

KLASIFIKASI EMOSI PADA TEKS DENGAN MENGGUNAKAN METODE DEEP LEARNING

Alan Tusa Bagus W, Dhomas Hatta Fudholi

Fakultas Teknik, Universitas Islam Indonesia, Yogyakarta, Indonesia
Email: 18523148@students.uii.ac.id, hatta.fudholi@uui.ac.id

Abstrak

Emosi merupakan reaksi dari tubuh manusia ketika mendapatkan reaksi terhadap sesuatu yang dapat disampaikan secara nonverbal yaitu dengan menunjukkan ekspresi wajah ataupun perilaku. Dalam dunia digital, lebih tepatnya media sosial yang berbasis teks, manusia mengekspresikan emosinya melalui teks dan perlu dilakukan analisa untuk mengetahui emosi yang disampaikan oleh orang tersebut. Untuk mengatasi masalah diatas, perlu melakukan metode klasifikasi pada kalimat-kalimat opini untuk mengetahui jenis emosi yang ada dalam teks tersebut, lalu emosi tersebut dikelompokkan menjadi sembilan kelas emosi yaitu marah, takut, sedih, netral, bahagia, tertarik, percaya, kaget, dan jijik. Semua jenis emosi tersebut dapat mewakili emosi pada manusia. Metode yang diterapkan pada penelitian ini adalah metode klasifikasi menggunakan BERT (Bidirectional Encoder Representations from Transformers). Dataset yang digunakan adalah data dari opini-opini seseorang di twitter yang dikumpulkan menjadi satu kedalam file csv. Klasifikasi emosi menggunakan BERT menghasilkan nilai akurasi sebesar 89.2% dengan sembilan kelas emosi pada data training, dan menghasilkan akurasi 76 % pada metode indoBERT.

Kata Kunci: Klasifikasi emosi, klasifikasi, BERT, opini

Abstract

Emotion is a reaction from the human body when it gets a reaction to something that can be conveyed non-verbally, namely by showing facial expressions or behavior. In the digital world, more precisely text-based social media, humans express their emotions through text and analysis needs to be done to find out the emotions conveyed by that person. To overcome the problem above, it is necessary to use a classification method on opinion sentences to find out the types of emotions that exist in the text, then these emotions are grouped into nine emotion classes, namely anger, fear, sadness, neutral, happy, interested, believe, surprised, and disgusting. All these types of emotions can represent emotions in humans. The method applied in this study is a classification method using BERT (Bidirectional Encoder Representations from Transformers). The dataset used is data from someone's opinions on Twitter which are collected into one csv file. The classification of emotions using BERT produces an accuracy value of 89.2% with nine emotion classes in the training data, and produces an accuracy of 76% on the indoBERT method.

Keywords: *Emotion classification, classification, BERT, opinion*

Received: 2021-10-20; Accepted: 2021-11-05; Published: 2021-11-20

Pendahuluan

Kehidupan manusia tidak terlepas dengan namanya emosi, emosi merupakan reaksi terhadap seseorang atau suatu kejadian, dan emosi memiliki peranan penting terhadap kehidupan sehari-hari. Emosi dibagi menjadi dua keluaran yaitu positif dan negatif, dari dua keluaran tersebut ada beberapa kategori di dalamnya seperti marah, senang, sedih, takut, dan sebagainya. Analisis emosi sudah menjadi pembahasan dalam berbagai disiplin ilmu, seperti ilmu kognitif, psikologi, dan bahkan media sosial (Rohman, Utami, & Raharjo, 2019).

Dengan adanya kemajuan teknologi, seseorang dapat dengan mudah mengekspresikan dirinya di sosial media, salah satunya beropini. Dalam bersosial media, kita tidak terlepas dengan namanya opini, opini terjadi karena adanya pesan dari komunikator kemudian terjadilah suatu diskusi ataupun reaksi terhadap isi pesan tersebut (Syarief, 2017). Dari opini tersebut biasanya orang-orang dapat mengeluarkan reaksi emosi dalam bentuk verbal maupun non-verbal.

Pada indentifikasi Emosi sendiri terbagi menjadi dua bentuk yaitu verbal dan non-verbal. Emosi verbal merupakan emosi dalam bentuk lisan ataupun tulisan, sedangkan non-verbal sendiri adalah emosi yang menggunakan Bahasa tubuh untuk mengekspresikannya, seperti raut wajah, tindakan, gerak-gerik tangan ataupun kaki, dan sebagainya. Mendeteksi emosi dalam dialog tekstual adalah masalah yang menantang karena tidak adanya ekspresi wajah dan modulasi suara (Chatterjee, Narahari, Joshi, & Agrawal, 2019). Cara mendeteksi emosi tersebut dengan melakukan klasifikasi emosi, klasifikasi sendiri adalah proses pengelompokan beberapa data sesuai kriteria.

Mendeteksi emosi memberikan peranan penting dalam berbagai aspek karena dapat diterapkan dalam berbagai bidang seperti pengambilan keputusan dalam lingkungan sosial maupun lingkungan bisnis (Fanesya, 2019). Pada era pandemi seperti ini, segala aktivitas yang biasanya tatap muka dialihkan semua ke daring, dari bidang politik, ekonomi, dan pendidikan. Salah satunya di dunia pendidikan, tenaga pengajar biasanya kesulitan untuk menyampaikan materinya dikarenakan mood murid atau siswa yang berbeda-beda, dan hasilnya akan buruk. Maka dari itu perlu adanya analisa emosi, analisis emosi dapat digunakan untuk mengetahui keadaan emosi pada siswa. Dari emosi tersebut tenaga pengajar dapat mengetahui mood siswanya, ini berguna untuk meningkatkan kualitas belajar mengajar tersebut.

Pada penelitian ini mengarah pada klasifikasi emosi pada teks opini yang digunakan untuk mengetahui jenis emosi dari opini yang telah diberikan. Opini tersebut diambil dari twitter. Penelitian ini menggunakan metode BERT (Bidirectional Encoder Representations from Transformers). BERT merupakan teknik berbasis jaringan saraf untuk pre-training natural language (Huang, Trabelsi, & Zaïane, 2019). Cara kerja BERT adalah mampu melatih model Bahasa berdasarkan seluruh rangkaian kata dalam

kalimat atau kueri. BERT memungkinkan model Bahasa untuk mempelajari konteks kata sesuai dengan yang ada disekitarnya.

Dalam penelitian-penelitian terdahulu, telah banyak sekali yang membahas tentang deteksi emosi. Sejauh ini, baru terdapat sedikit penelitian terkait dengan deteksi emosi dengan menggunakan algoritma BERT (Putri, 2020).

Metode Penelitian

Dalam penelitian ini, ada beberapa tahapan yang harus dilakukan untuk mendapatkan hasil yang bagus. Tahapan-tahapan tersebut terdiri dari pengumpulan data, pre-processing, labeling, modeling, dan evaluasi.

A. Pengumpulan Data

Data yang digunakan merupakan opini masyarakat di media sosial yang berupa teks, lalu data diambil dengan menggunakan teknik scraping. Setelah itu data dimasukkan kedalam file csv.

B. Pre-processing

Pada tahap ini data yang sudah dikumpulkan akan diproses untuk diambil informasi yang terkandung didalamnya (Harjanta, 2015). Pre-processing data dilakukan dengan cara menghapus data yang tidak perlu atau tidak sesuai agar membuat data lebih mudah untuk diproses, berikut ini adalah cara pre-processing yang dilakukan:

1. Menghapus symbol

Pada saat mengumpulkan dataset dengan melakukan scraping, terdapat banyak symbol-simbol tidak jelas dan tidak sesuai dengan konteks kalimat. Contoh simbolnya seperti (!, #, \$, <, &,\$, dsb).

2. Casefolding

Casefolding adalah tahapan menyamakan semua huruf pada data, dengan mengubah semua huruf menjadi huruf kecil

3. Remove slang

Remove slang adalah tahapan untuk menghapus kata-kata gaul yang kurang bermakna pada kalimat opini tersebut

4. Stemming

Stemming berfungsi untuk mengubah kata-kata ke bentuk akar (stem) (Wiryoadikusmo, Palit, & Andjarwirawan, 2018).

C. Labeling

Dataset yang sudah dibersihkan lalu dilakukan labeling satu persatu sesuai dengan label emosinya, ada Sembilan jenis emosi yang dipakai di penelitian ini.

data	label
Kurang ajar bgt sih lo!! Masuk rumah orang seenaknya	Marah
Seneng banget akhirnya lolos PTN	Bahagia
Gak tau lagi mesti gimana, semua pada ninggalin gw	Sedih
Wah produk terbaru apple keren banget jadi	Tertarik

pengen beli	
Mendingan kalian ikut seminar untuk mengisi hari libur kalian	Netral
Gw yakin dia gabakal ngelakuin itu	Percaya
Ihh jorok banget sih main di tempat sampah	Jijik
Wah ga nyangka bisa seramai ini	Kaget
Jadi merinding sendiri denger suara aneh diluar	Takut

D. Data pre perocessing

Data yang sudah diberisihkan dan diberi label akan diubah labelnya menjadi numerik, setelah diubah menjadi data numerik lalu diubah lagi menjadi data categorical. Data categorical tersebut akan dipisah menjadi data training dan data testing dengan perbandingan data training 80% dan 20% untuk data test.

E. Modeling

Setelah model dipre-processing lalu dimasukan kedalam model BERT sebagai layer input.

Layer (type)	Output shape	Param #
Input_ids (InputLayer)	[(None, 1000)]	0
BERT (TFBertMainLayer)	TFBaseModelOutputwithPoolingAndCrossAttentions(1 ast_hidden_state=(None, 100, 768), pooler_output=(None, 768), past_key_values=None, hidden_states=None, attentions=None, cross_attentions=None)	110617344
Pooled_output	(None, 768)	0
Emosi (Dense)	(None, 9)	6921

Total params : 110,624,265
 Trainable params : 110,624,265
 Non-trainable params : 0

F. Train Model

Setelah arsitektur model jadi, lalu model akan ditrain dengan hyperparameter sebagai berikut :

- Optimizer (adam)
- Loss (CategoricalCrossentropy)
- Validation_split = 0.2
- batch_size = 64
- epochs = 5

G. Evaluasi

Setelah model dilatih, yahap selanjutnya adalah evaluasi. Evaluasi disini berguna untuk mengetahui hasil dari model yang sudah ditrain apakah sudah baik atau belum dengan melihat hasil akurasi pada data test.

Hasil dan Pembahasan

A. Dataset

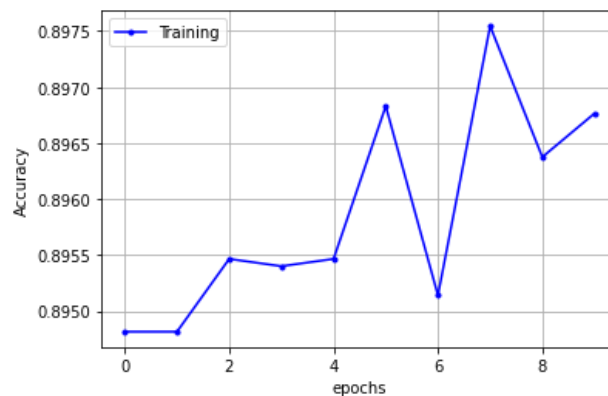
Setelah melakukan pengumpulan dataset, data telah terkumpul sebanyak 2700 data teksopini. Data ini diambil dari opini-opini orang di twitter secara acak, setelah dataset terkumpul lalu dipre-processing untuk membersihkan kata, simbol, dan tanda baca yang tidak perlu untuk mempermudah proses pelabelan data.

Label	Data
Marah	350
Bahagia	300
Sedih	300
Tertarik	250
Netral	350
Percaya	200
Jijik	300
Kaget	300
Takut	350

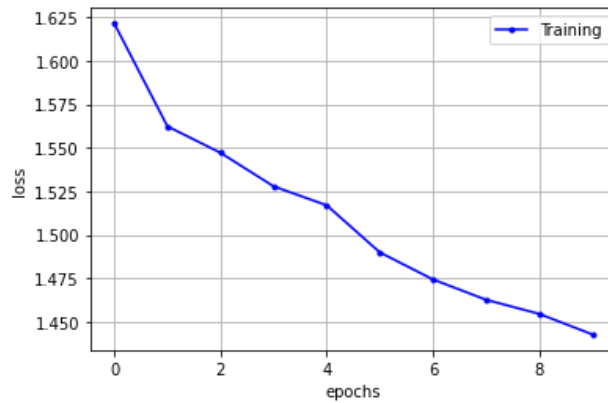
B. Hasil

Di bawah ini adalah hasil dari penelitian dan perbandingan menggunakan BERT Unceased dan indoBERT yakni sebagai berikut :

1. BERT Unceased



Gambar 1
BERT Unceased (accuracy)



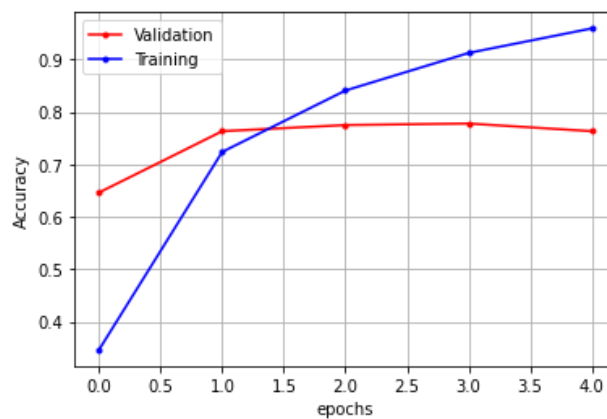
Gambar 2
BERT Unceased (Loss)

Pada gambar.4, saat model dilatih dengan menggunakan metode BERT Unceased menghasilkan akurasi sebesar 89.1% dan mendapatkan loss sebesar 1.6299 pada gambar.5

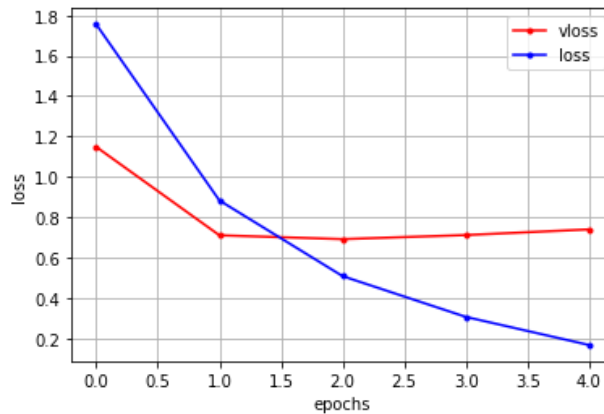
2. IndoBERT

IndoBERT sendiri adalah versi Indonesia dari model bert, dengan hasil sebagai berikut :

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
1	1.7553	0.3466	1.1481	0.6462
2	0.8826	0.7237	0.7100	0.7632
3	0.5076	0.8406	0.6910	0.7749
4	0.3057	0.9123	0.7111	0.7778
5	0.1664	0.9591	0.7383	0.7632



Gambar 3
IndoBERT (accuracy)



Gambar 4
IndoBERT (Loss)

IndoBERT menghasilkan akurasi sebesar 76% dan mendapatkan loss sebesar 0.1664.

Kesimpulan

Pada penelitian ini dilakukan klasifikasi emosi pada teks dengan banyak data sebesar 2700 data yang masing-masing data memiliki 9 label emosi (marah, sedih, takut, jijik, netral, senang, kaget, tertarik dan percaya). Dengan menggunakan BERT Unceasd dan IndoBERT yang mana hasilnya akan dibandingkan nilai akurasi. BERT Unceasd menghasilkan akurasi sebesar 89.1% sedangkan IndoBERT menghasilkan akurasi sebesar 76%. Perbandingan hasil yang dilakukan cukup jauh berbeda tetapi memiliki nilai akurasi yang cukup baik.

BIBLIOGRAFI

- Chatterjee, Ankush, Narahari, Kedhar Nath, Joshi, Meghana, & Agrawal, Puneet. (2019). Semeval-2019 task 3: Emocontext contextual emotion detection in text. *Proceedings of the 13th International Workshop on Semantic Evaluation*, 39–48. [Google Scholar](#)
- Fanesya, Fera. (2019). *Deteksi Emosi pada Twitter Menggunakan Metode Naive Bayes dan Kombinasi Fitur*. Universitas Brawijaya. [Google Scholar](#)
- Harjanta, Aris Tri Jaka. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Jurnal Informatika Upgris*, 1(1 Juni). [Google Scholar](#)
- Huang, Chenyang, Trabelsi, Amine, & Zaiiane, Osmar R. (2019). Ana at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and bert. *ArXiv Preprint ArXiv:1904.00132*. [Google Scholar](#)
- Putri, Cindy Alifia. (2020). Analisis Sentimen Review Film Berbahasa Inggris Dengan Pendekatan Bidirectional Encoder Representations from Transformers. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(2), 181–193. [Google Scholar](#)
- Rohman, Arif Nur, Utami, Ema, & Raharjo, Suwanto. (2019). Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing. *Jurnal Eksplora Informatika*, 9(1), 70–76. [Google Scholar](#)
- Syarief, Fauzi. (2017). Pemanfaatan Media Sosial Dalam Proses Pembentukan Opini Publik (Analisa Wacana Twitter Sby). *Jurnal Komunikasi*, 8(3). [Google Scholar](#)
- Wiryoadikusmo, Yohanes Adam Sastrodikoro, Palit, Henry Novianus, & Andjarwirawan, Justinus. (2018). Aplikasi Pendeteksi Unsur Hinaan dalam Komentar di Media Sosial Berbahasa Indonesia. *Jurnal Infra*, 6(1), 31–34. [Google Scholar](#)

Copyright holder:

Eko Prasetyo, Ivan Anindito Arista, Rudi Hermawan, Erlanda Pane (2021)

First publication right:

Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

