



TUGAS AKHIR - KM184801

***SENTIMENT ANALYSIS PELAKSANAAN WORK FROM HOME*** DI INDONESIA PADA MASA PANDEMI COVID-19  
**MENGGUNAKAN IndoBERT**

**BELLA RAHMATULLAH**  
**0611164000078**

Dosen Pembimbing  
Prof. Dr. tech. Drs. Mohammad Isa Irawan, MT.

Departemen Matematika  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember  
Surabaya 2021





**TUGAS AKHIR- KM184801**

***SENTIMENT ANALYSIS PELAKSANAAN  
WORK FROM HOME DI INDONESIA PADA  
MASA PANDEMI COVID-19  
MENGUNAKAN IndoBERT***

**BELLA RAHMATULLAH  
NRP 061116 4000 0078**

**Dosen Pembimbing :  
Prof. Dr. tech. Drs. Mohammad Isa Irawan, MT.**

**DEPARTEMEN MATEMATIKA  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember  
Surabaya 2021**





**FINAL PROJECT- KM184801**

***SENTIMENT ANALYSIS WORK FROM  
HOME IMPLEMENTATION IN INDONESIA  
DURING COVID-19 PANDEMIC USING  
IndoBERT***

**BELLA RAHMATULLAH  
NRP 061116 4000 0078**

**Supervisor :  
Prof. Dr. tech. Drs. Mohammad Isa Irawan, MT**

**DEPARTMENT OF MATHEMATICS  
Faculty of Science and Data Analytics  
Institut Teknologi Sepuluh Nopember  
Surabaya 2021**



## **LEMBAR PENGESAHAN**

**Sentiment Analysis Pelaksanaan Work From Home di Indonesia  
pada Masa Pandemi COVID-19 Menggunakan IndoBERT**  
*Sentiment Analysis Work From Home Implementation in Indonesia  
During COVID-19 Pandemic using IndoBERT*

### **Tugas Akhir**

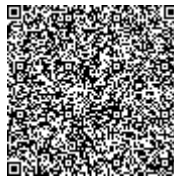
Diajukan untuk memenuhi salah satu syarat Untuk memperoleh gelar  
Sarjana Matematika Pada bidang studi Ilmu Komputer  
Program Studi S-1 Departemen Matematika  
Fakultas Sains dan Analitika Data  
Institut Teknologi Sepuluh Nopember Surabaya

Oleh:

Bella Rahmatullah  
NRP: 0611164000078

Menyetujui,

Prof. Dr. tech.. Drs. Mohammad Isa  
Irawan, MT  
19631225 198903 1 001



Mengetahui,  
Kepala Departemen Matematika  
FSAD ITS



Subchan, Ph.D  
NIP. 19710513 199702 1 001  
Surabaya, 23 Agustus 2021





***SENTIMENT ANALYSIS PELAKSANAAN WORK  
FROM HOME DI INDONESIA PADA MASA  
PANDEMI COVID-19 MENGGUNAKAN IndoBERT***

**Nama** : Bella Rahmatullah  
**NRP** : 0611164000078  
**Dosen Pembimbing** : Prof. Dr. tech. Drs. Mohammad  
Isa Irawan, MT.

**ABSTRAK**

Pembatasan Sosial Berskala Besar (PSBB) yang sekarang berubah nama menjadi Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) pada masa pandemi COVID-19 oleh Pemerintah Republik Indonesia membuat perusahaan, instansi, maupun organisasi memberlakukan kebijakan bekerja dari rumah atau yang sering disebut *Work From Home* (WFH). Pelaksanaan WFH pada masa pandemi tentunya memiliki keuntungan dan tantangan tersendiri bagi setiap pekerja. Opini mengenai pelaksanaan WFH kerap diungkapkan masyarakat melalui media sosial Twitter. Data opini dianalisis untuk mengetahui kecenderungan sentimen masyarakat sebagai pertimbangan kelanjutan pelaksanaan WFH pasca pandemi. Model *pre-trained Bidirectional Encoder Representations from Transformers* (BERT) merupakan model yang sedang populer di kalangan peneliti. Pada penelitian ini data opini dianalisis menggunakan *pre-trained model* BERT bahasa Indonesia, IndoBERT untuk analisis sentimen. Model yang digunakan menunjukkan performansi pembelajaran sebesar 95 % dan 68% pada saat pengujian. Dari hasil penelitian ini diketahui bahwa masyarakat Indonesia cenderung memiliki sentimen positif terhadap pelaksanaan *work from home* pada masa pandemi COVID-19.

***Kata Kunci : Analisis Sentimen, BERT, IndoBERT, WFH***

**SENTIMENT ANALYSIS WORK FROM HOME  
IMPLEMENTATION IN INDONESIA DURING  
COVID-19 PANDEMIC USING IndoBERT**

**Name of Student** : Bella Rahmatullah  
**NRP** : 06111640000078  
**Department** : Mathematics  
**Supervisor** : Prof. Dr. tech. Drs. Mohammad  
Isa Irawan, MT.

**ABSTRACT**

*The Large-Scale Social Restrictions (PSBB) which was now changed its name to Enforcement of Restriction on Community (PPKM) policy during the COVID-19 pandemic by Indonesia's government make companies, agencies, or organizations enforce a policy of Work From Home (WFH). The implementation of WFH has its own advantages and challenges for each worker. Opinions regarding the implementation of WFH are often expressed by public through social media Twitter. Opinion data was analyzed to determine the tendency of public statement as a consideration for the continuation of post-pandemic WFH implementation. The pre-trained model Bidirectional Encoder Representations of Transformers (BERT) is currently popular among researchers. In this study, opinion data were analyzed using pre-trained Indonesian BERT-based model, IndoBERT for sentiment analysis. The model shows learning performance in 95% for training phase and 68% for testing phase. From this study, it is known that Indonesian people tend to have positive sentiments towards implementation of work from home during COVID-19 pandemic.*

**Keyword** : BERT, IndoBERT, Sentiment Analysis, WFH



## KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Puji Syukur kehadiran Allah SWT atas limpahan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dan menyusun dokumentasi penelitian Tugas ini dengan judul “*Sentiment Analysis Pelaksanaan Work From Home di Indonesia pada Masa Pandemi COVID-19 Menggunakan IndoBERT*” sebagai salah satu syarat menyelesaikan Program Sarjana Departemen Matematika Fakultas Sains dan Analitika Data Institut Teknologi Sepuluh Nopember (ITS) Surabaya.

Tugas Akhir ini dapat terselesaikan dengan baik tidak lepas dari bimbingan, bantuan, doa, dan Kerjasama dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Kedua orang tua dan keluarga penulis yang selalu memberikan doa, motivasi dan dukungan moral dalam menyelesaikan Tugas Akhir ini,
2. Bapak Prof. Dr. tech. Drs. Mohammad Isa Irawan, MT. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta motivasi kepada penulis selama pengerjaan Tugas Akhir sehingga Tugas Akhir ini dapat terselesaikan.
3. Bapak Subchan, S.Si, M.Si, Ph.D. selaku Kepala Departemen Matematika ITS.
4. Ibu Alvida Mustika Rukmi, S.Si, M.Si. selaku dosen wali yang telah memberikan arahan akademik selama penulis menempuh Pendidikan di Departemen Matematika ITS.
5. Bapak Dr. Imam Mukhlash, S.Si, MT., Bapak Mohammad Iqbal S.Si., M.Si., Bapak Muhammad Luthfi Shahab, S.Si., M.Si. selaku dosen penguji yang telah memberikan saran untuk perbaikan Tugas Akhir ini.

6. Seluruh jajaran dosen dan staff Departemen Matematika FSAD ITS.
7. Teman-teman angkatan 2016 yang saling mendukung.
8. Teman-teman dari Resimen Mahasiswa ITS yang telah memberikan pengalaman baru dan suasana kekeluargaan selama penulis menjadi mahasiswa.
9. Semua pihak yang tidak bisa disebutkan, terima kasih telah membantu hingga penulis menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa penelitian Tugas Akhir ini masih jauh dari kesempurnaan, sehingga penulis mengharapkan kritik dan saran dari pembaca. Semoga hasil dari Tugas Akhir ini dapat memberikan manfaat bagi semua pihak yang berkepentingan.

Wassalamu'alaikum warahmatullahi wabarakatuh.

Surabaya, 22 Agustus 2021

**Penulis**

## DAFTAR ISI

| Contents  | Halaman |
|---|---------|
| HALAMAN JUDUL .....   | i       |
| TITLE PAGE .....  | iii     |
| LEMBAR PENGESAHAN .....   | v       |
| ABSTRAK.....  | vii     |
| <i>ABSTRACT</i> .....   | ix      |
| KATA PENGANTAR .....  | xi      |
| DAFTAR ISI .....  | xiii    |
| DAFTAR GAMBAR.....  | xv      |
| DAFTAR TABEL.....   | xvii    |
| DAFTAR LAMPIRAN.....  | xix     |
| PENDAHULUAN .....   | 1       |
| 1.1 Latar Belakang.....   | 1       |
| 1.2 Rumusan Masalah.....  | 5       |
| 1.3 Batasan Masalah .....   | 5       |
| 1.4 Tujuan.....   | 6       |
| 1.5 Manfaat.....  | 6       |
| 1.6 Sistematika Penulisan Tugas Akhir .....                                     | 6       |
| TINJAUAN PUSTAKA .....  | 9       |
| 2.1 Penelitian Terdahulu.....   | 9       |
| 2.2 <i>Work From Home (WFH)</i> .....   | 10      |
| 2.3 Analisis Media Sosial .....   | 14      |
| 2.4 Analisis Sentimen .....   | 14      |
| 2.5 <i>Bidirectional Encoder Representations from Transformers (BERT)</i> ..... | 15      |
| 2.5.1 Masked Language Modelling (MLM)....                                       | 18      |
| 2.5.2 <i>Next Sentence Prediction (NSP)</i> .....                               | 19      |

|                                    |  |    |
|------------------------------------|--|----|
| 2.6                                | <i>Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)</i> .... | 20 |
| 2.7                                | IndoBERT .....   | 21 |
| 2.8                                | <i>Confusion Matrix</i> .....  | 21 |
| METODOLOGI PENELITIAN .....        |  | 25 |
| 3.1                                | Studi Literatur .....  | 25 |
| 3.2                                | Pengumpulan Data .....   | 25 |
| 3.3                                | Praproses Data .....   | 25 |
| 3.4                                | Pelabelan Data .....   | 26 |
| 3.5                                | Implementasi IndoBERT .....  | 26 |
| 3.6                                | Uji Coba dan Evaluasi .....  | 26 |
| 3.7                                | Dokumentasi .....  | 26 |
| PERANCANGAN DAN IMPLEMENTASI ..... |  | 29 |
| 4.1                                | Akuisisi Data .....  | 29 |
| 4.2                                | <i>Pre-Processing</i> .....  | 30 |
| 4.2.1                              | Pembersihan Data .....   | 30 |
| 4.2.2                              | Menghapus Data Tweets Bahasa Asing .   | 32 |
| 4.3                                | Pelabelan Data .....   | 34 |
| 4.4                                | Split Data .....   | 35 |
| 4.5                                | Implementasi IndoBERT .....  | 35 |
| UJI COBA DAN EVALUASI .....        |  | 43 |
| 5.1                                | Data Uji Coba .....  | 43 |
| 5.2                                | Hasil Uji Coba .....   | 43 |
| PENUTUP .....                      |  | 49 |
| 6.1                                | Kesimpulan .....   | 49 |
| 6.2                                | Saran .....  | 49 |
| DAFTAR PUSTAKA .....               |  | 51 |
| LAMPIRAN .....                     |  | 55 |
| BIODATA PENULIS .....              |  | 79 |



## DAFTAR GAMBAR

Halaman

|  |    |
|--|----|
| <b>Gambar 2. 1</b> Arsitektur Transformers (Vashwani dkk, 2017) .....                            | 16 |
| <b>Gambar 2. 2</b> Representasi Input BERT (Devlin dkk, 2018) .....                              | 17 |
| <b>Gambar 2. 3</b> Ilustrasi Tahapan Pre-training dan Fine-tuning (Devlin dkk, 2018) .....       | 18 |
| <b>Gambar 2. 4</b> Proses Masked Language Modelling (MLM) .....                                  | 19 |
| <b>Gambar 2. 5</b> Ilustrasi <i>fine-tuning</i> BERT untuk Klasifikasi (Devlin dkk, 2018) .....  | 20 |
| <b>Gambar 3. 1</b> Diagram Alir Metode Penelitian.....   | 27 |
| <b>Gambar 4. 1</b> <i>WordCloud</i> .....  | 33 |
| <b>Gambar 4. 2</b> Pembagian Data yang Telah Dilabeli .....                                      | 35 |
| <b>Gambar 4. 3</b> <i>Vocabulary</i> IndoBERT .....  | 36 |
| <b>Gambar 4. 4</b> Ilustrasi <i>Fine-tuning</i> IndoBERT untuk Klasifikasi .....                 | 38 |
| <b>Gambar 4. 5</b> Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 5 Epoch.....    | 40 |
| <b>Gambar 4. 6</b> Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 10 Epoch.....   | 40 |
| <b>Gambar 4. 7</b> Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 15 Epoch.....   | 40 |
| <b>Gambar 4. 8</b> Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 5 Epoch.....    | 41 |
| <b>Gambar 4. 9</b> Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 10 Epoch.....   | 41 |
| <b>Gambar 4. 10</b> Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 15 Epoch ..... | 42 |

|  |    |
|--|----|
| <b>Gambar 5. 2</b> Diagram Confusion Matrix Model dengan Batch size 16 dan 5 epoch ..... | 44 |
| <b>Gambar 5. 3</b> Performansi Model dengan Batch size 16 dan 5 epoch .....              | 44 |
| <b>Gambar 5. 4</b> Diagram Confusion Matrix dengan Batch size 16 dan 10 epoch .....      | 44 |
| <b>Gambar 5. 5</b> Performansi Model dengan batch size 16 dan 10 epoch .....             | 45 |
| <b>Gambar 5. 6</b> Diagram Confusion Matrix dengan Batch size 16 dan 15 epoch .....      | 45 |
| <b>Gambar 5. 7</b> Performansi Model dengan batch size 16 dan 15 epoch .....             | 45 |
| <b>Gambar 5. 8</b> Diagram Confusion Matrix dengan Batch size 32 dan 5 epoch .....       | 46 |
| <b>Gambar 5. 9</b> Performansi Model dengan batch size 32 dan 5 epoch .....              | 46 |
| <b>Gambar 5. 10</b> Diagram Confusion Matrix dengan Batch size 16 dan 10 epoch .....     | 46 |
| <b>Gambar 5. 11</b> Performansi Model dengan batch size 32 dan 10 epoch.....             | 47 |
| <b>Gambar 5. 12</b> Diagram Confusion Matrix dengan Batch size 32 dan 15 epoch .....     | 47 |
| <b>Gambar 5. 13</b> Performansi Model dengan batch size 32 dan 15 epoch.....             | 47 |

## DAFTAR TABEL

|  | Halaman |
|--|---------|
| <b>Tabel 2. 1</b> <i>Confusion Matrix</i> .....    | 22      |
| <b>Tabel 4. 1</b> Contoh Hasil Akuisisi Data.....  | 30      |
| <b>Tabel 4. 2</b> Tahapan Pembersihan Dataset..... | 31      |
| <b>Tabel 4. 3</b> Contoh Pelabelan Tweets .....    | 34      |
| <b>Tabel 4. 4</b> Proses Tokenisasi IndoBERT.....  | 37      |
| <b>Tabel 4. 5</b> Hyperparameter IndoBERT .....    | 39      |
| <b>Tabel 4. 6</b> Data Tweets Salah Kelas .....    | 42      |
| <b>Tabel 5. 1</b> Hasil Klasifikasi Sentimen ..... | 48      |



## DAFTAR LAMPIRAN

|  | Halaman |
|--|---------|
| <b>Lampiran 1</b> Hasil Uji Coba Menggunakan Data Test .....       | 55      |
| <b>Lampiran 2</b> Klasifikasi data <i>tweets</i> menggunakan model | 68      |
| <b>Lampiran 3</b> Data <i>Tweets</i> .....                         | 73      |



## BAB I PENDAHULUAN

Pada bab ini berisi gambaran umum dari penulisan Tugas Akhir yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, serta sistematika penulisan dokumentasi Tugas Akhir ini.

### 1.1 Latar Belakang

*Severe Acute Respiratory Syndrom Coronaavirus 2* (SARS-COV-2) atau yang dikenal dengan COVID-19 ditemukan pertama kali di Wuhan, China pada tanggal 1 Desember 2019 (Huang dkk, 2020) dan menyebar ke 219 negara per akhir Februari 2021<sup>1</sup>. Penyebaran COVID-19 secara global membuat Organisasi Kesehatan Dunia (WHO) menyatakan kondisi ini sebagai pandemi COVID-19. Sebanyak 112.654.146 kasus COVID-19 yang telah terkonfirmasi di seluruh dunia sejak awal kemunculannya hingga Februari 2021. Dari 112,7 juta kasus tersebut, 88.239.672 pasien dinyatakan sembuh dan 2.496.749 dinyatakan meninggal dunia. Kasus COVID-19 di Indonesia pertama kali terjadi pada tanggal 2 Maret 2020 dan hingga akhir Februari 2021 mencapai 1.298.608 kasus<sup>2</sup>. WHO menyatakan bahwa *social distancing* sangat diperlukan untuk mengurangi penyebaran virus ini.

Beberapa negara melakukan *lockdown* agar lebih mudah dalam melakukan pengontrolan aktivitas masyarakat dalam rangka penerapan *social distancing* untuk mencegah penyebaran virus COVID-19 agar tidak semakin meluas (Walker dkk, 2020). Pelaksanaan *social distancing* di Indonesia tidak dilakukan dengan melakukan *lockdown* seperti di Tiongkok, Inggris, Malaysia, Thailand,

---

<sup>1</sup>COVID-19 *CORONAVIRUS PANDEMIC*. Tersedia: [worldometers.info/coronavirus](https://worldometers.info/coronavirus). (diakses 24 Februari 2021)

<sup>2</sup>[covid19.go.id](https://covid19.go.id) (diakses 24 Februari 2021)

dan Vietnam. Berdasarkan Peraturan Pemerintah No. 21 Tahun 2020, Pemerintah Republik Indonesia memberlakukan Pembatasan Sosial Berskala Besar (PSBB) yang dilakukan di beberapa Kota/Kabupaten Dengan adanya penerapan PSBB yang dilakukan terdapat kebijakan-kebijakan lain dari pemerintah daerah seperti penerapan jam malam, pembatasan jam operasional jumlah penumpang transportasi umum, dan pembatasan jumlah pengunjung dalam suatu ruangan, baik pertokoan, pasar pusat perbelanjaan, maupun perkanatoran. Pada 11 Januari 2021, pemerintah kemudian mengeluarkan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) merujuk pada Instruksi Menteri Dalam Negeri No.1 tahun 2021. Pada PPKM pembatasan kegiatan masyarakat lebih berfokus pada beberapa sektor, yaitu sektor perkantoran, kegiatan belajar mengajar, restoran dan tempat makan, mall atau pusat perbelanjaan dan tempat ibadah. Dengan diberlakukannya PPKM mengharuskan tempat kerja/perkantoran menerapkan *Work From Home* (WFH) sebesar 75% (tujuh puluh lima persen) dan *Work from Office* (WFO) sebesar 25% (dua puluh lima persen) dengan memberlakukan protokol kesehatan secara lebih ketat.

*Work From Home* (WFH) merupakan salah satu bentuk pengaplikasian dari *home-based telecommuting work* atau *telework*, dimana teknologi telekomunikasi dimanfaatkan oleh pekerja sehingga memungkinkan untuk melakukan tugas pekerjaannya dari jarak jauh (Ye, 2012). Dalam dua dekade terakhir *telework* mulai banyak dilakukan seiring dengan berkembangnya teknologi komunikasi dan informasi dan beberapa perusahaan dan bidang pekerjaan juga sudah melaksanakan kegiatan bekerja dari rumah sejak masa sebelum adanya pandemi COVID-19. Dampak positif dari *telecommuting work* untuk karyawan adalah mengurangi biaya transportasi ke tempat kerja, terhindar dari kemungkinan terjebak macet, waktu



kerja yang lebih fleksibel, serta dapat menyeimbangkan antara pekerjaan dengan kehidupan di luar kantor (Ansong dkk, 2018). Keuntungan dari *telework* juga bisa dirasakan oleh perusahaan karena dapat mengurangi biaya sewa tempat dan operasional kantor.

Sistem WFH tidak selalu cocok untuk setiap orang. Selain memberikan efek positif bagi karyawan dan perusahaan, pelaksanaan WFH juga memiliki tantangan tersendiri. Adanya keterbatasan mengenai sarana dan prasarana dalam pekerjaan, seperti perangkat teknologi dan akses internet bisa menjadi penghambat dalam melakukan WFH. Selain keterbatasan teknologi, terdapat faktor lain yaitu faktor psikologi dan fisik. Sistem bekerja di mana saja dapat dikaitkan dengan efek negatif yang dialami yaitu dimana pekerja akan merasa terkurung dengan koneksi sosial, perasaan kurang dihormati dalam organisasi sehingga karyawan harus secara aktif mengelola citra mereka dihadapan karyawan lain dan organisasi untuk menghilangkan kecurigaan rekan kerja mereka dan atasan bahwa mereka bermalas-malasan dengan membuat diri mereka selalu siaga didepan perangkat teknologi sepanjang waktu. Hal ini mengakibatkan dengan ketidakpuasan dengan pekerjaan hingga mempengaruhi kesehatan seseorang (Wang dkk, 2020).

Pelaksanaan WFH pada masa sebelum pandemi tentunya berbeda dengan ketika terjadi pandemi. Faktor lain seperti adanya PPKM dimana hampir seluruh kegiatan masyarakat dibatasi bisa mempengaruhi produktivitas para pekerja. Pekerja yang mampu mengontrol ritme kerja mereka dan kurang suka melakukan interaksi tatap muka mampu menjadi pekerja jarak jauh yang baik sedangkan beberapa pekerja yang tidak mampu melakukan WFH berpotensi mengalami *burnout* yang diakibatkan oleh gangguan kerja di rumah, *work-home interference* (WHI) (Delanoije dkk, 2018). Polemik yang terjadi dengan adanya

pelaksanaan WFH kerap diungkapkan oleh masyarakat ke dalam media sosial, salah satunya Twitter. Twitter merupakan salah satu jejaring media sosial *microblogging* yang banyak digunakan oleh peneliti untuk menganalisis opini masyarakat akan peristiwa atau topik tertentu (Li dkk, 2019). Pada penelitian ini dilakukan analisis sentimen masyarakat terhadap pelaksanaan WFH yang disampaikan melalui media social Twitter. Data *tweets* yang dikumpulkan kemudian dianalisa menggunakan pendekatan berbasis *machine learning*.

Beberapa metode *machine learning* yang sering digunakan untuk analisis sentimen antara lain Naïve Bayes, Linear Regression, dan Support Vector Machine (SVM). Deep Learning merupakan metode dari *machine learning* yang juga mampu untuk melakukan analisis sentimen. Metode *deep learning* membutuhkan waktu yang lebih lama pada saat training, tetapi kelebihan metode *deep learning* ialah dapat membuat model mempelajari konteks dari kata. Model pembelajaran *Bidirectional Encoder Representations from Transformers* (BERT) merupakan model yang sedang populer di kalangan peneliti. Tidak seperti model bahasa lainnya, BERT dibuat sebagai *pre-trained model* yang sudah di *training* secara *deep bidirectional* dari data teks yang tidak berlabel dengan cara menggabungkan konteks sisi kiri dan kanan layer. Dengan cara ini, model BERT dapat diadaptasi dengan tugas *machine learning* lain seperti klasifikasi, *question-answering* dengan hanya menambahkan satu layer saja (Devlin dkk, 2018). BERT pertama kali dikenalkan oleh J. Devlin dkk pada tahun 2018 dan digunakan oleh Google untuk *Search Engine Optimization* agar hasil pencarian sesuai dengan konteks yang diinput oleh *user*. Hingga kini model *pre-trained* BERT sudah tersedia dalam berbagai bahasa termasuk bahasa Indonesia. IndoBERT merupakan

*pre-trained* model bahasa Indonesia berdasarkan BERT (Wilie dkk, 2020).

Sejauh ini masih belum banyak penelitian mengenai analisis sentimen maupun klasifikasi teks bahasa Indonesia menggunakan BERT. Maka pada penelitian ini dilakukan analisis sentimen masyarakat dalam pelaksanaan WFH pada masa pandemi COVID-19 menggunakan *pre-trained model Bidirectional Encoder Representations from Transformers* (BERT) berbahasa Indonesia, IndoBERT dari data media sosial Twitter. Dari hasil penelitian ini diharapkan dapat mengetahui kecenderungan sentimen masyarakat dalam pelaksanaan WFH serta mengetahui kinerja model IndoBERT dalam analisis sentimen pelaksanaan WFH di Indonesia.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang, rumusan masalah dari penelitian tugas akhir ini adalah:

1. Bagaimana kecenderungan sentimen masyarakat dalam pelaksanaan *work from home* (WFH) di Indonesia pada masa pandemi COVID-19?
2. Bagaimana tingkat akurasi dan kinerja IndoBERT dalam melakukan analisis sentimen?

## **1.3 Batasan Masalah**

Berdasarkan permasalahan yang disebutkan di atas, batasan masalah pada penelitian tugas akhir ini adalah sebagai berikut.

1. Data yang digunakan adalah data *tweets* masyarakat Indonesia ketika pandemi COVID-19 dengan topik *work from home*.
2. Model yang digunakan pada penelitian ini adalah *Bidirectional Encoder Representations from Transformer* (BERT) bahasa Indonesia, IndoBERT.

#### **1.4 Tujuan**

Tujuan dari pengerjaan penelitian tugas akhir ini adalah sebagai berikut.

1. Mengetahui kecenderungan opini masyarakat mengenai pelaksanaan *work from home* di Indonesia pada saat pandemi COVID-19.
2. Mengetahui tingkat akurasi dan kinerja IndoBERT dalam menganalisis sentimen masyarakat mengenai pelaksanaan *work from home* di Indonesia.

#### **1.5 Manfaat**

Manfaat yang diharapkan dapat diperoleh dari tugas akhir ini antara lain.

1. Bagi perusahaan, dapat dimanfaatkan sebagai pertimbangan kelanjutan pelaksanaan *work from home* pasca pandemi COVID-19.
2. Bagi pembaca, dapat digunakan sebagai referensi baru untuk mengetahui performansi IndoBERT dalam melakukan analisis sentiment.

#### **1.6 Sistematika Penulisan Tugas Akhir**

Sistematika penulisan dalam laporan Tugas Akhir ini adalah sebagai berikut :

1. **BAB I PENDAHULUAN**  
Bab ini berisi tentang gambaran umum dari penulisan Tugas Akhir yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.
2. **BAB II TINJAUAN PUSTAKA**  
Bab ini berisi tentang materi-materi yang mendukung penelitian tugas akhir ini, antara lain penelitian terdahulu, *Sentiment Analysis*, BERT, dan IndoBERT.
3. **BAB III METODE PENELITIAN**

Pada bab ini dibahas tentang langkah-langkah dan metode yang digunakan untuk menyelesaikan penelitian pada Tugas Akhir ini.

4. BAB IV PERANCANGAN DAN IMPLEMENTASI

Pada bab ini dijelaskan mengenai perancangan dan implementasi penelitian tugas akhir dari tahap awal yaitu akuisisi data hingga tahap pembelajaran model.

5. BAB V UJI COBA DAN EVALUASI

Bab ini dijelaskan mengenai hasil uji coba model terhadap data uji dan evaluasi model menggunakan *confusion matrix*. Selain itu model juga akan diterapkan untuk melakukan klasifikasi sentimen terhadap data baru yang belum berlabel.

6. BAB VI PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari Analisa hasil dan pembahasan sebelumnya serta diberikan saran jika penelitian ini dikembangkan



## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam tinjauan pustaka, akan dijelaskan mengenai penelitian terdahulu yang berkaitan dengan penelitian Tugas Akhir ini, dasar teori mengenai analisis sentimen, *pre-trained model Bidirectional Representations of Transformers (BERT)*, dan IndoBERT.

#### **2.1 Penelitian Terdahulu**

Penelitian mengenai analisis sentimen sudah sering dilakukan pada berbagai macam bidang dan berbagai metode. Suvarna Kanakaraddi dkk (Kanakaraddi dkk, 2020) melakukan penelitian untuk membandingkan hasil sentimen analisis menggunakan algoritma *machine learning* yang berbeda. Peneliti menggunakan data training sebanyak 6.000 data tweet dan 2000 data tweet sebagai data test menggunakan *Support Vector Machine (SVM)*, *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Dari hasil penelitian diketahui bahwa analisis sentimen menggunakan SVM memiliki akurasi paling tinggi yaitu sebesar 79%.

Penelitian analisis sentimen menggunakan BERT pernah dilakukan oleh Usman Naseem dkk pada tahun 2020 (Naseem dkk, 2020). Dalam penelitian ini peneliti melakukan analisis sentimen dari data twitter dengan mengombinasikan beberapa word embedding yang berbeda, yaitu BERT, GloVe, Lexicon, dan POS untuk pengembangan  $DICE_T$  (*Transformers-based Deep Intelligence Contextual Embedding*) (Naseem dkk, 2020). Hasil dari penelitian ini menunjukkan bahwa performansi  $DICE_T$  lebih baik dari metode analisis sentimen lain yang pernah diterapkan pada dataset yang sama. Selain itu

DICE<sub>T</sub> juga memiliki performansi paling baik jika menerapkan BERT dalam kombinasi *word embedding* yang dipakai.

Penelitian lainnya menggunakan BERT<sub>BASE</sub> model dilakukan oleh Zhancheng Ren dkk untuk mendeteksi karakter seseorang dari data teks Twitter (Ren dkk, 2021). Data dibagi menjadi data training, data testing, dan data validation dengan rasio 6:2:2. Hasil menunjukkan bahwa penggunaan BERT<sub>BASE</sub> model untuk klasifikasi mendapatkan hasil yang lebih baik jika dibandingkan dengan model traditional neural network dengan peningkatan akurasi rata-rata sebesar 6,91% pada dataset MBTI, dan 6,04% pada dataset Big Five.

Penelitian mengenai klasifikasi data teks bahasa Indonesia menggunakan pre-trained BERT pernah dilakukan Putra dan Purwarianti (Putra dan Purwarianti, 2020). Dalam penelitian tersebut, peneliti menggabungkan data bahasa Inggris dan Indonesia untuk melakukan klasifikasi kalimat bahasa Indonesia yang mengandung ujaran kebencian menggunakan Multilingual BERT (M-BERT) dan mendapatkan akurasi sebesar 89,52 %. Dari hasil penelitian diketahui penambahan data teks bahasa Inggris dapat meningkatkan performansi model M-BERT pada data teks bahasa Indonesia.

Pada tahun 2020, Wilie dkk mempublikasikan hasil penelitian IndoBERT, monolingual pre-trained model BERT untuk bahasa Indonesia. Hasil dari penelitian menunjukkan model monolingual memiliki performansi yang lebih baik pada data teks bahasa formal maupun *colloquial* daripada multilingual model, walaupun ukuran IndoBERT lebih kecil 40% hingga 60% (Wilie dkk, 2020).

## **2.2 Work From Home (WFH)**



Perkembangan teknologi informasi dan komunikasi mempengaruhi banyak hal di dalam kehidupan manusia. Teknologi dimanfaatkan oleh banyak orang untuk membantu kegiatan yang dilakukan sehari-hari maupun pekerjaan mereka. *Work from home* atau bekerja dari rumah merupakan salah satu bentuk dari *telecommuting work* atau *telework*, dimana pekerja atau karyawan dapat melakukan pekerjaan dari mana saja dengan memanfaatkan teknologi informasi dan telekomunikasi (Ye, 2012). Maraknya pengaplikasian sistem bekerja dari mana saja dalam dua dekade terakhir (Ye, 2012) tidak lepas dari pengaruh internet sehingga semakin mudahnya orang berkomunikasi dan bertukar informasi, serta dukungan dari organisasi atau perusahaan yang memberikan kesempatan karyawannya untuk melakukan pekerjaan dari manapun

### **2.1.1 Keuntungan dan Tantangan Work From Home**

*Work from home* saat ini semakin diminati karena adanya 3 aspek besar yang mendasari yaitu karyawan, organisasi atau perusahaan serta lingkungan. Dari sisi karyawan, *work from home* diminati oleh pekerja dengan berbagai alasan, diantaranya yaitu menghemat pengeluaran untuk transportasi serta menghemat waktu yang diperlukan untuk berangkat ke tempat kerja. Selain itu, melakukan pekerjaan dari rumah juga erat kaitannya dengan fleksibilitas yang mungkin dilakukan oleh para pekerja tersebut (Wang dkk, 2020). *Work from home* juga dirasa mampu untuk meningkatkan hubungan antara pekerjaan dengan keluarga sehingga pekerja mendapatkan *work-life balance*.

Sistem *work from home* juga diminati oleh beberapa organisasi atau perusahaan. Organisasi atau perusahaan menilai bahwa adanya peningkatan produktifitas pekerja, ketika pekerja melakukan remote working. Selain itu sistem bekerja dari mana saja membantu perusahaan

memangkas biaya operasional perusahaan, seperti sewa tempat, biaya listrik, air, dan lainnya. *Work from home* juga memiliki dampak baik untuk lingkungan antara lain mengurangi kemacetan serta mengurangi emisi karbon yang dikeluarkan oleh kendaraan bermotor.

Sistem bekerja dari mana saja bukanlah hal baru di era digital, akan tetapi sistem ini memiliki tantangan sendiri seperti yang diungkapkan dalam penelitian yang dilakukan oleh Delanoije dan kawan-kawan. Delanoije mengatakan dalam pelaksanaan *work from home* ditemukan adanya kemungkinan gangguan kerja di rumah, *work from home interference* (WHI) (Delanoije dkk, 2019) yang berpotensi membuat pekerja mengalami *burnout*. Selain itu *work from home* berpotensi mengakibatkan adanya efek negatif seperti isolasi psikologis dan isolasi fisik. Isolasi psikologis merupakan perasaan terputusnya seseorang dari lingkungan sosialnya (Wang dkk, 2020). Isolasi psikologis memungkinkan karyawan merasa tidak puas akan hasil kerjanya dan berpotensi untuk mengganggu kesehatan karyawan tersebut. Isolasi fisik yang dialami karyawan berpotensi membuat karyawan merasa asing dan tidak dihormati terhadap satu sama lain. Hal ini dapat memicu adanya kecurigaan antar karyawan sehingga merasa perlu untuk mengelola citra mereka dengan senantiasa selalu aktif di depan perangkat komunikasinya sepanjang waktu (Wang dkk, 2020).

Teknologi informasi dan komunikasi memiliki peran vital dalam pelaksanaan *work from home*. Dalam penelitian yang dilakukan Daniel dan kawan-kawan pada negara dengan pendapatan perkapita yang tinggi, terdapat satu dari tiga pekerjaan yang dapat dikerjakan dari mana saja. Sedangkan pada negara berkembang dengan pendapatan perkapita yang kecil hanya terdapat satu dari dua puluh enam pekerjaan yang memungkinkan untuk dilakukan dari rumah yang diakibatkan oleh terbatasnya akses internet

yang memiliki peran penting dalam sistem ini (Delanoije dkk, 2020). Pada penerapannya sistem bekerja dari rumah juga memiliki ketergantungan terhadap sumber daya keuangan serta organisasi atau perusahaan yang memiliki komputer serta peralatan komunikasi dan perangkat lunak yang memadai. Hal inilah yang mengakibatkan banyak usaha kecil masih belum bisa menerapkan *work from home*.

### **2.1.2 Work From Home di Era Pandemi COVID-19**

Pandemi COVID-19 membuat perubahan besar-besaran yang tidak terduga hampir di semua lapisan masyarakat. Hal ini mengakibatkan adanya pergeseran mengenai sistem bekerja secara tatap muka yang dianut oleh sebagian besar perusahaan atau organisasi, berubah menjadi sistem bekerja dari rumah membuat pekerja yang terdampak harus mampu beradaptasi dengan cepat melaksanakan sistem ini. Sistem bekerja dari rumah membantu pemerintah untuk menerapkan *social distancing* sesuai anjuran WHO. *Work from home* mampu meminimalisir potensi penyebaran virus COVID-19 karena berkurangnya mobilitas pekerja.

Indonesia merupakan negara berkembang, *work from home* disambut baik oleh sebagian kalangan pekerja, namun masih ada 36% dari penduduk Indonesia yang belum melek teknologi (Setyawan dan Lestari, 2020). Tantangan lain yang dihadapi adalah ketika karyawan melakukan WFH sangat mudah sekali untuk diinterupsi oleh pekerjaan rumah sehari-hari. Hal ini menjadi rawan ketika karyawan tidak memiliki keterikatan dengan perusahaan karena tidak adanya pihak lain yang mengawasi seperti atasan dan rekan kerja (Setyawan dan Lestari, 2020). Selain itu, masalah psikologis juga menjadi perhatian. Adanya PPKM membuat sebagian besar karyawan harus tetap berada di rumah.

Meskipun terdapat tantangan yang dihadapi ketika melakukan *work from home* beberapa penelitian menunjukkan bahwa sistem *work from home* disambut positif oleh banyak pekerja. Sistem bekerja *work from home* yang semakin dianggap normal ketika pandemi COVID-19 terjadi membuat banyak pihak sadar bahwa hal ini tidak terlalu buruk untuk diterapkan di perusahaan (Dewi dan Ardiasi, 2020). WFH memiliki dampak yang berbeda-beda antara satu pekerja dengan yang lainnya sehingga perusahaan tetap membutuhkan strategi khusus untuk menangani masalah produktivitas karyawan.

### **2.3 Analisis Media Sosial**

Pengaruh perkembangan media social tidak bisa lepas dari perkembangan teknologi *mobile* dan internet. Penggunaan media sosial digunakan oleh para pengguna internet sebagai sarana berinteraksi, berkomunikasi dan berbagi informasi antara satu dengan yang lainnya. Platform media sosial sangat bermanfaat untuk berbagai pihak selain untuk pengguna dan perusahaan pemilik aplikasi tersebut, antara lain untuk dunia bisnis, pemerintahan, peneliti, dan organisasi nirlaba. Platform media sosial dapat digunakan sebagai *source* untuk mendapatkan data opini masyarakat untuk melakukan analisis.

### **2.4 Analisis Sentimen**

Analisis sentimen atau *opinion mining* adalah salah satu bentuk dari pengaplikasian *text mining* yang digunakan untuk mengetahui opini dari sekumpulan data tekstual mengenai peristiwa atau topik tertentu (Li dkk, 2019). Analisis sentimen merupakan kunci dari pengembangan *artificial intelligence* dari semua penelitian yang terkait dengan penggalian informasi dari data teks (Cambria dkk, 2017).

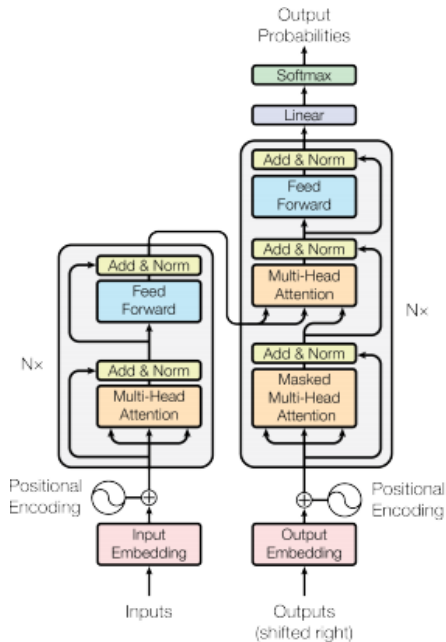
Fungsi utama dari analisis sentimen pada media sosial adalah untuk menggali informasi, mencari makna, serta opini pengguna dari data *posting*, baik berupa *caption*, *post*, komentar, *likes* yang diunggah. Dalam bisnis, penggunaan analisis opini dan sentimen sangat dibutuhkan dalam *role business intelligence* untuk mengetahui pendapat pelanggan mengenai produk atau jasa yang ditawarkan, sehingga dapat digunakan sebagai pertimbangan dalam mengembangkan produk. Tidak hanya itu, analisis setimen juga dapat membantu dalam menganalisa produk kompetitor. Dalam sistem pemerintahan, analisis sentimen dapat membantu pemerintah untuk mengatahui opini yang dikeluarkan masyarakat mengenai kebijakan yang dikeluarkan oleh pemerintah, serta memungkinkan menangkap data secara real-time untuk melihat perubahan preferensi publik dan menggabungkannya dengan sentimen yang dikeluarkan sebelumnya .

Proses penggalian sentimen atau opini publik terutama pada media sosial seperti Facebook, Twitter, atau kolom komentar memiliki tantangan tersendiri karena sifat dari media sosial tersebut yang memberi batasan karakter untuk ditulis ke dalam tiap postingan. Ekstraksi data tekstual menjadi masalah yang harus dihadapi dalam menggali sentimen karena data yang ada bersifat dinamis, terdapat penggunaan kata yang tidak baku, juga format data yang bervariasi.

## **2.5 *Bidirectional Encoder Representations from Transformers (BERT)***

*Bidirectional Encoder Representations from Transformers (BERT)* merupakan model representasi bahasa terlatih yang pertama kali dipopulerkan oleh Google pada 11 Oktober 2018. Tidak seperti model bahasa lainnya, BERT dibuat sebagai *pre-trained model* atau model yang

sudah dilatih secara *bidirectional* dari data teks yang tidak berlabel dengan cara menggabungkan konteks dari sisi kiri dan kanan layer. Dengan cara ini model BERT dapat *fine-tuning* dengan tambahan satu layer saja (Devlin dkk, 2020).

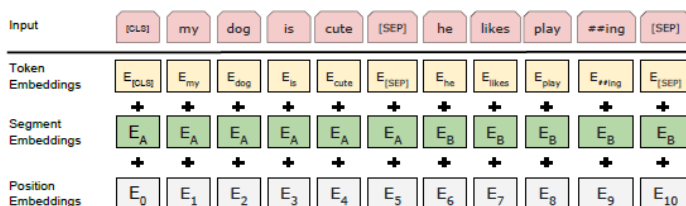


**Gambar 2.1** Arsitektur Transformers (Vashwani dkk, 2017)

Arsitektur BERT terdiri dari *multi-layer bidirectional* Transformer encoder. Transformer adalah sebuah mekanisme attention yang mempelajari hubungan kontekstual diantara kata atau sub kata dalam teks (Vashwani dkk, 2017). Transformer menggunakan mekanisme *self-attention* untuk memahami representasi

input dan output. Transformer memiliki 2 mekanisme terpisah, yaitu encoder untuk membaca teks input dan decoder yang menentukan prediksi. Tiap layer encoder memiliki dua sub layer, yang pertama adalah multi-head self-attention mechanism, dan yang kedua *fully connected feed forward network*. Pada encoder, setiap input akan melewati self-attention layer. Layer ini membantu encoder untuk melihat kata lain yang ada pada kalimat. Output dari attention layer kemudian digunakan sebagai input ke feed-forward neural network.

BERT menggunakan WordPiece embeddings dengan 30.000 token pada *vocabulary*. WordPiece embedding pertama kali dikembangkan untuk mengatasi masalah segmentasi bahasa Jepang dan Korea pada Google *speech recognition system* (Wu dkk, 2016). Segmentasi akan mengakibatkan banyak kata yang tidak ada pada *vocabulary*. WordPiece dilatih untuk mempelajari unit kata secara otomatis dari data yang besar tanpa menyebabkan *Out of Vocabulary* (OOV). Kata yang sering digunakan akan tetap disimpan sebagai satu kata yang utuh sedangkan kata yang jarang digunakan akan dipisahkan menjadi sub kata bahkan karakter.

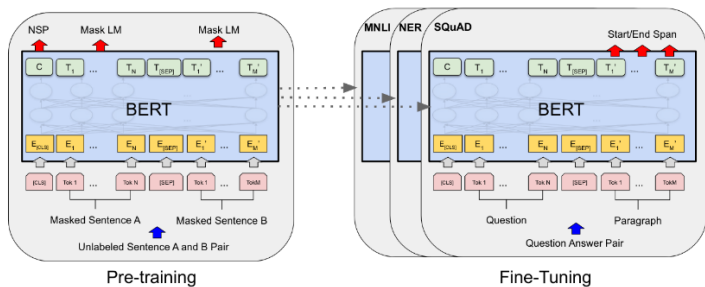


**Gambar 2. 2** Representasi Input BERT (Devlin dkk, 2018)

Token pertama dari setiap *sequence* adalah *special class classification* token ([CLS]). Untuk membedakan

setiap kalimat, ditambahkan token *separate* ([SEP]) pada akhir kalimat. Kemudian ditambahkan segment embedding pada setiap token sebagai pembeda antara kata dari kalimat A dan kata dari kalimat B. Selanjutnya ditambahkan positional embedding untuk menandai posisi token pada kalimat. Input untuk encoder BERT merupakan jumlahan dari Token Embedding, Segment embedding, dan Positional Embedding. Representasi input BERT ditunjukkan pada **Gambar 2.2**.

Terdapat 2 tahap dalam framework BERT, yaitu *pre-training* dan *fine-tuning*. Tahap *pre-training* merupakan tahap pada saat model mempelajari bahasa dan konteksnya dengan melakukan *Mask Language Modelling* (MLM) dan *Next Sentence Prediction* (NSP) secara bersamaan. MLM memungkinkan representasi untuk memadukan konteks kiri dan konteks kanan, yang memungkinkan kita untuk melatih transformator dua arah yang dalam dan NSP yang menggabungkan representasi pasangan *pre-train* teks.



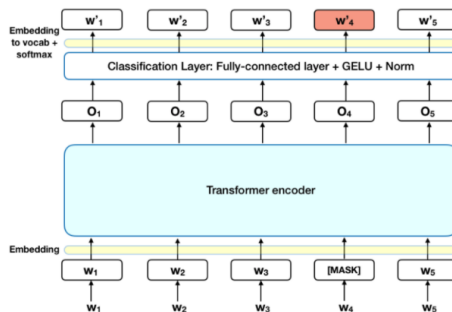
**Gambar 2.3** Ilustrasi Tahapan Pre-training dan Fine-tuning (Devlin dkk, 2018)

### 2.5.1 Masked Language Modelling (MLM)

Pada Masked Language Modelling, 15% kata dari setiap kalimat input dipilih secara acak untuk diganti



dengan token [MASK], kemudian model akan melakukan prediksi id *vocabulary* dari kata sebenarnya berdasarkan konteks kalimat. Untuk melakukan prediksi perlu ditambahkan layer classification di atas output encoder. Selanjutnya, dilakukan perkalian vektor output dengan embedding matriks kemudian mengubahnya menjadi dimensi *vocabulary* dan menghitung probability dari setiap kata di *vocabulary* menggunakan softmax. **Gambar 2.4** menunjukkan proses Masked LM.



**Gambar 2. 4** Proses Masked Language Modelling (MLM)<sup>3</sup>

### 2.5.2 Next Sentence Prediction (NSP)

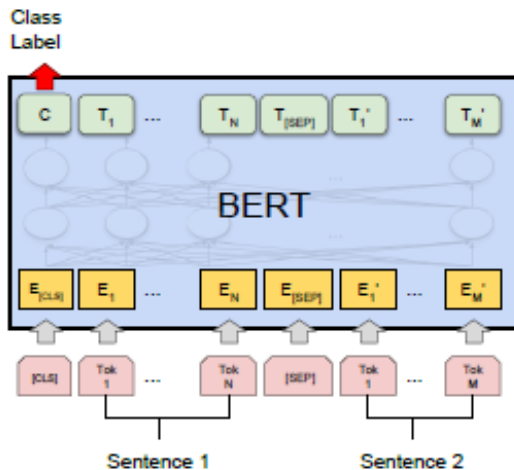
Pada saat training BERT, model menerima pasangan kalimat sebagai input dan dilatih untuk memprediksi kalimat kedua merupakan lanjutan dari kalimat pertama atau bukan. Pada saat training, 50% dari input adalah pasangan kalimat sebenarnya. Sedangkan 50% kalimat lain dipilih secara acak dari corpus sebagai kalimat kedua.

<sup>3</sup> BERT Explained: State of the art language model for NLP. <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270> (diakses pada 14 Agustus 2021).

Output dari pre-training adalah sebuah model BERT (*learned weight*) dan 30.000 vektor embedding. Vektor ini kemudian digunakan untuk merepresentasikan kata sebagai input pada model yang sudah dilatih. Model kemudian mentransformasikan representasi kata berdasarkan kata sekitarnya (konteks kalimat) yang sesuai dengan mengukur *cosine similarity* antara vector input dalam ruang vector.

## 2.6 *Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)*

Untuk melatih sebuah model bahasa, *classifier* perlu dilatih dengan sedikit perubahan pada model BERT selama fase pelatihan (*training*) yang disebut *fine-tuning*. *Fine-tuning* sangat mudah dilakukan karena mekanisme *self-attention* pada Transformer membuat BERT bisa membuat model untuk berbagi tugas, baik pada kalimat tunggal (*single sentence*) atau kalimat berpasangan, dengan menukar masukan dan keluaran yang sesuai.



**Gambar 2. 5** Ilustrasi *fine-tuning* BERT untuk Klasifikasi (Devlin dkk, 2018)

Pada tahap *fine-tuning*, hyperparameter pada saat *pre-training* adalah sama, kecuali batch size, learning rate, dan jumlah epoch. Dropout probability selalu 0.1. Nilai hyperparameter yang optimal untuk setiap *task* pada NLP selalu berbeda, namun model bekerja sangat baik pada saat batch size 16 dan 32, learning rate (Adam)  $5e-5$ ,  $3e-5$ ,  $2e-5$ , dan epoch sebanyak 2,3, dan 4 kali. (Devlin dkk, 2018)

## 2.7 IndoBERT

IndoBERT merupakan pre-trained model berdasarkan BERT untuk bahasa Indonesia. IndoBERT menggunakan dataset Indo4B yang berisi lebih dari 23 GB data teks bahasa Indonesia, terdiri dari 4 Milliar kata baik formal maupun *colloquial* (bahasa sehari-hari) dari berbagai sumber diantaranya media sosial, blog, berita, dan *website*[9]. IndoBERT menggunakan SentencePiece dengan Byte Pair Encoding (BPE) tokenizer sebagai metode pembentukan *vocabulary*. SentencePiece merupakan *language-independent* subword tokenizer dan detokenizer yang didesain untuk neural-based text processing.

Terdapat empat model pada IndoBERT yaitu IndoBERT<sub>BASE</sub>, IndoBERT<sub>LARGE</sub>, IndoBERT-lite<sub>BASE</sub>, IndoBERT-lite<sub>LARGE</sub>. Semua model IndoBERT dilatih menggunakan TPU v3-8 dalam dua fase. Pada fase pertama, model dilatih dengan *maximum sequence length* 128, dan pada fase kedua 512. Dari segi akurasi, IndoBERT<sub>LARGE</sub> memiliki akurasi yang lebih tinggi, namun *memory* yang dibutuhkan sangat besar.

## 2.8 Confusion Matrix

Untuk menguji kualitas dari klasifikasi dapat dilakukan dengan menggunakan *confusion matrix*. *Confusion Matrix* membandingkan klasifikasi hasil pengolahan dari sistem dengan hasil klasifikasi data yang sebenarnya dengan mengukur tingkat akurasi, presisi, dan

*recall*. Akurasi menghitung rasio dari prediksi yang benar dengan total data yang dievaluasi. Presisi menghitung tingkat keakuratan antara data yang diminta dengan jawaban atau hasil yang diberikan oleh sistem. Sedangkan *recall* digunakan untuk menghitung banyaknya data yang diklasifikasikan dengan benar dengan total data yang seharusnya ada dalam kelas tersebut. Berikut adalah contoh *confusion matrix* untuk *multi-class classification*.

**Tabel 2. 1** Contoh *Confusion Matrix* untuk multi-class classification.

| <i>Classes</i>               |   | <i>Predicted Classification</i> |    |    |
|------------------------------|---|---------------------------------|----|----|
|                              |   | A                               | B  | C  |
| <i>Actual Classification</i> | A | TN                              | FP | TN |
|                              | B | FN                              | TP | FN |
|                              | C | TN                              | FP | TN |

Sebagai contoh, prediksi benar terhadap kelas B merupakan kondisi True Positive (TP), sedangkan prediksi salah terhadap kelas B merupakan kondisi False Negative (FN). Kondisi False Positif (FP) terjadi ketika kelas lainnya diprediksi sebagai kelas B, dan True Negative (TN) jika kelas lain tidak diprediksi sebagai kelas B. Rumus untuk menghitung akurasi, presisi, dan recall terhadap kelas  $k$  disajikan sebagai berikut.

$$akurasi_k = \frac{TP_k + TN_k}{TP_k + FP_k + TN_k + FN_k} \quad (1)$$

$$presisi_k = \frac{TP_k}{TP_k + FP_k} \quad (2)$$

$$recall_k = \frac{TP_k}{TP_k + FN_k} \quad (3)$$

Selain akurasi, presisi, dan recall, pada tahap evaluasi juga dihitung F1-Score untuk mengetahui performansi model dengan cara mengkombinasikan nilai presisi dan recall. Nilai 1 menunjukkan bahwa model yang digunakan memiliki performansi yang baik dalam melakukan prediksi, sedangkan nilai 0 menunjukkan model memiliki performansi yang buruk. Pada binary classification, F-1 Score dirumuskan sebagai berikut.

$$F1 - Score = 2 * \frac{presisi * recall}{presisi + recall} \quad (4)$$

Pada kasus multi-class classification, perhitungan F1-Score harus memuat keseluruhan kelas dengan menghitung *Macro F1-Score* dan *Micro F1-Score* (Grandini dkk, 2020).

$$\begin{aligned} & \text{Macro F1 - Score} \\ & = 2 * \left( \frac{\text{Macro Average Precision} * \text{Macro Average Recall}}{\text{Macro Average Precision}^{-1} + \text{Macro Average Recall}^{-1}} \right) \end{aligned} \quad (5)$$

dimana,

$$\text{Macro Average Precision} = \frac{\sum_k^K \text{Precision}_k}{K} \quad (6)$$

$$\text{Macro Average Recall} = \frac{\sum_k^K \text{Recall}_k}{K} \quad (7)$$

$$\text{Micro Average F1 - Score} = \frac{\sum_k^K TP_k}{\text{Grand Total}} \quad (8)$$



## **BAB III**

### **METODOLOGI PENELITIAN**

Pada bab ini dijelaskan mengenai prosedur dan proses pelaksanaan tiap-tiap langkah yang dilakukan dalam menyelesaikan Tugas Akhir

#### **3.1 Studi Literatur**

Pada tahap studi literatur dilakukan kajian kepustakaan terhadap penelitian terkait analisis sentimen dan *pre-train* model *Bidirectional Encoder Representations from Transformer* (BERT) yang didapat dari beberapa buku, jurnal internasional, atau penelitian sebelumnya.

#### **3.2 Pengumpulan Data**

Data yang digunakan pada penelitian ini merupakan data opini masyarakat mengenai WFH selama masa pandemi COVID-19 yang diambil dari media sosial Twitter. Pengambilan data opini mengenai WFH dilakukan dengan cara melakukan *scrapping* data *tweets*. Data opini yang akan dianalisis merupakan *tweet* berbahasa Indonesia.

#### **3.3 Praproses Data**

Tahap selanjutnya yaitu tahap pra-pemrosesan data. Tahap ini bertujuan untuk memproses data mentah agar dapat digunakan sebagai data *training* penelitian. Tahapan yang dilakukan pada pra-pemrosesan data adalah:

a. *Case Folding*

Tahap *case folding* digunakan untuk merubah huruf pada kata-kata yang sudah diuraikan menjadi token diubah menjadi lowercase atau huruf kecil.

b. *Filtering*

Data yang tidak dibutuhkan seperti tanda baca, dan kata yang tidak alfabetis dihilangkan untuk mengurangi ukuran dataset agar proses training berjalan lebih cepat.

c. *Tokenizing*

*Tokenizing* adalah proses memisahkan rangkaian string menjadi potongan string, yang selanjutnya disebut token. Dalam penelitian ini, setiap kalimat maupun frasa akan dipisahkan menjadi sebuah kata berdasarkan *vocabulary* IndoBERT.

### 3.4 Pelabelan Data

Data yang dikumpulkan kemudian dilakukan pelabelan menggunakan TextBlob. Pelabelan data bertujuan agar model dapat belajar sehingga dapat menentukan data opini termasuk dalam kategori sentimen positif, negatif, maupun netral.

### 3.5 Implementasi IndoBERT

Setelah tahapan praproses data, dataset akan diubah menjadi input yang dapat diterima oleh IndoBERT yaitu dalam bentuk vektor representasi kata menggunakan IndoBERT Tokenizer. Kemudian dilakukan proses *fine-tuning* dimana *pre-train* model BERT diadaptasi untuk melakukan klasifikasi sentimen.

### 3.6 Uji Coba dan Evaluasi

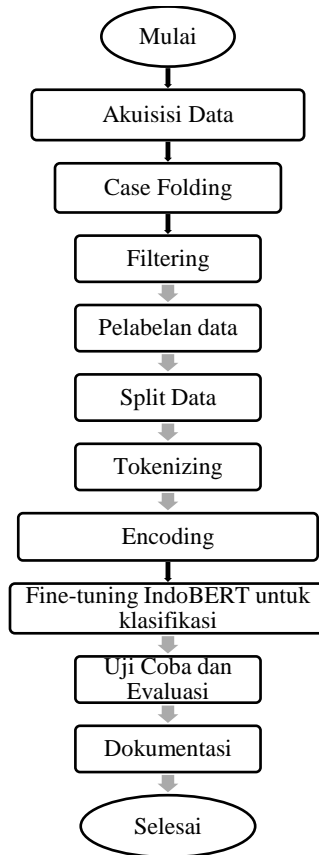
Setelah melalui tahapan implementasi IndoBERT, selanjutnya model di uji dengan menampilkan hasil klasifikasi sentimen dari data test. Klasifikasi sentimen yang dihasilkan kemudian dievaluasi menggunakan *confusion matrix*. Dari hal tersebut dapat dilihat akurasi dan performansi dari IndoBERT dalam melakukan analisis sentimen.

### 3.7 Dokumentasi

Pada tahap akhir dilakukan dokumentasi penelitian dalam bentuk buku tugas akhir dengan susunan yang telah ditentukan. Dokumentasi hasil penelitian ini diharapkan memberikan penjelasan yang baik sehingga dapat dengan



mudah dipahami jika penelitian ini nantinya akan dikembangkan. Gambaran tahapan penelitian tugas akhir ini disajikan sebagai berikut.



**Gambar 3. 1** Diagram Alir Metode Penelitian



## **BAB IV**

### **PERANCANGAN DAN IMPLEMENTASI**

Pada bab ini akan dijelaskan mengenai perancangan, serta implementasi sistem dari awal tahap pengumpulan data hingga implementasi IndoBERT.

#### **4.1 Akuisisi Data**

Tahap awal dimulai dengan mengumpulkan data yang dibutuhkan. Data yang dibutuhkan dalam penelitian tugas akhir ini merupakan data opini pengguna Twitter mengenai pelaksanaan *work from home* pada masa pandemi Covid-19 di Indonesia. Kata kunci yang digunakan adalah 'wfh'. Kata kunci 'wfh' dipilih karena wfh merupakan akronim dari '*work from home*' yang juga sering dipakai pengguna Twitter ketika memberikan opini. Proses pengambilan data dilakukan melalui Twitter *crawler* menggunakan *library tweepy* dengan memfilter pencarian *tweet* bahasa Indonesia dengan atribut username dan Tweets. Dataset ini kemudian disimpan dalam bentuk *file csv* untuk mempermudah proses pengolahan data. Dari hasil pengumpulan data yang dimulai pada tanggal 15 Mei 2021 hingga 30 Juni 2021 didapatkan sebanyak 290.458 data. Supaya data dapat digunakan secara optimal, maka dilakukan pengurangan atribut dengan menghilangkan atribut user karena tidak digunakan dalam proses analisis sentimen. Berikut contoh hasil akuisisi data.

**Tabel 4. 1** Contoh Hasil Akuisisi Data

| <b>user</b>              | <b>Tweet</b>   |
|--------------------------|--|
| DailySocial.id           | b'Selain mengusung spesifikasi yang mumpuni, fitur unggulan HP EliteOne 800 G8 AiO juga mencakup webcam berteknologi face tracking dan sistem noise reduction berbasis AI<br><a href="https://t.co/psW17VJLuW">https://t.co/psW17VJLuW</a> ' |
| Ria                      | b'Oi @FirstMediaCares dari jam 1 internet mati. Saya merasa tertunggu karena sedang WFH. Jadi service delay ke klien. Minta tolong segera diperbaiki jaringannya.'   |
| Athirah Syuhada          | b'Hari tak WFH ni la byk parcel sampai \xf0\x9f\x99\x84'   |
| Rhee                     | b'@chandrazyifaw Wkwkwkw wfh udah setahun parah bget.. Beli baju nya daster aja udah itu mah.. \xf0\x9f\x98\x85 jangan tutup lah kasian bos gue.. Yg gantiin gue si nitaaa.. \xf0\x9f\x91\x8f\xf0\x9f\x91\x8f\xf0\x9f\x98\x81'               |
| Gory from #FinancialGory | b'RT @aidy_syazreef: @ikhwandayne sama weh. siod tertengok tweet @azraeimuhamad nasib la wfh tak kantai menangis.'   |

## 4.2 *Pre-Processing*

Data hasil crawling yang merupakan data mentah harus diolah terlebih dahulu. *Pre-processing* data dilakukan untuk menyiapkan data mentah menjadi data yang lebih bersih agar pemrosesan data lebih efektif dan efisien.

### 4.2.1 **Pembersihan Data**

Pada dataset penelitian masih terdapat simbol, karakter, *hashtag*, *mention*, dan *retweet* yang tidak mengandung opini. Maka dilakukan proses filterasi dengan menghapus karakter atau kata yang tidak diperlukan agar

data tidak terlalu besar dan proses training dapat berjalan lebih efektif. Berikut merupakan tahapan pembersihan dataset.

1. Mengubah representasi *unicode* kedalam bentuk yang mudah dibaca
2. Menghilangkan atribut *retweet* dengan menghapus 'RT'
3. Menghilangkan atribut *mentios* dengan menghapus akun yang di mentions
4. Menghapus hashtag
5. Menghapus emoji
6. Menghapus URL
7. Menghapus seluruh tanda baca
8. Mengubah semua kata menjadi huruf kecil (*case folding*)

Berikut merupakan contoh hasil pembersihan data tweet

**Tabel 4. 2** Tahapan Pembersihan Dataset

| Aktivitas/<br>Kondisi | Hasil   |
|-----------------------|---|
| Tweet Awal            | b'RT @cahyo_semburat: Keislaman Yang Rahasia\n\n#WFH #SYAIR #EMHA \n\x0\x9f\x91\x87\x0\x9f\x91\x87 #Mbahnun \nhttps://t.co/hqJNmLBjuv <a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a> ' |
| Decoding              | RT @cahyo_semburat: Keislaman Yang Rahasia\n\n#WFH #SYAIR #EMHA \n👉👉 #Mbahnun \nhttps://t.co/hqJNmLBjuv <a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a>                                 |
| Menghapus             | @cahyo_semburat: Keislaman Yang   |

|                      |   |
|----------------------|---|
| Retweet              | Rahasia\n\n#WFH #SYAIR<br>#EMHA \n👉👉 #Mbahnun<br>\nhttps://t.co/hqJNmLBjuv<br><a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a>                     |
| Menghapus mentions   | : Keislaman Yang<br>Rahasia\n\n#WFH #SYAIR<br>#EMHA \n👉👉 #Mbahnun<br>\nhttps://t.co/hqJNmLBjuv<br><a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a> |
| Menghapus Hashtag    | : Keislaman Yang Rahasia \n\n👉👉<br>\nhttps://t.co/hqJNmLBjuv<br><a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a>                                   |
| Menghapus Emoji      | : Keislaman Yang Rahasia \n\n<br>\nhttps://t.co/hqJNmLBjuv<br><a href="https://t.co/SPLp5PSPAC">https://t.co/SPLp5PSPAC</a>                                     |
| Menghapus URL        | : Keislaman Yang Rahasia \n\n   |
| Menghapus Tanda Baca | Keislaman Yang Rahasia  |
| Case Folding         | keislaman yang rahasia  |

Setelah didapatkan data yang bersih, kemudian dilakukan penghapusan duplikasi data. Duplikasi data didapat dari tweet asli dan retweet yang sebelumnya tidak dianggap sebagai data duplikat. Kemudian tweet yang tidak berkaitan dengan WFH, seperti tweet ‘keislaman yang rahasia’ juga akan dihapus. Dari 290.457 data, didapatkan sebanyak 88.483 data tweet bersih.

#### 4.2.2 Menghapus Data Tweets Bahasa Asing

Dari hasil akuisisi data diketahui masih terdapat tweet bahasa asing yang dikenali sebagai tweet bahasa Indonesia. **Gambar 4.1** menunjukkan WodCloud yang



### 4.3 Pelabelan Data

Setelah didapatkan data tweet bahasa Indonesia bersih, kemudian dilakukan pelabelan data menggunakan TextBlob. Sebanyak 15.000 data dari 68.538 data tweets bahasa Indonesia. Berikut contoh hasil pelabelan tweets dengan library TextBlob.

**Tabel 4. 3** Contoh Pelabelan Tweets

| <b>Tweets</b>  | <b>Polaritas</b> | <b>Label Sentimen</b> |
|--|------------------|-----------------------|
| oi dari jam 1 internet mati saya merasa tertanggu karena sedang wfh jadi service delay ke klien minta tolong segera diperbaiki jaringannya | -0.2             | negative              |
| wkwkwkw wfh udah setahun parah bget beli baju nya daster aja udah itu mah jangan tutup lah kasian bos gue yg gantiin gue si nitaaa         | -0.6333333333    | negative              |
| gncity kayaknya dari kemarin berasa wfh sih udah puas aku diketawain adek aku cuti tapi depan laptop mulu                                  | 0.6              | positive              |
| i need wfh kerja sambil tiduran  | 0                | neutral               |
| waduh apa dah wfh petugas kppn nya ya senen selasa depan terakhir pencairan sp2d thr   | -0.1             | negative              |



Tweets yang memiliki nilai polaritas kurang dari nol dinilai sebagai sentiment negatif, *tweets* dengan nilai polaritas lebih besar dari nol dinilai memiliki sentiment positif, sedangkan tweet dengan nilai 0 akan dinilai sebagai tweet dengan sentiment netral.

#### 4.4 Split Data

Sebelum melakukan tahap klasifikasi, data yang telah dilabeli dibagi menjadi 3 yaitu data train, data test, dan data validation. Data train digunakan pengembangan model, data test digunakan untuk menguji dan melihat keakuratan model, sedangkan data validation digunakan untuk memvalidasi kinerja model dan meminimalisir overfitting yang sering terjadi. *Scikit learn* merupakan library python yang sering digunakan untuk split data. Pada penelitian ini sebanyak 15.000 data berlabel digunakan untuk membangun model, sedangkan 53.538 data lainnya yang tidak berlabel akan digunakan untuk menerapkan model melakukan klasifikasi sentimen.

Pembagian data ke dalam data training dan data testing adalah sebanyak 60% data training, 20% data validation, dan 20% data testing.

| 60%        | 20%             | 20%        |
|------------|-----------------|------------|
| Data train | Data validation | Data test  |
| 9.000 data | 3.000 data      | 3.000 data |

**Gambar 4. 2** Pembagian Data yang Telah Dilabeli

#### 4.5 Implementasi IndoBERT

Dataset yang digunakan harus disesuaikan dengan representasi input yang diterima oleh IndoBERT. Setelah didapatkan dataset yang bersih hasil preprocessing,

selanjutnya dilakukan penambahan token khusus yaitu [CLS] pada awal kalimat, dan token [SEP] pada tiap akhir kalimat sebagai token pemisah antar kalimat. Kemudian dilakukan pengkodean (*encoding*) menggunakan tokenizer sesuai indeks *vocabulary* dari *pre-trained* IndoBERT. IndoBERT telah memiliki *vocabulary* yang nantinya akan digunakan sebagai proses transfer learning.

```
vocabulary1.txt X
100 w
101 ##om
102 ##ia
103 ##alah
104 ##ab
105 ##ot
106 ##si
107 ##aj
108 f
109 2
110 ##up
111 1
112 dalam
113 pr
114 tid
115 ##bu
116 ##ama
117 at
118 ##ok
119 tidak
120 ad
```

**Gambar 4. 3** *Vocabulary* IndoBERT

BERT menerima input dengan panjang maksimal 512 input. Jika kalimat lebih pendek akan dilakukan padding dengan menambahkan token [PAD] dan jika kalimat lebih panjang maka akan dilakukan *truncate* atau pemisahan kata. Untuk kata yang tidak ada pada *vocabulary* (*out-of vocabulary*) dilakukan pemisahan

menjadi sub kata (*subword*) dengan symbol `##`. Sebagai contoh proses penyesuaian input yang dapat diterima IndoBERT disajikan pada tabel berikut.

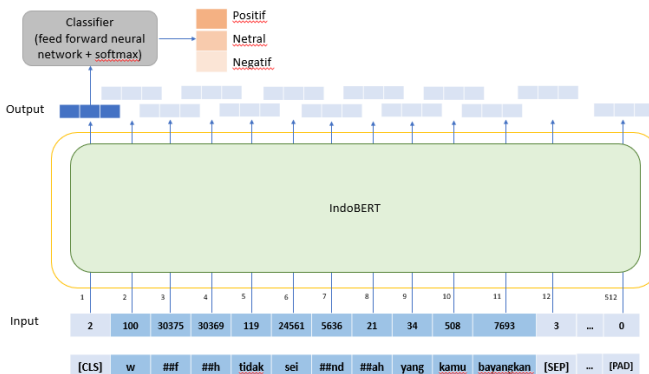
**Tabel 4. 4** Proses Tokenisasi IndoBERT

|                         |  |
|-------------------------|--|
| Kalimat                 | wfh tidak seindah yang kamu bayangkan  |
| Hasil tokenisasi        | 'w', '##f', '##h', 'tidak', 'sei', '##nd', '##ah', 'yang', 'kamu', 'bayangkan'                   |
| Penambahan token khusus | '[CLS]', 'w', '##f', '##h', 'tidak', 'sei', '##nd', '##ah', 'yang', 'kamu', 'bayangkan', '[SEP]' |
| Encoding                | 2, 100, 30375, 30369, 119, 24561, 5636, 21, 34, 508, 7693, 3                                     |
| Attention mask          | 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1  |

Berdasarkan tabel , kalimat ‘wfh tidak seindah yang kamu bayangkan’ kemudian dilakukan tokenisasi yang sesuai dengan *vocabulary* IndoBERT. Kata ‘wfh’ merupakan kata utama pada topik penelitian ini. Kata wfh dipisahkan menjadi sub kata (*sub word*) ‘w’, ‘##f’, ‘##h’ karena kata ‘wfh’ tidak ada dalam *vocabulary*. Penambahan kata pada *vocabulary* dapat menambah ukuran *vocabulary*. Ukuran embedding matrix juga harus dirubah karena adanya penambahan token kemudian melakukan *pre-training*. Maka dari itu pada penelitian ini hanya menggunakan *vocabulary* dan pre-trained model yang sudah tersedia.

Setelah dilakukan tokenisasi kemudian ditambahkan token khusus yaitu token [CLS] di awal kalimat dan token [SEP] di akhir kalimat. Selanjutnya pada tiap-tiap token dikodekan sesuai dengan indeks *vocabulary*. Token [CLS] memiliki id 2, token ‘w’ memiliki id 100, dan seterusnya. Attention mask digunakan untuk membedakan nilai token kata dengan nilai padding. Toked padding [PAD] memiliki nilai 0 sedangkan token kata diberi nilai 1. Untuk memudahkan proses tokenisasi dibuat suatu data loader untuk masing-masing dataset yaitu data loader untuk training, data loader untuk testing, dan data loader untuk validasi.

Selanjutnya input yang telah disesuaikan kemudian diteruskan ke dalam jaringan BERT yaitu tumpukan 12 layer Transformers Encoder. Tiap layer encoder memiliki dua sub layer, yang pertama adalah multi-head self-attention mechanism, dan yang kedua *fully connected feed forward network*. Setelah melewati semua encoder, didapatkan vector output dari tiap token. Namun hanya vector output dari token [CLS] yang akan digunakan sebagai vector input untuk classifier.



**Gambar 4. 4** Ilustrasi *Fine-tuning* IndoBERT untuk Klasifikasi

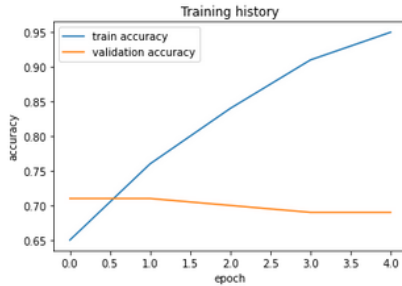
Proses *fine-tuning* IndoBERT untuk klasifikasi dilakukan dengan menambahkan layer klasifikasi. Library Transformers memiliki kelas BertForSequenceClassification yang didesain untuk tugas klasifikasi. Kelas BertForSequenceClassification bekerja dengan cara memasukkan output dari pooler untuk menghitung logits. Nilai logits yang dihasilkan kemudian digunakan untuk mendapatkan nilai prediksi menggunakan softmax.

Pada penelitian ini, penulis menggunakan model IndoBERT<sub>BASE</sub> phase 1 yang memiliki 12 *layer encoder*, 768 *hidden nodes*, dan 124 juta parameters. Berdasarkan paper, *hyperparameter* yang direkomendasikan untuk mendapatkan performa yang optimal adalah menggunakan batch size 16 atau 32, learning rate sebesar 5e-5, 3e-5, 2e-5 dengan optimizer adam dengan epoch 2,3, dan 4 (Devlin dkk, 2020). **Tabel 4.5** menunjukkan hyperparameter yang digunakan pada penelitian Tugas Akhir ini.

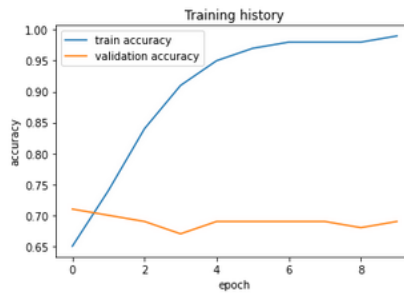
**Tabel 4.5** Hyperparameter IndoBERT

| No. | Hyperparameter | Ukuran       |
|-----|----------------|--------------|
| 1.  | Batch Size     | 16, dan 32   |
| 2.  | Learning Rate  | 2e-5         |
| 3.  | Epoch          | 5,10, dan 15 |

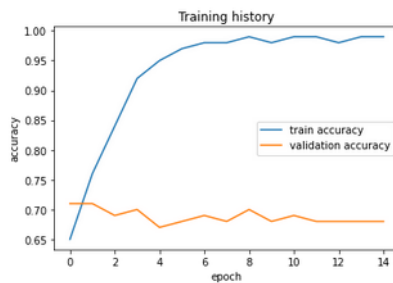
**Gambar 4.5**, **Gambar 4.6** , dan **Gambar 4.7** menunjukkan grafik perbandingan akurasi training dan akurasi validasi dari model menggunakan batch size 16 dengan iterasi 5 epoch, 10 epoch, dan 15 epoch.



**Gambar 4. 5** Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 5 Epoch



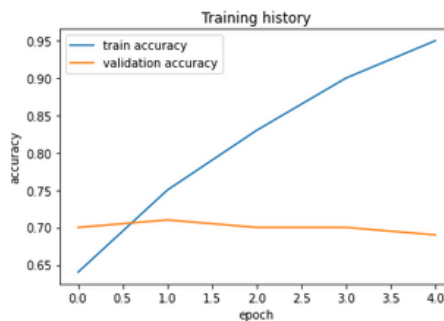
**Gambar 4. 6** Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 10 Epoch



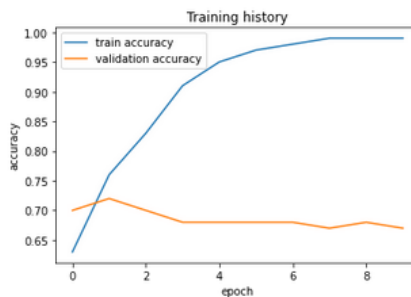
**Gambar 4. 7** Grafik Akurasi Training dan Validasi dengan Batch Size 16 dan 15 Epoch

Grafik akurasi model yang dihasilkan menunjukkan bahwa akurasi data training sangatlah baik karena hanya tingkat akurasi training hampir sangatlah baik yaitu sebesar 95% pada saat iterasi 5 mencapai 100%. Akan tetapi pada saat validasi tingkat akurasi lebih rendah dan cenderung sedikit fluktuatif. Hal ini menunjukkan bahwa model mengalami *overfitting*.

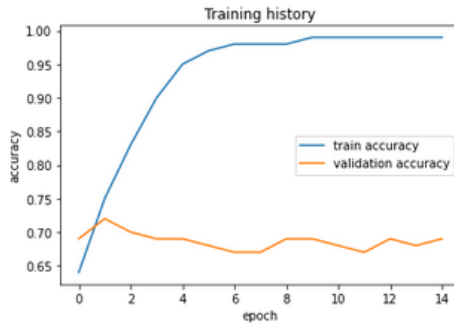
**Gambar 4.8**, **Gambar 4.9**, dan **Gambar 4.10** menunjukkan grafik perbandingan akurasi training dan akurasi validasi dari model menggunakan batch size 32 dengan iterasi 5 epoch, 10 epoch, dan 15 epoch.



**Gambar 4.8** Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 5 Epoch



**Gambar 4.9** Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 10 Epoch



**Gambar 4. 10** Grafik Akurasi Training dan Validasi dengan Batch Size 32 dan 15 Epoch

Grafik akurasi model juga menunjukkan bahwa akurasi data training sangatlah baik akan tetapi pada saat validasi tingkat akurasi lebih rendah. Dengan batch size 32 model juga mengalami *overfitting*. *Overfitting* dapat terjadi ketika model terlalu baik dalam melakukan klasifikasi terhadap data training akan tetapi sangat buruk untuk mengklasifikasi data test. Beberapa hal yang dapat menyebabkan *overfitting* antara lain jumlah data yang terlalu sedikit, jumlah variable input yang terlalu sedikit, atau kualitas data yang kurang baik. Pada penelitian ini, hal yang menyebabkan *overfitting* yaitu kualitas data yang kurang baik. Ditemukan data ulasan negative yang berlabelkan positif pada data tweet. **Tabel 4.6** menunjukkan beberapa data yang salah kelas.

**Tabel 4. 6** Data Tweets Salah Kelas

| Tweets   | Label   |
|--|---------|
| yeeeeee besok besok udh liburrr dari diri yang sudah lama tidak lg mencicipi wfh | negatif |
| alhamdulillah ada jadwal wfh   | netral  |
| mencabar ya wfh banyak gangguann   | positif |



## **BAB V**

### **UJI COBA DAN EVALUASI**

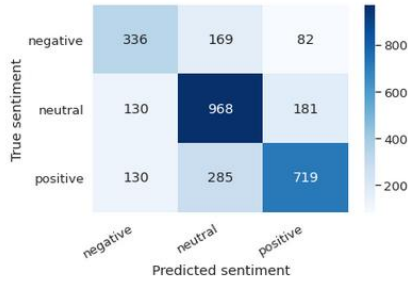
Pada bab ini dijelaskan mengenai pengujian dan evaluasi dari penelitian. Pengujian yang dilakukan adalah menguji model pada data testing kemudian dievaluasi menggunakan *confusion matrix*

#### **5.1 Data Uji Coba**

Data test adalah data yang digunakan pada saat pengujian model. Data test terdiri dari 3.000 data dengan 1.134 data sentimen positif, 587 data sentimen negatif, dan 1.279 data sentimen netral.

#### **5.2 Hasil Uji Coba**

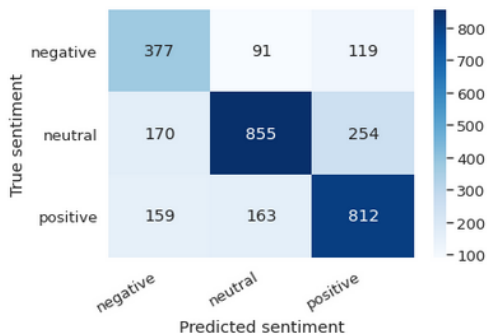
Setelah dilakukan proses training maka selanjutnya dilakukan uji coba untuk mengetahui performa model terhadap data baru. Berdasarkan hasil implementasi model terhadap data testing didapatkan akurasi sebesar 67%, 68%, dan 68% dari model menggunakan batch size 16 dengan iterasi 5 epoch, 10 epoch, dan 15 epoch. Sedangkan pada model dengan batch size 32 akurasi yang didapatkan sebesar 67%, 68%, dan 67% pada iterasi 5 epoch, 10 epoch, dan 15 epoch. Diagram *confusion matrix* dan performansi model pada tiap-tiap hasil percobaan sebagai berikut.



**Gambar 5. 1** Diagram Confusion Matrix Model dengan Batch size 16 dan 5 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.56      | 0.57   | 0.57     | 587     |
| neutral      | 0.68      | 0.76   | 0.72     | 1279    |
| positive     | 0.73      | 0.63   | 0.68     | 1134    |
| accuracy     |           |        | 0.67     | 3000    |
| macro avg    | 0.66      | 0.65   | 0.65     | 3000    |
| weighted avg | 0.68      | 0.67   | 0.67     | 3000    |

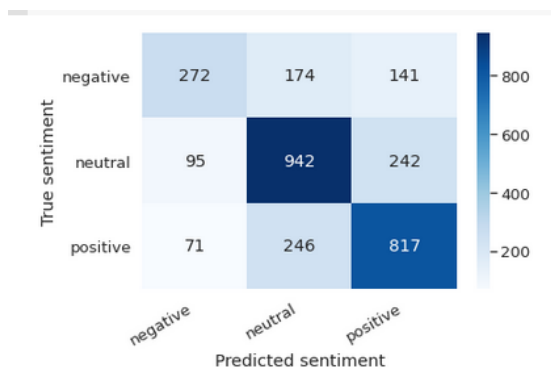
**Gambar 5. 2** Performansi Model dengan Batch size 16 dan 5 epoch



**Gambar 5. 3** Diagram Confusion Matrix dengan Batch size 16 dan 10 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.53      | 0.64   | 0.58     | 587     |
| neutral      | 0.77      | 0.67   | 0.72     | 1279    |
| positive     | 0.69      | 0.72   | 0.70     | 1134    |
| accuracy     |           |        | 0.68     | 3000    |
| macro avg    | 0.66      | 0.68   | 0.67     | 3000    |
| weighted avg | 0.69      | 0.68   | 0.68     | 3000    |

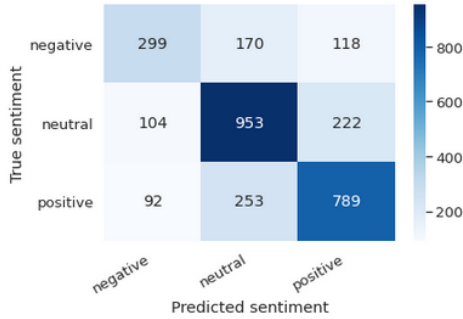
**Gambar 5. 4** Performansi Model dengan batch size 16 dan 10 epoch



**Gambar 5. 5** Diagram Confusion Matrix dengan Batch size 16 dan 15 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.62      | 0.46   | 0.53     | 587     |
| neutral      | 0.69      | 0.74   | 0.71     | 1279    |
| positive     | 0.68      | 0.72   | 0.70     | 1134    |
| accuracy     |           |        | 0.68     | 3000    |
| macro avg    | 0.66      | 0.64   | 0.65     | 3000    |
| weighted avg | 0.67      | 0.68   | 0.67     | 3000    |

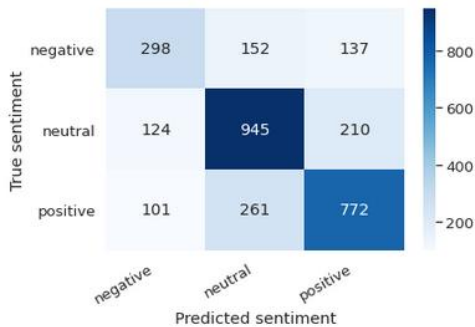
**Gambar 5. 6** Performansi Model dengan batch size 16 dan 15 epoch



**Gambar 5. 7** Diagram Confusion Matrix dengan Batch size 32 dan 5 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.60      | 0.51   | 0.55     | 587     |
| neutral      | 0.69      | 0.75   | 0.72     | 1279    |
| positive     | 0.70      | 0.70   | 0.70     | 1134    |
| accuracy     |           |        | 0.68     | 3000    |
| macro avg    | 0.67      | 0.65   | 0.66     | 3000    |
| weighted avg | 0.68      | 0.68   | 0.68     | 3000    |

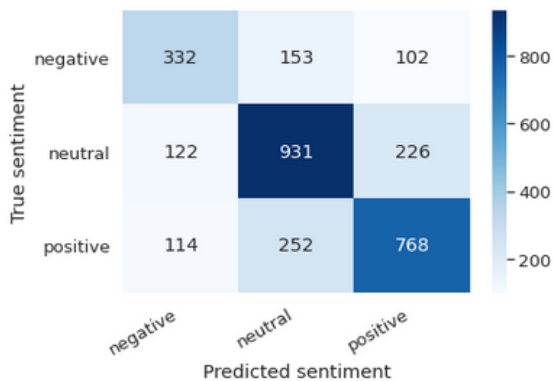
**Gambar 5. 8** Performansi Model dengan batch size 32 dan 5 epoch



**Gambar 5. 9** Diagram Confusion Matrix dengan Batch size 32 dan 10 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.57      | 0.51   | 0.54     | 587     |
| neutral      | 0.70      | 0.74   | 0.72     | 1279    |
| positive     | 0.69      | 0.68   | 0.69     | 1134    |
| accuracy     |           |        | 0.67     | 3000    |
| macro avg    | 0.65      | 0.64   | 0.65     | 3000    |
| weighted avg | 0.67      | 0.67   | 0.67     | 3000    |

**Gambar 5. 10** Performansi Model dengan batch size 32 dan 10 epoch



**Gambar 5. 11** Diagram Confusion Matrix dengan Batch size 32 dan 15 epoch

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.58      | 0.57   | 0.57     | 587     |
| neutral      | 0.70      | 0.73   | 0.71     | 1279    |
| positive     | 0.70      | 0.68   | 0.69     | 1134    |
| accuracy     |           |        | 0.68     | 3000    |
| macro avg    | 0.66      | 0.66   | 0.66     | 3000    |
| weighted avg | 0.68      | 0.68   | 0.68     | 3000    |

**Gambar 5. 12** Performansi Model dengan batch size 32 dan 15 epoch

Berdasarkan diagram confusion matrix dapat dilihat bahwa model memiliki performansi yang mirip dan cukup baik dalam mengklasifikasikan data. Rata-rata akurasi yang didapatkan model dari tiap-tiap percobaan sebesar 67,5%. Dari enam percobaan diketahui bahwa hanya dengan iterasi 5 epoch sudah menunjukkan performansi yang cukup baik. Model dengan batch size 32 memiliki akurasi 0.01% lebih baik dalam mengklasifikasikan dengan benar. Dari 3000 data uji, model mampu mengklasifikasikan 2.041 data sentimen dengan benar. Oleh karena itu model ini dipilih untuk diterapkan pada data yang tidak berlabel untuk mendapatkan hasil klasifikasi sentimen.

Dari 53.538 data, sebanyak 21.984 data diklasifikasi sebagai sentimen positif, 20.516 data sentimen netral, dan 9.038 data sentimen negatif. Berikut ini merupakan ringkasan hasil klasifikasi sentimen terhadap data yang tidak berlabel.

**Tabel 5. 1 Hasil Klasifikasi Sentimen**

| <b>Sentimen</b> | <b>Jumlah Data</b> |
|-----------------|--------------------|
| Positif         | 21.984             |
| Netral          | 20.516             |
| Negatif         | 9.038              |

Hasil penerapan model pada data opini yang tidak berlabel menunjukkan bahwa lebih banyak opini dikenali sebagai sentimen positif. Hal ini menunjukkan bahwa kecenderungan opini yang disampaikan melalui media sosial Twitter lebih banyak mengandung sentiment positif terhadap pelaksanaan WFH pada masa pandemi COVID-19.

## **BAB VI PENUTUP**

Pada bab ini berisi kesimpulan yang diperoleh dari peneliti Tugas Akhir ini serta saran perbaikan penelitian.

### **6.1 Kesimpulan**

Berdasarkan penelitian yang dilakukan maka dapat disimpulkan antara lain:

1. Masyarakat Indonesia cenderung memiliki sentimen positif terhadap pelaksanaan *work from home* di Indonesia pada masa Pandemi COVID-19. Kecenderungan sentimen positif terhadap pelaksanaan WFH menunjukkan bahwa WFH bisa dipertimbangkan untuk diterapkan pasca pandemi COVID-19.
2. Performansi model IndoBERT dengan pembagian data 60% data training, 20% data validation, dan 20% data testing mendapatkan nilai akurasi sebesar 98% untuk data training, dan akurasi rata-rata sebesar 68% untuk data testing. Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model IndoBERT mampu melakukan klasifikasi sentiment.

### **6.2 Saran**

Saran yang diberikan untuk penelitian selanjutnya

1. Pada penelitian ini kata 'wfh' merupakan kata yang sering digunakan, akan tetapi kata 'wfh' tidak terdapat pada *vocabulary*. Penulis menyarankan pada penelitian selanjutnya untuk melakukan percobaan dengan menambahkan kata pada *vocabulary* sesuai dengan topik penelitian untuk mengetahui pengaruhnya terhadap performansi model.

2. Melakukan penelitian dengan jumlah data yang seimbang antara sentimen positif, negative, dan netral, dan menggunakan data yang memiliki kualitas baik agar didapatkan performansi yang optimal.
3. Melakukan penelitian menggunakan model IndoBERT yang lain seperti IndoBERT<sub>LITE</sub> IndoBERT<sub>LARGE</sub> untuk kemajuan penelitian NLP dalam Bahasa Indonesia.



## DAFTAR PUSTAKA

- Ansong, E. and Boateng, R. 2018. Organisational adoption of telecommuting: Evidence from a developing country. *The Electronic Journal of Information Systems in Developing Countries*, 84(1), p.e12008.
- Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. eds. 2017. *A practical guide to sentiment analysis*. Cham, Switzerland: Springer International Publishing.
- Daniel, D., Nicolas, N., Ozden, C., Rijkers, B., Viollaz, M. and Winkler, H., 2020. Who on earth can work from home?.
- Delanoeije, J., Verbruggen, M. and Germeys, L. 2019. Boundary role transitions: A day-to-day approach to explain the effects of home-based telework on work-to-home conflict and home-to-work conflict. *Human Relations*, 72(12), pp.1843-1868.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Dewi, I. and Adiarsi, G.R. 2020. Leveraging the Covid-19 Crisis in Making Work-From-Home a Mainstream Practice in the Oil and Gas Industry.
- Grandini M., Bagli, E., Visani, G. 2020. Metrics for multi-class classification: An Overview. Arxiv:2008.05756.

- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), pp.497-506.
- Kanakaraddi, S. G., Chikaraddi, A. K., Gull K. C., and Hiremath., P. S. 2020. Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms. *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 287-292, doi: 10.1109/ICICT48043.2020.9112546.
- Li, Z., Fan, Y., Jiang, B., Lei, T. and Liu, W. 2019. A survey on sentiment analysis and opinion mining for social multimedia. *Multimedia Tools and Applications*, 78(6), pp.6939-6967.
- Naseem, U., Razzak, I., Musial, K., and Imran, M. 2020. Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. *Future Generation Computer Systems*, Vol. 113, pp. 58-69.
- Putra, I., F., Purwarianti, A. 2020. Improving Indonesian Text Classification Using Multilingual Language Model. <https://arxiv.org/abs/2009.05713>
- Ren, Z., Shen, Q., Diao, X., and Xu, H. 2021. A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management*, Vol. 58, Issue 3.
- Setyawan, F.E.B. and Lestari, R., 2020. Challenges of Stay-At-Home Policy Implementation During the Coronavirus (Covid-19) Pandemic in Indonesia. *Jurnal Administrasi Kesehatan Indonesia*, 8(2), pp.15-20.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Walker, P., Whittaker, C., Watson, O., Baguelin, M., Ainslie, K., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Cattarino, L. and Cucunuba Perez, Z. 2020. *Report 12: The global impact of COVID-19 and strategies for mitigation and suppression.*
- Wang, W., Albert, L. and Sun, Q. 2020. Employee isolation and telecommuter organizational commitment. *Employee Relations: The International Journal.*
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. 2020. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1<sup>st</sup> Conference of the Asia-Pacific Chapter of the Association for Computational Linguistic and the 10<sup>th</sup> International Joint Conference on Natural Language Processing.*
- Wu, Yonghui., Schuster, M., Chen, Z., Le, Q. 2016. Google's Neural Machine Translation System: bridging the gap between human and machine translation. <https://arxiv.org/pdf/1609.08144>.
- Ye, L.R. 2012. Telecommuting Implementation for success. *International Journal of Business and Social Science*, 3(15).



## LAMPIRAN

### Lampiran 1 Hasil Uji Coba Menggunakan Data Test

| No. | Dataset   | Label    | Hasil Klasifikasi |
|-----|---|----------|-------------------|
| 1   | time bulan posa wfh<br>smua benda jiran masak<br>dpt d cium bau   | neutral  | neutral           |
| 2   | lho bangun mepet juga<br>bisa anak wfh  | negative | negative          |
| 3   | oiii kan udah gw ceritain<br>maapin yak ga<br>nyambangin kedatangan<br>lo terpaksa wfh lagiii   | negative | negative          |
| 4   | terdampak secara<br>ekonomi karena covid 19<br>kurang banyak sabtu<br>minggu libur wfh<br>syukuri saja  | negative | negative          |
| 5   | untung aing wfh tadi  | positive | positive          |
| 6   | hari tu dah elok 30 wfh   | neutral  | positive          |
| 7   | terus ibu dpt kabar<br>temen se kantor dan se<br>ruangan nya cov dan ibu<br>ketemu beliau 2 hari<br>sebelum dpt kabar itu<br>ibu parno dong dan<br>emang sama kantor<br>langsung di suruh wfh<br>seminggu selama<br>dirumah ibu gak keluar<br>rumah samsek biasanya<br>belanja sendiri ini tu<br>nyuruh aku yg keluar | neutral  | positive          |

| No. | Dataset  | Label    | Hasil Klasifikasi |
|-----|--|----------|-------------------|
| 8   | wfh membuatku semakin malas dan semakin rajin bermain zpt  | positive | neutral           |
| 9   | kuota office 50 tapi ga ada wfh dan solusinya pindah meja biar 50 gatel  | neutral  | neutral           |
| 10  | wah asik dong shin kantornya lgsg lockdown wfh tuh   | positive | positive          |
| 11  | bojoku kan wkwwk ak risi nek ndelok pas nganggo seragam kui tujune saiki wfh                                       | positive | positive          |
| 12  | wfh jadi mandinya nanti aja wkwk   | neutral  | neutral           |
| 13  | 18 emang kalian yg wfh berasa lebih cape ya dari wfo   | positive | positive          |
| 14  | beatnya ada wfh kenapa la aku xde  | neutral  | neutral           |
| 15  | bener kerasa banget wfh malah makin byk kerjaan  | positive | positive          |
| 16  | lsamalah mcm kes aku org lain blh jer nk wfh tp tang aku byk bunyik plak sebabnya aku admin apa2 org cari like wth | negative | neutral           |
| 17  | terus gak bisa charge laptop terus wfh terus mesti ke rumah mbah atau rumah siapa gitu buat numpang                | neutral  | neutral           |

| No. | Dataset   | Label    | Hasil Klasifikasi |
|-----|---|----------|-------------------|
| 18  | wfh tapi nk masuk office<br>hmm   | neutral  | neutral           |
| 19  | klaster perkantoran dan tempat kerja ngeri ngeri mimin baca berita itu tuh rasanya dag dig dur der gitu mungking banyak dari tempat kerja kamu terapkan sistem kerja wfh lagi kan ya sudah punya cara buat pantau karyawan selama wfh belum | positive | negative          |
| 20  | wfh sampai abang jamil die  | neutral  | neutral           |
| 21  | akif21 sorry ye saya dah 4 bulan wfh tk keluar mana and keluar setakat bungkus makan beli brg dapur sahaja jangan tuduh bukan2 sbb kita tk thu org kt luar or persekitaran luar tu mungkin ada virus  | negative | negative          |
| 22  | krn publik sudah terbiasa dg wfh bikin cabin di desa adalah opsi menarik yg penting tercover jaringan 4g rumah kecil tanah sekitarnya dijadiin kolam ikan dan tempat nanam sayuran semua kegiatan ekonomi                                   | positive | positive          |

| No. | Dataset  | Label    | Hasil Klasifikasi |
|-----|--|----------|-------------------|
|     | dijalankan via jaringan digital cita citaku  |          |                   |
| 23  | konsep wfh tapi tetap harus masuk kantor dan finger datang pulang di kantor tuh gimanya  | neutral  | neutral           |
| 24  | nasib baik harini wfh  | positive | positive          |
| 25  | tim gue sebagai desainer di divisi setim ber3 smua punya laptop yg capable utk desain bahkan edit video waktu trial wfh taon lalu juga lancar2 aja tp si bos gabolehin 1 divisi utk wfh dgn alesan ga percaya kita beneran kerja di rmh padahal dia sendiri sampe skrg msh wfh | positive | positive          |
| 26  | b orang rumah pada pergi aku sendiri wfh aja n nsaatnya update can host awkwkwk  | neutral  | neutral           |
| 27  | weyy best ngat dpt wfh aku kena mari opis rodok hidung tiap minggu   | positive | positive          |
| 28  | yeeeeee besok besok udh liburr dari diri yang sudah lama tidak lg mencicipi wfh  | negative | positive          |
| 29  | baru sehari ni aku nk wfh belum pkul 10 lgi ni   | neutral  | negative          |



| No. | Dataset  | Label    | Hasil Klasifikasi |
|-----|--|----------|-------------------|
|     | dh carut kt office xpaa<br>ptg nnt aku masok xkn<br>nk handle office skjp<br>xdpt adoyaii  |          |                   |
| 30  | nyari kursi buat wfh dari<br>sebelum puasa sampe<br>skrng buset dah ga nemu<br>yg cocok  | positive | neutral           |
| 31  | thx to pandemic mau<br>ngantor boleh tapi izin<br>dulu mau wfh juga oke<br>udah setahun lebih wfh<br>all division semoga<br>setelah pandemi pun<br>akan tetep fleksibel<br>kerjanya  | positive | positive          |
| 32  | gini enaknya wfh   | positive | neutral           |
| 33  | pernaahh seminggu wfh<br>di bali sekalian jalan2<br>amp honeymoon walau<br>udh cuti tp ttp aja masih<br>disuruh meeting online<br>tp aslii asik bgttt buka<br>laptop online meeting<br>dgn pemandangan sawah<br>ubud pantai kuta kafe2<br>lucu | positive | positive          |
| 34  | ossas sama anjir<br>wkwkwk untuk wfh   | neutral  | neutral           |
| 35  | wfh sambil netflix<br>multitasking juga ga si  | neutral  | neutral           |
| 36  | agak capek yaallah lg<br>wfh mau call tp   | negative | negative          |

| No. | Dataset   | Label    | Hasil Klasifikasi |
|-----|---|----------|-------------------|
|     | indihomenya begini trs  |          |                   |
| 37  | kenapa kamu nggak pernah nyalain kamera saat meeting wfh ini  | neutral  | neutral           |
| 38  | wfh selang 3 hari bapak dia   | neutral  | neutral           |
| 39  | wfh hari ini sungguh produktif  | positive | positive          |
| 40  | enaknya sistem yg baru bisa diakses dari mana aja karena cloud basis wfh sebulan juga siap asal ada koneksi internet bisa kerja oke dah mulai ada chemistry   | positive | positive          |
| 41  | salah emang wfh hari ini bau masakan lebaran  | negative | neutral           |
| 42  | ga heran kalo nantinya wfh diperpanjang sampe 2030  | neutral  | neutral           |
| 43  | wfh la aku teman  | neutral  | neutral           |
| 44  | mencabar ya wfh banyak gangguann  | positive | positive          |
| 45  | and dh almost 3 month aku jaga ngan adik and abah nasib baik wfh penat tu xyah kate bukan penat jaga penat urus urusan lain byangkan kalau hri keje nk buat keje yg mmg bnyak pastu nk masak nk kemas umah x termasuk | negative | negative          |

| No. | Dataset   | Label   | Hasil Klasifikasi |
|-----|---|---------|-------------------|
|     | kalau time mak tngah sakit  |         |                   |
| 46  | mus jhon saya ga keberatan masalah thr ga ada komponen tukin d dalamnya cuma ga enak dibaca aja dibilang selama wfh kerja cm 2 hari beberapa jam gila aja manusia yang gapernah ditodong parang pas masuk hutan kayak anda ini sembarangan bacot sama ratain semua buruk pokoknya | neutral | negative          |
| 47  | pembimas katolik kristoforus sinselius beserta seluruh asn bimas katolik baik yang wfh maupun wfo mengikuti pembinaan secara luring dan daring melalui aplikasi zoom selasa 11 05   | neutral | neutral           |
| 48  | work aku lagi butuh orang yg punya skill di ads landing page dan website rep aja kalo bs 100 wfh bs part time   | neutral | neutral           |
| 49  | pasusu baru universe taynew two tweet au teta wfh vs ndut wfh   | neutral | neutral           |

| No.  | Dataset   | Label    | Hasil Klasifikasi |
|------|---|----------|-------------------|
| ⋮    |   |          |                   |
| ⋮    |   |          |                   |
| 2961 | ketika kasi wfh malas kerja   | neutral  | positive          |
| 2962 | syukurlah negative mari mudik lebih tepatnya pindah wfh dari jakarta ke rumah ortu di kampung halaman                                 | positive | positive          |
| 2963 | enakan wfh bisa liat athena main  | positive | neutral           |
| 2964 | wfh kashtalu  | neutral  | neutral           |
| 2965 | dzai wfh naman buas   | positive | neutral           |
| 2966 | wfh ajaa wkk  | neutral  | negative          |
| 2967 | maaf kak kami didaerah sgt membutuhkan jaringan klo prosesnya lama sangat merugikan kami pak apalagi wfh tlg dibantu agar lbh cpt kak | negative | negative          |
| 2968 | efek kelamaan wfh mba wkwk  | negative | negative          |
| 2969 | last day wfh tapi takut dilanjutkan agik dilema mok isik borang mdm berik psl tarikh abis wfh   | negative | positive          |
| 2970 | almond aslinya pas wfh ngerasa kek gini juga  | neutral  | neutral           |
| 2971 | besok wfh mau coba  | neutral  | neutral           |

| No.  | Dataset   | Label    | Hasil Klasifikasi |
|------|---|----------|-------------------|
|      | jalan ajalah  |          |                   |
| 2972 | cape abis kelar wfh trs masak abis ngadem lanjut mandi  | negative | positive          |
| 2973 | untuk para akuntan wanita wfh ini tentunya membawa keberkahan sendiri punya waktu lebih banyak bersama anak anak dan  | neutral  | neutral           |
| 2974 | wfh dong biar dikira tetangga melihara tuyul  | neutral  | positive          |
| 2975 | ternyataaaaaa gue pernah setangguh itu dan sampe sekarang selalu amazing sama diri sendiri yang bisa sekuat dan sesabar ini hadapi anak anak bayik yang lagi cranky dengan wajah penuh senyuman di tengah tengah kesibukan mamak yang wfh | positive | positive          |
| 2976 | senin pagi udah siap tinggal berangkat dan baru tau kalo ternyata masih wfh wfo dan gua wfh sia sia mandi pagi  | negative | negative          |
| 2977 | orang orang wfh tapi dikira miara babi ngepet duh   | negative | negative          |
| 2978 | hari ini kaya apes bgt udh kereta brgkt telat   | negative | negative          |

| No.  | Dataset   | Label    | Hasil Klasifikasi |
|------|---|----------|-------------------|
|      | pulang ga kebagian kereta di kerjaan isinya org komplek semua minta dikirim sbelum lebaran istirahat cuman 8 menit krn emg ga sempet pdhl mestinya pgn wfh tp gatau knp kok pgn ke kantor |          |                   |
| 2979 | jerih payah pemerintah apaan orang kerjanya main zuma bedanya main zumanya di rumah karena lagi wfh   | negative | neutral           |
| 2980 | hoooh selang seling sih iki seminggu wfh 2x   | neutral  | positive          |
| 2981 | bos ku lg baik bgt masa tgl 10 12 wfh   | positive | neutral           |
| 2982 | cm halo sore info dong tempat magang dijabodetabek yang engga wfh wfo makasih   | neutral  | negative          |
| 2983 | going magz uda dianggurin di kantor 5 hari plus barang dari bar uda 2 hari derita wfh unboxingnya delay lamaaaa   | negative | negative          |
| 2984 | mau ayam sambel ijo samping bri ih kangen kerja normal pandemi ini cepet berakhir doooooong capek wfh   | negative | negative          |

| No.  | Dataset  | Label    | Hasil Klasifikasi |
|------|--|----------|-------------------|
| 2985 | wfh aja dulu besok ya please kaki udah lemes bgt   | negative | neutral           |
| 2986 | gue biasanya cm jawab ngga bu wfh bodo amat dia ngerti wfh apa kaga wkwkwkwk   | negative | neutral           |
| 2987 | baru jg masuk sehari udah dikasih wfh aja  | neutral  | positive          |
| 2988 | bad news untuk athirah ijad wfh hehehe tengah kuak lentang atas katil ni best betullah   | positive | neutral           |
| 2989 | alhamdulillah ada jadwal wfh   | neutral  | positive          |
| 2990 | bagus bt bgini jk la company yg miti benarkan buka semua pekerja wajib swab test and ditanggung majikan settle konfem semua d sruh wfh   | positive | positive          |
| 2991 | xda wfh kerja smpai lebam haha   | positive | positive          |
| 2992 | kalau mami sukanya nyanyi hobi bu andi meriam matalata kalau papi alat musiknya lobo solo singer favoritnya kamu apa ingat dan tahukah siapa yang disukai orang tuamu wfh itu cuma untuk juragan | neutral  | neutral           |

| No.  | Dataset  | Label    | Hasil Klasifikasi |
|------|--|----------|-------------------|
|      | bukan pegawai polri tni campus   |          |                   |
| 2993 | suami hari ini wfo dan aku wfh aku udah selesai kerja suami masih otw pulang ohhh jadi inikah rasanya sendirian di rumah nunggu suami pulang kerja   | neutral  | positive          |
| 2994 | please total lockdown tutup semua sektor pekerjaan sementara fully wfh   | neutral  | positive          |
| 2995 | kena bawak blek printer siap2 pa wfh   | positive | positive          |
| 2996 | dia cuma kasi liat chat nya doang yaa gue iya iya aja dong ya terus pagi ini gue kan wfh ya gue lupa kalo ada seminar dan gue jadi host nya untungnya temen gue nanya di grup udah masuk apa belum ternyata dia udah ada di seminar dan dia minta bantuan temen magang satu lagi | neutral  | neutral           |
| 2997 | pud doakan lah bos nisa suruh wfh  | neutral  | positive          |
| 2998 | bila dah wfh ni no hp pun bertebaran la diseluruh pelusuk dunia  | neutral  | positive          |



| No.  | Dataset   | Label    | Hasil Klasifikasi |
|------|---|----------|-------------------|
|      | ya  |          |                   |
| 2999 | mal penuh bangun harusnya kita masih wfh  | positive | neutral           |
| 3000 | hai hiiii butuh buah tangan buat dikirim ke kampung snack di kala wfh wfo temen nonton bola yuuk pre order snack kacang a | neutral  | neutral           |

**Lampiran 2** Klasifikasi data *tweets* menggunakan model

| No. | Tweets   | Sentimen |
|-----|--|----------|
| 1   | wfh pero miss ko pa din ang bbq  | neutral  |
| 2   | tidur duluan besok wfh pagi jam 6  | positive |
| 3   | admin sosial media ada ga kak yang bisa wfh gitu   | neutral  |
| 4   | pulanginya bisa kak ntr baliknya ke sini yg ga bisa mana aku ga bisa wfh juga  | neutral  |
| 5   | beneran ngalamin beberapa bulan lalu apalagi wfh begini tp krn punya tanggung jawab selalu mikir gabisa resign gitu aja akhirnya ambil cuti buat libur sambil benerin cv portfolio terus nyari tempat kerja lain krn emg mau improve skill dan bersyukur udh dapet | negative |
| 6   | ice enti intlo vanta cheyadam kuda wfh ah neku   | neutral  |
| 7   | tengah hari nyetel dangdut puyeng bgt wfh dikasi dangdut lagi panas2 gueee pen beli spiker gue trus gue volume full dan stel lagu korea ampe magrib mampus lu jd kpopers   | positive |
| 8   | b singapura lockdown semua wfh dilarang dine in di resto   | negative |
| 9   | pbb mau wfh atau wfo kinerjanya buat perdamaian palestina dan israel sama aja yah  | neutral  |
| 10  | dunia kerja pasca covid 1 wfh akan bertahan ini berefek pada bisnis sewa kantor restoran toko retail dan transportasi  | neutral  |
| 11  | semenjak wfh udah yu dadah bubyet setrika  | neutral  |

| No. | Tweets   | Sentimen |
|-----|--|----------|
| 12  | negi prayag kameeno sham ko rakhiyega vivek ji pl cz mera wfh rahega till 6pm  | neutral  |
| 13  | b 1 singapura lockdown semua wfh dilarang dine in di resto   | negative |
| 14  | tidur lu jangan wfh mulu wkwk kasian jamban2 ga lu urusin  | negative |
| 15  | wfh tu apa you   | neutral  |
| 16  | harusnya bsok sabtu masuk cuma dari management bilang wfh aja biar karyawan bisa nikmatin bnyak waktu bersama keluarg wkwk liburan troooooossss  | positive |
| 17  | ababil senior ada yg suami istri di k l sama tapi beda dirjen ya sama aja sih mas gantian wfh wfo nya jagain anak tetep pusing si ibu satu direktorat sm saya ya mbaknya tetep pusing ngurus anak apalagi klo berdua rapat offline kemana gt pada mencar solusinya ya pake nanny | positive |
| 18  | 4 fakta lockdown sebulan singapura wfh lagi gara gara diamuk covid 19  | neutral  |
| 19  | singapura mengumumkan penerapan pembatasan ketat seluruh aktivitas perkantoran juga disetop dan warga singapura akan wfh dine in juga dilarang   | negative |
| 20  | hrp ya setelah senin juga tidak apa sahabat pas wfh aja baru keluar cacar aku nih  | neutral  |
| 21  | emang pada tolol biarin aja emosi gw udh ikut vaksin setaun lebih wfh  | positive |

| No. | Tweets  | Sentimen |
|-----|---|----------|
|     | selalu prokes tp banyak yg acuh kan gitu  |          |
| 22  | h 2 lebaran harusnya masih quality time sama keluarga saudara temen dll ini malah harus banget berangkat dan kerja di kereta padahal besok mah wfh ijin wfh hari ini doang ngeles mulu bos aing | negative |
| 23  | aelah sampe lupa mandi sih gara2 betah wfh dahlah dijama besok  | positive |
| 24  | ahh ya ya dia mah pokoknya klo bisa dilakuin di dorm kenapa harus keluar jamaah taat wfh banget   | positive |
| 25  | dennywattimena mana ada wfh kek gituu   | neutral  |
| 26  | wfh la irukaravangala ena sollrathu   | neutral  |
| 27  | waktu itu gue magang di salah satu media kerjanya bikin konten grafis buat postingan socmed biasa aja sih soalnya wfh jadi bisa nyantai   | positive |
| 28  | hai gue dkv konsentrasi animasi juni thn lalu gue magang di studio startup magangnya online alias wfh enak bgt  | positive |
| 29  | foxconn melihat lonjakan 2q pada ledakan wfh taipei foxconn taiwan pada hari jumaat meramalkan kenaikan pendapatan suku kedua sebanyak 15 ketik   | neutral  |
| 30  | sebulan wfh haih laaa   | neutral  |
| 31  | makanya malah bnyk yg kerja lembur gratisan semenjak wfh haha   | positive |
| 32  | keren bgt abis wfh gtu langsung nulis nath aku dah mumet misuh2   | positive |

| No. | Tweets  | Sentimen |
|-----|---|----------|
|     | marah2 pasti  |          |
| 33  | biasanya ditanyain lebi minat kerja di lapangan ato di kantor baru prefer wfh apa wfo in some emergency cases kamu hrs masuk tp ga sering                         | positive |
| 34  | kituh nya wfh tah   | neutral  |
| 35  | wfh sampai waktu yang tidak akan ditentukan   | neutral  |
| 36  | esok kerja lusa cuti tulat kerja then wfh   | neutral  |
| 37  | abis dikasih long weekend dari kantor dan setelah berapa lama wfh akhirnya terbang lagi ada kerjaan diluar pulau tes lagi ntar capek di colok colok terus o lt lt | negative |
| 38  | taknak balik bangi kenapa kena kerjaaaaaaa takleh wfh keeeees   | neutral  |
| 39  | pokoknya nnti kalo aku wfh harus sedia snack snack sama speaker buat dengerin lagu semoga dapet kerja aja dulu deh ah   | positive |
| 40  | onlyfans cukup wfh cuan lancar  | positive |
| 41  | bangga pada diri ini yang sanggup hidup 14 hari tanpa ncus jadi ibu sambil wfh  | neutral  |
| 42  | magang online wfh program magang ancora foundation dapat uang saku deadline 21 mei 2021 pukul 23 59 daftar di gratis gabung di channel telegram                   | positive |
| 43  | kantor suami udh inform kembali wfh lagi laptop komputer bawa rumah lagi  | neutral  |

| No.    | Tweets   | Sentimen |
|--------|--|----------|
| 44     | wfh makin gampang ngantuknyaa  | neutral  |
| 45     | poto mamah mam kalakuan si bungsu kalau pegang hape trs dapet poto terbaru mamah langsung edit aja lebaran taun ini hanya si bungsu yg temenin mamah di kampung halaman karena si bungsu wfh                           | positive |
| 46     | di sisi lain malas mandi contohnya saat wfh tidak mengharuskan pergi ke kantor juga punya risiko utamanya yang berkaitan dengan kebersihan pribadi beberapa risiko yang mungkin muncul adalah                          | negative |
| ⋮      |  |          |
| ⋮      |  |          |
| 53.535 | udahh wfh lagi   | neutral  |
| 53.536 | inilah yang memantapkan saya untuk wfh corona makin liar setelah mudik dan lebaran rs udah overload ambulance tiap jam lewat kenalan temen bahkan sahabat deket juga udah pada kena yok peduli yok minimal pake masker | positive |
| 53.537 | sekantor tumbang nahloh mana nih yang bikin kebijakan splitting wfh wfo dikata wfh aja udah  | neutral  |
| 53.538 | jgn dulu resign masa ky gini belum tentu wfh dikantor yg baru senyaman yg ini dan cuanya ky yg ini sehat   | positive |

**Lampiran 3 Data Tweets**

| No. | User                      | Tweet  |
|-----|---------------------------|--|
| 1   | DailySocial.id            | b'Selain mengusung spesifikasi yang mumpuni, fitur unggulan HP EliteOne 800 G8 AiO juga mencakup webcam berteknologi face tracking dan sistem noise reduction berbasis AI<br><a href="https://t.co/psW17VJLuW">https://t.co/psW17VJLuW</a> ' |
| 2   | Ria                       | b'Oi @FirstMediaCares dari jam 1 internet mati. Saya merasa tertanggung karena sedang WFH. Jadi service delay ke klien. Minta tolong segera diperbaiki jaringannya.'   |
| 3   | Athirah Syuhada           | b'Hari tak WFH ni la byk parcel sampai \xf0\x9f\x99\x84'   |
| 4   | Rhee                      | b'@chandrasyifaw Wkwkwkw wfh udah setahun parah bget.. Beli bajunya daster aja udah itu mah.. \xf0\x9f\x98\x85 jangan tutup lah kasian bos gue.. Yg gantiin gue si nitaaa..<br>\xf0\x9f\x91\x8f\xf0\x9f\x91\x8f\xf0\x9f\x98\x81'             |
| 5   | Gory from #Financial Gory | b'RT @aidy_syazreef: @ikhwandayne sama weh. siod tertengok tweet @azraeimuhamad nasib la wfh tak kanto menangis.'  |
| 6   | ðŸ• ±                     | b'Wfh pun wfh la janji i nak settle down kechap'   |
| 7   | irvani utami              | b'@_gncity Kayaknya dari kemarin berasa wfh sih\n\nUdah puas aku diketawain adek aku cuti tapi depan laptop mulu<br>\xf0\x9f\xa4\xa3\xf0\x9f\xa4\xa3'  |
| 8   | Aidy                      | b'@ikhwandayne sama weh. siod  |

| No. | User   | Tweet   |
|-----|--|---|
|     | Syazreef   | tertengok tweet @azraeimuhamad nasib la wfh tak kantai menangis.'   |
| 9   | â€¢  | b'I need wfh, kerja sambil tiduran<br>\\xf0\\x9f\\xa5\\xb4'   |
| 10  | .  | b'harini aku nak soping baju..tapi panas..terus malas.\\n\\nasal wfh je panas..asal wfh je panas'                                       |
| 11  | Û...Ø-<br>Û...Ø <sup>-</sup><br>Ø£Û†Û^Ø<br>±  <br>×ž×•×—<br>×ž×“<br>×• × ×•×•<br>×” ×□×œ<br>×•×•×”×•<br>×‘×™ | b'Celcom ada problem ke? Internet tak dapat refresh dari tadi. Buat hal la. Dah la aku WFH. Satu line ni je yang ada. Aduiiii'          |
| 12  | Recobeng   | b'@MargasayaWang Waduh. Apa dah WFH petugas KPPN nya ya\\nSenen selasa depan terakhir pencairan sp2d THR'                               |
| 13  | A Stark  | b'@hrdbacot Yes. Bebas2 aja sih. Mau WFH silakann, mau WFH juga gapapa'   |
| 14  | Agustinus Seco   | b'@puspitacsr Wuh... masih di jogja ternyata, tak kira jadi anak rantau dirimu. Ga pernah main gimana maksudnya? Aku wfh terus iki....' |
| 15  | ã, ·ãf†ã,£K<br>omceu   | b'@normmmsss kalo curhat temen; lebih baik mudik biar ga bayar kosan<br>\\xf0\\x9f\\x98\\x85 soalnya kantornya wfh ampe akhir taon'     |
| 16  | el   | b'wfh tp ttep dpet uang makan\\xf0\\x9f\\x99\\x86\\xe2\\x80\\x8d\\  |



| No. | User  | Tweet   |
|-----|---|---|
|     |   | xe2\x99\x80\xef\xb8\x8f'  |
| 17  | Chem  | b'@RyaWiedy WFH & LFH aja mulai skrg<br>\xf0\x9f\x91\x8c\xf0\x9f\x98\xa5'   |
| 18  | Jamell<br>Puasa<br>Genab<br>Sebolan<br>ðŸˆ¶ | b'WFH pun tak yaaaa. Tapi still dibayar gaji. Celaka tak celaka lah.<br><a href="https://t.co/R6o1JoI9tN">https://t.co/R6o1JoI9tN</a>   |
| 19  | Juhi<br>Wagmare                             | b'kaam se zyada time to vpn issue leleta hai\n#wfh'   |
| 20  | akram<br>rakib                              | b'@zulhilmiyusop Tktau plak sebab work stress, pmt ak pun ramai depression smpai amik medication. Wfh burnout'  |
| 21  | dindd.                                      | b'@alxxdke_ Mangat bumiil~ :)') sehat2 yaah. Dah berapa weeks bun?<br>\nAku sing gak hamil ae sambat wfh.<br>\nSambat kademen... Wkwkwk.'   |
| 22  | Aldike<br>Kiw Kiw                           | b'@adindaprasty Pol :)')\nAku jg amaze sm mereka yg masi dikasi kekuatan nggembol anak + wfo\nAku seng wfh ae sambat terus :(('   |
| 23  | J.  | b'@irvani28 Kamu keitung wfh atau cuti sih kak hari ini?\xf0\x9f\x98\x82'   |
| 24  | med<br>ðŸˆ§^ðŸˆ›>ðŸˆ<br>Ÿ□Ÿ                 | b'outfit wfh selama seminggu terakhir ini ialah mukena. terima kasih krn meetingnya boleh off kamera.\n\nsebenarnya lebih ke kedinginan sama suhu ruangan sendiri<br>\xf0\x9f\xa5\xb2 dimatiin, kegerahan, yaudah kita pake mukena ajahhh hehe dsr org jakarta baru \xf0\x9f\x99\x83' |
| 25  | attacking                                   | b'@yacil_ Kalo wfh masih dicari ya  |

| No. | User                | Tweet   |
|-----|---------------------|---|
|     | midfielder          | wajar la, kecuali cuti'   |
| 26  | who dis             | b'alhamdulillah bgt wkww, kantor kemarin ada acara bukber, dan karena aku satu2nya yang WFH, tbtb hari ini ditransfer uang mentahannya dong, katanya biar ikut juga meskipun jauh\xf0\x9f\x98\x82'  |
| 27  | Bond Joeurve        | b'@jjsxz Best Klau ade partner sebelah....lg suka wfh....haha'  |
| 28  | joy                 | b'@CJCalimaki seyeng wfh kami \xf0\x9f\x98\x9e'   |
| 29  | den                 | b'Lagi sibuk sibuknya wfh, semua serba online \xf0\x9f\x98\x8c\xf0\x9f\x98\x8c\xe2\xad\x90 <a href="https://t.co/BOgt54yx5F">https://t.co/BOgt54yx5F</a> '  |
| 30  | Alfridho Yuliananda | b'WFH sambil puasa itu tantangannya cuma 1: ngantuk!'   |
| 31  | DADDY POOH          | b'@KifliMally @Celcom aritu digi buat hal, nasib baik la data phone pakai lain, wifi umah pakai lain, kalau sama jahanam! da la kene WFH 1 hape xleh buat nanti'                                    |
| 32  | Awan Nano           | b'@jjsxz Sebab tu suka alternate sebab wfh sepanjang masa mmg rasa tak produktif \xf0\x9f\x98\xa9'  |
| 33  | Awan Nano           | b'RT @jjsxz: Wfh mostly tido\nDh katil sebelah'   |
| 34  | ðŸ• .               | b'is there anything or anyone more stupid than orang yg jelas2 sakit tp tetep masuk kerja???? lo compromising safety n health orang2 sekitar lo bgsd, pdhl dibolehin wfh kenapa si hobi amat masuk' |

| No. | User          | Tweet   |
|-----|---------------|---|
| 35  | RiMayuka      | b'Mco 3.0 kita wfh lagi.. hahaha takde beza nya pon<br>\xf0\x9f\xa4\xa3\xf0\x9f\xa4\xa3'  |
| 36  | chris         | b'Wfh mostly tido\nDh katil sebelah'  |
| 37  | Ai            | b'RT @bangdoel54: Jumat begini enak juga ternyata WFH..gk susah nyari tempat solat jumat'   |
| 38  | ðŸ••          | b'@kimchari__ Ku pernah merasakannya tapi tidak di tambah dengan wfh wkwkwkk'   |
| 39  | í,☐ē, kimi    | b'mari kita membalas dm orang-orang, gini amat ya mhs smt akhir nyambi kerja ditambah wfh ruwet pala ku'  |
| 40  | Hidayat Johan | b'Byk kli da prsan ble time wfh.. \nTgh2 buat kje tpi tgkp ni, tgh2 hari buta buln puasa ni mst ada bau asap rkok gudang garam.. Kuat gler.. Pdhal duk kjiaranan mlyu kiri, knn, dpn, blkg sume mlyu.. Aduhai.. Sesak nfas den..'                       |
| 41  | Namaku        | b'Wfh hari ini bawaan nya ngantuk, mager, rebahan sambil nungguin muter" mulu ini BPRD gua smpe ketiduran 1 jam \xf0\x9f\x98\xaa'   |
| 42  | FF            | b'tontonan wfh konser-konser begini, jadi kepikiran kalo nanti ada lagi, pengen nonton live. \nErwin Gutawa Orchestra ft All Artists - Closing (Konser Salute Erwin Gut...<br><a href="https://t.co/lxHxX0jME">https://t.co/lxHxX0jME</a> via @YouTube' |
| 43  | Havis         | b'Salah satu alasan lebih suka WFH adalah kursi dan meja kerja di rumah   |

| No.    | User                           | Tweet   |
|--------|--------------------------------|---|
|        |                                | lebih nyaman daripada yang di kantor. sing di kantor bikin boyoken wkwk.'   |
| 44     | Dee<br>ðŸŒ€'â€œâ€<br>~ï,â€~ï,• | b'Waktunya kerja ya kerja, waktunya mudik ya mudik semua ada timeline nya eh tiba2 pas waktunya mudik dilarang mudik. Untung sy kebagian WFH kalau tidak kemungkinan ga bisa mudik juga.'             |
| 45     | Tamara                         | b'Hari ini jadi wfh gara2 ada yg positif mana satu ruangan lg yaAllah.. udahlah gue orgnya parno kebangetan fix besok mau swab antigen aja\xf0\x9f\x98\xad'   |
| 46     | rin82                          | b'Hari ini berasa ada yg kurang.\nTernyata blm \xf0\x9f\x92\x84 lipstik-an.\nNyahahahaaa nasib WFH, lipstik ga habis2. Ga ada alesan buat beli baru.<br>\xf0\x9f\x91\x84\x9f\x8d\x92\x9f\x9f\x92\x84' |
| :      | :                              |   |
| 66.538 | crazybadboy88                  | b'@seasalteed Jgn dulu resign masa ky gini belum tentu wfh dikantor yg baru senyaman yg ini dan cuanya ky yg ini...sehat”   |

## BIODATA PENULIS



Penulis memiliki nama lengkap Bella Rahmatullah, lahir di Surabaya pada tanggal 28 Maret 1998. Penulis berasal dari Surabaya, bertempat tinggal di Jl. Pakis Gunung II-F/DALAM No. 15, Surabaya. Pendidikan yang pernah ditempuh yaitu SDN Pakis I Surabaya (2004-2010), SMP Negeri 10 Surabaya (2010-2013), SMA Negeri 18

Surabaya (2013-2016). Kemudian penulis melanjutkan studi di Departemen Matematika ITS. Semasa menempuh jenjang Pendidikan S-1, penulis juga berorganisasi di Resimen Mahasiswa ITS. Selain itu, penulis juga pernah melaksanakan kerja praktik di PT. Terminal Petikemas Surabaya pada tahun 2019. Penulis menyadari bahwa Tugas Akhir ini tidak lepas dari kekurangan. Untuk itu penulis mengharapkan kritik dan saran mengenai Tugas Akhir ini yang dapat disampaikan melalui email [bellarahmatullah@gmail.com](mailto:bellarahmatullah@gmail.com).