

---

**Efektivitas Variabel Demografi Pengguna Twitter dalam Prediksi Pilpres Indonesia 2014 dan 2019****Muhammad Misbah Effendy<sup>1</sup>, Taufik Edy Sutanto<sup>2</sup>, Muhaza Liebenlito<sup>3</sup>**muhammad.misbah20@mhs.uinjkt.ac.id<sup>1</sup><sup>1,2,3</sup>UIN Syarif Hidayatullah Jakarta

---

**Informasi Artikel**

Diterima : 27 Okt 2023

Direview : 15 Nov 2023

Disetujui : 30 Des 2023

---

**Kata Kunci**Analisis demografi,  
BERT, Pilpres Indonesia,  
Random Forest, Twitter

---

**Abstrak**

Twitter menjadi platform media sosial yang digunakan untuk menyuarakan dukungan terhadap calon pada Pemilihan Presiden (Pilpres) Indonesia 2014 dan 2019. Meskipun banyak analisis telah dilakukan pada Twitter, penelitian mengenai demografi pengguna masih jarang dilakukan. Masalah yang dihadapi dalam analisis demografi adalah keterbatasan informasi di Twitter. Selain analisis sentimen, penelitian ini menggunakan informasi demografi yaitu lokasi dan jenis kelamin yang diprediksi melalui nama pengguna Twitter. Menggunakan model BERT untuk analisis sentimen dan Random Forest untuk prediksi, data Pilpres 2014 diolah untuk memprediksi Pilpres 2019, yang kemudian dibandingkan dengan data resmi Komisi Pemilihan Umum (KPU). Dengan menggunakan variabel demografi, model memprediksi lebih baik 13.34% dibandingkan dengan model tanpa variabel demografi, dari 63.33% menjadi 76.67%. Berdasarkan hasil analisis variabel indikator, jumlah penduduk menjadi variabel demografi dengan nilai *importance* tertinggi. Hasil penelitian ini dapat digunakan sebagai acuan untuk meningkatkan hasil prediksi permasalahan lain yang serupa dan belum menggunakan informasi demografi.

---

**Keywords***BERT, Demographic Analysis, Indonesian Presidential Election, Random Forest, Twitter*

---

**Abstrak**

*Twitter has emerged as a social media platform used to express support for candidates in the Pilpres of 2014 and 2019. Although extensive analyses on Twitter have been conducted, studies focused on user demographics remain sparse. A primary challenge in demographic analysis is the limited information available on Twitter. Beyond sentiment analysis, this study leverages demographic information, specifically location and gender, inferred from Twitter usernames. Employing the BERT model for sentiment analysis and Random Forest for prediction, the 2014 Pilpres data was processed to forecast the 2019 election outcomes, which were subsequently compared to data from the KPU. By incorporating demographic variables, the model enhanced its prediction accuracy by 13.34%, moving from 63.33% to 76.67%. Based on the variable analysis, population size emerged as the most influential demographic variable. The findings of this study can serve as a benchmark to improve predictions for similar issues not yet utilizing demographic information.*

## A. Pendahuluan

Media sosial, khususnya Twitter, telah berkembang pesat sebagai platform komunikasi politik di Indonesia. Gambar 1 dan 2 menunjukkan peta demografi pengguna Twitter berdasarkan jumlah pria dan wanita per provinsi, terlihat bahwa Pulau Jawa mendominasi dengan jumlah pengguna yang paling tinggi. Hal ini mengilustrasikan pentingnya analisis demografis dalam memahami dinamika komunikasi publik, khususnya dalam konteks politik. Seiring dengan pertumbuhan ini, pemahaman mendalam tentang demografi pengguna Twitter menjadi semakin relevan untuk memahami preferensi dan kecenderungan pemilih dalam lanskap politik.



**Gambar 1.** Peta Demografi Jumlah Pria per Provinsi



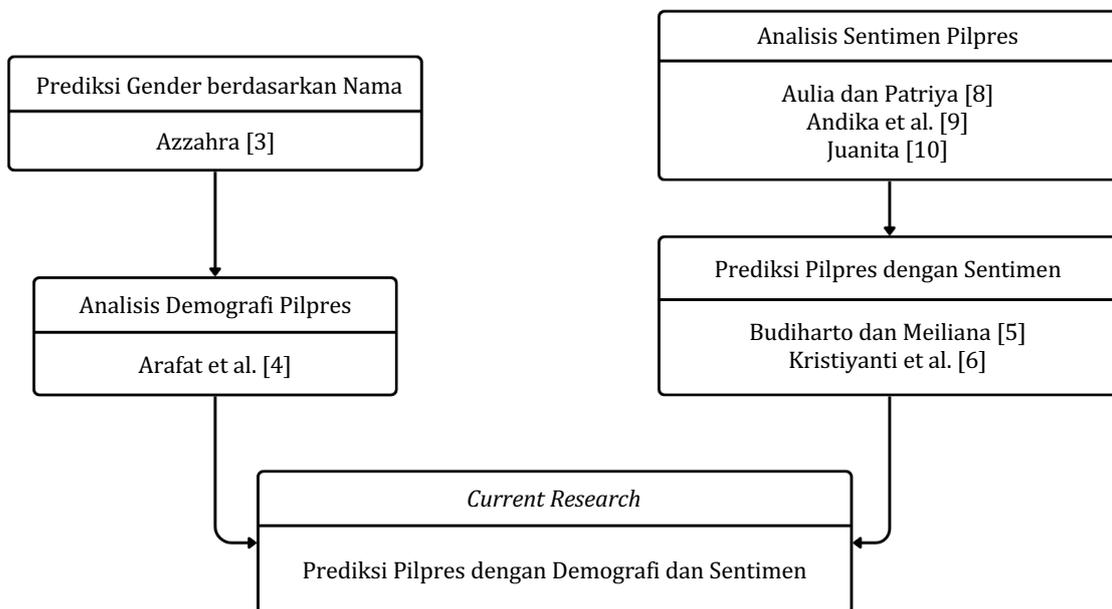
**Gambar 2.** Peta Demografi Jumlah Wanita per Provinsi

Salah satu tantangan utama dalam melakukan analisis demografi pada media sosial adalah keterbatasan informasi demografi yang biasanya tidak disediakan oleh media sosial. Twitter, salah satu media sosial yang menyajikan beberapa data mengenai lokasi pengguna, tidak semua pengguna memilih untuk membagikannya, sehingga membuat data tersebut tidak lengkap. Selain itu, informasi spesifik seperti gender tidak tersedia. Hal ini menimbulkan masalah dalam model prediksi yang membutuhkan informasi demografi. Namun, seiring dengan perkembangan ilmu kecerdasan buatan saat ini menawarkan solusi inovatif.

Informasi demografi dapat diprediksi menggunakan machine learning atau melalui augmentasi data eksternal. Sebagai contoh, jenis kelamin pengguna dapat diprediksi berdasarkan gambar profil atau nama tampilan mereka [1] [2]. Penelitian yang dilakukan oleh Azzahra [3] mengeksplorasi 15.349 nama dari berbagai daerah di Indonesia berhasil memprediksi jenis kelamin dengan akurasi hingga 93,58% menggunakan model BiLSTM. Pendekatan ini membuka peluang baru dalam memanfaatkan informasi demografi.

Pendekatan serupa telah diadopsi dalam penelitian yang dilakukan oleh Arafat et al. [4]. Melalui metode klasifikasi teks dan kamus nama serta jenis kelamin yang berasal dari data pemilih KPU, penelitian tersebut dapat mengekstraksi informasi demografi dari pengguna Twitter. Hal ini menunjukkan potensi adanya pendekatan yang memanfaatkan data eksternal dan *machine learning* untuk analisis demografi di media sosial.

Meskipun pendekatan demografi memberikan wawasan mengenai karakteristik pengguna, beberapa peneliti lainnya juga telah memanfaatkan data media sosial seperti melakukan prediksi hasil pemilihan presiden. Sebagai contoh, penelitian yang dilakukan oleh Budiharto dan Meiliana [5] dan Kristiyanti et al. [6] yang memprediksi hasil pemilihan berdasarkan analisis sentimen dari tweet yang dikumpulkan, tanpa menekankan pada variabel demografi. Berbeda dengan penelitian-penelitian sebelumnya, penelitian ini mengintegrasikan augmentasi informasi demografi menggunakan kecerdasan buatan dan prediksi hasil pemilihan presiden. Selain itu, penelitian ini menggunakan data latih Pilpres 2014 untuk melakukan prediksi di Pilpres 2019 yang belum pernah dilakukan sebelumnya.



**Gambar 3.** Diagram Studi Pustaka

Dalam penelitian yang dilakukan oleh Arafat et al. [4], Twitter menjadi alat kampanye politik, namun memiliki keterbatasan berupa ketiadaan informasi demografi. Fokus penelitian tersebut adalah mengidentifikasi pola demografi pendukung setiap kandidat presiden Indonesia 2019. Dengan metode klasifikasi

teks dan kamus berdasarkan data pemilih tetap KPU, hasil menunjukkan mayoritas pengguna aktif yang berdiskusi mengenai pemilihan presiden 2019 adalah perempuan dengan kategori usia milenial. Metode tersebut menjadi landasan penting bagi eksplorasi lebih lanjut dalam penelitian ini.

Selain analisis demografi, Budiharto dan Meiliana [5] memanfaatkan tweet dari kandidat presiden Indonesia, Jokowi dan Prabowo, serta *hashtag* yang relevan selama periode Maret hingga Juli 2018 untuk analisis sentimen. Hal serupa dilakukan oleh Kristiyanti et al. [6] yang menggunakan analisis sentimen untuk memprediksi hasil pemilihan presiden 2019. Kristiyanti et al. [6] memanfaatkan algoritma *Support Vector Machine* (SVM) yang dikombinasikan dengan teknik seleksi fitur *Particle Swarm Optimization* (PSO) dan *Genetic Algorithms* (GA). Metode SVM dengan kombinasi PSO mendapatkan akurasi mencapai 86,20%. Namun, akurasi ini merupakan akurasi dari sentimen analisis, dan bukan hasil prediksi Pilpres seperti yang dilakukan dalam penelitian ini.

Harkan dan Eriyanto [7] melakukan analisis akurasi dari indeks Google Trends yang mewakili kedua calon untuk memprediksi penghitungan suara nyata dari KPU. Hasilnya menunjukkan bahwa penggunaan data dari Google Trends untuk memprediksi hasil Pilpres Indonesia 2019 tidak tepat. Hal ini menunjukkan signifikansi permasalahan prediksi hasil Pilpres yang tidak mudah.

Terlihat pada Gambar 3, dalam lanskap penelitian seputar pemilihan presiden Indonesia, beragam pendekatan telah diterapkan, mulai dari analisis demografi hingga analisis sentimen [8] [9] [10]. Namun, ada hal yang belum banyak dieksplorasi, yaitu kombinasi analisis demografi dengan analisis sentimen untuk memprediksi hasil pemilihan dan juga masih minimnya penelitian yang mencoba memprediksi hasil Pilpres di masa depan. Dengan mengintegrasikan kedua pendekatan tersebut, penelitian ini akan dapat menentukan apakah variabel demografi dapat meningkatkan performa model ketika melakukan prediksi, terutama pada kasus peramalan atau *forecasting*.

Analisis demografi dalam penelitian ini terinspirasi dari model yang telah dikembangkan oleh Arafat et al. [4]. Namun pada penelitian tersebut hanya fokus melakukan analisis demografi, tidak melakukan prediksi Pilpres, dan tidak membandingkan data dari KPU.

Analisis sentimen dalam penelitian ini menggunakan model BERT (Bidirectional Encoder Representations from Transformers) yang pertama kali diperkenalkan oleh Devlin et al. [11], inovasi baru dalam representasi bahasa, yang dirancang untuk pre-training representasi bidirectional dari teks tanpa label dengan mengkondisikan kedua konteks kiri dan kanan dalam semua lapisannya. BERT telah menunjukkan peningkatan signifikan dalam tugas-tugas seperti klasifikasi teks ilmiah. Sebagai contoh, model CovBERT, yang berdasarkan pada model BERT, telah digunakan untuk mengotomatisasi proses tinjauan literatur pada publikasi ilmiah terkait Covid-19 [12]. BERT muncul sebagai model pembelajaran mesin unggulan yang populer dalam beberapa tahun terakhir, mampu menangani berbagai tugas NLP (Natural Language Processing) seperti klasifikasi teks. Eksperimen menunjukkan superioritas BERT dalam berbagai skenario, menegaskan kredibilitasnya sebagai teknik utama yang digunakan dalam masalah NLP [13]. Selain itu, ada juga varian BERT khusus untuk bahasa Indonesia yang dikenal sebagai IndoBERT, yang telah dilatih khusus pada korpora berbahasa

Indonesia dan telah menunjukkan hasil yang baik dalam pemahaman konteks dan nuansa bahasa lokal [14].

## B. Metode Penelitian

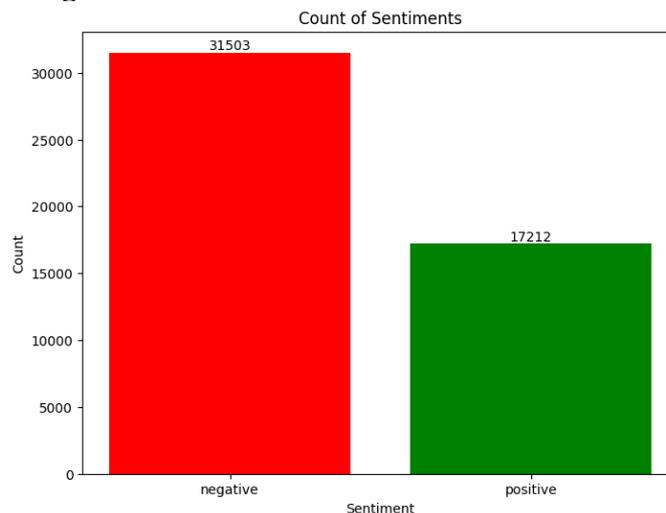
Sumber data yang digunakan berasal dari media sosial Twitter, khususnya *tweet* yang berkaitan dengan Pemilihan Presiden Indonesia pada tahun 2014 dan 2019 dengan *keyword* "jokowi" dan "prabowo". Proses awal melibatkan penyaringan dari total 7 juta *tweet* tahun 2014 dan 3 juta *tweet* tahun 2019. Kriteria penyaringan fokus pada *tweet* yang memuat informasi nama pengguna serta kode wilayah pengguna yang kemudian dapat diidentifikasi sebagai lokasi provinsi. Dari proses penyaringan tersebut, data tahun 2014 tereduksi menjadi 83.173 *tweet*, sementara tahun 2019 menghasilkan 63.216 *tweet* yang memenuhi kriteria. Alasan pemilihan data dari dua periode pemilihan presiden yang berbeda adalah agar data *tweet* dari tahun 2014 dapat dijadikan sebagai dasar pelatihan (*training set*) dalam membangun dan menyesuaikan model analisis. Di sisi lain, data dari tahun 2019 dijadikan sebagai bahan uji untuk memprediksi hasil pemilihan. Melalui pendekatan ini, harapannya adalah model yang telah terlatih dengan data tahun 2014 mampu memprediksi kecenderungan hasil pemilihan tahun 2019 dengan efektif, berdasarkan analisis sentimen dan demografi pengguna Twitter.

## C. Hasil dan Pembahasan

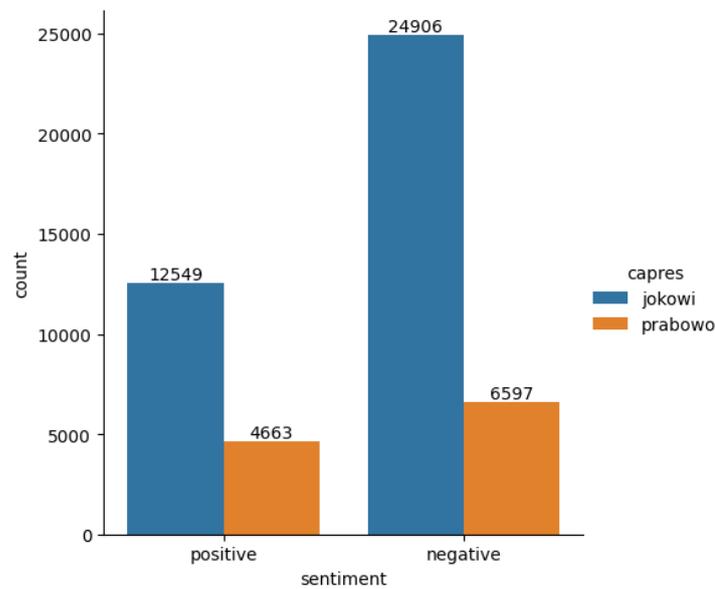
### 1. Analisis Sentimen

Pada tahap pelabelan sentimen, dimanfaatkan model BERT khusus yang telah dilatih untuk klasifikasi sentimen teks berbahasa Indonesia. Model tersebut diidentifikasi dengan nama "*mdhugol/indonesia-bert-sentiment-classification*" [15]. Hasil dari klasifikasi ini membagi teks ke dalam dua label yaitu positif dan negatif.

Sebaran sentimen untuk masing-masing calon memberikan gambaran awal tentang opini publik terhadap mereka berdasarkan data Twitter yang dikumpulkan. Data sentimen positif dan negatif dibagi berdasarkan calon presiden untuk mendapatkan pemahaman yang lebih mendalam tentang sebaran sentimen terhadap masing-masing calon.



**Gambar 4.** Jumlah *Tweet* Positif dan Negatif



**Gambar 5.** Jumlah *Tweet* Positif dan Negatif setiap Calon

Terlihat pada Gambar 4 dan 5 bahwa secara keseluruhan data didominasi oleh *tweet* berlabel negatif. Ini memungkinkan adanya isu kontroversial yang sedang hangat diperbincangkan atau kampanye negatif dari masing-masing pendukung. Di era media sosial, kampanye negatif bisa menyebar cepat dan memengaruhi opini publik [16].

Dalam proses pelatihan model BERT untuk analisis sentimen, berbagai metrik kinerja digunakan seperti *loss* dan *accuracy*. Setiap iterasi atau *epoch* pelatihan dievaluasi terhadap dataset validasi untuk memastikan model tidak hanya belajar dari data latih tetapi juga berupaya untuk melakukan generalisasi dengan baik untuk data uji.

- *Loss*: Indikator utama dari seberapa baik model memprediksi sentimen selama fase pelatihan. Nilai kerugian yang lebih rendah menunjukkan prediksi yang lebih baik.
- *Accuracy*: Mengukur persentase tebakan yang benar dari model pada dataset validasi.

Setelah melalui proses pelatihan, model sudah memiliki akurasi yang baik pada data validasi dengan nilai 0.94. Ini menandakan bahwa model telah berhasil memahami pola dalam data pelatihan dan mampu mengeneralisasikannya dengan efektif ke data yang belum dilihatnya.

## 2. Analisis Demografi Prediksi Gender

Menggunakan model yang sudah ada [4], variabel nama pengguna diprediksi untuk menentukan gender dari setiap pengguna. Penting untuk memeriksa kinerja model klasifikasi gender yang digunakan. Untuk memastikan validitas hasil klasifikasi, dilakukan proses pengambilan sampel sebanyak 10 kali pada setiap iterasinya:

- Sebanyak 30 baris data dipilih secara acak dari kumpulan data.
- Menggunakan regresi logistik sebagai model klasifikasi, mengingat hanya terdapat dua variabel target, yaitu pria dan wanita.

- Model dilatih dengan subset data yang dipilih dan akurasi dicatat.

Setelah 10 iterasi selesai, dapat dihitung rata-rata dari seluruh skor akurasi untuk mendapatkan gambaran akurasi keseluruhan dari model klasifikasi yang digunakan. Diperoleh hasil rata-rata dari seluruh skor akurasi yaitu 0.88. Dengan mengikuti pendekatan ini, dapat dikuatkan bukti bahwa pelabelan gender dalam analisis demografi dilakukan dengan cara yang akurat dan dapat diandalkan. Selanjutnya, hasil ini dapat digunakan untuk analisis lebih lanjut.

### 3. Hasil Feature Engineering

Setelah melalui analisis sentimen dan analisis demografi, kedua hasil analisis tersebut digabung menjadi satu tabel. Tabel ini memberikan gambaran tentang struktur data akhir yang siap digunakan untuk pemodelan, memastikan bahwa setiap aspek penting dari analisis sebelumnya telah diintegrasikan dengan baik. Selain gender dan sentimen, ditambahkan data eksternal berupa hasil suara Pilpres 2014 yang bersumber dari KPU [17], jumlah penduduk, dan jumlah penduduk yang terhubung internet pada tahun 2014 [18].

*Keterangan: J untuk Jokowi dan P untuk Prabowo.*

**Tabel 1.** Contoh Data Hasil Feature Engineering

Provinsi	Pria	Wanita	Negatif J	Negatif P	Positif J	Positif P	Hasil Suara J	Hasil Suara P	Jumlah Penduduk	Jumlah Penduduk Internet
Aceh	804	504	643	185	310	170	913309	1089290	4906800	556431
Bali	1	0	0	0	1	0	1535110	614241	4104900	920729
Banten	396	219	343	64	161	47	2398631	3192671	11704900	2466222
Bengkulu	6	3	3	2	4	0	523669	433173	1844800	277827
...	...	...	...	...	...	...	...	...	...	...

Tabel 1 merupakan cuplikan dari hasil secara keseluruhan, di mana terdapat 31 provinsi di Indonesia yang ada pada tabel. Setiap provinsi dilengkapi dengan informasi demografis, sentimen, dan data eksternal yang relevan. Variasi data antar provinsi menunjukkan dinamika yang berbeda-beda, mencerminkan keragaman sosial, budaya, dan politik dari masing-masing wilayah. Adanya data eksternal seperti hasil suara Pilpres 2014, jumlah penduduk dan jumlah penduduk yang terhubung internet memberikan konteks tambahan dalam memahami preferensi dan kecenderungan pemilih di setiap provinsi. Penting untuk dicatat bahwa hasil suara Pilpres 2014, yang bersumber dari KPU, digunakan sebagai variabel target dalam pemodelan prediksi. Oleh karena itu, data tersebut memiliki peran krusial dalam mengukur sejauh mana model dapat memprediksi dengan akurat berdasarkan variabel-variabel yang ada. Dengan data yang komprehensif ini, diharapkan pemodelan lebih lanjut dapat menghasilkan prediksi yang lebih baik.

### 4. Prediksi Hasil Pemilihan

Dalam upaya memprediksi hasil pemilihan, analisis dilakukan dengan menggunakan dua set data berbeda: satu set data menghasilkan dari *feature engineering* yang mempertimbangkan variabel demografi, dan satu set data lainnya tanpa memasukkan variabel demografi. Untuk masing-masing set data ini, telah diuji tiga model klasifikasi yang berbeda untuk menentukan pendekatan mana yang paling akurat dalam meramalkan hasil pemilihan berdasarkan fitur-fitur yang telah diperoleh dari analisis sebelumnya.

**Tabel 2.** Akurasi Prediksi Hasil Pemilihan

	Random Forest	SVM	Regresi Logistik
Dengan variabel demografi	76.67%	60.00%	56.67%
Tanpa variabel demografi	63.33%	56.67%	56.67%

Dari Tabel 2 dapat diamati bahwa variabel demografi berperan cukup signifikan dalam meningkatkan akurasi prediksi, khususnya saat menggunakan model Random Forest. Ini menunjukkan bahwa variabel demografi memiliki dampak positif dan informatif terhadap keakuratan prediksi dalam konteks pemilihan presiden.

- Random Forest: Model ini mendapatkan manfaat terbesar dari penambahan variabel demografi dengan kenaikan akurasi sebesar 13.34% dibandingkan tanpa variabel demografi. Hal ini mungkin menunjukkan bahwa model Random Forest mampu memanfaatkan relasi antar variabel dalam data demografi.
- SVM: Meskipun ada peningkatan dalam akurasi saat memasukkan variabel demografi, peningkatannya tidak sebesar Random Forest. Ini mungkin mengindikasikan bahwa variabel demografi memberikan informasi tambahan yang berguna bagi model, tetapi SVM mungkin tidak seefisien Random Forest dalam memanfaatkannya.
- Regresi Logistik: Akurasi model ini stabil baik saat menggunakan variabel demografi maupun tanpa demografi. Hal ini menunjukkan bahwa, bagi model Regresi Logistik, penambahan variabel demografi mungkin tidak memberikan informasi tambahan yang signifikan untuk meningkatkan keakuratan prediksi.

Menganalisis lebih lanjut pada model Random Forest dengan kinerja terbaik, berikut adalah signifikansi setiap fitur dalam memengaruhi akurasi prediksi:

**Tabel 3.** Feature Importance

Feature	Importance
Pria	0.0473
Wanita	0.0906
Negative J	<b>0.1987</b>
Negative P	0.0912
Positive J	0.1204
Positive P	0.1665
Jumlah Penduduk	<b>0.1692</b>
Jumlah Penduduk Terhubung Internet	0.1161

Berdasarkan Tabel 3, variabel demografi seperti jumlah penduduk dan jumlah penduduk terhubung internet memiliki nilai *importance* yang cukup tinggi. Jumlah penduduk dapat mencerminkan sebaran geografis pemilih, sementara jumlah penduduk yang terhubung internet menunjukkan seberapa aktif dan terinformasi pemilih di wilayah tersebut. Sementara itu, variabel demografi seperti gender (pria dan wanita) memiliki nilai *importance* yang lebih rendah dibandingkan dengan variabel lain, meskipun mereka tetap memberikan kontribusi penting.

#### D. Simpulan

Dalam analisis pemilihan presiden berdasarkan data dari Twitter, penambahan variabel demografi memberikan dampak signifikan terhadap akurasi

model prediksi. Dengan mengintegrasikan variabel demografi, terjadi peningkatan akurasi sebesar 13.34%, yaitu dari 63.33% tanpa variabel demografi menjadi 76.67% dengan variabel demografi. Dari semua variabel demografi yang dianalisis, jumlah penduduk menonjol sebagai variabel dengan nilai importance tertinggi.

Saran untuk penelitian selanjutnya, meskipun jumlah penduduk menonjol sebagai variabel dengan nilai importance tertinggi, variabel demografi lainnya seperti distribusi pendidikan, pendapatan rata-rata, atau tingkat urbanisasi mungkin juga memiliki dampak terhadap preferensi pemilihan. Menganalisis bagaimana variabel-variabel ini memengaruhi kecenderungan pemilih dapat memberikan wawasan tambahan.

### **E. Ucapan Terima Kasih**

Penulis menyampaikan ucapan terima kasih kepada Bapak Taufik Edy Sutanto dan Bapak Muhaza Liebenlito sebagai Dosen Pembimbing yang telah memberikan arahan, saran, dan masukan yang sangat berharga selama proses penelitian berlangsung. Semoga hasil dari penelitian ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan dan menjadi inspirasi bagi peneliti-peneliti lainnya di masa yang akan datang.

### **F. Referensi**

- [1] R. C. B. Rego, V. M. L. Silva, dan V. M. Fernandes, "Predicting Gender by First Name Using Character-level Machine Learning," 2021, doi: 10.48550/ARXIV.2106.10156.
- [2] S. Sakib *dkk.*, "Predicting Gender from Human or Non-human Social Media Profile Photos by using Transfer Learning," dalam *2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, India: IEEE, Jan 2023, hlm. 1–7. doi: 10.1109/ICCECE51049.2023.10085525.
- [3] N. Azzahra, "Implementasi deep learning untuk memprediksi jenis kelamin berdasarkan nama (studi kasus : nama-nama orang Indonesia)," UIN Syarif Hidayatullah Jakarta, 2022. [Daring]. Tersedia pada: <https://repository.uinjkt.ac.id/dspace/handle/123456789/65563>
- [4] T. A. Arafat, I. Budi, R. Mahendra, dan D. A. Salehah, "Demographic Analysis of Candidates Supporter in Twitter During Indonesian Presidential Election 2019," dalam *2020 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia: IEEE, Nov 2020, hlm. 1–6. doi: 10.1109/ICISS50791.2020.9307598.
- [5] W. Budiharto dan M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *J. Big Data*, vol. 5, no. 1, hlm. 51, Des 2018, doi: 10.1186/s40537-018-0164-1.
- [6] D. A. Kristiyanti, Normah, dan A. H. Umam, "Prediction of Indonesia Presidential Election Results for the 2019-2024 Period Using Twitter Sentiment Analysis," dalam *2019 5th International Conference on New Media Studies (CONMEDIA)*, Bali, Indonesia: IEEE, Okt 2019, hlm. 36–42. doi: 10.1109/CONMEDIA46929.2019.8981823.
- [7] A. A. Harkan dan Eriyanto, "Predicting the Results of the 2019 Indonesian Presidential Election with Google Trends: Analysis of Accuracy, Precision, and

- Its Opportunity,” dipresentasikan pada Asia-Pacific Research in Social Sciences and Humanities Universitas Indonesia Conference (APRISH 2019), Jakarta, Indonesia, 2021. doi: 10.2991/assehr.k.210531.001.
- [8] G. N. Aulia dan E. Patriya, “Implementasi Lexicon Based dan Naive Bayes pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019,” *J. Ilm. Inform. Komput.*, vol. 24, no. 2, hlm. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.
- [9] L. A. Andika, P. A. N. Azizah, dan R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, hlm. 34, Jul 2019, doi: 10.13057/ijas.v2i1.29998.
- [10] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, hlm. 552, Jul 2020, doi: 10.30865/mib.v4i3.2140.
- [11] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018, doi: 10.48550/ARXIV.1810.04805.
- [12] M. Khadhraoui, H. Bellaaj, M. B. Ammar, H. Hamam, dan M. Jmaiel, “Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study,” *Appl. Sci.*, vol. 12, no. 6, hlm. 2891, Mar 2022, doi: 10.3390/app12062891.
- [13] S. González-Carvajal dan E. C. Garrido-Merchán, “Comparing BERT against traditional machine learning text classification,” 2020, doi: 10.48550/ARXIV.2005.13012.
- [14] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” dalam *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, hlm. 757–770. doi: 10.18653/v1/2020.coling-main.66.
- [15] mdhugol, “Indonesian BERT Base Sentiment Classifier,” 2021. Diakses: 29 Agustus 2023. [Daring]. Tersedia pada: <https://huggingface.co/mdhugol/indonesia-bert-sentiment-classification>
- [16] F. Syarief, “Pemanfaatan Media Sosial dalam Proses Pembentukan Opini Publik (Analisa Wacana Twitter SBY),” *J. Komun.*, 2017.
- [17] “Hasil Penghitungan Perolehan Suara dari Setiap Provinsi dan Luar Negeri dalam Pemilu Presiden dan Wakil Presiden Tahun 2014,” Jul 2014. Diakses: 29 Agustus 2023. [Daring]. Tersedia pada: [https://www.kpu.go.id/koleksigambar/PPWP\\_-\\_Nasional\\_Rekapitulasi\\_2014\\_-\\_New\\_-\\_Final\\_2014\\_07\\_22.pdf](https://www.kpu.go.id/koleksigambar/PPWP_-_Nasional_Rekapitulasi_2014_-_New_-_Final_2014_07_22.pdf)
- [18] *Statistik Indonesia 2014*. Badan Pusat Statistik, 2014. Diakses: 29 Agustus 2023. [Daring]. Tersedia pada: <https://www.bps.go.id/publication/2014/05/05/8d2c08d9d41aa8c02fad22e7/statistik-indonesia-2014.html>