



Analisis Prediksi Hasil Pemilu Legislatif DPR RI DKI Jakarta Tahun 2024 Menggunakan Metode *Random Forest* dan *Gradient Boosting*

* Rangga Febrian Effendy¹, Agung Budi Susanto², Sajarwo Anggai³

^{1,2,3}Teknik Informatika, Program Pascasarjana Universitas Pamulang, Tangerang Selatan, Banten

Email: ¹ ranggafebrian@gmail.com, ² dosen02680@unpam.ac.id, ³ sajarwo@gmail.com

ABSTRACT

In general elections, it is closely related to predictions, predictions play an important role in obtaining results in future legislative elections. Predicting general election results can be done through a series of processes to find patterns and knowledge from a set of data using data mining techniques. To get accurate prediction results in the future, a method is needed that can be used as predictive modeling. This research aims to find out the results of model testing and predictions for the 2024 DPR RI DKI Jakarta legislative election using random forest and gradient boosting methods and to find out patterns and knowledge from the prediction results themselves. Based on the model testing results, the gradient boosting method has an accuracy value of 95.8%, precision 72.2% and recall 61.9%. Meanwhile, random forest has an accuracy value of 95.4%, precision 63.6% and recall 33.3%. The pattern and knowledge from the prediction results is that the elected legislative candidates on average are in serial numbers 1 and 2, have valid votes starting from 63,529, are male and have a doctoral degree.

Keywords: Election; Legislative; Prediction; Orange; Data Mining.

ABSTRAK

Dalam pemilihan umum, erat kaitannya dengan prediksi, prediksi memainkan peran penting dalam mendapatkan hasil perolehan pemilu legislatif di masa depan. Prediksi hasil pemilihan umum dapat dilakukan melalui serangkaian proses untuk menemukan pola dan pengetahuan dari sekumpulan data menggunakan teknik *data mining*. Untuk mendapatkan hasil prediksi yang akurat di masa depan, diperlukan suatu metode yang dapat dijadikan sebagai pemodelan prediksi. Penelitian ini bertujuan untuk mengetahui hasil pengujian model dan prediksi pemilu legislatif DPR RI DKI Jakarta tahun 2024 menggunakan metode *random forest* dan *gradient boosting* serta mengetahui pola dan pengetahuan dari hasil prediksi itu sendiri. Berdasarkan hasil pengujian model, metode *random forest* memiliki nilai *accuracy* 95.2%, *precision* 80% dan *recall* 38.1%. Sedangkan *gradient boosting* memiliki nilai *accuracy* 95.8%, *precision* 72.2% dan *recall* 61.9%. Adapun pola dan pengetahuan dari hasil prediksi yaitu calon anggota legislatif yang terpilih rata-rata berada pada nomor urut 1 dan 2, memiliki suara sah mulai dari 63.529, berstatus petahana dan berpendidikan S3.

Kata kunci: Pemilu; Legislatif; Prediksi; Orange; Data Mining.

1. PENDAHULUAN

Penelitian mengenai prediksi hasil pemilu telah dilakukan oleh beberapa peneliti, diantaranya oleh Gunawan Budi Sulistyو melakukan penelitian penerapan *decision tree C4.5* berbasis *particle swarm optimization* untuk memprediksi hasil Pileg DPRD Provinsi

Sulawesi Selatan dengan hasil akurasi 93,70% [1]. Kemudian penelitian yang dilakukan oleh Achmad Syaiful Rizal dan Moch. Lutfi mengenai prediksi hasil Pileg Kabupaten Sidoarjo yang menggunakan metode *k-nearest neighbor* digabungkan dengan metode *backward elimination* menghasilkan akurasi sebesar 96,03% [2]. Berdasarkan penelitian sebelumnya, didapatkan hasil prediksi beserta akurasi, namun belum diketahui mengenai pola dan pengetahuan dari prediksi itu sendiri, sehingga dalam penelitian ini dimungkinkan untuk didapatkan pola dan pengetahuan dari prediksi.

Kemudian penelitian mengenai prediksi lainnya dilakukan oleh Dede Husen mengenai prediksi kebakaran hutan menggunakan metode *random forest* menghasilkan tingkat akurasi 100%. Kemudian penelitian yang dilakukan oleh Kartina Diah Kusuma Wardhani mengenai prediksi resiko diabetes menggunakan metode *extreme gradient boosting* memiliki nilai akurasi 98,7% [3].

Berdasarkan penelitian sebelumnya, kedua metode *random forest* dan *gradient boosting* memiliki tingkat akurasi yang tinggi pada objek penelitian prediksi kebakaran hutan dan prediksi resiko diabetes. Dengan kedua metode tersebut dapat dimungkinkan dipergunakan sebagai metode dalam penelitian ini untuk mendapatkan hasil prediksi dengan tingkat akurasi yang tinggi.

2. METODE

Penelitian ini menggunakan teknik *supervised learning* yaitu teknik klasifikasi menggunakan data sekunder terdiri dari data Pileg 2019 sebanyak 311 *record* dan data Pileg 2024 sebanyak 373 *record*. Penelitian bertujuan untuk memperkirakan hasil atau kejadian di masa depan berdasarkan analisis data historis atau hubungan antar variabel.

2.1. Analisis Kebutuhan Data

Untuk kebutuhan data penelitian akan digunakan dataset yang disusun dari kumpulan data, sehingga dataset tersebut dapat diolah oleh aplikasi *Orange* untuk dilakukan analisis. Dataset tersebut akan disajikan dalam bentuk Tabel 1.

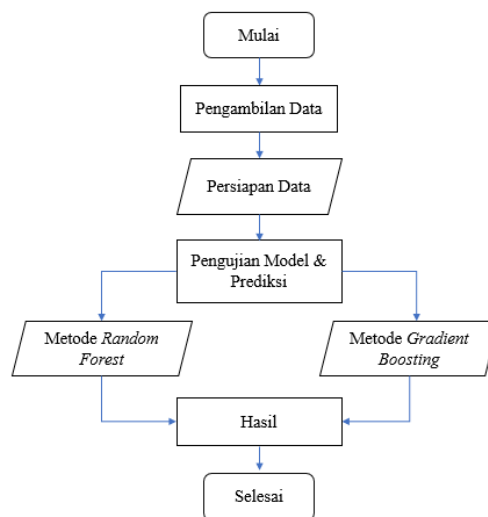
Tabel 1. Kebutuhan Data

No	Nama Atribut	Tipe	Role	Value
1.	NAMA DAPIL	Kategori	Fitur	DKI Jakarta I, DKI Jakarta II, DKI Jakarta III
2.	NAMA PARTAI	Kategori	Fitur	PKB, Gerindra, PDIP, Golkar, Nasdem, Buruh, Gelora, PKS, PKN, Hanura, Garda, PAN,

No	Nama Atribut	Tipe	Role	Value
				PBB, Demokrat, PSI, Perindo, PPP, Ummat
3.	NAMA CALEG	Text	Meta	
4.	NOMOR URUT	Numerik	Fitur	1, 2, 3, 4, 5, 6, 7, 8
5.	JENIS KELAMIN	Kategori	Fitur	L, P
6.	PENDIDIKAN	Kategori	Fitur	-, S1, S2, S3
7.	SUARA SAH CALON	Numerik	Fitur	
8.	SUARA SAH PARTAI	Numerik	Fitur	
9.	TOTAL SUARA	Numerik	Fitur	
10.	PAR. TRESHOLD	Kategori	Fitur	LOLOS, TIDAK LOLOS
11.	STATUS PETAHANA	Kategori	Fitur	PETAHANA, BUKAN
12.	TERPILIH / TIDAK	Kategori	Target	YA, TIDAK

2.2. Perancangan Penelitian

Tahapan ini adalah tahap dimana dibuatkan suatu rancangan untuk penelitian mulai dari pengambilan data sampai selesai. Secara umum pada penelitian ini dilakukan dengan mengikuti tahapan seperti pada Gambar 1.



Gambar 1. Perancangan Penelitian

Pada tahap pertama yaitu pengambilan data, sumber data dari *website kpu.go.id*. data yang dipilih adalah data sekunder. Data sekunder adalah data yang sudah tersedia yang digunakan oleh peneliti guna kepentingan penelitiannya. Data yang dipilih bersifat umum atau publik dimana semua dapat mengakses data tersebut.

Setelah data – data tersebut diperoleh, dilakukan persiapan data yang pertama pembersihan data dengan cara menghilangkan gelar dan jabatan pada nama Caleg, selanjutnya data yang sudah dilakukan pembersihan dikumpulkan dan disusun dengan menghilangkan atribut yang tidak relevan sehingga menjadi dataset yang dapat diproses oleh aplikasi *Orange*.

Pada tahapan pengujian model dilakukan terhadap metode yang diusulkan yaitu metode *random forest* dan *gradient boosting*, tahapan pemilihan metode dipertimbangkan untuk mendapatkan metode yang tepat dan akurat. Tahap ini merupakan tahap pembentukan model klasifikasi *supervised learning*. Proses ini menggunakan aplikasi *Orange* dengan menggunakan *widget test and score*, *confusion matrix*, kurva ROC dan hasil prediksi. Kemudian tahap hasil yang terakhir yaitu hasil dari pengujian model, hasil dari prediksi beserta perbandingan hasil akurasi dan yang terakhir pola dan pengetahuan dari hasil prediksi.

2.3. Teknik Data Mining

2.3.1. Random Forest

Metode *random forest* merupakan suatu metode *ensemble* yang berarti meningkatkan akurasi metode klasifikasi dengan cara mengombinasikan metode klasifikasi. Proses *random forest* diawali dengan teknik dasar *data mining* yaitu *decision tree*. Selanjutnya, pada *decision tree* input dimasukkan pada bagian atas (*root*) sebelum turun ke bagian bawah (*leaf*) untuk menentukan kelas pada data tersebut. *Random forest* juga dapat dikatakan sebagai pengklasifikasian yang terdiri dari kumpulan pengklasifikasi *decision tree* di mana masing-masing *tree* memberikan unit suara untuk kelas paling populer pada input x . Jadi, bisa dikatakan *random forest* terdiri dari sekumpulan *decision tree* dimana kumpulan *decision tree* tersebut digunakan untuk mengklasifikasikan data ke dalam suatu kelas [4].

2.3.2. Gradient Boosting

Ensemble learning algorithm menjadi salah satu metode yang digunakan dalam *machine learning* dengan menggunakan banyak model *machine learning* yang bekerja bersama untuk menghasilkan prediksi yang tepat. *Gradient boosting* merupakan algoritma klasifikasi *machine learning* yang menggunakan *ensemble* dari *decision tree* untuk memprediksi nilai. *Gradient boosting* termasuk *supervised learning* berbasis *decision tree* yang dapat digunakan untuk klasifikasi. *Gradient boosting* dimulai dengan menghasilkan pohon klasifikasi awal dan terus menyesuaikan pohon baru melalui minimalisasi fungsi kerugian [5].

2.4. Pengujian Algoritma

2.4.1. Confusion Matrix

Confusion matrix adalah tabel untuk menggambarkan kinerja sampel klasifikasi yang digunakan untuk mengukur kinerja pada kumpulan data. *Confusion matrix* memberikan informasi mengenai klasifikasi aktual dan prediksi yang telah dibuat oleh kinerja sistem sehingga hasil dari suatu sistem sering dievaluasi menggunakan tes *confusion matrix* [6]. Dengan metode *confusion matrix* dapat dilakukan perhitungan dan menghasilkan nilai *accuracy*, *recall* dan *precision* [7].

- a. *Accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Dengan rumus sebagai berikut:

$$Accuracy = (TP + TN) / (TP+FP+FN+TN)$$

- b. *Recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Dengan rumus sebagai berikut:

$$Recall = (TP) / (TP + FN)$$

- c. *Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dengan rumus sebagai berikut:

$$Precision = (TP) / (TP+FP)$$

2.4.2. Kurva ROC

Untuk dapat melihat akurasi secara manual dilakukan perbandingan klasifikasi menggunakan kurva ROC hasil eksperisi dari *confusion matrix*. Didalam kurva ROC terdapat AUC (*Area Under the Curve*). AUC adalah luas area di bawah kurva ROC atau integral dari fungsi ROC. AUC memiliki nilai antara 0,0 dan 1,0 dengan tingkat keakuratan klasifikasi sebagai berikut [8]:

- a. 0,90 sd 1,00 = sangat baik
- b. 0,80 sd 0,90 = baik
- c. 0,70 sd 0,80 = sama
- d. 0,60 sd 0,70 = rendah
- e. 0,50 sd 0,60 = gagal

2.5. Knowledge Presentation

Dalam proses *data mining*, terdapat tahapan yang disebut *Knowledge Presentation*, merupakan tahap terakhir dimana pengetahuan yang telah ditemukan secara

visual dapat ditampilkan. Tahap penting ini menggunakan teknik visualisasi untuk membantu pengetahuan dan menginterpretasikan hasil dari *data mining* [9]. Dalam penelitian ini visualisasi data mining menggunakan *widget* bernama *box plot*. *Box plot* adalah sebuah *widget* pada software *Orange* yang menunjukkan distribusi nilai atribut. Merupakan *widget* visualisasi yang berfungsi untuk memeriksa anomali dengan variabel yang dapat dipilih berdasarkan kategori [10].

3. HASIL DAN PEMBAHASAN

3.1. Pengambilan Data

Data yang digunakan dalam penelitian terdiri dari data Pileg DPR RI Dapil DKI Jakarta I, II dan III tahun 2019 dan 2024, data tersebut diambil dari website KPU yang berisi data rekapitulasi hasil penghitungan perolehan suara Caleg DPR RI Dapil DKI Jakarta I, II dan III. Kemudian digunakan data tambahan dari hasil publikasi Pileg tahun 2019 dan 2024, yaitu data lolos dan tidak lolos *parliamentary treshold* dan data jenis kelamin. Agar mendapatkan hasil prediksi yang akurat maka diperlukan data tambahan yaitu pendidikan terakhir yang didapatkan dari gelar pada nama Caleg.

3.2. Persiapan Data

Setelah data – data tersebut diperoleh, dilakukan pembersihan data dengan cara menghilangkan gelar dan jabatan pada nama Caleg, selanjutnya data yang sudah dilakukan pembersihan dikumpulkan dan disusun dengan menghilangkan atribut yang tidak relevan sehingga menjadi dataset yang dapat diproses oleh aplikasi *Orange*.

Tabel 2. Dataset Pileg 2019

Nama Dapil	Nama Partai	No. Urut	Nama Caleg	JK	Pend.	Suara Caleg	Suara Partai	Total Suara	PT	Terpilih / Tidak
DKI Jakarta I	PKB	1	IMAM NAHRAWI	L	S1	29.909	30.322	99.114	Lolos	Tidak
DKI Jakarta I	PKB	2	YUSUF MUJENIH	L	S1	5.503	30.322	99.114	Lolos	Tidak
...

(Sumber: KPU, 2024)

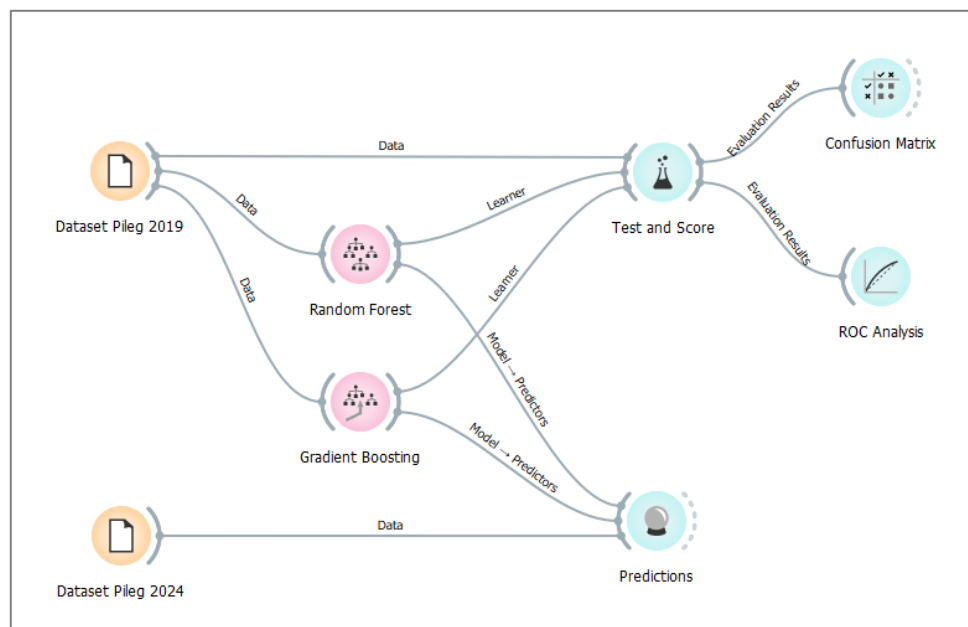
Tabel 3. Dataset Pileg 2024

Nama Dapil	Nama Partai	No. Urut	Nama Caleg	JK	Pend.	Suara Caleg	Suara Partai	Total Suara	PT
DKI Jakarta I	PKB	1	HASBIALLAH ILYAS	L	-	23.598	12.863	42.929	Lolos
DKI Jakarta I	PKB	2	DARUSSALAM	L	S1	10.599	12.863	42.929	Lolos
...

(Sumber: KPU, 2024)

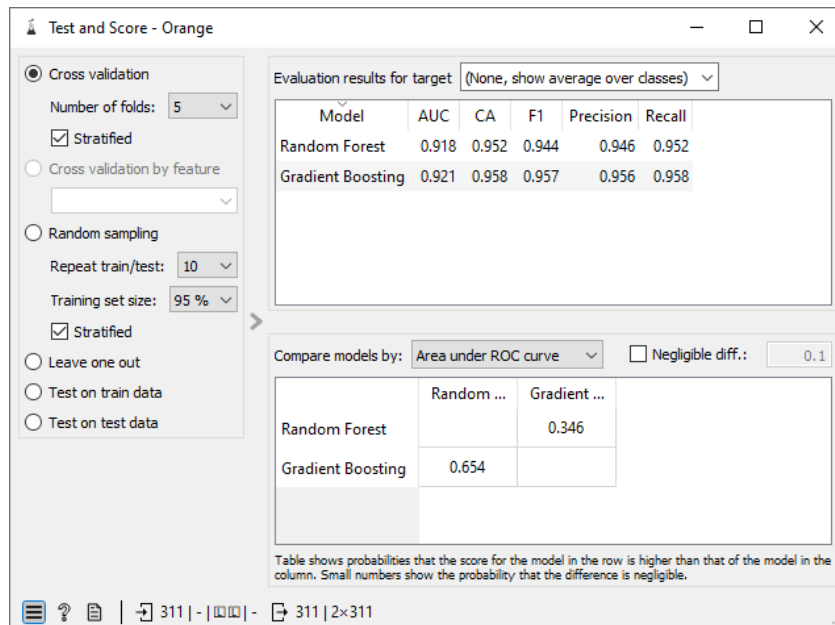
3.3. Pengujian Model dan Prediksi

Data yang telah melalui tahap persiapan data selanjutnya dilakukan pengujian untuk mendapatkan model prediksi terbaik. Pada Gambar 2. merupakan *workflow* proses pengujian model dan prediksi. Proses perhitungan keberhasilan model prediksi menggunakan *widget Test and Score* dan hasil prediksinya dapat dilihat menggunakan *widget Predictions*.

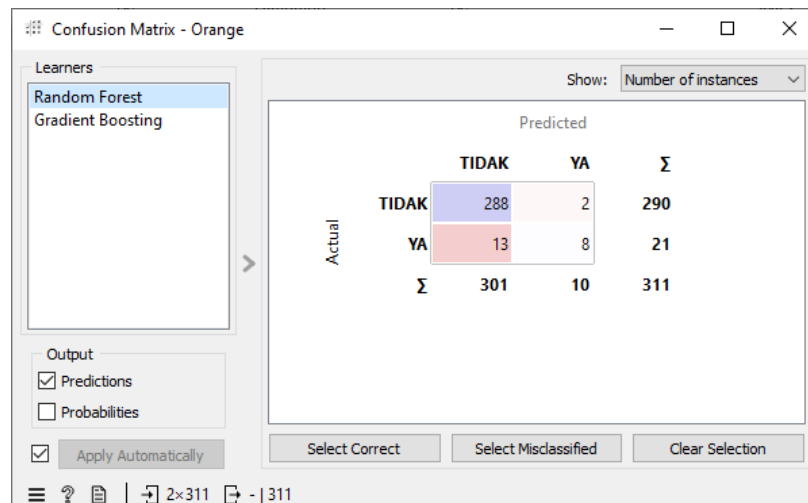


Gambar 2. *Workflow* Pengujian Model dan Prediksi

Pada Gambar 3. merupakan tampilan dari *widget Test and Score* aplikasi *Orange*. Pada penelitian ini proses pengujian menerapkan *K-Fold Cross Validation* (K=5) yang dapat diatur pada *widget Test and Score*.



Gambar 3. Hasil *Test and Score*



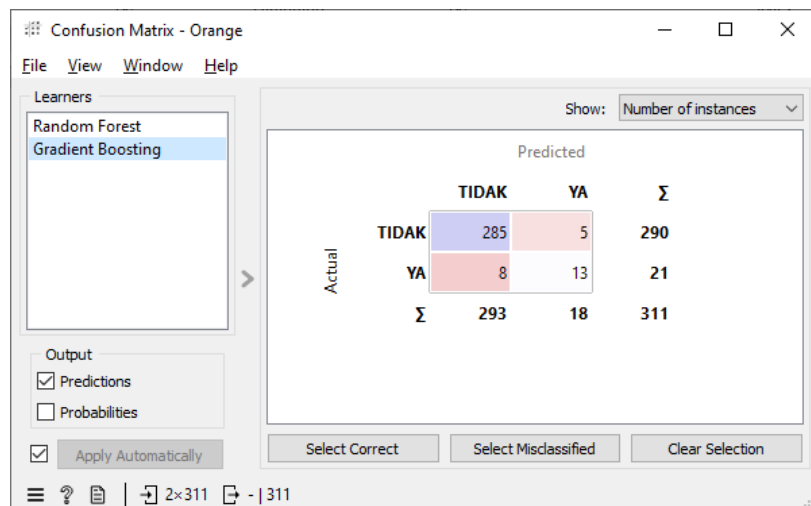
Gambar 4. *Confusion Matrix Random Forest*

Pada Gambar 4. menunjukkan bahwa nilai TP = 8, TN = 288, FP = 2 dan FN = 13. Nilai metode *random forest* untuk *accuracy*, *precision* dan *recall* berdasarkan nilai *confusion matrix* adalah sebagai berikut:

$$Accuracy = \frac{6+288}{6+288+2+13} \times 100\% = 95,2\%$$

$$Precision = \frac{8}{8+2} \times 100\% = 80\%$$

$$Recall = \frac{8}{8+13} \times 100\% = 38,1\%$$



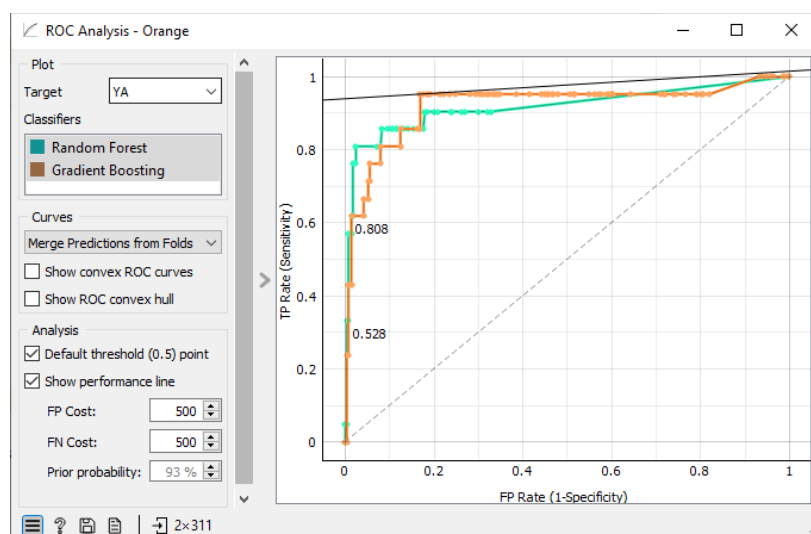
Gambar 5. Confusion Matrix Gradient Boosting

Pada gambar 5. menunjukkan bahwa nilai TP = 13, TN = 285, FP = 5 dan FN = 8. Nilai metode *gradient boosting* untuk *accuracy*, *precision* dan *recall* berdasarkan nilai *confusion matrix* adalah sebagai berikut:

$$Accuracy = \frac{13+285}{13+285+5+8} \times 100\% = 95,8\%$$

$$Precision = \frac{13}{13+5} \times 100\% = 72,2\%$$

$$Recall = \frac{13}{13+8} \times 100\% = 61,9\%$$



Gambar 6. Kurva ROC

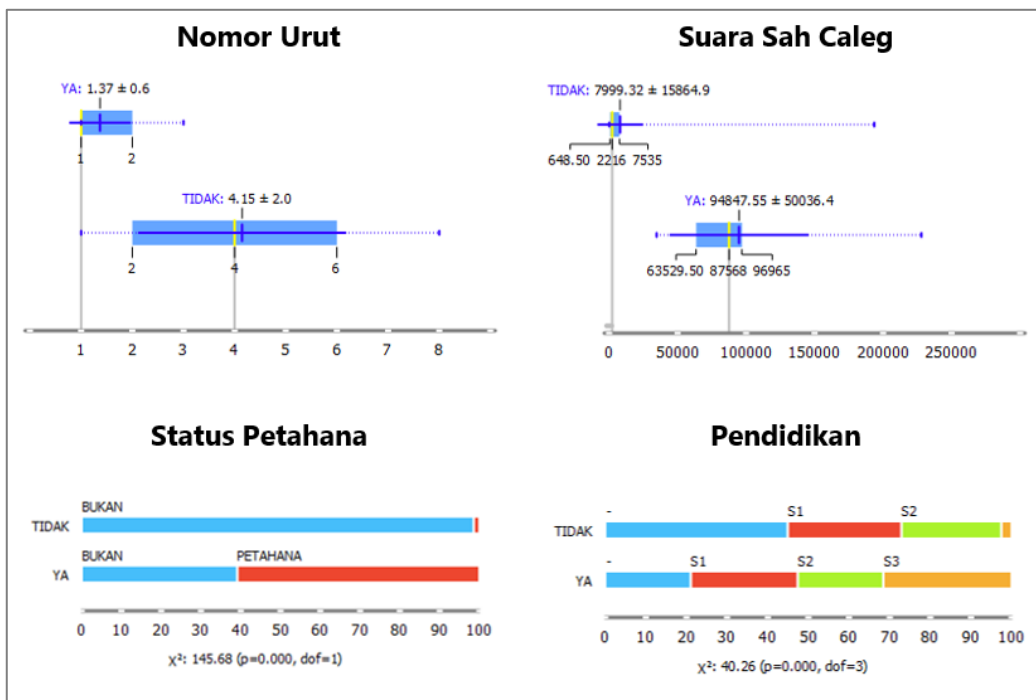
Pada Gambar 6. menunjukkan hasil evaluasi model dari kurva ROC metode *random forest* sebesar 0.528 dan *gradient boosting* sebesar 0.808.

Show probabilities for (None)		Predictions - Orange											
Random Forest	Gradient Boosting	NAMA CALEG	NAMA DAPIL	NAMA PARTAI	NOMOR URUT	JENIS KELAMIN	PEND.	ARA SAH CAI	ARA SAH PAR	TOTAL SUARA	R. TRESHC	PETAHANA	
1	YA	MEILANI LEIMENA SUHARLI	DKI JAKARTA II	Demokrat	1	P	-	61046	40331	148440	LOLOS	YA	
2	YA	HASBIALLAH ILVAS	DKI JAKARTA I	PKB	1	L	-	80895	41515	145309	LOLOS	TIDAK	
3	YA	ANIS BYARWATI	DKI JAKARTA I	PKS	2	P	S3	64304	97256	390441	LOLOS	YA	
4	YA	KURBIASIH MUFIDAYATI	DKI JAKARTA II	PKS	2	P	S3	62755	117876	483058	LOLOS	YA	
5	YA	HIMMATUL ALYAH	DKI JAKARTA II	Gerindra	1	P	S2	94543	139787	329827	LOLOS	YA	
6	YA	AHMAD SAHRONI	DKI JAKARTA III	Nasdem	1	L	S2	163292	38295	239287	LOLOS	YA	
7	YA	ERWIN AKSA	DKI JAKARTA III	Golkar	2	L	-	186897	38256	313619	LOLOS	TIDAK	
8	YA	CHARLES HONORIS	DKI JAKARTA III	PDIP	2	L	-	97016	106386	386980	LOLOS	YA	
9	YA	EKO HENDRO PURNOMO	DKI JAKARTA I	PAN	1	L	S1	93673	24364	149427	LOLOS	YA	
10	YA	ADANG DARADJATUN	DKI JAKARTA III	PKS	1	L	S1	95773	84988	322200	LOLOS	YA	
11	YA	PUTRA NABABAN	DKI JAKARTA I	PDIP	1	L	-	105559	54010	288461	LOLOS	YA	
12	YA	DARMADI DURIANTO	DKI JAKARTA III	PDIP	1	L	S3	95553	106386	386980	LOLOS	YA	
13	YA	HIDAYAT NUR WAHID	DKI JAKARTA II	PKS	1	L	S3	227974	117876	483058	LOLOS	YA	
14	YA	MARDANI	DKI JAKARTA I	PKS	1	L	S3	176584	97256	390441	LOLOS	YA	
15	YA	HABIBUROKHMAN	DKI JAKARTA I	Gerindra	1	L	S3	96914	64805	196708	LOLOS	YA	
16	TIDAK	SONDANG TAMPUBOLON	DKI JAKARTA I	PDIP	2	P	S2	68370	54010	288461	LOLOS	YA	
17	TIDAK	LULA KAMAL	DKI JAKARTA II	PAN	4	P	S1	36667	35765	207623	LOLOS	TIDAK	
18	TIDAK	NURWAYAH	DKI JAKARTA II	Demokrat	3	P	S1	34661	27983	133307	LOLOS	TIDAK	
19	TIDAK	IDA FAUZİYAH	DKI JAKARTA II	PKB	1	P	S2	76092	61336	202696	LOLOS	TIDAK	
20	TIDAK	MUHAMMAD IQBAL	DKI JAKARTA II	PKS	4	L	S3	38806	117876	483058	LOLOS	TIDAK	
21	TIDAK	ADI KURNIA SETIADI	DKI JAKARTA I	Demokrat	1	L	S3	39234	32029	99383	LOLOS	TIDAK	
22	TIDAK	UYA KUYA	DKI JAKARTA II	PAN	2	L	-	81463	35765	207623	LOLOS	TIDAK	
23	TIDAK	ABRAHAM SRIDJAJA	DKI JAKARTA II	Golkar	3	L	S2	76945	58839	259126	LOLOS	TIDAK	
24	YA	SANTOSO	DKI JAKARTA III	Demokrat	1	L	S2	29899	27983	133307	LOLOS	YA	
25	YA	ERIKO SOTARDUGA	DKI JAKARTA II	PDIP	1	L	S1	48737	112307	368762	LOLOS	YA	
26	YA	MASINTON PASARIBU	DKI JAKARTA II	PDIP	7	L	-	50992	112307	368762	LOLOS	YA	

Gambar 7. Hasil Prediksi

3.4. Pola dan Pengetahuan Hasil Prediksi

Setelah mendapatkan hasil prediksi, tahapan selanjutnya adalah pencarian informasi berupa pola dan pengetahuan dari masing – masing dataset menggunakan teknik visualisasi *box plot* yang hasilnya dapat dilihat pada Gambar 8.



Gambar 8. Pola dan Pengetahuan Hasil Prediksi

Pada Gambar 8 menunjukkan grafik keterpilihan berdasarkan nomor urut. Caleg yang terpilih rata-rata berada pada nomor urut 1 dan 2. Grafik keterpilihan berdasarkan

suara sah Caleg. Caleg yang terpilih rata-rata memiliki suara sah mulai dari 63.529. Grafik keterpilihan berdasarkan status petahana. Caleg yang terpilih rata-rata berstatus petahana sebanyak 60%. Grafik keterpilihan berdasarkan pendidikan. Caleg yang terpilih rata – rata berpendidikan S3.

4. KESIMPULAN

Hasil pengujian model dan prediksi menggunakan metode *random forest* dan *gradient boosting* pada Pileg DPR RI DKI Jakarta tahun 2024 yaitu *gradient boosting* memiliki hasil yang lebih baik dibandingkan dengan *random forest* berdasarkan pengujian *test and score*, *confusion matrix* dan kurva ROC. Dengan perbandingan *gradient boosting* memiliki nilai *accuracy* 95.8%, *precision* 72.2% dan *recall* 61.9%. Sedangkan *random forest* memiliki nilai *accuracy* 95.4%, *precision* 63.6% dan *recall* 33.3%. Adapun pola dan pengetahuan mengenai hasil prediksi pada Pileg DPR RI DKI Jakarta tahun 2024 yaitu Caleg yang terpilih rata-rata berada pada nomor urut 1 dan 2. Caleg yang terpilih rata-rata memiliki suara sah mulai dari 63.529. Caleg yang terpilih rata-rata berstatus petahana sebanyak 60%. Caleg yang terpilih rata-rata berpendidikan S3.

5. DAFTAR PUSTAKA

- [1] G. B. Sulistyono, “Penerapan Decision Tree C.45 Berbasis Particle Swarm Optimization untuk Prediksi Hasil Pemilu Legislatif,” *J. Speed-Sentra Penelit. Eng. dan Edukasi*, vol. 12, no. 1, pp. 21–26, 2020, [Online]. Available: www.rapidminer.com.
- [2] Achmad Saiful Rizal and Moch. Lutfi, “Prediksi Hasil Pemilu Legislatif Menggunakan Algoritma K-Nearest Neighbor Berbasis Backward Elimination,” *J. Resist. (Rekayasa Sist. Komputer)*, vol. 3, no. 1, pp. 27–41, 2020, doi: 10.31598/jurnalresistor.v3i1.517.
- [3] H. Priyono, R. Sari, and T. Mardiana, “Klasifikasi Pemilihan Jurusan Sekolah Menengah Kejuruan Menggunakan Gradient Boosting Classifier,” *J. Inform.*, vol. 9, no. 2, pp. 131–139, 2022, doi: 10.31294/inf.v9i2.12654.
- [4] I Made Pasek Windu Merta, “Mengenal Metode Random Forest dalam Machine Learning,” *Website*, 2022.
- [5] A. R. Febryananda, “Mengenal Teknik Gradient Boosting dalam Algoritma

- Machine Learning,” 2022, [Online]. Available: https://lab_adrk.ub.ac.id/id/mengenal-teknik-gradient-boosting-dalam-algoritma-machine-learning/
- [6] D. Safitri, S. S. Hilabi, and F. Nurapriani, “Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining,” *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 75–81, 2023, doi: 10.36341/rabit.v8i1.3009.
- [7] S. Adi and A. Wintarti, “Komparasi Metode Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Dan Random Forest (RF) Untuk Prediksi Penyakit Gagal Jantung,” vol. 10, no. 02, pp. 258–268, 2022.
- [8] G. D. M. Zulma, Angelika, and N. Chamidah, “Perbandingan Metode Klasifikasi Naive Bayes, Decision Tree Dan K-Nearest Neighbor Pada Data Log Firewall,” *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. April, pp. 679–688, 2021.
- [9] Mustika *et al.*, *Data Mining dan Aplikasinya*. Bandung: penerbitwidina, 2021.
- [10] D. W. Hoffman, “Box Plot Shows distribution of attribute values.”, [Online].