

# ANALISIS PERBANDINGAN ALGORITMA NAÏVE BAYES DAN DECISION TREE ID3 UNTUK MEMPREDIKSI PRODUKTIFITAS MESIN WECKERLE

Fried Sinlae<sup>1\*</sup>, Adi Muhajirin<sup>2</sup>

Universitas Bhayangkara Jakarta Raya, Ilmu Komputer, Informatika, Jakarta  
Jl Raya Perjuangan Bekasi Utara, Kota Bekasi, Jawa Barat

<sup>1\*</sup>[fried.sinlae@dsn.ubharajaya.ac.id](mailto:fried.sinlae@dsn.ubharajaya.ac.id)

<sup>2</sup>[adi.muhajirin@dsn.ubharajaya.ac.id](mailto:adi.muhajirin@dsn.ubharajaya.ac.id)

**Abstrak** — Tingkat keakuratan data dalam kehidupan sehari-hari sangat diperlukan karena dilihat dari perkembangan teknologi informasi yang terus maju. Menganalisis pengolahan data menjadi informasi yang diketahui menggunakan sistem data mining. Algoritma yang umum digunakan untuk prediksi adalah Naive Bayes dan Decision Tree ID3. Tujuan dari penelitian ini adalah untuk membandingkan algoritma Naïve Bayes dan algoritma Decision Tree ID3 terhadap akurasi studi kasus prediksi produktivitas mesin Weckerle di PT. Rudy Soetadi. Metode yang digunakan adalah mempelajari literatur dari berbagai sumber terkait dan mendapatkan pemahaman tentang data yang tersedia pada sumber yang relevan tentang Naïve Bayes classifier dan Decision Tree ID3 dalam sistem data mining. Hasil dari penelitian ini adalah Naïve Bayes nilai keakuratan data untuk menunjukkan confidence data uji yang sedang diolah yang diterapkan menunjukkan angka desimal. Sedangkan pada algoritma Decision Tree mendapatkan hasil pengukuran confidence data uji dalam memprediksi produktivitas mesin weckerle yaitu hanya menunjukkan angka 1 pada produktif dan 0 pada tidak produktif.

**Kata kunci** — Naïve Bayes, Decision Tree ID3, Algoritma

*Abstract* — The level of data accuracy in everyday life is very necessary because it is seen from the development of information technology that continues to advance. Analyze data processing into known information using a data mining system. Commonly used algorithms for prediction are Naive Bayes and ID3 Decision Tree. The purpose of this research is to compare the Naïve Bayes algorithm and the ID3 Decision Tree algorithm to the accuracy of the Weckerle machine productivity prediction case study at PT Rudy Soetadi. The method used is to study literature from various related sources and gain an understanding of the data available on relevant sources about Naive Bayes classifier and Decision Tree ID3 in data mining systems. The results of this study are Naïve Bayes data accuracy value to show the confidence of the test data being processed which is applied to show decimal numbers. While the Decision Tree algorithm gets the results of measuring the confidence of test data in predicting the productivity of the weckerle machine, which only shows the number 1 on productive and 0 on unproductive.

*Keywords* — Naïve Bayes, decision tree, Algorithm

## I. PENDAHULUAN

Peningkatan pada akurasi data dalam kehidupan saat ini sangat dibutuhkan karena melihat perkembangan teknologi informasi yang semakin pesat. Menurut (Huda, 2010) dalam penentuan tiap keputusan dalam kondisi tertentu hal yang perlu diperhatikan yaitu informasinya, maka dari itu tersedianya informasi telah menjadi media untuk menganalisa dan merangkum pengetahuan dari data yang berguna dalam pengambilan keputusan. Namun untuk mengambil keputusan pengetahuan dari data pada suatu informasi saja tidak cukup. Dibutuhkan juga sebuah Analisa data untuk menghasilkan bahan pertimbangan dari informasi yang telah disediakan. Analisa pengelolaan data menjadi informasi yang bisa memberikan pengetahuan yaitu caranya dengan menggunakan sistem *data mining* (Fikri, 2013).

Pada sebuah masalah dapat dilihat dari kecenderungan dari segi aturan ataupun perkiraannya di waktu yang akan datang yaitu dengan menggunakan *data mining*. Penerapan langkah-langkah dan teknik pada *data mining* dalam kehidupan nyata ada beragam, salah satunya yaitu dengan teknik klasifikasi. Menurut (Kurniawan, 2018) klasifikasi ini merupakan bentuk dasar dari analisis data, sedangkan menurut (Arpit Bansal, 2017) klasifikasi yaitu teknik yang digunakan untuk keanggotaan kelompok menurut data-data yang sudah tersedia. Sebelumnya telah banyak penelitian untuk memprediksi kelulusan dengan metode klasifikasi, salah satunya yaitu dengan algoritma Naïve Bayes dan algoritma *decision tree*. Menurut (Kurniawan, 2018) algoritma Naïve Bayes ini sendiri saat ini memang sedang populer hal ini karena tingkat akurasi dari kedua algoritma ini tinggi. Oleh karena hal tersebut penelitian ini memiliki bertujuan dalam membandingkan tingkat keakuratan dari algoritma Naïve Bayes dan algoritma *decision tree* ID3 dalam hal keakuratan produktifitas mesin weckerle pada PT. Rudy Soetadi.

## II. TINJAUAN PUSTAKA

Penelitian ini adalah perbandingan antara metode Naïve Bayes dan decision tree ID3 dalam klasifikasi data *mining* untuk memprediksi produktifitas mesin weckerle pada PT. Rudy Soetadi. Ada beberapa pustaka yang penulis gunakan sebagai berikut :

### A. Penambangan Data atau *Data Mining*

Data *mining* merupakan pencarian pada pola atau aturan dari sebuah data yang kapasitasnya sangat besar untuk menemukan informasi baru (Davies, 2004). Data *mining* juga dapat disebut sebagai proses yang digunakan untuk menentukan sebuah struktur data (Roiger, 2017). Karakteristik dari data *mining* (Davies, 2004) yaitu :

1. Berkaitan dengan penemuan sesuatu yang baru yang tidak diketahui sebelumnya terkait pola data tertentu.
2. Data yang digunakan merupakan data yang besar untuk membuat hasil yang besar.
3. Dapat digunakan untuk penentuan strategi dan keputusan yang kritis.

### B. Algoritma Naïve Bayes

Naïve Bayes merupakan pengklasifikasian atau penggolongan data yang menghitung kemungkinan dari dataset yang tersedia (Saleh, 2015). Sedangkan menurut (Bustami, 2018) Naïve Bayes merupakan pengklasifikasian untuk memprediksi peluang masa depan dengan metode probabilitas dan statistik sesuai dengan pengalaman di waktu sebelumnya.

Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(H|X) \cdot P(H)}{P(X)}$$

Keterangan :

$X$  : Data dengan *class* yang belum diketahui

$H$  : Hipotesis data  $X$  merupakan suatu *class* spesifik

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

$P(H)$  : Probabilitas hipotesis  $H$  (*prior probability*)

$P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$

$P(X)$  : Probabilitas  $X$

### C. Algoritma Decision Tree

*Decision tree* dibentuk dari tiga simpul, simpul *root*, simpul perantara, dan simpul *leaf*. Simpul *leaf* memuat keputusan akhir atau kelas target untuk suatu pohon keputusan. Simpul *root* adalah titik awal suatu keputusan (Ayung Candra Padmasari, 2019).

## III. METODOLOGI PENELITIAN

Pada penelitian ini terdapat pedoman agar penelitian dapat mendapatkan hasil yang maksimal dan menghindari kesalahan dari tujuan penelitian ini. Adapun tahapan dari penelitian adalah sebagai berikut :

Langkah-langkah dari metodologi penelitian dijelaskan sebagai berikut :

a. Identifikasi Masalah

Tahapan ini melaksanakan penjelasan terhadap masalah penelitian dengan menjelaskan poin-poin penting permasalahan lalu penulis mendapatkan landasan untuk menjabarkan masalah penelitian.

b. Tinjauan Pustaka

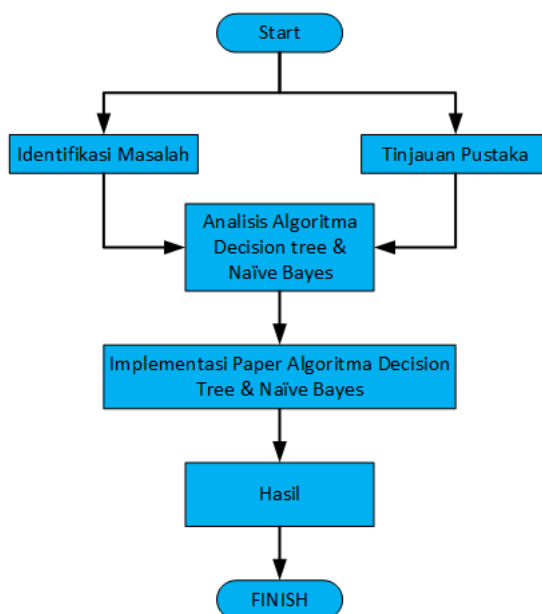
Penelitian dilakukan secara literatur dari berbagai referensi terkait serta memahami data-data yang ada pada referensi yang berkaitan dengan topik metode klasifikasi algoritma Naïve Bayes dan *Decision Tree* ke dalam sebuah sistem penambangan data atau data *mining*. Dilakukan juga pengkajian tentang teori yang berkaitan dengan topik penelitian yang sudah dilakukan pada penelitian yang ada serta perkembangan berbagai teori dan referensi saat ini.

c. Analisis

Pada tahap ini dilakukan proses pengkajian, penguraian dan pemecahan terhadap suatu masalah dan tujuan penelitian.

d. Implementasi Paper

Pada tahap ini dilakukan penerapan hasil analisis kedalam sebuah paper atau jurnal.



**Gambar 1. Metode Penelitian**

#### IV. HASIL DAN PEMBAHASAN

Pada penelitian ini analisa dengan melakukan perbandingan dua metode ialah algoritma *Decision Tree* dan metode algoritma Naïve Bayes ke dalam sebuah sistem penambangan data atau *data mining*. Disini kita membandingkan kedua metode dengan menggunakan data dari dua sumber yang mengambil tema mengenai prediksi produktifitas mesin weckerle pada PT. Rudy Soetadi. Sumber pertama pada data produktifitas mesin weckerle dibawah ini menggunakan perhitungan algoritma *Decision Tree*. Proses pada pengolahan data ini menggunakan metode *Decision Tree* berfungsi membangun sebuah pohon keputusan terdapat beberapa proses yaitu:

1. Menghitung hasil penjumlahan data, penjumlahan data ini berdasarkan banyaknya atribut hasil dengan memenuhi syarat yang sudah ditetapkan.
2. Menetapkan atribut tersebut dan digunakan untuk Node. Node adalah salah satu atribut yang nilai gainnya tertinggi dari atribut lain.
3. Membuat percabangan untuk setiap anggota dari Node.
4. Mengecek apabila ada anggota Node yang memiliki nilai nol, tapi jika hasil ada yang memiliki nilai nol maka tentukanlah mana yang tepat untuk menjadi daun dari pohon keputusan. Lakukan terus sampai keseluruhan nilai *entropy* anggota dari Node memiliki nilai nol sehingga proses berhenti.
5. Jika terlihat nilai *entropy* lebih dari nol yang berasal dari salah satu anggota dari Node, maka ulangi proses sebelumnya dari awal sampai semua Node memiliki nilai nol.

Tabel 1. Data Sampel

FINISH GOODS	DOWN TIME	REJECT PRODUCT	OUTPUT PRODUKSI	KELAYAKAN
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
SEDANG	ADA	BANYAK	BANYAK	PRODUKTIF
SEDANG	TIDAK ADA	BANYAK	BANYAK	PRODUKTIF
RENDAH	ADA	BANYAK	BANYAK	TIDAK PRODUKTIF
RENDAH	ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	TIDAK ADA	SEDIKIT	SEDIKIT	TIDAK PRODUKTIF
TINGGI	TIDAK ADA	BANYAK	BANYAK	PRODUKTIF
SEDANG	ADA	SEDIKIT	BANYAK	PRODUKTIF
RENDAH	ADA	BANYAK	BANYAK	TIDAK PRODUKTIF
RENDAH	ADA	BANYAK	BANYAK	TIDAK PRODUKTIF
RENDAH	TIDAK ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
SEDANG	TIDAK ADA	SEDIKIT	SEDIKIT	TIDAK PRODUKTIF
SEDANG	TIDAK ADA	SEDIKIT	BANYAK	PRODUKTIF

FINISH GOODS	DOWN TIME	REJECT PRODUCT	OUTPUT PRODUKSI	KELAYAKAN
TINGGI	TIDAK ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
RENDAH	TIDAK ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
SEDANG	ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
RENDAH	TIDAK ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
SEDANG	ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	SEDIKIT	TIDAK PRODUKTIF
SEDANG	TIDAK ADA	SEDIKIT	SEDIKIT	TIDAK PRODUKTIF
RENDAH	TIDAK ADA	SEDIKIT	BANYAK	TIDAK PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	SEDIKIT	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
SEDANG	ADA	SEDIKIT	BANYAK	PRODUKTIF
SEDANG	ADA	SEDIKIT	BANYAK	PRODUKTIF
SEDANG	TIDAK ADA	BANYAK	BANYAK	PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF
RENDAH	ADA	SEDIKIT	SEDIKIT	TIDAK PRODUKTIF
TINGGI	ADA	BANYAK	BANYAK	PRODUKTIF

Setelah didapatkan data sampel tersebut maka melakukan proses perhitungan jumlah data, *entropy* dan gain. Dimana hasilnya ada pada pada tabel dibawah ini :

Tabel 2. Perhitungan jumlah data, *entropy* dan gain

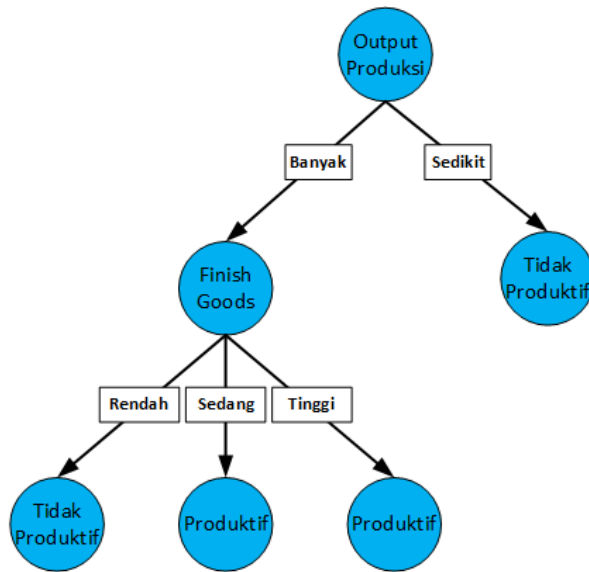
NO DE	JUMLAH	PROD UKTIF	TIDAK PRODUKTIF	ENTR OPY	GAIN
1 TOTAL	35	21	14	0,97095	
FINISH GOODS					0,446569331
TINGGI	15	12	3	0,72193	
SEDANG	11	9	2	0,68404	

NO DE		JUMLAH	PROD UKTIF	TIDAK PRODUKTIF	ENTR OPY	GAIN
	RENDAH	9	0	9	0,00000	
	DOWN TIME					0,052411577
	ADA	23	16	7	0,88654	
	TIDAK ADA	12	5	7	0,97987	
	REJECT PRODUCT					0,011891174
	BANYAK	25	16	9	0,94268	
	SEDIKIT	10	5	5	1,00000	
	OUTPUT PRODUKSI					0,517872341
	BANYAK	25	21	4	0,63431	
	SEDIKIT	10	0	10	0,00000	

NODE		JUMLAH	PRODUKTIF	TIDAK PRODUKTIF	ENTROPY	GAIN
2	OUTPUT PRODUKSI	25	21	4	0,63431	
	FINISH GOODS					0,63430955
	TINGGI	12	12	0	0,00000	
	SEDANG	9	9	0	0,00000	
	RENDAH	4	0	4	0,00000	
	DOWN TIME					0,00007476
	ADA	19	16	3	0,62925	
	TIDAK ADA	6	5	1	0,65002	
	REJECT PRODUCT					0,00007476
	BANYAK	19	16	3	0,62925	
	SEDIKIT	6	5	1	0,65002	

Setelah melalui beberapa proses perhitungan menggunakan metode pohon keputusan dari data kemudian didapatkan hasil dari nilai gain tertinggi yaitu pada *output* produksi. Maka *output* produksi disini

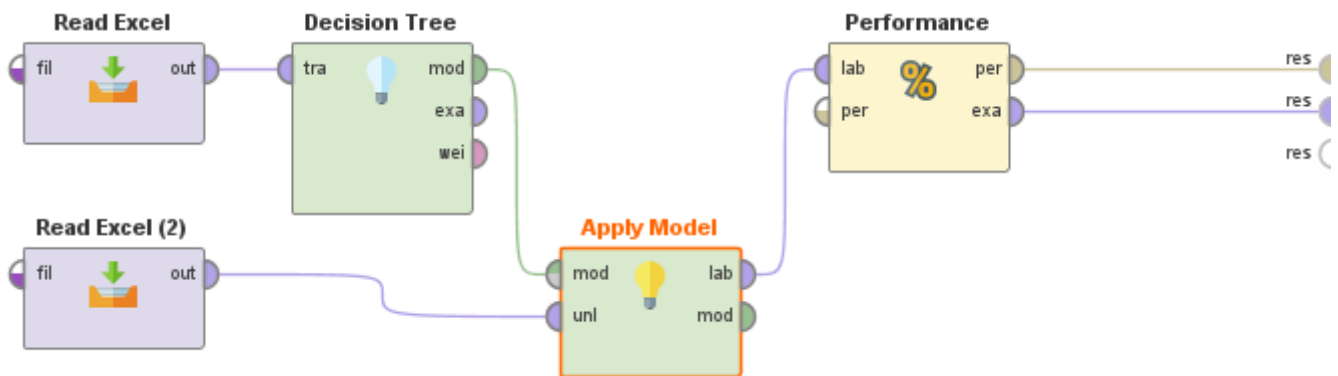
ditempatkan menjadi sebuah akar dari pohon keputusan. Dan *finish goods* produksi menjadi faktor yang menentukan produktifitas mesin weckerle. Terlihat pada gambar berikut:



**Gambar 2. Pohon Keputusan**

Pada Gambar 2 dijelaskan bahwa terdapat 2 macam anggota yaitu anggota banyak (produktif) dan anggota sedikit (tidak produktif). Dimana *Finish Goods* terdapat 3 anggota yaitu rendah (tidak produktif), sedang (produktif), tinggi (produktif). Dan mengapa pohon keputusan ini hanya sampai pada *Finish Goods*, hal ini dikarenakan nilai tersebut berada antara Anggota produktif dan tidak produktif terdapat nilai 0, maka keputusannya bisa langsung didapatkan. Kemudian juga terlihat *reject product* dan *downtime* tidak mempengaruhi produktifitas mesin weckerle PT. Rudy Soetadi.

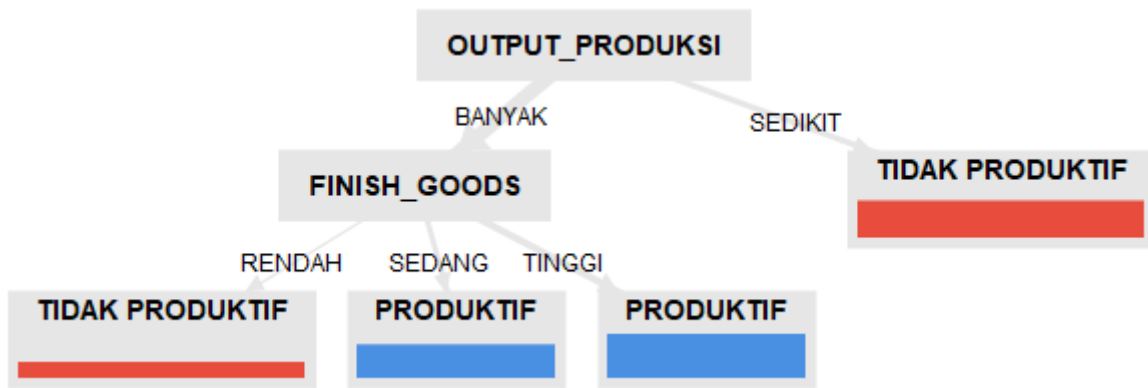
Pengujian selanjutnya yaitu kepada data sampel menggunakan tools yang ada pada aplikasi excel yaitu tools Rapidminer dimulai dari proses koneksi antara basis data sampel ke operator dan selanjutnya validasi seperti pada Gambar 3 berikut:



(Sumber: Penulis)

**Gambar 3. Koneksi Decision Tree Pada Tools Rapidminer 9.9.002**

Dari proses koneksi Rapidminer diatas hasil yang didapat sama seperti dengan proses perhitungan secara manual yang ada pada Gambar 2 sehingga mendapatkan hasil pohon keputusan seperti dibawah ini :



(Sumber: Penulis)

**Gambar 4. Pohon Keputusan Pada *Tools Rapidminer 9.9.002***

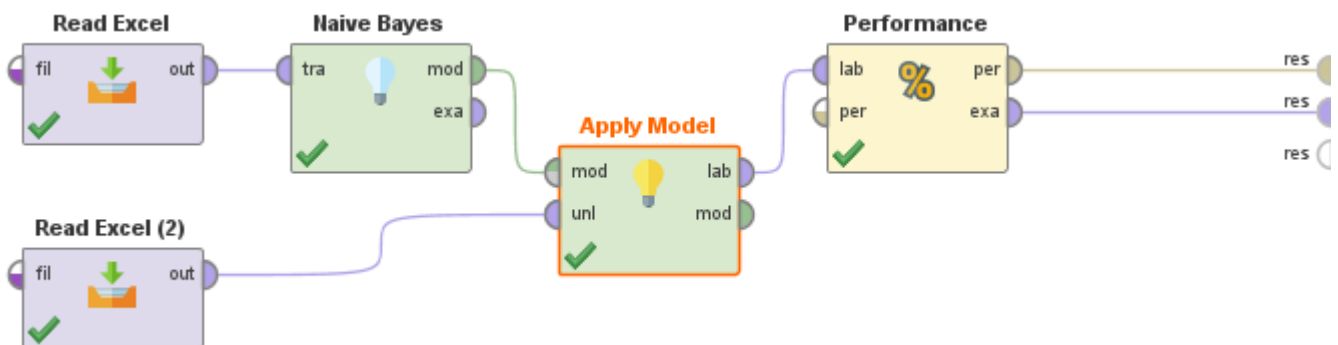
Berikut adalah *screenshot* hasil pengukuran data uji terhadap *Apply Model* algoritma *Decision Tree* dalam memprediksi produktifitas mesin weckerle terlihat bahwa *confidence* produktif 1 dan tidak produktif 0 karena mengikuti pohon keputusan pada gambar 4 dimana *downtime* dan *reject product* diabaikan.

Row No.	KELAYAKAN	prediction(KELAYAKAN)	confidence(PRODUKTIF)	confidence(TIDAK PRODUKTIF)	FINISH_GOODS	DOWN_TIME	REJECT_PRODUCT	OUTPUT_PRODUKSI
1	?	PRODUKTIF	1	0	TINGGI	TIDAK ADA	SEDIKIT	BANYAK

(Sumber: Penulis)

**Gambar 5. Prediksi Data Uji *Decision Tree* Pada *Tools Rapidminer 9.9.002***

Pada studi kasus yang kedua ini menerapkan metode Naïve Bayes dengan pengolahan data masih pada tema yang sama. Pada metode ini terdapat proses pengujian aplikasi tetap menggunakan *Rapidminer*. Dengan aplikasi *Rapidminer* ini akan dikelompokkan menurut atribut yang terpilih yaitu ada atribut-attribut dan dataset yang sama seperti metode *Decision Tree*.



(Sumber: Penulis)

**Gambar 6. Koneksi Naïve Bayes Pada *Tools Rapidminer 9.9.002***

Berikut adalah *screenshot* hasil pengukuran data uji terhadap *Apply Model* algoritma Naïve Bayes dalam memprediksi produktifitas mesin weckerle terlihat bahwa *confidence* produktif 0,816 dan tidak produktif 0,184 karena mengikuti rumus probabilitas pada Naïve Bayes sehingga dapat diimplementasikan



ke dalam angka berbeda dengan algoritma decision tree hanya mengikuti pohon keputusan dan mengabaikan atribut yang tidak dimasukkan ke dalam hitungan.

Row No.	KELAYAKAN	prediction(KELAYAKAN)	confidence(PRODUKTI...	confidence(TIDAK PRODUKTIF)	FINISH_GOODS	DOWN_TIME	REJECT_PRODUCT	OUTPUT_PRODUKSI
1	?	PRODUKTIF	0.816	0.184	TINGGI	TIDAK ADA	SEDIKIT	BANYAK

**Gambar 7. Prediksi Data Uji Naïve Bayes Pada Tools Rapidminer 9.9.002**

## V. KESIMPULAN

Dari hasil penjelasan pada uraian diatas dapat disimpulkan bahwa penggunaan metode algoritma Naïve Bayes nilai keakuratan data untuk menunjukkan *confidence* data uji yang sedang diolah yang diterapkan menunjukkan angka desimal. Sedangkan pada algoritma *Decision Tree* mendapatkan hasil pengukuran *confidence* data uji dalam memprediksi produktifitas mesin weckerle yaitu hanya menunjukkan angka 1 pada produktif dan 0 pada tidak produktif. Maka hal ini memunjukkan bahwa untuk memprediksi kelulusan algoritma Naïve Bayes memiliki *confidence* tingkat keakuratan yang lebih tinggi dibandingkan algoritma *Decision Tree*.

## REFERENSI

- Arpit Bansal, M. S. (2017). Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining. *International Journal of Computer Applications*, 35-40.
- Ayung Candra Padmasari, A. K. (2019). Penerapan Model Decision Tree untuk Rancangan Game Multiplayer Berbasis Jaringan (Uka-Uka Treasure Hunter). *Jurnal Edsence Vol. 1 No. 1*, 19-24.
- Bustami. (2018). PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI. *JURNAL INFORMATIKA*, 884-898.
- Davies, P. B. (2004). *Database Systems Third Edition*. New York: Plgrave Macmillan.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 36-41.
- Fikri, A. (2013). *Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression (Skripsi)*. Semarang: Fakultas Ilmu Komputer Universitas Dian Nuswantoro.
- Huda, N. M. (2010). *APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA (Skripsi)*. Semarang: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Diponegoro.
- Kurniawan, Y. I. (2018). PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 455-463.
- Lucidchart, T. (2018, May 23). *4 Phases of Rapid Application Development Methodology*. Retrieved 11 7, 2020, from Lucidchart: <https://www.lucidchart.com/blog/rapid-application-development-methodology>
- Mardi, Y. (2016). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 213-219.
- Martin, J. (1991). *Rapid Application Development*. New York: Macmillan Publishing.

- Nik, M. N., Nor, A. A., & Hazlifah, M. R. (2010). Implementing Rapid Application Development. In M. N. Nik, *Implementing Rapid Application Development (RAD) Methodology in Developing Practical Training Application System* (pp. 1664-1667). Kuala Lumpur, Malaysia: IEEE.
- Ricardo M. Bastos, D. D. (2002). *Extending UML Activity Diagram for Workflow Modeling in Production Systems*. Hawaii: IEEE.
- Roiger, R. J. (2017). *Data Mining: A Tutorial-Based Primer*. CRC Press.
- Saleh, A. (2015). KLASIFIKASI METODE NAIVE BAYES DALAM DATA MINING UNTUK MENENTUKAN KONSENTRASI SISWA ( STUDI KASUS DI MAS PAB 2 MEDAN ). *Konferensi Nasional Pengembangan Teknologi Informasi dan Komunikasi*, 200-208.
- Syahputra, Z. (2015). Penerapan Permodelan UML Sistem Informasi Perpustakaan pada Universitas Islam Indragiri Berbasis Client Server. *Jurnal SISTEMASI*, 57-64.
- Unhelkar, B. (2018). *Software Engineering with UML*. Boca Raton: Auerbach Publications/CRC Press.