

# 1. PENDAHULUAN

## Latar Belakang

Covid19 atau Coronavirus Disease 2019, merupakan penyakit menular yang pertama kali ditemukan di Wuhan China pada tahun 2019. Gejala pasien pertama Covid-19 termasuk demam, malaise, batuk kering, dan dispnea, didiagnosis sebagai pneumonia [1]. Seluruh belahan dunia perlahan-lahan tertular penyakit ini. Di Indonesia Covid-19 ditemukan pada bulan maret 2020. Sejak saat itu, pemerintah segera melakukan tindakan pencegahan dengan pembatasan sosial. Seluruh kegiatan tatap muka berganti menjadi via internet, termasuk kegiatan sekolah dan pekerjaan. Kebutuhan memberitakan perkembangan dan penanggulangan covid-19 disampaikan melalui televisi dan media sosial. Media sosial yang digunakan termasuk melalui portal berita, facebook, instagram dan twitter sangat berperan dalam pertukaran informasi seputar covid-19.

Twitter merupakan media sosial yang populer di dunia, jejaring sosial berbasis microblog didirikan pada tahun 2006 oleh John Dorsey. Pengguna twitter di Indonesia menjadi negara dengan pengguna twitter paling banyak di dunia [2]. Postingan twitter disebut *tweet* mampu mengirim pesan dengan teks maksimal 280 karakter, foto dan video pendek. *Tweet* yang hanya mampu memuat 280 karakter membuat pengguna yang ingin mengirim pesan lebih panjang melakukan beberapa hal seperti menulis dengan mengubah kata menjadi singkatan atau akronim. Singkatan dan akronim adalah dua hal yang berbeda. Secara proses morfologi, singkatan merupakan bentuk kata baru dari hasil penanggalan satu atau lebih terhadap kombinasi leksem [3]. Sedangkan akronim adalah gabungan dari huruf awal nama kata, nama judul, atau frasa dan dilafalkan seperti kata baru [4].

Pertukaran informasi di Twitter sangat cepat dan dapat dilakukan semua pengguna. Keadaan ini menyebabkan berita palsu juga ikut tersebar. Berita palsu merupakan berita yang tidak benar dan tidak kredibel. Berita palsu dapat disebarkan secara sengaja maupun secara tidak sengaja. Berita palsu yang disebarkan secara sengaja biasanya karena memiliki tujuan tertentu seperti propaganda, politik atau untuk menguntungkan pihak tertentu. Persebaran berita palsu Covid-19 sangat merugikan banyak pihak karena berita palsu covid-19 bisa saja mengenai cara pencegahan, cara pengobatan, PSBB (Pembatasan Sosial Berskala Besar) dan vaksin. Maka daripada itu, dibutuhkan suatu sistem yang mampu mendeteksi berita palsu covid-19.

Transformer merupakan arsitektur yang mengubah satu urutan ke urutan (sequence to sequence) dengan bantuan *encoder-decoder* [5]. Model transformer yang populer akhir-akhir ini adalah BERT. BERT atau Bidirectional Encoder Representation from Transformer merupakan model *deep learning* yang pertama kali dikenalkan oleh Google [6]. Sejak dikenalkan pada tahun 2018, BERT telah menjadi baseline dalam penelitian mengenai NLP. BERT telah dilatih (pretrained) menggunakan teks *corpus* wikipedia sehingga mampu membantu komputer memahami bahasa ambigu dengan memanfaatkan teks disekitarnya.

Pada penelitian ini, kami mengusulkan deteksi berita palsu Covid-19 menggunakan arsitektur Transformer yaitu BERT. Kami menambahkan proses augmentasi data untuk meningkatkan performa BERT sehingga mampu meningkatkan nilai evaluasi 4 metrik : akurasi, presisi, recall dan f1-score. Tujuan penelitian ini adalah untuk mendeteksi berita palsu sehingga mampu meminimalkan penyebaran berita palsu Covid-19. Kontribusi utama dari penelitian ini adalah menunjukkan bahwa proses augmentasi data dapat meningkatkan performa bert dalam mendeteksi berita palsu.

## Topik dan Batasannya

Pada penelitian ini kami menggunakan dataset dari postingan twitter berbahasa Indonesia dari penelitian sebelumnya [7].

## Tujuan

Penelitian ini bertujuan untuk meningkatkan performa BERT dalam mendeteksi berita palsu Covid-19 dengan menambahkan proses augmentasi data. Adapun augmentasi data yang akan digunakan adalah augmentasi data berbasis akronim, berbasis pemeriksaan ejaan dan berbasis typo.

**Organisasi Tulisan**

Adapun susunan untuk penelitian ini yaitu bagian pertama membahas pendahuluan, bagian kedua membahas studi terkait, bagian ketiga membahas sistem yang dibangun, bagian keempat membahas evaluasi dan bagian kelima membahas kesimpulan.