

**PERBANDINGAN ALGORITMA C4.5 DENGAN C4.5 BAGGING DALAM  
MEMPREDIKSI DAN ANALISA DATA SET PEMILIH MINYAK PELUMAS MESIN  
(OLI) PADA KEDARAAN RODA DUA**

**Tri Ginanjar Laksana\*, Fahrudin Mukti Wibowo**

Jurusan Informatika, Sekolah Tinggi Telematika Telkom Purwokerto

Jl. D.I. Panjaitan No. 128 Purwokerto 53147

HP. 0821.3338.8278/ 0898.0476.695

e-mail : anjarlaksana@st3telkom.ac.id

**Abstrak**

*Semakin berkembangnya produsen pelumas kendaraan roda 2 (dua), baik motor jenis matik dan manual. oleh karena itu, untuk memprediksi pola pemilihan merk pelumas mesin kendaraan roda dua yang banyak beredar dipasaran, hal tersebut memberikan kesulitan pada para pengguna roda dua untuk melakukan perawatan mesin kendaraannya. Berdasarkan hasil penelitian terdahulu terdapat 4 analisa anomali masalah pada pelumas mesin kendaraan (oli), dalam permasalahan yang terjadi bahwa, 1. Tekanan oli terlalu rendah terhadap mesin, 2. Tekanan oli terlalu tinggi terhadap mesin, 3. Temperatur oli terlalu tinggi dan 4. Temperatur oli terlalu rendah. Objek pada penelitian ini adalah pengguna kendaraan roda 2 baik matic ataupun manual, data yang digunakan adalah data primer melalui kuisioner (pertanyaan). Dalam peper penelitian ini menggunakan metode data mining, dimana membandingkan antara algoritma C4.5 (Decisoin tree) dan algoritma C4.5 berbasis bagging, hal tersebut dilakukan melalui perhitungan accuracy, RMSE, dan Classification Error untuk mendapatkan hasil kesesuaian diantara algoritma tersebut. Hasil klasifikasi diharapkan dapat digunakan dalam menganalisa pola pemilihan minyak pelumas kendaraan roda dua yang sesuai.*

**Kata kunci :** Algoritma, Analisa, C4.5, C4.5 Bagging, Oli.

## 1. PENDAHULUAN

Berdasarkan hasil penelusuran yang dilakukan, baik berupa literature di jurnal (study pustaka), browsing, artikel, majalah dan koran, semakin berkembangnya produsen pelumas kendaraan roda 2 (dua), baik motor jenis matik dan manual. dari data yang diperoleh dilapangan ada beberapa jenis merk pelumas (oli) kendaraan, seperti tabel dibawah ini :

**Tabel 1. Tabel Berbagai Jenis Oli (Pelumas)**

No	Merk Oli (Pelumas)	No.	Merk Oli (Pelumas)
1	Oli AHM (Astra Honda Motor)	7	Pertamina Enduro
2	Castrol Oil	8	Oli Motor Pertamina Fastron
3	Evalube	9	Pertamina Mesran
4	Federal	10	Repsol
5	Motul	11	Oli Motor Shell
6	Yamalube	12	Top One

*Sumber Data : Hargaupdate.com, 2017*

Berdasarkan tabel 1. Dapat diketahui merk oli (pemunas motor) yang beredar dimasyarakat update sampai 15 Maret 2017, dengan banyaknya pemilihan pelumas oli, memiliki kemampuan dan ke khasan untuk menjaga mesin motor agar tetap baik, dalam jangka waktu yang lama. Algoritma Klasifikasi C4.5 dan C4.5 berbasis bagging telah banyak digunakan untuk melakukan prediksi dan analisis, dan memiliki keakurasian dan kemampuan dalam melakukan kecepatan pembuatan pemodelan dalam menganalisa dan memprediksi, penelitian terdahulu yang dilakukan oleh ari wibowo tahun 2015, dengan judul Analisis Perbandingan Kinerja Metode Klasifikasi Dalam Data Mining, masalah yang diangkat Untuk mengetahui sebaran dan karakteristik nasabah yang putus di tengah jalan perlu dilakukan pengelompokan/klasifikasi sesuai dengan karakteristiknya, Algoritma/ Modetode yang digunakan Random Forest, Bagging dan Boosting, Hasil yang diperoleh tingkat akurasi prediksi paling baik adalah metode Bagging CART. Dimana metode tersebut bisa melakukan prediksi dengan tingkat kebenaran/akurasi mencapai 90%, sementara metode yang lain

hanya memiliki tingkat akurasi kurang dari 85%. (Wibowo, 2015), kemudian penelitian yang dilakukan oleh Sukmawati Anggraeni Putri tahun 2017. Hasil terbukti memperoleh hasil yang baik untuk menangani imbalance class dan noise attribute untuk prediksi cacat software, serta dapat meningkatkan hasil akurasi dari prediksi cacat pada software (Putri, 2017), dan Hasil dari penelitian menunjukkan bahwa metode yang diusulkan membuat peningkatan yang signifikan pada dua algoritma klasifikasi yaitu decision tree dengan P value of t-Test sebesar 0,0184 dan k-nearest neighbor dengan P value of t-Test sebesar 0,0292, akan tetapi tidak signifikan pada naïve bayes dengan P value of t-Test sebesar 0,9236. Setelah diterapkan teknik bagging pada lima dataset medis, naïve bayes memiliki akurasi paling tinggi untuk dataset breast-cancer sebesar 96,14% dengan AUC sebesar 0,984, heart-disease sebesar 84,44% dengan AUC sebesar 0,911 dan pima diabetes sebesar 74,73% dengan AUC sebesar 0,806. Sedangkan k-nearest neighbor memiliki akurasi yang paling baik untuk dataset liver-disorder sebesar 62,03% dengan AUC sebesar 0,632 dan vertebral column dengan akurasi sebesar 82,26% dengan AUC sebesar 0,867 (Rizki Tri Prasetio, 2015). Penelitian terdahulu tentang C4.5 dan Bagging pernah juga dilakukan oleh : (Fransiska Amalia Kurniawan, 2011), (Amalia, 2007), (Wibowo, 2015), (Safuan, 2015), (Ari Wibowo, 2012), (Rizki Tri Prasetio, 2015), (Syarif, 2015), (Made Leo Radhitya, 2016), (Putri, 2017), (Fitriyani, 2015), (Yusuf s, Nugroho, 2015)

Berdasarkan penelitian terdahulu dapat disimpulkan bahwa, teknik bagging terbukti dapat digunakan untuk meningkatkan akurasi algoritma yang digunakan, penelitian ini akan menguji coba teknik bagging dalam meningkatkan akurasi algoritma C4.5 didalam data set pengguna pelumas (Oli) pada motor, dimana data yang diperoleh adalah data primer yang diperoleh dari hasil survey, kepada 50 pengguna kendaraan bermotor, dimana menggunakan 10 atribut, diantaranya Kode pengguna, Jenis Kelamin, Umur Pengguna, Hoby, Status, Daerah Tinggal, Jenis Motor, Merk Motor, Jenis Oli yang digunakan.

## 2. METODOLOGI

Dalam melakukan penelitian yang telah dilakukan, maka disusunlah tahapan dalam untuk menyelesaikan permasalahan adalah sebagai berikut :

### 2.1. Studi Literatur

Melakukan proses pencarian melalui jurnal, majalah, koran, dan buku – buku, internet, perpustakaan, yang terkait dengan penelitian menggunakan algoritma C4.5 dan C4.5 berbasis bagging, Berdasarkan tujuan penelitian, maka penelitian yang dilakukan adalah dengan metode deskriptif kuantitatif yaitu suatu bentuk penelitian yang berdasarkan data yang dikumpulkan selama penelitian secara sistematis mengenai fakta-fakta yang akan diteliti. Metode ini bertujuan untuk memberikan analisa dan prediksi, menggunakan perbandingan algoritma C4.5 dan C4.5 Berbasis Bagging. Data yang dikumpulkan diperoleh dari data pimer yang dilakukan melalui penyebaran angket kepada pengguna kendaraan roda 2 baik matik maupun non matik (manual).

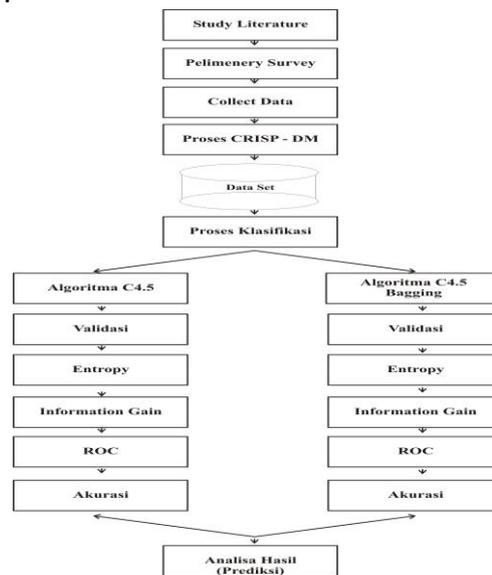
### 2.2. Penentuan Atribut

Berdasarkan data yang diperoleh, penelitian ini menggunakan atribut – atribut yang akan digunakan untuk perbandingan algoritma yang akan digunakan antara C4.5 dan C4.5 Berbasis bagging dalam data mining ini ditentukan sesuai dengan kebutuhan analisis. Adapun atribut yang digunakan dapat dilihat pada tabel 2. Atribut dan jenis variable yang digunakan, dimana variabel sebagai berikut :



### 2.6. Perancangan Tahapan Pemodelan Data Mining

Tahapan pemodelan yang dilakukan dalam penelitian ini adalah sebagai berikut, yang mana memiliki tahapan – tahapan :



Gambar 2. Tahapan Penelitian

## 3. PEMBAHASAN DAN HASIL

### 3.1 Pembahasan

Data set yang dipergunakan dalam penelitian ini, dapat dilihat pada tabel 3. Dibawah ini, terdapat 50 responden pengguna oli (pelumas kendaraan roda 2) yang telah memberikan jawaban data kuesionernya dari berbagai latar belakang yang berbeda – beda. Dimana atribut yang digunakan dalam pemodelan analisis dan prediksi pengguna oli adalah Id, Jenis Kelamin, Umur, Hobi, Status, Desa, Jenis Motor, Merk, jenis bahan bakar, oli.

Dari beberapa penelitian sebelumnya maka peneliti menggunakan metode penelitian yang akan digunakan dalam eksperimen ini adalah membandingkan algoritma C4.5 dengan C4.5 berbasis *Bagging*. Tahapan – tahapan yang dilakukan dalam penelitian ini, digunakan sesuai dengan gambar 2. Tahapan penelitian, dimana menggunakan beberapa analisa dan pemodelan serta perhitungan, seperti, perhitungan akurasi, perhitungan RMSE dan Klasifikasi error, dimana dipergunakan dalam mencari kesesuaian penggunaan algoritma yang digunakan dalam penelitian ini.

Tabel 3. Tabel Data Set

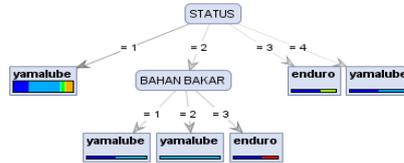
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
No	Nama	Jenis Kelamin	Tempat, Tanggal Lahir	Usia	Agama	Hobi	Status	Alamat	Kecamatan	Kelurahan	RT/RW	Jenis Kendaraan	Merk Kendaraan	Bahan Bakar	Pelumas/Oli
1	Chuffah Nur Zahmadyanti	P	Cirebon, 23 Desember 1999	17 tahun	Islam	Berenang	Lajang	J. sunai gunung jat gg. H Makoem	gunung jat	gunung jat	001/004	matic	honda vario	premium	enduro
2	Salman	L	Kebumen, 25 Mei 1989	28 tahun	Islam	Seni	Menikah	J. samadikun	kejaksan	kejaksan	008/008	Bebek	yamaha jupiter mx	premium	yamalube
3	Musliman	L	Kebumen, 06 Desember 1990	26 tahun	Islam	Main Musik	Lajang	J. samadikun	kejaksan	kejaksan	008/006	Bebek	yamaha jupiter mx	petralite	yamalube
4	Tarmiyah	P	Cirebon, 01 Januari 1989	27 tahun	Islam	Masak	Menikah	J. samadikun	kejaksan	kejaksan	008/006	matic	yamaha X-Ride	petralite	yamalube
5	Wulan Purwasih	P	Cirebon, 15 Mei 1995	21 tahun	Islam	Main Musik	Lajang	J. griya cempaka arum	talun	wanasaba lor	006/004	matic	honda scopy	pertamax	enduro
6	Nurlia Fitri	P	Kuningan, 20 Februari 1996	20 tahun	Islam	Sport	Lajang	Blok cikadu desa sidawangi	sumber	sidawangi	006/003	matic	suzuki nex	pertamax	SGO
7	Lestiyawati	P	Cirebon, 12 Februari 1995	21 tahun	Islam	Jalan-Jalan	Lajang	J. pamtran gg. tambak no. 55	kejaksan	kejaksan	003/003	matic	yamaha soul	premium	yamalube
8	Agus Priyandri	L	Jakarta, 17 Agustus 1992	24 tahun	Islam	Main Musik	Lajang	J. jenda a yanki kersembaja	lemah wungkuk	segambiran	001/003	matic	yamaha mio	pertamax	yamalube
9	Rodiyah	P	Cirebon, 16 November 1993	21 tahun	Islam	Membaca	Janda	J. nyi gede cangkring	plared	bagasari	008/007	matic	honda beat	petralite	enduro
10	Tri Widiasari	P	Cirebon, 30 Maret 1994	22 tahun	Islam	Baca Novel	Lajang	Perum J. go bromo D. xiii	harjamukti	larangan	003/003	matic	yamaha mio j	pertamax	yamalube
11	Ibnu Alfian Refaldi	L	Tegal, 12 September 2000	16 tahun	Islam	Renang	Lajang	kepompong Blok. kavling Salam	talun	kepomongan	003/003	matic	yamaha mio z	petralite	yamalube
12	Abimaynu Rizki	L	Cirebon, 06 Agustus 2000	16 tahun	Islam	Berenang	Lajang	DS. sambene	leunene iati	leunene iati	006/004	matic	yamaha X-Ride	premium	yamalube

### 3.2 Hasil

#### Model Algoritma C4.5

Dalam penelitian ini akan dilakukan analisis menggunakan metode klasifikasi data mining. Metode yang diusulkan untuk pengolahan data pengguna oli (pelumas mesin motor) adalah penggunaan algoritma C4.5, serta di tingkatkan dengan teknik *Bagging*, setelah diolah dan menghasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan validation, dengan 10x validasi dari data akan di pakai untuk testing. Kemudian dilakukan evaluasi dan validasi hasil dengan confusion matrix dan kurva ROC. Tahap selanjutnya adalah membandingkan hasil akurasi dan RMSE serta klasifikasi error dari setiap model, sehingga diperoleh model dari

metode klasifikasi analisa penggunaan oli motor dengan nilai akurasi dan RMSE serta klasifikasi error. Dalam tahapan ini akan dilakukan beberapa langkah-langkah metode yang diusulkan yaitu: 1. Algoritma C4.5 yaitu salah satu algoritma dalam metode decision tree yang merubah data menjadi pohon keputusan menggunakan rumus perhitungan entropi. 2. Teknik Bagging yaitu peningkatan teknik klasifikasi. Dari hasil pengujian algoritma C4.5 didapatkan pohon keputusan dibawah ini :



Gambar 3. Pohon Keputusan C4.5

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar 4, akurasi C4.5 sebesar 44 %. dimana Jumlah *True Positive* (TP) adalah enduro 1, Yamalube 2 Untuk *False Negative* (FN) adalah yamalube 14, enduro 21, untuk *False Positive* (FP) adalah 14 dan Untuk *True Negative* (TN) adalah 21.

accuracy: 44.00% +/- 17.44% (mikro: 44.00%)								
	true enduro	true yamalube	true SGO	true primaXP	true TOP 1	true castrol	true mpx2	class precision
pred. enduro	1	2	0	0	0	0	0	33.33%
pred. yamalube	14	21	1	2	2	4	1	46.67%
pred. SGO	0	0	0	0	0	0	0	0.00%
pred. primaXP	0	0	0	0	0	0	0	0.00%
pred. TOP 1	1	0	0	0	0	0	0	0.00%
pred. castrol	0	0	0	0	0	0	0	0.00%
pred. mpx2	0	1	0	0	0	0	0	0.00%
class recall	6.25%	87.50%	0.00%	0.00%	0.00%	0.00%	0.00%	

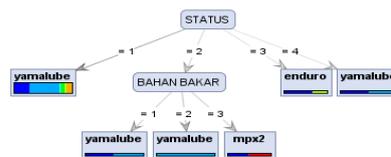
Gambar 4. Hasil Akurasi Algoritma C4.5

Tabel 4. Tabel Nilai Accuracy, RMSE dan Klasifikasi Error Algoritma C4.5

Nilai (%)	
Accuracy	44
RMSE	0,687
Klasifikasi Error	56

**Model Algoritma C4.5 Bagging**

Berdasarkan hasil pengujian algoritma C4.5 bagging didapatkan pohon keputusan dibawah ini :



Gambar 5. Pohon Keputusan C4.5 Bagging

Confusion matrix membentuk matriks yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negatif. Pada gambar 4, akurasi C4.5 bagging sebesar 46 %. dimana Jumlah *True Positive* (TP) adalah enduro 2, Yamalube 3 Untuk *False Negative* (FN) adalah yamalube 14, enduro 3, untuk *False Positive* (FP) adalah 14 dan Untuk *True Negative* (TN) adalah 21.

	true enduro	true yamalube	true SGO	true primaXP	true TOP 1	true castrol	true mpx2	class prec
pred. enduro	2	3	0	0	0	0	0	40.00%
pred. yamalube	14	21	1	2	2	4	1	46.67%
pred. SGO	0	0	0	0	0	0	0	0.00%
pred. primaXP	0	0	0	0	0	0	0	0.00%
pred. TOP 1	0	0	0	0	0	0	0	0.00%
pred. castrol	0	0	0	0	0	0	0	0.00%
pred. mpx2	0	0	0	0	0	0	0	0.00%
class recall	12.50%	87.50%	0.00%	0.00%	0.00%	0.00%	0.00%	

Gambar 6. Hasil Akurasi Algoritma C4.5 Bagging

**Tabel 5. Tabel Nilai Accuracy, RMSE dan Classification Error Algoritma C4.5 Bagging**

Nilai (%)	
Accuracy	46
RMSE	0,690
Classification Error	54

#### 4. KESIMPULAN

Berdasarkan hasil yang diperoleh dari penelitian yang telah dilakukan maka dapat disimpulkan telah diperoleh analisa dan prediksi pemilihan pelumas mesin motor (oli) menggunakan perbandingan algoritma C4.5 dan C4.5 berbasis Bagging, dihasilkan beberapa petunjuk bahwa variabel yang paling tinggi pengaruhnya terhadap pemilihan pelumas mesin motor (oli) adalah status, dan rata – rata Bahan bakar Hal ini dibuktikan dengan variabel rata – rata status menempati sebagai simpul akar pada node didalam diagram akar yang terbentuk. Berdasarkan perhitungan akurasi, RMSE dan Klasification Error, algoritma C4.5 Bagging lebih sesuai dibandingkan dengan metode C4.5, nilai *perhitungan akurasi 46%, Klasification Error sebesar 54, dan RMSE sebesar 0,690* maka algoritma C4.5 Bagging lebih baik dibandingkan dengan algoritma C4.5 dengan nilai akurasi 46%.

#### UCAPAN TERIMAKASIH

Ucapan terimakasih kepada bagian LPPM terutama pihak kampus ST3 Telkom Purwokerto, yang telah mendanai publikasi paper penelitian ini, saya berdoa mudah – mudahan dapat bermanfaat bagi pembaca (khususnya) dan masyarakat pada (umumnya).

#### DAFTAR PUSTAKA

- Amalia, R. (2007). Analisis Algoritma C4.5 dengan Bosoting Untuk Klasifikasi Teks pada SMS Customer Support System (CSS). *Telkom University, 2*(Data Mining), 1–8.
- Ari Wibowo, A. P. (2012). Penerapan Bagging Untuk Memperbaiki Hasil Prediksi Nasabah Perusahaan Asuransi X. *Informatika, 2*(Bagging), 1–8.
- Fitriyani, R. S. W. (2015). Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes. *Software Engineering, 1*(Naive Bayes), 101–108.
- Fransiska Amalia Kurniawan, D. (2011). Analisis dan Implementasi Random Forest dan Regression Tree (CART) Untuk Klasifikasi pada Misuse Intrusion Detection System. *Fakultas Teknik Informatika, (Data Mining)*, 1–7.
- Made Leo Radhitya, A. H. (2016). Sistem Informasi Geografis Risiko Kemunculan Rip Current Menggunakan Decision Tree C4.5. *IJCCS, 10*(Decision Tree C4.5), 1–12.
- Putri, S. A. (2017). Integrasi Teknik Smote Bagging Dengan Information Gain Pada Naive Bayes Untuk Prediksi Cacat Software. *Ilmu Pengetahuan Dan Teknologi Komputer, 2*(Prediksi Cacat Software), 22–31.
- Rizki Tri Prasetyo, P. (2015). Penerapan Teknik Bagging Pada Algoritma Klasifikasi Untuk Mengatasi Ketidak Seimbangan Kelas Dataset Medis. *Informatika, 2*(Bagging), 395–403.
- Safuan, D. (2015). Penanganan Fitur Kontinyu dengan Feature Discretization Berbasis Expectation Maximization Clustering untuk Klasifikasi Spam Email Menggunakan Algoritma ID3. *Journal of Intelligent Systems, 1*(Data Mining), 148–155.
- Syarif, M. (2015). Penerapan Particle Swarm Optimization Untuk Menentukan Kredit Kepemilikan Rumah Dengan Menggunakan. *Tehno Nusa Mandiri, 12*(Algoritma C4.5), 47–56.
- Wibowo, A. (2015). Analisis Perbandingan Kinerja Metode Klasifikasi Dalam Data Mining. *Integrasi, 7*(Data Mining), 23–30.
- Yusuf s, Nugroho, D. (2015). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. *Ilmu Komputer Dan Informatika, 1*(Klasifikasi dan Klastering), 3–8.