



### ANALISIS PERBANDINGAN KINERJA DECISION TREE DAN NAIVE BAYES DALAM PENENTUAN PENYEBAB KERUSAKAN RADIATOR

Muhtajuddin Danny, Nurhadi Surojudin

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa  
utat@pelitabarung.ac.id | nurhadi@pelitabarung.ac.id

#### Abstrak

Klasifikasi dalam data mining merupakan salah satu teknik yang bertujuan untuk menempatkan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Saat ini kerusakan radiator diperiksa dengan cara manual, radiator setelah diproduksi dikirim ke customer dan dalam masa penggunaannya terkadang terjadi kerusakan.

Penelitian ini difokuskan untuk mengetahui kinerja terbaik dari beberapa algoritma klasifikasi dalam data mining yaitu decision tree J48 dan naïve bayes. Evaluasi dilakukan berdasarkan pada perbandingan tingkat akurasi dengan menggunakan true positive dan false positive dalam confusion matrix yang dihasilkan dari masing-masing algoritma serta menggunakan correct dan incorrect instance untuk mengetahui metode yang paling efisien dari kedua algoritma tersebut.

Hasil percobaan menunjukkan bahwa algoritma decision tree J48 lebih baik jika dibandingkan dengan naïve bayes sehingga dapat disimpulkan jika algoritma decision tree J48 lebih efisien untuk menyelesaikan penentuan penyebab kerusakan radiator.

**Kata kunci :** Kerusakan Radiator, *Data Mining, Decision Tree, Naïve Bayes*

#### Abstract

*Classification in data mining is one technique that aims to place objects into one of several categories that have been set previously. Currently radiator damage is checked manually, radiator after production is sent to the customer and in the period of its use sometimes damage occurs.*

*This research is focused to know the best performance of some classification algorithm in data mining that is decision tree J48 and naïve bayes. The evaluation is based on a comparison of accuracy levels using true*

*positive and false positives in the confusion matrix generated from each algorithm and using the correct and incorrect instance to find out the most efficient method of the two algorithms.*

*The experimental results show that decision tree J48 algorithm is better compared to naïve bayes so it can be concluded if decision tree J48 algorithm is more efficient to solve the determination of cause of radiator damage.*

**Keywords:** *Radiator Damage, Data Mining, Decision Tree, Naïve Bayes.*

#### 1. Pendahuluan

Data mining merupakan serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Dalam data mining terdapat banyak teknik dalam pengerjaannya,

diantaranya yaitu algoritma decision tree, naïve bayes, jaringan saraf tiruan dan masih banyak lainnya.

Radiator merupakan salah satu pendingin mesin dengan air, pada sistem pendingin air (water cooling system), panas dari pembakaran gas dalam silinder pada sistem pendingin air sebagian diserap oleh air pendingin. Secara prinsip penukar panas (heat exchanger). Panas hasil pembakaran akan diserap oleh air pendingin yang disirkulasikan masuk

radiator. Air pendingin dalam radiator didinginkan oleh udara. Udara melewati radiator karena adanya kipas udara.

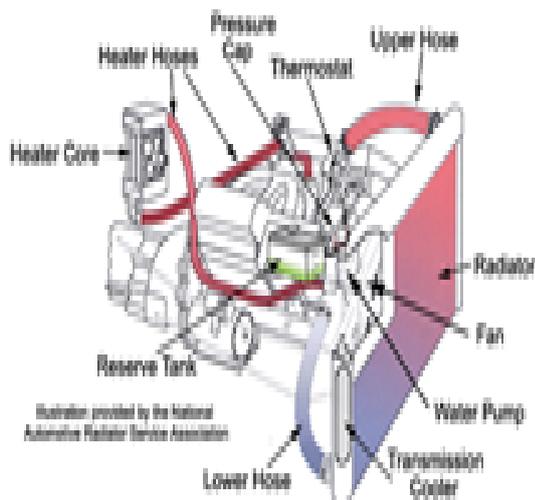
Radiator setelah diproduksi dikirim ke customer dan dalam masa penggunaannya terkadang terjadi kerusakan. Saat ini penyebab kerusakan radiator diperiksa dengan cara manual. Penulis akan membandingkan tingkat akurasi untuk algoritma decision tree dan naïve bayes. Penulis akan mengevaluasi hasil dari kedua algoritma tersebut untuk penentuan penyebab kerusakan radiator. Tujuan penulisan ini adalah untuk membandingkan keakuratan dengan algoritma decision tree dan algoritma naïve bayes untuk penentuan penyebab kerusakan radiator.

## 2. Tinjauan Pustaka

### 2.1. Sistem Pendingin Air

Sistem pendingin air secara garis besar ada dua macam yaitu :

1. Sistem alam (*Natural circulation*)  
Sistem pendingin ini terjadi akibat perbedaan berat jenis air panas dengan air dingin
2. Sirkulasi dengan tekanan  
Sirkulasi jenis ini hampir sama dengan sirkulasi dengan aliran hanya ditambahkan pompa air untuk mempercepat terjadinya sirkulasi air pendingin



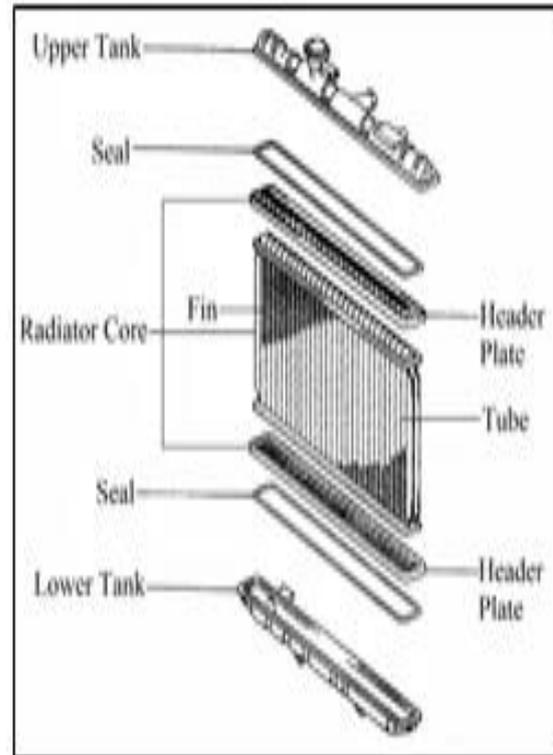
**Gambar 1.** Skema Sistem Pendingin Air

Sumber : NARSA (National Automotive Radiator Service Association)

Sistem pendingin air memiliki komponen-komponen yang bekerja secara integrasi satu dengan yang lainnya, dimana komponen-komponen tersebut antara lain: radiator, pompa air, kipas (fan), katup thermostat, mantel pendingin (water jacket) dan cairan pendingin.

### 2.2. Komponen Radiator

Radiator berfungsi untuk menampung dan mendinginkan cairan pendingin yang telah menjadi panas setelah menyerap dari komponen mesin. Radiator terdiri dari beberapa komponen seperti Gambar 2.2 yang mempunyai fungsi masing-masing sebagai berikut:



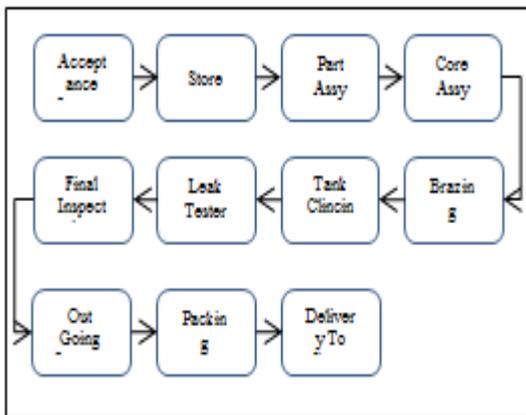
**Gambar 2.** Komponen Radiator

- *Upper Tank.*  
Tangki atas atau *upper tank* berfungsi untuk transit air panas yang berasal dari blok mesin.
- *Seal.*  
Berfungsi untuk mencegah kebocoran dari *upper tank* maupun *lower tank*.
- *Radiator Core.*  
Inti radiator atau *radiator core* berfungsi untuk menyerap panas pada air pendingin yang telah panas, yang kemudian akan didinginkan oleh kipas dan udara luar yang timbul karena kendaraan berjalan.
- *Header Plate.*  
Berfungsi menyatukan *fin* dan *tube*.
- *Fin.*  
Sirip pendingin atau *fin* pada inti radiator berfungsi untuk membuang panas air yang berada pada *tube* saat kipas menyala dan kendaraan berjalan.
- *Tube.*  
Pipa air atau *tube* pada inti radiator berfungsi untuk mengalirkan air dari *upper tank* ke *lower tank*.

- *Lower Tank*.  
Tangki bawah atau *lower tank* berfungsi untuk menampung air yang telah didinginkan oleh kipas dan udara melalui inti radiator dan selanjutnya disalurkan ke mesin melalui *water pump*.

**2.3. Proses Pembuatan Radiator**

Dilihat dari tankinya radiator dibedakan menjadi 2 tipe, yaitu tipe plastic tank dan aluminium tank, berikut flow proses pembuatan radiator tipe plastic tank seperti pada Gambar 2.3

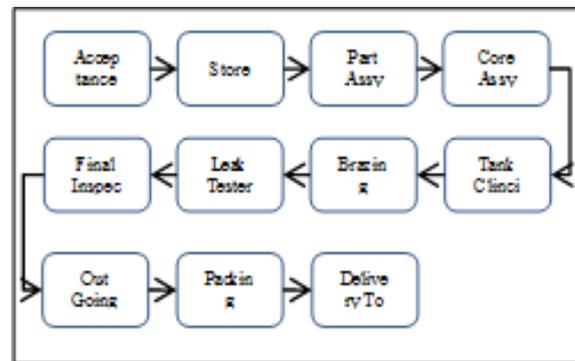


**Gambar 3.** Flow Proses Pembuatan Radiator Plastic Tank

1. *Acceptance Inspection*  
Merupakan proses pemeriksaan *part* dari *supplier* yang dilakukan oleh *inspector quality control departement* sebelum *part* masuk ke proses *assembly*
2. *Store*  
Merupakan tempat penyimpanan *part* OK hasil pemeriksaan *inspector*
3. *Part Assy*  
Merupakan proses *assy part* satu dengan *part* lainnya
4. *Core Assy*  
Merupakan proses *assy fin, tube* dan *header plate*
5. *Brazing*  
Merupakan proses perekatan *core radiator* dengan mesin pemanas suhu tinggi yang berfungsi untuk merekatkan *part* satu dengan *part* lainnya agar tidak bocor.
6. *Tank Clincing*  
Merupakan proses penyatuan *tanki radiator* dengan *core radiator* baik *tanki atas* maupun *tanki bawah*.
7. *Leak Tester*  
Merupakan proses pengecekan kebocoran radiator yang dilakukan 100% dengan mesin *leak tester*
8. *Final Inspection*  
Merupakan proses visual cek radiator yang dilakukan oleh *inspector*
9. *Out Going Inspection*

10. *Packing*  
Merupakan proses pengepakan radiator yang siap di kirim ke *customer*
11. *Delivery To Customer*  
Radiator yang telah dipastikan OK dikirm ke *customer*.

Untuk flow proses pembuatan radiator aluminium tank seperti Gambar 2.4



**Gambar 4.** Flow Proses Pembuatan Radiator Aluminium Tank

Proses pembuatan radiator kedua tipe tersebut yang membedakan adalah pada proses *tank clincing*, dimana untuk tipe *plastic tank* proses *tank clincing* dilakukan setelah *brazing*. Untuk tipe *aluminium tank* proses *clincing* dilakukan sebelum proses *brazing* karena tidak menggunakan *seal*.

**2.4. Jenis Kerusakan Radiator**

Radiator yang telah dikirim ke *customer* ada beberapa penyebab kerusakan pada radiator diantaranya sebagai berikut :

- **Korosi**  
Kerusakan radiator disebabkan karena korosi dapat dilihat dari bentuk kerusakannya antar lain : *tube* bocor yang terkikis dari dalam dan didalam radiator terdapat serbuk putih. Kerusakan radiator disebabkan karena korosi terjadi pada radiator tipe tangki plastik dan tipe tanki aluminium.
- **Getaran Berlebih**  
Radiator rusak disebabkan karena getaran berlebih dapat dilihat dari kerusakannya antara lain : *tube* bocor, *fin* pecah dan *tube* retak. Kerusakan radiator disebabkan karena getaran berlebih terjadi pada radiator tipe tangki plastik dan tipe tangki aluminium.
- **Seal Abnormal**  
Kerusakan radiator disebabkan karena *seal abnormal* dapat dilihat dari bentuk kerusakannya antara lain : *tank header plate* bocor. Kerusakan radiator disebabkan karena *seal abnormal* terjadi pada radiator tipe tangki plastik.

**2.5 Data Mining**

Data mining adalah serangkaian proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar (Prasetyo, 2012). Menurut Giudici & Figini (2009) data mining merupakan proses terpadu dari analisis data yang terdiri dari serangkaian kegiatan yang berjalan berdasarkan definisi tujuan yang akan dianalisis, dengan analisis datanya sampai interpretasi dan evaluasi hasil, berbagai tahapan dari proses data mining adalah sebagai berikut:

1. Definisi tujuan untuk analisis  
Suatu pernyataan yang jelas tentang masalah dan tujuan yang akan dicapai merupakan hal yang paling penting dalam membentuk analisis dengan benar. Hal ini menjadi salah satu bagian paling sulit dari proses karena menentukan metode yang akan digunakan. Oleh karena itu tujuan harus jelas dan tidak boleh ada ruang untuk keraguan atau ketidakpastian.
2. Seleksi, organisasi, dan pra-perawatan data  
Setelah tujuan dianalisa dan diidentifikasi perlu adanya pengumpulan atau pemilihan data yang diperlukan untuk analisis. Sumber data yang ideal adalah data *warehouse* perusahaan, sebuah “*store room*” data histori yang sudah tidak lagi digunakan. Jika tidak ada data *warehouse* maka pasar data dapat diciptakan dari *overlapping* dari berbagai sumber data perusahaan.
3. Eksplorasi analisis data dan mentransformasinya  
Pada tahap ini melibatkan analisis eksplorasi awal dari data, yang sangat serupa dengan teknik *Online Analytical Process (OLAP)*. Tahap ini dapat mengakibatkan transformasi dari variabel asli dalam rangka lebih memahami fenomena atau metode statistik yang digunakan. Analisis eksplorasi dapat menyoroti data *anomaly*, data yang berbeda dari data yang lain.
4. Spesifikasi dari metode statistik  
Berbagai macam metode statistik yang dapat digunakan dan juga banyak algoritma yang tersedia, sehingga sangat penting membuat klasifikasi dari metode yang sudah tersedia. Pemilihan metode yang akan digunakan dalam membuat analisa tergantung dari masalah yang sedang dipelajari atau pada jenis data yang tersedia. Berbagai macam metode tersebut dikelompokkan menjadi dua kelas utama sesuai dengan tahap yang berbeda dari analisis data yaitu:
  - a. Metode Deskriptif  
Tujuan utama dari metode di kelas ini adalah untuk menggambarkan kelompok data dengan cara yang ringkas. Dalam metode deskriptif tidak ada hipotesis sebab akibat antara variabel-variabel yang tersedia. Yang termasuk dalam kelompok

ini misalnya metode asosiasi, model *log-linear*, model grafis).

- b. Metode Prediksi  
Metode dalam kelas ini mempunyai tujuan untuk menggambarkan satu atau lebih variabel yang dilakukan dengan mencari aturan klasifikasi atau prediksi berdasarkan data. Aturan-aturan ini membantu memprediksi atau mengklasifikasikan hasil masa depan salah satu atau lebih variabel respon atau target dalam kaitannya dengan apa yang terjadi pada variabel penjelas atau masukan. Yang termasuk dalam metode ini antara lain *neural network* dan *decision tree* dan termasuk juga model regresi linear dan logistik.
5. Analisis data berdasarkan metode yang dipilih  
Setelah menentukan metode statistik yang akan digunakan selanjutnya menerjemahkan ke dalam algoritma yang sesuai untuk mendapatkan hasil yang dibutuhkan berdasarkan data yang tersedia.
  6. Evaluasi dan perbandingan dari metode yang digunakan dan pilihan model akhir untuk analisis.

Interpretasi dari model yang dipilih dan penggunaannya dalam *decision process*.

#### 2.5.1 Algoritma *Decision Tree*

Menurut Maimon dan Rokach (2014), pohon keputusan (*decision tree*) adalah salah satu metode klasifikasi yang dinyatakan sebagai partisi rekursif. Pohon keputusan terdiri dari node yang membentuk pohon yang berakar, semua node memiliki satu masukan. Node yang keluar disebut node tes. Node yang lain disebut node keputusan atau sering disebut node daun. Setiap simpul internal membagi dua atau lebih sub-ruang sesuai dengan kategori atribut dan akan dipartisi sesuai dengan nilai kategori kasus. Kasus-kasus tersebut membentuk pohon keputusan, yang menghasilkan *problem solving*.

Pohon keputusan adalah sebuah flowchart seperti struktur tree, dimana tiap internal node menunjukkan sebuah tes pada sebuah atribut, setiap cabang menunjukkan hasil dari test, dan leaf node menunjukkan kelas. Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah root node (titik awal) yang dipakai oleh user untuk mengambil tindakan. Dari root node ini, user memecahkan sesuai dengan algoritma *decision tree*. Hasil akhirnya adalah sebuah pohon keputusan dengan setiap cabangnya menunjukkan kemungkinan skenario dari keputusan yang diambil serta hasilnya.

Metode penelitian pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang dihubungkan dengan cabang-cabang dari simpul akar sampai ke node daun

(akhir). Pada node keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan.

Decision tree merupakan metode yang paling efisien untuk menyaring sesuatu lewat pohon keputusan apakah suatu data lolos atau tidak terhadap saringan dengan proses yang cukup cepat. Ada beberapa tahap dalam membuat sebuah pohon keputusan yaitu : (Saputra, 2014)

1. Menyiapkan data training yang sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar diambil dari atribut yang dipilih, yaitu dengan cara menghitung nilai *gain* yang tertinggi akan menjadi akar pertama. Sebelum menghitung *gain* dari atribut hitunglah nilai *entropy*.

Hal yang harus dilakukan dalam metode *decision tree* adalah menghitung *entropy* dan *information gain*. Untuk menghitung nilai *entropy* digunakan rumus:

$$Entropy(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_n \log_2 p_n \dots \dots \dots (2.1)$$

Dimana :

*Entropy(y)* : Ruang (data) sample yang digunakan untuk training

$p_1, p_2, \dots, p_n$  : Masing-masing menyatakan proporsi kelas 1, kelas 2, ..., kelas *n* dalam output

2.5.2 Algoritma *Naïve Bayes*

*Naïve Bayes* merupakan salah satu metode klasifikasi yang dapat memprediksi probabilitas keanggotaan dari suatu *class*. Nilai dari suatu *class* di metode *Naïve Bayes* ini bersifat independen, tidak tergantung dengan atribut-atribut lainnya (Han, Kamber, Pei, 2012 ). Berikut merupakan persamaan dari teorema *bayes*

$$\frac{P(H|X)}{P(X)} = \frac{P(X|H).P(H)}{P(X)} \dots \dots \dots (2.2)$$

Dimana :

X: Data dengan class belum diketahui

H: Hipotesis data X

P(H|X): Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

P(H): Probabilitas awal (*priori*) H

P(X|H): Probabilitas X berdasar kondisi pada hipotesis H

P(X): Probabilitas awal (*priori*) X

2.5.3 Akurasi

Nilai *accuracy* merupakan prosentase jumlah data yang diklasifikasi secara benar secara keseluruhan (Andriyani, 2013). Akurasi diperlukan untuk evaluasi dan mengukur keakuratan dari hasil klasifikasi, semakain besar nilai akurasi maka

semakin baik tingkat klasifikasinya. Dan untuk menghitung nilai akurasinya dengan rumus:

$$Akurasi = \frac{Jumlah\ Klasifikasi\ Benar}{Jumlah\ Data\ Uji} \times 100\% \dots \dots \dots (2.3)$$

2.6 WEKA

WEKA adalah sebuah paket tools machine learning praktis. WEKA merupakan singkatan dari Waikato Environment for Knowledge Analysis, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi (Cunningham, 1995). WEKA mampu menyelesaikan masalah-masalah data mining di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan machine learning. Perangkat lunak ini ditulis dalam hirarki class Java dengan metode berorientasi objek dan dapat berjalan hampir di semua platform.

WEKA mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia implementasi algoritma-algoritma pembelajaran state-of-the-art yang dapat diterapkan pada dataset dari command line. WEKA mengandung tools untuk pre-processing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. User dapat melakukan preprocess pada data, memasukkannya dalam sebuah skema pembelajaran, dan menganalisa classifier yang dihasilkan dan performansinya semua itu tanpa menulis kode program sama sekali.

Ada banyak metodologi Data Mining yang dapat dilakukan pada WEKA, salah satu yang populer adalah pohon keputusan (decision tree). Pohon keputusan atribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan.

3. Metodologi

Metode yang akan digunakan untuk menyelesaikan penelitian ini adalah jenis penelitian eksperimen yaitu melakukan pengujian tingkat akurasi terbaik antara algoritma decision tree dan *naïve bayes* dalam penentuan penyebab kerusakan radiator. Data eksperimen diambil dari data customer claim radiator suatu perusahaan manufacture.

3.1 Pengumpulan Data

Dalam penelitian ini metode pengumpulan data yang digunakan yaitu :

1. Metode Observasi  
Melakukan pengamatan langsung ke perusahaan *manufacture* untuk memperoleh data yang dibutuhkan.
2. Metode Wawancara  
Mengadakan wawancara dengan pihak-pihak yang berkaitan langsung dengan permasalahan yang sedang dibahas pada

penelitian ini untuk memperoleh gambaran dan penjelasan secara mendasar.

3. Metode Studi Pustaka

Metode ini dengan mengumpulkan referensi dari literature-literatur yang bisa mendukung penelitian sebagai landasan teori dan dasar pedoman dalam pembuatan laporan.

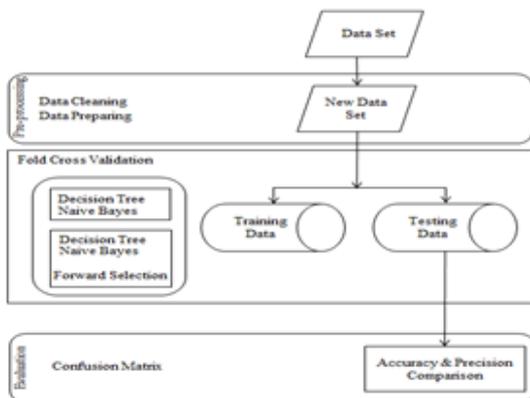
3.2 Perangkat Lunak

Setelah menentukan dataset untuk penelitian, kemudian mencari perangkat lunak yang digunakan untuk alat bantu penelitian. Dari sejumlah perangkat lunak dan dataset yang ada, peneliti memilih menggunakan perangkat lunak WEKA dan dataset customer claim radiator .

Weka adalah perangkat lunak yang beorientasi objek, open source berbasis java yang bisa digunakan untuk klasifikasi dalam penentuan kerusakan radiator dengan menganalisis dan membandingkan tingkat akurasi kedua algoritma dalam penelitian ini.

3.3 Alur Penelitian

Secara umum alur penelitian yang dilakukan mengacu pada kerangka peneltian seperti Gambar 3.1 berikut ini:



Gambar 5. Alur Penelitian

• Data Set.

No	Spesifikasi Pengukuran	Nilai
1	Correctly Classified Instances	69 atau 98.5714 %
2	Incorrectly Classified Instances	1 atau 1.4286 %
3	Kappa statistic	0.9664
4	Mean absolute error	0.014
5	Root mean squared error	0.0837
6	Relative absolute error	6.1873%
7	Root relative squared error	25.3274%
8	Total Number of Instances	70

Data set diambil dari data real disuatu perusahaan manufacture yaitu data customer claim

• New Data Set.

Data set dari data real disuatu perusahaan manufacture kemudian dilakukan pembersihan data untuk menentukan atribut.

• Training Data.

New data set kemudian dibagi menjadi 2 yaitu ; sebagai training data dan testing data. Data training digunakan untuk data preprocess

• Testing Data.

Testing data untuk proses klasifikasi dengan algoritma yang digunakan dalam penelitian ini

• Accuracy.

Proses klasifikasi akan menghasilkan tingkat akurasi suatu algoritma yang digunakan

4. Pembahasan dan Hasil

Bab ini membahas tentang hasil penelitian yang dilakukan mulai dari dataset yang digunakan dan hasil yang diperoleh. Dalam penelitian ini dilakukan percobaan terhadap dataset customer claim radiator dengan dua algoritma yang terdapat pada WEKA yaitu decision tree dan naïve bayes.

4.1. Dataset

Penelitian ini menggunakan dataset yaitu dataset customer claim radiator yang terdiri dari 8 atribut dataset yang digunakan :

Tabel 1. Atribut Dataset Penelitian

No	Nama Atribut	Deskripsi
1	Tipe	1, 2
2	Odometer	Dalam Kilometer
3	Fin Pecah	yes, no
4	Tube Retak	yes, no
5	Tank Header Plate Bocor	yes, no
6	TubeBocor	yes, no
7	Terdapat Serbuk Putih	yes, no
8	Penyebab	Korosi, Getaran Berlebih, Seal Abnormal

4.2 Hasil Evaluasi

4.2.1 Hasil Evaluasi Algoritma Decision Tree J48

Hasil evaluasi dataset customer claim radiator dengan menggunakan classifier tree J48

1) Training Set

Evaluasi klasifikasi data customer claim dengan pilihan use training set dengan menggunakan training data yang berjumlah 70 instance pada WEKA menghasilkan data pada Tabel 2.

Tabel 2. Hasil Evaluasi Training Set Tree J48

Tabel 2. menunjukkan akurasi yang diperoleh adalah 98,5714% dengan test record yang diklasifikasi secara benar sebanyak 69, jumlah test record yang diklasifikasi secara tidak benar sebanyak 1 atau 1,4286%, dengan hasil meansquared error adalah 0,0837%.

## 2) Supplied Test Set

Classifier yang telah terbentuk pada tahap training set selanjutnya diuji dengan menggunakan data test dengan 37 instance menghasilkan data yang dapat pada Tabel 4.3

Tabel 4.3 Hasil Evaluasi Supplied Test Set Tree J48

No	Spesifikasi Pengukuran	Nilai
1	<i>Correctly Classified Instances</i>	36 atau 97.2973 %
2	<i>Incorrectly Classified Instances</i>	1 atau 2.7027 %
3	<i>Kappa statistic</i>	0.9578
4	<i>Mean absolute error</i>	0.0179
5	<i>Root mean squared error</i>	0.1166
6	<i>Relative absolute error</i>	5.9367%
7	<i>Root relative squared error</i>	27.126%
8	<i>Total Number of Instances</i>	37

Tabel 4.3 menunjukkan akurasi yang diperoleh adalah 97,2973% dengan test record yang diklasifikasi secara benar sebanyak 36, jumlah test record yang diklasifikasi secara tidak benar sebanyak 1 atau 2,7027%, dengan hasil mean squared error adalah 0,1166%.

## 5. Kesimpulan

Berdasarkan hasil pengujian yang dilakukan oleh peneliti maka dapat disimpulkan sebagai berikut:

1. Dalam penelitian ini algoritma *decision tree J48* memiliki akurasi yang lebih besar dari pada algoritma *naïve bayes*.
2. Dengan menggunakan dataset yang sedikit dalam klasifikasi penentuan penyebab kerusakan radiator, algoritma *naive bayes* memiliki tingkat akurasi lebih baik.
3. Semakin banyak dataset yang digunakan pada algoritma *decision tree* memiliki tingkat akurasi tinggi, tetapi pada algoritma *naïve bayes* semakin banyak dataset yang digunakan memiliki tingkat akurasi rendah.
4. Algoritma *decision tree* dengan pilihan tes *cross validation* dengan jumlah *fold*s sebanyak 8, 10 dan 15 tingkat akurasinya sama.

5. Algoritma *naive bayes* dengan pilihan tes *cross validation* dengan jumlah *fold*s 8 dan *fold*s 15 memiliki tingkat akurasi lebih tinggi dibandingkan dengan jumlah *fold*s 10.
6. Atribut yang paling berpengaruh adalah terdapat serbuk putih.

## Daftar Pustaka

- [1] Andriani, Anik., (2013). Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: Amik "Bsi Yogyakarta". Seminar Nasional : Program Studi Manajemen Informatika, AMIK BSI Jakarta.
- [2] Ashari, A., Paryudi, I., Tjoa, A. M., "Performance comparison between naive bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool." IJACSA) International Journal of Advanced Computer Science and Applications 4.11 (2013).
- [3] Cunningham, S.J., (1995). "Machine learning and statistics: A matter of perspective". Working Paper Series 95/11, Departement of Computer Science, University of Waikato (Hamilton, New Zealand).
- [4] Dimitoglou, G., Adams, J.A., Jim, C.M., "Comparison of the C4. 5 and a naïve bayes classifier for the prediction of lung cancer survivability". Journal of Computing Press, ISSN (online): 2151-9617, Vol. 4 Issue 8 August 2012, New York USA.
- [5] Diwandari, Saucha., Setiawan, Noor Akhmad., (2015). Perbandingan Algoritma J48 Dan Naive Bayes Untuk Klasifikasi Diagnosa Penyakit Pada Soybean. Seminar Nasional : Program Studi Teknologi Informasi Dan komunikasi, Universitas Gadjah Mada.
- [6] Giudici., Figini ., (2009). Applied Data Mining for Business and Industry, 2nd Edition.
- [6] Han, Jiawei., Kamber, Micheline., Pei, Jian., (2012). Data Mining : Concepts and Techniques. 3rd ed. MK Publishers.
- [7] Maimon, O., Rokach, L., (2014). Data Mining with Decision Trees\_ Theory and Applications. 2nd ed.
- [8] Patil, T. R., Sherekar, S. S., "Performance analysis of naive bayes and J48 classification algorithm for data classification." International Journal of Computer Science and Applications 6.2 (2013): 256-261.
- [9] Patil, T. R., Sherekar, S. S., "Performance analysis of naive bayes and J48 classification algorithm for data classification." International Journal of Computer Science and Applications 6.2 (2013): 256-261.

- [10] Prasetyo, Eko., "Data Mining: Konsep dan Aplikasi Menggunakan Matlab." (2012). Journal of Computer Science Issues 9.1 (2012): 381-387.
- [11] Sani, Khairul., Winarno, Wing Wahyu., Fauziati, Silmi., 2016, Analisis Perbandingan Algoritma Clasification Untuk Authentication Uang Kertas: Studi Kasus Banknote Authentication, Jurnal Informatika Vol.10 No.1 Januari 2016.
- [12] Santra, A. K., and S. Jayasudha. "Classification of web log data to identify interested users using Naïve Bayesian classification." International
- [13] Saputra, Rizal Amegia., (2014). Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi. Seminar Nasional : Program Studi Manajemen Informatika, AMIK BSI Sukabumi.