

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam era digital penyebaran berbagai informasi dan opini semakin dipercepat, sarkasme seringkali digunakan untuk menarik perhatian pengguna internet. Definisi sarkasme menurut kamus Cambridge adalah penggunaan komentar yang jelas-jelas memiliki arti yang berlawanan dengan apa yang dikatakannya, yang dibuat dengan tujuan untuk menyakiti perasaan seseorang atau untuk mengkritik sesuatu dengan cara yang lucu. Sarkasme digunakan oleh berbagai kalangan baik pengguna internet biasa maupun tokoh masyarakat, dalam perdebatan *online* atau ketika membahas suatu peristiwa publik atau topik hangat dan kontroversial [1]. Ketika melakukan percakapan dengan seseorang secara langsung, dengan memperhatikan berbagai konteks seperti gerak tubuh, nada suara, dan perasaan pembicara, seseorang dapat dengan cepat mengidentifikasi sarkasme [2]. Namun sarkasme sulit diidentifikasi dalam komunikasi tertulis karena ciri-ciri tersebut tidak terlihat jelas dalam tulisan yang sarkastik [3]. Sarkasme dalam internet dapat ditemukan dalam berbagai bentuk, seperti opini dalam sebuah komentar, sebuah *posting* yang mengkritik, hingga *headline* sebuah berita. Penggunaan sarkasme dalam *headline* berita banyak menimbulkan dampak negatif seperti ambiguitas makna, kesalahanpahaman pembaca, bahkan disinformasi apabila tidak dipahami dengan benar [4]. Terlebih juga dampak pada ranah sosial, politik, atau ekonomi, sarkasme dalam berita dapat memperkeruh opini publik dan memicu interpretasi yang menyesatkan. Kesalahan dalam mengidentifikasi satire sebagai fakta dapat menyesatkan persepsi publik terhadap isu serius dan memicu polarisasi opini secara tidak sehat [5]. Karena itu upaya untuk mendeteksi sarkasme secara otomatis merupakan hal yang penting, karena kemampuan untuk mengidentifikasi sarkasme memastikan interpretasi yang lebih akurat dari sentimen mendasar yang disampaikan dalam suatu teks [6].

BERT (*Bidirectional Encoder Representations from Transformers*) adalah model bahasa berbasis *Transformer encoder* yang dilatih secara pre-training pada tugas masked language modeling sehingga dapat menangkap konteks dua arah (kiri dan kanan) dari setiap token, arsitekturnya terdiri dari lapisan *embedding* diikuti oleh beberapa blok *self-attention multi-head* dan *feed-forward (transformer)*

encoder stack) yang memungkinkan representasi kontekstual kaya untuk tiap kata [7]. Karena representasi BERT bersifat *contextualized* yang berarti nilai *embedding* berubah menurut konteks kalimat dan bisa dilakukan *finetuning* pada dataset dengan tugas spesifik, BERT umumnya lebih unggul daripada metode tradisional berbasis *bag-of-words* atau *embedding* statis (mis. Word2Vec/FastText) pada tugas yang memerlukan nuansa pragmatik seperti sarkasme, BERT mampu menyerap petunjuk seperti konteks serta pola bahasa yang menandai ironi atau ketidaksesuaian makna literal [8]. Dalam penelitian khusus deteksi sarkasme, studi-studi terbaru menunjukkan bahwa BERT (yang sudah *difinetuning*) / varian seperti RoBERTa meningkatkan akurasi dibanding *pipeline* tradisional dan metode klasik (*feature engineering + SVM*), terutama ketika model dipadu dengan fitur kontekstual atau data tambahan seperti metadata atau konteks percakapan [9]. Selain itu, kemampuan *transfer learning* BERT memudahkan adaptasi ke dataset sarkasme yang relatif kecil melalui *finetuning*, sehingga menjadi pilihan praktis dan efektif dalam riset sarkasme modern [10].

Berikut beberapa penelitian menggunakan BERT dalam mendekripsi sarkasme, Bhardwaj dan Prusty [11] menyusun model *hybrid* yang memadukan *embeddings* BERT dengan CNN dan LSTM, menghasilkan akurasi hingga 99,63% dan F1-score 99,56% pada dataset *headline* berita sarkastik. Studi oleh Parameswaran [12] dalam *shared task* ALTA menunjukkan bahwa model BERT memperbaiki performa deteksi target sarkasme hingga 15,2% dibanding penelitian sebelumnya pada *corpus* Reddit. Fitur BERT (konteks *bidirectional* dan *self-attention*) dan hasil penelitian dari berbagai sumber inilah yang menjadi landasan mengapa BERT merupakan pilihan dalam mendekripsi sarkasme. Dalam penelitian ini akan digunakan google BERT base uncased yang di *finetune* pada dataset sarkasme yang sudah di label untuk melihat performa BERT dalam mendekripsi sarkasme.

1.2 Rumusan Masalah

Mengacu pada latar belakang yang telah dijelaskan, permasalahan yang dirumuskan meliputi:

1. Bagaimana mengimplementasikan model BERT dalam mendekripsi sarkasme pada *headline* berita
2. Bagaimana hasil performa BERT dalam mendekripsi sarkasme pada *headline*

berita dalam pengukuran accuracy pada training, validasi, dan testing

1.3 Batasan Permasalahan

Mengacu pada rumusan masalah yang telah dijelaskan, batasan permasalahan yang dibuat meliputi:

1. Dataset yang digunakan merupakan yang sudah di label, yang tersedia pada kaggle. Kumpulan *headline* berita pada dataset ini dikumpulkan dari dua situs web berita, yaitu TheOnion untuk *headline* yang mengandung sarkasme dan HuffPost untuk *headline* berita yang tidak mengandung sarkasme [13]
2. *Headline* berita yang digunakan berbahasa inggris
3. Jumlah *headline* berita dalam dataset adalah sebanyak 26.708

1.4 Tujuan Penelitian

Dengan mempertimbangkan rumusan masalah penelitian, dapat diberi beberapa tujuan penelitian sebagai berikut:

1. Mengimplementasikan BERT untuk mendeteksi sarkasme pada *headline* berita
2. Mengetahui hasil performa BERT dalam mendeteksi sarkasme pada *headline* berita dalam pengukuran accuracy pada training, validasi, dan testing

1.5 Manfaat Penelitian

Berikut beberapa manfaat penelitian yang sesuai dengan tujuan penelitian ini:

1. Model BERT dapat digunakan sebagai alat yang efisien untuk mendeteksi sarkasme
2. Model dapat digunakan oleh pihak terkait untuk mendeteksi sarkasme sesuai dengan datasetnya yaitu *headline* berita

1.6 Sistematika Penulisan

Penulisan skripsi ini meliputi lima bab dengan sistematika penulisan sebagai berikut :

- Bab 1 PENDAHULUAN

Merangkum latar belakang masalah mengenai deteksi sarkasme pada headline sebuah berita dan BERT sebagai model yang disarankan dalam pembuatan model deteksi sarkasme, rumusan masalah terkait topik judul penelitian, batasan masalah, tujuan penelitian, serta manfaat yang dapat diambil dari penelitian.

- Bab 2 LANDASAN TEORI

Pada bab ini dituliskan teori-teori yang mendukung penelitian, antara lain adalah tahapan-tahapan dalam melakukan deteksi sarkasme menggunakan model BERT seperti text preprocessing.

- Bab 3 METODOLOGI PENELITIAN

Pada bab ini dituliskan dan diilustrasikan dalam bentuk flowchart metodologi yang akan digunakan dalam penelitian antara lain adalah studi literatur, pengumpulan data, dan perancangan sistem.

- Bab 4 HASIL DAN DISKUSI

Pada bab ini dituliskan hasil dan proses membangun model deteksi sarkasme menggunakan BERT serta pembahasan performa dari model BERT.

- Bab 5 KESIMPULAN DAN SARAN

Pada bab ini dituliskan kesimpulan yang didapat dari hasil penelitian serta saran yang dapat digunakan bagi penelitian selanjutnya.

UNIVERSITAS
MULTIMEDIA
NUSANTARA