



## Perbandingan Akurasi *Random Forest Classifier* dalam Memprediksi Kelayakan Kendaraan berdasarkan Jumlah *Decision tree* dan *Selection Feature*

Laily Nur Qomariyati<sup>1</sup>, Sifa Nurpadillah<sup>2</sup>, Nurya Fahu Rosyidin<sup>3</sup>, Fauzi Sofyan<sup>4</sup>, Luthfi Rizqi Mubarak<sup>5</sup>

<sup>1,3,4,5</sup> Faculty of Electrical Engineering, Telkom University, Bandung, Indonesia

<sup>2</sup> Fakultas Teknik, Universitas Garut, Garut, Indonesia

Korespondensi: [lailynrq@student.telkomuniversity.ac.id](mailto:lailynrq@student.telkomuniversity.ac.id)

### ARTICLE HISTORY

Received:11-06-2024

Revised:22-6-2024

Accepointed:28-6-2024

### Abstrak

Pasca pandemi, terdapat peningkatan permintaan dan kebutuhan kendaraan yang ditandai dengan peningkatan penjualan. Konsumen memiliki banyak pertimbangan dalam membeli kendaraan. Produsen harus memahami kebutuhan dan pertimbangan konsumen agar kendaraan yang diluncurkan dapat diterima dengan baik. Penelitian ini bertujuan memprediksi kelayakan kendaraan menggunakan *Random Forest Classifier* berdasarkan jumlah *decision tree* dan pemilihan fitur. Model *Random Forest* diterapkan dengan variasi jumlah *decision tree* untuk mengidentifikasi pengaruhnya terhadap akurasi. Hasil menunjukkan bahwa akurasi meningkat signifikan dengan penambahan jumlah *decision tree* dan mencapai stabil pada sekitar 50-100 pohon dengan akurasi maksimum 94.57%. Selanjutnya, penelitian mengevaluasi pengaruh jumlah fitur input terhadap akurasi. Hasil menunjukkan peningkatan akurasi dari 69.53% dengan satu fitur hingga 94.57% dengan semua fitur. Fitur-fitur utama seperti *Safety*, *Persons*, *Buying*, *Maint*, dan *Lug\_boot* terbukti sangat penting, dengan penggunaan lima fitur ini menghasilkan akurasi 92.64%. Analisis *importance score* menunjukkan bahwa menghilangkan fitur dengan *importance* tinggi, seperti *Safety*, menurunkan akurasi secara signifikan. Kesimpulan penelitian ini adalah bahwa pemilihan jumlah *decision tree* yang optimal dan fitur yang tepat sangat penting untuk meningkatkan akurasi model *Random Forest*. Penggunaan sekitar 50-100 *decision tree* dan pemilihan fitur berdasarkan *importance score* dapat memberikan prediksi yang akurat dan efisien. Penelitian ini memberikan panduan praktis bagi produsen kendaraan dalam memanfaatkan data *mining* dan *machine learning* untuk memahami preferensi konsumen dan meningkatkan efektivitas produksi.

**Kata kunci:** *Random Forest*, *Decision tree*, Feature Selection, Data Mining, Machine Learning, Klasifikasi Kendaraan

## Comparison of the Accuracy of *Random Forest Classifier* in Predicting Vehicle Eligibility Based on the Number of *Decision tree* and Feature Selection

### **Abstract**

Post-pandemic, there has been an increase in vehicle demand and usage, indicated by rising sales. Consumers have many considerations when purchasing vehicles. Manufacturers must understand consumer needs and considerations to ensure their vehicles are well-received. This study aims to predict vehicle feasibility using a Random Forest Classifier based on the number of decision trees and feature selection. The Random Forest model was applied with varying numbers of decision trees to identify their impact on accuracy. Results showed that accuracy significantly increased with more decision trees and stabilized around 50-100 trees, reaching a maximum accuracy of 94.57%. The study also evaluated the impact of the number of input features on accuracy. Results indicated an accuracy increase from 69.53% with one feature to 94.57% with all features. Key features such as Safety, Persons, Buying, Maint, and Lug\_boot proved to be highly important, with the use of these five features yielding an accuracy of 92.64%. Feature importance analysis revealed that removing high-importance features, like Safety, significantly reduced accuracy. The conclusion of this study is that selecting the optimal number of decision trees and the right features is crucial for improving the Random Forest model's accuracy. Using approximately 50-100 decision trees and selecting features based on their importance score can provide accurate and efficient predictions. This research offers practical guidance for vehicle manufacturers in leveraging data mining and machine learning to understand consumer preferences and enhance production effectiveness.

**Key words:** Random Forest, Decision tree, Feature Selection, Data Mining, Machine Learning, Vehicle Classification

## **1. Pendahuluan**

Manusia tidak dapat terlepas dari kehadiran transportasi dalam kehidupannya. Kemudian perkembangan penjualan kendaraan mengalami peningkatan sejak tahun 2020 baik itu penjualan retail maupun dealer resmi [1]. Hal tersebut menunjukkan, konsumsi masyarakat akan kendaraan sudah membaik seiring pemulihan ekonomi di Tanah Air. Banyak dari mereka yang mulai memberi kendaraan baru, untuk menunjang kebutuhan aktivitas harian [2]. Para konsumen mempertimbangkan banyak hal dalam pemilihan alat transportasi yang sesuai dengan kebutuhannya [3]. Berdasarkan kenyamanannya, konsumen yang lebih memilih kendaraan roda empat. Banyak hal yang akan dipertimbangkan konsumen dalam memilih mobil. Dengan demikian, untuk dapat bertahan dalam bisnisnya, produsen pun harus mempertimbangkan kembali tingkat penerimaan mobil oleh para konsumen. Sehingga produksi mobil akan lebih efektif, kerugian yang minimal dan profit yang meningkat. Bagaimana produsen dapat menentukan keinginan konsumen dalam memilih mobil yang diinginkannya? Data penerimaan mobil oleh konsumen yang telah ada sebelumnya dapat digunakan untuk memprediksi keinginan konsumen akan penggunaan mobil.

Prediksi penerimaan mobil oleh konsumen pada produksi mobil berikutnya dapat dibantu dengan pengaplikasian *data mining* dan *Artificial Intelligence*. *Artificial Intelligence* akan membantu manusia dalam menentukan keputusan sesuai kebutuhannya. *Data Mining* adalah salah satu bidang dalam *Artificial Intelligence* [4]. *Machine learning* digunakan

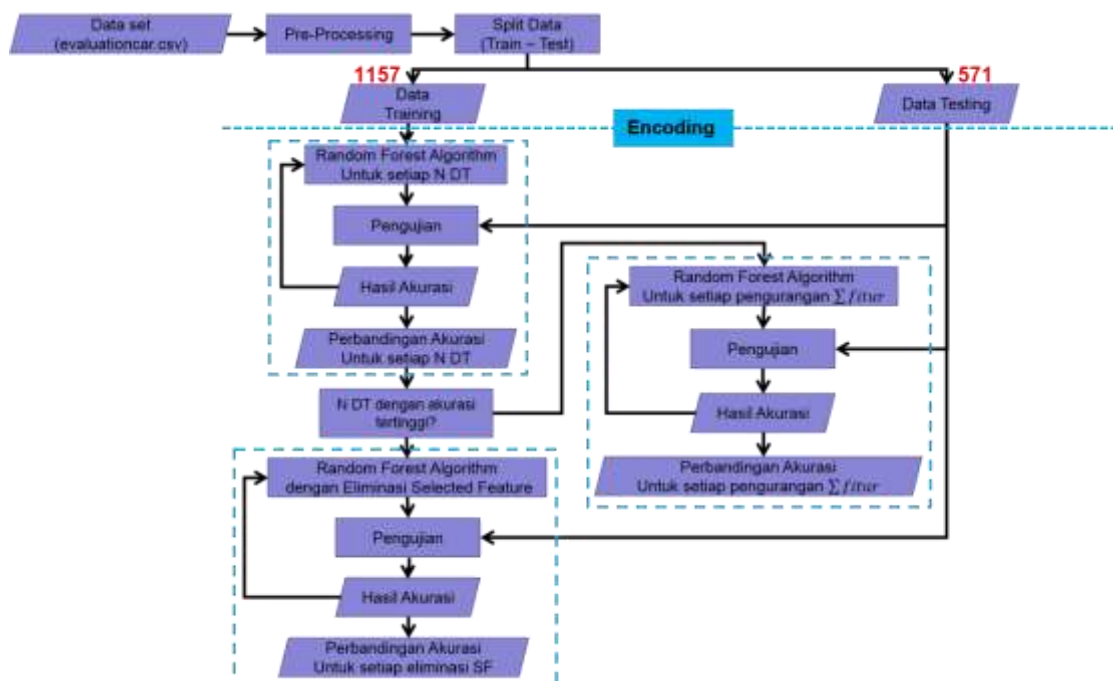
dalam *data mining* untuk menentukan model atau pola dari data yang akan digunakan. *Machine learning* dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [5]. Model yang dihasilkan dari *data training* akan diuji dengan *data testing* untuk menguji akurasi.

Pada penelitian ini *machine learning* yang digunakan untuk memprediksi kelayakan kendaraan adalah *Random Forest Classifier*. *Random Forest* adalah pengembangan dari *decision tree* [3]. *Random forest classifier* lebih baik dalam melaksanakan tugas prediksi dibanding regresi linier. Hal tersebut dikarenakan *random forest* dengan mudah mengadaptasi non-linieritas pada data [6]. Bahkan, *Random forest* juga memberikan hasil yang lebih baik dibandingkan *support vector regression* [7], dengan algoritma K-Nearest Neighbor juga memberikan hasil lebih baik dalam mengevaluasi *car dataset* [8]. *Random Forest* merupakan salah satu metode atau model *ensemble learning* yang merupakan kombinasi dari beberapa *decision tree* atau pohon yang atribut *root*-nya dipilih secara acak, kemudian untuk hasil klasifikasinya didapat dari pengambilan suara terbanyak (*votting*) dari jumlah pohon yang telah ditetapkan [9]. Pada penelitian ini akan dibandingkan akurasi dari *Random Forest Classifier* yang digunakan dalam memprediksi kelayakan kendaraan berdasarkan jumlah *decision tree* dan *selection feature*-nya. Prediksi dilakukan berdasarkan dataset public dari UCI Repository, <https://github.com/lpfgarcia/ucipp/blob/master/uci/car-evaluation.arff>.

Terdapat beberapa penelitian terdahulu yang membahas klasifikasi menggunakan dataset “*car evaluation*” ini. Pertama, penelitian yang membandingkan 3 model machine learning, *decision tree*, *naïve bayes* dan *artificial neural network(ANN)* untuk mendapatkan model dengan akurasi tertinggi [10]. Penelitian kedua, membahas analisis performa dari model klasifikasi *naïve bayes*, *decision tree* dan *rotation forest* [11]. Ketiga, mencari akurasi terbaik dari model *decision tree* dan *Random Forest* dengan pemilihan *splitting criteria* yang ada dalam parameter *decision tree* dan *Random Forest* [3].

## 2. Metode

Penelitian ini membahas tentang metode klasifikasi pada *data mining* dengan variasi penggunaan jumlah *decision tree* dan *selected feature* pada algoritma *Random Forest*. Pada penelitian ini, dataset digunakan untuk memprediksi kelayakan kendaraan yang nantinya akan bersesuaian dengan keinginan konsumen untuk dapat membeli kendaraan tersebut. Dalam setiap prediksinya dilakukan variasi jumlah *decision tree* yang digunakan. Kemudian akan dilakukan analisa akurasi *Random Forest Classifier* tersebut untuk setiap prediksinya berdasarkan jumlah *decision tree*-nya. Untuk jumlah *decision tree* yang menghasilkan akurasi paling tinggi kemudian dilakukan perbandingan akurasi pada eliminasi jumlah *feature input* untuk mengetahui pengaruh jumlah *feature input* yang digunakan terhadap akurasi. Selain itu, pada kondisi jumlah *feature input* yang sama dilakukan pengujian ulang dengan metode *Selection Feature* untuk mengetahui pengaruh *selection feature* terhadap akurasi. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian perbandingan akurasi *Random Forest Classifier* berdasarkan jumlah *Decision tree* dan *Selected feature*.

Pada pemilihan dataset digunakan dataset public yaitu “evaluationcar.csv”. Dataset tersebut mempunyai 1728 record dan 7 atribut yang terdiri dari 6 fitur input dan 1 label output. Adapun deskripsi dan nilai setiap atribut ditunjukkan pada tabel 1.

Tabel 1. Deskripsi dan nilai dari atribut.

Kategori	Nama	Deskripsi	Nilai
Input (Feature)	<i>Buying</i> <i>Maint.</i> <i>Doors</i> <i>Persons</i> <i>Lug_boot</i> <i>Safety</i>	Harga pembelian Harga perawatan Banyaknya pintu Kapasitas membawa orang Ukuran bagasi Estimasi keamanan mobil	<i>v-high, high, med, low</i> <i>v-high, high, med, low</i> <i>2, 3, 4, 5-more</i> <i>2, 4, more</i> <i>small, med, big</i> <i>low, med, high</i>
Output (Label)	<i>class</i>	Klasifikasi prediksi	<i>Unacc, acc, good, v-good</i>

Langkah kedua adalah *preprocessing*, yakni merubah tipe data menjadi *numeric*. Ketiga, membagi dataset menjadi *data training* dan *data testing*, dengan jumlah *data training* 1157 dan *data testing* 571. Keempat, data train dilatih menggunakan algoritma *Random Forest* untuk setiap variasi jumlah *decision tree* tertentu. Kemudian data akurasi untuk setiap variasinya dibandingkan. Selanjutnya, untuk jumlah *decision tree* yang mempunyai akurasi paling tinggi, akan dilakukan pelatihan kembali menggunakan algoritma *Random Forest*. Akan tetapi, pelatihan dilakukan dengan eliminasi fitur tertentu. Kemudian dilakukan pengujian untuk setiap eliminasi fitur tersebut dan dibandingkan akurasinya.

Dalam konteks penelitian menggunakan *Random Forest Classifier*, akurasi adalah metrik evaluasi yang umum digunakan untuk mengukur seberapa baik model tersebut dalam melakukan prediksi. Akurasi dihitung dengan menggunakan persamaan berikut [12]:

$$Akurasi = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Jumlah Prediksi}} \dots\dots\dots (1)$$

Dalam konteks klasifikasi biner atau multi-kelas, rumus ini dapat dinyatakan sebagai [13], [14]:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2)$$

Dengan;

TP (*True Positives*) adalah jumlah prediksi positif yang benar.

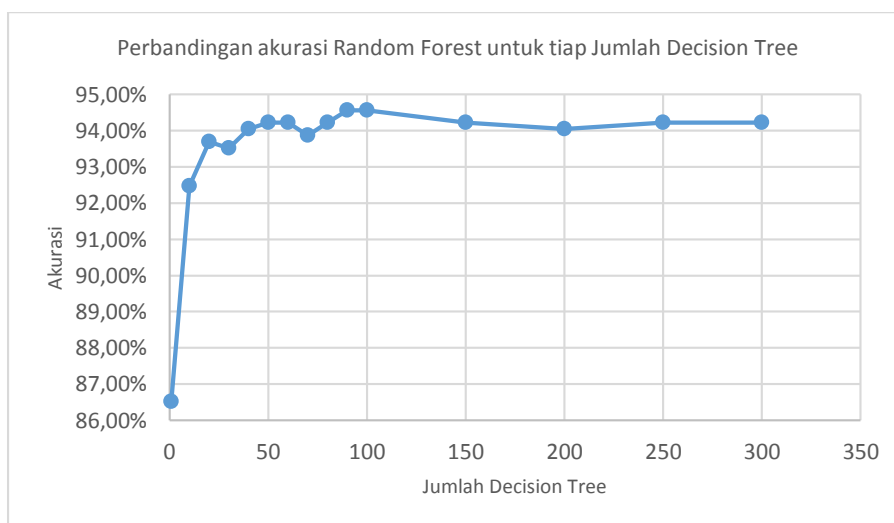
TN (*True Negatives*) adalah jumlah prediksi negatif yang benar.

FP (*False Positives*) adalah jumlah prediksi positif yang salah.

FN (*False Negatives*) adalah jumlah prediksi negatif yang salah.

### 3. Hasil dan Pembahasan

Hasil dari penelitian ditunjukkan pada Gambar 2, yakni perbandingan nilai akurasi untuk setiap pengujian model dengan variasi jumlah *Decision tree* yang digunakan. *Trendline* untuk hasil perbandingannya terlihat cenderung meningkat seiring pertambahan jumlah *decision tree* yang digunakan.



Gambar 2. Perbandingan akurasi *Random Forest* untuk tiap jumlah *Decision tree*.

Berdasarkan Gambar 2 dapat dilihat bahwa Grafik menunjukkan akurasi meningkat tajam dari penggunaan 1 *decision tree* hingga 10 *decision tree*. Setelah itu, peningkatan akurasi mulai melambat, dengan sedikit fluktuasi. Akurasi tertinggi dicapai pada penggunaan 90 dan 100 *decision tree*, yaitu 94.57%. Setelah 100 *decision tree*, akurasi cenderung stabil di sekitar 94.22% hingga 94.57%.

Akurasi meningkat secara signifikan dari 1 *decision tree* (86.51%) ke 10 *decision tree* (92.47%). Hal ini menunjukkan bahwa menambahkan lebih banyak pohon keputusan pada tahap awal secara drastis memperbaiki performa model. Ini karena *Random Forest* bekerja dengan prinsip *majority voting*, di mana lebih banyak pohon memberikan prediksi yang lebih stabil dan akurat. Setelah mencapai 40 *decision tree*, peningkatan akurasi mulai melambat. Akurasi tertinggi dicapai pada 90 dan 100 *decision tree* (94.57%).

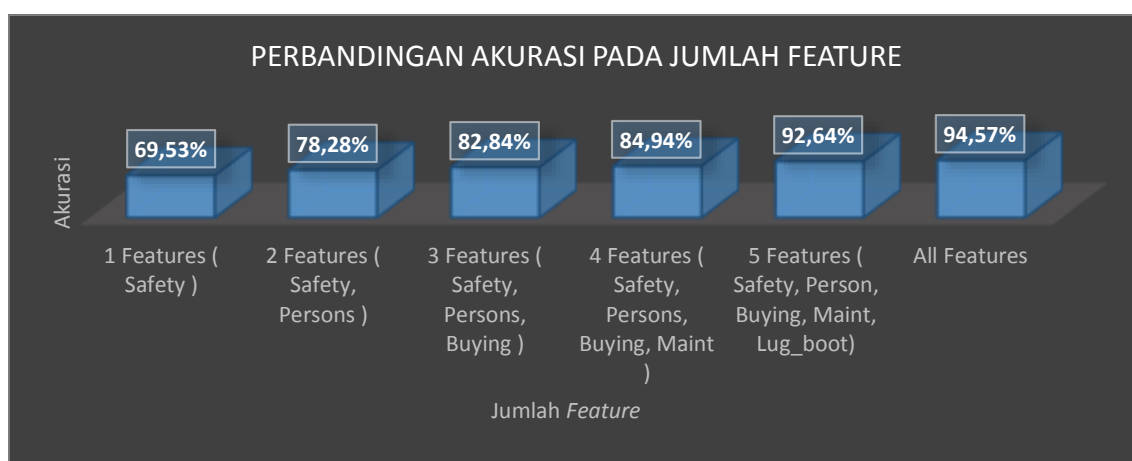


Peningkatan jumlah pohon di atas 100 tidak memberikan peningkatan akurasi yang signifikan dan bahkan cenderung stabil di sekitar 94.22% hingga 94.57%. Hal ini menunjukkan bahwa setelah mencapai titik tertentu, menambahkan lebih banyak pohon tidak memberikan manfaat tambahan yang berarti. Dari hasil penelitian ini, jumlah *decision tree* yang optimal berada di sekitar 90 hingga 100 *decision tree*. Jumlah ini memberikan akurasi tertinggi dan stabil. Penambahan *decision tree* lebih dari jumlah ini tidak memberikan peningkatan akurasi yang signifikan, sehingga dapat dianggap sebagai titik optimal untuk model ini.

Langkah berikutnya dalam penelitian ini adalah menganalisis pengaruh jumlah fitur input dan metode *selection feature* terhadap akurasi. Berdasarkan hasil yang ada, menggunakan *selection feature* bisa membantu dalam mengurangi kompleksitas model tanpa mengurangi akurasi secara signifikan. Pada saat jumlah *decision tree* N=100 tersebut, dilakukan *training* ulang menggunakan algoritma *Random Forest* dengan kondisi:

#### a. Pengurangan jumlah *feature input*

Penelitian ini melanjutkan analisis dengan menguji pengaruh pengurangan jumlah fitur input terhadap akurasi model *Random Forest*, di mana jumlah *decision tree* telah ditetapkan sebanyak 100 (N=100). Hasil *training* memberikan akurasi untuk tiap pengurangan jumlah *feature input* seperti ditunjukkan pada Gambar 3.



Gambar 3. Perbandingan akurasi *Random Forest Classifier* untuk setiap jumlah *feature*.  
(Note: Pengurangan jumlah fitur input ini dilakukan pada N = 100.)

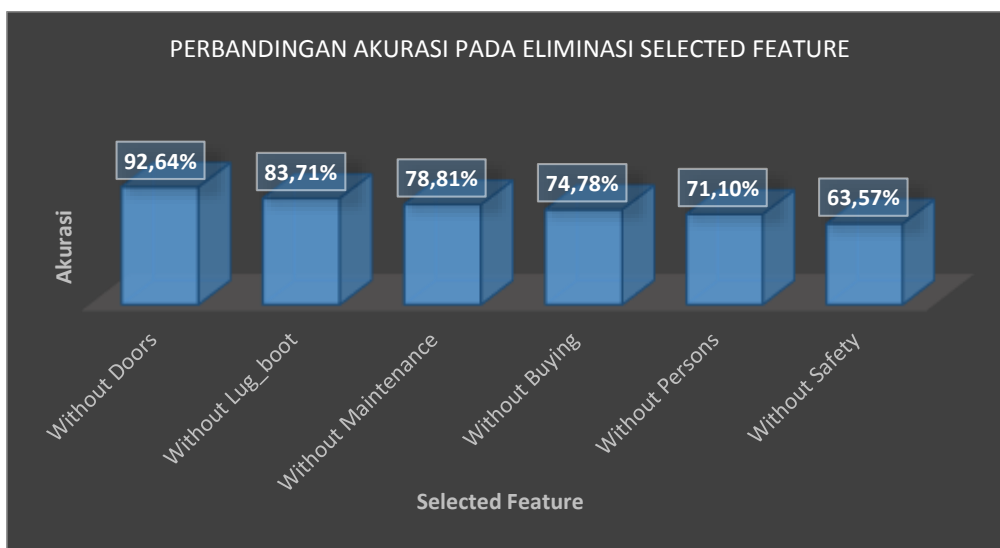
Berdasarkan Gambar 3, dapat dilihat bahwa terdapat peningkatan akurasi seiring dengan bertambahnya jumlah fitur yang digunakan dalam model. Akurasi meningkat secara bertahap dari 69.53% dengan hanya satu fitur (*Safety*) hingga mencapai 94.57% saat semua fitur digunakan. Ini menunjukkan bahwa setiap fitur tambahan memberikan informasi yang signifikan yang membantu model membuat prediksi yang lebih akurat.

Analisis ini menunjukkan bahwa penambahan fitur input secara signifikan meningkatkan akurasi model *Random Forest*. Fitur-fitur utama seperti *Safety*, *Persons*, *Buying*, *Maint*, dan *Lug\_boot* memberikan kontribusi besar terhadap peningkatan akurasi. Penggunaan semua fitur memberikan akurasi tertinggi, namun penggunaan 5 fitur utama juga sudah memberikan hasil yang hampir maksimal dengan kompleksitas yang lebih rendah. Hal ini

penting untuk dipertimbangkan dalam implementasi model di dunia nyata, di mana keseimbangan antara akurasi dan efisiensi komputasi menjadi faktor kunci.

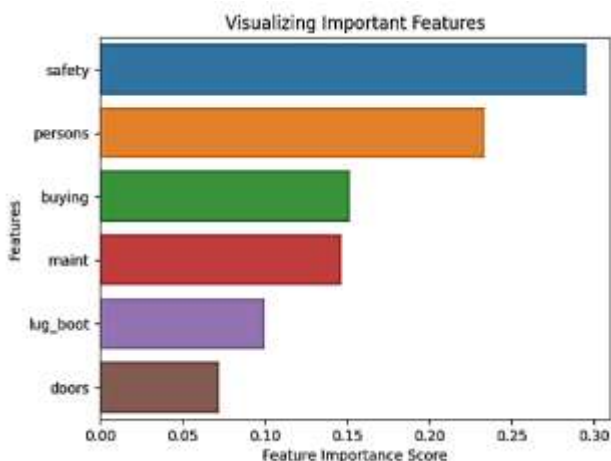
**b. Eliminasi selection feature pada jumlah feature input yang sama.**

Pada peng-eliminasian *selected feature* kami menghilangkan fitur tertentu pada saat training. Kemudian dilakukan pengujian kembali. Pemilihan fitur input dapat membuat model random forest yang lebih cepat [15]. Hasil perbandingan akurasi untuk setiap kondisi pengeliminasian tersebut dapat dilihat pada Gambar 4. Berdasarkan Gambar tersebut akurasi terbesar diperoleh ketika menghilangkan fitur *doors* pada saat pelatihan, kemudian akurasi menurun mulai dari *Lug\_boot*, *Maintenance*, *Buying*, *Persons* dan akurasi paling rendah pada saat eliminasi *Safety*.



Gambar 4. Perbandingan akurasi *Random Forest Classifier* pada peng-eliminasian *selected feature*.

Setiap fitur tersebut mempunyai “*feature important score*” tersendiri. Adapun nilai *important features* tersebut ditunjukkan pada Gambar 5. Dapat terlihat bahwa urutan nilai *important features* dari yang paling besar adalah *Safety*, *Persons*, *Buying*, *Maintenance*, *Lug\_boot*, dan paling rendah adalah *doors*.



Gambar 5. Visualisasi *Important features*.

Berdasarkan Gambar 5 dan Gambar 4, dapat terlihat bahwa ketika mengeliminasi suatu fitur maka akan akurasi menurun. Kemudian penurunan akurasi tersebut berkorelasi antara “*feature importance score*” dengan hasil perbandingan akurasi *Random Forest* pada saat menghilangkan fitur tertentu. Dimana, ketika kita mengeliminasi fitur yang mempunyai “*feature importance score*” rendah maka penurunan tingkat akurasi nya tidak terlalu besar. Yakni, pada saat mengeleminasi fitur *doors* yang mempunyai “*feature importance score*” paling kecil, menghasilkan akurasi yang paling besar. Sebaliknya, untuk fitur *Safety* yang mempunyai “*feature importance score*” paling besar, menghasilkan akurasi paling kecil. Hal tersebut menunjukkan bahwa semakin besar “*feature importance score*” maka penurunan tingkat akurasi semakin besar pula.

Korelasi Antara Feature Importance dan Akurasi:

- a. Fitur dengan Feature Importance Tinggi: *Safety* (0.295): Menghilangkan fitur ini menurunkan akurasi secara signifikan ke 63.57%. Ini menunjukkan bahwa fitur *Safety* sangat penting dalam memprediksi kelayakan kendaraan. *Persons* (0.234): Menghilangkan fitur ini menurunkan akurasi menjadi 71.10%. Fitur ini juga penting karena informasi mengenai kapasitas penumpang berkontribusi besar dalam pengambilan keputusan.
- b. Fitur dengan Feature Importance Menengah: *Buying* (0.152): Menghilangkan fitur ini menurunkan akurasi menjadi 74.78%. Harga pembelian merupakan faktor penting, tetapi tidak sepenting *Safety* atau *Persons*. *Maint* (0.147): Menghilangkan fitur ini menurunkan akurasi menjadi 78.81%. Biaya pemeliharaan memberikan kontribusi yang signifikan, namun sedikit lebih rendah dibandingkan dengan fitur *Buying*.
- c. Fitur dengan Feature Importance Rendah: *Lug\_boot* (0.10): Menghilangkan fitur ini menurunkan akurasi menjadi 83.71%. Kapasitas bagasi penting namun tidak sebesar fitur lain. *Doors* (0.072): Menghilangkan fitur ini hanya menurunkan akurasi menjadi 92.64%, menunjukkan bahwa jumlah pintu adalah fitur yang paling tidak penting di antara semua fitur.

Penurunan akurasi ketika menghilangkan fitur tertentu menunjukkan seberapa penting fitur tersebut dalam membantu model membuat prediksi yang akurat. Semakin tinggi nilai *feature importance*, semakin besar pengaruhnya terhadap akurasi model. Fitur *Safety* dan *Persons* memiliki nilai *importance* yang tinggi dan eliminasi fitur ini menyebabkan penurunan akurasi yang signifikan. Ini menunjukkan bahwa kedua fitur ini memberikan informasi yang kritis untuk model. Fitur *Buying* dan *Maint* juga penting, namun tidak sebesar *Safety* dan *Persons*. Fitur ini memberikan informasi yang cukup signifikan tetapi tidak dominan. Fitur *Lug\_boot* dan *Doors* memiliki pengaruh yang lebih rendah. Menghilangkan fitur-fitur ini tidak menurunkan akurasi secara drastis, menunjukkan bahwa fitur-fitur ini memberikan kontribusi yang lebih kecil.

#### 4. Kesimpulan

Penelitian ini menyelidiki pengaruh jumlah *decision tree* dan jumlah serta pemilihan fitur input terhadap akurasi model *Random Forest* dalam memprediksi kelayakan kendaraan. Penambahan jumlah *decision tree* secara umum meningkatkan akurasi model. Akurasi meningkat signifikan dari 86.51% dengan 1 *decision tree* hingga mencapai stabil pada



sekitar 94.57% dengan 100 *decision tree* dan seterusnya. Setelah mencapai sekitar 50 *decision tree*, akurasi model cenderung stabil, dengan variasi kecil dalam peningkatan akurasi. Ini menunjukkan bahwa setelah titik tertentu, menambah lebih banyak *decision tree* tidak memberikan peningkatan akurasi yang signifikan. Penambahan fitur input secara bertahap meningkatkan akurasi model. Akurasi meningkat dari 69.53% saat hanya menggunakan fitur *Safety* hingga mencapai 94.57% saat menggunakan semua fitur. Fitur-fitur seperti *Safety*, *Persons*, *Buying*, *Maint*, dan *Lug\_boot* memberikan kontribusi signifikan terhadap akurasi model. Penggunaan lima fitur utama ini memberikan akurasi mendekati maksimum, yakni 92.64%, yang sangat dekat dengan akurasi saat menggunakan semua fitur (94.57%). Fitur *Safety* memiliki importance tertinggi dan menghilangkan fitur ini menurunkan akurasi menjadi 63.57%. Hal ini menunjukkan bahwa fitur ini sangat penting dalam menentukan akurasi model. Menghilangkan fitur *Persons*, *Buying*, dan *Maint* juga menurunkan akurasi, namun tidak sebesar *Safety*. Fitur *Lug\_boot* dan *Doors* memiliki importance yang lebih rendah, dan menghilangkan fitur ini tidak terlalu berdampak signifikan pada akurasi. Dalam praktik, penting untuk menyeimbangkan antara akurasi dan kompleksitas model. Penggunaan lima fitur utama memberikan akurasi yang sangat dekat dengan maksimum sambil mengurangi kompleksitas komputasi. Pemilihan fitur berdasarkan *importance score* sangat penting untuk meningkatkan performa model tanpa menambah kompleksitas yang tidak perlu. Fitur-fitur dengan importance tinggi harus dipertahankan untuk menjaga akurasi model. Menggunakan sekitar 50-100 *decision tree* sudah cukup untuk mencapai akurasi yang optimal tanpa menambah kompleksitas komputasi yang tidak perlu. Fokus pada fitur-fitur dengan *importance score* tinggi seperti *Safety*, *Persons*, *Buying*, *Maint*, dan *Lug\_boot* untuk meningkatkan akurasi model. Hindari mengeliminasi fitur dengan importance tinggi karena dapat menurunkan akurasi secara signifikan.

### Ucapan Terima Kasih

Kami ingin menyampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penyelesaian penelitian ini. Kami berterima kasih kepada Bapak Dr.Eng. Willy Anugrah Cahyadi, ST., M.T. dan Ibu Khilda Afifah, S.Pd., M.T., Ph.D atas bimbingan, saran, dan dukungannya selama proses penelitian ini. Penghargaan khusus kami sampaikan kepada rekan-rekan peneliti, yang telah berkontribusi dalam pengumpulan data dan analisis. Ucapan terima kasih ini juga ditujukan kepada semua pihak yang tidak dapat disebutkan satu per satu, tetapi turut membantu dan mendukung kelancaran penelitian ini.

### Daftar Pustaka

- [1] K. Perindustrian, "Upaya Industri Otomotif Bangkit Dari Pandemi," in *Buku Analisis Kinerja Industri Otomotif*, II., Kementrian Perindustrian Republik Indonesia, 2022.
- [2] M. Thoifur, "Penjualan Mobil Baru di Indonesia Meningkat karena Ini," *Viva.co.id*, 2022. [viva.co.id/otomotif/1503524-penjualan-mobil-baru-di-indonesia-meningkat-karena-ini](https://viva.co.id/otomotif/1503524-penjualan-mobil-baru-di-indonesia-meningkat-karena-ini) (accessed Jul. 07, 2023).
- [3] A. Nugroho, "Analisa Splitting Criteria Pada Decision Tree dan Random Forest

- untuk Klasifikasi Evaluasi Kendaraan,” *JSITIK J. Sist. Inf. dan ...*, 2022, [Online]. Available: <https://jurnal.ciptamediaharmoni.id/index.php/jsitik/article/view/154>
- [4] Q. Pham, “Artificial Intelligence ( AI ) and Big Data for Coronavirus ( COVID-19 ) Pandemic : A Survey on the State-of-the-Arts,” vol. 8, 2020.
- [5] D. E. Goldberg and J. H. Holland, “Genetic Algorithms and Machine Learning,” in *Machine Learning*, Netherlands: Kluwer Academic Publishers, 1988, pp. 95–99.
- [6] M. Schonlau and R. Y. Zou, “The random forest algorithm for statistical learning,” *Stata J.*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.
- [7] H. Lee, J. Wang, and B. Leblon, “Using Linear Regression, Random Forests, and Support Vector Machine with Unmanned Aerial Vehicle Multispectral Images to Predict Canopy Nitrogen Weight in Corn,” *Remote Sensing*, vol. 12, no. 13, 2020. doi: 10.3390/rs12132071.
- [8] V. Ramya and K. Ganapathy, “Evaluation of Vehicle Quality Performance using Random forest in Comparison with KNN to measure the Accuracy, Recall, and Precision,” in *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, 2022, pp. 550–555. doi: 10.1109/ICIEM54221.2022.9853143.
- [9] R. Genuer, J. M. Poggi, R. Genuer, and J. M. Poggi, *Random forests*. Springer, 2020. doi: 10.1007/978-3-030-56485-8\_3.
- [10] J. Awwalu, A. Ghazvini, and A. A. Bakar, “Performance Comparison of Data Mining Algorithms : A Case Study on Car Evaluation Dataset,” vol. 13, no. 2, pp. 78–82, 2014.
- [11] M. Das and R. Dash, “Performance Analysis of Classification Techniques for Car Data Set Analysis,” pp. 549–553, 2020.
- [12] N. A. Ahmed, “What is A Confusion Matrix in Machine Learning? The Model Evaluation Tool Explained,” *Datacamp.com*, 2023. <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning> (accessed Jun. 22, 2024).
- [13] A. Marquand *et al.*, “A Novel Approach to Probabilistic Biomarker-Based Classification Using Functional Near-Infrared Spectroscopy,” *Hum. Brain Mapp.*, vol. 34, May 2013, doi: 10.1002/hbm.21497.
- [14] M. J. Raihan, M. A.-M. Khan, S.-H. Kee, and A. Nahid, “Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP,” *Sci. Rep.*, vol. 13, Apr. 2023, doi: 10.1038/s41598-023-33525-0.
- [15] M. A. M. Hasan, M. Nasser, S. Ahmad, and K. I. Molla, “Feature Selection for Intrusion Detection Using Random Forest,” *J. Inf. Secur.*, vol. 07, no. 03, pp. 129–140, 2016, doi: 10.4236/jis.2016.73009.