

ANALYSIS SENTIMENT ON THE ACCEPTANCE OF CPNS 2021 ON TWITTER SOCIAL MEDIA USING TEXTBLOB

Widi Astuti¹; Elly Firasari²; F. Lia Dwi Cahyanti³; Fajar Sarasati⁴;

^{1,4} Bisnis Digital; ^{2,3} Sistem Informasi;
^{1,2,3,4} Universitas Nusa Mandiri, Jakarta, Indonesia
^{1,2,3,4} www.nusamandiri.ac.id

widiastuti.wtu@nusamandiri.ac.id¹; elly.efa@nusamandiri.ac.id²; flia.fdc@nusamandiri.ac.id³;
fajar.fss@nusamandiri.ac.id⁴;



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract— Information technology is developing rapidly with the development of hardware and hardware developed by the world's largest companies. These advances have a significant impact on human life. Many jobs in human life use the help of technology. data mining technology, one of which is used in the field of research. Data mining extracts valuable information by analyzing the presence of certain patterns or relationships from large amounts of data. Indonesian government agencies routinely organize the recruitment and selection of Candidates for Civil Servants (CPNS). Almost every year the government opens CPNS formations and it is never empty of applicants. Prospective Civil Servants (abbreviated as CPNS) are employees who have just passed the first stage of the selection test for Candidates for Civil Servants. Minister for Administrative Reform and Bureaucratic Reform (PANRB) Tjahjo Kumolo said the government would eliminate the recruitment of prospective civil servants (CPNS) in 2020. In order to find out the enthusiasm of the people in Indonesia, this study took data from reviews on the Twitter application with the topic of CPNS Acceptance. 2021. Using Textblob sentiment analysis, the sentiment class used was positive, negative and neutral sentiment class, and the data mining method used was Logistic Regression.

Keywords : Data mining, CPNS acceptance, Textblob, SVM, Logistic Regression.

Intisari— Teknologi informasi berkembang pesat dengan pengembangan perangkat keras dan perangkat lunak yang dikembangkan oleh perusahaan terbesar di dunia. Kemajuan tersebut memberikan dampak yang signifikan terhadap kehidupan manusia. Banyak pekerjaan dalam kehidupan manusia menggunakan bantuan teknologi. teknologi data mining salah satunya yang digunakan dalam bidang penelitian. Data mining mengekstrak informasi berharga dengan menganalisis keberadaan pola atau hubungan tertentu dari sejumlah besar data. Instansi pemerintah Indonesia secara rutin menyelenggarakan rekrutmen dan seleksi Calon Pegawai Negeri Sipil (CPNS). Hampir tiap tahun pemerintah membuka formasi CPNS dan tidak pernah sepi pelamar. Calon Pegawai Negeri Sipil (disingkat CPNS) adalah pegawai yang baru lulus tes seleksi penerimaan Calon Pegawai Negeri Sipil tahap pertama. Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (PANRB) Tjahjo Kumolo mengatakan pemerintah akan meniadakan rekrutmen calon pegawai negeri sipil (CPNS) tahun 2020. Dalam rangka ingin mengetahui antusiasme masyarakat di indonesia penelitian ini mengambil data dari ulasan-ulasan di

aplikasi Twitter dengan topik Penerimaan CPNS 2021. Menggunakan sentimen analisis Textblob kelas sentimen yang digunakan kelas sentimen positif, negatif dan netral, dan metode data mining yang digunakan adalah Logistic Regression.

Kata Kunci: Klasifikasi, Penerimaan CPNS, Textblob, Logistic Regression.

PENDAHULUAN

Instansi pemerintah di Indonesia secara berkala menyelenggarakan rekrutmen dan seleksi Calon Pegawai Negeri Sipil (CPNS). Hampir tiap tahun pemerintah membuka formasi CPNS dan tidak pernah sepi pelamar. Calon Pegawai Negeri Sipil (disingkat CPNS) adalah pegawai yang baru lulus seleksi penerimaan calon pegawai negeri sipil tahap pertama. Calon PNS tidak memenuhi kewajiban 100% gaji PNS. Anda akan menerima persentase 80% berdasarkan CPNSSK yang ditentukan berdasarkan hukum yang berlaku di Indonesia. Namun, pada saat seleksi akreditasi CPNS 2018, hanya 128.000 dari 1.700.000 pendaftar yang lolos dinilai gagal, sehingga tidak memenuhi jumlah individu yang layak menduduki jabatan pemerintahan. Kami masih memiliki beberapa tes seleksi kompetisi lapangan (SKB), sehingga kami telah sepenuhnya lulus seleksi CPNS oleh masing-masing institusi. (Nurafni Kusumawardhani, 2019). Selain itu, proses seleksi CPNS di Indonesia juga dinilai sangat buruk, dengan pendaftaran yang kompleks dikombinasikan dengan pilihan tradisional daripada berdasarkan keahlian dan kompetensi secara keseluruhan, sehingga mengakibatkan korupsi, kolusi dan nepotisme. (Sulaiman, 2021).

Indonesia merupakan pengguna aktif media sosial, berdasarkan We are social dalam tahun 2020, Indonesia mencapai nomor 175,4 Juta orang mengakses internet. Dan Twitter adalah salah satu platform media yg paling banyak dipakai sang warga Indonesia. Menurut asal

berdasarkan We are Social & Hootsuite tahun 2020, twitter menempati urutan ke lima pada kategori yang paling sering digunakan dengan jumlah presentasi 56% setelah Youtube, Whatsapp, Facebook dan Instagram (We Are Social dan Hootsuite, 2020). Berdasarkan Twitter merupakan Sebuah situs web yang menyediakan layanan microblogging online yang memungkinkan pengguna untuk berbagi konten yang saat ini dibatasi hingga 280 karakter. (Informatika, 2018). Fitur tweet yang sering digunakan untuk menuliskan pemikiran, serta opini oleh para pengguna platform yang dapat digunakan penulis dalam pengolahan informasi yang sangat berguna. Adanya perbedaan pendapat tersebut sangat penting dan juga merupakan salah satu pengaruh terpenting pada perilaku manusia untuk mendapatkan hasil dari emosi.

Dengan latar belakang tersebut, penulis membahas tentang analisis opini publik tentang penerimaan CPNS di Twitter di media sosial. Hal ini sangat menarik untuk ditelaah dan menemukan opini publik tentang antusiasme masyarakat di tengah pandemi yang sedang berlangsung, apakah reaksinya positif, negatif atau netral.

BAHAN DAN MODEL

A. Teknik Pengumpulan Data

Ada beberapa cara untuk mengumpulkan data:

1. Crawling data

Crawling data adalah proses mengambil atau mengunduh data dari server Twitter dalam format data pengguna dan tweet menggunakan antarmuka pemrograman aplikasi (API) Twitter. (Brata Mas Pintoko, 2018). Objek pada penelitian ini yaitu Penerimaan CPNS 2021 yg sedang berlangsung pada Indonesia. Total lebih menurut 3500 data.

2. Preprocessing

Pada tahap ini dilakukan tindak lanjut dari data yang di tarik dengan merapikan data, sortir data, membersihkan data dari duplikat sampai data siap di gunakan untuk proeses seanjutnya. Data yang dibersihkan adalah data yang Masih banyak simbol dan kata-kata yang tidak perlu (Brata Mas Pintoko, 2018).

3. Sentimen Analysis menggunakan Textblob

Sentiment Analisis adalah metode untuk memahami dan mengekstraksi data opini secara otomatis dan memproses data tekstual untuk menangkap emosi yang terkandung dalam opini tersebut. Pengelompokan data analisis sentimen ini memiliki beberapa pendapat dan jawaban, termasuk positif, negatif, dan netral. (Nova Tri Romadloni, Imam Santoso, 2019). Analisis sentimen juga dapat disebut sebagai suatu proses untuk menemukan suatu makna dari perilaku, opini, pandangan, dan emosi dari suatu teks, perkataan, *tweets*, dan *database* dengan sumber dari *Natural Language Processing* (NLP) (Hernikawati, 2021).

Sedangkan Textblob adalah salah satu library Python2 dan Python3 yang bisa Anda gunakan untuk mengolah data teks. Textblob mudah diakses dan dapat digunakan untuk pembuatan prototipe cepat (Hernikawati, 2021).

4. Modelling

Pemodelan dilakukan dengan tujuan untuk menguji keakuratan prediksi sistem berdasarkan data model yang dibuat. (Andini, 2021). klasifikasi dengan menggunakan algoritma Logistic Regression. Hasil dari proses klasifikasi kemudian dibandingkan untuk mengetahui algoritma mana yang terbaik akurasi.

Saat mengembangkan metode penelitian eksperimental

menggunakan CRISPDM, metode penelitian standar yang digunakan dalam penambangan data, terdiri dari enam tahap: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, dan evaluasi. , Dan ada langkah-langkah penyebaran. (Binsar, 2020). Model yang digunakan dalam penelitian ini antara lain:

1. Business Understanding

Pada tahap ini dilakukan pemahaman terhadap objek penelitian yang dilakukan dengan menemukan informasi melalui media social twitter tentang Penerimaan CPNS 2021 yang sedang berlangsung di Indonesia mengekspresikan berbagai macam pendapat baik negatif maupun positif pada tweet pengguna media sosial.

2. Data Understanding

Pada tahap data understanding adalah proses memahami data yang digunakan sebagai bahan yang diselidiki dan memungkinkan Anda untuk melanjutkan ke langkah berikutnya, pra-pemrosesan. Pada fase ini proses pengambilan data mentah berjalan sesuai dengan atribut yang dibutuhkan. Data yang digunakan dalam survei ini berasal dari kolom kicauan di media sosial Twitter. Di bawah ini adalah Tabel 1 dari data yang diambil selama fase ini.

Tabel 1. Proses Data Understanding

Date	Tweet
04 Juli 2021	'Seleksi CASN Kemenparekraf/ Baparekraf mulai masa pendaftaran 2-21 Juli 2021 #CPNS2021 https://t.co/KF2uRLbamu ',D isporapar Jateng #JatengGayeng
05 Juli 2021	'Buruan! Pendaftaran CASN Pesisir Selatan Tahun 2021 Telah Dibuka https://t.co/ajZp0f719c ',Hari an Haluan

Date	Tweet
06 Juli 2021	'LINK Pendaftaran CASN Kemenparekraf Tahun 2021 Dengan 186 Jumlah Kebutuhan Formasi Dari Berbagai Lulusan https://t.co/xQnT22pWqq ', Portal Kudus - Pikiran Rakyat
07 Juli 2021	'Pendaftaran Seleksi CASN Kemenag Dibuka Hingga 21 Juli 2021 https://t.co/prEsA90iBs ', rad arbangsa.com

Sumber: (Astuti, 2022)

3. Data Preparation

Tahap data preparation merupakan tahap dengan proses penyiapan data digunakan dalam data dibersihkan dan Siap dapatkan digunakan dalam penelitian. Dalam text mining tahapan awal yang akan dilakukan adalah tahap text preprocessing. Berikut merupakan langkah yang dilakukan dalam text preprocessing:

4. Modelling

Dalam sesi Modelling ini hendak dicoba metode pengklasifikasian informasi yang sangat akurat. Buat menyamakan ataupun mengkomparasi, pada riset ini hendak digunakan algoritma SVM serta Logistic Regression.

5. Evaluation

Model yang terbentuk akan diuji menggunakan confusion matrix yang akan mengetahui tingkat akurasi. Confusion matrix akan menggambarkan Hasil akurasi berkisar dari prediksi positif positif, prediksi positif palsu, prediksi negatif positif, dan prediksi negatif palsu. Akurasi dihitung dari semua prediksi yang benar (baik prediksi positif maupun negatif). Dibandingkan dengan seluruh data testing. Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Pengujian juga diukur dengan menggunakan ROC Curve. ROC Curve akan menggambarkan kelas

positif dalam bentuk kurva. Pengujian dilakukan dengan menghitung nilai area under the curve (AUC), dan semakin tinggi nilai kurva AUC dan ROC maka akan semakin baik model klasifikasi yang terbentuk.

6. Deployment

Sesi ini merupakan sesi terakhir dari CRISP DM dan merupakan hasil dari semua tahapan yang sebenarnya digunakan sebelumnya. Maknanya merupakan melaksanakan suatu bersumber pada pengetahuan yang didapatkan dari aktivitas mining terhadap informasi. Pelaksanaan dalam riset ini hendak dibesarkan dengan PHP serta MySQL.

HASIL DAN PEMBAHASAN

- a. Proses @#Annotation removal adalah proses untuk menghilangkan teks yang memiliki anotasi @ dan #. Proses preprocessing ini dilakukan dengan menggunakan python yaitu menggunakan regular expression perintah `re.sub("([@#][A-Za-z0-9]+)|(\w+:\w+\/\S+)", "", text)`. Sehingga tweet dengan mention user dan hashtag hilang. Tabel 2 menunjukkan perbandingan teks sebelum dan sesudah dilakukannya proses @#Annotation Removal.

Tabel 2. Perbandingan teks sebelum dan sesudah proses @#Annotation Removal

@#Annotation Removal	
Data sebelum	Data sesudah
#'@himynameisokk y Pertanyaan seputar Pendaftaran CASN 2021 agar disampaikan melalui WA Helpdesk 087884112321', BKP SDM_PURWAKARTA	pertanyaan seputar pendaftaran casn agar disampaikan melalui wa helpdesk, BKPSD M_PURWAKARTA A,

Sumber: (Astuti, 2022)

Tokenization: Regexp

Proses Tokenization merupakan data tweet pada setiap kata akan dipisahkan berdasarkan berdasarkan spasi yang ditemukan (Brata Mas Pintoko, 2018). Tokenization peneliti gunakan untuk memisahkan kata atau huruf dari tanda baca dan simbol (Eka Rini Yulia, 2021). Regexp adalah proses untuk menghilangkan tanda baca dan angka sehingga hasilnya adalah kata. Proses preprocessing ini dilakukan dengan menggunakan library Regexp di python. Tabel 3 menunjukkan perbandingan teks sebelum dan sesudah dilakukannya proses Tokenization: Regexp.

Tabel 3. Hasil Proses Tokenization

Tokenization: Regexp	
Data sebelum	Data sesudah
'@Kota_Tangerang Hallo min, mau tanya. Untuk menanyakan terkait pendaftaran CASN kemana ya? Adakah nomor yang bisa dihubungi. Terima kasih"	tangerang hallo min mau tanya untuk menanyakan terkait pendaftaran casn kemana ya adakah nomor yang bisa dihubungi terima kasih

Sumber: (Astuti, 2022)

c. Indonesian Stemming

Proses Indonesian Stemming adalah Konversi data tweet dari preposisi ke kata dasar (Brata Mas Pintoko, 2018). Proses preprocessing ini dilakukan dengan menggunakan library python Sastrawi. Tabel 4 menunjukkan perbandingan teks sebelum dan sesudah proses stemming bahasa Indonesia..

Tabel 4. Proses Indonesian Stemming

Indonesian Stemming	
Data sebelum	Data sesudah
Buat SobatBKN, yang punya	buat sobatbkn, yang punya banyak tanya putar daftar

banyak pertanyaan seputar pendaftaran CASN 2021 pada portal SSCASN, akan ada sesi tanya jawab live pada akun Instagram BKN BKNgoidofficial, di waktu yang tertera berikut Jangan sampai kelewatan ya	casn 2021 pada portal sscasn, akan ada sesi tanya jawab live pada akun instagram bkn bkngoidofficial, di waktu yang tera ikut jangan sampai lewat ya
---	---

Sumber: (Astuti, 2022)

d. Indonesian Stop Word Removal

Proses Indonesian Stop word removal adalah Proses menghilangkan kata-kata umum yang biasanya terlihat dalam jumlah banyak dan dianggap tidak berarti. Beberapa contohnya adalah dg, rt, with, ny, d, klo, if, amp, let, make, say, no, ga, krn, nya, ya, si, know, no, uh, for, yes, And soon. Proses prapemrosesan ini dilakukan menggunakan pustaka stopword Python NLTK. Tabel 5 menunjukkan perbandingan teks sebelum dan sesudah prosedur penghapusan stopword bahasa Indonesia. Tabel 5. Hasil Stop Word Removal

Indonesian Stop Word Removal	
Data sebelum	Data sesudah
BKNgoid halo selamat siang admin BKN saya mau nanya untuk pendaftaran akun seleksi CASN untuk tempat lahir	@bkngoid halo selamat siang admin bkn nanya pendaftaran akun seleksi casn lahir ktp dgn ijazah berbeda. ktp

saya antara ktp dgn ijazah berbeda. Ktp depok ijazah bogor. Kira-kira bagaimana ya solusinya? Terima kasih, Tyas Murti Lestari	depok ijazah bogor. solusinya? terima kasih, tyas murti lestari
---	--

Sumber : (Astuti, 2022)s

e. N Chars Filter

Proses N Chars Filter yaitu untuk menghapus kata yang kurang dari satu suku kata. Tabel 6 Menampilkan perbandingan teks sebelum dan sesudah proses penyaringan N-karakter.

Tabel 6. Hasil Nchars Filter

N Chars Filter	
Data sebelum	Data sesudah
b'Segera Daftar! Pemkab Pessel Buka Pendaftaran CASN Tahun 2021 https://t.co/ZLsD8Q4oDr r',BERITAMINANG	segera daftar pemkab pessel buka pendaftaran casn tahun,BERIT AMINANG

Sumber: (Astuti, 2022)

F. Stop Word Removal

Pada tahap ini, proses penghilangan stopword adalah untuk mengecek apakah stopword tersebut mengandung sebuah kata. Proses ini menghilangkan istilah-istilah yang tidak berarti dan tidak relevan. Istilah yang diambil dari tingkat tokenisasi diperiksa dalam daftar stopword. Jika sebuah kata dimasukkan dalam daftar stopword, kata itu tidak akan diproses.

G. Textblob Analysis Sentiment

Pada proses ini yaitu menghitung sentiment dari setiap tweet untuk memproses data tekstual. Pada tahap ini menggunakan teknik Natural Language Processing (NLP) dan data tweet

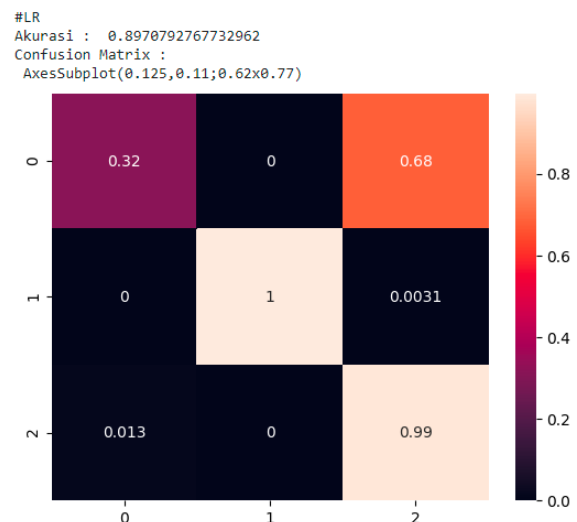
umumnya tergolong kedalam unsupervised Learning. Diperlukan NLP untuk mengidentifikasi opini dan sentimen dan mengklasifikasikannya menjadi label positif, negatif, atau netral. Library python yang digunakan untuk identifikasi analisis sentimen pada penelitian ini adalah Textblob. Dan dari masing-masing tweet tersebut ditetapkan polaritasnya apakah bermakna positif, negatif, atau netral.

Tabel 7. Klasifikasi Identifikasi Analisis Sentimen *Textblob*

Sentiment	Count	Percentage
Positive	1.527	42,5%
Neutral	1.597	44,4%
Negative	471	13,1%
Total	3.595	100%

Sumber: (Astuti, 2022)

Gambar 1. Hasil Klasifikasi Identifikasi Analisis Sentimen *Textblob* dibawah ini.



Gambar 1. Hasil Klasifikasi Logistic regression

Sumber: (Astuti, 2022)

Akurasi yang dicapai adalah 89,70% data analisis sentimen positif, negatif, dan netral terkait Asumsi CPNS 2021. 32% data tweet negatif sesuai dengan ekspektasi. Data tweet negatif yang termasuk dalam prediksi positif adalah 68%, dan data tweet negatif yang diprediksi netral adalah 0%. Tweet

negatif berisi 0,01% data tweet positif, 99n data tweet positif sesuai ekspektasi, 100% data tweet netral sesuai ekspektasi, dan 0,0031% tweet netral yang masuk positif.

Kesimpulan

Hasil penelitian ini telah ditemukan kesimpulan berdasarkan data yang telah di dapatkan dari media sosial twitter mengenai Penerimaan CPNS 2021 pada masyarakat indonesia adalah sebagai berikut 42,5% masyarakat bersentimen Positif, dan 44,4% bersentimen Netral dan 13,1% bersentimen Negatif. Kesimpulan ini di buktikan dengan keakuratan pengujian metode data mining Logistic Regression dengan akurasi sebesar 89,70%. Dari hasil tersebut dapat diartikan bahwa masyarakat indonesia cenderung bersentimen positif terhadap berita penerimaan CPNS.

REFERENSI

- Andini, G. R. dan R. (2021). Analisis Respon Masyarakat Pada Platform Media Sosial Twitter Terhadap Tokoh Politik, Jenderal TNI (PURN.) Gatot Nurmantyo. *Jurnal Ilmiah Akuntansi Dan Keuangan*, 4(Vol. 4 No. 2 (2021): FairValue: Jurnal Ilmiah Akuntansi dan Keuangan), 12. <https://doi.org/https://doi.org/10.32670/fairvalue.v4i2.650>
- Astuti, W. et all. (2022). ANALYSIS SENTIMENT ON THE ACCEPTANCE OF CPNS 2021 ON TWITTER SOCIAL MEDIA USING TEXTBLOB, 17(1), 1–6.
- Binsar, T. M. dan F. (2020). Cross-Industry Standard Process for Data Mining (CRISP-DM).
- Brata Mas Pintoko, K. M. L. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. e-proceeding of Engineering.
- Eka Rini Yulia, K. S. (2021). Implementasi Particle Swarm Optimization (PSO) pada Analysis Sentiment Review Aplikasi Trafi Menggunakan Naive Bayes (NB). *Jurnal Teknik Komputer AMIK BSI*.
- Hernikawati, D. (2021). Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis. *Jurnal Ilmu Pengetahuan Dan Teknologi Komunikasi*, 23(1), 21–31.
- Informatika, K. K. dan. (2018). *Memaksimalkan Penggunaan Media Sosial dalam Lembaga Pemerintah*. Jakarta: Dorektorat Jenderal Informasi dan Komunikasi Publik, Kementerian Komunikasi dan Informatika.
- Nova Tri Romadloni, Imam Santoso, S. B. (2019). Perbandingan Metode Naive Bayes , KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *Jurnal IKRA-ITH Informatika*.
- Nurafni Kusumawardhani, A. dan R. L. (2019). Problematika Seleksi CPNS 2018 dalam Pengangkatan CPNS yang tidak Memenuhi Passing Grade. *Civil Service*.
- Sulaiman, dan D. R. (2021). Opini Peserta Terhadap Transparansi Penerimaan CPNS Melalui Metode Computer Assisted Test (CAT). *Jurnal Eksos*.
- We Are Social dan Hootsuite. (2020). *Indonesian Digital Report 2020*.