

Deteksi Konteks Intensi dan Emosi pada Tweet PPKM Bahasa Indonesia Menggunakan Bidirectional Encoder Representations from Transformers Secara Semi-Supervised Learning

Darren Ngoh
Fakultas Ilmu Komputer
Universitas Indonesia
Depok, Jawa Barat
darren@ristek.cs.ui.ac.id

Ardhani Dzaky Ralfiano
Fakultas Ilmu Komputer
Universitas Indonesia
Depok, Jawa Barat
ardhanidzaky@ristek.cs.ui.ac.id

Fikri Aufaa Zain
Fakultas Ilmu Komputer
Universitas Indonesia
Depok, Jawa Barat
fikri.aufaa@ui.ac.id

Abstract—Pandemi Covid-19 telah mengubah keseharian kehidupan kita terutama di Indonesia. Berbagai cara penanganan pandemic yang telah diterapkan pemerintah tentunya memiliki respon yang berbeda di tengah masyarakat. Twitter sebagai salah satu media sosial yang digunakan di Indonesia telah menjadi salah satu tempat dimana masyarakat Indonesia menyampaikan opini, keluhan, dan hal lainnya tak terkecuali respon dan tanggapan terkait penanganan pandemic khususnya di saat PPKM ini. Respon tersebut tentunya memiliki intensi dan sentiment terhadap konteks yang berkaitan erat dengan penanganan pandemic seperti kebijakan, vaksinasi, dan dinamika masyarakat selama pandemic. Fokus penelitian ini adalah untuk mengetahui dan memberikan gambaran secara umum tentang bagaimana respon masyarakat terhadap penanganan pandemi di Indonesia. Penelitian ini menggunakan metode *Traditional Machine Learning* (SGDClassifier) serta *Deep Learning* secara *Semi-Supervised Learning* dengan data yang didapatkan dari media sosial Twitter. Data tersebut merupakan *tweet* yang diambil dengan rentang waktu dua minggu terhitung mulai tanggal 8 Juli 2021 sampai dengan 22 Juli 2021 dengan jumlah *tweet* per hari yang diambil sejumlah 900 *tweets*. Berdasarkan hasil yang diperoleh akurasi klasifikasi menggunakan *Deep Learning* lebih tinggi dibandingkan dengan menggunakan *Traditional Machine Learning* kemudian *tweets* yang telah diklasifikasikan masih didominasi oleh *tweets* yang memiliki sentimen negatif dibandingkan *tweets* dengan sentimen positif.

Keyword: BERT, Cyclical Learning Rate, Semi-Supervised Learning, Transfer Learning

I. PENDAHULUAN

A. Latar Belakang

Penerapan PPKM merupakan salah satu bentuk dari upaya pemerintah untuk mengurangi laju kenaikan angka Covid-19 di Indonesia. Terjadinya lonjakan kasus Covid-19 di Indonesia setelah libur natal dan tahun baru 2021 membuat pemerintah menerapkan PPKM.

Penerapan PPKM yang berlangsung sampai saat ini bersifat dinamis dimana terjadi beberapa perubahan pada aturan maupun wilayah yang menerapkan PPKM sehingga memiliki efek terhadap kegiatan masyarakat sehari-hari. Hal tersebut tentunya memunculkan respon di tengah masyarakat yang memiliki intensi dan emosi yang berbeda terhadap konteks yang berkaitan dengan pelaksanaan PPKM.

Dalam masa PPKM ini Twitter menjadi salah satu platform media sosial dimana setiap pribadi bebas menuangkan aspirasi dalam bentuk *tweets*. *Tweets* tersebut tentunya mencakup opini, berita/info, keluhan, saran, dan hal lainnya yang berasal dari pengguna itu sendiri yang dapat

dianalisa. Dalam masa pandemi ini tentunya hasil analisis tersebut dapat menjadi salah satu gambaran keberhasilan pemerintah dalam pelaksanaan PPKM.

B. Tujuan dan Manfaat

Penelitian ini bertujuan untuk memberikan gambaran secara umum tentang respon masyarakat terhadap konteks yang memiliki kaitan erat dengan penanganan pandemic Covid-19 di Indonesia. Selain itu, penelitian ini memiliki manfaat untuk menjadikan respon masyarakat sebagai salah satu refleksi kebijakan yang telah dilaksanakan sehingga dapat menjadi pertimbangan dalam menentukan kebijakan selanjutnya.

C. Batasan Penelitian

Data yang akan digunakan dalam penelitian ini memiliki beberapa batasan sebagai berikut:

- *Tweet* yang digunakan berbahasa Indonesia.
- *Tweet* yang digunakan hanya *tweet* yang di-posting dalam periode 8 Juli 2021 sampai 22 Juli 2021.
- *Tweet* yang digunakan hanya *tweet* yang memiliki kata kunci seperti 'PPKM' atau 'vaksin' atau 'lockdown' atau 'covid' atau 'corona' atau 'pandemi'.
- Jika *tweet* memiliki gambar hanya teks yang akan diproses.

II. STUDI LITERATUR

Pada penelitian ini tim peneliti menggunakan beberapa metode dalam mengolah data mulai dari data diambil sampai dengan data menjadi sebuah input untuk dipelajari kedalam sebuah model.

A. BERT

BERT merupakan *machine learning framework* yang menggunakan *transformers*. *Transformers* digunakan untuk mempelajari hubungan antara dua kata pada suatu tulisan. Mekanisme yang ada di dalam *transformers* dibagi menjadi dua yaitu *encoder* untuk membaca *input* dan *decoder* sebagai hasil prediksi.

Encoder pada *transformers* membaca seluruh urutan kata sekaligus sehingga dianggap sebagai dua arah. Karakteristik ini yang membuat model mempelajari konteks sebuah kata berdasarkan semua lingkungannya [1].

Untuk Bahasa Indonesia sendiri terdapat model BERT yang telah di *pre-trained* sebelumnya menggunakan korpus Indo4B yaitu IndoBERT Base IndoBERT Lite Base. IndoBERT Lite Base mempunyai *vocab* sebesar 30.000 dan IndoBERT Base mempunyai *vocab* sebesar 30.522 [2].

B. Cyclical Learning Rate

CLR merupakan sebuah metodologi yang digunakan untuk melatih *neural network* dengan *learning rate* yang bervariasi selama proses *training*. Dalam metode CLR kita dapat mendefinisikan *learning rate minimum* dan *maximum* serta *step size* yang diperlukan untuk *learning rate minimum* mencapai *maximum* dan sebaliknya [3].

C. Pseudo-labeling

Pseudo-labeling merupakan salah satu pendekatan yang dapat dipilih dalam *semi-supervised learning*. Cara kerja teknik pseudo-labeling ini adalah dengan menggunakan model yang dibangun dari proses klasifikasi menggunakan *labeled data* untuk melakukan prediksi terhadap *unlabeled data*. Hasil prediksi dari *unlabeled data* ini yang dimaksud sebagai pseudo-label data yang mana label hasil prediksi dianggap sebagai label sebenarnya dari data. Kedua data tersebut kemudian dimanfaatkan untuk membuat model *classifier* baru.

D. Stochastic Gradient Descent Classifier

SGDClassifier merupakan salah satu tipe linear model yang disediakan pada library *scikit-learn*. *Estimator* ini sering digunakan dalam permasalahan *machine learning* yang berskala besar serta untuk keperluan *text classification* dalam NLP. SGDClassifier bekerja dengan cara mengimplemen sebuah garis *learning routine* penurunan gradien stokastik secara iteratif untuk mendapatkan penurunan gradien yang optimal.

III. HASIL

A. Dataset

Untuk pengumpulan data untuk pembuatan model *deep learning* ini, dilakukan *scraping tweets* pada platform Twitter dengan rentang waktu pengambilan data pada tanggal 8 sampai 22 Juli 2021 dengan 900 jumlah *tweets* per harinya. Pengambilan *tweets* memanfaatkan Twitter API dan modul Tweepy dengan kata kunci 'PPKM' atau 'vaksin' atau 'lockdown' atau 'covid' atau 'corona' atau 'pandemi' yang menggunakan Bahasa Indonesia dengan *filter retweets*.

Data yang telah diperoleh diproses terlebih dahulu sebelum dapat menjadi *input data* yang akan dimasukkan kedalam model. Beberapa pemrosesan data yang dilakukan sebelumnya adalah

- Pengubahan *username* yang terdapat dalam *tweet* menjadi <USERNAME>.
- Pengubahan URL yang terdapat dalam *tweet* menjadi <URL>.
- Pengubahan angka yang terdapat dalam *tweet* menjadi <NUM>.
- Menghapus *hashtag* yang terdapat dalam *tweet*.
- Menghapus *emoji* yang terdapat dalam *tweet*.

Tim peneliti juga menerapkan *pseudo-labeling* pada tahap ini untuk menghemat waktu serta menghindari *manual labour* yang tidak diperlukan. Tim peneliti hanya memberi label secara manual pada 5000 data lalu menerapkan teknik *pseudo-labeling* pada 6300 data lain untuk mendapatkan data

train yang cukup besar, di angka 11300. Terakhir, data dibagi menjadi *train set* dan *test set* dengan perbandingan 80 banding 20.

B. Hasil Pengujian

Untuk *training* model pada pengujian kali ini tim peneliti menggunakan *epoch* sebesar 5 untuk setiap modelnya serta menggunakan CLR untuk *learning rate* dengan *tuning learning rate minimum* sebesar $5e-7$ dan nilai *learning rate maximum* $5e-5$ dengan *step* sebesar 100. Sedangkan untuk *optimizer* kami menggunakan Adam dengan *loss function* *Sparse Categorical Cross Entropy*.

Berikut merupakan hasil pengujian dari ketiga model klasifikasi yaitu *context*, *intent*, dan *sentiment* menggunakan pendekatan *traditional machine learning* sebagai *baseline model* dan pendekatan *deep learning*. Untuk *traditional machine learning*, digunakan SGDClassifier karena model ini umum digunakan untuk masalah *text classification*.

TABLE I. PERFORMA MODEL

Performa Model			
Model	Context	Intent	Sentiment
SGDClassifier (Baseline)	0.63	0.45	0.66
IndoBERT Lite Base (phase-2)	0.73	0.35	0.64
IndoBERT Base (phase-2)	0.86	0.65	0.8

^a. Skor menggunakan F-1 score.

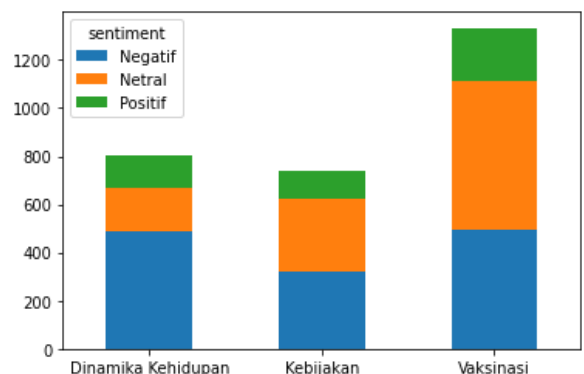
Fig. 1. Perbandingan performa setiap model.

C. Analisis

Pemodelan menggunakan *traditional machine learning* menghasilkan performa yang belum maksimal sehingga tim peneliti mencoba pemodelan selanjutnya menggunakan model IndoBERT. Performa yang didapatkan menggunakan IndoBERT Base (phase-2) lebih tinggi dibanding IndoBERT Lite Base (phase-2). *Context* memiliki selisih akurasi sebesar 0.13 lalu *intent* memiliki selisih akurasi sebesar 0.3 dan terakhir *sentiment* memiliki selisih akurasi sebesar 0.16.

Kemudian, dengan diterapkannya CLR tim peneliti dapat mendapatkan akurasi maksimal dengan *step* yang lebih rendah. Tetapi, dengan adanya *data imbalance* pada *dataset* menyebabkan beberapa *label* memiliki akurasi yang kurang maksimal.

Tim peneliti juga melakukan analisis pada hasil *sentiment* terhadap *context* pada *tweets*. Hal ini dapat memberikan gambaran kecil bagaimana respon masyarakat pada *context* yang berkaitan.



Selanjutnya tim peneliti mengambil data yang telah dikategorikan yang memiliki frekuensi lebih dari 100. Berdasarkan data tersebut didapatkan bahwa *tweet* yang telah dikategorikan masih didominasi oleh *tweet* sentimen negatif atau netral dan hanya satu kategori yang memiliki sentiment positif

Intent	Context	Sentiment	Count
Berita/Info	Vaksinasi	Netral	774
Opini	Dinamika Kehidupan	Negatif	562
Berita/Info	Vaksinasi	Negatif	518
Berita/Info	Kebijakan	Netral	448
Keluhan	Dinamika Kehidupan	Negatif	424
Keluhan	Kebijakan	Negatif	321
Berita/Info	Vaksinasi	Positif	244
Opini	Vaksinasi	Negatif	243
Berita/Info	Dinamika Kehidupan	Negatif	196
Keluhan	Vaksinasi	Negatif	160
Opini	Dinamika Kehidupan	Netral	119
Opini	Kebijakan	Negatif	115

Analisa serupa dapat diterapkan pada berbagai kebijakan untuk meninjau respon masyarakat luas. Dengan data yang minimal, hasil yang tim peneliti peroleh dapat dikatakan sudah dapat memetakan respon masyarakat secara tepat dan akurat sehingga apabila dilakukan hal serupa dengan data yang lebih besar tentunya dapat mendapatkan hasil yang lebih maksimal sehingga klasifikasi yang dihasilkan lebih tepat dan akurat.

IV. PENUTUP

A. Kesimpulan

Berdasarkan hasil pengujian, algoritma *deep learning* khususnya menggunakan *pre-trained* model IndoBERT Base P2 memberikan performa terbaik jika dibandingkan dengan IndoBERT Lite Base P2 dan algoritma *traditional machine learning*, yaitu SGDClassifier. IndoBERT Lite Base P2 memberikan hasil yang tidak jauh berbeda dengan SGDClassifier dengan selisih rata-rata akurasi sebesar 2% dan skor F1 7%, sedangkan IndoBERT Base P2 memberikan hasil akurasi yang cukup memuaskan untuk ketiga kategori label di angka akurasi 88% untuk label *context*, 80% untuk label *intent*, 85% untuk label *sentiment* dan skor F1 86% untuk label *context*, 65% untuk label *intent*, 80% untuk label *sentiment*.

Melalui eksperimen ini, tim peneliti menemukan bahwa respon masyarakat yang dituangkan pada tweets dapat

dikategorikan dan dapat menjadi masukan/*feedback* untuk pemerintah terhadap aksi yang diambil semasa PPKM. Penerapan *deep learning* model BERT terbukti dapat memberikan performa lebih baik untuk melakukan prediksi *context*, *intent*, dan sentimen dari suatu *tweet*. Selain itu, teknik *pseudo-labeling* sebagai pendekatan *semi-supervised learning* juga terbukti memadai untuk dilakukan karena tidak mempengaruhi performa dari model, sehingga teknik ini dapat digunakan untuk mengatasi masalah kurangnya data berlabel.

B. Saran

- Melakukan eksperimen pada teknik *deep learning* lainnya seperti LSTM dan RNN.
- Memperbanyak data *train* dan *validation* agar tidak terjadinya *data imbalance* yang dapat menyebabkan kurang maksimalnya akurasi model.
- Melakukan eksperimen dengan teknik *semi-supervised learning* lainnya seperti *bootstrapping*.
- Menggunakan model *pre-trained* lainnya seperti IndoBERT Large yang memiliki lebih banyak parameter.

REFERENCES

- [1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*. <http://arxiv.org/abs/1810.04805>
- [2] Mokrii, I., Boytsov, L., & Braslavski, P. (2021). A Systematic Evaluation of Transfer Learning and Pseudo-labeling with BERT-based Ranking Models. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2081–2085. <https://doi.org/10.1145/3404835.3463093>
- [3] Nalini, M. K., & Radhika, K. R. (2020). Comparative analysis of deep network models through transfer learning. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1007–1012. <https://doi.org/10.1109/I-SMAC49090.2020.9243469>
- [4] Smith, L. N. (2017). Cyclical Learning Rates for Training Neural Networks. *ArXiv:1506.01186 [Cs]*. <http://arxiv.org/abs/1506.01186>
- [5] Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: A transfer learning based approach. *Journal of Big Data*, 7(1), 1. <https://doi.org/10.1186/s40537-019-0278-0>
- [6] 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. *arXiv:2009.05387 [cs]*. <http://arxiv.org/abs/2009.05387>