

5. HASIL DAN PEMBAHASAN

5.1. Hasil Eksperimen

Pada bagian ini akan ditunjukkan beberapa temuan dari eksperimen yang dilakukan selama pengerjaan skripsi ini. Tujuan dari penelitian ini adalah untuk menghasilkan model yang memiliki tingkat akurasi tinggi dalam mendeteksi penyakit mental.

5.1.1. Model Klasifikasi Biner

Pada bagian ini dilakukan pembagian *dataset* menjadi 2 bagian, *training* dan *testing*, dengan rasio 80:20 dan data *validation* yang merupakan 20% dari *training dataset*. Model yang diuji menggunakan empat jenis arsitektur model, yaitu *Long Short-Term Memory* (LSTM), *Bidirectional Long Short-Term Memory* (BiLSTM), CNN-LSTM, dan CNN-BiLSTM. Pengukuran performa dari model memanfaatkan perhitungan akurasi.

5.1.1.1. Arsitektur Model

Untuk mengatasi tantangan dalam klasifikasi teks biner, digunakan empat arsitektur model machine learning yang berbeda. Setiap arsitektur memiliki keunggulan dan karakteristik tersendiri dalam menangani data teks. Model-model tersebut adalah LSTM, BiLSTM, CNN+LSTM, dan CNN+BiLSTM. Berikut adalah penjelasan singkat mengenai masing-masing arsitektur yang digunakan dalam penelitian ini:

5.1.1.1.1. LSTM

Tujuan utama pengujian parameter dalam model LSTM adalah untuk menentukan lapisan konfigurasi mana yang paling cocok untuk tugas klasifikasi. Jumlah unit, *dropout*, dan *recurrent dropout* dicoba untuk mengoptimalkan model dan meningkatkan akurasi dan kinerjanya dalam menangani tugas termasuk klasifikasi. Dalam pengujian ini layer-layer lain diatur tetap dengan setting berikut ini:

Embedding, Dense(units=1, activation='sigmoid')

Tabel 5.1

Pengujian LSTM

Layer	Akurasi Terbaik	Epoch terbaik
-------	-----------------	---------------

LSTM(32)	0.70741	6
LSTM(64)	0.73054	8
LSTM(128)	0.76051	5
LSTM(256)	0.88662	13
LSTM(512)	0.87168	19
LSTM(256), Dropout(0.1)	0.76604	15
LSTM(256), Dropout(0.2)	0.81416	19
LSTM(256), Dropout(0.3)	0.86947	8
LSTM(256), Dropout(0.4)	0.69137	11

5.1.1.1.2. BiLSTM

Tujuan utama pengujian parameter dalam model BiLSTM adalah untuk menentukan lapisan konfigurasi mana yang paling cocok untuk tugas klasifikasi. Sama seperti sebelumnya, jumlah unit, *dropout*, dan *recurrent dropout* dicoba untuk mengoptimalkan model dan meningkatkan akurasi dan kinerjanya dalam menangani tugas klasifikasi. Dalam pengujian ini layer-layer lain diatur tetap dengan setting berikut ini:

Embedding, Dense(units=1, activation='sigmoid')

Tabel 5.2

Pengujian BiLSTM

Layer	Akurasi	Epoch terbaik
Bidirectional(LSTM(32))	0.90210	8
Bidirectional(LSTM(64))	0.90929	10
Bidirectional(LSTM(128))	0.82688	13
Bidirectional(LSTM(256))	0.83186	6
Bidirectional(LSTM(512))	0.86173	7
Bidirectional(LSTM(64)), Dropout(0.1)	0.84900	11
Bidirectional(LSTM(64)), Dropout(0.2)	0.87555	11
Bidirectional(LSTM(64)), Dropout(0.3)	0.86504	5
Bidirectional(LSTM(64)), Dropout(0.4)	0.87832	12

5.1.1.1.3. CNN LSTM

Tujuan utama pengujian parameter dalam model CNN LSTM adalah untuk menentukan konfigurasi lapisan *Conv1D* mana yang paling cocok untuk tugas klasifikasi. Jumlah *hidden layer*, dan *kernel size* dicoba untuk mengoptimalkan model dan meningkatkan akurasi dan kinerja model LSTM dalam menangani tugas klasifikasi. Dalam pengujian ini layer-layer lain diatur tetap dengan setting berikut ini:

Embedding, LSTM(256), Dense(units=1, activation='sigmoid')

Tabel 5.3

Pengujian CNN LSTM

Layer	Akurasi	Epoch terbaik
Conv1D(filters=32, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.81858	14
Conv1D(filters=64, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.85564	6
Conv1D(filters=128, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.82135	11
Conv1D(filters=64, kernel_size=4, activation='relu'), MaxPooling1D(pool_size=2)	0.83960	5
Conv1D(filters=64, kernel_size=5, activation='relu'), MaxPooling1D(pool_size=2)	0.84237	13
Conv1D(filters=64, kernel_size=6, activation='relu'), MaxPooling1D(pool_size=2)	0.80863	17
Conv1D(filters=64, kernel_size=6, activation='relu'), MaxPooling1D(pool_size=2), Dropout(0.2)	0.85288	6
Conv1D(filters=64, kernel_size=6, activation='relu'), Dropout(0.2), MaxPooling1D(pool_size=2)	0.85785	10
Conv1D(filters=64, kernel_size=6, activation='relu'), Dropout(0.4), MaxPooling1D(pool_size=2)	0.87998	20

5.1.1.1.4. CNN BiLSTM

Tujuan utama pengujian parameter dalam model CNN BiLSTM adalah untuk menentukan konfigurasi lapisan *Conv1D* mana yang paling cocok untuk tugas klasifikasi. Jumlah *hidden layer*, dan *kernel size* dicoba untuk mengoptimalkan model dan meningkatkan akurasi dan kinerja model BiLSTM dalam menangani tugas klasifikasi. Dalam pengujian ini layer-layer lain diatur tetap dengan setting berikut ini:

Embedding, Bidirectional(LSTM(32)), Dense(units=1, activation='sigmoid')

Tabel 5.4

Pengujian CNN BiLSTM

Layer	Akurasi	Epoch terbaik
Conv1D(filters=32, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.88496	10
Conv1D(filters=64, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.86338	10
Conv1D(filters=128, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2)	0.89215	6
Conv1D(filters=128, kernel_size=2, activation='relu'), MaxPooling1D(pool_size=2)	0.91925	6
Conv1D(filters=128, kernel_size=4, activation='relu'), MaxPooling1D(pool_size=2)	0.89104	7
Conv1D(filters=128, kernel_size=5, activation='relu'), MaxPooling1D(pool_size=2)	0.85564	10
Conv1D(filters=128, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2), Dropout(0.2)	0.89767	6
Conv1D(filters=128, kernel_size=3, activation='relu'), Dropout(0.2), MaxPooling1D(pool_size=2)	0.86892	14
Conv1D(filters=128, kernel_size=3, activation='relu'), MaxPooling1D(pool_size=2), Dropout(0.3)	0.89657	5
Conv1D(filters=128, kernel_size=3,	0.88827	13

activation='relu'), MaxPooling1D(pool_size=2), Dropout(0.4)		
---	--	--

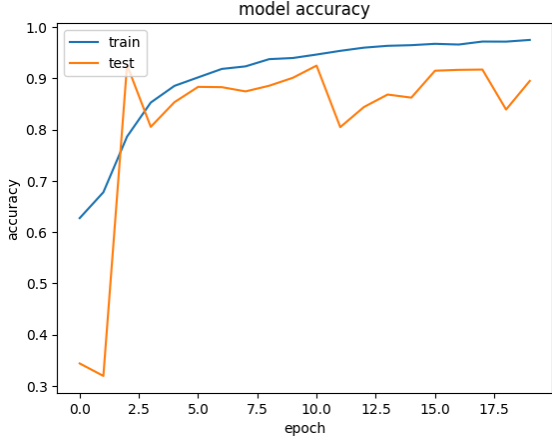
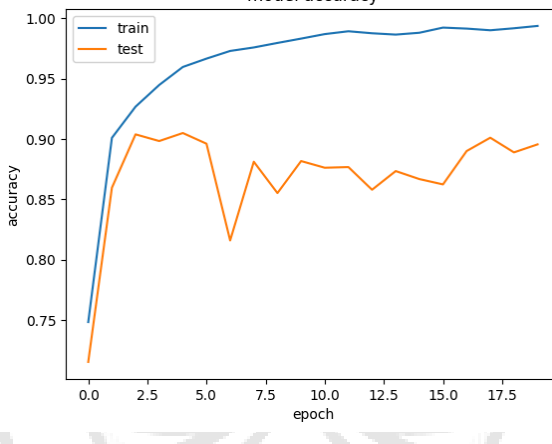
5.1.1.2. ADHD

Pengujian pada label ADHD dilakukan menggunakan 4 macam model dengan arsitektur yang menghasilkan akurasi terbaik sebelumnya. Hasil pengujian sebagai berikut:

Tabel 5.5

Hasil Label ADHD

No.	Model	Grafik	Val Accuracy
1.	LSTM		0.91040
2.	BiLSTM		0.92644

3.	CNN-LSTM		0.92588
4.	CNN - BiLSTM		0.90487

5.1.1.3. Anxiety

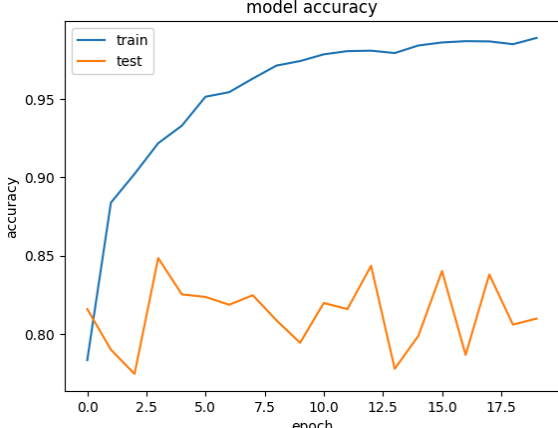
Pengujian pada label *anxiety* dilakukan menggunakan 4 macam model dengan arsitektur yang menghasilkan akurasi terbaik sebelumnya. Hasil pengujian sebagai berikut:

Tabel 5.6

Hasil Label *Anxiety*

No.	Model	Grafik	Val_Accuracy
-----	-------	--------	--------------

1.	LSTM		0.63938
2.	BiLSTM		0.83905
3.	CNN-LSTM		0.78153

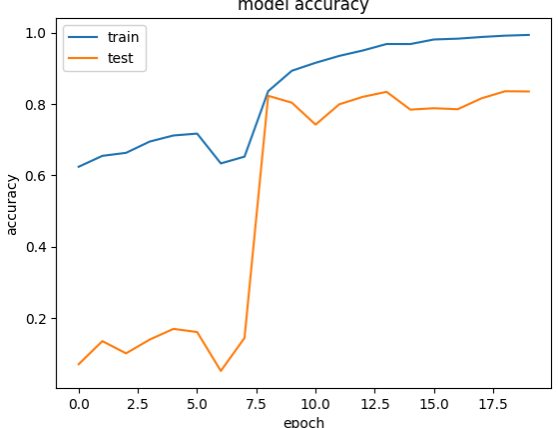
4.	CNN - BiLSTM		0.84845
----	---------------------	--	----------------

5.1.1.4. Bipolar

Pengujian pada label *bipolar* dilakukan menggunakan 4 macam model dengan arsitektur yang menghasilkan akurasi terbaik sebelumnya. Hasil pengujian sebagai berikut:

Tabel 5.7

Hasil Label *Bipolar*

No.	Model	Grafik	Val_Accuracy
1.	LSTM		0.83573

2.	BiLSTM	<p>model accuracy</p> <p>accuracy</p> <p>epoch</p> <p>train</p> <p>test</p>	0.86560
3.	CNN-LSTM	<p>model accuracy</p> <p>accuracy</p> <p>epoch</p> <p>train</p> <p>test</p>	0.86394
4.	CNN - BiLSTM	<p>model accuracy</p> <p>accuracy</p> <p>epoch</p> <p>train</p> <p>test</p>	0.86836

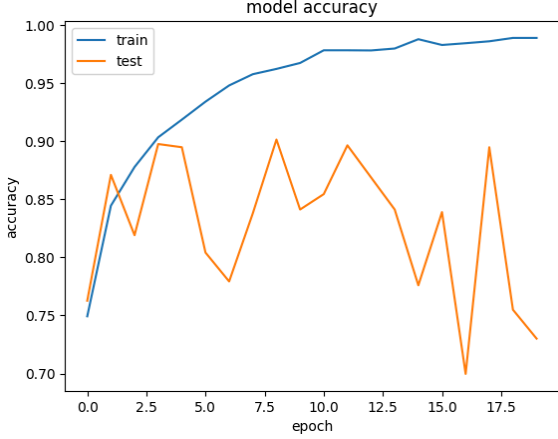
5.1.1.5. Depression

Pengujian pada label *depression* dilakukan menggunakan 4 macam model dengan arsitektur yang menghasilkan akurasi terbaik sebelumnya. Hasil pengujian sebagai berikut:

Tabel 5.8

Hasil Label *Depression*

No.	Model	Grafik	Val_Accuracy
1.	LSTM		0.89215
2.	BiLSTM		0.89602
3.	CNN-LSTM		0.83352

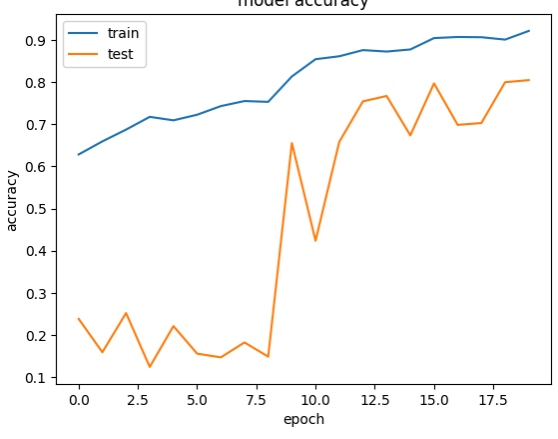
4.	CNN - BiLSTM		0.90155
----	---------------------	--	----------------

5.1.1.6. PTSD

Pengujian pada label PTSD dilakukan menggunakan 4 macam model dengan arsitektur yang menghasilkan akurasi terbaik sebelumnya. Hasil pengujian sebagai berikut:

Tabel 5.9

Hasil Label PTSD

No.	Model	Grafik	Val_Accuracy
1.	LSTM		0.80476

2.	BiLSTM		0.87696
3.	CNN-LSTM		0.85122
4.	CNN - BiLSTM		0.88717

5.1.1.7. Analisis hasil

Pengujian berikut dilakukan dengan tujuan untuk mencari arsitektur model apa yang memiliki performa terbaik, baik secara keseluruhan atau dalam setiap kelas.

Tabel 5.10

Hasil klasifikasi biner tiap kelas

Kelas	LSTM	BiLSTM	CNN + LSTM	CNN+BiLSTM
ADHD	0.9104	0.92644	0.92588	0.90487
Anxiety	0.63938	0.83905	0.78153	0.84845
Bipolar	0.83573	0.8656	0.86394	0.86836
Depression	0.89215	0.89602	0.83352	0.90155
PTSD	0.80476	0.87696	0.85122	0.88717
Rata-rata	0.84845	0.88649	0.84237	0.89436

Berdasarkan tabel 5.1 dapat disimpulkan bahwa pemilihan model yang tepat bergantung pada karakteristik spesifik dari setiap kelas. Dengan nilai rata-rata tertinggi, model kombinasi CNN dan BiLSTM mampu meraih tingkat performa yang lebih tinggi dibandingkan model lainnya.

5.1.2. Pengujian Generasi Data CLM

Pada bagian ini dilakukan pengujian hasil teks yang dihasilkan model. Model yang diuji adalah *Causal Language Model* dengan menggunakan GPT-2 sebagai model *pretrained* dan *Conditional Generation Model* dengan menggunakan T5 sebagai model *pretrained*. Pengujian dilakukan dengan mencoba menampilkan hasil generasi teks dari masing-masing model dengan kelas yang dituju

Tabel 5.11

Hasil teks model generator

Model	Hasil Teks
<i>Conditional Generation Model</i>	ADHD: what are some ways you manage adhd? the answer came my friends today stating about their adhd diagnosis if you dont have adhd. well, it doesnt look like they're being treated for a medical condition - what are the best ways to get help with adhd, how can they help you? Im sure they are having difficulty stuttering, and I don't feel like the symptoms were a result of my adhd because of my depression. its

	been awhile since i had my dr prescription medication to prevent me getting into a job and have to get into work and get caught up with..
<i>Conditional Generation Model</i>	Anxiety: help getting started this weekend. for the first time in my career that was i have tv anxiety about it from my parents' experiences of stressors and anxiety because the worst part was, the numb feeling and the anxiety around me wasnt ok at all. my kids were high school sports, but my anxiety was just going downhill. ive only been there. 4 years of university, never did college. if the job would be fun, not impossible, and i feel like its the end of me trying. my parents has no idea how to cope. for my job, being in the beginning..
<i>Causal Language Model</i>	ADHD: comitalin-3 hours of adderall, so i have adhd? i was diagnosed with adhd, so, i was just diagnosed with adder all over my life. different meds, my adhd is the doctor has been a lot of a lot. i am not going to take it. i have been taking medication for a few months. i was not sure i started taking a job for my ad and a month and i was really bad. [bipolar] i am now. i'm feeling very hard to do my life, but i am just feel like i am i do have a panic attacks. i feel like to feel like this feeling more than i have no matter what to be a therapist because i have to be my meds and i just feel more. i just not really just feel so anxious. i don't know if it. but i feel so i feel guilty for a lot more and i'm in this. i can't even feel like my own. i do feel like everything, i'm having things and i know i just don't even have an option to do is going to do not. i want to do, but. i need to get out of the past. i get more often feel like..
<i>Causal Language Model</i>	Anxiety: now im not sure if this is so hard for me, i cant even have a new job, but i was a little bit of a lot of anxiety. i wasnt doing well, and i was not sure i was in my life. i am just being a few weeks ago and i just have an anxiety that i was having a lot. [anax. i feel like a lot more than i was still very depressed and i dont know how much. i would have a hard time. i just want to keep my anxiety, because i am not sure. i cant do, and have the only have to have to get really have to do anything. i'm tired of this? [an off theres nothing to do i have to see my anxiety and i feel my life, but it. i dont have been really need to be able to have some of my life or get better. i have any help. i do it out of this, i am, and not a lot that my head that my parents. i get the time. the time, i do have been on my head. i can't have to tell my life of this. i don't get so bad, but my mind is that i feel i feel this..

Dari hasil tabel 5.11, teks yang dihasilkan model *Conditional Generation Model* T5 memiliki tata bahasa yang lebih baik dibandingkan model *Causal Language Model* GPT-2. Hal ini disebabkan karena sifat CLM yang memberi data berdasarkan hasil augmentasi data latihan yang diberikan.

Dari data yang dihasilkan dari proses diatas, akan diambil sebanyak 3 kelompok sampel data secara acak dari teks yang dihasilkan kedua model untuk dilakukan survei model mana yang menghasilkan teks yang lebih lebih koheren, mudah dibaca, dan lebih merepresentasikan orang yang mengalami gangguan mental. Daftar lengkap dari kelompok sampel yang diambil bisa dilihat pada Lampiran 1.

Tabel 5.12

Hasil survei generasi teks

Kelompok sampel data	GPT2	T5
I	9%	91%
II	18%	82%
III	9%	91%

Dari hasil survei terhadap 11 responden, rata-rata responden mengatakan bahwa teks yang dihasilkan oleh model pre-trained Seq2Seq T5 lebih koheren, mudah dibaca, dan lebih merepresentasikan orang yang mengalami gangguan mental dari pada teks yang dihasilkan oleh model pre-trained GPT2.

5.1.3. Model *Sampling Multi-label*

Pada bagian ini dilakukan pengujian pada model *sampling*. Model ini merupakan penggabungan dua jenis model yang memiliki hasil terbaik, yaitu model kombinasi CNN dengan LSTM untuk tugas klasifikasi dan model *Conditional Generation Model* T5 yang sudah dilatih sebelumnya. Model dilakukan *sampling* dengan target label berupa list *subset* dari set label yang ada.

Tabel 5.13

Komposisi data hasil generasi dua target kelas

Target kelas	Komposisi kelas dicapai (setiap 200 data)
A = {adhd}; B = {anxiety}	$ A \cup B = 194$; $ A \cap B = 28$
A = {adhd}; B = {bipolar}	$ A \cup B = 147$; $ A \cap B = 32$
A = {adhd}; B = {depression}	$ A \cup B = 171$; $ A \cap B = 20$

A = {adhd}; B = {ptsd}	$ A \cup B = 179; A \cap B = 44$
A = {anxiety}; B = {bipolar}	$ A \cup B = 186; A \cap B = 34$
A = {anxiety}; B = {depression}	$ A \cup B = 193; A \cap B = 45$
A = {anxiety}; B = {ptsd}	$ A \cup B = 190; A \cap B = 48$
A = {bipolar}; B = {depression}	$ A \cup B = 180; A \cap B = 30$
A = {bipolar}; B = {ptsd}	$ A \cup B = 163; A \cap B = 34$
A = {depression}; B = {ptsd}	$ A \cup B = 175; A \cap B = 19$

Tabel 5.14

Komposisi data hasil generasi tiga target kelas

Target kelas	Komposisi kelas dicapai (setiap 500 data)
A = {adhd}; B = {anxiety}; C = {bipolar}	$ A \cup B \cup C = 489; A \cap B \cap C = 20$
A = {adhd}; B = {anxiety}; C = {depression}	$ A \cup B \cup C = 496; A \cap B \cap C = 24$
A = {adhd}; B = {anxiety}; C = {ptsd}	$ A \cup B \cup C = 495; A \cap B \cap C = 33$
A = {adhd}; B = {bipolar}; C = {depression}	$ A \cup B \cup C = 471; A \cap B \cap C = 36$
A = {adhd}; B = {bipolar}; C = {ptsd}	$ A \cup B \cup C = 477; A \cap B \cap C = 42$
A = {adhd}; B = {depression}; C = {ptsd}	$ A \cup B \cup C = 463; A \cap B \cap C = 15$
A = {anxiety}; B = {bipolar}; C = {depression}	$ A \cup B \cup C = 490; A \cap B \cap C = 49$
A = {anxiety}; B = {bipolar}; C = {ptsd}	$ A \cup B \cup C = 491; A \cap B \cap C = 37$
A = {anxiety}; B = {depression}; C = {ptsd}	$ A \cup B \cup C = 490; A \cap B \cap C = 38$
A = {bipolar}; B = {depression}; C = {ptsd}	$ A \cup B \cup C = 477; A \cap B \cap C = 27$

Tabel 5.15

Komposisi data hasil generasi empat dan lima target kelas

Target kelas	Komposisi kelas dicapai (setiap 500 data)
A={anxiety}; B={bipolar}; C={depression}; D = {ptsd}	$ A \cup B \cup C \cup D = 498; A \cap B \cap C \cap D = 6$
A = {adhd}; B = {bipolar}; C={depression}; D = {ptsd}	$ A \cup B \cup C \cup D = 471; A \cap B \cap C \cap D = 3$

A = {adhd}; B = {anxiety}; C={depression}; D = {ptsd}	$ A \cup B \cup C \cup D = 481$; $ A \cap B \cap C \cap D = 3$
A = {adhd}; B = {anxiety}; C={bipolar}; D = {ptsd}	$ A \cup B \cup C \cup D = 490$; $ A \cap B \cap C \cap D = 2$
A = {adhd}; B = {anxiety}; C = {bipolar}; D = {depression}	$ A \cup B \cup C \cup D = 491$; $ A \cap B \cap C \cap D = 5$
A = {adhd}; B = {anxiety}; C = {bipolar}; D = {depression}; E = {ptsd}	$ A \cup B \cup C \cup D \cup E = 482$; $ A \cap B \cap C \cap D \cap E = 0$

Tabel 5.12, 5.13, dan 5.14 menunjukkan komposisi dari data yang berhasil di-generate dengan metode *single-to-multi-label*. Berdasarkan hasil tabel tersebut, semakin banyak target label yang diinginkan, maka kemunculan data yang sesuai dengan target akan menurun. Dari data yang dihasilkan dari proses diatas, akan diambil sebanyak 25 sampel data secara acak untuk diverifikasi oleh pakar psikolog. Daftar lengkap dari sampel yang diambil bisa dilihat pada Lampiran 2.

Berdasarkan hasil validasi yang dilakukan pakar psikolog, diperoleh skor 0.86 berdasarkan metrik Jaro-Winkler untuk kemiripan hasil klasifikasi yang dilakukan oleh pakar dengan klasifikasi oleh mesin. Selain itu, berikut beberapa ulasan terhadap data yang dihasilkan mesin berdasarkan pakar psikolog:

- Banyak kalimat yang tidak jelas dan tidak koheren.
- Untuk hasil klasifikasi mesin ada beberapa yang bisa termasuk tetapi tidak disebutkan.
- Banyak kalimat *self diagnosed* penyakit tertentu dan validator tidak tahu apakah orang yang mengatakan itu paham apa itu penyakit yang dideritanya atau tidak.

5.1.4. Model Klasifikasi *Multi-label*

Pengujian ini dilakukan pada model klasifikasi *multi-label* varian BERT untuk mengetahui pengaruh performa model terhadap dari data train yang memiliki *multi-label*. Maka dilakukan perbandingan antara model yang hanya menggunakan data *single-label* sebagai data latih dengan model yang menggunakan data tambahan berupa *multi-label* sebagai data latih.

Tabel 5.16

Hasil klasifikasi multi-label varian BERT

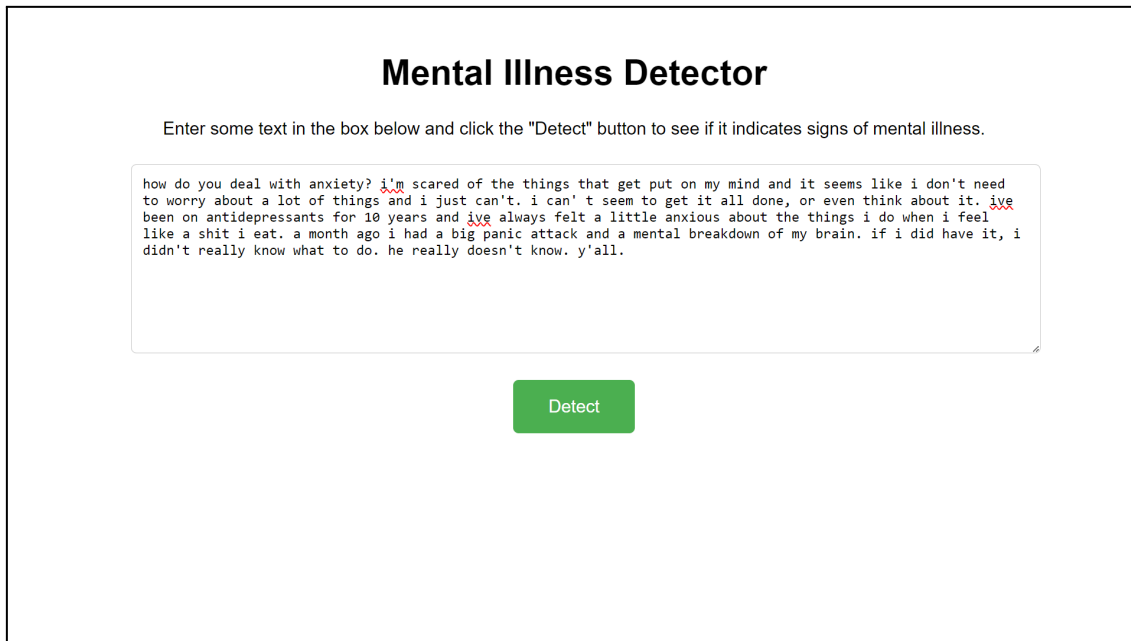
Data dan Model <i>pretrained</i>	Accuracy	Precision	Recall	F1-score
<i>Single-label</i> + BERT	0.81	0.88	0.52	0.66
<i>Multi-label</i> + BERT	0.86	0.91	0.67	0.77
<i>Multi-label</i> + ALBERT	0.91	0.86	0.85	0.83
<i>Multi-label</i> + RoBERTa	0.91	0.87	0.86	0.84

Tabel 5.15 menunjukkan bahwa penggunaan set data *multi-label* membuat performa dari model varian BERT meningkat. Dari hasil tersebut, model *pre-trained* RoBERTa memiliki performa tertinggi diantara semua model varian BERT.

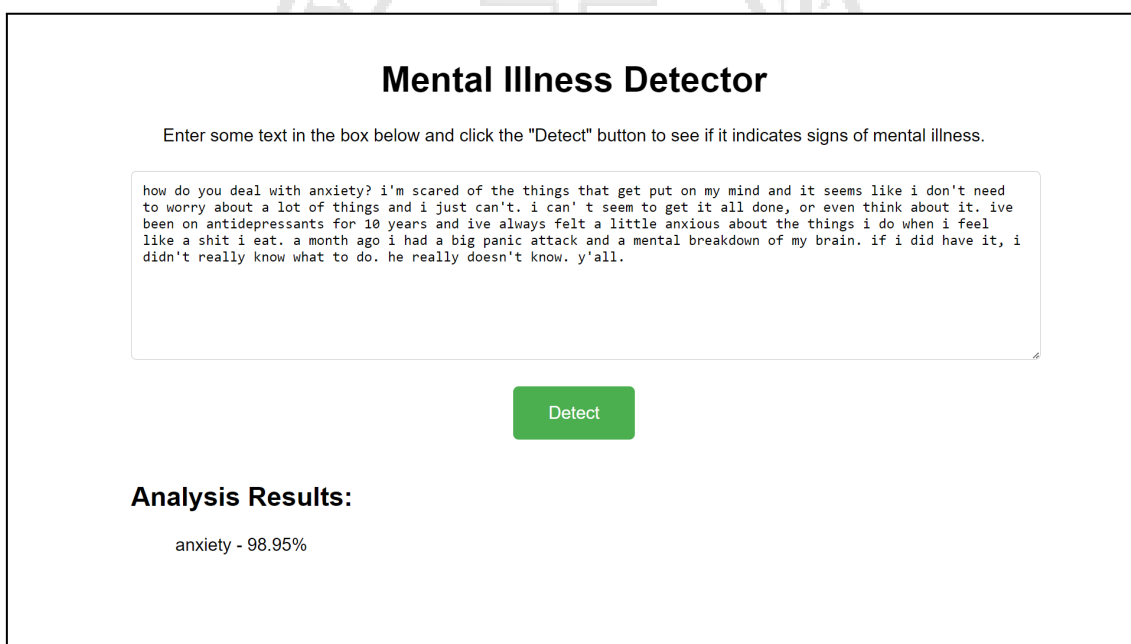
5.2. Pengujian Aplikasi Web

Pengujian ini dilakukan pada aplikasi web pendeteksi gangguan mental. Terdapat beberapa macam gangguan mental yang akan dideteksi, diantaranya ADHD, *anxiety*, bipolar, depresi, dan PTSD.

Gambar 5.1 Tampilan awal aplikasi web



Gambar 5.2 Tampilan setelah pengguna memasukkan teks



Gambar 5.3 Tampilan setelah program mendeteksi teks masukan pengguna

Pada Gambar 5.1, Aplikasi akan menampilkan halaman utama. Ketika pengguna memasukkan teks yang ingin dideteksi, maka pengguna dapat memasukkan teksnya ke dalam bagian input seperti pada Gambar 5.2. Kemudian setelah pengguna memilih “Detect”, maka

program akan mulai mendeteksi teks dan akan memunculkan hasil analisisnya seperti Gambar 5.3.

