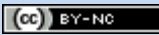



# Analisis Sentimen pandangan public terhadap profesi PNS (Pegawai Negeri Sipil) dari Twitter menerapkan Indonesian Roberta Base Sentiment Classifier

Alwi Jaya

<sup>a</sup> Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231, Indonesia  
alwiyein@gmail.com;

INFORMASI ARTIKEL	ABSTRAK
<p>Diterima : 17 – 01 – 2023 Direvisi : 19 – 02 – 2023 Diterbitkan : 31 – 03 – 2023</p> <p><i>Kata Kunci:</i> <i>BERT,</i> <i>Tweet,</i> <i>Pegawai Negeri Sipil,</i> <i>Sentiment.</i></p>	<p>Penelitian ini dilakukan untuk mendapatkan opini masyarakat terkait profesi Pegawai Negeri Sipil (PNS) yang beredar di media sosial twitter dengan menggunakan metode algoritma Bidirectional Encoder Representations from Transformers (BERT) yang di mana dalam Bahasa Indonesia berarti representasi dua arah encoder. BERT berguna untuk mengolah representasi dua arah yang berada dalam teks tanpa nama dengan menggabungkan sisi kanan dan kiri pada sebuah konteks dalam segala bagian yang di gunakan untuk menentukan positif, negatif atau netralnya pandangan masyarakat terkait profesi PNS yang beredar di media social twitter dengan meharapkan Hasil berupa pandangan masyarakat terhadap profesi PNS dengan mengumpulkan 394 tweet yang diambil dari media sosial twitter kita bisa mendapatkan hasil kesimpulan berupa berapa persen orang yang memandang positif profesi PNS dan berapa persen orang yang memandang negatif profesi PNS.</p> <p> </p>

## I. Pendahuluan

Selama bertahun-tahun, profesi Pegawai Negeri Sipil (PNS) adalah profesi yang sangat banyak diminati[1]. Sehingga tidak mengherankan bahwa banyak sekali opini yang tersebar di masyarakat tentang profesi PNS apalagi dengan mengingat bahwa pesatnya penggunaan internet di Indonesia dengan alasan utama adalah untuk komunikasi singkat sebesar 41,1%, mesin pencarian informasi terkait 11,5%, sosial media sebesar 38% dan lainnya[2]. Platform media sosial yang digunakan masyarakat Indonesia selain Facebook, Instagram, dan Youtube adalah Twitter[2].

Twitter menempati peringkat kelima sebagai media sosial yang sering diakses oleh masyarakat Indonesia[3] sehingga Twitter menjadi media sosial yang cukup populer untuk menanggapi opini yang beredar di masyarakat baik itu opini terkait politik, hobi, tren bahkan terkait profesi PNS [2]. Selain itu Twitter juga dapat digunakan untuk tujuan penelitian yang dapat menganalisis data menggunakan text mining, yang didapat dari tweet pengguna Twitter [4].

Dengan melihat bagaimana opini masyarakat Indonesia yang tersebar di media sosial Twitter dengan melihat kemunculan kata yang berkaitan dengan Pegawai Negeri Sipil atau Aparatur Sipil Negara cukup menonjol hal inilah yang mendasari topik analisis sentimen pandangan publik terhadap profesi Pegawai Negeri Sipil melalui media Twitter.

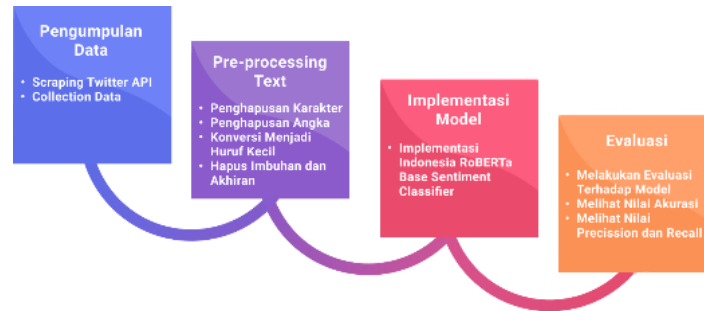
Analisis sentimen ini biasanya digunakan untuk memantau aktivitas media sosial karena memberikan gambaran yang lebih luas tentang opini publik. Di media sosial[2] terutama pada media sosial twitter yang dimana sangat populer sehingga statistik pengguna twitter memiliki 106 juta akun pada bulan april 2010 [5].

Dalam tulisan ini, penulis akan menggunakan metode Robustly Optimized BERT Pretraining Approach (RoBERTa). Model RoBERTa yang digunakan adalah RoBERTa Base Sentiment Classifier Indonesia, yang melabeli sentimen opini masyarakat Indonesia di media sosial Twitter. Data yang diberi label digunakan

sebagai data pelatihan dan pengujian untuk menerapkan Long Short Term Memory untuk melakukan pengujian dan analisis yang akurat. Pengolahan data dengan bahasa pemrograman Python [6].

## II. Metode

Setiap data scientist tentu saja membutuhkan sebuah metode untuk memecahkan masalah data[4]. Hal inilah yang menjadi acuan, rencana dan prosedur dalam melakukan penelitian ini untuk mencapai tujuan penelitian[7]. Oleh karena itu di sini penulis menggunakan sebuah algoritma BERT yang bisa di liat sebagai berikut:



### 2.1 Pengumpulan Data

Metode yang digunakan untuk mengumpulkan data tweet dari twitter menggunakan metode scraper[8]. Aktivitas scraping ini menggunakan library "twint"[9]. Twint adalah alat scraping Twitter yang ditulis dengan Python yang memungkinkan Anda untuk scraping tweet tanpa menggunakan API KEY[6]. Keuntungan dari alat Twint ini adalah Anda dapat mengambil hampir semua tweet, menggunakannya secara anonim tanpa mendaftarkan akun Twitter, dan tidak memiliki batasan harga[6]. Disini peneliti menggunakan kata kunci "Aparatur Sipil Negara" dan "Pegawai Negeri Sipil" dengan total banyak tweet yaitu 395 tweet.

### 2.2 Pre-Processing Text

Pada tahapan ini adalah tahapan di mana pengolahan data mentah di ubah menjadi data yang siap digunakan[10]. Tujuan dari preprocessing data adalah untuk menghilangkan data yang tidak perlu seperti noise, piktogram, URL, angka, dll. dan untuk menyatukan bentuk data kata. Ini membersihkan data kata sebelum diproses oleh proses selanjutnya[6].

Ketika data yang didapatkan masih banyak ditemukan kata yang tidak berstruktur maka dilakukanlah proses case folding untuk mengubah semua huruf menjadi huruf kecil atau lower case[11], setelah itu dilakukanlah proses selanjutnya yaitu Cleaning data seperti menghapus data yang tergadakan, menghilangkan symbol url, angka, string ASCII, emote, singkatan seperti 'yg', 'dg', 'tdk', dll. dan tanda baca seperti '©', 'TM'[6].

### 2.3 Implementasi Model

Data telah dibersihkan sebelumnya akan dilakukan text processing dan pelabelan dengan menerapkan model yang telah dilatih sebelumnya menggunakan Model Klasifikasi Indonesia RoBERTa Base Sentiment Classifier[6] sebanyak 5 epoch yang dapat pada Tabel 1 dibawah ini:

**Table 1.** Hasil Trained 5 Epoch

Epoch	Akurasi	F1	Presisi	Recall
1	92,85%	89,85%	90,98%	89,6%
2	93,41%	90,11%	92,52%	88,27%
3	94,28%	92,09%	92,75%	91,51%
4	94,36%	92,42%	92,64%	92,20%
5	94,20%	92,5%	92,40%	91,73%

Klasifikasi RoBERTa Base Sentiment Indonesia adalah model klasifikasi emosi teks berdasarkan model RoBERTa. Model ini awalnya merupakan model RoBERTa Base Indonesia yang telah dilatih sebelumnya, tetapi telah disempurnakan menggunakan dataset SmSA Indonesia, yang terdiri dari komentar dan ulasan yang

ditulis dalam bahasa Indonesia dengan akurasi mencapai 94,36% dan F1-makro 92,42% Seperti yang ditunjukkan pada [Tabel 2](#). Pada set uji benchmark, model mencapai akurasi 93,2% dan F1-makro 91,02% Seperti yang ditunjukkan pada [Tabel 3](#)[6].

**Table 2.** Hasil Setelah training

Akurasi	F1-Makro
94,36%	92,42%

**Tabel 3.** Hasil Pada Set Uji Benchmark

Akurasi	F1-Makro
93,2%	91,02%

## 2.4 Evaluasi Model

Evaluasi model akan dilakukan terhadap model yang telah di implementasi menggunakan Klasifikasi Indonesia RoBERTa Base Sentiment Classifier nantinya akan digunakan untuk mendefinisikan, mengilustrasikan dan menarik kesimpulan tentang kinerja metode klasifikasi Indonesia RoBERTa [12] yang akan berupa visualisasi grafik dan tabel untuk melihat nilai akurasi dan persentasi respon publik terhadap pekerjaan Pegawai Negeri Sipil berdasarkan jenis sentimen yang tersebar di media sosial twitter baik itu nilai negative, positif ataupun neutral.

## III. Hasil dan Pembahasan

### 3.1 Scraping

Hasil yang didapatkan dari menggunakan teknik scraping yang di mana kata kunci PNS dan Pegawai Negeri Sipil digunakan untuk mengumpulkan data raw dengan jumlah data kata kunci Aparatur Sipil Negara sebanyak 192 tweet dan kata kunci Pegawai Negeri Sipil sebanyak 203 tweet Sesudah proses dilakukan baru didapatkan 38 atribut yang terdiri dari id, conversation\_id, created\_at, dll, tetapi hanya beberapa atribut yang diambil yang di mana atribut yang diambil adalah atribut yang memiliki peran penting seperti tweet, created\_at, dan juga username[6] sehingga didapatlah jumlah tweets yang siap digunakan untuk pre-processing dan analisis yaitu sebanyak 395 data. Hasil Scraping Dengan Kata Kunci PNS dan Pegawai Negeri Sipil ditunjukkan pada [Tabel 4](#).

**Tabel 4.** Hasil Scraping Dengan Kata Kunci PNS dan Pegawai Negeri Sipil

Sebelum	Sesudah
@jeliwis Request lek @kuliundip @brokersmgg	request lek
Menyambut pemekaran setelah disahkannya UU Dae...	menyambut pemekaran disahkannya undang- undang daerah
...	...
Hal tersebut di sampaikan Sekretaris Jenderal...	sekretaris jenderal kemenkumham komisaris jenderal...
Jakarta – Calon Pegawai Negeri Sipil (CPNS) ke...	jakarta calon pegawai negeri sipil cpns...

### 3.2 Pre-Processing Text

Tahap pre-processing atau bisa dibilang adalah tahap persiapan yang di mana tahapan ini adalah tahapan yang sangat penting untuk penelitian menggunakan text mining [13]. Untuk mendapatkan hasil analisis yang baik tentu saja harus menyiapkan data mentah yang didapat dari proses scraping, data scraping inilah yang akan disiapkan terlebih dahulu sebelum di proses. misalnya salah satu tahapan dari pre-processing seperti mengubah semua ukuran teks menjadi huruf kecil (lowercase). Hal ini penting karena kata-kata yang terdiri dari huruf yang sama dibaca oleh sistem sebagai kata yang berbeda jika huruf tersebut memiliki ukuran font yang

berbeda. Sebagai contoh kata ‘Baik’ dan ‘baik’ [14]. Jika tidak diolah seperti itu maka akan berpengaruh terhadap waktu dalam memproses data di komputer, dan yang lebih parahnya data akan menjadi ambigu dan tidak jelas.

Dibawah ini ada contoh penerapan pre-processing dari data yang didapat melalui scraping ditunjukkan pada Tabel 5:

**Tabel 5.** Sebelum dan Sesudah Pre-processing

Sebelum	Sesudah
@jeliwis Request lek @kuliundip @brokersmngg	request lek
Menyambut pemekaran setelah disahkannya UU Dae...	menyambut pemekaran disahkannya undang- undang daerah
...	...
Hal tersebut di sampaikan Sekretaris Jenderal...	sekretaris jenderal kemenkumham komisaris jenderal...
Jakarta – Calon Pegawai Negeri Sipil (CPNS) ke...	jakarta calon pegawai negeri sipil cpns...

Dari perbandingan di atas, kita dapat melihat bahwa data di bagian sebelum dan sesudah miliki perbedaan yang cukup signifikan yang di mana terdapat penghapusan tanda baca, simbol, teks ganda emot, dan juga penghapusan stopword. Semua itu dilakukan dengan menggunakan python dengan memanfaatkan library yang tersedia di python seperti Sastrawi, regex, NLTK dsb.

### 3.3 Indonesian RoBERTa Base Sentiment Classifier

Data yang sudah di bersihkan selanjutnya akan di implementasikan ke dalam model pretrained Indonesia RoBERTa Base Sentiment Classifier yang dimana digunakan untuk menghasilkan label dan akurasi dari prediksi label. Hasil implementasi dari implementasi Indonesia RoBERTa Base Sentiment Classifier ditunjukkan di bawah pada Gambar 1 dan Tabel 6. Proses prediksi ini memakan waktu kurang lebih 8 menit untuk melakukan klasifikasi sentimen terhadap 1093 tweet yang telah di bersihkan sebelumnya.



**Gambar 1.** Implementasi Indonesia RoBERTa Base Sentiment Classifier

**Tabel 6.** Hasil Pelabelan Dari Indonesia RoBERTa Base Sentiment Classifier

Tweet_Remove_Punct	Label	Akurasi
request lek	neutral	0.985138
menyambut pemekaran disahkannya undang-undang...	neutral	0.998972
...	...	...
sobat baraya cpns bapas palangka raya dapatkan...	neutral	0.998939
rabu Thomas tembungge spi luhkan kabupaten me...	neutral	0.998921





- BUDIDARMA*, vol. 6, no. 1, p. 535, 2022, doi: 10.30865/mib.v6i1.3527.
- [8] R. Kurniawan and A. Apriliani, "Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari twitter berbasis web scraper," *Analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari twitter berbasis web scraper*, vol. 5, no. 1, 2020.
- [9] M. Nurmalasari, N. Anggita Temesvari, S. Ni, and matul Maula, "Analisis Sentimen terhadap Opini Masyarakat dalam Penggunaan Mobile-JKN untuk Pelayanan BPJS Kesehatan Tahun 2019," *Heal. Inf. Manag. J. ISSN*, vol. 8, no. 1, pp. 2655–9129, 2020.
- [10] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [11] K. K. Agustiningsih, E. Utami, O. Muhammad, and A. Alsyabani, "Sentiment Analysis of COVID-19 Vaccines in Indonesia on Twitter Using Pre-Trained and Self-Training Word Embeddings," *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Information)*, vol. 15, no. 1, pp. 39–46, 2022, doi: 10.21609/jiki.v15i1.1044.
- [12] A. E. Putra and W. Maharani, "Depression Levels Detection Through Twitter Tweets Using RoBERTa Method," vol. 3, no. 4, pp. 453–459, 2022, doi: 10.47065/josh.v3i4.1872.
- [13] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [14] D. Fimoza, "Analisis Sentimen Terhadap Film Indoensia Dengan Pendekatan BERT," *Expert Syst. Appl.*, vol. 118, pp. 272–299, 2019, [Online]. Available: <http://repositori.usu.ac.id/handle/123456789/30445%0Ahttps://linkinghub.elsevier.com/retrieve/pii/S0957417418306456>
- [15] G. Gumelar, Q. Ain, R. Marsuciati, S. A. Bambang, and A. Sunyoto, "Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance," *Sisfotek*, pp. 250–255, 2021.