



HIBRID METHOD MENGGUNAKAN DATA MINING DAN NAIVE BAYES MODEL UNTUK PREDIKSI STUDI KASUS KERUSAKAN LAMPU EFEK

Sarwo

Sekolah Tinggi Manajemen Informatika dan Komputer Mercusuar
sarwo@mercusuar.ac.id

Abstrak

Sudah banyak penelitian dengan berbagai metode untuk memprediksi kasus, dalam penelitian ini penulis menggunakan metode Naif Bayes, banyak juga menerapkan metode ini, dengan bermacam-macam studi kasus, data mining adalah kumpulan pengetahuan yang diperoleh dari satu set data, banyak pula peneliti di data mining dengan berbagai kasus dan metode.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode. "Data mining adalah analisis data survei set untuk menemukan hubungan tak terduga dan meringkas data dengan cara yang berbeda, yang dapat dipahami dan bermanfaat untuk pemilik data" (Larose, 2006)

dalam penelitian ini penulis menggabungkan dua metode adalah metode data mining metode standar Naif Bayes metode ini disebut METODE HYBRID DATA MINING DAN MODEL Naive Bayes UNTUK PREDIKSI STUDI KASUS KERUSAKAN RINGAN EFFECT, diharapkan oleh beberapa kombinasi dari kedua metode ini , akurasi dalam memprediksi efek kerusakan ringan dapat ditingkatkan, dalam penelitian ini penulis WEKA sebagai perangkat lunak untuk mengukur atau pengujian

Kata kunci: efek cahaya, data mining, Naif algoritma klasifikasi Bayes, Weka

Abstract

It's been a lot of research with a variety of methods to predict a case, in this study the authors use methods Naive Bayes, many also implement this method, with an assortment of case studies, data mining is a collection of knowledge that was obtained from a data set , many Similarly researchers in data mining with various cases and methods .

Data mining is the process of searching for a pattern or interesting information in the selected data by using techniques or methods. " Data mining is the analysis of the survey data sets to find unsuspected relationships and summarize data in a way that is different, that is understandable and useful to the data owner " (Larose , 2006)

in this study the authors combine the two methods is the method of data mining standard method Naive Bayes this method is called HIBRID METHOD OF DATA MINING AND Naive Bayes MODEL FOR

PREDICTION OF CASE STUDY OF DAMAGE LIGHT EFFECT, expected by some combination of these two methods , the accuracy in predicting light damage effects can be increased , in this study the authors WEKA as software for measuring or testing

Keywords : *light effect , data mining , Naive Bayes classifier algorithm, Weka*

1. Pendahuluan

1.1. Latar Belakang

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang bisnis, industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia bisnis juga dapat menghasilkan data yang

berlimpah mengenai data dan proses pengolahan data yang dihasilkan.

Dalam memprediksi kebutuhan *sparepart* dan jenis kerusakan lampu efek akan dimanfaatkan *dataset* yang sudah tersimpan kemudian dianalisa dan diringkas kedalam informasi yang berguna (Andyan, 2015), dimana informasi tersebut bisa digunakan untuk meningkatkan efektivitas dari efisiensi alur kerja pada divisi teknik. Proses penggalian informasi baru dari data set yang ada dinamakan dengan *data mining*. *Data Mining* adalah ekstraksi pola yang menarik dari data dalam jumlah besar. Salah satu teknik *data mining* adalah klasifikasi. Teknik klasifikasi yaitu membangun suatu model yang bisa mengklasifikasikan suatu objek berdasar atribut-atributnya. Kelas target sudah tersedia dalam data sebelumnya, sehingga fokusnya adalah bagaimana mempelajari data yang ada agar klasifikator bisa mengklasifikasikan sendiri.

Metode yang digunakan untuk penelitian ini adalah metode *Naïve Bayes*. Klasifikasi dengan metode *Naïve Bayes* memperlihatkan keakurasian yang tinggi dan kecepatan yang baik ketika di jalankan pada database yang besar. Maka dari itu berdasarkan latar belakang diatas diperlukan penelitian dengan judul **HIBRID METHOD MENGGUNAKAN DATA MINING DAN NAIVE BAYES MODEL UNTUK PREDIKSI STUDI KASUS KERUSAKAN LAMPU EFEK.**

1.2. Identifikasi

Dari permasalahan yang diperoleh, maka identifikasi masalahnya adalah sebagai berikut:

- Karena sulitnya memprediksi kebutuhan *sparepart*, *sparepart* yang dibutuhkan sering kali tidak ada sehingga proses pengerjaan servis menjadi lama karena harus menunggu pemesanan kebutuhan *sparepart* dari pihak luar perusahaan.
- Metode Data mining standart, memiliki tingkat akurasi yang rendah, sehingga diperlukan suatu metode agar hasilnya meningkat.

1.3. Ruang Lingkup

Dalam penulisan penelitian ini, Penulis membatasi masalah pada pembuatan sistem pengolahan data perbaikan lampu efek pada divisi teknik yang meliputi:

- Prediksi perbaikan lampu efek berdasarkan gejala kerusakan menggunakan algoritma *naïve bayes classifier*.
- Ketersediaan *sparepart* untuk proses perbaikan lampu efek berdasarkan hasil evaluasi kerusakan lampu efek.

1.4. Rumusan Masalah

Berdasarkan permasalahan, identifikasi dan ruang lingkup masalah diatas, maka rumusan masalahnya adalah:

- Bagaimana proses pengolahan data sehingga dapat menghasilkan informasi yang akurat dan berguna?
- Bagaimana penerapan data mining dengan algoritma *Naïve Bayes Classifier* untuk evaluasi kerusakan lampu efek?

1.5. Tujuan dan Manfaat

Dengan adanya penerapan *data mining* ini maka penulis mempunyai tujuan dan mendapatkan manfaat dari penulisan laporan skripsi ini, maka tujuan penelitian dan manfaat penulisannya sebagai berikut:

1.5.1. Tujuan

Tujuan yang ingin dicapai dari penulisan skripsi ini adalah untuk mendapatkan informasi yang berguna tentang prediksi waktu jenis kerusakan lampu efek berdasarkan gejala kerusakan dengan teknik *data mining*.

1.5.2. Manfaat

Manfaat yang diperoleh dari hasil penelitian ini yaitu:

- Memudahkan *user* dalam pengolahan data dan prediksi kebutuhan *sparepart* dan jenis kerusakan lampu efek
- Informasi yang didapatkan cepat dan akurat
- Dapat mengenali dan menganalisa kekurangan dan kelebihan dengan metode yang digunakan

1.6. Metodologi

- Metode *Naïve Bayes Classifier*
Suatu interpretasi dari kalkulus yang memuat konsep probabilitas sebagai derajat dimana suatu pernyataan dipercaya benar. Metode ini digunakan sebagai alat pengambilan keputusan untuk memperbaharui suatu informasi.
- Literatur
Untuk mendukung dalam proses penulisan, penulis memanfaatkan literatur-literatur yang berkaitan dengan tema skripsi baik dalam bentuk buku pustaka, informasi perusahaan obyek skripsi.

2. Landasan Teori

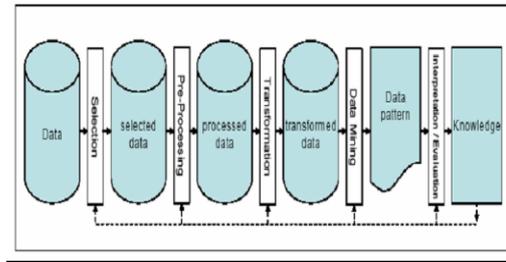
2.1. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. "*Data*

mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data” (Larose, 2006). “*Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar” (Larose, 2006).

Adapun proses KDD sebagai berikut:

- a. *Data Selection*
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.
- b. *Pre-processing/Cleaning*
Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi *eksternal*.
- c. *Transformation*
Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
- d. *Data mining*
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode dan algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
- e. *Interpretation/Evaluation*
Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Penjelasan di atas dapat direpresentasikan pada gambar berikut:



Gambar 1. Proses dari Data Mining

2.2. Operasi Data Mining

Operasi data mining menurut sifatnya dibedakan menjadi dua, yaitu bersifat prediksi (*prediction driven*) untuk menjawab pertanyaan apa dan sesuatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan, analisis multidimensi, OLAP (*Online Analytic Processing*) serta analisis statistik. Penemuan (*prediction driven*) bersifat transparan dan untuk menjawab pertanyaan “mengapa?”. Operasi penemuan digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi *database*, analisis keterkaitan dan deteksi deviasi.

2.3. Pengelompokkan Data Mining

Dengan definisi *data mining* yang luas, ada banyak jenis metode analisis yang dapat digolongkan dalam *data mining*. Metode-metode *data mining* diantaranya adalah:

a. Association Rules

Aturan asosiasi ingin memberikan informasi tersebut dalam bentuk hubungan “*if-then*” atau “jika-maka”. Aturan ini dihitung dari data yang sifatnya *probabilistic* (Santosa, 2007). Dalam asosiasi terdapat istilah *antecedent* dan *consequent*, *antecedent* untuk mewakili bagian “jika” dan *consequent* untuk mewakili bagian “maka”. Dalam analisis ini, *antecedent* dan *consequent* adalah sekelompok *item* yang tidak punya hubungan secara bersama (Santosa, 2007).

b. Decision Tree

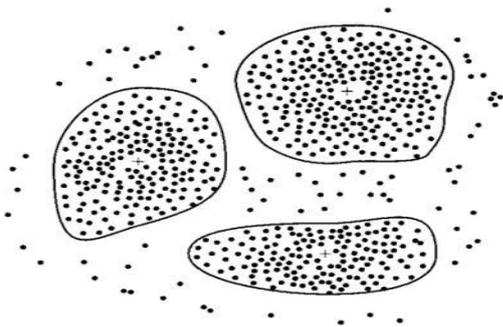
Decision tree sesuai digunakan untuk kasus-kasus yang keluarannya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri yang cocok untuk diterapkannya *decision tree* adalah sebagai berikut:

1. Data dinyatakan dengan pasangan atribut dan nilainya
2. Label/keluaran data biasanya bernilai diskrit

3. Data mempunyai *missing value* (nilai dari suatu atribut tidak diketahui).

c. Clustering

Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data/obyek ke dalam *cluster* (*group*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin seperti diilustrasikan pada gambar II.3. Dalam *clustering*, metode ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antar klaster sejauh mungkin.



Gambar 2. Clustering

Ini berarti obyek dalam satu *cluster* sangat mirip satu sama lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain. Dalam metode ini tidak diketahui sebelumnya berapa jumlah *cluster* dan bagaimana pengelompokannya (Santosa, 2007).

d. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data ke dalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis.

Klasifikasi dapat dilakukan dengan berbagai metode yang menghasilkan representasi model yang berbeda pula, antara lain:

1. Klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan kriteria pembagian kelas dan kelas-kelas. Beberapa algoritma yang menggunakan metode ini adalah algoritma ID3, SLIQ, SPRINT, dan PUBLIC.

2. Klasifikasi dengan *memory-based reasoning*, yaitu metode klasifikasi yang digabungkan dengan penalaran berbasis memori.
3. *Neural network*, yaitu metode klasifikasi yang menggunakan model jaringan syaraf tiruan (JST) yang direpresentasikan dalam bentuk neuron-neuron, bobot neuron dan struktur JST.
4. *Naive Bayes dan Bayesian Belief Networks*
5. *Klasifikasi* berdasarkan kaidah (*rule*). Metode ini terbagi menjadi dua; yang pertama mengekstrak kaidah secara langsung dari data (misalnya algoritma RIPPER), sedangkan yang kedua mengekstrak data dari model klasifikasi lain (misalnya algoritma C4.5) seperti pohon keputusan, *neural network*, dll.

e. Data Warehouse

Menurut Vidette Poe, “*data warehouse* merupakan *database* yang bersifat analisis dan *read only* yang digunakan sebagai fondasi dari sistem penunjang keputusan”.

Menurut Paul Lane, “*data warehouse* merupakan *database relasional* yang didesain lebih kepada query dan analisa dari pada proses transaksi, biasanya mengandung history data dari proses transaksi dan bisa juga data dari sumber lainnya. *Data warehouse* memisahkan beban kerja analisis dari beban kerja transaksi dan memungkinkan organisasi menggabung/konsolidasi data dari berbagai macam sumber”.

Dari definisi-definisi yang dijelaskan tadi, dapat disimpulkan *data warehouse* adalah *database* yang saling bereaksi yang dapat digunakan untuk *query* dan analisis, bersifat orientasi subjek, terintegrasi, *time-variant*, tidak berubah yang digunakan untuk membantu para pengambil keputusan.

f. Lampu Efek

Berdasarkan wawancara yang saya lakukan terhadap Yoka Afriyanto selaku *Programmer* lampu efek pada PT. CIPTA TPI (MNC TV) “lampu efek dapat diartikan sebagai lampu ada mesinnya. Lampu tersebut dapat berganti warna cahaya dan dapat berganti *gobo* (bentuk sinar sorot). Fungsi dari lampu efek merupakan membentuk sebuah atmosfer panggung agar tercipta sebuah suasana panggung sesuai kebutuhan produksi”.

Berdasarkan (Manual Book FineArt, 2009) standar umur penggunaan lampu efek adalah 750 jam. Kondisi ini diharuskan dilakukan perawatan dan pengecekan lampu efek. Lampu

efek *FineArt* menggunakan daya 1.200 W pada setiap unit yang digunakan.

g. Algoritma Naïve Bayes Classifier

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Larose, 2006).

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

Dimana Variabel *C* merepresentasikan kelas, sementara variabel *F1 ... Fn* merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (*Posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas *C* (disebut juga *likely hood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{evidence}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menggunakan aturan perkalian sebagai berikut:

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor -faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk (*F1, F2 ... Fn*) saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(P_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)}$$

Untuk *i≠j*, sehingga:

$$P(F_i|C, F_j) = P(F_i|C)$$

Dari persamaan diatas dapat disimpulkan bahwa asumsi independensi *naif* tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1|C)P(F_2|C)P(F_3|C)\dots \\ &= P(C)\prod_{i=1}^n P(F_i|C) \end{aligned}$$

Persamaan diatas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss*:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Keterangan:

P : Peluang

Xi: Atribut ke *i*

Xi: Nilai atribut ke *i*

Yi: Kelas yang dicari

yj: Sub kelas *Y* yang dicari

μ: Mean, menyatakan rata rata dari seluruh atribut

σ: Deviasi standar, menyatakan varian dari seluruh atribut

Teorema *Naive Bayesian* memiliki beberapa kelebihan dan kekurangan yaitu sebagai berikut:

Keuntungan *Naive Bayesian*:

- a. Menangani kuantitatif dan data diskrit

- b. Kokoh untuk titik noise yang diisolasi, misalkan titik yang dirata-ratakan ketika mengestimasi peluang bersyarat data.
- c. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
- d. Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang
- e. Cepat dan efisiensi ruang
- f. Kokoh terhadap atribut yang tidak relevan

Kekurangan Naive Bayesian:

- a. Tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga
- b. Mengasumsikan variabel bebas

h. Alat Bantu Implementasi

Tools WEKA

Weka adalah aplikasi *data mining open source* berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi/formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan *data mining* tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan (Susanto).

Tools yang dapat digunakan untuk *pre-processing* dataset membuat *user* dapat berfokus pada algoritma yang digunakan tanpa terlalu memperhatikan detail seperti pembacaan data dari file-file, implementasi algoritma *filtering*, dan penyediaan kode untuk evaluasi hasil.

Pengembangan WEKA mengikuti model *releases* Linux: digit kedua yang genap menunjukkan *release* yang stabil dan digit kedua yang ganjil menunjukkan *release* ‘pengembangan’ (misalnya 3.0.x adalah *release* stabil, sedangkan 3.1.x adalah *release* yang sedang dikembangkan). Beberapa versi awal dari WEKA:

- a. WEKA 3.0 : “versi buku” yang sesuai dengan deskripsi buku data mining.
- b. WEKA 3.2 : “versi GUI” yang menambahkan GUI dari CLI awal.

- c. WEKA 3.3 : “versi pengembangan” dengan berbagai peningkatan.

Praktek data mining ini menggunakan WEKA versi 3.6.9.



Gambar 3. Tampilan WEKA

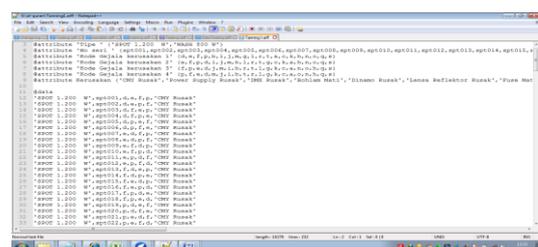
Empat tombol diatas dapat digunakan untuk menjalankan Aplikasi:

- a. *Explorer* digunakan untuk menggali lebih jauh data dengan aplikasi WEKA
- b. *Experimenter* digunakan untuk melakukan percobaan dengan pengujian statistik skema belajar
- c. *Knowledge Flow* digunakan untuk pengetahuan pendukung
- d. *Simple CLI* antar muka dengan menggunakan tampilan *command-line* yang memungkinkan langsung mengeksekusi perintah weka untuk Sistem Operasi yg tidak menyediakan secara langsung.

3. Analisa Dan Pembahasan

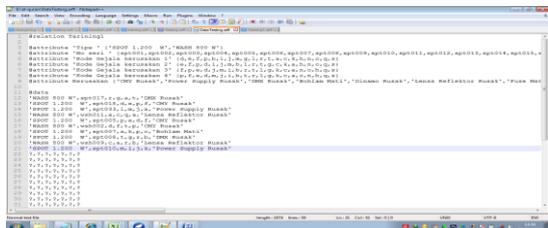
3.1. Analisis

Sebelum pengolahan data dalam WEKA dilakukan, data *training* dan data *testing* yang disimpan dalam Microsoft Excell dengan format .csv harus diubah dalam format .arff terlebih dahulu agar dapat digunakan pada aplikasi WEKA guna mengetahui beberapa informasi tentang tiap atribut. Berikut adalah data *training* yang berjumlah 192 *record* dan 7 atribut yang telah diubah menjadi format .arff yang dibuka dengan *Notepad++*:



Gambar 4. Data Training Format .Arff

Berikut adalah data *testing* yang telah diubah menjadi format *.arff* yang dibuka dengan *Notepad++*:



Gambar 5. Data *Testing* Format *.Arff*

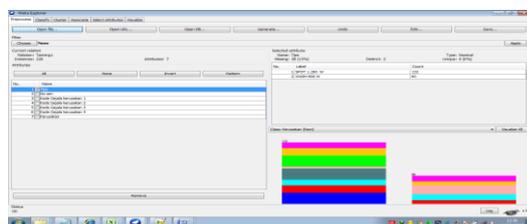
Dataset yang menjadi data *testing* yaitu 10 *record* dan memiliki atribut yang sama dengan data *training*.

3. 2. Preprocess

Sebelum proses data mining dilakukan, perlu dilakukan *preprocess* pada data *training* untuk memeriksa dan memperbaiki jika ada kesalahan cetak. Berikut tahapan *preprocess* data training berdasarkan atribut-atributnya.

a. Preprocess Atribut Tipe Lampu Efek

Data preprocess Atribut tipe lampu efek berbentuk *.arff* di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut tipe lampu efek:



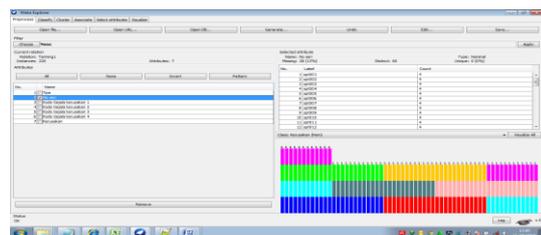
Gambar 6. Preprocess Atribut Tipe Lampu Efek

Pada gambar di atas dapat dilihat bahwa atribut Tipe Lampu Efek:

- 1. Spot 1.200 W = 132
- 2. Wash 800 W = 60

b. Preprocess Atribut No Seri

Data preprocess Atribut tipe no seri berbentuk *.arff* di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut tipe no seri:



Gambar 7. Preprocess Atribut nomor seri

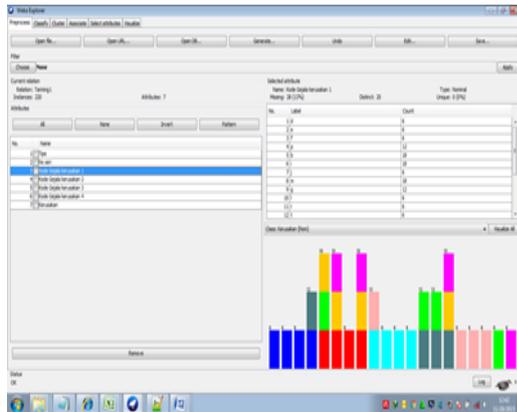
Pada gambar di atas dapat dilihat bahwa atribut nomor seri:

1. spt001 = 4	15. spt015 = 3	29. spt029 = 3
2. spt002 = 4	16. spt016 = 3	30. spt030 = 3
3. spt003 = 4	17. spt017 = 3	31. spt031 = 3
4. spt004 = 4	18. spt018 = 3	32. spt032 = 3
5. spt005 = 4	19. spt019 = 3	33. spt033 = 3
6. spt006 = 4	20. spt020 = 3	34. spt034 = 3
7. spt007 = 4	21. spt021 = 3	35. spt035 = 3
8. spt008 = 4	22. spt022 = 3	36. spt036 = 3
9. spt009 = 4	23. spt023 = 3	37. spt037 = 3
10. spt010 = 4	24. spt024 = 3	38. spt038 = 3
11. spt011 = 4	25. spt025 = 3	39. spt039 = 3
12. spt012 = 4	26. spt026 = 3	40. spt040 = 3
13. spt013 = 3	27. spt027 = 3	41. wsb001 = 3
14. spt014 = 3	28. spt028 = 3	42. wsb002 = 3
43. wsb003 = 3	49. wsb009 = 3	55. wsb015 = 3
44. wsb004 = 3	50. wsb010 = 3	56. wsb016 = 3
45. wsb005 = 3	51. wsb011 = 3	57. wsb017 = 3
46. wsb006 = 3	52. wsb012 = 3	58. wsb018 = 3
47. wsb007 = 3	53. wsb013 = 3	59. wsb019 = 3
48. wsb008 = 3	54. wsb014 = 3	60. wsb020 = 3

c. Preprocess Atribut Kode Gejala Kerusakan 1
 Data preprocess Atribut kode gejala kerusakan 1 berbentuk *.arff* di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut kode gejala kerusakan 1.

d. Preprocess Atribut Kode Gejala Kerusakan 1
 Data preprocess Atribut kode gejala kerusakan 1 berbentuk *.arff* di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut kode gejala kerusakan 1:

- 1. d = 6 6. i = 18 11. l = 12 16. h = 6
- 2. e = 6 7. j = 6 12. r = 6 17. n = 6
- 3. f = 6 8. m = 18 13. t = 6 18. o = 6
- 4. p = 12 9. g = 12 14. a = 6 19. q = 6
- 5. b = 18 10. l = 6 15. c = 12 20. s = 6

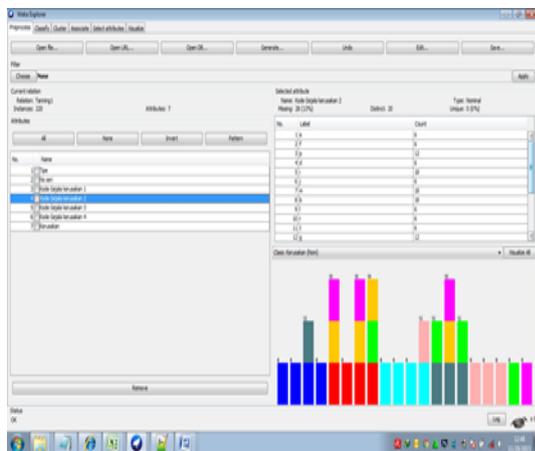


Gambar 6. Preprocess Atribut Kode Gejala Kerusakan 1

Pada gambar di atas dapat dilihat bahwa atribut gejala kerusakan 1:

- 1. d = 6 6. i = 18 11. l = 12 16. h = 6
- 2. e = 6 7. j = 6 12. r = 6 17. n = 6
- 3. f = 6 8. m = 18 13. t = 6 18. o = 6
- 4. p = 12 9. g = 12 14. a = 6 19. q = 6
- 5. b = 18 10. l = 6 15. c = 12 20. s = 6

- e. Preprocess Atribut Kode Gejala Kerusakan 2
Data preprocess Atribut kode gejala kerusakan 2 berbentuk .arff di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut kode gejala kerusakan 2:

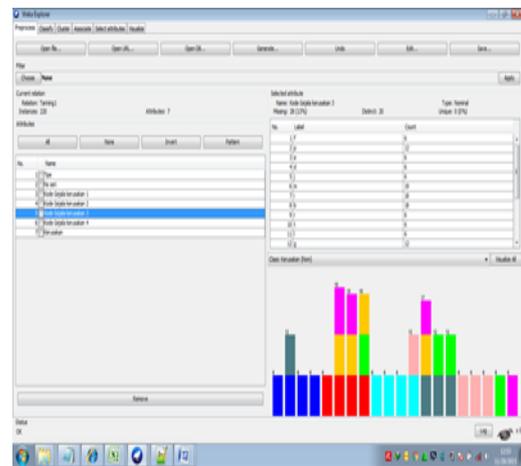


Gambar 7. Preprocess Kode Gejala Kerusakan 2

Pada gambar di atas dapat dilihat bahwa atribut gejala kerusakan 2:

- 1. e = 6 6. j = 6 11. t = 6 16. h = 6
- 2. f = 6 7. m = 18 12. g = 12 17. n = 6
- 3. p = 12 8. b = 18 13. c = 12 18. o = 6
- 4. d = 6 9. l = 6 14. k = 18 19. q = 6
- 5. i = 18 10. r = 6 15. a = 12 20. s = 6

- f. Preprocess Atribut Kode Gejala Kerusakan 3
Data preprocess Atribut kode gejala kerusakan 3 berbentuk .arff di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya. Berikut isi atribut preprocess atribut kode gejala kerusakan 3:



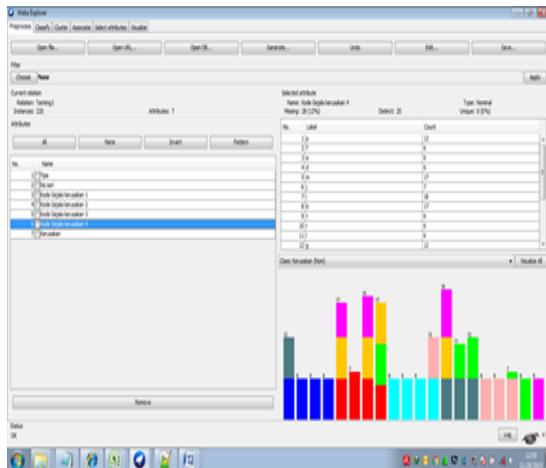
Gambar 8. Preprocess Atribut Kode Kerusakan 3

Pada gambar di atas dapat dilihat bahwa atribut gejala kerusakan 3:

- 1. f = 6 6. m = 18 11. g = 12 16. n = 2
- 2. p = 12 7. i = 18 12. k = 18 17. h = 2
- 3. e = 6 8. b = 18 13. c = 12 18. a = 2
- 4. d = 6 9. r = 6 14. a = 12 19. q = 2
- 5. j = 6 10. t = 6 15. n = 6 20. s = 2

- g. Preprocess Atribut Kode Gejala Kerusakan 4
Data preprocess Atribut kode gejala kerusakan 4 berbentuk .arff di masukan kedalam aplikasi weka dan dapat dilihat isi atribut didalamnya.

Berikut isi atribut preprocess atribut kode gejala kerusakan 4:

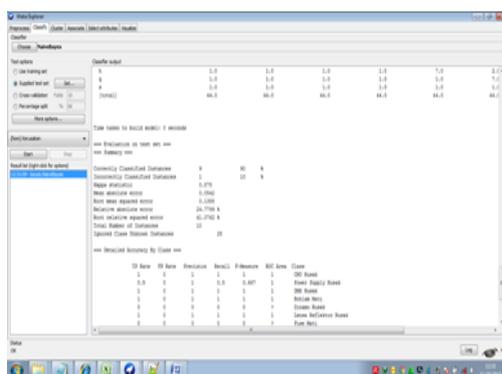


Gambar 9. Preprocess Atribut Kode Kerusakan 4

- 1. p = 12
- 2. f = 6
- 3. e = 6
- 4. d = 6
- 5. m = 18
- 6. j = 6
- 7. i = 18
- 8. b = 18
- 9. t = 6
- 10. r = 6
- 11. l = 6
- 12. g = 18
- 13. k = 6
- 14. c = 12
- 15. a = 12
- 16. o = 6
- 17. n = 6
- 18. h = 6
- 19. q = 6
- 20. s = 6

3.3. Hasil

Disini akan dilakukan proses prediksi kerusakan lampu efek berdasarkan gejala pada data testing yang sudah dipersiapkan berjumlah 10 record dengan metode *Naïve Bayes Classifier*.



Gambar 10. Hasil Prediksi Data Testing

Dari gambar di atas, dapat dilihat hasil presentase akurasi yaitu:

Persentase akurasi

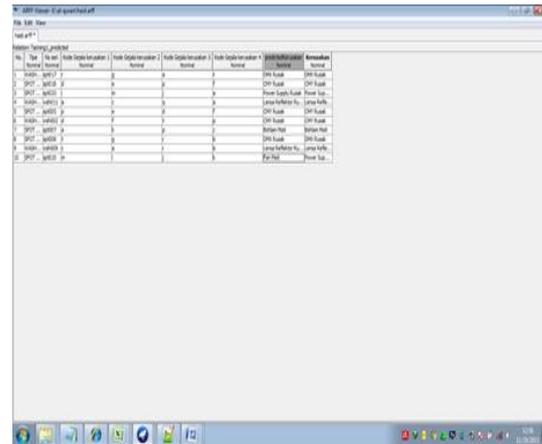
$$= \frac{\text{Banyak prediksi yang benar}}{\text{Total banyaknyadata}} \times 100\%$$

Persentase akurasi:

$$= 9/10 * 100\%$$

$$= 90\%$$

Untuk lebih jelasnya, hasil prediksi dapat dilihat pada ARFF-Viewer seperti gambar di bawah ini:



Gambar 11. ARFF-Viewer Hasil Data Testing

Dari gambar di atas dapat dibandingkan hasil prediksi kerusakan lampu efek menggunakan metode algoritma *Naïve Bayes Classifier* pada WEKA dengan perhitungan manual, bahwa didapatkan total prediksi yang benar sebesar 90%, dan kesalahan prediksi hanya sebesar.

4. Kesimpulan

Berdasarkan uraian dari bab-bab sebelumnya, maka penulis mengambil kesimpulan antara lain:

- a. Pengolahan data gejala kerusakan yang terjadi dapat dijadikan informasi yang akurat dalam mengevaluasi dan memprediksi kerusakan lampu efek.
- b. Penerapan *data mining* menggunakan algoritma *naïve bayes classifier* tepat dalam memprediksi dan mengevaluasi kerusakan lampu efek. Karena pada penerapannya berdasarkan probabilitas banyaknya gejala kerusakan yang terjadi sehingga menghasilkan hasil prediksi yang mendekati akurat.

Daftar Pustaka

[1] Bramer, M. (2007). *Principles of Data Mining*. London: Springer.

[2] Dahlan, U. A. (n.d.). Retrieved Oktober 17, 2015, from <http://id.portalgaruda.org/>

- ?ref=browse&mod=viewarticle&article=329081.
- [3] Larose, D. (2006). *Discovering Knowledge in database*. New jersey: John wiley and Sons INC.
- [4] *Manual Book FineArt*. (2009). Guangzhou: FineArt .Inc.
- [5] Munir, R. (2011). *Algoritma dan Pemrograman* (rev. ed.). Bandung: Informatika.
- [6] Pramudiono, I. (n.d.). Retrieved Oktober 15, 2015, from <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip>.