

Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naïve Bayes

Fattya Ariani¹, Amir², Nur Alam³, Khairul Rizal⁴

¹STMIK Nusa Mandiri
e-mail: fattya.fty@nusamandiri.ac.id

²AMIK BSI Pontianak
e-mail: amir.amr@bsi.ac.id

³AMIK BSI Jakarta
e-mail: nur.nra@bsi.ac.id

⁴AMIK BSI Sukabumi
e-mail: khairul.krl@bsi.ac.id

Cara Sitasi: Sunge, A. S. (2018). Optimasi Algoritma C4.5 Dalam Prediksi Web Phishing Menggunakan Seleksi Fitur Genetic Algoritma. *Paradigma*, XX(2), 27-32. doi:10.31294/p.v%vi%i.4021

Abstract - Salah satu asset yang berharga dalam sebuah perusahaan adalah sumber daya manusia (SDM). Karyawan adalah sdm tersebut. Karyawan menjadi peran penting. Dalam perusahaan besar pasti mempunyai banyak karyawan. Didalam perusahaan terdapat tiga jenis karyawan yaitu, *outsourcing*, karyawan kontrak dan karyawan tetap. Untuk mendapatkan karyawan tetap dilakukan proses penyaringan melalui penilaian yang dilakukan oleh bagian SDM. Dalam proses tersebut mengalami kendala dalam segi teknis dan efektifitas karena penilaian manual dan kadang keputusannya dinilai secara subyektif. Oleh karena itu dilakukan penelitian untuk klasifikasi penetapan status karyawan dengan metode Naïve Bayes dan Rapidminer. Penelitian ini menggunakan lima atribut yaitu : absensi, attitude, psikotes, wawancara dan diangkat. yang digunakan sebagai labelnya adalah atribut diangkat. Hasilnya adalah perhitungan manual dan penggunaan rapidminer menghasilkan nilai yang sama. Dengan nilai $c1(\text{diangkat})$ adalah 0.567 sedangkan untuk nilai $c2$ (tidak diangkat) adalah 0.433. Dan nilai yang tertinggi adalah diangkat. Jadi pola yang didapatkan dari proses ini untuk penetapan status karyawan dengan atribut dan nilai, absensi : bagus, attitude: baik, psikotes: tinggi, dan wawancara : baik.

Keywords: Naïve Bayes, klasifikasi, karyawan

PENDAHULUAN

Semakin besar sebuah perusahaan, maka semakin banyak pula jumlah karyawan yang bekerja diperusahaan tersebut. Dalam perusahaan biasanya terdiri dari beberapa macam pegawai, seperti karyawan *outsourcing*, karyawan kontrak dan karyawan tetap. Perusahaan kadang melakukan pengangkatan karyawan kontak atau tetap dalam jangka waktu satu tahun sekali. Bagian SDM melakukan penilaian terhadap hal tersebut. Penilaian dilakukan untuk menjaring karyawan yang berkompeten. Bagian SDM mengalami beberapa kendala dalam proses penilaian penetapan status dalam segi teknis ataupun efektifitas. Karena penilaian dinilai secara manual oleh kepala bidang. Proses ini memakan waktu lama dan keputusannya sering dinilai secara subyektif.

Klasifikasi adalah proses untuk menemukan atau fungsi yang menjelaskan atau membedakan konsep

atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan "jika-maka". Dalam teknik klasifikasi terdapat beberapa algoritma yang bisa digunakan antara lain decision tree, Naïve Bayes, adaptive Naïve Bayes, logistic regression dan support vector machine.

Untuk mengklasifikasi gejala diabetes secara cepat dan akurat, tentu diperlukan data-data yang valid dengan metode yang handal agar kesalahan dalam proses klasifikasi dapat diminimalisir. Lalu data tersebut tentu harus diproses agar bisa diterjemahkan menjadi diagnose (Rahman, Darmawidjadja, & Alamsah, 2017). Kinerja algoritma Bayesian lebih efisien dalam mengklasifikasikan Network IDS (NIDS) dibandingkan ANN. (Panda & Patra, 2007). Naïve Bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada

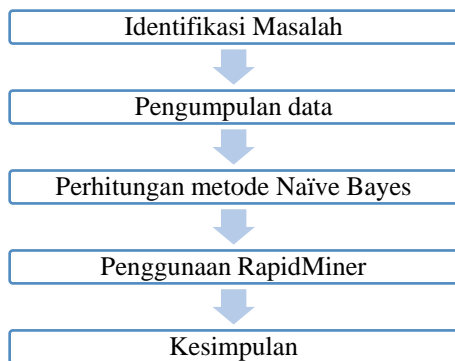
data training (Mukminin & Riana, 2017). Penerapan algoritma pengklasifikasi Naïve Bayes dapat mengurangi data *noise* pada *dataset* berukuran besar dan memiliki banyak kelas atau multi kelas sehingga akurasi klasifikasi algoritma decision tree dapat meningkat (Khadafy & Wahono, 2015). Pengklasifikasi Naïve Bayes sangat sederhana, efisien dan merupakan teknik machine learning yang populer untuk klasifikasi teks, serta memiliki performa yang baik pada banyak domain. Namun, Naïve Bayes memiliki kekurangan yaitu sangat sensitif pada fitur yang terlalu banyak, yang mengakibatkan akurasi klasifikasi menjadi rendah. (Muthia, 2014)

Rumus Naïve Bayes adalah:

$$f_c(E) = \underset{c \in C}{argmax} P(c) \prod_{j=1}^n P(E_j|c)$$

METODOLOGI PENELITIAN

Metode penelitian yang dilakukan dapat dilihat pada gambar 1.



Gambar 1. Kerangka Penelitian

Penelitian ini dilakukan dengan beberapa langkah :

1. Identifikasi Masalah
Masalah yang ada adalah menentukan pola klasifikasi penetapan status karyawan agar dapat mengembangkan keputusan yang baik dan tidak subjektif lagi dalam penentapannya.
2. Pengumpulan Data
Data yang digunakan adalah data yang didapat dari perusahaan pada tahun 2016. Yaitu data untuk penetapan status karyawan.
3. Pengolahan data manual menggunakan metode Naïve Bayes
Langkah selanjutnya adalah pengolahan data menggunakan metode Naïve Bayes. Langkah ini melakukan perhitungan manual dengan menggunakan rumus Naïve Bayes dari data yang telah didapat sebelumnya.
4. Pengolahan data menggunakan rapid miner

Lalu selanjutnya adalah penggunaan aplikasi rapid miner. Data yang didapat dimasukan kedalam aplikasi rapid miner dan di proses dengan operasi Naïve Bayes.

5. Kesimpulan

Kesimpulan akan dijlaskan setelah didapatkan hasil dari pengolahan data bak manual dan juga penggunaan rapid miner.

HASIL DAN PEMBAHASAN

Data mentah atau data mentah yang didapat :

Tabel 1. Data Penetapan Karyawan

Kary	Absen	Attitude	Psikotes	wawancara	Diangkat
A1	bagus	baik	tinggi	baik	ya
A2	bagus	baik	tinggi	baik	ya
A3	bagus	buruk	tinggi	baik	tidak
A4	bagus	baik	sedang	baik	ya
A5	cukup	baik	sedang	buruk	ya
A6	bagus	baik	rendah	buruk	tidak
A7	cukup	baik	tinggi	baik	ya
A8	cukup	baik	sedang	baik	ya
A9	cukup	baik	sedang	buruk	ya
A10	cukup	baik	rendah	buruk	tidak
A11	cukup	baik	sedang	baik	ya
A12	bagus	baik	sedang	buruk	ya
A13	cukup	baik	tinggi	buruk	tidak
A14	cukup	baik	tinggi	baik	ya
A15	cukup	buruk	tinggi	baik	tidak
A16	bagus	buruk	sedang	baik	tidak
A17	bagus	baik	sedang	buruk	ya
A18	bagus	baik	rendah	buruk	tidak
A19	cukup	baik	rendah	baik	ya
A20	cukup	buruk	tinggi	buruk	tidak
A21	bagus	baik	tinggi	baik	ya
A22	cukup	buruk	sedang	buruk	tidak
A23	bagus	baik	tinggi	buruk	ya
A24	cukup	baik	sedang	buruk	ya
A25	bagus	baik	sedang	baik	ya
A26	bagus	buruk	rendah	baik	tidak
A27	bagus	buruk	rendah	baik	tidak
A28	bagus	buruk	tinggi	baik	tidak
A29	bagus	baik	tinggi	baik	ya
A30	cukup	baik	rendah	buruk	tidak

Lalu ada empat atribut dan anggotanya

1. Absensi [bagus | cukup]
2. Attitude [baik | buruk]
3. Psikotes [tinggi | sedang | rendah]
4. Wawancara [baik | buruk]
5. diangkat [ya | tidak]

Gunakan sebagai labelnya adalah diangkat [ya | tidak]

A. Perhitungan Manual

Data yang didapat lalu di rekap seperti berikut:

Tabel 2. Rekap data

Diangkat	ya	17
	tidak	13
Absensi	bagus	16
	cukup	14
Attitude	baik	22
	buruk	8
psikotes	tinggi	12
	sedang	11
	rendah	7
Wawancara	baik	17
	buruk	13

Data yang ditentukan atributnya adalah :

Absensi = bagus

Attitude = bagus

Psikotes = tinggi

Wawancara = baik

Untuk diangkat: c1 = ya dan c2 = tidak

Perhitungannya adalah sebagai berikut :

$$fc1(E) = \frac{17}{30} \times \frac{16}{30} \times \frac{22}{30} \times \frac{12}{30} \times \frac{17}{30} = \frac{1220736}{24300000}$$

$$fc2(E) = \frac{13}{30} \times \frac{16}{30} \times \frac{22}{30} \times \frac{12}{30} \times \frac{17}{30} = \frac{933504}{24300000}$$

Untuk ya diangkat (C1)

$$\frac{1220736}{24300000} = \frac{\frac{1220736}{24300000}}{\frac{1220736}{24300000} + \frac{933504}{24300000}}$$

$$= \frac{1220736}{2154240}$$

$$= \frac{1220736}{24300000} \times \frac{24300000}{2154240} = \frac{1220736}{2154240} = 0,567$$

Untuk tidak (C2)

$$\frac{933504}{24300000} = \frac{\frac{933504}{24300000}}{\frac{1220736}{24300000} + \frac{933504}{24300000}}$$

$$= \frac{1220736}{24300000}$$

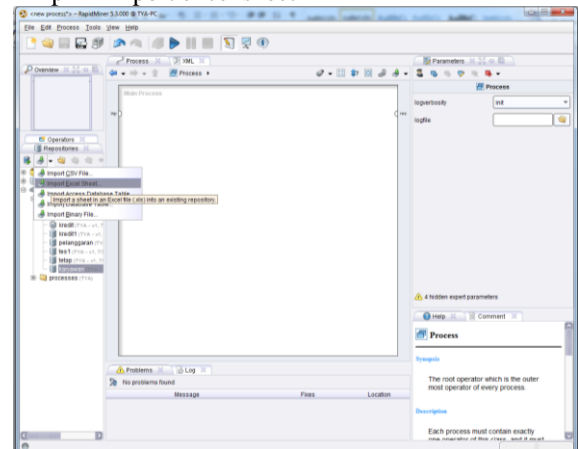
$$= \frac{933504}{24300000} \times \frac{24300000}{2154240} = \frac{933504}{2154240} = 0,433$$

Nilai probabilitas yang terbesar adalah c1 (ya diangkat) maka yang dipilih adalah c1

B. Penggunaan Rapidminer

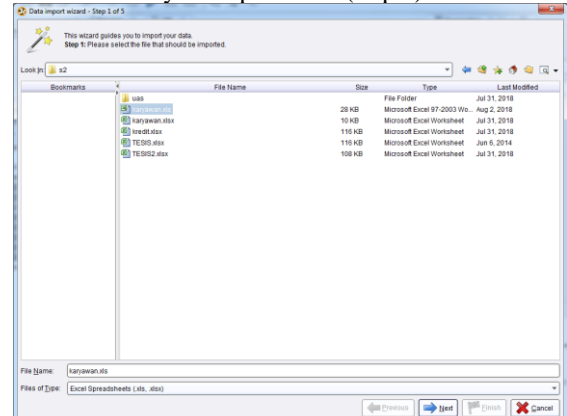
Berikut adalah langkah-langkah dalam penggunaan rapidminer.

1. Import data yang akan diolah. Pilih import lalu pilih import excel sheet



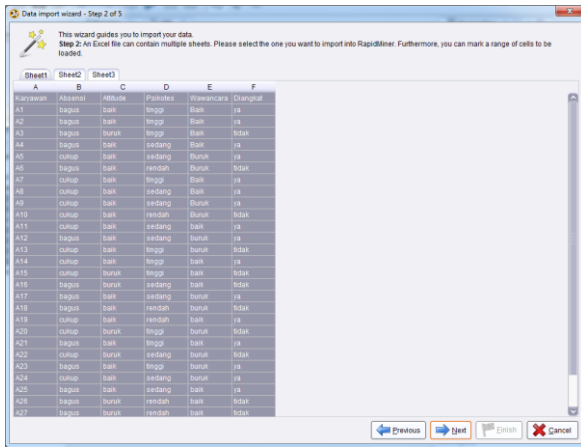
Gambar 2. Import data

2. Pilih datanya lalu pilih next (step 1)



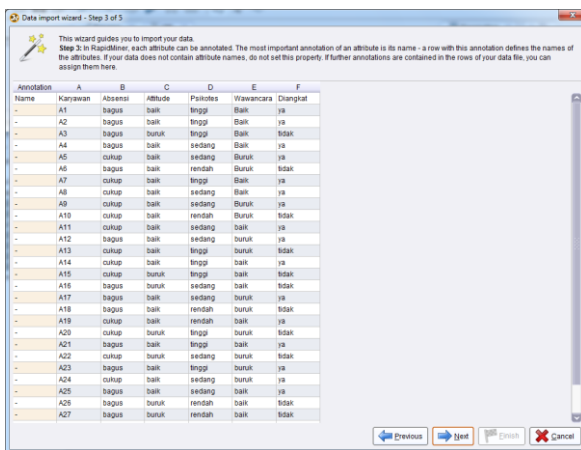
Gambar 3. Pemilihan data (step 1)

3. Setelah tampil datanya (step 2), pilih next lagi



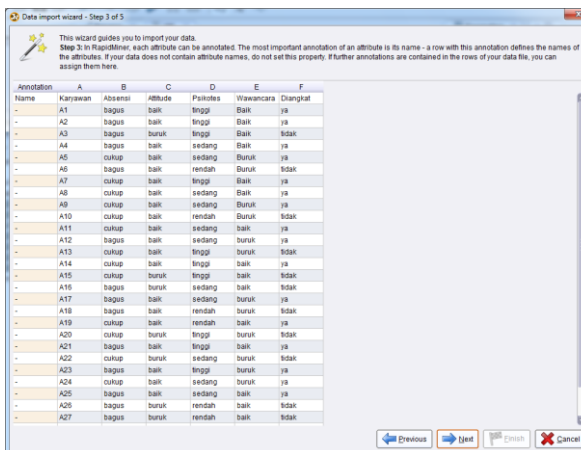
Gambar 4. Step 2 import data

4. Akan muncul tampilan sebagai berikut lalu pilih next (step 3)



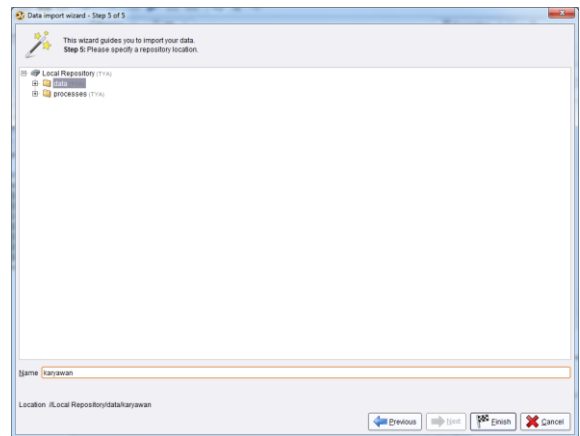
Gambar 5. Step 3 import data

5. Lalu kita dapat mengatur datanya, seperti penamaan, tipe dan lain-lain (step 4). Setelah diatur pilih next.



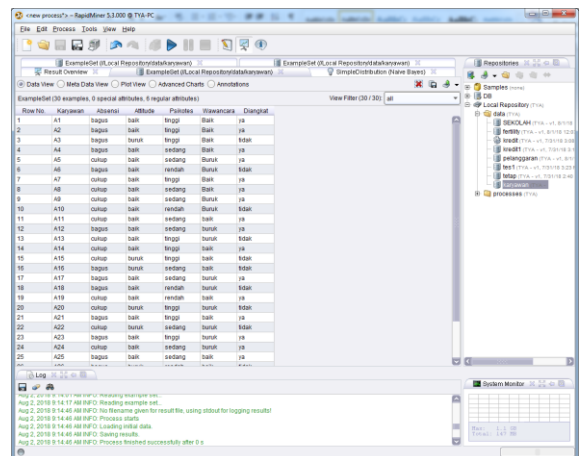
Gambar 6. Step 4 import data

6. Tentukan lokasi penyimpanannya dan simpan dengan name file yang diinginkan. lalu pilih finish (step 5)



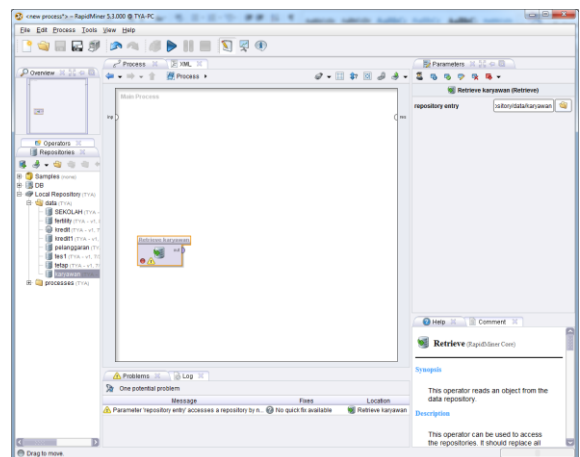
Gambar 7. Step 5 import data

7. Jika benar akan tampil datanya. Dengan penjelasan jumlah data dan jumlah atributnya.



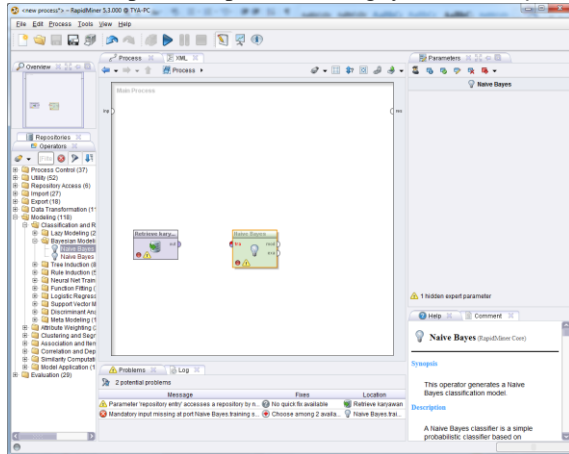
Gambar 8. Hasil import data

8. Pilih menu design untuk melakukan proses.



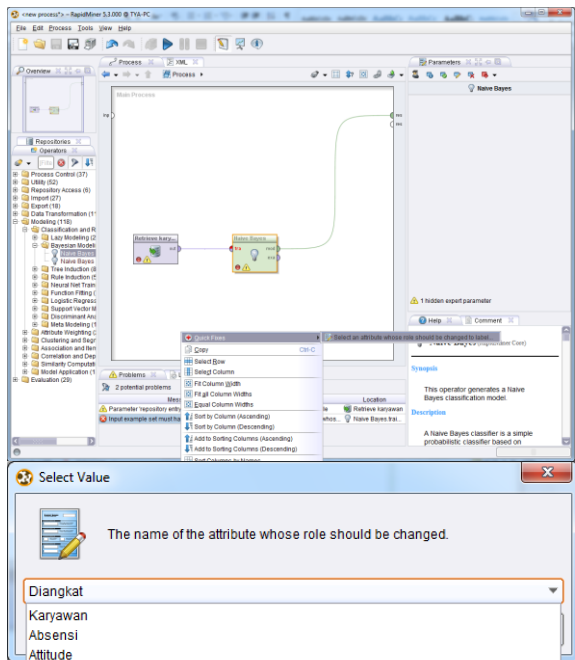
Gambar 9. Editor main process

9. Tarik data dari repository (data yang sebelumnya kita simpan) dan pilih modelingnya (naive bayes)



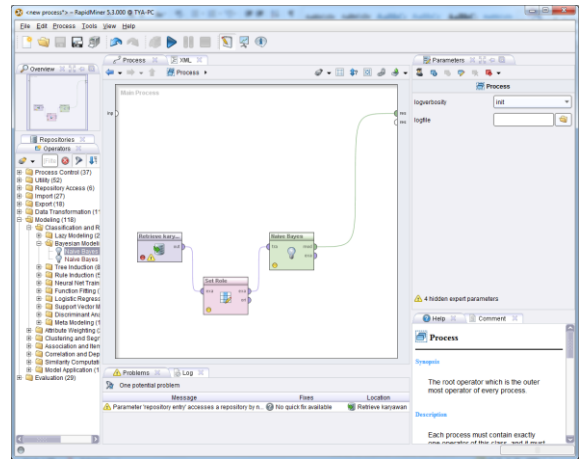
Gambar 10. Penarikan data kedalam main process

10. Lalu kita lanjutkan lagi prosesnya. Pilih operation, karna metode yang digunakan adalah naïve bayes. Maka pilih modeling-clasification-bayesian modeling. Pilih naïve bayes. Lalu tarik naïve bayes kedalam main process, Jika masih ada tanda merah atau ada error. Lihat pada jendela problem, klik kanan pada kesalahannya. Pilih quickfixs lalu pilih select an atribut whose role should be chage label (pilih yang akan dijadikan labelnya)



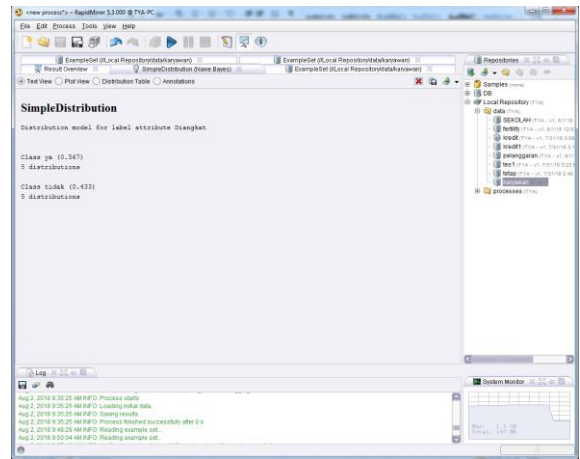
Gambar 11. Pemilihan medel dan label

11. Jika sudah akan tampil gambar seperti berikut ini. Semua telah terhubung.



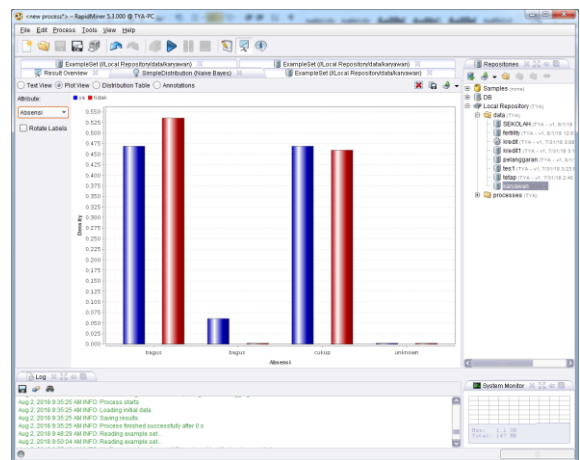
Gambar 12. Hasil main process

12. Jalankan program untuk melihat hasilnya. Berikut ini hasil dalam bentuk text view. Bisa kita lihat hasil yang ditampilkan oleh aplikasi rapidminer sama dengan perhitungan manual.



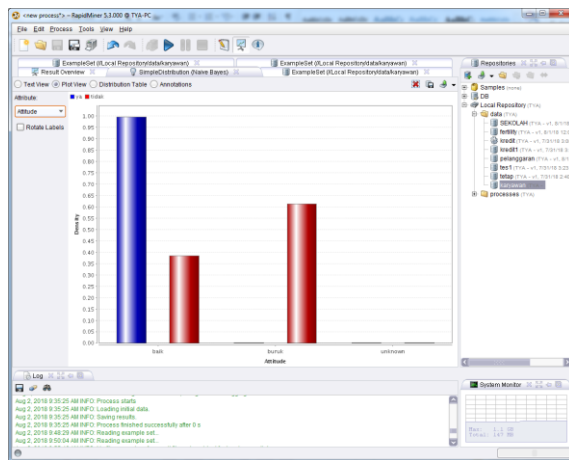
Gambar 13. Hasil running program

13. Berikut ini hasil dalam bentuk tampilan diagram (post view) berdasarkan absensi



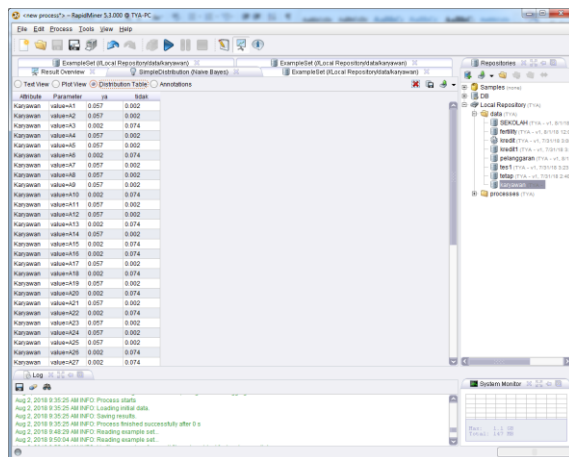
Gambar 14. Hasil dalam bentuk post view berdasarkan absensi

14. Berikut ini hasil dalam bentuk tampilan diagram (post view) berdasarkan attitude



Gambar 15. Hasil dalam bentuk post view berdasarkan attitude

15. Dan ini adalah tampilan dalam bentuk distribution table



Gambar 16. Hasil dalam bentuk distribution table

C. Hasil

Dari hasil tersebut dapat dilihat dalam bentuk text view, bahwa hasil yang di dapat dari perhitungan manual dan yang dihasilkan rapidminer nilainya sama. Karena nilai probalitas yang tertinggi adalah c1 atau diangkat karyawan. Maka didapatkan pola untuk penetapan status karyawan diangkat adalah

berdasarkan atribut dan nilai, absensi: bagus, attitude: bagus, psikotes: tinggi dan wawancara : baik.

KESIMPULAN

Berdasarkan proses klasifikasi penetapan status karyawan, maka penulis dapat simpulkan, pengangkatan karyawan dengan sistem yang manual menuai kendala karena waktu dan subyektif pemilihan. Dengan penelitian klasifikasi penetapan status karyawan didapat pola yang tepat. Metode *Naive Bayes* dan *Rapidminer* sangat mudah digunakan. Untuk penelitian selanjutnya diharapkan dapat mengembangkan atribut yang lebih baik dengan menggunakan metode yang lain untuk klasifikasi.

REFERENSI

Khadafy, A. R., & Wahono, R. S. (2015). Penerapan Naive Bayes Untuk Mengurangi Data Noise Pada Klasifikasi Multi Kelas Dengan Decision Tree. *Journal of Intelligent Systems*, 1(2), 136–142.

Mukminin, A., & Riana, D. (2017). Komparasi Algoritma C4 . 5 , Naive Bayes Dan Neural Network Untuk Klasifikasi Tanah. *Jurnal Informatika*, 4(1), 21–31.

Muthia, D. A. (2014). ANALISIS SENTIMEN PADA REVIEW BUKU MENGGUNAKAN ALGORITMA NAIVE BAYES. *Jurnal Paradigma*, XVI(1), 8–16.

Panda, M., & Patra, M. R. (2007). Network Intrusion Detection Using Naive Bayes. *International Journal of Computer Science and Network Security*, 7(12), 258–263. <https://doi.org/10.1.1.128.936>

Rahman, M. F., Darmawidjaja, M. I., & Alamsah, D. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36–45.