Menganalisis Konten Negatif *Cyberbullying* pada Media Sosial Twitter dengan Metode RoBERTa

Tugas Akhir

diajukan untuk memenuhi salah satu syarat

memperoleh gelar sarjana

dari Program Studi Informatika

Fakultas Informatika

Universitas Telkom

1301194233

Muh Akib A Yani

Program Studi Sarjana S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2023

LEMBAR PENGESAHAN

Menganalisis Konten Negatif *Cyberbullying* pada Media Sosial Twitter dengan Metode RoBERTa

Analyzing the Negative Content of Cyberbullying on Twitter Social Media with the RoBERTa Method

NIM : 1301194233

Muh Akib A Yani

Tugas akhir ini telah diterima dan disahkan untuk memenuhi sebagian syarat memperoleh gelar pada Program Studi Sarjana Informatika

Fakultas Informatika

Universitas Telkom

Bandung, 23 Januari 2023

Menyetujui

Pembimbing I,

Dr. Warih Maharani

NIP: 01780020

Ketua Program Studi
Sarjana Informatika,

Dr. Erwin Budi Setiawan, S.Si., M.T.
NIP: 00760045

LEMBAR PERNYATAAN

Dengan ini saya, Muh Akib A Yani, menyatakan sesungguhnya bahwa Tugas Akhir saya dengan judul Menganalisis Konten Negatif *Cyberbullying* pada Media Sosial Twitter dengan Metode RoBERTa beserta dengan seluruh isinya adalah merupakan hasil karya sendiri, dan saya tidak melakukan penjiplakan yang tidak sesuai dengan etika keilmuan yang belaku dalam masyarakat keilmuan. Saya siap menanggung resiko/sanksi yang diberikan jika di kemudian hari ditemukan pelanggaran terhadap etika keilmuan dalam buku TA atau jika ada klaim dari pihak lain terhadap keaslian karya,

Bandung, 23 Januari 2023

Yang Menyatakan

Muh Akib A Yani

Menganalisis Konten Negatif Cyberbullying pada Media Sosial Twitter dengan Metode RoBERTa

Muh Akib A Yani[1], Warih Maharani[2]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]muhammadakib@students.telkomuniversity.ac.id, [2]wmaharani@telkomuniversity.ac.id

Abstrak

Media sosial adalah sebuah sarana komunikasi berbasis daring yang dapat mempermudah penggunanya dalam berinteraksi antar sesama pengguna tanpa adanya batasan wilayah dan waktu. Indonesia yang memiliki angka pengguna sosial media tertinggi di dunia. *Platform* sosial media Twitter merupakan salah satu tempat dimana para penggunanya dapat mencurahkan seluruh isi hati dalam bentuk cuitan. Interaksi yang bebas serta beragam pada Twitter memiliki pengaruh yang cukup besar dalam kondisi psikologis penggunanya. *Cyberbullying* atau perundungan secara daring adalah suatu tindakan penghinaan atau menyakiti perasaan orang lain secara sengaja dan berulang–ulang pada sosial media, pesan, atau dengan cara lain. Pada penelitian ini menggunakan metode klasifikasi RoBERTa untuk mendeteksi tweet yang mengandung *cyberbullying* dengan skor terbaik *accuracy* 86.9% dan *f1-score* 77.5%.

Kata kunci : media sosial, twitter, *cyberbullying*, RoBERTa.

Abstract

Social media is an online-based communication tool that can make it easier for users to interact among fellow users without any area and time restrictions. Indonesia has the highest number of social media users in the world. The Twitter social media platform is a place where users can pour out their whole hearts in the form of tweets. Free and diverse interactions on Twitter have a considerable influence on the psychological condition of its users. Cyberbullying or online bullying is an act of humiliating or hurting other people's feelings intentionally and repeatedly on social media, messages, or in other ways. In this study, the RoBERTa classification method was used to detect cyberbullying tweets with a best accuracy score of 86.9% and an f1-score of 77.5%.

Keywords: social media, twitter, cyberbullying, RoBERTa

1. Pendahuluan

1.1. Latar Belakang

Media sosial adalah sebuah sarana komunikasi berbasis daring yang dapat mempermudah penggunanya dalam berinteraksi antar sesama pengguna tanpa adanya batasan wilayah dan waktu. Indonesia yang memiliki angka pengguna sosial media tertinggi di dunia. Kementerian Komunikasi dan Informatika (Kemenkominfo) menyatakan ada sebanyak 95% dari 63 juta pengguna internet adalah pengguna sosial[1]. Salah satu media sosial yang sering digunakan para pengguna yaitu Twitter, *Country Industry Head Twitter* Indonesia mengklaim bahwa negara Indonesia merupakan negara dengan pertumbuhan aktif pengguna harian sosial media Twitter yang paling tinggi[1].

Namun tidak semua pengguna dapat menggunakan media sosial Twitter dengan bijak. Tak sedikit pengguna menggunakan media sosial Twitter untuk melakukan Tindakan yang dapat merugikan pengguna lain seperti penyebaran berita hoaks, penipuan, hingga pelecehan atau perundungan secara daring (*cyberbullying*) masalah ini timbul seiring dengan perkembangan media sosial akhir-akhir ini. *Cyberbullying* dapat diartikan sebagai penggunaan media sosial oleh individu atau sekolompok pengguna untuk melecehkan pengguna lain yang dapat mengakibatkan pengguna lain merasakan efek negatif terhadap korban. Satu studi yang dilakukan oleh badan amal anti-intimidasi nasional menunjukkan bahwa dua dari tiga anak berusia 13-22 tahun yang disurvei telah menjadi korban perundungan secara daring (*cyberbullying*)[2].

Oleh karena itu, penelitian ini bertujuan untuk mendeteksi perundungan secara daing (*cyberbullying*) pada *platform* Twitter. Yang menjadi tolak ukur dalam penelitian ini berdasarkan timbal balik serta interaksi cuitan pengguna yang ada di Twitter. Sebelumnya telah dilakukan penelitan yang dilakukan oleh Yinhan Liu mengenai perluasan arsitektur BERT yang telah dikembangkan dan diberi nama RoBERTa, penelitian tersebut dilakukan sebagai pengembangan arsitektur BERT yang menunjukkan hasil secara signifikan kurang terlatih dibandingkan dengan model setelahnya. Maka dari itu metode yang akan digunakan pada penelitian ini adalah RoBERTa yang

dimana RoBERTa merupakan pelatihan ulang BERT dengan metodologi pelatihan yang ditingkatkan, memungkinkan menerima data yang lebih banyak, dan daya komputasi. Selain itu, metode ini menghilangkan *Next Sentence Prediction* (NSP) pada BERT dan menggunakan *Dynamic Masking*. Dengan beberapa kelebihan yang dimiliki oleh RoBERTa, membuat metode ini dapat diandalkan dengan peningkatan performa dibandingkan dengan metode BERT[3].

1.2. Topik dan Batasan

Penelitian ini membahas terkait deteksi tweet masyarakat Indonesia yang mengandung *cyberbullying* dengan kata kunci "gendut" , "makan", "pemerintah", "kecebong", "anj*ng", "tol*l" dan beberapa kata yang mengandung kata kasar serta menggunakan kata umpatan hewan melalui media sosial twitter dengan metode klasifikasi BERT yang telah dikembangkan yaitu RoBERTa.

1.3. Tujuan

Tujuan dilakukannya penelitian ini adalah untuk mengetahui performansi metode klasifikasi RoBERTa dengan mencari nilai akurasi terbaik terhadap cuitan masyarakat Indonesia pada media sosial twitter yang mengandung *cyberbullying*.

1.4. Organisasi Tulisan

Pada laporan penelitian ini, akan dibahas studi literatur yang terkait dan berhubungan yang menjadi acuan penelitian pada Bab 2. Kemudian sistem yang dibangun pada penelitian ini dapat dilihat pada Bab 3. Hasil dari penelitian yang telah dilakukan dapat dilihat pada Bab 4. Serta kesimpulan dari penelitian ini tertera pada Bab 5.

2. Studi Terkait

Media sosial adalah sebuah wadah komunikasi berbasis daring yang memungkinkan para penggunanya bisa dengan mudah saling berinteraksi tanpa adanya batasan jarak, waktu, dan tempat. Adapun dampak positif dari media sosial yaitu memudahkan para pengguna untuk berinteraksi dengan banyak orang, memperluas pergaulan, jarak dan waktu bukan lagi masalah, lebih mudah dalam mengekspresikan diri, penyebaran informasi dapat berlangsung secara cepat, biaya lebih murah[9]. Twitter adalah salah satu microblogging yang dikembangkan oleh Twitter, Inc. Disebut microblogging karena platform ini memungkinkan penggunanya mengirim dan membaca pesan seperti blog pada umumnya. Cuitan tersebut dinamakan tweet, yaitu teks tulisan yang memiliki batas jumlah 140 karakter yang dipublikasikan pada halaman profil pengguna[8].

Cyberbullying atau perundungan secara daring adalah suatu tindakan penghinaan atau menyakiti perasaan orang lain secara sengaja dan berulang–ulang pada sosial media, pesan, atau dengan cara lain. Kegiatan cyberbullying di platform media sosial Twitter dilakukan dengan cara mempublikasikan cuitan yang mengandung kata-kata kasar atau kata-kata hinaan, hingga kata-kata yang menjurus kepada penghinaan terhadap SARA[1].

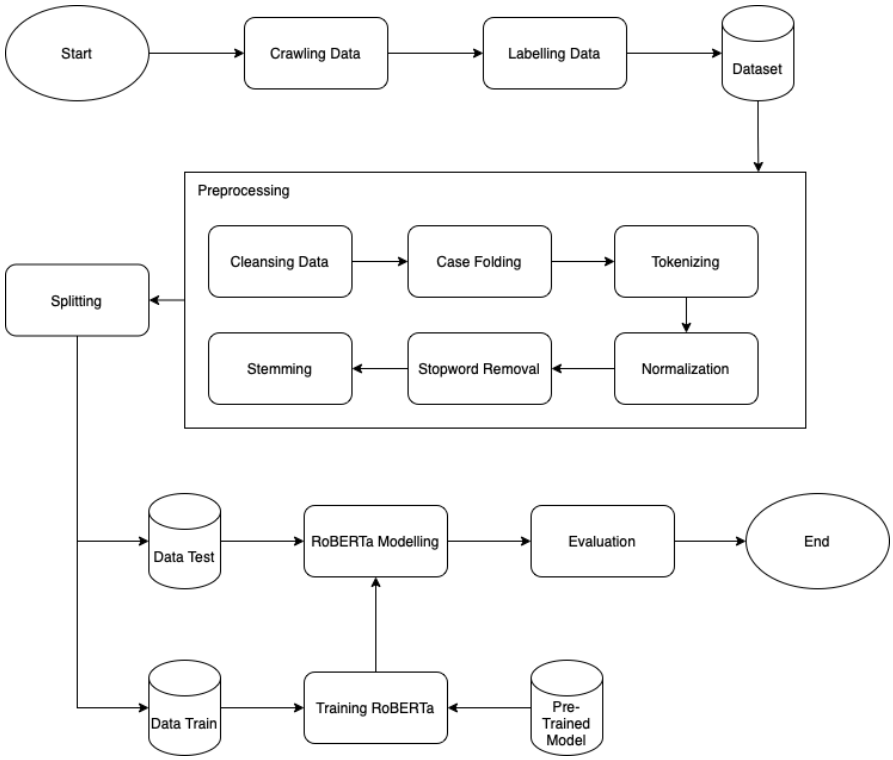
Pada penelitian mengenai deteksi *cyberbullying* telah dilakukan dengan berbagai metode klasifikasi. A.Saravanaraj, J *et al*[2] melakukan penelitian mengenai deteksi *cyberbullying* pada media sosial Twitter menggunakan metode klasifikasi *Naïve Bayes* dan *Random Forest*. Tahapan kegiatan penilitiannya terdiri dari pengumpulan data, pra-pemrosesan, klasifikasi, dan evaluasi. Penelitian ini menghasilkan nilai akurasi sebesar 90%.

Nassharih Abdulloh *et al*[4] melakukan penelitian mengenai deteksi *cyberbullying* pada cuitan media sosial Twitter menggunakan 4 metode klasifikasi yaitu *Multinomial Naïve Bayes*, *Linear SVM*, *Logistic Regression*, dan *KNN*. Tahapan kegiatan penelitiannya terdiri dari pengumpulan data, *preprocessing*, ekstraksi fitur, klasifikasi, dan evaluasi. Penelitian ini menghasilkan nilai akurasi sebesar 0.961 untuk algoritma *Multinomial Naïve Bayes*, 0.994 untuk algoritma *Logistic Regression*, 0.997 untuk algoritma *Linear SVM*, dan 0.918 untuk algoritma *KNN*.

Lida Ketsbaia *et al*[6] melakukan penelitian mengenai deteksi *cyberbullying* menggunakan dua *dataset* yang bersumber dari media sosial Twitter dengan memberi 3 label yaitu "*hate speech*", "*offensive*", "*neither hate-speech or offensive*" dan menggunakan 3 metode klasifikasi yaitu *RoBERTa*, *XLNet* dan *DistilBERT*. Tahapan kegiatan penelitiannya terdiri dari pengumpulan data, *preprocessing*, fitur *tf-idf*, teknik penyeimbangan *dataset* yaitu *SMOTE* dan *Random Under Sampling*, fitur seleksi, *Logistic Regression*, *Particle Swarm Optimisation*, *Genetic Algorithm*, metodek klasifikasi, dan evaluasi. Dan didapatkan hasil dari akurasi untuk *dataset* pertama yaitu 86.78% untuk algoritma *DistilBERT*, 87.78% untuk algoritma *XLNet*, 86.43% untuk algoritma *RoBERTa*. Sedangkan untuk *dataset* yang kedua didapatkan nilai akurasi sebesar 94% untuk algoritma *DistilBERT*, 94.47% untuk algoritma *XLNet*, 93.89% untuk algoritma *RoBERTa*.

3. Sistem yang Dibangun

Pada penelitian ini, sistem yang dibangun mampu mendeteksi *cyberbullying* pada cuitan twitter. Terdapat beberapa tahap untuk melakukan deteksi yaitu, *crawling*, *labelling data*, *preprocessing data*, *splitting data* dan terbagi menjadi dua yaitu *training dataset* dan *testing dataset* yang dimana *training dataset* digunakan untuk melatih algoritma yang akan digunakan sedangkan *testing dataset* digunakan untuk mengetahui performa algoritma yang sudah dilatih sebelumnya, klasifikasi RoBERTa, dan evaluasi. Adapun skema pada penelitian ini yang dapat dilihat pada gambar 1.



Gambar 1. Flowchart Perancangan Sistem

3.1. Crawling Data

Pada peneliatian ini, dataset merupakan data yang diambil berdasarkan cuitan -tweet masyarakat Indonesia pada media sosial twitter dengan kata kunci "gendut" , "makan", "pemerintah", "kecebong", "anj*ng", "tol*l" dan beberapa kata yang mengandung kata kasar serta menggunakan kata umpatan hewan.

Proses *crawling data* dilakukan menggunakan website netlytic dengan format *Comma Separated Value* (CSV) yang berisikan opini masyarakat Indonesia terkait konten-konten yang ada pada media sosial twitter.

3.2. Labelling Data

Setelah dilakukan *crawling data* selanjutnya *dataset* yang telah dikumpulkan akan melalui proses *labelling data* secara manual dimana data dibagi menjadi 2 label yaitu positif dan negatif. Pemberian label dilakukan oleh 3 orang dengan tujuan untuk mengurangi subjektivitas dalam melakukan pelabelan. Pemberian label dilakukan dengan cara memperhatikan kata yang terdapat pada data *tweets* yang sudah di *crawling*, apabila di dalam data *tweets* terdapat kata kasar atau yang mengandung *cyberbullying* maka diberikan label negatif atau 1, apabila data tersebut tidak mengandung kata kasar atau tidak mengandung *cyberbullying* maka diberikan label positif atau 0.

Cyberbullying atau kata kasar memiliki artian umpatan, cacian, ejakan, atau kata yang mengandung kata umpatan hewan.

Tabel 1. Contoh Pelabelan Data

Label	Tweet
Positif	kemarennya lagi dibilang gendut sama keluarga nyokap

Negatif	orang seperti fadli zonk , harusnya di keluarkan dr parlemen, kerja g becus, hobi bacot dan nyinyir, cuma doyan makan doang, tuh sampe gendut , otaknya menciut

3.3. BERT

BERT adalah metode klasifikasi yang merupakan model representasi bahasa yang dirancang untuk melatih representasi dua arah dari teks yang tidak berlabel dengan menyesuaikan konteks kiri dan kanan di semua lapisan. Representasi *Encoder* Dua Arah dari *Transformers* (BERT) mengoptimalkan *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP) dalam proses *Pre-Trained*. *Masked Model Language* (MLM) adalah proses untuk memprediksi kata-kata yang muncul dari komentar sebelumnya. *Next Sentence Prediction* (NSP) adalah hilangnya klasifikasi biner yang berfungsi memunculkan dua kata yang saling mengikuti dalam sebuah teks. BERT adalah model representasional berbasis finetuning pertama yang mengungguli banyak arsitektur. Analisis pelatihan model BERT mengeksplorasi dan menghitung opsi penting untuk melatih model BERT sambil mempertahankan model arsitektural. Ini dimulai dari melatih model BERT dengan konfigurasi yang sama dengan BERT dasar (pelatihan²L = 12, H= 768, A = 12, params 110M)[11].

3.4. RoBERTa

Robustly Optimized BERT Approach (RoBERTa) adalah versi replikasi dari pendekatan *Pre-Trained* dari *BERT*, yang telah dioptimalkan dan dapat mendeteksi teks yang tidak memiliki notasi dengan jumlah maksimum 160GB. RoBERTa menghapus *Next Sentence Prediction* (NSP) dan menambahkan penyembunyian kata dinamis selama periode pelatihan. Perubahan dan fitur ini mengidentifikasi peningkatan kinerja dibandingkan dengan BERT di banyak tugas NLP, termasuk klasifikasi teks. Dengan demikian, RoBERTa menggunakan lebih banyak data untuk melatih, meningkatkan ukuran *batch*, menghilangkan prediksi kerugian berikutnya, dan mengganti masking statis dengan masking dinamis pada tahap pra-pelatihan untuk lebih meningkatkan performa. RoBERTa akan memberikan hasil yang lebih optimal dibandingkan BERT karena adanya modifikasi tersebut. Hal ini dibuktikan dengan penelitian menggunakan arsitektur BERT besar L=24, H=1024, A=16, parameter 355M[12].

3.5. *Preprocessing*

Setelah dilakukan pengumpulan data, langkah selanjutnya adalah melakukan pengubahan data. Berikut beberapa tahap dalam melakukan pengubahan data :

1) *Cleansing Data*

*Cleansing Da*ta merupakan proses membersihkan atau menghapus hal-hal yang tidak diperlukan seperti tanda baca, angka, URL, dan kata-kata yang dianggap tidak penting[5]. Contoh data yang telah melewat proses cleansing dapat dilihat pada tabel 2.

Tabel 2. *Cleansing Data*

Tweet	***Cleansing***
Memakan banyak tapi gak gendut : Kuli Indon https://t.co/4bLEdvTvEc	Memakan banyak tapi gak gendut Kuli Indon

2) *Case Folding*

Pada proses ini kata yang memiliki *uppercase* atau huruf besar diganti menjadi *lowercase* atau huruf kecil. Proses ini dilakukan untuk menghindari adanya duplikasi yang dibedakan dari huruf besar kecil (*case- sensitive*). Data yang telah melewati *case folding* akan terlihat seperti pada tabel 3.

Tabel 3. *Case Folding*

Cleansing	***Case Folding***
Memakan banyak tapi gak gendut Kuli Indon	memakan banyak tapi gak gendut kuli indon

3) *Tokenizing*

Tokenization merupakan proses membagi atau memecah suatu kalimat yang sebelumnya dipisah oleh spasi menjadi kata-kata yang menyusunnya. *Tokenization* ini perlu dilakukan untuk memudahkan klasifikasi. Contoh data yang telah di–*tokenize* terdapat pada tabel 4.

Tabel 4. ***Tokenizing***

Case Folding	*Tokenizing*
memakan banyak tapi gak gendut kuli indon	[‘memakan’, ‘banyak’, ‘tapi’, ‘gak’, ‘gendut’, ‘kuli’, ‘indon’]

4) Normalization

Pada tahap normalization dilakukan pengubahan kata-kata yang disingkat, kata yang salah dalam penulisan, dan kata–kata yang tidak formal menjadi kata yang baku dan sesuai penulisan KBBI. Kamus yang dipakai merupakan kamus kata KBBI yang telah dibuat sebelumnya. Contoh *normalization* dapat dilihat pada tabel 5.

Tabel 5. ***Normalization***

Tokenizing	*Normalization*
[‘memakan’, ‘banyak’, ‘tapi’, ‘gak’, ‘gendut’, ‘kuli’, ‘indon’]	[‘memakan’, ‘banyak’, ‘tapi’, ‘tidak, ‘gendut’, ‘kuli’, ‘indon’]

5) Stopword Removal

Stopword Removal merupakan tahap penghapusan kata sambung yang sering muncul. Kata-kata ini biasanya memiliki fungsi namun tidak memiliki makna yang berarti dan tidak memberi bobot pada suatu opini atau kalimat. Contoh stopword removal dapat dilihat pada tabel 6.

Tabel 6. ***Stopword Removal***

Normalization	*Stopword Removal*
[‘memakan’, ‘banyak’, ‘tapi’, ‘tidak, ‘gendut’, ‘kuli’, ‘indon’]	[‘memakan’, ‘banyak’, ‘tapi’, ‘tidak, ‘gendut’]

6) Stemming

Stemming merupakan tahap membersihkan kata imbuhan yang meliputi awalan, akhiran atau gabungan keduanya. Dengan stemming, kata yang memiliki kata dasar yang sama akan dianggap memiliki token yang sama. Hal ini membantu dalam meningkatkan kinerja pemrosesan data. Hasil data stemming dapat dilihat pada tabel 7.

Tabel 7. ***Stemming***

Stopword Removal	*Stemming*
[‘memakan’, ‘banyak’, ‘tapi’, ‘tidak, ‘gendut’]	[‘makan’, ‘banyak’, ‘tapi’, ‘tidak, ‘gendut’]

3.6. RoBERTa Tokenize

Setelah *dataset* dibagi menjadi *data train* dan *data test* selanjutnya dilakukan *encoding* untuk setiap *dataset*. Model *pre-trained* pada model RoBERTa diperlukan untuk melakukan sebuah *encoding* pada *dataset*. Pada penelitian ini digunakan model *pre-trained* RoBERTa dari ayameRushia yang menggunakan pemodelan *masked language modeling*. Model ini sebelumnya dilatih menggunakan Wikipedia Indonesia dan input dari model ini adalah sebuah kata, kalimat, dan paragraf. Dalam proses *encoding* terdapat tiga *input*-an yang akan dimasukkan dalam model yaitu *input_ids*, *attention_mask*, dan *token_type_*ids. Pada tahap ini *dataset* akan di- *tokenize* di dalam proses RoBERTa *modelling*. Proses ini akan disesuaikan dengan model *pre-trained* RoBERTa yang dimana model tersebut akan mendeteksi kata atau kalimat dari *dataset*. Kata atau kalimat yang ada pada dalam *dataset* akan digabungkan ketika model *pre-trained* dapat memahami kalimat dan akan dipisah ketika tidak dapat memahami kalimat yang ada pada *dataset*[7]. Dapat dilihat contoh pada tabel 8.

Tabel 8. ***Tokenize*** **RoBERTa**

Tahapan	**Hasil**

Dataset	ayoo diet lagi biar gak dibilang nambah gendut
Tokenize RoBERTa	['ay', 'oo', 'Ġdiet', 'Ġlagi', 'Ġbiar', 'Ġga', k', 'Ġdibilang', 'Ġn', 'ambah', 'Ġgend', 'ut']

3.7. Confusion Matrix

Selanjutnya tahapan terakhir yaitu mengevaluasi tahapan-tahapan yang telah diproses sebelumnya. Pada tahap ini diperlukan untuk dapat menguji tingkat akurasi dari metode sebelumnya. Untuk mendapatkan hasil evaluasi, diperlukan pengukuran dengan metode Confusion Matrix yang memiliki 4 karakteristik, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Kinerja suatu matriks diukur berdasarkan *accuracy*, *precision*, *recall*, dan *f1-sco*re yang dapat diuji berdasarkan TP, TN, FP, dan FN[10]. Berikut ini merupakan contoh dari perhitungan nilai evaluasi :

1) *Accuracy*
   Accuracy mempresentasikan rasio yang akan diklasifikasikan dengan rumus :

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2) *Precision*
   Precision merupakan nilai yang didapatkan dari akurasi suatu kelas dengan jumlah total prediksi untuk kelas tersebut. Tujuan dari precision adalah untuk melihat persentase relevansi dari hasil klasifikasi dengan rumus :

$$precision = \frac{TP}{TP + FP} \quad (2)$$

3) *Recall*
   Recall merupakan nilai yang didapatkan dari akurasi prediksi suatu kelas dengan jumlah total fakta untuk kelas tersebut. Recall dapat dihitung dengan rumus :

$$recall = \frac{TP}{TP + FN} = \frac{TN}{P} \quad (3)$$

4) *F1-Score*
   F1-Score merupakan perhitungan evaluasi yang dilakukan dengan menggabungkan kedua nilai precision dan recall. F1-Score dapat dihitung dengan rumus :

$$F1 = \frac{2.(Recall.Precision)}{Recall + Precision} \quad (4)$$

4. Evaluasi

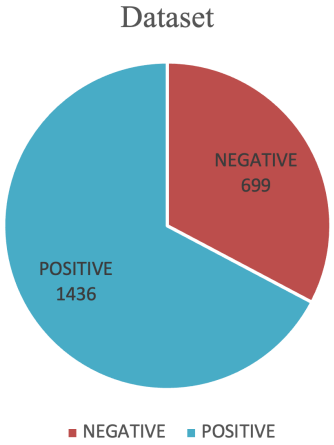
4.1. Data

Data yang digunakan pada penelitian ini berjumlah 3034 tweet berbahasa Indonesia dengan kata kunci “gendut” , “makan”, “pemerintah”, “kecebong”, “anj*ng”, “tol*l” dan beberapa kata yang mengandung kata kasar serta menggunakan kata umpatan hewan yang telah di *crawling* sebelumnya menggunakan *website* netlytic. Setelah itu dilakukan *preprocessing* untuk menyaring data yang akan diuji dan setelah didapatkan *dataset* akhir yang berjumlah 2135 tweet.

Tabel 9. Persebaran Data

Label	Jumlah Data
Positif	1436
Negatif	699

Pemberian label dilakukan berdasarkan tweet. Label 1 untuk data yang mengandung *cyberbullying* atau negatif dan 0 untuk data yang tidak mengandung *cyberbullying* atau positif.



4.2. Skenario dan Hasil Pengujian

Pada penelitian ini dilakukan 2 skenario perbandingan *preprocessing*, yaitu *Preprocessing* yang berakhir pada bagian *case folding* atau tidak lengkap dan *preprocessing* sampai pada tahap akhir atau *stemming*. Setelah dilakukan dua skenario *preprocessing* dilanjutkan sampai dengan tahap klasifikasi dengan melakukan 5 perbandingan *train dataset* dan *test dataset* dan didapatkan hasil sebagai berikut :

Tabel 10. Perbandingan *preprocessing* dan *full preprocessing* pada *data train*

Preprocessing ***(Cleansing Data, Case Folding)***		***Full Preprocessing***	
Data Train	***Accuracy***	***Data Train***	***Accuracy***
90%	**86.92%**	90%	**82.24%**
80%	83.23%	80%	81.97%
70%	84.09%	70%	80.81%
60%	83.14%	60%	80.44%
50%	81.27%	50%	77.06%

Setelah dilakukan 2 skenario perbandingan *preprocessing* didapatkan nilai akurasi tertinggi pada data yang telah di-*train* yaitu pada *data train* sebesar 90%, setelah mendapatkan nilai akurasi tertinggi selanjutnya dilakukan pengujian dengan 2 skenario, yaitu :

1) Pengujian performansi menggunakan *preprocessing* (*cleansing data, case folding)* dengan *data train* sebesar 90%
2) Pengujian performansi menggunakan *full preprocessing* dengan *data train* sebesar 90%

4.2.1. Skenario Pengujian Pertama

Skenario pertama dilakukan dengan *preprocessing* yang tidak lengkap atau hanya *cleansing data* dan *case folding* dengan menggunakan *data train* sebesar 90% dengan menguji ukuran *batch size* sebesar 8, 16, dan 32 dengan melakukan 5 kali percobaan. Hasilnya dapat terlihat dari tabel 11, 12, dan 13.

Tabel 11. ***Preprocessing batch size 8***

EPOCH	*Accuracy*
1	82.7%
2	84.1%
3	84.5%
4	**86.9%**
5	83.6%

Tabel 12. ***Preprocessing batch size 16***

EPOCH	*Accuracy*
1	85.1%
2	85.5%
3	**86.9%**
4	80.8%
5	85.1%

Tabel 13. ***Preprocessing batch size 32***

EPOCH	*Accuracy*
1	80.8%
2	83.1%
3	**85.1%**
4	83.1%
5	83.6%

4.2.2. Skenario Pengujian Kedua

Skenario pertama dilakukan dengan *full preprocessing* menggunakan *data train* sebesar 90% dengan menguji ukuran *batch size* sebesar 8, 16, dan 32 dengan melakukan 5 kali percobaan. Hasilnya dapat terlihat dari tabel 14, 15, dan 16.

Tabel 14. ***Full preprocessing batch size 8***

EPOCH	*Accuracy*
1	78.0%
2	70.5%
3	**85.1%**
4	77.5%
5	82.7%

Tabel 15. ***Full preprocessing batch size 16***

EPOCH	*Accuracy*
1	81.7%
2	83.6%
3	83.1%
4	84.5%
5	**85.9%**

Tabel 16. ***Full preprocessing batch size 32***

EPOCH	*Accuracy*
1	80.8%
2	83.2%
3	**85.1%**
4	83.2%
5	83.6%

4.3. Analisis Hasil Pengujian

Tabel 17. Perbandingan akurasi

	Accuracy	*F1-Score*
***Preprocessing batch size* 8**	86.9%	76.3%

***Preprocessing batch size* 16**	86.9%	77.5%
***Preprocessing batch size* 32**	85.1%	71.6%
***Full preprocessing batch size* 8**	85.1%	75.3%
***Full preprocessing batch size* 16**	85.9%	75.7%
***Full preprocessing batch seze* 32**	85.1%	70.1%

Terdapat perbandingan skor performansi pada tabel 17, dapat ditemukan bahwa ukuran *batch size* sebesar 16 pada pengujian menggunakan *preprocessing* dan *full preprocessing* mendapatkan nilai akurasi dan *f1-score* sebesar 86,9% nilai akurasi dan 77.5% nilai *f1-score* untuk *preprocessing* sedangkan untuk *full preprocessing* mendapatkan skor 85.9% akurasi dan 75.7% *f1-score*. Dapat dipastikan bahwa ukuran *batch size* sebesar 16 memiliki skor lebih tinggi dibandingkan *batch size* sebesar 8 dan 32.

Berdasarkan hasil pengujian dari dua skenario yang telah dilakukan, ditemukan bahwa *dataset* yang diproses melalui tahap *preprocessing* tanpa melalui tahap *normalization*, *stopwrod removal*, dan *stemming* mendapatkan nilai akurasi yang lebih tinggi dibandingkan *dataset* yang diproses secara *full preprocessing*, dikarenakan *dataset* yang diuji tanpa penghapusan kata yang dianggap tidak baku dan tanpa melalui proses membersihkan kata imbuhan dapat meningkatkan tingkat deteksi pada metode klasifikasi RoBERTa. Rasio perbandingan *dataset* yang telah dilatih dengan *dataset* yang tidak dilatih mempengaruhi nilai akurasi, semakin banyak *dataset* yang dilatih semakin tinggi juga akurasi yang didapatkan. Dan *batch size* pada pengujian mempengaruhi nilai akurasi yang didapatkan dengan semakin berkurang ukuran *batch size* maka kemungkinan tingkat akurasi akan semaking tinggi yang didapatkan karena algoritma konvergen lebih cepat tetapi akan menghasilkan *noise* pada komputasi lebih besar.

5. Kesimpulan

Penelitian ini dilakukan untuk mendeteksi opini masyarakat Indonesia terkait konten yang ada pada media sosial Twitter dengan kata kunci "gendut" , "makan", "pemerintah", "kecebong", "anj*ng", "tol*l" dan beberapa kata yang mengandung kata kasar serta menggunakan kata umpatan hewan. Pada penelitian ini dilakukan dua perbandingan *preprocessing* (*cleansing data* dan *case folding*) dan *full preprocessing* (*cleansing data*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*) berdasarkan kedua pengujian tersebut didapatkan hasil nilai akurasi dari *preprocessing* lebih tinggi dibandingkan dengan menggunakan *full preprocessing*. Pada model RoBERTa sangat mempengaruhi kalimat yang ada pada *dataset* sehingga jika ada kata yang dihapus pada sebuah kalimat maka akan mempengaruhi deteksi dari model RoBERTa yang menyebabkan nilai akurasinya berkurang.

Pada penelitian selanjutnya dapat dilakukan pengujian dengan menggunakan beberapa metode klasifikasi yang berbeda dengan cara membandingkan dengan metode klasifikasi RoBERTa agar dapat memberoleh nilai akurasi yang berbeda dari setiap metode.

Daftar Pustaka

[1] H. F. Fadli and A. F. Hidayatullah, "Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM." (accessed May 13, 2022)

[2] A. Saravanaraj, J. I. Sheeba, and S. P. Devaneyan, "AUTOMATIC DETECTION OF CYBERBULLYING FROM TWITTER." [Online]. Available: https://www.researchgate.net/publication/333320174 (accessed May 13, 2022)

[3] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.11692 (accessed May 13, 2022)

[4] N. Abdulloh and A. F. Hidayatullah, "Deteksi Cyberbullying pada Cuitan Media Sosial Twitter." [Online]. Available: https://t.co/wjwvdPTRBa (accessed May 13, 2022)

[5] Hanson Siagian, Putra Pandu Adikara, Sigit Adinugroho, "Deteksi Konten Negatif di Twitter Menggunakan Support Vector Machine dan Pemisahan Hashtag dengan Algoritme Pipeline" April. 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik (accessed May 13, 2022)

[6] L. Ketsbaia and X. Chen, "Evaluation of Cyberbullying using Optimized Multi-Stage ML Framework and NLP." (accessed May 14, 2022)

[7] Cagri Toraman, Eyup Halit Yilmaz, Furkan Şahinuç, dan Oguzhan Ozcelik, "Impact of Tokenization on Language Models: An Analysis for Turkish" April. 2022, [Online]. Available: https://arxiv.org/abs/2204.08832

[8] N. M. Anggreini, "PEMANFAATAN MEDIA SOSIAL TWITTER DI KALANGAN PELAJAR SMK NEGERI 5 SAMARINDA," eJournal Sosiatri-Sosiologi, vol. 2016, no. 2, pp. 239–251, 2016. (accessed May 16, 2022)

[9] A. S. Cahyono, "PENGARUH MEDIA SOSIAL TERHADAP PERUBAHAN SOSIAL MASYARAKAT DI INDONESIA," Publiciana, vol. 2016, no. 9, 2016. (accessed May 16, 2022)

[10] P. Singh, N. Singh, K. K. Singh, and A. Singh, "Diagnosting Of Disease Using Machine Learning," Mach. Learn. Internet Med. Things Healthc., pp. 89-111, Jan. 2021, doi: 10.1016/B978-0-12-821229-5.00003-3. (accessed May 20, 2022)
[11] Wang X., Chen S., Li T., Li W., Zhou Y., and Zheng J, "Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods : Content Analysis. " July. 2020, [Online]. Available: https://medinform.jmir.org/2020/7/e17958/
[12] Devlin J., Chang M., Lee K., and Toutanova K, " BERT: Pre-training of deep bidirectional transformers for language understanding." May. 2019, [Online]. Available: https://arxiv.org/abs/1810.04805