

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah membawa banyak perubahan dalam kehidupan sehari-hari, termasuk dalam upaya melestarikan budaya. Salah satu bentuk budaya yang penting untuk dijaga adalah lagu daerah, karena mengandung nilai sejarah, kearifan lokal, dan ciri khas suatu masyarakat. Sayangnya, masih banyak lagu daerah yang belum terdokumentasikan dengan baik, sehingga sulit ditemukan, dianalisis, atau dipelajari, terutama oleh generasi muda dan masyarakat umum.

Indonesia sendiri memiliki beragam lagu daerah dari berbagai provinsi. Misalnya, Jawa Barat dikenal dengan lagu-lagu seperti "*Bubuy Bulan*" dan "*Es Lilin*", sementara dari Sumatera Utara ada lagu "*Butet*" dan "*Anju Ahu*". Namun, sebagian besar lagu-lagu ini masih belum terdigitalisasi dan hanya disimpan dalam bentuk fisik atau diingat secara turun-temurun. Jika tidak segera dikelola dengan baik, warisan budaya ini bisa saja hilang. Karena itu, dibutuhkan sistem yang tidak hanya menyimpan, tetapi juga mampu mengekstraksi informasi penting dari lirik lagu daerah agar dapat dianalisis dan dimanfaatkan lebih lanjut.

Berdasarkan hasil kuesioner yang diisi oleh 33 responden, terlihat bahwa tingkat pengenalan terhadap lagu daerah masih terbatas. Sebagian besar responden, yaitu 54,5%, hanya mengenal sebagian lagu daerah, sementara 45,5% lainnya mengaku kurang mengenal atau sekadar mengetahui judul lagunya saja. Dari sisi kebiasaan mendengarkan, hanya 6,1% responden yang menyatakan rutin mendengarkan lagu daerah, sedangkan sisanya, yakni 93,9%, jarang bahkan hampir tidak pernah mendengarkannya. Kendala lain yang cukup menonjol adalah kesulitan dalam memahami makna lirik, yang dialami oleh 78,8% responden, sementara 21,2% responden merasa tidak mengalami masalah berarti.

Terkait kebutuhan akan adanya sistem yang dapat membantu menjelaskan makna lagu daerah, 57,6% responden menilai sistem tersebut sangat penting, sedangkan 42,4% responden menilai cukup penting. Selain itu, mayoritas responden juga menunjukkan minat yang tinggi terhadap pengembangan sistem ini. Mereka menyatakan tertarik atau sangat tertarik jika sistem mampu menyediakan fitur pencarian berdasarkan tema, ekstraksi kata kunci penting, serta ringkasan makna lirik lagu. Temuan ini mengindikasikan adanya kebutuhan nyata terhadap sistem ekstraksi informasi lirik lagu daerah yang diharapkan dapat mempermudah masyarakat dalam mengakses sekaligus memahami nilai budaya yang terkandung di dalamnya.

Ekstraksi informasi adalah proses mengidentifikasi dan mengambil elemen-elemen penting dari suatu dokumen atau teks sehingga dapat mewakili isi secara ringkas dan jelas. Dalam konteks lirik lagu daerah, proses ini dapat digunakan untuk mendapatkan kata kunci utama, frasa bermakna, atau tema yang menggambarkan karakteristik lagu tersebut. Hasil ekstraksi ini dapat dimanfaatkan untuk keperluan analisis, pengelompokan, atau dokumentasi budaya secara lebih terstruktur.

Beberapa penelitian sebelumnya telah mengkaji teknik ekstraksi informasi berbasis teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Misalnya, penelitian oleh Ibnu Surya Wibowo, Arita Witanti, dan Indah Susilawati (2024) melakukan *keyword extraction* pada judul berita online di Indonesia yang diambil dari tiga situs berita populer (Kompas.com, Detik.com, dan Tempo.co). Proses ekstraksi dimulai dari *text preprocessing* (*case folding*, *tokenizing*, *stopword removal*, dan *stemming*) hingga pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil perhitungan menghasilkan 30 kata kunci teratas, dengan tingkat akurasi mencapai 99,76% dan F1-Score sebesar 99,67%. [1]

Penelitian tersebut membuktikan bahwa metode *Term Frequency-Inverse Document Frequency* (TF-IDF) mampu memberikan bobot tinggi pada kata-kata yang sering muncul dalam satu dokumen namun jarang muncul di dokumen lain,

sehingga efektif untuk mengidentifikasi kata kunci yang relevan. Meskipun demikian, penerapannya masih terbatas pada domain berita online. Belum ada penelitian yang secara khusus mengadaptasikan metode ini untuk ekstraksi informasi dari lirik lagu daerah yang memiliki karakteristik bahasa, makna, dan nilai budaya yang berbeda dibandingkan teks berita. Oleh karena itu, penelitian ini berupaya mengisi kesenjangan tersebut dengan mengimplementasikan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* untuk mengekstraksi informasi penting dari lirik lagu daerah sebagai upaya pelestarian budaya.

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan karena mampu memberikan bobot pada kata berdasarkan frekuensi kemunculannya, sehingga kata yang lebih mewakili isi lagu akan menonjol. Sementara itu, *Cosine Similarity* dapat menghitung tingkat kemiripan antara teks hasil ekstraksi dengan referensi atau kategori tertentu, sehingga memudahkan proses analisis dan pengelompokan. Kombinasi kedua metode ini diharapkan dapat menghasilkan sistem ekstraksi informasi yang akurat, cepat, dan relevan.

Dengan adanya sistem ekstraksi informasi lirik lagu daerah ini, diharapkan para peneliti budaya, pendidik, musisi, maupun masyarakat umum dapat dengan mudah mengakses, memahami, dan memanfaatkan informasi penting dari lagu daerah. Selain itu, sistem ini juga dapat menjadi sarana edukasi bagi generasi muda agar lebih mengenal dan mencintai kekayaan budaya Indonesia, sekaligus menjadi pijakan untuk pengembangan sistem serupa pada domain kebudayaan lainnya di masa depan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini berupaya menjawab dua pertanyaan utama. Pertama, bagaimana merancang sebuah sistem yang mampu mengekstraksi informasi dari lirik lagu daerah secara terstruktur dan efisien sehingga dapat menjadi sarana pendukung dalam pelestarian budaya Indonesia. Sistem ini diharapkan mampu membantu pengguna memahami isi dan makna lagu melalui pengolahan teks yang menghasilkan informasi yang relevan

dan mudah dipahami. Kedua, bagaimana penerapan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* dan *Cosine Similarity* dapat dimanfaatkan untuk mendukung proses ekstraksi informasi pada lirik lagu daerah melalui pembobotan kata dan pengukuran tingkat kemiripan terhadap *corpus* lagu. Kedua metode tersebut digunakan untuk mengidentifikasi kata-kata yang memiliki bobot penting serta menentukan tingkat kemiripan antara lirik lagu yang diinput dengan corpus, sehingga dapat mendukung proses analisis dan interpretasi makna lagu secara lebih tepat.

1.3 Batasan Masalah

Penelitian ini dibatasi pada pengolahan data berupa lirik lagu daerah sebagai objek kajian, tanpa melibatkan unsur musik, melodi, atau aspek audio lainnya. Sistem yang dikembangkan hanya menggunakan metode TF-IDF dan *Cosine Similarity* untuk proses pembobotan kata, pengukuran kemiripan teks, serta penentuan relevansi lirik berdasarkan kata kunci *input* pengguna. Proses analisis linguistik dilakukan secara otomatis menggunakan kombinasi NLP (*POS tagging, dependency parsing, dan model multilingual*) serta pendekatan *rule-based* sebagai mekanisme *fallback*, sehingga hasil yang diperoleh berupa ekstraksi informasi linguistik dan ringkasan makna semantik dari lirik lagu. Pembatasan ini bertujuan untuk menjaga fokus penelitian pada pengujian efektivitas metode pemrosesan teks dan kualitas informasi semantik yang dihasilkan oleh sistem.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah untuk membuat sistem ekstraksi informasi lirik lagu daerah menggunakan *Term frequency – inverse document frequency (TF-IDF)* dan *Cosine Similarity*.

1.5 Manfaat

Adapun manfaat dari sistem pencarian informasi lirik lagu daerah menggunakan *Term frequency – inverse document frequency (TF-IDF)* dan *cosine Similarity* sebagai berikut:

1. Memudahkan masyarakat, pelajar, peneliti, maupun pecinta musik daerah dalam memahami makna lirik lagu tanpa harus membaca dan menafsirkan seluruh teks secara manual.
2. Menyediakan sarana edukasi yang dapat membantu generasi muda mengenal dan mencintai lagu daerah dengan cara yang lebih modern dan mudah diakses.
3. Membantu pelestarian budaya Indonesia melalui digitalisasi dan pengolahan lirik lagu daerah sehingga informasi budaya dapat disimpan, diolah, dan dibagikan secara luas.
4. Sebagai salah satu syarat untuk menyelesaikan Program Studi Serjana Terapan Rekayasa Perangkat Lunak.