

## PREDIKSI LABA PENJUALAN MENGGUNAKAN METODE ALGORITMA C4.5 PADA PT.BASUNJAYA NASTARI

Ilman Kadori<sup>1</sup>, Bei Harira Irawan<sup>2</sup>, Mustamil<sup>3</sup>

<sup>1,2,3</sup>STMIC MIC Cikarang  
Cikarang Bekasi, Indonesia

[ilmankadori@gmail.com](mailto:ilmankadori@gmail.com) - [beiharira@gmail.com](mailto:beiharira@gmail.com) - [mustamil25@gmail.com](mailto:mustamil25@gmail.com)

### Abstrak

Era komputerisasi saat ini membawa dampak dalam banyaknya data penjualan oleh sebab itu dibutuhkan sebuah metode perhitungan laba penjualan yang efisien dan tepat untuk bisa selalu bertahan dalam ketatnya persaingan. Masalah yang dialami PT. Basunjaya Nastari sebagai perusahaan yang bergerak dibidang penyedia jasa dan jual beli bahan kimia adalah menentukan keuntungan penjualan barang untuk *customer* mana yang memiliki performansi, baik kualitas dan kuantitas, sehingga perusahaan bisa memprioritaskan *customer* tersebut dalam memenuhi permintaan yang dibutuhkan. Metode teknik klasifikasi dan prediksi data mining digunakan dalam memecahkan masalah tersebut yaitu *Decision Tree*. Teknik ini banyak digunakan karena dapat menyajikan data prediksi penjualan yang lebih cepat. Evaluasi hasil klasifikasi dalam prediksi keuntungan memiliki tingkat akurasi sebesar 81,71% dari data *training* dan tingkat akurasi sebesar 85,83% dari data *testing* dengan *Confusion Matrix*. Jika dilihat dengan kurva ROC menggunakan data *training* dengan akurasi *Excellent Classification* sebesar 1,000 dan data *testing* dengan akurasi *Excellent Classification* sebesar 0,925.

**Kata kunci:** *prediksi laba, keuntungan penjualan, decision tree, algoritma C4.5*

### Abstract

*In the current era of computerization has an impact on the amount of sales data, therefore we need an efficient and precise method of calculating sales profits to be able to always survive the intense competition. Problems experienced by PT. Basunjaya Nastari as a company engaged in providing services and buying and selling chemicals is to determine the advantages of selling goods to customers who have performance, both in quality and quantity, so that the company can prioritize these customers in meeting the required demands, classification techniques and data prediction methods. mining is used to solve this problem, the Decision Tree. This technique is widely used because it can provide faster sales prediction data. Evaluation of the classification results in profit prediction has an accuracy rate of 81.71% of the training data and an accuracy rate of 85.83% of the testing data with the Confusion Matrix. If you look at the ROC curve using training data with an accuracy of Excellent Classification of 1,000 and testing data with an accuracy of Excellent Classification of 0.925.*

**Keywords:** *profit prediction, sales profit, decision tree, C4.5 algorithm*

## PENDAHULUAN

PT. Basunjaya Nastari merupakan perusahaan yang bergerak di bidang penyedia jasa *Waste Water Treatment* dan jual beli bahan kimia yang terletak di daerah kawasan industri Cikarang Jawa Barat. Perusahaan ini memiliki beberapa *customer* yang antara lain juga perusahaan-perusahaan besar di kawasan industri. Masalah yang dialami PT Basunjaya Nastari adalah menentukan keuntungan penjualan barang untuk *customer* mana yang memiliki performasi, baik kualitas dan kuantitas, sehingga perusahaan bisa memprioritaskan *customer* tersebut dalam memenuhi permintaan yang dibutuhkan. Data penjualan yang ada belum diolah secara optimal untuk dijadikan referensi mencari performasi *customer* unggul guna meningkatkan penjualan lebih optimal. Belum adanya suatu metode yang tepat dan efisien dalam menentukan prediksi penjualan yang didapat dari hasil penjualan kepada *customer*.

Penelitian yang dilakukan oleh Rika Nofitri dan Muhammad Ardiansyah Sembiring (2017) yang membahas tentang analisa kinerja algoritma C4.5 dalam memprediksi pencapaian profit untuk menemukan *new knowledge* (pengetahuan baru) berupa *rule* dalam memprediksi pencapaian profit. Proses penelitian ini menghasilkan 6 buah *rule* sebagai landasan dalam memprediksi pencapaian profit dengan tingkat akurasi sebesar 92,60%+/- 1,36%.

Penelitian lain oleh Sandrawira Anggraini dkk (2018) tentang optimalisasi transaksi penjualan ban, memiliki pengaruh pada penentuan persediaan ban di gudang CV Roda Inti Mas. Dari penelitian ini jumlah persediaan ban dapat

diatur dengan baik, sehingga tidak lagi terjadi kekurangan ataupun kelebihan stok. Hal ini membantu dalam pengaturan optimalisasi transaksi ban.

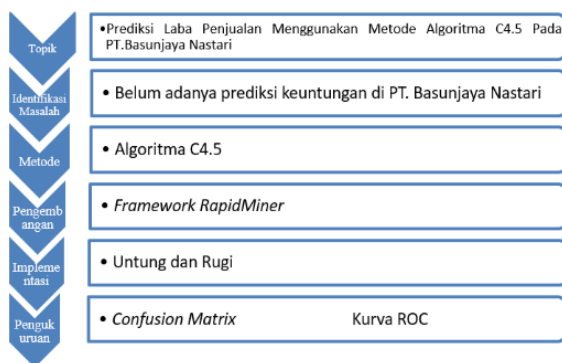
Penelitian lain oleh Goldie Gunadi dan Dana Indra Sensuse (2012) tentang penerapan metode data mining *Market Basket Analysis* terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan *frequent pattern growth (fp-growth)* pada percetakan PT Gramedia. Penelitian dilakukan dengan mencari sejumlah *frequent itemset* dan dilanjutkan dengan pembentukan aturan-aturan asosiasi (*association rules*). Simpulan didapatkan bahwa tingkat kekuatan aturan-aturan asosiasi yang dihasilkan oleh algoritma Apriori lebih besar dibandingkan dengan yang dihasilkan oleh algoritma FP-growth. Hal ini disebabkan oleh rendahnya nilai *support* yang menggambarkan terjadinya kombinasi *items* dalam aturan terhadap keseluruhan data transaksi terkini penjualan produk buku. Tingkat akurasi dari algoritma Apriori terhadap FP-growth adalah sebesar 257,4543%.

Penelitian lain oleh E. Buulolo, N. Silalahi, F. F and R. Rahim (2017) tentang prediksi dampak gempa bumi menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa algoritma C4.5 dapat memprediksi dampak gempa berdasarkan data seismik yang pernah terjadi dimana dampak gempa dipengaruhi oleh beberapa karakteristik atau kondisi gempa yaitu skala, durasi, jarak dari pantai dan pusat gempa.

## METODE PENELITIAN

Pengumpulan data dilakukan langsung di PT. Basunjaya Nastari. Data yang telah terkumpul berjumlah 192. Data

kemudian dibagi menjadi 2 yaitu data *training* sejumlah 154 data dan data *testing* sejumlah 38 data dengan perbandingan 80% dan 20%. Hasil data *training* digunakan untuk memperoleh hasil laba penjualan dalam bentuk pohon keputusan. Sedangkan data *testing* digunakan untuk mengukur tingkat akurasi dari hasil klasifikasi tersebut. Kerangka pemikiran yang digunakan pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Pemikiran

**HASIL DAN PEMBAHASAN**

Tabel 1 menunjukkan kategori penjualan pada PT. Basunjaya Nastari dari proses klasifikasi *Knowledge Discovery in Databases* (KDD).

Tabel 1. Tabel Kategori Penjualan

Barang	Warna Barang	Variasi Barang	Jenis Barang	Kemasan	Satuan	Harga	Customer	Kelas
1	Coklat	Ya	Padat	Pcs	Ltr	Rendah	PT. YKK Zepco	Rugi
2	Putih	Tidak	Cair	Zack	Kg	Sedang	PT. IVPV	Untung
3	Abu-Abu			Jerican		Tinggi	PT. Daido	
4	Kuning			Tangki			PT. Sutek	
5	Hitam						PT. RPT	
6	Biru						PT. Yamazaki	
7	Oren						PT. MUI	
8	Merah						PT. Sugiy	
9							PT. Masasi	

Hasilnya kemudian dikategorikan dengan variabel, atribut dengan nilai Rugi dan Untung yang kemudian dijadikan data *training* dan data *testing*.

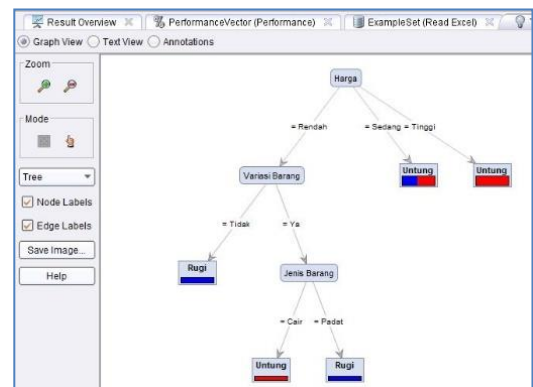
Tabel 2. Data Training

Row No	Kelas	Barang	Warna Bara	Variasi Bar	Jenis Barang	Kemasan	Satuan	Customer	Harga
1	Rugi	1	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Rendah
2	Untung	2	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Tinggi
3	Rugi	3	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Sedang
4	Untung	4	Coklat	Ya	Cair	Jerican	Ltr	PT. IVPV	Rendah
5	Untung	5	Coklat	Ya	Cair	Jerican	Ltr	PT. Daido	Tinggi
6	Untung	5	Coklat	Ya	Cair	Jerican	Ltr	PT. Sutek	Sedang
7	Rugi	7	Coklat	Tidak	Padat	Zack	Kg	PT. RPT	Rendah
8	Untung	8	Coklat	Tidak	Padat	Zack	Kg	PT. Yamazaki	Tinggi
9	Untung	9	Coklat	Tidak	Padat	Zack	Kg	PT. MUI	Sedang
10	Rugi	10	Coklat	Ya	Padat	Zack	Kg	PT. Sugiy	Rendah
11	Rugi	11	Coklat	Ya	Padat	Zack	Kg	PT. Masasi	Tinggi
12	Rugi	12	Coklat	Ya	Padat	Zack	Kg	PT. RPT	Sedang
13	Rugi	13	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Rendah
14	Untung	14	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Tinggi
15	Rugi	15	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Sedang
16	Untung	16	Coklat	Ya	Cair	Tangki	Ltr	PT. YKK Zepco	Rendah
17	Untung	17	Coklat	Ya	Cair	Tangki	Ltr	PT. YKK Zepco	Tinggi
18	Untung	18	Coklat	Ya	Cair	Tangki	Ltr	PT. RPT	Sedang
19	Rugi	19	Coklat	Tidak	Padat	Pcs	Kg	PT. RPT	Rendah

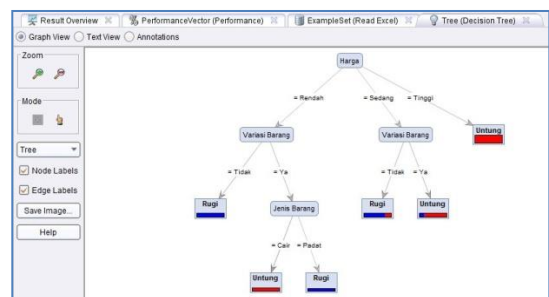
Tabel 3. Data Testing

Row No	Kelas	Barang	Warna Bara	Variasi Bar	Jenis Barang	Kemasan	Satuan	Customer	Harga
1	Rugi	1	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Rendah
2	Untung	2	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Tinggi
3	Rugi	3	Coklat	Tidak	Cair	Jerican	Ltr	PT. YKK Zepco	Sedang
4	Untung	4	Coklat	Ya	Cair	Jerican	Ltr	PT. IVPV	Rendah
5	Untung	5	Coklat	Ya	Cair	Jerican	Ltr	PT. Daido	Tinggi
6	Untung	5	Coklat	Ya	Cair	Jerican	Ltr	PT. Sutek	Sedang
7	Rugi	7	Coklat	Tidak	Padat	Zack	Kg	PT. RPT	Rendah
8	Untung	8	Coklat	Tidak	Padat	Zack	Kg	PT. Yamazaki	Tinggi
9	Untung	9	Coklat	Tidak	Padat	Zack	Kg	PT. MUI	Sedang
10	Rugi	10	Coklat	Ya	Padat	Zack	Kg	PT. Sugiy	Rendah
11	Untung	11	Coklat	Ya	Padat	Zack	Kg	PT. Masasi	Tinggi
12	Rugi	12	Coklat	Ya	Padat	Zack	Kg	PT. RPT	Sedang
13	Rugi	13	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Rendah
14	Untung	14	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Tinggi
15	Rugi	15	Coklat	Tidak	Cair	Tangki	Ltr	PT. YKK Zepco	Sedang
16	Untung	16	Coklat	Ya	Cair	Tangki	Ltr	PT. YKK Zepco	Rendah
17	Untung	17	Coklat	Ya	Cair	Tangki	Ltr	PT. YKK Zepco	Tinggi
18	Untung	18	Coklat	Ya	Cair	Tangki	Ltr	PT. RPT	Sedang
19	Rugi	19	Coklat	Tidak	Padat	Pcs	Kg	PT. RPT	Rendah

Berikut hasil pohon keputusan data *training* dan data *testing*.



Gambar 2. Pohon Keputusan Data Training



Gambar 3. Pohon Keputusan Data Testing

Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, stabilitas dan interpretabilitas (Vecellis, 2009). Lalu datanya diolah dan diuji tingkat akurasinya dengan metode *Confussion Matrix* dan kurva ROC/AUC (*Area Under Cover*).

**1. Confusion Matrix Data Training**

Perhitungan akurasi data *training* dari 154 data, 42 diklasifikasikan keuntungan ternyata *true* rugi, 7 data diprediksi *true* rugi tetapi ternyata *true* untung, 21 data diprediksi *true* rugi tetapi ternyata sebanyak 84 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 81,71%.

	true Rugi	true Untung	class precision
pred. Rugi	42	7	85.71%
pred. Untung	21	84	80.00%
class recall	66.67%	82.31%	

**Gambar 4.**Perhitungan Akurasi

Perhitungan akurasi berdasarkan *precision*, 42 diklasifikasikan keuntungan ternyata *true* rugi, 7 data diprediksi *true* rugi tetapi ternyata *true* untung, 21 data diprediksi *true* rugi tetapi ternyata sebanyak 84 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 83,20%.

	true Rugi	true Untung	class precision
pred. Rugi	42	7	85.71%
pred. Untung	21	84	80.00%
class recall	66.67%	82.31%	

**Gambar 5.**Evaluasi *Precision* Data *Training*

Perhitungan akurasi berdasarkan *recall*, 42 diklasifikasikan keuntungan ternyata *true* rugi, 7 data diprediksi *true* rugi tetapi ternyata *true* untung, 21 data diprediksi

*true* rugi tetapi ternyata sebanyak 84 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 91,81%.

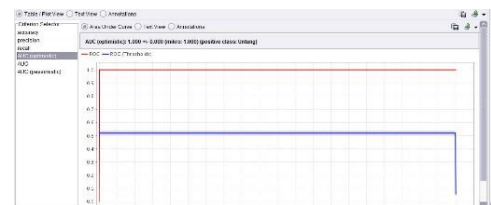
	true Rugi	true Untung	class precision
pred. Rugi	42	7	85.71%
pred. Untung	21	84	80.00%
class recall	66.67%	82.31%	

**Gambar 6.**Evaluasi *Recall* Data *Training*

*Performance* keakurasian AUC (Gorunescu, 2011) dapat diklasifikasikan menjadi lima kelompok yaitu:

1. 0.90 – 1.00 = *Excellent Clasification*
2. 0.80 – 0.90 = *Good Clasification*
3. 0.70 – 0.80 = *Fair Clasification*
4. 0.60 – 0.70 = *Poor Clasification*
5. 0.50 – 0.60 = *Failure*

AUC *Optimistic* metode Algoritma C4.5 menggunakan data *training* sebesar 1,00 dengan tingkat akurasi kategori *Excellent Clasification*.



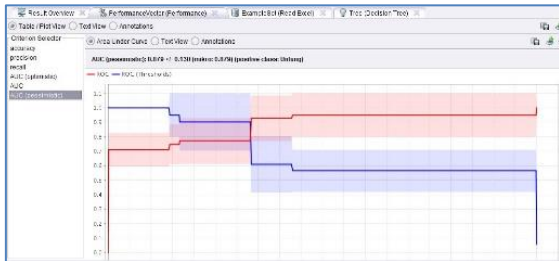
**Gambar 7.**Evaluasi AUC *Optimistic*

AUC *Optimistic* metode Algoritma C4.5 menggunakan data *training* sebesar 0,914 dengan tingkat akurasi kategori *Excellent Clasification*.



**Gambar 8.**Evaluasi AUC

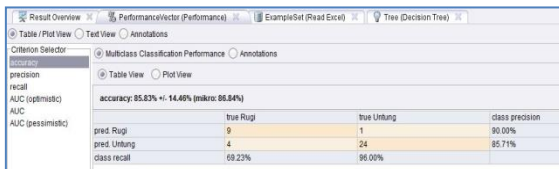
AUC *Pessimistic* metode Algoritma C4.5 menggunakan data *training* sebesar 0,879 dengan tingkat akurasi kategori *Good Clasification*.



Gambar 9. Evaluasi AUC *Pessimistic*

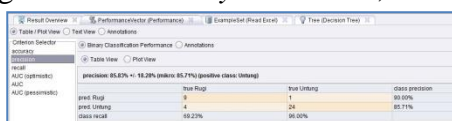
### 2. Confusion Matrix Data Testing

Perhitungan akurasi data *testing* dari 38 data, 9 diklasifikasikan keuntungan ternyata *true* rugi, 1 data diprediksi *true* rugi tetapi ternyata *true* untung, 4 data diprediksi *true* rugi tetapi ternyata sebanyak 24 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 85,83%.



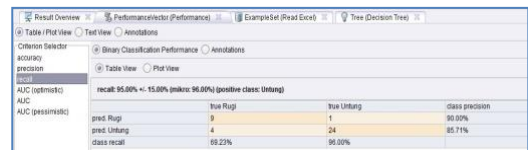
Gambar 10. Evaluasi Accuracy Data Testing

Perhitungan akurasi data *testing* dari 38 data, 9 diklasifikasikan keuntungan ternyata *true* rugi, 1 data diprediksi *true* rugi tetapi ternyata *true* untung, 4 data diprediksi *true* rugi tetapi ternyata sebanyak 24 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 85,83%.



Gambar 11. Evaluasi Precision Data Testing

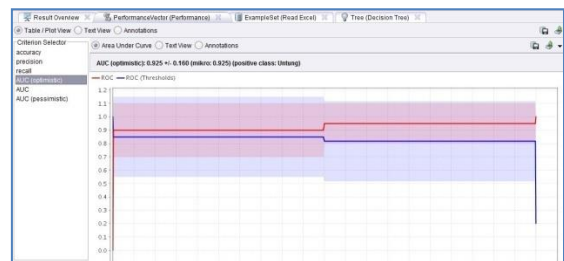
Perhitungan akurasi data *testing* dari 38 data, 9 diklasifikasikan keuntungan ternyata *true* rugi, 1 data diprediksi *true* rugi tetapi ternyata *true* untung, 4 data diprediksi *true* rugi tetapi ternyata sebanyak 24 diprediksi sesuai *true* untung, dengan nilai *accuracy* sebesar 95,00%.



Gambar 12. Evaluasi Recall Data Testing

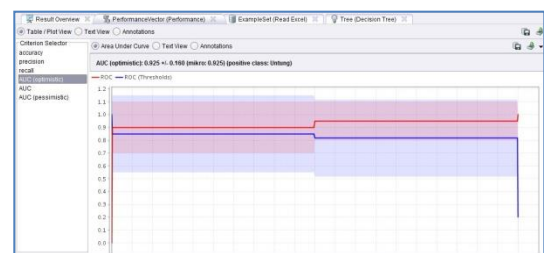
### 3. ROC/AUC (Area Under Cover) Data Testing

AUC *Optimistic* metode Algoritma C4.5 menggunakan data *training* sebesar 0,925 dengan tingkat akurasi kategori *Excellent Clasification*.



Gambar 13. Evaluasi AUC *Optimistic*

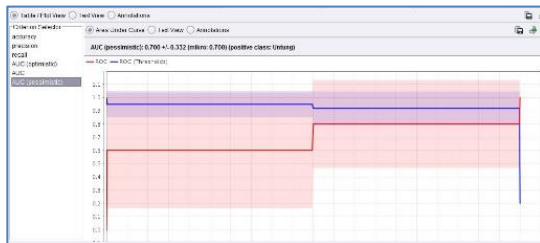
AUC metode Algoritma C4.5 menggunakan data *training* sebesar 0,550 dengan tingkat akurasi kategori *Failure*.



Gambar 14. Evaluasi AUC



AUC metode Algoritma C4.5 menggunakan data *training* sebesar 0,700 dengan tingkat akurasi kategori *Poor Classification*.



Gambar 15. Evaluasi AUC Pessimistic

#### 4. Accuracy Customer

Hasil yang didapat dari pengolahan data *training accuracy customer* terlaris menggunakan metode Algoritma C4.5 sebesar 76,45% yaitu pada PT. RPT (Resin Plating Technology) menggunakan data *training*.

	Yuk PT. Ynk 2	Yuk PT. LUPP	Yuk PT. Dado	Yuk PT. Sate	Yuk PT. RPT	Yuk PT. Yamo	Yuk PT. M&J	Yuk PT. Saph	Yuk PT. Manu	class predict
pred PT. Ynk 2	1	4	1	4	4	1	0	0	0	55.91%
pred PT. LUPP	0	1	1	2	0	0	0	0	0	0.00%
pred PT. Dado	1	0	1	3	0	0	0	0	0	0.00%
pred PT. Sate	0	0	0	2	1	0	0	0	0	0.00%
pred PT. RPT	4	1	0	52	1	3	3	0	0	74.29%
pred PT. Yamo	0	0	0	3	0	0	0	0	0	0.00%
pred PT. M&J	0	0	0	1	0	2	1	2	2	28.57%
pred PT. Saph	0	0	0	2	0	1	3	1	1	42.86%
pred PT. Manu	0	0	0	2	0	0	2	0	2	50.00%
class recall	55.91%	0.00%	0.00%	0.00%	75.47%	0.00%	28.57%	42.86%	40.00%	

Gambar 16. Accuracy Customer

#### KESIMPULAN DAN SARAN

Berdasarkan tujuan penelitian dan hasil beserta pembahasan diperoleh kesimpulan bahwa hasil yang dicapai dari evaluasi hasil klasifikasi dalam prediksi keuntungan memiliki tingkat akurasi sebesar 81,71% dari data *training* dan tingkat akurasi sebesar 85,83% dari data *testing* dengan *Confusion Matrix*. Jika melihat dengan kurva ROC menggunakan data *training* dengan akurasi kategori Excellent Classification sebesar 1,000 dan data *testing* dengan akurasi kategori Excellent Classification sebesar 0,925.

#### DAFTAR PUSTAKA

Anggraini Sandrawira dkk. (2018). *Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4.5*. Jurnal Ilmu Teknik Elektro Komputer dan Informatika (JITEKI) Vol. 4, No.2.

E. Buulolo, N. Silalahi, F. F and R. Rahim (2017). *C4.5 Algorithm To Predict the Impact of the Earthquake*. International Journal of Engineering Research & Technology (IJERT), vol. 6, no. 2, pp. 10-15.

Gorunescu. (2011). *Data Mining Concepts, Models and Techniques*. Romania: Springer-Verlag Berlin Heidelberg.

Gunadi Goldiedkk (2012). *Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) Studi Kasus Percetakan PT. Gramedia*. Jurnal Telematika. Volume 4. Hal 118 – 132.

Hermawati, Fajar Astuti. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.

Kamagi, D. H., & Hansun, S. (2014). *Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa*. ULTIMATICS, Vol. VI, No. 1 | Juni 2014, VI(1), 15–20.

Rika Nofitri, Muhammad Ardiansyah Sembiring (2017). *Analisa Kinerja Algoritma C4.5 Dalam Memprediksi Pencapaian Profit*. Seminar Nasional Multidisiplin Ilmu.

Vecellis, Carlo. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Willey & Son