melalui agregasi *output* yang diterima dari berbagai *decision tree* yang tidak berkorelasi yang disusun untuk menghasilkan *segregasi* [15]. ETC digunakan dikarenakan memiliki waktu pelatihan yang lebih singkat dan hasil akurasi yang lebih baik untuk teks besar seperti lirik lagu [16].

Machine learning sendiri masih memiliki kelemahan dalam menangkap kiasan dalam lirik lagu. Hal ini membuat *deep learning* dibutuhkan dalam mengklasifikasi konten seksual. Model *deep learning* yang dapat digunakan dalam mengklasifikasi konten seksual adalah *Long Short-Term Memory (LSTM)*. LSTM merupakan model *deep learning* yang digunakan untuk menjalankan data *sequential* dimana memiliki 3 *gate* yaitu *input gate*, *output gate*, dan *forget gate* [17]. LSTM digunakan dikarenakan memiliki kemampuan menangkap bahasa kiasan dan metafora dalam lagu [14].

Kedua model ini dapat digabungkan dengan *voting classifier* dengan menggunakan *soft voting*. *Soft voting* digunakan untuk meningkatkan akurasi model serta untuk menggabungkan kelebihan dari model yang digunakan [18]. Salah satu contoh *voting classifier* adalah *soft voting* dimana menggabungkan hasil rata-rata dari semua model untuk mencari akurasi model. Contoh dari *soft voting* adalah penelitian yang telah dilakukan oleh Xiaoyuan Chen dengan *Long Short-Term Memory (LSTM)* sebagai *deep learning* dan *Extra Trees Classifier (ETC)* sebagai

machine learning untuk mendeteksi lirik lagu eksplisit dengan akurasi 96,92% [14].

Penelitian lainnya mengenai *voting classifier* adalah dalam penelitian *COVID-19 tweet sentiment classification* menunjukkan bahwa *soft voting* antara ETC dan CNN untuk *Word2Vec* dengan fitur *TextBlob* memiliki nilai *accuracy* yang lebih tinggi sebesar 0,97 dibandingkan fitur *VADER* sebesar 0,95 [19]. Terdapat pula penelitian mengenai *offensive language detection* dimana membandingkan antara beberapa *machine learning* dengan *soft voting* dimana *soft voting* dengan 5 *machine learning* dengan akurasi terbaik memiliki tingkat akurasi terbaik di 95% dan *deep learning* dengan *soft voting* dimana *soft voting* dengan 3 *deep learning* dengan akurasi terbaik memiliki akurasi 95% [20].

Berdasarkan hasil penelitian sebelumnya, penelitian akan berfokus dalam mengklasifikasi konten seksual dalam lirik lagu. Penelitian ini dibuat untuk membantu proses pemilihan lagu yang tidak mengandung kata seksual. Penelitian ini dapat digunakan bagi masyarakat khususnya di dalam industri penyiaran dimana dapat menjadi alat bantu untuk melakukan filter lagu yang mengandung kata seksual. Dengan hasil prediksi, masyarakat dapat memilih lagu yang tidak mengandung kata seksual. Dengan menggunakan algoritma *machine learning Extra Tree Classification* dan algoritma *deep learning LSTM* sebagai algoritma pengujian dan hasil dari kedua algoritma ini akan digabungkan dengan metode *Voting Ensemble*. Penelitian ini diharapkan dapat membantu masyarakat dalam memilih lagu tanpa kata seksual yang dapat dinikmati oleh seluruh kalangan.

1.2 Rumusan Masalah

Berdasarkan latar belakang, dapat diidentifikasi rumusan masalah yang merupakan.

1. Bagaimana cara klasifikasi konten seksual dalam lirik lagu dalam suatu lagu dengan menggunakan *soft voting*?
2. Bagaimana hasil performa dari penerapan algoritma *soft voting*, untuk mengklasifikasi konten seksual dalam lagu tersebut, diukur dari *accuracy*, *precision*, *recall* dan *f1-score* ?

1.3 Batasan Permasalahan

Berdasarkan rumusan masalah, dapat diidentifikasi batasan masalah yang merupakan.

1. Penelitian ini menggunakan data lagu-lagu dari data penelitian Xiaoyuan Chen [14]
2. Data lagu yang digunakan adalah lagu dengan lirik berbahasa Inggris, berjumlah 5 ribu.
3. Model *soft voting* yang digunakan adalah model *soft voting* gabungan antara model ETC dan model LSTM.

1.4 Tujuan Penelitian

Berdasarkan batasan masalah, dapat diidentifikasi tujuan penelitian yang merupakan.

1. Melakukan klasifikasi konten seksual dalam lirik lagu dalam suatu lagu dengan menggunakan *soft voting*
2. Mengukur hasil performa dari penerapan algoritma *soft voting*, untuk analisis mengklasifikasi konten seksual dalam lagu tersebut, diukur dari *accuracy*, *precision*, *recall* dan *f1-score*

1.5 Manfaat Penelitian

Manfaat dalam penelitian yang dapat diperoleh adalah sebagai berikut.

1. Bagi peneliti, penelitian ini bermanfaat untuk menambah dan menerapkan ilmu mengenai mengklasifikasi konten seksual pada lirik lagu untuk memprediksi kata seksual dalam suatu lagu, dengan basis algoritma *voting classifier*.
2. Bagi pengguna, prediksi kata seksual dalam lirik lagu berdasarkan mengklasifikasi konten seksual dalam lirik lagu akan membantu pengguna dalam pemilihan lagu dalam televisi dan radio, sehingga lagu yang dipilih dapat digunakan dalam program siaran dengan mengikuti undang-undang dari KPI.

1.6 Sistematika Penulisan

- Bab 1 PENDAHULUAN
Bab pertama membahas mengenai latar belakang, rumusan masalah, batasan

masalah, tujuan, manfaat, dan sistematika penulisan.

- Bab 2 LANDASAN TEORI

Bab kedua membahas kumpulan teori sebagai dasar penelitian yang mencakup lirik lagu, kata seksual, *preprocessing*, *voting classifier*, ETC, LSTM, *feature extraction*, *k-fold cross validation*, dan metrik evaluasi

- Bab 3 METODOLOGI PENELITIAN

Bab ketiga membahas tahapan penelitian yang dilakukan selama proses penelitian. Tahapan ini dibuat dengan *flowchart* dengan menggambarkan proses penelitian selama penelitian.

- Bab 4 HASIL DAN DISKUSI

Bab keempat membahas mengenai hasil dari penelitian yang dilakukan. Hasil dari penelitian diberikan dengan menggunakan potongan kode serta gambar untuk menunjukkan hasil penerapan algoritma.

- Bab 5 KESIMPULAN DAN SARAN

Bab kelima menjelaskan mengenai kesimpulan dari penelitian yang telah dilakukan, serta saran untuk pengembangan penelitian lebih lanjut.

