

PEMODELAN KLASIFIKASI PREFERENSI IDEOLOGI PARTAI POLITIK PADA PEMILIH PEMULA DI PROVINSI BANTEN

Agung Satrio Wicaksono^{1*}, Moh. Rizky Godjali², Ika Arinia Indriyany³, Weksi Budiaji⁴,
Aulia Ikhsan⁵

^{1,4,5} Program Studi Statistika, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa

^{2,3} Program Studi Ilmu Pemerintahan, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Sultan
Ageng Tirtayasa

*e-mail: agungatriow@untirta.ac.id

DOI: 10.14710/j.gauss.13.2.421-430

Article Info:

Received: 2024-09-09

Accepted: 2024-12-11

Available Online: 2024-12-12

Keywords:

*first-time voters; political party
ideology; election; classification
modeling; random forest*

Abstract: First-time voters in the 2024 election are expected to play an important role in determining the future direction of Indonesia, by bringing fresh perspectives and new aspirations for change amidst increasingly complex political dynamics. This study aims to apply classification modeling using political party preference data on first-time voters in Banten Province. Data were collected through a survey with responses from 3 different party ideologies, namely the Secular National Party (PNS), the Religious National Party (PNR), and the Islamic Party (PI). The analysis stages include data preprocessing, forming a Random Forest classification model, evaluating model performance using accuracy, precision, recall, and F1 score measures, and mapping the importance of features from the Random Forest results. The SMOTE technique is used to handle class imbalance, where the majority of party ideologies from the data obtained are PNS, which is 50%. The results obtained from 10-fold cross-validation with multiclass classification show an accuracy of 64.00%, with precision, recall, and F1 score values of 61.88%, 62.33%, and 60.41%. The variable importance of political party ideology is the ideological background of the voters themselves, with a Mean Decrease Gini (MDG) score is 91.28.

1. PENDAHULUAN

Demokrasi memegang peranan sentral dalam menjaga stabilitas dan kesejahteraan suatu negara. Di dalamnya, pemilihan umum menjadi salah satu pilar penting dalam proses pengambilan keputusan politik. Pemilih pemula, merupakan segmen khusus dari masyarakat yang baru pertama kali memasuki usia kelayakan untuk memberikan suara mereka dalam proses demokratisasi (Azirah, 2019). Dalam konteks Indonesia, kelompok ini menunjukkan kecenderungan yang signifikan dalam membentuk arah dan kebijakan politik di tingkat nasional maupun lokal. Provinsi Banten, sebagai salah satu wilayah dengan pertumbuhan pemilih pemula yang dinamis, menawarkan lanskap yang menarik untuk memahami perubahan dan kecenderungan perilaku politik.

Komisi Pemilihan Umum Republik Indonesia (KPU-RI) menyebutkan bahwa jumlah pemilih pada Pemilu 2024 di Indonesia didominasi oleh pemilih muda, yang mana hampir 113 juta pemilih dari lebih 204 juta pemilih adalah generasi milenial dan generasi Z. Dari jumlah tersebut, jumlah generasi Z mencapai 60 juta orang. Dengan jumlah tersebut, Generasi Z merupakan kelompok pemilih terbesar yang akan mendominasi Pemilu 2024 di Indonesia dan sebagian besar merupakan pemilih pemula (Sitanggang *et al.*, 2024).

Dalam dinamika politik dan demokrasi di Indonesia yang telah berjalan pasca Reformasi 1998, Partai-Partai bertransformasi dalam keyakinan ideologi politik yang beragam. Santi (2018) menyebut bahwa partai politik dapat dikelompokkan menjadi tiga kelompok besar,

yaitu kelompok Partai Nasionalis Religius (PNR), Partai Nasionalis Sekuler (PNS), dan Partai Islam (PI). Dengan ideologinya masing-masing partai politik akan mempunyai identitas yang jelas, hal itulah kemudian yang memudahkan partai politik tersebut dalam mendapatkan massa pendukung, termasuk pemilih pemula (Santi, 2018). Pemahaman pemilih pemula tentang kompleksitas ideologi politiknya saat ini banyak dipengaruhi oleh nilai-nilai pribadi, stigma dari lingkungan sekitar, tradisi keluarga, serta popularitas melalui media sosial dan digital.

Random Forest merupakan salah satu algoritma yang cukup populer digunakan dalam pemodelan klasifikasi. Dalam hal ini, pemodelan klasifikasi dapat memetakan antara karakteristik pemilih pemula dengan ideologi partai politik yang dipilihnya, yang dalam hal ini adalah PNR, PNS, dan PI. *Random Forest* merupakan salah satu algoritma yang memungkinkan untuk menyelesaikan permasalahan *multiclass classification*. Algoritma ini menggunakan banyak *decision tree* untuk menghasilkan prediksi yang lebih akurat. *Random Forest* cenderung bekerja dengan baik pada data yang memiliki banyak fitur dan kategori (Mahmuda, 2024). Penelitian yang dilakukan oleh Iman, dkk. 2022 berjudul “Perbandingan Algoritma Klasifikasi *Random Forest* dan *Extreme Gradient Boosting* pada Dataset Cuaca Provinsi DKI Jakarta Tahun 2018” menunjukkan keunggulan metode *Random Forest* jika dibandingkan dengan *XGBoost* dalam menyelesaikan masalah *multiclass classification*, yang mana data cuaca digunakan dalam penelitian tersebut (Iman *et al.*, 2022). Selain itu, penelitian yang dilakukan oleh Junus, dkk. 2023 berjudul “Klasifikasi Menggunakan Metode Support Vector Machine Dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus” juga menunjukkan keunggulan algoritma klasifikasi *Random Forest* dengan nilai akurasi, presisi, recall dan F1 score yang dihasilkan lebih tinggi dibandingkan algoritma SVM (Junus, Tarno and Kartikasari, 2023). Oleh karena itu, penelitian ini menggunakan metode *Random Forest* untuk menyelesaikan permasalahan *multiclass classification*.

Penelitian ini membahas tentang penerapan model klasifikasi untuk memetakan preferensi ideologi partai politik berdasarkan karakteristik pemilih pemula di Provinsi Banten. Model yang diperoleh dapat menjadi acuan bagi partai politik di Provinsi Banten untuk menentukan strategi pengumpulan suara yang cocok sesuai dengan masing-masing ideologi partai politiknya, khususnya bagi kalangan pemilih pemula.

2. TINJAUAN PUSTAKA

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau kategori dari entitas berdasarkan fitur atau atributnya, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan (Sihombing and Yulianti, 2021). Klasifikasi merupakan bagian dari jenis *supervised learning*. Klasifikasi pada suatu data secara umum terdiri dari dua proses. Proses pertama akan dilaksanakan pembangunan atau pembentukan model klasifikasi berdasarkan data sebelumnya. Proses kedua dalam klasifikasi adalah menentukan akurasi dari model klasifikasi yang dilaksanakan apakah layak untuk diterapkan pada data tersebut ataupun untuk data baru (Rizal, Fatekurohman and Anggraeni, 2024). Salah satu algoritma yang digunakan dalam pemodelan klasifikasi adalah *Random Forest*.

Random Forest merupakan suatu algoritma klasifikasi yang terdiri dari kumpulan model klasifikasi berstruktur pohon (*tree*) $\{h(x,k), k=1,\dots\}$ yang mana $\{k\}$ merupakan vektor acak yang terdistribusi identik dan independen, dan setiap pohon memberikan suara satuan untuk kelas yang paling populer pada masukan x (Ramayanti *et al.*, 2023). Algoritma *Random Forest* merupakan pengembangan dari Metode *Classification and Regression Tree* (CART). *Random Forest* menerapkan metode *bootstrap aggregating* (bagging) dan *random feature*

selection. Dalam *Random Forest*, pohon CART banyak terbentuk sedemikian sehingga membentuk kumpulan pohon atau hutan. Kemudian dilakukan analisis terhadap kumpulan pohon tersebut sehingga dapat digunakan untuk mengklasifikasikan respon, baik biner maupun *multiclass* (Hadiana, Soleh and Sartono, 2020).

Algoritma *Random Forest* untuk regresi dan klasifikasi adalah sebagai berikut (Hastie, Tibshirani and Friedman, 2017):

1. untuk $b = 1$ hingga B :
 - a. Gambarkan sampel *bootstrap* Z^* berukuran N dari data latih;
 - b. Bangun pohon *random-forest* T_b dari data *bootstrap*, dengan mengulang langkah-langkah berikut secara rekursif untuk setiap simpul terminal pohon, hingga ukuran simpul minimum n_{min} tercapai:
 - i. Pilih m peubah secara acak dari p peubah;
 - ii. Ambil peubah/titik pisah terbaik diantara m peubah tersebut;
 - iii. Pisahkan simpul tersebut menjadi dua simpul cabang.

2. Keluarkan kumpulan pohon $\{T_b\}_1^B$, untuk membuat prediksi pada titik baru x :

- a. Regresi:

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

dengan $\hat{f}(x)$ merupakan prediksi regresi *random-forest* pada titik baru x , T merupakan pohon, B merupakan banyaknya pohon yang dibangun.

- b. Klasifikasi:

Misalkan $\hat{C}_b(x)$ adalah prediksi kelas dari pohon *random-forest* ke- b , maka:

$$\hat{C}_{rf}^B(x) = \text{majority vote} \{ \hat{C}_b(x) \}_1^B \quad (2)$$

Salah satu permasalahan yang kerap dihadapi dalam pemodelan klasifikasi baik biner maupun *multiclass* yaitu permasalahan ketidakseimbangan kelas (*class imbalance*). Salah satu metode yang cukup populer digunakan dalam menangani kasus ini yaitu dengan membuat amatan sintetik melalui *oversampling* kelas amatan minoritas, atau yang biasa disebut *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini diaplikasikan sehingga amatan dengan kelas yang lebih sedikit atau amatan minoritas memiliki proporsi yang lebih seimbang dengan amatan mayoritas (Fernandez *et al.*, 2018). Prosedur untuk membuat amatan sintetik dalam SMOTE adalah sebagai berikut:

- i. Tentukan tetangga terdekat k , yang mana tetangga tersebut juga merupakan observasi kelas minoritas;
- ii. Pilih secara acak j dari k observasi pada poin (i), di mana nilai j bergantung pada jumlah *oversampling* yang telah ditentukan;
- iii. Buat observasi sintetis baru pada garis lurus observasi minoritas dengan tetangga yang dipilih.

Amatan sintetik x_s dihitung dengan interpolasi linier sebagai berikut:

$$x_s = x + \lambda(x_{neighbor} - x) \quad (3)$$

dengan x merupakan amatan kelas minoritas, $x_{neighbor}$ yaitu amatan tetangga terdekat dari kelas minoritas yang dipilih secara acak, dan λ adalah angka acak antara 0 dan 1, yang menghasilkan posisi di antara amatan asli dan amatan tetangganya.

Teknik *k-Fold Cross Validation* (KCV) merupakan salah satu pendekatan yang paling banyak digunakan untuk pemilihan model dan estimasi *error* dari sebuah pemodelan klasifikasi (Anguita *et al.*, 2012). KCV terdiri dari pemisahan himpunan data menjadi k subhimpunan, kemudian secara berulang beberapa di antaranya digunakan untuk membuat

model, beberapa lainnya dimanfaatkan untuk menilai kinerja atau yang disebut dengan evaluasi model. Evaluasi ini dilakukan dengan melihat ukuran kebaikan model yang dibangun dari gugus data latih dan dievaluasi dengan menggunakan data uji. *Confusion matrix* merupakan salah satu teknik yang dapat dilakukan dalam melihat ukuran kebaikan model. *Confusion matrix* akan menghasilkan berbagai ukuran kinerja dari model yang telah dibuat. Ukuran yang sering digunakan adalah *accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy mengukur proporsi prediksi yang benar dari seluruh prediksi yang dibuat. *Precision* mengukur proporsi prediksi kelas positif yang benar dari seluruh prediksi kelas positif yang dibuat. *Recall (sensitivity, true positive rate)* mengukur proporsi prediksi kelas positif yang benar dari seluruh data yang seharusnya positif. Sedangkan *F1-score* adalah rataan harmonis dari *precision* dan *recall*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - score = \frac{2(TP)}{2(TP+FP+FN)} \quad (7)$$

dengan TP adalah prediksi kelas positif yang benar (*True Positive*), TN adalah prediksi kelas negatif yang benar (*True Negative*), FP adalah prediksi kelas positif yang salah (*False Positive*), FN adalah prediksi kelas negatif yang salah (*False Negative*) (Fatmawati and Narti, 2022).

Identifikasi *feature importance* merupakan salah satu solusi untuk mempermudah perolehan informasi dalam *Random Forest*. Jika *feature importance* dapat diidentifikasi, maka dapat diperoleh metode penyeleksian variabel yang berpengaruh penting terhadap pembentukan pohon dalam *Random Forest*. Salah satu pengukuran *feature importance* dapat dihitung dari *Mean Decrease Gini* (MDG) (Ramadhan, Susetyo, and others, 2019). Misal suatu p peubah penjelas dengan $h=(1,2,\dots,p)$, maka pengukuran tingkat kepentingan peubah penjelas X_h dapat dihitung melalui rumus berikut (Christy and Suryowati, 2021):

$$MDG(X_h) = \frac{1}{k} \sum_{t=1}^k \sum_{t \in (T), v(t)=h} \frac{N_n(t)}{n} \Delta x^{(t)} \quad (8)$$

dengan $\Delta x^{(t)}$ yaitu indeks gini untuk peubah penjelas X_h pada pohon ke- k , h merupakan fitur yang sedang diukur tingkat kepentingannya; $N_n(t)$ yaitu jumlah sampel keseluruhan pada *leaf*; n adalah jumlah total sampel dalam *dataset*; serta k adalah banyaknya pohon dalam *random forest*.

3. METODE PENELITIAN

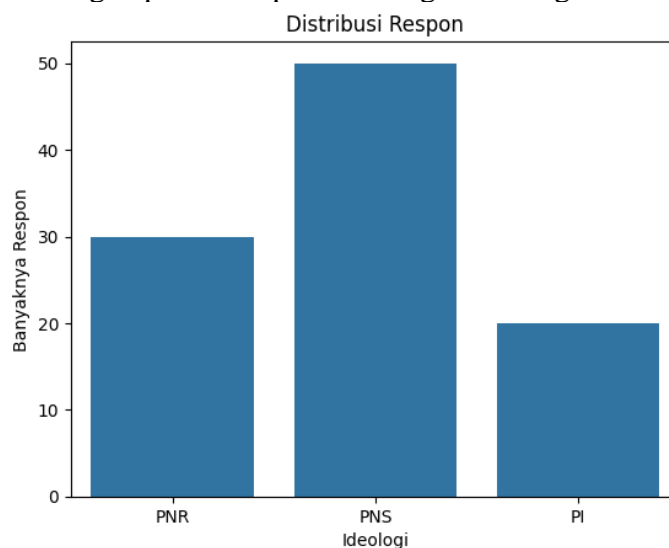
Data yang didapatkan dalam penelitian ini merupakan data primer yang dikumpulkan melalui survei yang dilakukan di Provinsi Banten. Responden merupakan pemilih pemula yang memiliki keterikatan emosional dengan suatu partai tertentu. Kriteria tersebut dianggap paling mewakili sebagai syarat untuk menjadi responden. Ideologi partai (PNS, PNR, PI) dalam penelitian ini merupakan variabel respon. Sedangkan variabel prediktor dalam penelitian ini yaitu karakteristik pemilih pemula yang diukur dari 10 variabel, yaitu: B1-Kesamaan daerah; B2-Kesamaan agama; B3-Latar belakang partai politik; B4-Ideologi Pemilih; B5-Informasi peserta pemilu; B6-Ketokohan peserta pemilu; B7-Popularitas peserta pemilu; B8-Kemampuan intelektual peserta pemilu; B9-Isu dan janji politik, serta; B10-Kasus hukum yang menjerat.

Setelah data dikumpulkan, data dianalisis menggunakan bantuan *software* Python melalui tahapan-tahapan berikut:

- i. Eksplorasi dan visualisasi data: untuk memahami struktur, pola, sebaran, dan karakteristik dari data yang telah dikumpulkan;
- ii. *k-Fold Cross Validation* (KCV): membagi data menjadi 10 bagian dengan proporsi kelas yang sama;
- iii. Pembentukan gugus data latih dan data uji: setiap data yang telah dibagi pada tahap (ii) bergantian menjadi data uji, sedangkan bagian sisanya membentuk gugus data latih untuk membuat model;
- iv. SMOTE: penanganan kasus ketidakseimbangan data dengan membuat amatan sintetik pada setiap gugus data latih;
- v. Pembentukan model *Random Forest*: Model dilatih menggunakan gugus data latih yang telah dibentuk. Model akan mempelajari hubungan antara variabel respon yaitu ideologi partai politik dengan prediktor yaitu karakteristik pemilih pemula;
- vi. Prediksi: Model yang telah dibuat kemudian diuji menggunakan data uji. Hasil prediksi kemudian digunakan untuk mengukur kebaikan model;
- vii. Mengukur performa model: Performa model *Random Forest* diukur menggunakan data uji. Pengukuran kebaikan model diukur dari nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Keempatnya digunakan untuk mengukur seberapa baik model dapat mengklasifikasikan respon dengan benar;
- viii. Ulangi langkah (iii) hingga (vii). Pengulangan ini dilakukan hingga setiap subset atau bagian data telah menjadi data uji;
- ix. Menghitung ukuran kebaikan model secara keseluruhan: rata-rata *accuracy*, *precision*, *recall*, dan *F1-score* untuk setiap iterasinya.
- x. *Feature importance*: mengukur tingkat kepentingan prediktor dalam mengklasifikasikan respon dengan menghitung *Mean Decrease Gini* (MDG).

4. HASIL DAN PEMBAHASAN

Gambar 1 menunjukkan hasil distribusi variabel respon dari data yang telah diperoleh. Dapat dilihat bahwa terjadi kasus ketidakseimbangan data, yang mana 50% data dengan label PNS, sedangkan PNR dan PI masing-masing sebesar 30% dan 20%. Hal ini menunjukkan kecenderungan politik responden dengan ideologi PNS.

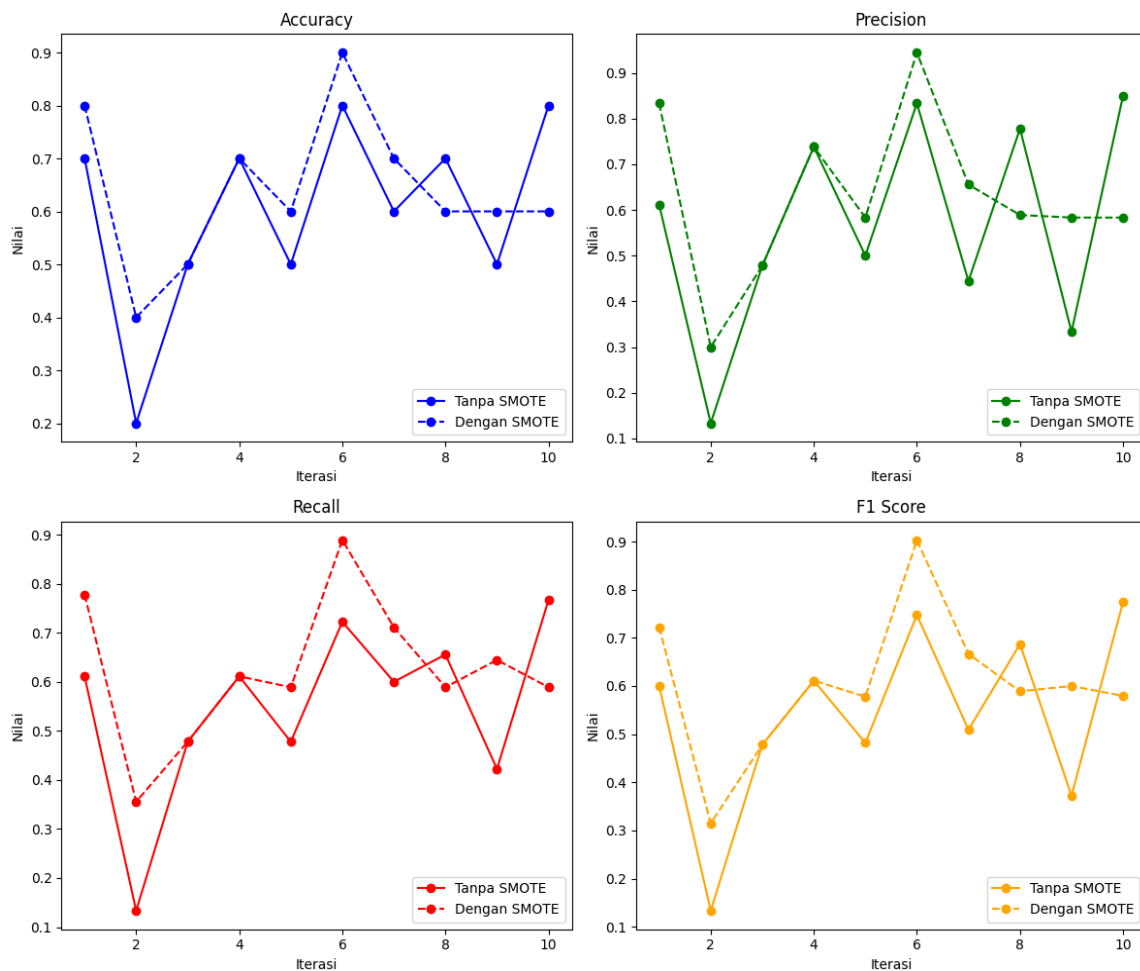


Gambar 1. Distribusi Variabel Respon

Selanjutnya, algoritma klasifikasi *Random Forest* digunakan dalam pemodelan dengan KCV sebanyak 10 lipatan. Metode KCV digunakan untuk melihat kestabilan hasil model yang diukur dari seberapa jauh data yang tersedia berada di sekitar rata-rata. Pada setiap pengulangan, dilakukan penanganan ketidakseimbangan data terhadap data latih dengan menggunakan SMOTE. Ukuran kebaikan model yang digunakan yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* merupakan perbandingan prediksi benar terhadap keseluruhan data. *Precision* merupakan perbandingan antara *True Positive* (TP) dengan banyaknya data yang diprediksi positif. *Recall* merupakan perbandingan antara TP dengan banyaknya data yang sebenarnya positif. Sedangkan *F1-score* merupakan rata-rata harmonik dari *precision* dan *recall*. Hasil dari Tabel 1 dan Gambar 2 menunjukkan bahwa penerapan SMOTE dapat meningkatkan ukuran kebaikan model.

Tabel 1. Rata-rata Ukuran Kebaikan Model

Rata-rata	Accuracy	Precision	Recall	F1-Score
Tanpa SMOTE	0.6	0.569	0.547	0.539
Dengan SMOTE	0.64	0.628	0.623	0.604

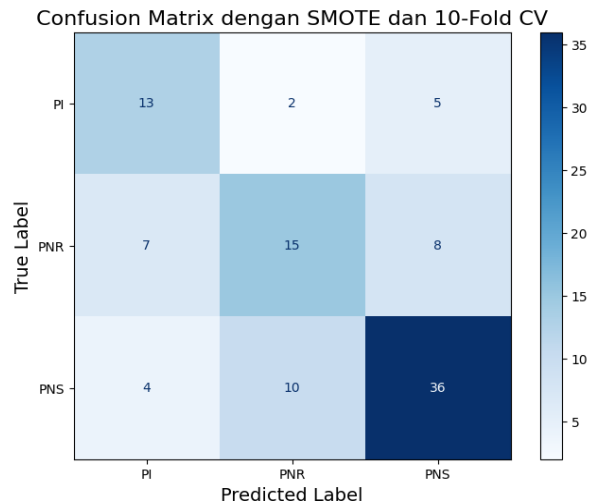


Gambar 2. Perbandingan Ukuran Kebaikan Model

Selanjutnya, setelah dilakukan pemodelan klasifikasi menggunakan algoritma *Random Forest* dengan 10-Fold CV dan SMOTE, didapatkan hasil *Classification Report* dan *Confusion Matrix* yang masing-masing disajikan pada Tabel 2 dan Gambar 3.

Tabel 2. *Classification Report*

	Precision	Recall	F1-Score	Support
PI	0.54	0.65	0.59	20
PNR	0.56	0.50	0.53	30
PNS	0.73	0.72	0.73	50
accuracy			0.64	100
macro avg	0.61	0.62	0.61	100
weighted avg	0.64	0.64	0.64	100

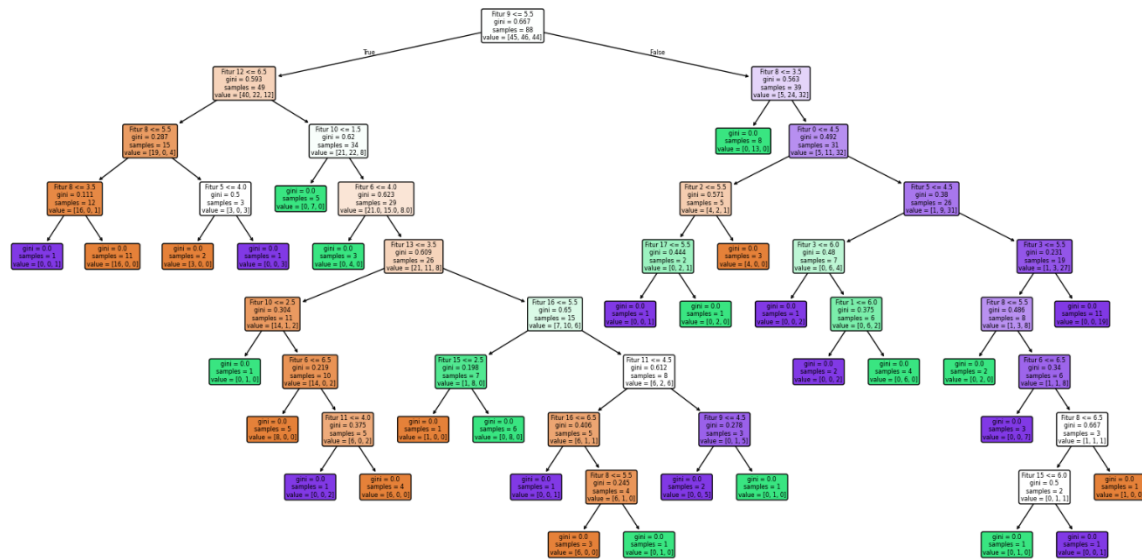


Gambar 3. Plot Hasil *Confusion Matrix*

Berdasarkan Gambar 3 dan Tabel 2 dapat diketahui bahwa dari 20 amatan dengan kategori PI, 13 diantaranya diprediksi dengan benar (65%). Kemudian dari 30 amatan dengan kategori PNR, 15 diantaranya diprediksi dengan benar (50%). Sedangkan dari 50 amatan dengan kategori PNS, 36 diantaranya diprediksi dengan benar (72%). Rata-rata akurasi yang diperoleh dari model tersebut yaitu sebesar 64%.

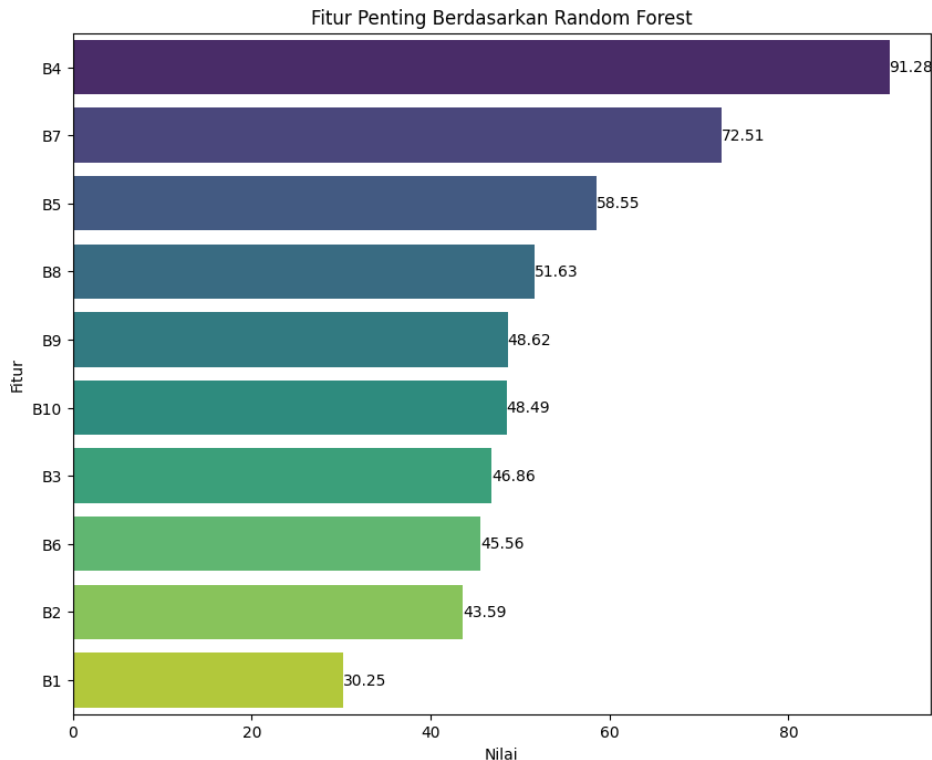
Gambar 4 menunjukkan visualisasi salah satu pohon dari *Random Forest*. Visualisasi ini membantu memahami bagaimana pohon individual dalam *Random Forest* membuat keputusan. Pada pohon berikut, terlihat bahwa *Node* teratas menunjukkan fitur 9 merupakan variabel yang digunakan untuk memisahkan data. Nilai *threshold* ≤ 5.5 merupakan batas untuk memisahkan data. Samples bernilai 88 merupakan banyaknya data yang mencapai *node* tersebut. Sedangkan nilai Gini 0.667 menunjukkan bahwa *node* tersebut memiliki ketidakhomogenan yang cukup tinggi. Artinya, proporsi sampel dari berbagai kelas di *node* ini cukup merata, tanpa dominasi oleh kelas tertentu.

Visualisasi Salah Satu Pohon dari Random Forest



Gambar 4. Visualisasi Salah Satu Pohon dari *Random Forest*

Prediktor yang berpengaruh dapat diketahui melalui nilai *feature importance* yang didapatkan dari hasil *random forest*.



Gambar 5. *Feature Importance*

Hasil *feature importance* didapatkan dari perhitungan *Mean Decrease Gini* (MDG). Gambar 4 menunjukkan bahwa B4-Ideologi Pemilih memiliki *importance value* atau tingkat kepentingan tertinggi untuk menentukan pilihan pemilih pemula terhadap ideologi partai politik dengan nilai 91.28. Selain itu, popularitas, serta informasi mengenai peserta pemilu juga menjadi pertimbangan penting lainnya bagi para pemilih pemula dalam menentukan pilihannya.

5. KESIMPULAN

Pemodelan klasifikasi menggunakan algoritma *Random Forest* melalui penerapan 10-Fold CV dan SMOTE menghasilkan *accuracy* sebesar 64%, dengan *precision*, *recall*, dan *F1 score* masing-masing yaitu sebesar 61.88%, 62.33%, dan 60.41%. Hasil tersebut juga memperlihatkan penanganan kasus ketidakseimbangan data dengan SMOTE mampu memberikan rata-rata ukuran kebaikan model yang lebih baik. Model tersebut juga menunjukkan bahwa ideologi pemilih merupakan hal yang paling menentukan dalam keterpilihan suatu partai. Selain itu, popularitas dan informasi mengenai peserta pemilu juga mempunyai peranan dalam menentukan pilihan pemilih pemula. Hal ini menunjukkan bahwa pemilih pemula di Provinsi Banten masih perlu diberikan edukasi untuk menjadi pemilih yang bijak agar tidak hanya mengandalkan popularitas, tetapi juga faktor lain yang melibatkan kemampuan calon, seperti misalkan isu dan janji politik yang dibawa agar sesuai dengan perkembangan kondisi wilayah Banten saat ini, serta kemampuan intelektual yang dimiliki. Di sisi lain, penting untuk partai politik dalam memberikan informasi yang lengkap mengenai latar belakang, identitas, dan gagasan yang dibawa oleh calon. Hal ini sejalan dengan perkembangan gen-Z dan milenial yang kerap berinteraksi dengan media sosial dan digital sebagai sumber informasi utama.

UCAPAN TERIMA KASIH

Tulisan ini merupakan hasil penelitian yang didanai oleh Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Sultan Ageng Tirtayasa. Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu kelancaran penelitian dan penulisan ini.

DAFTAR PUSTAKA

- Anguita, D. *et al.* (2012) ‘The “K” in K-fold Cross Validation’, *Computational Intelligence* [Preprint].
- Azirah (2019) ‘PARTISIPASI POLITIK PEMILIH PEMULA DALAM PESTA DEMOKRASI’, *Politica: Jurnal Hukum Tata Negara dan Politik Islam*, 6(2), pp. 86–100. Available at: <https://doi.org/10.32505/politica.v6i2.2735>.
- Christy, E. and Suryowati, K. (2021) ‘ANALISIS KLASIFIKASI STATUS BEKERJA PENDUDUK DAERAH ISTIMEWA YOGYAKARTA MENGGUNAKAN METODE RANDOM FOREST’, *Jurnal Statistika Industri dan Komputasi*, 6(1), pp. 69–76.
- Fatmawati, F. and Narti, N. (2022) ‘Perbandingan algoritma C4. 5 dan Naive Bayes dalam klasifikasi tingkat kepuasan mahasiswa terhadap pembelajaran daring’, *JTIM: Jurnal Teknologi Informasi dan Multimedia*, 4(1), pp. 1–12.
- Fernandez, A. *et al.* (2018) ‘SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary’, *Journal of Artificial Intelligence Research*, 61, pp. 863–905. Available at: <https://doi.org/10.1613/jair.1.11192>.
- Hadiana, R., Soleh, A.M. and Sartono, B. (2020) ‘Two-Stage Statistical Downscaling Modeling with Multi-Class Random Forest on Rainfall Prediction’, in *Proceedings of the Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia. Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019*,

- Bogor, Indonesia, Bogor, Indonesia: EAI. Available at: <https://doi.org/10.4108/eai.2-8-2019.2290525>.
- Hastie, T., Tibshirani, R. and Friedman, J. (2017) 'The elements of statistical learning: data mining, inference, and prediction'. Springer.
- Iman, A.H. *et al.* (2022) 'Perbandingan Algoritma Klasifikasi Random Forest dan Extreme Gradient Boosting pada Dataset Cuaca Provinsi DKI Jakarta Tahun 2018', in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, pp. 793–803.
- Junus, C.Z.V., Tarno, T. and Kartikasari, P. (2023) 'Klasifikasi Menggunakan Metode Support Vector Machine Dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus', *Jurnal Gaussian*, 11(3), pp. 386–396.
- Mahmuda, S. (2024) 'Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube', *JURNAL JENDELA MATEMATIKA*, 2(01), pp. 21–31. Available at: <https://doi.org/10.57008/jjm.v2i01.633>.
- Ramadhan, A., Susetyo, B., and others (2019) 'Penerapan Metode Klasifikasi Random Forest Dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan', *Jurnal Pendidikan dan Kebudayaan*, 4(2), pp. 169–182.
- Ramayanti, F. *et al.* (2023) 'Application of Random Forest to Identify for Poor Households in West Sumatera Province', *UNP Journal of Statistics and Data Science*, 1(2), pp. 97–104. Available at: <https://doi.org/10.24036/ujsds/vol1-iss2/31>.
- Rizal, N., Fatekurohman, M. and Anggraeni, D. (2024) 'KLASIFIKASI PENENTUAN LOKASI STRATEGIS OUTLET BANK SYARIAH INDONESIA DENGAN METODE NAÏVE BAYES CLASSIFIER', *Jurnal Gaussian*, 12(4), pp. 477–486. Available at: <https://doi.org/10.14710/j.gauss.12.4.477-486>.
- Santi, V.M. (2018) 'Pengembangan Model Regresi Logistik Multinomial untuk Klasifikasi Politik pada Pemilihan Umum', *Jurnal Statistika dan Aplikasinya*, 2(1), pp. 37–43. Available at: <https://doi.org/10.21009/JSA.02105>.
- Sihombing, P.R. and Yuliati, I.F. (2021) 'Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia', *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), pp. 417–426. Available at: <https://doi.org/10.30812/matrik.v20i2.1174>.
- Sitanggang, P.B. *et al.* (2024) 'The Relationship between Political Image and Voting Behavior in First-time Voters: Demographics of Generation Z Voters in the 2024 Election in Indonesia', *INTERNATIONAL JOURNAL OF SOCIAL SCIENCE AND EDUCATION RESEARCH STUDIES*, 04(06). Available at: <https://doi.org/10.55677/ijssers/V04I6Y2024-01>.