

## Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid

Aditya Arisusanto<sup>1\*</sup>, Nana Suarna<sup>2</sup>, Gifthera Dwilestari<sup>3</sup>

Program Studi Teknik Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia

Email: <sup>1\*</sup>adit026osd@gmail.com, <sup>2</sup>st\_nana@yahoo.com, <sup>3</sup>gifthera.ikmi@gmail.com

Email Coressponding Author: adit026osd@gmail.com

**Abstrak**— Perkembangan teknologi dan komunikasi dari waktu ke waktu dirasakan semakin meningkat pesat, yang mendorong penggunaan dan pemanfaatan perkembangan teknologi di berbagai bidang dan aspek kehidupan, contohnya seperti penggunaan *smartphone*. Klasifikasi handphone yang dilakukan yaitu dengan menggunakan metode random forest dengan optimize parameter grid. Masalah belum mengetahui berapa nilai akurasi terhadap klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid, belum mengetahui pengaruh nilai parameter *maximal\_depth* dan jenis parameter *criterion* *gain\_ratio*, *accuracy* dan *gini\_index* pada klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid. Tujuan ingin mengetahui berapa nilai akurasi terhadap klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid. Ingin mengetahui pengaruh nilai parameter *maximal\_depth* dan jenis parameter *criterion* *gain\_ratio*, *accuracy* dan *gini\_index* pada klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid. Metode klasifikasi yang akan diterapkan adalah metode *Random forests dengan Optimize Parameter Grid*. Hasilnya penelitian ini yaitu dapat mengetahui berapa nilai akurasi terhadap klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid dan dapat mengetahui pengaruh nilai parameter *maximal\_depth* dan jenis parameter *criterion* *gain\_ratio*, *accuracy* dan *gini\_index* pada klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid Dengan nilai *accuracy* tertinggi sebesar 88,50%.

**Kata Kunci:** klasifikasi, handphone, algoritma decision tree random forest, optimize parameter grid

**Abstract**— The development of technology and communication from time to time is felt to be increasing rapidly, which encourages the use and utilization of technological developments in various fields and aspects of life, such as the use of smartphones. The classification of mobile phones that is carried out is by using the random forest method with optimize grid parameters. Problems do not know the accuracy value of the classification of cellphone price data using the Random Forest Algorithm with the Optimize Parameter Grid, do not know the effect of the *maximal\_depth* parameter value and the type of *criterion* parameter *gain\_ratio*, *accuracy* and *gini\_index* on the classification of cellphone price data using the Random Forest Algorithm with the Optimize Parameter Grid. Purpose want to know the accuracy value of the classification of mobile phone price data using the Random Forest Algorithm with Optimize Parameter Grid. Want to know the effect of the *maximal\_depth* parameter value and the type of *criterion* parameter *gain\_ratio*, *accuracy* and *gini\_index* on the classification of cellphone price data using the Random Forest Algorithm with Optimize Parameter Grid. The classification method that will be applied is the Random Forests method with Optimize Parameter Grid. The results of this research are being able to find out the accuracy of the classification of cellphone price data using the Random Forest Algorithm with the Optimize Parameter Grid and being able to find out the effect of the *maximal\_depth* parameter value and the type of parameter *criterion* *gain\_ratio*, *accuracy* and *gini\_index* on the classification of cellphone price data using the Random Forest Algorithm with Optimize Parameter Grid With the highest accuracy value of 88.50%.

**Keywords:** classification, handphone, decision tree random forest algorithm, optimize grid parameters

### 1. PENDAHULUAN

Pada masa sekarang ini perkembangan teknologi dan komunikasi dari waktu ke waktu dirasakan semakin meningkat pesat, yang mendorong penggunaan dan pemanfaatan perkembangan teknologi di berbagai bidang dan aspek kehidupan, contohnya seperti penggunaan *smartphone*. Masyarakat jaman sekarang bahkan kalangan menengah kebawah bisa memiliki *smartphone*. Perkembangan teknologi ini memicu persaingan oleh berbagai kalangan pebisnis *smartphone* yang ada dalam melakukan penjualan, sehingga membuat tiap pebisnis *smartphone* harus memiliki fitur dan performa yang lebih baik agar tetap bisa bersaing dalam teknologi *smartphone*. Menetapkan harga merupakan sesuatu yang sangat penting bagi perusahaan yaitu memperoleh penghasilan dan laba untuk perusahaan agar tetap bisa berjalan. Permasalahan yang umum biasanya seperti bagaimana penjualan barang di masa yang akan datang lebih laku. Untuk mempermudah toko dalam penyediaan stok barang juga diperlukan sebuah analisa dalam memprediksi dan mengklasterkan barang berdasarkan harga serta perlu di analisa dataset harga barang menggunakan data mining.

Penelitian yang dilakukan oleh: Vanissa Wanika Siburian, Ika Elvina Mulyana, menyimpulkan bahwa, hasil percobaan yang dilakukan didapatkan tingkat akurasi prediksi dengan menggunakan metode Random Forest sebesar

81%. Selain akurasi didapatkan pula nilai precision, nilai recall, dan nilai dari f1-score yang sama sebesar 81%. Kemudian tingkat akurasi tertinggi pada variabel respon terdapat pada kategori harga murah. (Siburian & Mulyana, 2018). Penelitian terdahulu yang dilakukan oleh: Ahmadi Irmansyah Lubis, Umri Erdiansyah, Rosma Siregar, dalam jurnalnya yang berjudul : *Komparasi Akurasi Pada Naïve Bayes Dan Random Forest Dalam klasifikasi penyakit liver menyimpulkan bahwa : hasil yang diperoleh setelah pengujian pada penelitian ini dengan menggunakan Indian Liver Patient Dataset (ILPD) diketahui dari hasil penelitian yaitu dengan metode Naïve Bayes diperoleh nilai akurasi 55.80 % dan metode Random Forest memperoleh nilai akurasi 70.60%. Sehingga dapat disimpulkan bahwa Naïve Bayes dan Random Forest memiliki performa kinerja yang berbeda di dalam proses perhitungan klasifikasi pada kasus penyakit liver yang dimana Random Forest lebih unggul peroleh akurasi yang dihasilkan.* (Ahmadi Irmansyah Lubis, Umri Erdiansyah, 2022).

Berdasarkan latar belakang di atas, permasalahan yang didapat dalam penelitian ini adalah “Bagaimana mengetahui berapa nilai akurasi terhadap klasifikasi data harga handphone dengan menggunakan algoritma Random Forest dengan Optimize Parameter Grid, Belum dapat mengetahui pengaruh nilai parameter maximal\_depth dan jenis parameter criterion gain\_ratio, accuracy dan gini\_index pada klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid. Sehingga perlu dilakukan analisa klasterkan harga handphone dengan Algoritma Random Forest.

Fokus masalah dari penelitian ini adalah Belum mengetahui berapa nilai akurasi terhadap klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid terhadap analisa nilai parameter maximal\_depth dan jenis parameter criterion gain\_ratio, accuracy dan gini\_index pada klasifikasi data harga handphone menggunakan Algoritma Random Forest dengan Optimize Parameter Grid. Berdasarkan latarbelakang permasalahan yang telah diutarakan maka penulis membuat penelitian dengan judul “Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid” dengan alasan yaitu bertujuan untuk menggali informasi dalam berapa nilai akurasi terhadap klasifikasi data harga handphone dengan menggunakan Algoritma Random Forest dengan Optimize Parameter Grid.

## 2. KERANGKA TEORI

### 2.1 Random Forest

Random Forest adalah pengembangan dari metode Decision Tree yang menggunakan beberapa Decision Tree, dimana setiap Decision Tree telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk resisting outliers, serta efisien untuk penyimpanan sebuah data. Mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi. (Supriyadi et al., 2020)

### 2.2. Knowledge Discovery in Database (KDD)

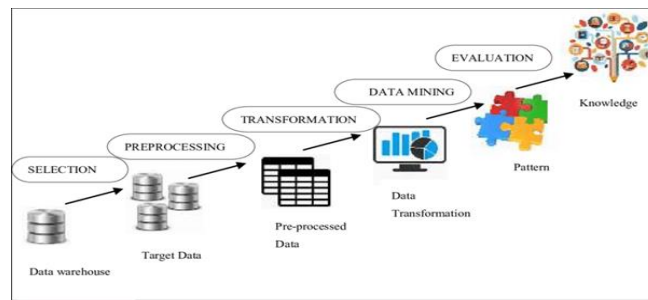
KDD adalah sebuah proses mengubah data dalam jumlah besar menjadi sebuah Data mining informasi yang berharga. Kualitas dari data akan sangat berpengaruh pada data mining. Data yang besar memiliki potensi data hilang dan tidak konsisten (Ferdyani et al., 2022).

### 2.3 Ponsel/Telepon

Ponsel atau telepon genggam adalah salah satu bukti produk kecanggihan teknologi yang semakin maju, karena pada fungsi umumnya. Ponsel adalah sebuah alat komunikasi jarak jauh. Namun kini ponsel bisa multifungsi dan mengubah ke segala aspek kehidupan manusia mulai dari baca berita, main game, pemotretan, dan lain sebagainya. (Muhayat et al., 2022). Handphone merupakan perlengkapan telekomunikasi elektronik 2 arah yang dapat dibawa dengan mudah serta mempunyai keahlian untuk mengirimkan pesan berbentuk suara. Penafsiran tersebut ialah penafsiran HP secara universal (Indah Nuryani, 2021)

## 3. METODE PENELITIAN

Menganalisis dataset rumah tidak layak huni dalam penerapan data mining ini menggunakan proses tahapan *Knowledge Discovery in Databases* (KDD) yang terdiri dari Data, *Data Cleaning*, *Data transformation*, *Data mining*, *Pattern evolution*, *knowledge*, seperti terlihat gambar 1:



**Gambar 1.** Tahapan *Knowledge Discovery in Databases* (KDD)

## Tahapannya *Knowledge Discovery in Databases* (KDD) sebagai berikut

1. **Data**  
Data set electronic. diperoleh dari data yang bersumber dari *repository Kaggle.go.id*. Data yang digunakan pada penelitian ini ialah sebanyak 936 dataset electronic.
2. **Seleksi Data (*Data Selection*)**  
Proses ini memilih dan menyeleksi data elektronik yang diperlukan. Data tersebut dikelompokkan menjadi dataset. Pada umumnya, data yang diperoleh, baik dari database maupun survey, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik.
3. ***Data transformation***  
Bagian ini adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.
4. **Data mining**  
Bagian ini yaitu proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Databases* (KDD) secara keseluruhan.
5. ***Evaluation***  
Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola pola yang khas maupun model dievaluasi untuk menilai apakah memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai maka ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

## 4. HASIL

Hasil penelitian yang dilakukan dalam pembahasan ini yaitu akan menguraikan proses bagaimana mengelompokkan atau clusterisasi dataset elektronik dengan menggunakan *Algoritma Fuzzy C-Mean*. Pengelompokkan ini dilakukan dengan proses pengujian menggunakan machine learning yaitu *Software RapidMiner Studio*.

### 1. Data

Data dalam bentuk file excel 2003 yang bernama dataset harga HP diperoleh dari site <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>. Jumlah record sebanyak 2000 record, terdiri dari 21 atribut seperti tampak pada tabel 1.

**Tabel 1.** Atribut default dari file electronic item price

No.	Atribut	Type Data	Keterangan
1.	battery_power	integer	Total energi yang dapat disimpan baterai dalam satu waktu diukur dalam mAh
2.	blue	integer	Ada bluetooth atau tidak
3.	clock_speed	real	Kecepatan di mana mikroprosesor mengeksekusi instruksi
4.	dual_sim	integer	Apakah mendukung dual sim atau tidak
5.	fc	integer	Kamera depan dalam mega piksel
6.	four_g	integer	Apakah 4G atau tidak
7.	int_memory	integer	Memori Internal dalam Gigabyte
8.	m_dep	real	Kedalaman Seluler dalam cm
9.	mobile_wt	integer	Berat ponsel
10.	n_cores	integer	Jumlah inti prosesor

11.	pc	integer
12.	px_height	integer
13.	px_width	integer
14.	ram	integer
15.	sc_h	integer
16.	sc_w	integer
17.	talk_time	integer
18.	three_g	integer
19.	touch_screen	integer
20.	wifi	integer
21.	price_range	integer

## 2. Selection

Dataset yang dibaca dalam bentuk *file* excel, menggunakan operator *Read Excel* pada Rapid Miner.

## 3. Preprocessing

Proses *cleansing* atau pembersihan data yang *missing* atau memiliki nilai yang tidak konsisten pada langkah *preprocessing*. Sebelum melakukan proses ini, dilakukan analisa terlebih dahulu apakah atribut pada dataset yang dipilih memiliki nilai *missing* atau tidak serta memiliki data yang konsisten atau tidak. Dari hasil *result* dari statistik dataset seperti tampak pada gambar 2 dibawah ini, diketahui bahwa tidak ada atribut yang memiliki nilai *missing*. Untuk memeriksa konsisten atau tidak konsistennya dataset yang digunakan diperiksa per-*record* secara langsung dan menunjukkan bahwa dataset memiliki data yang konsisten terhadap nilainya.

Attribute	Type	Min	Max	Average
id	Integer	0	2000	1000.500
price_range	Integer	0	3	1.500
battery_power	Integer	0	1998	1238.518
blue	Integer	0	1	0.495
clock_speed	Real	0	3	1.522
dual_sim	Integer	0	1	0.509
fc	Integer	0	19	4.309
four_g	Integer	0	1	0.521
int_memory	Integer	0	64	32.047
m_dep	Real	0	1	0.502
mobile_wt	Integer	0	200	140.249

Gambar 2. Result dari statistik dataset

## 4. Transformation

Tipe data pada label berjenis numerik sedangkan algoritme klasifikasi *Random Forest* jenis label harus bertipe nominal, maka diperlukan transformation untuk mengubah data bertipe *numeric* menjadi *polynomial* menggunakan operator *Numerical to Polynomial*

## 5. Data mining

Pada langkah data mining karena menggunakan *Optimize Parameter (Grid)*, maka operator *Optimize Parameter (Grid)* digunakan lebih awal sebelum menggunakan operator lainnya. Operator *Optimize Parameter (Grid)* merupakan jenis operator sub proses, maka didalam operator ini dipasang operator *Split Data* untuk membagi data menjadi *data training* dan *data testing*.

## 6. Evaluation

Evaluasi yang dilakukan terhadap hasil kegiatan eksperimen terhadap dataset yang diperoleh hasil seperti tabel 2 sebagai berikut.

Tabel 2. Akurasi Hasil Data mining

iteration	maximal_depth	criterion	accuracy
1	1	gain_ratio	0.2375
2	2	gain_ratio	0.5650
3	3	gain_ratio	0.7400
4	4	gain_ratio	0.7550
5	5	gain_ratio	0.7650

6	6	gain_ratio	0.7625
7	7	gain_ratio	0.7675
8	8	gain_ratio	0.7750
9	9	gain_ratio	0.7825
10	10	gain_ratio	0.7900
.....	.....	....	.....
90	30	gini_index	0.8650

Dari ketiga parameter *criterion*, diperoleh nilai *accuracy* terbesar pada *criterion gini\_index* sebesar 0,8850 (88,5%) pada *maximal\_dept* 12 dan 25.

Tabel 3. Akurasi Hasil Data mining

iteration	maximal_depth	criterion	accuracy
26	26	gain_ratio	0.8650
42	12	accuracy	0.8550
51	21	accuracy	0.8550
72	12	gini_index	<b>0.8850</b>
85	25	gini_index	<b>0.8850</b>

## 5. KESIMPULAN

Berdasarkan hasil penelitian pengelompokkan Dataset handphone menggunakan *Algoritma Random Forest*, maka hasilnya dapat disimpulkan sebagai berikut :

1. Nilai akurasi terhadap klasifikasi data harga handphone menggunakan *Algoritma Random Forest* dengan *Optimize Parameter Grid* diperoleh sebesar 88,50%
2. Nilai *accuracy* tertinggi sebesar 88,50%, diperoleh dari nilai parameter *maximal\_depth* optimal sebesar 12 dan 25, serta parameter *criterion* yang optimal pada *gini\_index* pada klasifikasi data harga handphone menggunakan *Algoritma Random Forest* dengan *Optimize Parameter Grid*.

## DAFTAR PUSTAKA

- Ahmadi Irmansyah Lubis, Umri Erdiansyah, R. S. (2022). Komparasi Akurasi Pada Naïve Bayes Dan Random Forest Dalam klasifikasi penyakit liver. (*Journal of Computing Engineering, System and Science*), vol. 7 Jan.
- Alfian Faiz Izzulhaq, S. (2020). KLASIFIKASI PENJUALAN APLIKASI ANDROID MENGGUNAKAN ALGORITMA C4.5. *SENDIU*, 2019, 978–979.
- Amaliah, S., & Nusrang, M. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng. 4(2), 121–127. <https://doi.org/10.35580/variandsium31>
- Dhika, H., Veris, H., Studi, P., Informatika, T., Gedong, K., Rebo, P., & Timur, J. (2021). PENERAPAN ALGORITMA DECISION TREE DALAM PENJUALAN HANDPHONE. *JRKT (Jurnal Rekayasa Komputasi Terapan)*, 01(04), 239–246.
- Ferdyandi, M., Setiawan, N. Y., & Bachtiar, F. A. (2022). Prediksi Potensi Penjualan Makanan Beku berdasarkan Ulasan Pengguna Shopee menggunakan Metode Decision Tree Algoritma C4 . 5 dan Random Forest ( Studi Kasus Dapur Lilis ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(2), 588–596.
- Gultom, Z. A., Kom, K. E. S., Kom, M., Ahmad, D., & Kom, C. S. (2019). Implementasi Data Mining untuk Memprediksi Penghasilan Penjualan Handphone di Toko Ponsel Takasimura dengan Metode Rough Set. *Jurnal Cyber Tech*, 1–14.
- Indah Nuryani, D. D. (2021). Analisis clustering pada pengguna brand hp menggunakan metode k-means. *Prosiding Seminar Nasional Ilmu Komputer*, 1(1), 190–211.
- Muhayat, T., Kom, S., Si, M., Chamidah, N., Kom, S., Kom, M., & Prediksi, K. (2022). Prediksi harga Smartphone menggunakan Algoritma Multiple Linear Regression. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, e-ISSN 296, 506–525.
- Putra, S. H. (2018). Klasifikasi Harga Cell Phone menggunakan Metode K-Nearest Neighbor ( KNN ). *Computer Science and ICT*, 4(1), 978–979.
- Siburian, V. W., & Mulyana, I. E. (2018). Prediksi Harga Ponsel Menggunakan Metode Random Forest. *Computer Science and ICT*, 4(1), 978–979.
- Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *JURNAL ILMIAH EKONOMI DAN BISNIS*, 13(2), 67–75.
- Yoga Religial, Agung Nugroho2, W. H. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *JURNAL RESTI*, 1(10), 187–192.