



PENERAPAN DATA MINING UNTUK ANALISA KUALITAS PRODUK WELDING DENGAN ALGORITMA NAÏVE BAYES DAN C4.5 PADA PT. KARYA BAHANA UNIGAM

Ismasari Nawangsih¹, Junisa Sahar²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹ismasari.n@pelitabangsa.ac.id, ²junisasahar123@gmail.com

Abstrak

PT Karya Bahana Unigam merupakan *automotive seat company*, berdasarkan hasil pengujian algoritma C.45 dan Naïve bayes yang digunakan menghasilkan klasifikasi dari pengujian data. Hasil dari penelitian atau observasi dengan menggunakan data yang berbeda dengan jumlah data yang sama tetapi memiliki atribut, label dan kelas yang berbeda maka algoritma naïve bayes memiliki hasil klasifikasi optimal karena tingkat *accuracy*, *precision* dan *recall* memiliki rata-rata diatas 70% sampai 90%, karena data secara umum dapat dikategorikan sebagai data yang baik karena data memiliki jumlah atribut dan kelas yang lengkap dan memiliki informasi setiap atribut, label dan kelas pada data yang digunakan sehingga dapat memberikan informasi dalam meningkatkan kualitas pada tahap pengujian dari algoritma C.45 memiliki nilai *accuracy* 96.00%, *Precision* 95.00% dan *Recall* 79.17% sedangkan dari algoritma Naïve Bayes *accuracy* 99.33%, *Precision* 96.00% dan *Recall* 100.00%.

Kata Kunci : Data Mining, C.45, Naïve Bayes

Abstract

PT Karya Bahana Unigam is an automotive seat company, based on the results of testing the C.45 and Naïve Bayes algorithms which are used to produce classifications from data testing. The results of research or observations using different data with the same amount of data but having different attributes, labels and classes, the Naïve Bayes algorithm has optimal classification results because the levels of accuracy, precision and recall have an average of above 70% to 90%, because data in general can be categorized as good data because data has a complete number of attributes and classes and has information on each attribute, label and class in the data used so that it can provide information in improving quality at the testing stage of the C.45 algorithm. 96.00%, Precision 95.00% and Recall 79.17% while the Naïve Bayes accuracy algorithm is 99.33%, Precision 96.00% and Recall 100.00%.

Keywords: Data Mining, C.45, Naïve Bayes

1. Pendahuluan

Pada era globalisasi ini, perusahaan sangat membutuhkan suatu hasil kerja yang memiliki nilai produktivitas yang baik sehingga nilai perusahaan akan semakin meningkat. Perbaikan kualitas produk welding dan kuantitas harus terus dilakukan oleh perusahaan, baik dengan melakukan pengendalian kualitas langsung pada hasil produk maupun dengan melakukan kegiatan rutin untuk menganalisis pengendalian kualitas tersebut. Kualitas suatu produk welding merupakan salah satu kriteria yang menjadi pertimbangan pelanggan dalam memilih suatu produk. Kualitas produk merupakan indikator penting bagi perusahaan untuk dapat berdiri ditengah ketatnya persaingan dalam dunia industri. Kualitas produk welding semata-mata ditentukan oleh konsumen sehingga kepuasan konsumen hanya dapat dicapai dengan memberikan kualitas yang baik. Kualitas suatu produk dibangun perusahaan dengan memperhatikan kebutuhan dan keinginan *customer* karena suatu pabrik industri tidak akan eksis apabila produk yang dibuat atau dipesan tidak sesuai dengan

keinginan konsumen.

Pada PT. Karya Bahana Unigam (KBU) selaku perusahaan yang bergerak di bidang industri otomotif sangat memperhatikan kualitas dari produktivitasnya. Demi suatu kualitas dan kuantitas dari produk yang dihasilkan dengan baik, PT KBU secara rutin setiap harinya melakukan kegiatan *QC*, guna mengevaluasi masalah-masalah yang ada di lapangan. Dengan salah satu produk yang dihasilkan yaitu frame jok mobil, yang dimana dalam pembuatannya menggunakan tenaga robot dan manual yang masih bisa teridentifikasi welding *NG* atau *Not Good* yaitu product tidak bagus seperti cacat, produk rusak atau *NG* dan posisi welding tidak tepat. Di PT. Karya Bahana Unigam (KBU) produk *NG* masih menjadi tanggung jawab perusahaan, lebih tepatnya disebut sebagai kerugian perusahaan dikarenakan oleh pihak dalam perusahaan, bukan dari supplier. Produk *NG* ini dapat disebabkan oleh *human error*, karena kerusakan mesin yang digunakan sehingga barang rusak, dan bisa juga karena kesalahan teknik pada proses produksi dan kelalaian dari operator.

Dalam menghadapi masalah yang terjadi dalam perusahaan maka harus mengambil keputusan dengan tepat dalam menetapkan strategi produk yang akan dijual, untuk mendapatkan keputusan yang tepat diperlukan data barang yang cukup untuk dianalisa. Pada pertumbuhan yang sangat pesat dalam dunia teknologi dari akumulasi data telah menciptakan kondisi kaya akan data tetapi minim informasi. Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah yang besar yang diharapkan dapat mengatasi kondisi tersebut. Data mining sendiri memiliki beberapa teknik salah satunya yaitu klasifikasi. Teknik klasifikasi terdiri dari beberapa metode, dan Decision Tree adalah bagian dari metode klasifikasi. Terdapat berbagai macam metode untuk melakukan Data Mining yang digunakan seperti Naïve Bayes dan C.45. Metode yang terdapat didalam data mining yang akan digunakan dalam penelitian ini adalah algoritma C4.5, dimana algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*Decision Tree*). Berdasarkan permasalahan yang dipaparkan diatas, peneliti mencoba untuk menerapkan teknik klasifikasi menggunakan algoritma Naïve Bayes, C4.5 sebagai metode penyelesaian masalah yang ada.

2. Landasan Pemikiran

Analisa atau analisis atau Analysis adalah suatu usaha untuk mengamati secara detail sesuatu hal atau benda dengan cara menguraikan komponen-komponen pembentuknya atau penyusunnya untuk di kaji lebih lanjut. Analisa berasal dari kata Yunani kuno *analisis* yang artinya melepaskan. *Analisis* terbentuk dari dua suku kata, yaitu *ana* yang berarti kembali, dan *luain* yang berarti melepas sehingga jika di gabungkan maka artinya adalah melepas kembali atau menguraikan. Kata *analisis* ini di serap kedalam bahasa Inggris menjadi *analysis* yang kemudian di serap juga ke dalam bahasa Indonesia menjadi analisis.[4]

Data mining adalah proses menemukan

hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar[5].

Algoritma C4.5 (Decision Tree) Algoritma C.45 adalah salah satu metode klasifikasi yang paling populer karena mudah di interpretasi manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Misalkan untuk menentukan main tenis, kriteria yang digunakan adalah cuaca, angin, iklim dan temperatur. Manfaat utama menggunakan pohon keputusan adalah kemampuannya untuk membreak down proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambilan keputusan akan menjadi lebih menginterpretasikan solusi permasalahan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan sehingga sangat bagus sebagai langkah awal pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain[8].

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [9], yaitu: Menyiapkan data training. Data training biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu. Menentukan akar dari pohon. akar akan diambil dari atribut yang terpilih dengan cara menghitung nilai Gain dari masing-masing atribut, nilai Gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai Gain dari atribut, hitung dahulu nilai entropy yaitu:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

Kemudian hitung nilai Gain dengan metode *information gain*:

Algoritma Naive Bayes Algoritma Naive Bayes Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang $P(h|X)$ ini peluang posterior untuk h dan sebaliknya perior $P(h)$ [3].

Naive Bayes mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas label kelas label kelas y dengan tiap set atribut $X = \{X_1, X_2, \dots, X_d\}$ terdiri dari d atribut. Tahapan algoritma naive bayes:

1. Menyiapkan data training.
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu $X = X_1, X_2, X_3 \dots X_n$
3. N adalah gambaran dari ukuran yang dibuat di test dari n atribut yaitu: $A_1, A_2, A_3 \dots A_n$
4. M adalah kumpulan kategori yaitu $X = C_1, C_2, C_3 \dots C_m$
5. Diberikan data test X yang tidak diketahui kategorinya, maka classifier akan memprediksi bahwa X adalah milik kategori dengan posterior probability tertinggi berdasarkan kondisi X.
6. Naive bayes classifier menandai bahwa test X yang tidak diketahui tadi ke kategori C1 jika dan hanya jika $P(C_i|X) > P(C_j|X)$ untuk $1 \leq j \leq m, j \neq i$
7. Kemudian kita perlu memaksimalkan $P(C_i|X) P(C_i|X) = P(X|C_i) \cdot P(C_i) P(X)$.
8. Dimana x adalah nilai-nilai atribut dalam sampel X dan probabilitas $P(x_1|C_i), P(x_2|C_i), \dots, P(x_n|C_i)$,

RapidMiner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan program Java di bawah lisensi GNU Public Licence dan Rapid Miner dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuankoding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction dan Neural Network. Banyak metode yang disediakan oleh Rapid Miner mulai dari kllasifikasi, klustering, asosiasi dan lain sebagainya[9].

3. Metode Penelitian

Tahap penelitian menggunakan pengujian

model c4.5 dan naive bayes untuk mengukur kualitas welding yang memenuhi customer.

3.1 Pengumpulan Data

Pengumpulan data cara atau teknik yang dapat digunakan oleh peneliti untuk mengumpulkan data pada PT Karya Bahana Unigam. Atribut atau variabel adalah sifat atau nilai dari suatu objek, atau kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti untuk dipelajari kemudian ditarik kesimpulannya, teknik pengumpulan data yang digunakan dalam penelitian dalam melakukan analisis dan mencari pola agar memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat langkah – langkah dalam tahapan penelitian dan tahap ini menjelaskan tentang bagaimana dan dari mana sumber data didapatkan yaitu PT Karya Bahana Unigam.

3.2 Pengolahan Data

Pengolahan data pada penelitian ini menggunakan teknik klasifikasi pada data mining. Data yang akan dijadikan dataset dalam penelitian ini adalah data dari PT Karya Bahana Unigam, data tersebut akan dibagi menjadi 80% data training dan 20% data testing dan memisahkan data menjadi training dan testing agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data.

Pada tahap data cleaning merupakan proses pembersihan dari data yang akan dipakai untuk penghapusan data dengan membuang missing value, duplikasi data, dan memeriksa inkonsistensi data dan memperbaiki kesalahan pada data. Proses pembersihan data dilakukan secara manual dengan bantuan software spreadsheet.

Tabel 1 Data Cleaning

Nama Barang	Part No.	Customer	Model Barang	Proses	4M Factor
FRM_ASM_2NDB_LH_UN	M0177120	PT. NKS	4L45W	Pressing	Method
FRM_ASM_2NDB_LH_UN	M0177120	PT. NKS	4L45W	Welding	Machine
FRM_ASM_2NDB_RH_UN	M0165510	PT. NKS	4L45W	Welding	Method
FRM_ASM_2NDB_RH_UN	M0165510	PT. NKS	4L45W	Welding	Machine
...

3.3 Data Selection

Data Selection merupakan proses pemilihan data dari sekumpulan data operasional yang ada sebelum masuk ke tahap mining data maupun informasi.

Tabel 2 Data Dan Selection

Nama Barang	Part No.	Customer	Model Barang	Proses	4M Factor
FRM_ASM_2NDB_LH_UN	M0177120	PT. NKS	4L45W	Pressing	Method
FRM_ASM_2NDB_LH_UN	M0177120	PT. NKS	4L45W	Welding	Machine
FRM_ASM_2NDB_RH_UN	M0165510	PT. NKS	4L45W	Welding	Method
FRM_ASM_2NDB_RH_UN	M0165510	PT. NKS	4L45W	Welding	Machine
FRM_ASM_2NDC_RH_UN	M0165510	PT. NKS	4L45W	Welding	Material
FRM_ASM_2NDC_LH_UN	M0165510	PT. NKS	4L45W	Welding	Machine

3.4 Data Transformation

Tahap Data Transformation merupakan proses mengubah format data awal menjadi sebuah format data standar untuk proses pembacaan data dengan algoritma pada program maupun tool yang digunakan.

Tabel 3 Data Transformation

Nama Barang	Model Barang	Proses	4M Factor	Category
FRM_ASM_2NDB_LH_UN	4L45W	Pressing	Method	Lengkap
FRM_ASM_2NDB_LH_UN	4L45W	Welding	Machine	Welding NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Method	Welding NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Machine	Welding NG
...	...			

3.5 Pemodelan

Pada tahap ini dilakukan pemodelan data, metode yang dipakai pada penelitian ini adalah algoritma *C.45* dipilih karena salah satu kelebihanannya adalah dapat menangani data numerik dan diskret. Algoritma *C.45* menggunakan rasio perolehan (*gain rasio*). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep *entropy* untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasi.

3.6 Pengujian Data Menggunakan Metode

Pengujian metode dilakukan untuk mengetahui hasil perhitungan yang dianalisis untuk mengetahui apakah fungsi bekerja dengan baik atau tidak. Setelah data dihitung secara manual, kemudian data diuji menggunakan tools RapidMiner untuk memastikan apakah hasil perhitungan berjalan baik dalam menentukan klasifikasi stock barang dan mendapatkan hasil *accuracy, precision dan recall* dari pengujian data produk welding dari hasil yang diperoleh *rapidminer*, berikut ini adalah data sample

Tabel 4 Data Testing

Nama Barang	Model Barang	Proses	4M Factor	Category
FRM_ASM_2NDB_LH_UN	4L45W	Pressing	Method	Lengkap
FRM_ASM_2NDB_LH_UN	4L45W	Welding	Machine	Welding NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Method	Welding NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Machine	Welding NG
...	...			

C.45 Nilai *accuracy* dihitung dengan cara menjumlah data benar yang bernilai positif (True Positive) ditambah dengan nilai Negatif (True Negatif) dibagi dengan jumlah data benar yang bernilai positif

(True Positive), Negatif (True Negatif) dan ditambah dengan data salah yang bernilai positif (False Positif), Negatif (False Negatif).

Pembahasan dilakukan guna untuk mendapatkan nilai *accuracy, precision, recall* dengan algoritma *C.45*. Akurasi adalah tingkat kedekatan hasil prediksi dengan hasil fakta. Presisi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Serta *recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

3.7 Evaluasi Pengujian

Evaluasi dapat dilakukan dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian benar-benar sesuai dengan pembahasan. Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian menghasilkan nilai *accuracy, precision, recall* benar dan sesuai hasil pembahasan, pengujian dilakukan untuk mengukur keakuratan hasil dari tiap model yang diusulkan. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Pengukuran akurasi terhadap model dengan menggunakan confusion matrix yang menitik beratkan pada kelasnya. Confusion matrix merupakan tabel untuk mencatat hasil kerja klasifikasi.

4. Pembahasan

4.1 Hasil Penelitian

4.1.1 Data Uji

Dari hasil yang ada kemudian dikategorikan dengan *variable, atribut*, dari proses tersebut kemudian dihitung dengan algoritma Naïve Bayes, *C.45* untuk mengetahui prediksi stock barang, hasil *accuracy, recall dan precision*, sebagai dataset kemudian dikategorikan dengan *variable, atribut* dan kemudian dijadikan data *training dan testing*, dari proses tersebut kemudian di hitung dengan algoritma dengan Naïve Bayes dan *C.45*.

4.1.2 K-Fold Cross Validation

Cross Validation merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian data training dan sebagian data testing. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik sampling random untuk data training (85%) dan data testing (15%).

Tabel 5 Model K-Fold Cross Validation

K- Fold Cross Validation	Data training	Data Testing
1	750	150
2	750	150

Dari hasil diatas diperoleh *data testing* sebanyak 150 data, maka sisanya dijadikan *data training* sebanyak $900 - 150 = 750$ data.

Data tersebut disimpan dalam format excel workbook yang selanjutnya diubah menjadi data.frame dengan perintah read.excel. Berikut Ini adalah data uji atau data testing untuk di olah ke dalam tools *Rapid Mener*.

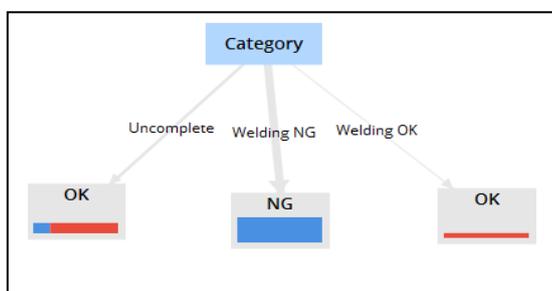
Tabel 6 DataTesting

Nama Barang	Model Barang	Proses	4M Factor	Category	Klasifikasi
FRM_ASM_2NDB_LH_UN	4L45W	Pressing	Method	Lengkap	OK
FRM_ASM_2NDB_LH_UN	4L45W	Welding	Machine	Welding NG	NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Method	Welding NG	NG
FRM_ASM_2NDB_RH_UN	4L45W	Welding	Machine	Welding NG	NG
FRM_ASM_2NDC_RH_UN	4L45W	Welding	Material	Welding NG	NG
FRM_ASM_2NDC_LH_UN	4L45W	Welding	Machine	Welding OK	OK

4.1.3 Perhitungan C.45

Tabel 7 Perhitungan Algoritma C.45

Langkah	Jumlah Kasus	OK	NG	ENTROPY	INFORMATION GAIN
Total	20	4	16	0.778632154	
Proses					0
Welding	20	6	16	0.778632154	
4M Factor					
Machine	10	3	7	0.881290899	0.01976006
Man	5	1	4	0.721928095	
Material	3	1	2	0.918297834	
Method	2	0	2	0	
Category					
Incomplete	5	4	1	0.721928095	0.59812013
Welding NG	11	0	11	0	
Welding OK	2	2	0	0	
Model Barang					0.10104094
4L45W	14	4	10	0.863120569	
TD	6	2	4	0.918297834	



Gambar 1 Pohon Keputusan

4.1.4 Perhitungan Naïve Bayes

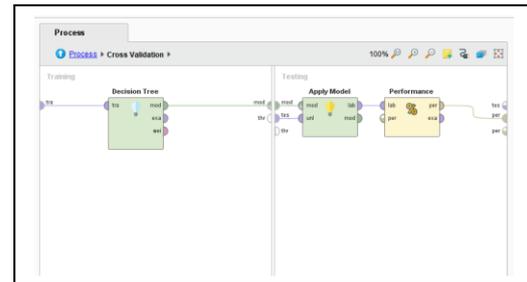
Perhitungan Naïve Bayes :

Kesimpulan: $0.0170898438 < 0.15$ maka hasil test = "OK".

4.2 Proses Pengujian Data (Rapid Mener)

Melakukan select attributes yaitu untuk mengetahui hasil prediski dari RapidMiner, hasil perhitungan manual dan hasil uji di rapid mener .

Untuk mempermudah dalam pembacaan data planing produksi dan delivery order baru yang diperlihatkan pada Gambar 1. Maka perlu di masukan tools algoritma naïve bayes, apply model untuk memodelakan hasil dari algoritma tersebut dan performance untuk mencari tingkat accuracy, percision, recall. Berikutini hasil ya :



Gambar 2 Proses Algoritma

4.2.1 Rapid Miner

Langkah kedua, dilakukan implementasi Naïve Bayes dan C.45 dengan menggunakan tools Rapid miner, Berikut adalah tahapan dalam penerapan algoritma Naïve Bayes dan C.45:

1. Menentukan nilai *accuracy*, *recal* dan *percision*
2. Menentukan performance

Langkah terakhir program yang telah dikembangkan adalah mengolah data yang sudah ada. Pada tahap akhir ini dilakukan evaluasi hasil pengolahan data dengan program yang telah dikembangkan.

Dataset yang digunakan berupa dataset berformat excel workbook yang kemudian dibaca menggunakan *Rapid miner*.

Setelah melakukan pengujian dataset welding, algoritma C.45 dan Naïve bayes yang digunakan menghasilkan klasifikasi dari pengujian data. Pada dataset sample ini ditentukan nilai *accuracy*, *precision* dan *recall* dari data sampel perhitungan sebanyak 150 data welding, pada tahap pengujian dari algoritma C.45 memiliki nilai *accuracy* 96.00%, *Precision* 95.00% dan *Recall* 79.17% sedangkan dari algoritma Naïve Bayes *accuracy* 99.33%, *Precision* 96.00% dan *Recall* 100.00%, berdasarkan hasil dari pengujian algoritma Naïve Bayes memiliki nilai accuracy tinggi dalam pengolahan data welding mendapatkan hasil yang baik.. Meningkatnya akurasi ini dapat mempermudah dalam pengambilan keputusan dan upaya pencegahan dalam meningkatkan kualitas

produk welding.

Hasil dari penelitian atau observasi dengan menggunakan data yang berbeda dengan jumlah data yang sama tetapi memiliki atribut, label dan kelas yang berbedamaka algoritma naïve bayes memiliki hasil klasifikasi optimal karena tingkat *accuracy*, *precision* dan *recall* memiliki rata-rata diatas 70% sampai 90%, karena data secara umum dapat dikategorikan sebagai data yang baik karena data memiliki jumlah atribut dan kelas yang lengkap dan memiliki informasi setiap atribut, label dan kelas pada data yang digunakan sehingga dapat memberikan informasi dalam meningkatkan kualitas.

5. Penutup

Berdasarkan hasil penelitian ini maka didapat beberapa kesimpulan sebagai berikut:

Berdasarkan hasil pengujian algoritma C.45 dan Naïve bayes yang digunakan menghasilkan klasifikasi dari pengujian data., pada tahap pengujian dari algoritma C.45 memiliki nilai *accuracy* 96.00%, *Precision* 95.00% dan *Recall* 79.17% sedangkan dari algoritma Naïve Bayes *accuracy* 99.33%, *Precision* 96.00% dan *Recall* 100.00%. Hasil dari penelitian atau observasi dengan menggunakan data yang berbeda dengan jumlah data yang sama tetapi memiliki atribut, label dan kelas yang berbeda maka algoritma naïve bayes memiliki hasil klasifikasi optimal karena tingkat *accuracy*, *precision* dan *recall* memiliki rata-rata diatas 70% sampai 90%, karena data secara umum dapat dikategorikan sebagai data yang baik karena data memiliki jumlah atribut dan kelas yang lengkap dan memiliki informasi setiap atribut, label dan kelas pada data yang digunakan sehingga dapat memberikan informasi dalam meningkatkan kualitas.

Daftar Pustaka

[1] R. Fitriana and N. Anisa, "Perancangan Pebaikan

Kualitas Produk Baut dan Sekrup Menggunakan Metode Six Sigma dan Data Mining di PT . A," vol. 9, no. 1, pp. 46–53, 2019, doi: 10.25105/jti.v9i1.4786.

- [2] M. Bakri, "Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K- Means," *J. Teknoinfo*, vol. 11, no. 1, p. 6, 2017, doi: 10.33365/jti.v11i1.3.
- [3] H. K. Siradjuddin, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Tingkat Kualitas Kesuburan (Fertility)," *J. Ilm.*, no. September, pp. 1–14, 2015, doi: 10.13140/RG.2.2.14204.18565.
- [4] A. Maukar, "Peningkatan Kualitas Pada Painted Head Saat Proses Tampo," no. March, 2018.
- [5] Suyanto, *Data Mining*. Yogyakarta: Informatika, 2017.
- [6] Retno Tri vulandari, *Data Mining*. Yogyakarta: Gava Media, 2017.
- [7] G. widi N. Dicky Nofriansyah, *Algoritma Data Mining Dan pengujian*. Yogyakarta: Cv Budi Utama, 2015.
- [8] S. Haryati, A. Sudarsono, and E. (2015) Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [9] O. Villacampa, "(Weka - Thesis) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," *ProQuest Diss. Theses*, no. 63, p. 188, 2015.
- [10] Y. Silalahi, Kristiani Desri., Murfi., Satria, "Studi Perbandingan Pemilihan Fitur untuk Support Vektor Machine pada Klasifikasi Penilaian Resiko Kredit," vol. 1, no.2, pp.119-136,2017