

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Konsumsi masyarakat terhadap musik setiap harinya cukup tinggi. Dilansir dari CNN Indonesia, orang Indonesia dapat mendengarkan berbagai macam lagu di platform Spotify, selama 3 jam per hari. Direktur Spotify untuk Asia, Sunita Kaur, mengatakan bahwa Indonesia merupakan negara yang paling aktif mendengarkan musik [1]. Konsumsi masyarakat terhadap musik dapat semakin meningkat dengan adanya berbagai platform *streaming* musik. Masyarakat dapat lebih mudah mengakses berbagai macam lagu, dengan menggunakan platform *streaming* musik.

Namun, keberadaan platform *streaming* musik belum tentu menguntungkan bagi semua musisi. Menurut pengamat persaingan pasar di Inggris, persaingan industri musik menjadi lebih ketat dengan adanya platform *streaming* musik. Sistem *streaming* musik dianggap hanya menguntungkan musisi papan atas, yang berasal dari label musik terkenal. Hanya sebagian kecil musisi yang mengalami keuntungan, sedangkan sebagian besar musisi mengalami kerugian [2]. Untuk menghadapi masalah ini, para musisi harus membuat strategi baru. Faktor-faktor yang mempengaruhi popularitas suatu lagu perlu lebih diperhatikan oleh para musisi, sehingga lagu ciptaan mereka dapat menjadi populer dan menghasilkan keuntungan.

Berbagai macam faktor dapat menyebabkan suatu lagu menjadi populer. Faktor-faktor tersebut dapat berkaitan dengan elemen-elemen musik suatu lagu. Salah satu dari elemen musik lagu, yang dapat mempengaruhi popularitas lagu adalah lirik dari lagu tersebut. Lagu dengan pola lirik tertentu lebih disukai oleh masyarakat, sehingga lagu tersebut menjadi lebih populer [3]. Berdasarkan beberapa penelitian di bidang *deep learning*, diketahui bahwa fitur lirik mempunyai pengaruh yang cukup signifikan pada popularitas lagu, bahkan lebih signifikan dibandingkan dengan fitur audio [4]. Adapun akurasi prediksi popularitas lagu dengan menggunakan fitur lirik lagu lebih baik, dibandingkan dengan akurasi prediksi popularitas lagu tanpa menggunakan fitur lirik lagu [5][4].

Peranan lirik pada popularitas suatu lagu dapat diteliti dengan analisis sentimen lirik lagu tersebut. Adapun analisis sentimen merupakan studi komputasi yang mempelajari sentimen atau emosi pada suatu teks. Pada penelitian "*Predicting*

*a Hit Song with Machine Learning: Is There an Apriori Secret Formula?*”, prediksi popularitas lagu dengan menggunakan analisis sentimen pada lirik lagu dan analisis fitur audio menghasilkan tingkat akurasi yang lebih tinggi, dibandingkan prediksi popularitas lagu hanya dengan analisis fitur audio. Akurasi tertinggi yang dapat diperoleh adalah 51%, dengan menggunakan algoritma *Naïve Bayes* [6]. Pada penelitian lain yang serupa, dilakukan prediksi popularitas lagu dengan fokus hanya pada lirik lagu. Akurasi yang diperoleh untuk prediksi popularitas lagu mencapai 73%, dengan menggunakan algoritma BERT [7].

*Bidirectional Encoder Representations from Transformers* (BERT) merupakan sebuah *pretrained language model*, sebuah algoritma *deep learning* dapat digunakan untuk proyek *Natural Language Processing* (NLP). Adapun BERT merupakan algoritma *deep learning* yang diperkenalkan oleh Google pada akhir tahun 2018 [8]. Algoritma BERT banyak digunakan pada penelitian analisis sentimen. BERT digunakan pada penelitian analisis sentimen *customer review* di Amazon, dimana BERT mengungguli performa *Naive Bayes*, *Long Short Term Memory* (LSTM), *Support Vector Machine* (SVM) [9]. Hal serupa juga ditemukan pada penelitian *review* produk di *e-commerce*, dimana performa BERT mengungguli performa *Recurrent Neural Network* (RNN), *Long Short Term Memory* (LSTM), dan *Gated Recurrent Unit* (GRU) [10]. Performa model BERT masih dapat bersaing dengan performa model NLP yang lebih baru, bahkan seringkali mengungguli model-model NLP yang lebih baru. Contohnya seperti pada penelitian analisis sentimen publik mengenai metode vaksin *Human Papillomavirus* (HPV) di Twitter, BERT berhasil mengungguli *Generative Pre-trained Transformer* (GPT) [11]. Dengan demikian, dapat diketahui bahwa BERT merupakan algoritma yang unggul untuk kepentingan analisis sentimen [12].

Berdasarkan uraian di atas, pada penelitian ini dikembangkan sebuah model berbasis algoritma BERT untuk prediksi popularitas lagu, berdasarkan analisis sentimen pada lirik lagu. Penelitian ini dilakukan dengan dasar kebutuhan para musisi untuk dapat mengembangkan lagu yang terkenal, dengan memperhatikan faktor lirik lagu. Dengan penelitian ini, masyarakat khususnya para penulis lagu dapat memprediksi popularitas sebuah lagu, dengan fokus dari segi lirik lagu. Berdasarkan hasil prediksi, penulis lagu dapat mengevaluasi lirik lagu ciptaannya agar lagu tersebut sesuai dengan selera masyarakat dan lebih berpeluang untuk menjadi terkenal. Algoritma yang dipakai adalah BERT, karena BERT merupakan algoritma *deep learning* yang unggul dalam tugas analisis sentimen. Selain itu, penelitian ini dilakukan sebagai bentuk pengembangan lebih lanjut dari penelitian

terdahulu, dimana akurasi sebesar 73% telah dicapai untuk prediksi popularitas lagu dengan menggunakan BERT. Data yang digunakan untuk pelatihan model merupakan data lagu-lagu berbahasa Inggris, berjumlah 52 ribu data. Performa model ini diukur menggunakan indikator *accuracy*, *precision*, *recall*, dan *f1-score*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, berikut merupakan rumusan masalah yang akan diteliti lebih lanjut.

1. Bagaimana analisis sentimen lirik lagu dapat memprediksi popularitas suatu lagu, dengan menggunakan algoritma BERT?
2. Bagaimana hasil performa dari penerapan algoritma BERT, untuk analisis sentimen lirik lagu dalam rangka memprediksi popularitas lagu tersebut, diukur dari *accuracy*, *precision*, *recall* dan *f1-score* ?

## 1.3 Batasan Masalah

Berikut adalah batasan masalah dari penelitian ini.

1. Penelitian ini menggunakan data lagu-lagu dari *playlist* Top 50 di berbagai negara, yang tersedia di platform penyedia musik Spotify. *Dataset* lagu dibentuk menggunakan API Spotify dan API Genius [13]. *Dataset* ini dapat diakses pada situs web penyedia *dataset* publik Kaggle [14].
2. Data lagu yang digunakan adalah lagu dengan lirik berbahasa Inggris, berjumlah 52 ribu.
3. Model BERT yang digunakan adalah model BERT *bert-based-uncased*. Model BERT ini dilatih dengan data teks bahasa Inggris non-kapital.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Menganalisis sentimen lirik lagu untuk memprediksi popularitas suatu lagu, dengan menggunakan algoritma BERT.

2. Mengukur hasil performa dari penerapan algoritma BERT, untuk analisis sentimen lirik lagu dalam rangka memprediksi popularitas lagu tersebut dengan *accuracy*, *precision*, *recall* dan *f1-score*.

### 1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi peneliti, penelitian ini bermanfaat untuk menambah dan menerapkan ilmu mengenai analisis sentimen pada lirik lagu untuk memprediksi popularitas suatu lagu, berbasis algoritma BERT.
2. Bagi pengguna, prediksi popularitas lagu berdasarkan analisis sentimen lirik lagu akan membantu pengguna dalam pembuatan dan pengembangan lirik lagu, sehingga lagu yang dibuat dapat berpeluang tinggi untuk menjadi populer.

### 1.6 Sistematika Penulisan

Sistematika penulisan laporan skripsi ini adalah sebagai berikut.

- **BAB 1 PENDAHULUAN**

Pada bab ini, diuraikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan pada laporan skripsi ini.

- **BAB 2 LANDASAN TEORI**

Pada bab ini, dijelaskan teori-teori yang digunakan sebagai landasan dari penelitian ini. Teori-teori yang dimaksud merupakan teori mengenai lirik lagu, analisis sentimen, *preprocessing*, *Bidirectional Encoder Representations from Transformers* (BERT), dan metrik evaluasi.

- **BAB 3 METODOLOGI PENELITIAN**

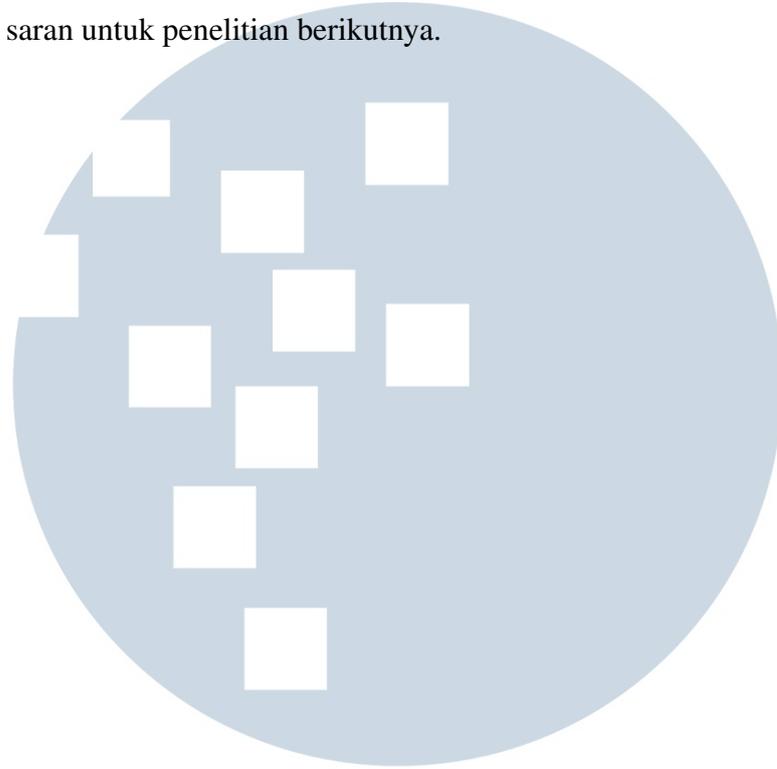
Pada bab ini, diuraikan tahap-tahap penelitian secara urut. Berbagai diagram dan gambar disertakan untuk menjelaskan perancangan model dan perancangan desain antarmuka, pada penelitian ini.

- **BAB 4 HASIL DAN DISKUSI**

Pada bab ini, dijelaskan hasil implementasi pada penelitian ini. Hasil *output* dari penelitian ini akan ditulis dan dianalisis pada bab ini.

- **BAB 5 SIMPULAN DAN SARAN**

Pada bab ini, dijabarkan kesimpulan dari penelitian yang telah dilakukan, serta saran untuk penelitian berikutnya.



**UMMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA