



ANALISIS DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES DALAM MEMPREDIKSI PEMBELIAN MATERIAL PLASTIK INJECTION (STUDI KASUS: PT SURYA TECHNOLOGY INDUSTRI)

Muhtajuddin Danny¹, Dede Yusup²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

¹utat@pelitabangsa.ac.id

Absstraksi

Untuk mempermudah setiap pengambilan keputusan dalam proses pembelian material plastik injection dibagian penjualan PT. Surya Technology Industri yang selama ini pembelian masih belum terkontrol dengan produksi ataupun penjualan ke *customer* sehingga mengakibatkan stok terlalu banyak dan tidak bisa terjual. PT. Surya Technology Industri yang bergerak di bidang Manufacturing memiliki jenis bahan baku lebih dari 400 jenis sehingga perusahaan kesulitan untuk memprediksi kebutuhan yang akan datang dan dibutuhkan perusahaan. Tujuan dari penelitian ini adalah menganalisis data pembelian material plastik *injection* untuk hasil yang lebih akurat dan efisien. Teknik data mining dengan metode algoritma Naïve Bayes digunakan dalam penelitian ini untuk melakukan klasifikasi sehingga menghasilkan keputusan dengan nilai probabilitas serta aturan-aturan yang berguna sebagai masukan dalam menentukan proses pengambilan keputusan. Dari 1120 dataset pembelian material plastik *injection* dilakukan pengujian sebanyak sepuluh kali, pembagian pengujian dengan data training dan data testing yang berbeda dengan ketentuan 10% data uji dan 90% data latih. Berdasarkan hasil yang diperoleh menunjukkan bahwa algoritma Naïve Bayes dalam melakukan klasifikasi data pembelian material plastik *injection* dengan menggunakan *tools RapidMiner* yang memiliki nilai *accuracy*, *precision* dan *recall* tertinggi yaitu pengujian pada proporsi 90% (1008 data) training dan 10% (112 data) testing menggunakan *cross validation 10-fold* dengan *accuracy* 88.93%, *precision* 92.88% dan *recall* 87.77%. Hal ini menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang cukup baik dalam melakukan Prediksi sehingga dapat di implemetasikan untuk proses pengambilan keputusan bagi perusahaan.

Kata Kunci: Algoritma Naïve Bayes, Data Mining, RapidMiner, Klasifikasi

Abstract

To facilitate every decision making in the process of purchasing plastic injection material in the sales department of PT. Surya Technology Industri, which has not been controlled with production or sales to customers, resulting in too much stock and cannot be sold. PT. Surya Technology Industry, which is engaged in Manufacturing, has more than 400 types of raw materials so that companies find it difficult to predict future needs and what the company needs. The aim of this research is to analyze the data for purchasing plastic injection materials for more accurate and efficient results. Data mining techniques with the Naïve Bayes algorithm method are used in this study to classify so as to produce decisions with probability values and useful rules as input in determining the decision-making process. Of the 1120 datasets on the purchase of plastic injection material, testing was carried out ten times,

the distribution of the test with different training data and testing data with the provisions of 10% test data and 90% training data. Based on the results obtained, it shows that the Naïve Bayes algorithm in classifying data on the purchase of plastic injection materials using the Rapid Miner tools has the highest accuracy, precision and recall values, namely testing at the proportion of 90% (1008 data) training and 10% (112 data) testing using cross validation 10-fold with 88.93% accuracy, 92.88% precision and 87.77% recall. This shows that the Naïve Bayes algorithm has a fairly good performance in making predictions so that it can be implemented for the decision-making process for the company.

Keywords : Naïve Bayes Algorithm, Data Mining, RapidMiner, Classification

1. Pendahuluan

Kemajuan teknologi adalah sesuatu yang tidak bisa dihindari dalam kehidupan ini, karena kemajuan teknologi akan berjalan sesuai dengan kemajuan ilmu pengetahuan. Dengan adanya teknologi yang berkembang pesat pada saat ini tentunya akan mempengaruhi dalam dunia industri. Perusahaan Manufaktur adalah sebuah badan usaha yang mengoperasikan mesin, peralatan dan tenaga kerja dalam suatu medium proses untuk mengubah bahan-bahan mentah menjadi barang jadi yang memiliki nilai jual. Karena itulah dalam pembelian bahan baku plastik *injection* harus dilakukan secara obyektif dan efisien agar proses seleksi dapat bersaing dalam dunia industri yang diharapkan perusahaan. PT. Surya Teknologi Industri yang bergerak di bidang Manufacturing memiliki jenis bahan baku lebih dari 400 jenis sehingga perusahaan kesulitan untuk memprediksi kebutuhan yang akan datang dan dibutuhkan perusahaan. Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna bagi perusahaan. Banyaknya parameter dalam menentukan material yang akan di beli atau tidak dibeli dengan menggunakan data sebelumnya yang sudah pernah terjadi.

Ada beberapa penelitian yang membahas tentang prediksi pembelian material namun metode yang digunakan ialah algoritma C4.5 yang pernah dilakukan oleh peneliti sebelumnya antara lain dilakukan oleh Fitriana Harahap pada tahun 2015. Dalam penelitiannya yang berjudul "Penerapan Data Mining dalam Memprediksi Pembelian cat".

Data mining mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna bagi perusahaan. Banyaknya parameter dalam menentukan diperpanjang atau tidaknya kontrak kerja karyawan menyebabkan ketepatan dan kecepatan dalam penilaian kerja kurang terpenuhi. Metode klasifikasi data mining merupakan sebuah teknik yang dilakukan untuk memprediksi class atau properti dari data itu sendiri. Adapun metode klasifikasi data mining memiliki beberapa algoritma salah satunya yaitu algoritma C4.5. Penyeleksian ini dilakukan sesuai kriteria-kriteria klasifikasi utama dengan menggunakan algoritma C4.5 yang dapat mengelola nilai inputan yang sesuai dengan kriteria-kriteria pada penilaian kinerja karyawan yang mempunyai nilai prioritas tertentu.

Berdasarkan uraian di atas maka peneliti tertarik untuk memprediksi pembelian material plastik *injection* pada PT. Surya Technology Industri. Namun penelitian ini membentuk model klasifikasi keputusan menggunakan metode algoritma *Naïve Bayes* guna membantu untuk memprediksi pembelian material plastik *injection* di PT. Surya Technology Industri.

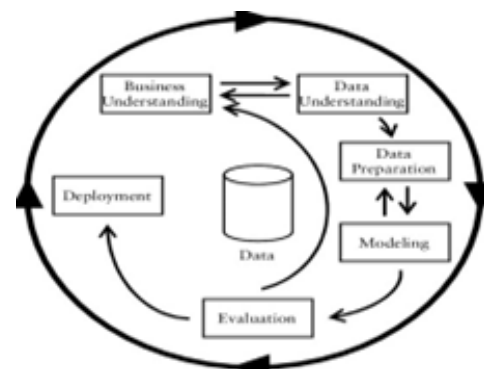
2. Landasan Pemikiran

2.1. Data Mining

Data Mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang

sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui [3].

Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa industry menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek data mining memiliki siklus hidup yang terbagi dalam 6 (enam) fase. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif dan fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Proses data mining menurut CRISP-DM dapat dilihat pada Gambar 2.3.



Gambar 1. Data Mining Menurut CRISP-DM

Dalam CRISP-DM, siklus hidup data mining yang terbagi dalam 6 (enam) fase yaitu [4]:

1. Fase pemahaman bisnis (*Business Understanding Phase*)
2. Fase pemahaman data (*Data Understanding Phase*)
3. Fase pengolahan data (*Data Preparation Phase*)
4. Fase pemodelan (*Modeling Phase*)
5. Fase evaluasi (*Evaluation Phase*)
6. Fase penyebaran (*Deployment Phase*)

2.2. Algoritma *Naïve Bayes*

Naïve Bayes merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. *Naïve Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. *Naïve Bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistic yang dikemukakan oleh ilmuwan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

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

$$P(H|X) = \frac{P(X|H) \cdot (P(H))}{P(H)} \dots \dots (2.1)$$

Dimana :

X : data dengan class yang belum diketahui

H : hipotesis data menggunakan suatu class spesifik

$P(H|X)$: probabilitas hipotesis H berdasar kondisi X (parteriori probabilitas) $P(H)$: probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: probabilitas X bedasarkan kondisi pada hipotesis $HP(X)$: probabilitas H

3. Metode Penelitian

Metode yang akan digunakan untuk menyelesaikan penelitian ini adalah penelitian eksperimen. Desain eksperimen yang digunakan adalah *Cross Standard Industry Process for Data Mining (CRISP-DM)* yaitu melakukan pengujian tingkat precision, recall dan akurasi dengan algoritma Naïve Bayes dalam menentukan pembelian material yang harus dibeli dan tidak dibeli.

Tahapan eksperimen sesuai dengan standar CRISP-DM adalah sebagai berikut:

1. Fase pemahaman bisnis (*Business Understanding Phase*)

Tahapan pertama yang dilakukan dalam penelitian ini ada survey lapangan, melakukan observasi di PT Surya Technology Industri. Berdasarkan hasil observasi ditemukan bahwa dalam proses memprediksi pembelian material plastic injection membutuhkan ketepatan supaya barang bias terjual dengan maksimal tidak menjadi stok digudang material.

2. Fase pemahaman data (*Data Understanding Phase*)

Tahap berikutnya adalah pengumpulan data. Data yang telah dikumpulkan dalam penelitian ini adalah data pembelian material plastic injection di PT Surya Technology Industri dari tahun 2017-2019. Jumlah data adalah 1120 data yang meliputi data nomor, tahun, bulan, customer, jenis resin, color, supplier, nama material, qty material, harga material, total amount pembelian material.

3. Fase pengolahan data (*Data Preparation Phase*)

Persiapan data yang dilakukan dalam penelitian ini adalah mengolah data mentah menjadi data yang dapat diolah dengan data mining, yaitu klasifikasi. Sampel data asli untuk data prediksi pembelian material plastic injection dapat dilihat pada gambar 3.2 sebelum dilakukan data preparation.

Dari data-data yang telah diperoleh, maka akan ditentukan suatu variabel yang menjadi variable keputusan dalam pembelian material. Diketahui beberapa faktor yang menjadi penentu dalam pembelian material plastik injection oleh PT. Surya Technology Industri adalah tingkat animo customer terhadap produk material yang dapat dilihat berdasarkan hasil penjualan material dengan warna material tertentu pad PT, Surya Technology Industri, jika hasil penjualan material tersebut tinggi maka animo customer tinggi terhadap material tersebut. Kompetisi supplier juga menjadi indikator

dalam pembelian material dimana persaingan supplier dalam menawarkan produk mereka kepada PT. Surya Technology Industri.

Dengan memberikan material dengan kualitas rendah, sedang dan tinggi dengan penawaran harga yang berbeda. Tingkat kompetisi dikatakan tinggi jika hasil penjualan material dari supplier tersebut tinggi dan banyak diminati oleh customer. Berdasarkan analisa tersebut dapat ditentukan variabel-variabel yang digunakan dalam penentuan pembelian material dengan mempertimbangkan faktor di atas yaitu : kualitas, amino, harga dankompetisi.

Tabel 2. Atribut Data Penelitian

Atribut	Proses	Penjelasan
Nomor	Data cleaning	Nomor pembelian material
Tahun	Data cleaning	Tahun pembelian material
Bulan	Data cleaning	Bulan pembelian material
Customer	Data cleaning	Pelanggan untuk tujuan penjualan material
Jenis Resin	Data cleaning	Jenis atau grade pada suatu material
Color	Data cleaning	Warna material plastic injection
Supplier	Digunakan sebagai atribut	Bahan baku yang di pasok untuk diproses di PT Surya Technology Industri
Nama Material	Data cleaning	Nama material plastic injection
Qty Pembelian	Digunakan sebagai atribut	Quantity pembelian material resin
Harga Material	Digunakan sebagai atribut	Harga material per-kg menggunakan IDR.
Amount Pembelian	Digunakan sebagai atribut	Amount pembelian material plastic injection
Pembelian	Digunakan sebagai label	Data pembelian / Invoice di PT Surya Technology Industri

4. Modelling

Model yang diusulkan untuk klasifikasi adalah menggunakan algoritma *Naïve Bayes*. Untuk membuat model atau rule dan pengukuran akurasi dataset dibagi menjadi dua subset, yaitu data training dan data testing. Data training merupakan data yang digunakan untuk proses pemodelan klasifikasi. Selanjutnya data yang terbentuk akan diujikan kembali menggunakan data testing. Desain pemodelan ditunjukkan pada gambar 3.3 berikut.

Tabel 3. Pembagian Data

Ekperimen	Data set		Akurasi
1	10 %	90% Data Latih	
2	90 %	10 %	90% Data Latih
3	90%	10 %	90% Data Latih
4	90% Data Latih	10 %	90% Data Latih
5	90% Data Latih	10 %	90% Data Latih
6	90% Data Latih	10 %	90% Data Latih
7	90% Data Latih	10 %	90% Data Latih
8	90% Data Latih	10 %	90%
9	90% Data Latih	10 %	90 %
10	90% Data Latih	10 %	
		Rata - Rata	

5. Evaluation

Dalam tahapan ini dilakukan evaluasi dengan melihat hasil yang dicapai oleh model, yaitu untuk mengetahui tingkat *accuracy*, *precision*, dan *recall* dari metode klasifikasi yang dibuat menggunakan teknik confusion matrix. *Confusion matrix* adalah salah satu alat ukur berbentuk matrix 2x2 yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi dataset terhadap kelas pada algoritma yang dipakai.

Tabel 4. Confusion Matrix

		Predicated Class	
		YES	NO
Actual Class	YES	True Positive (TP)	False Negative (FN)
	NO	False Positive (FP)	True Negative (TN)

6. Deployment

Hasil pemodelan ini adalah analisa yang mengarah ke *Decision Support System* (DSS) dapat digunakan oleh perusahaan untuk memberikan pertimbangan dalam penentuan langkah selanjutnya menangani masalah proses seleksi perpanjangan kontrak kerja dan juga dapat digunakan untuk rujukan pemodelan berikutnya serta dapat digunakan sebagai pengambil keputusan.

3.6 Perangkat Lunak

Perangkat lunak merupakan sarana pendukung lainnya bagi penelitian ini. Pada penelitian yang akan dilakukan untuk menyelesaikan penelitian ini, instrument peralatannya yang berupa *software* meliputi:

1. *Microsoft Office Excel 2010 Microsoft excel* merupakan

Software yang digunakan untuk membuat tabel, dalam software ini juga dapat menggunakan rumus- rumus guna menjumlahkan, mengurangi, mencari nilai rata-rata, mencari nilai maximal ataupun minimal, serta menampilkan persyaratan suatu kondisi secara langsung tanpa harus menuliskan nya satu persatu.

2. *RapidMiner*

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi.

Tools yang digunakan dalam *RapidMiner* ini adalah:

- a. *Panel operator view*

Panel operator view merupakan induk langkah kerja dari proses analisa.

- b. *Panel repository view*

Merupakan komponen utama dalam *Design Perspective* selain *Operator View*. *View* ini dapat digunakan untuk mengelola dan menata proses Analisis menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber datadan yang berkaitan dengan meta data.

- c. *Process view*

Process View menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubunglangkah -langkah tersebut.

- d. *Parameter view*

Digunakan untuk mengatur fungsionalitas dari beberapa operator yang membutuhkannya agar dapat *dijalankan* sesuai dengan proses analisa.

4. Pembahasan

Tabel 5. Sample Data Hasil Pra-Proses

Ekperimen	Data set		Akurasi
1	10 %	90% Data Lelah	
2	90 %	10 %	90% Data Lelah
3	90%	10 %	90% Data Lelah
4	90% Data Lelah	10 %	90% Data Lelah
5	90% Data Lelah	10 %	90% Data Lelah
6	90% Data Lelah	10 %	90% Data Lelah
7	90% Data Lelah	10 %	90% Data Lelah
8	90% Data Lelah	10 %	90%
9	90% Data Lelah	10 %	90 %
10	90% Data Lelah	10 %	90 %
Rata - Rata			

Adapun Pra-Proses yang dilakukan dalam mempertimbangkan faktor diatas diambil berdasarkan sampel.

1. Mengelompokkan Kualitas Material Plastik *Injection*.

Tabel 6. Klasifikasi Kualitas

Klasifikasi	Kualitas
<1.000kg	Rendah
1.000kg - 5.000kg	Sedang
>5.000kg	Tinggi

2. Mengelompokkan Harga Material Pengelompokkan Harga material plastik *injection* diklasifikasikan menjadi tiga kelas yaitu, harga dikatakan Mahal jika harga lebih besar dari Rp. 75.000 /kg, harga dikatakan normal jika harga mencapai Rp35.000/kg sampai dengan 75.000/kg dan harga dikatakan murah jika harga dibawah dari Rp. 35.000/kg. Berikut harga material plastik

injection dalam range yang tampak seperti tabel 4.7 dibawah ini :

Tabel 7. Klasifikasi Harga

Harga	Klasifikasi
>75.000/kg	Mahal
35.000/kg -75.000/kg	Normal
<35.000/kg	Murah

- a. Mengelompokkan Animo Pengelompokkan Animo diambil berdasarkan hasil penjualan per produk material plastik injection yang dipasarkan dengan berbagai cara yang dilakukan pihak supplier material. Animo dikatakan kurang jika hasil penjualan per produk mencapai Rp.100.000.000, Animo dikatakan cukup jika hasil penjualan mencapai Rp. 100.000.000 sampai Rp.500.000.000 dan animo dikatakan bagus jika hasil penjualan lebih besar dari Rp. 500.000.000 yang dapat dilihat pada range tabel 4.8

Tabel 8. Klasifikasi Animo

Animo	Klasifikasi
>Rp.500.000.000	Bagus
Rp.100.000.000 - Rp.500.000.000	Cukup
<Rp.100.000.000	Kurang

- b. Mengelompokkan Kompetisi Pengelompokkan kompetisi diambil berdasarkan hasil penjualan produk material plastik injection per supplier yang memasarkan produk material tersebut. Kompetisi diklasifikasikan menjadi bagus, cukup dan rendah. Kompetisi dikatakan bagus jika hasil penjualan per supplier mencapai lebih besar dari Rp. 6.000.000.000, kompetisi dikatakan cukup jika hasil penjualan persupplier mencapai Rp. 2.000.000.000 sampai Rp. 6.000.000.000 dan kompetisi dikatakan rendah jika hasil penjualan kurang dari Rp 2.000.000.000, yang dapat dilihat pada range tabel 4.9

Tabel 9. Klasifikasi Kompetisi

Kompetisi	Klasifikasi
>Rp.6.000.000.000	Bagus
Rp.2.000.000.000 - Rp.6.000.000.000	Cukup
<Rp.2.000.000.000	Rendah

4.1 Penghitungan Dataset Cross Validation 10-Fold

Dari dataset 1120 data pemeblian material plastik injection dilakukan pengujian sebanyak 10 kali atau bisa disebut 10-fold. Pembagian pengujian dengan data training dan data testing yang berbeda, dimana masing-masing proposi pembagiannya datanya.

4.4.1 Penghitungan Pertama (1- Fold)

Pengujian yang pertama data training sejumlah 1008 data pengambilan data dari nomor 113 sampai 1120 dan data testing sejumlah 112 data dari nomor 1 sampai 112 dengan perbandingan 90% dan 10% (1-Fold). Tampilan seperti tabel 4.12 berikut ini:

Tabel 10. Data Training (90%)

Fold	NO	KUALITAS	AMINO	HARGA	KOMPETISI	PEMBELIAN
1-FOLD	1					
	112					
2-FOLD	113	RENDAH	KURANG	MURAH	RENDAH	TIDAK
	224	RENDAH	CUKUP	NORMAL	SEDANG	TIDAK
3-FOLD	225	SEDANG	BAGUS	MURAH	TINGGI	BELI
	336	SEDANG	BAGUS	MAHAL	TINGGI	BELI
4-FOLD	337	RENDAH	KURANG	MURAH	TINGGI	TIDAK
	448	RENDAH	BAGUS	MURAH	TINGGI	BELI
5-FOLD	449	RENDAH	BAGUS	MURAH	TINGGI	BELI
	560	TINGGI	BAGUS	MURAH	TINGGI	BELI
6-FOLD	561	TINGGI	BAGUS	MURAH	TINGGI	BELI
	672	TINGGI	BAGUS	MURAH	RENDAH	BELI
7-FOLD	673	RENDAH	BAGUS	NORMAL	RENDAH	BELI
	784	SEDANG	BAGUS	MURAH	RENDAH	BELI
8-FOLD	785	TINGGI	BAGUS	MURAH	RENDAH	BELI
	896	SEDANG	BAGUS	NORMAL	SEDANG	BELI
9-FOLD	897	RENDAH	BAGUS	MURAH	TINGGI	BELI
	1008	RENDAH	CUKUP	MAHAL	SEDANG	TIDAK
10-FOLD	1009	RENDAH	CUKUP	MAHAL	SEDANG	TIDAK
	1120	RENDAH	BAGUS	MURAH	TINGGI	BELI

4.4.2 Hasil Penghitungan Dataset Cross Validation 10-Fold

Dari dataset 1120 data pembelian material plastik injection dilakukan pengujian sebanyak 10 kali atau bisa disebut 10-fold. Pembagian pengujian dengan data training dan data testing yang berbeda, dimana masing-masing proposi pembagiannya datanya. Adapun hasil yang telah dipaparkan diatas bisa dilihat dari table

3.5 dibawah ini :

Tabel 13. Hasil Cross Validation 10-Fold (90% : 10%)

EXPERIMEN	DATA SET	AKURASI
1-Fold	up Data Latih	89.29%
2-Fold	Data Latih up	91.96%
3-Fold	Data Latih up	85.71%
4-Fold	Data Latih up	75.89%
5-Fold	Data Latih up	91.07%
6-Fold	Data Latih up	86.61%
7-Fold	Data Latih up	98.21%
8-Fold	Data Latih up	91.07%
9-Fold	Data Latih up	88.39%
10-Fold	Data Latih up	91.07%
	Rata - Rata	88.93%

Berdasarkan Tabel 4.22 hasil dari data cross validation 10-fold, hasil accuracy fold pertama 89.29%, hasil accuracy fold kedua 91.96%, hasil accuracy fold ketiga 85.71%, hasil accuracy fold keempat 75.89%, hasil accuracy fold kelima 91.07%, hasil accuracy fold keenam 86.61%, hasil accuracy fold ketujuh 98.21%, hasil accuracy fold kedelapan 91.07%, hasil accuracy fold kesembilan 88.39%, hasil accuracy fold kesepuluh 91.07%.

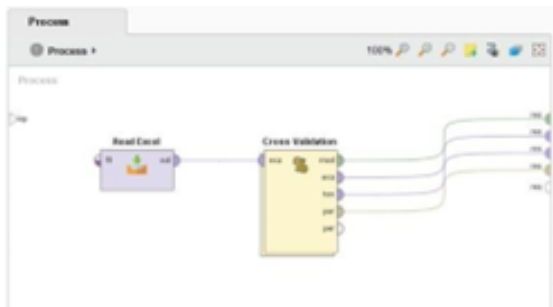
Hasil nilai rata –rata accuracy yang didapat ialah 88.93%, karena ada 44 data dengan class aktual TIDAK namun diprediksi BELI dan ada 80 data di dari

class aktual BELI namun diprediksi TIDAK, tampilan seperti tabel 4.6 berikut ini:

4.4.3. Hasil Pengujian dengan RapidMiner

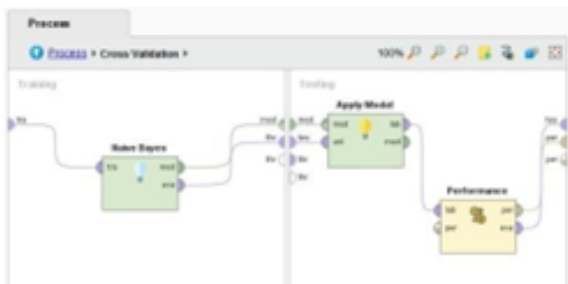
Pada penelitian ini peneliti menggunakan tools *RapidMiner* untuk menguji kebenaran dari hasil pengolahan data training yang dihitung secara manual. Pada tools *RapidMiner* terdapat panel *classifier output* berisi informasi dari hasil *running data* yang diinputkan yang merupakan presetase nilai akurasi, presisi dan *recall*.

1. Proses persiapan data training
 Dalam pengujian algoritma *Naïve Bayes* menggunakan tools *Rapid Miner*, data training pembelian *material injection* yang sudah dipersiapkan dari excel .
2. Menghubungkan data training dengan tools *RapidMiner*
 Buka tools *RapidMiner*, kemudian cari *read excel* di kolom pencarian *operators* lalu *drag-drop* pada kolom proses kemudian klik *import configuration wizard* di menu parameter lalu cari *file* yang sudah kita siapkanlalu klik *next*.



Gambar 2. Tampilan *RapidMiner Model Cross Validation*

3. Memilih Algoritma *Naïve Bayes*



Gambar 3. Tampilan *RapidMiner* dengan *Naïve Bayes*

	Real TIDAK	Real BELI	Real precision
pred TIDAK	402	80	84.00%
pred BELI	44	574	92.88%
class recall	88.93%	87.77%	

Gambar 4. Tampilan *RapidMiner* hasil *accuracy*

Berdasarkan gambar 4.5 di atas, sample data yang diambil untuk dianalisa. Untuk jumlah data 1120 memiliki *accuracy* 88.93% dengan model *cross validation*, penghitungan sebanyak 10-fold mendapatkan nilai ketepatan prediksi sebanyak 996 data dan yang tidak tepat sebanyak 124 data. Dimana

ada 80 data dari class aktual BELI namun di prediksi TIDAK sedangkan 44 data dari class aktual TIDAK namun di prediksi BELI, sehingga terjadi perbedaan data dalam penghitungan *cross validation* tersebut.

	Real TIDAK	Real BELI	Real precision
pred TIDAK	402	80	84.00%
pred BELI	44	574	92.88%
class recall	88.93%	87.77%	

Gambar 5. Tampilan persentasi *precision* model *cross validation*

Berdasarkan tabel 4.6 di atas, sample data yang diambil untuk dianalisa. Untuk jumlah data 1120 record dengan menggunakan *cross validation* algoritma *naïve bayes* memiliki presisi 92.89%,

	Real TIDAK	Real BELI	Real precision
pred TIDAK	402	80	84.00%
pred BELI	44	574	92.88%
class recall	88.93%	87.77%	

Gambar 6. Tampilan persentasi *recall* model *cross validation*

Berdasarkan tabel 4.7 di atas, sample data yang diambil untuk dianalisa. Untuk jumlah data 1120 record dengan menggunakan *cross validation* algoritma *naïve bayes* memiliki *recall* 87.78%.

4.2 Evaluasi dan Analisa Data

Tabel 14. *Confusion Matrix* Hasil *Accuracy, Precision dan Recall*

		CLASS										
		Prediksi BELI	TIDAK	Prediksi BELI	TIDAK							
1-FOLD	Confusion Table	BELI	47	4	6-FOLD	Confusion Table	BELI	47	3			
		TIDAK	8	53			TIDAK	12	50			
				CLASS								
2-FOLD	Confusion Table	BELI	53	0	7-FOLD	Confusion Table	BELI	95	0			
		TIDAK	9	50			TIDAK	2	15			
				CLASS								
3-FOLD	Confusion Table	BELI	52	7	8-FOLD	Confusion Table	BELI	52	1			
		TIDAK	9	44			TIDAK	9	50			
				CLASS								
4-FOLD	Confusion Table	BELI	41	18	9-FOLD	Confusion Table	BELI	64	6			
		TIDAK	9	44			TIDAK	7	35			
				CLASS								
5-FOLD	Confusion Table	BELI	50	5	10-FOLD	Confusion Table	BELI	79	0			
		TIDAK	5	52			TIDAK	10	29			
				T. CLASS								
TOTAL 1-10 FOLD	Confusion Table	T. Prediksi BELI	574	44	Accuracy	88.93%	Precision	92.88%	Recall	87.77%	TRUE	POSITIVE
		BELI	574	44								
		TIDAK	80	402								
				Accuracy	11.07%	Precision	84.06%	Recall	90.56%	TRUE	NEGATIVE	

Berdasarkan Pada tabel 4.23 nilai *confusion matrix dataset cross validation* menunjukkan bahwa

klasifikasi BELI memiliki 574 data diklasifikasikan kedalam *class* yang sesuai yaitu sebagai *class* BELI dan 80 data diklasifikasikan kedalam *class* yang tidak sesuai yaitu *class* TIDAK, sedangkan untuk nilai confusion matrix klasifikasi TIDAK memiliki 422 data sebagai *class* TIDAK dan 44 data diklasifikasikan kedalam *class* yang tidak sesuai, dari data tersebut didapatkan nilai *accuracy* 88.93%, nilai *presision* 92.88% dan nilai *recall* 87.77%.

Evaluasi dan analisis data dilakukan untuk melihat hasil yang dicapai oleh model, apakah model yang digunakan sudah menghasilkan klasifikasi pembelian material plastik injection yang sesuai. Evaluasi dan analisis data dilakukan dengan memasukan data pembelian kemudian membandingkan dan menghitung selisih antara hasil data testing klasifikasi pembelian material plastik injection yang dihasilkan menggunakan algoritma Naïve Bayes pada tools RapidMiner dengan data pembelian material yang sebenarnya.

Hasil *accuracy* yang didapatkan ialah 88.93%, dikarenakan ada 124 data yang tidak sesuai diantaranya 80 data *class* aktual BELI yang diprediksi TIDAK dan ada 44 data *class* aktual TIDAK yang diprediksi BELI, sehingga *accuracy* yang didapatkan ialah 88.93, adapun data yang salah prediksi sebagai berikut :

Tabel 15. Prediksi Pembelian yang tidak sesuai dengan *class* aktual

NO	KUALITAS	AMINO	HARGA	KOMPETISI	PEMBELIAN	CLASS Pred.	BELI Pred.	TIDAK Pred.
1	SEDANG	BAGUS	MURAH	TINGGI	BELI	TIDAK	0.001	0.040
2	RENDAH	CUKUP	MURAH	SEDANG	BELI	TIDAK	0.000	0.097
3	RENDAH	BAGUS	NORMAL	SEDANG	BELI	TIDAK	0.029	0.146
4	SEDANG	BAGUS	NORMAL	SEDANG	BELI	TIDAK	0.007	0.070
5	SEDANG	BAGUS	MURAH	TINGGI	BELI	TIDAK	0.004	0.047
6	SEDANG	BAGUS	MURAH	TINGGI	BELI	TIDAK	0.029	0.146
7	SEDANG	BAGUS	MURAH	SEDANG	BELI	TIDAK	0.000	0.031
8	SEDANG	BAGUS	NORMAL	SEDANG	BELI	TIDAK	0.000	0.065
9	RENDAH	BAGUS	NORMAL	SEDANG	BELI	TIDAK	0.000	0.065
10	SEDANG	BAGUS	MURAH	RENDAH	BELI	TIDAK	0.015	0.098
11	TINGGI	BAGUS	NORMAL	TINGGI	BELI	TIDAK	0.001	0.040
12	SEDANG	BAGUS	NORMAL	SEDANG	BELI	TIDAK	0.000	0.056
13	RENDAH	BAGUS	MURAH	TINGGI	BELI	TIDAK	0.006	0.084
14	RENDAH	BAGUS	MURAH	TINGGI	BELI	TIDAK	0.006	0.084
15	RENDAH	CUKUP	MURAH	SEDANG	BELI	TIDAK	0.000	0.097
16	RENDAH	BAGUS	MURAH	RENDAH	BELI	TIDAK	0.000	0.046
---	---	---	---	---	---	---	---	---
80	RENDAH	CUKUP	MURAH	TINGGI	BELI	TIDAK	0.029	0.136

NO	KUALITAS	AMINO	HARGA	KOMPETISI	PEMBELIAN	CLASS Pred.	BELI Pred.	TIDAK Pred.
1	RENDAH	KURANG	NORMAL	TINGGI	TIDAK	BELI	0.156	0.017
2	RENDAH	CUKUP	NORMAL	SEDANG	TIDAK	BELI	0.043	0.000
3	RENDAH	KURANG	MURAH	RENDAH	TIDAK	BELI	0.181	0.001
4	RENDAH	KURANG	MURAH	TINGGI	TIDAK	BELI	0.023	0.000
5	RENDAH	CUKUP	MURAH	TINGGI	TIDAK	BELI	0.022	0.000
6	RENDAH	CUKUP	NORMAL	SEDANG	TIDAK	BELI	0.043	0.000
7	RENDAH	CUKUP	NORMAL	TINGGI	TIDAK	BELI	0.023	0.000
8	RENDAH	KURANG	MURAH	SEDANG	TIDAK	BELI	0.008	0.000
9	RENDAH	CUKUP	MURAH	TINGGI	TIDAK	BELI	0.035	0.001
10	RENDAH	CUKUP	MURAH	TINGGI	TIDAK	BELI	0.030	0.010
11	RENDAH	CUKUP	MAHAL	RENDAH	TIDAK	BELI	0.023	0.000
12	RENDAH	KURANG	NORMAL	TINGGI	TIDAK	BELI	0.081	0.012
---	---	---	---	---	---	---	---	---
44	SEDANG	CUKUP	MURAH	SEDANG	TIDAK	BELI	0.017	0.007

Berdasarkan Tabel 4.24 diatas nilai *accuracy*, *presision* dan *recall* tidak akan mendapatkan 100% karena masih ada label *class* aktual berbeda dengan *class* prediksi, perbedaan tersebut didapatkan dengan penghitungan probabilitas yang didapatkan menggunakan algoritma Naïve Bayes, *accuracy* mendapatkan tingkat keakuratan 88.93% karena masih ada data yang tidak sesuai dengan perhitungan dengan data, ada 124 data yang tidak sesuai sehingga tingkat keakuratan yang didapat tidak mencapai 100% sedangkan *precision* mendapat nilai presisi dengan nilai 92.88% karena terdapat data *class* aktual

TIDAK namun dalam penghitungan probabilitas menunjukkan nilai BELI dan untuk tingkat *recall* memperoleh nilai 87.77% dikarenakan data yang tidak sesuai ialah 80 data dengan *class* aktual BELI namun dalam penghitungan probabilitas menunjukkan nilai TIDAK. berikut hasil yang didapatkan menggunakan tool RapidMiner sebagai berikut :

PerformanceVector

```
PerformanceVector:
accuracy: 88.93% +/- 2.73% (micro average: 88.93%)
ConfusionMatrix:
True: TIDAK BELI
TIDAK: 422 80
BELI: 44 574
precision: 92.88% +/- 2.09% (micro average: 92.88%) (positive class: BELI)
ConfusionMatrix:
True: TIDAK BELI
TIDAK: 422 80
BELI: 44 574
recall: 87.78% +/- 3.48% (micro average: 87.77%) (positive class: BELI)
ConfusionMatrix:
True: TIDAK BELI
TIDAK: 422 80
BELI: 44 574
AUC (optimistic): 0.937 +/- 0.015 (micro average: 0.937) (positive class: BELI)
AUC: 0.932 +/- 0.017 (micro average: 0.932) (positive class: BELI)
AUC (pessimistic): 0.926 +/- 0.018 (micro average: 0.926) (positive class: BELI)
```

Gambar 7. Detail Akurasi Hasil Pengujian Dataset Cross Validation

5. Penutup

Berdasarkan dari proses klasifikasi 1120 *record dataset* pembelian material plastik injection yang telah dilakukan pengujian menggunakan *cross validation* sebanyak 10-fold dan pembahasan yang telah dijelaskan maka dapat disimpulkan bahwa pengujian klasifikasi dengan menggunakan algoritma *naive bayes* pada tools RapidMiner yang memiliki nilai *accuracy*, *precision*, dan *recall* yang cukup tinggi yaitu pengujian menggunakan *cross validation 10- fold* mendapatkan nilai *accuracy* 88.93%, *precision* 92.88% dan *recall* 87.77%. Hal ini menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang cukup baik dalam melakukan klasifikasi sehingga dapat di implementasikan dan dikembangkan menjadi sistem penentu keputusan untuk membantu dalam prediksi pembelian material plastik injection di PT. Surya Technology Industri.

Adapun saran yang diberikan untuk menyempurnakan penelitian selanjutnya yaitu, tidak semua kasus atau permasalahan harus diselesaikan dengan satu algoritma pada *data mining*. Karena belum tentu algoritma yang digunakan merupakan algoritma yang paling akurat. Oleh karena itu penelitian selanjutnya perlu dilakukan komparasi beberapa algoritma untuk menentukan algoritma yang paling akurat.

Daftar Pustaka

[1] Alam, M. A., *Student Guide Series : Microsoft Excel*, Gramedia, Jakarta, 2006.
 [2] Andri, Kunang, Y. N., & Murniati, S., *Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang*. Seminar Nasional Informatika 2013 (semnasIF 2013), 56-

- 63, 2013.
- [3] Betrisandi, *Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis BackWard Elimination*, ILKOM Jurnal Ilmiah Volume 9 Nomor 1, IV(3), 96-100, 2017.
- [4] Bustami, *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*, Jurnal Informatika Vol. 8, No. 1, 6(2), 884-898, 2014.
- [5] Fauzan, R., Riadi, J., & Rinandi, F. R., *Identifikasi Siswa Bermasalah Menggunakan Metode Naive Bayes Di SMK Muhammadiyah 3 Banjarmasin*, Poros Teknik Volume 9, No. 2, XIV(3), 11-15, 2017.
- [6] Gorunescu, F., *Data Mining: Concepts, Models and Techniques*, Springer-Verlag Berlin Heidelberg, Romania, 2019.
- [7] Ikhsan, R., Pusnawati, ., E., & Siswandi, A., *Penentuan Tingkat Penjualan Mobil di Indonesia Dengan Menggunakan Algoritma Naive Bayes*, Prosiding Seminar Nasional Teknologi dan Sains (SNasTekS), 55- 70, 2019.
- [8] Imandasari, T., Irwan, E., Windarto, A. P., & Wanto, A., *Algoritma Naive Bayes Dalam Klasifikasi Volume 10 Nomor 4 Juni 2020 ISSN : 2407-3903*
- [9] *Lokasi Pembangunan Sumber Air*, Prosiding Seminar Nasional Riset Information, -, 750-761, 2019.
- [10] Kurniawan, D. A., & Kriestanto, D., *Implementasi Algoritma C4.5 Dalam Menentukan Lokasi Prioritas Penyuluhan Program Keluarga Berencana Di Kecamatan Dumai Timur*, Jurnal Informatika dan Komputer (JIKO) – Vol. 1, No. 1, 8(16), 19-23, 2016.
- [11] Kusriani, & Luthfi, E. T., *Algoritma Data Mining*, ANDI OFFSET, Yogyakarta , 2009.
- Muflikhah, L., Ratnawati, D. E., & Rekyan, *Data Mining*, UB Press, Malang , 2018.
- [12] Nofriansyah, D., & Nurcahyo, G.W., *Algoritma Data Mining Dan Pengujian*, CV Budi Utama, Yogyakarta, 2019.
- [13] Nuraeni, N., *Penentuan Kelayakan Kredit Dengan Algoritma Naive Bayes Classifier: Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC*. Teknik Komputer AMIK BSI Vol. III, No. 1., VI(2), 98-106, 2017.
- [14] Nurfiqih, & Noris, S., *Sistem Aplikasi Untuk Menentukan Kelayakan Area Distribusi Majalah Dengan Metode Naive Bayes Classifier Pada PT. Media Komunikasi Serpong*. Jurnal Prosiding Seminar Nasional Informatika, 15(2), 68-77, 2018.
- [15] Pratiwi, R. W., & Nugroho, Y. S., *Prediksi Rating Film Menggunakan Metode Naive Bayes*, Jurnal Teknik Elektro Vol. 8 No. 2, II(7), 60-63, 2016.
- [16] Priyono, F., & Indrajit, R. E., *Perbandingan Teknik Klasifikasi Untuk Prediksi Status Kontrak Kerja Karyawan*, Seminar Nasional Sains dan Teknologi 2017, 1-4, 2017.
- [17] Shenia, A., & Rouf, I., *Buku Pintar Menguasai Microsoft Excel untuk Pemula*, Media Kata, Jakarta, 2011 Siregar, A. M., & Puspabhuana, A., *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*, CV Kekata Group, Surakarta, 2018.
- [18] Sulianta, F., & Juju, D., *Data Mining: Meramalkan Bisnis Perusahaan*, PT Elex Media Komputindo, Jakarta, 2010.
- [19] Witten, H. I., Frank, Eibe, & Hal, *Data Mining: Pratical Machine Learning Tools and Techniques*, Third Edition. Burlington: Morgan, 2011.
- [20] Yudistira, D. T., *Penentuan Klasifikasi Status Gizi Orang Dewasa Dengan Algoritma Naive Bayes Classification (Studi Kasus Puskesmas Jiken)*. Scan, IX(2), 33- 39, 2014.