

Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4.5

Sandrawira Anggraini, Sarjon Defit, Gunadi Widi Nurcahyo

Teknik Informatika, Universitas Putra Indonesia YPTK Padang

Jalan Raya Lubuk Begalung Padang, Sumatera Barat, Indonesia 25221

e-mail: sandrawiraanggraini07@gmail.com, sarjond@yahoo.uk, gunadiwidi@yahoo.co.id

Abstract

This study aims to optimize tire sales transactions. This optimization will affect the determination of tire inventory in the warehouse. The method used in this study is C4.5. Processed data is data on tire sales transactions on the CV. Roda Inti Mas. The results obtained in this study are able to manage the amount of inventory well, so that there is no shortage or excess stock. So, this research is very helpful in regulating tire transaction optimization.

Keywords: Sales; Request; Transaction; Stock; C4.5

Abstrak

Penelitian ini bertujuan untuk mengoptimalkan transaksi penjualan ban. Optimasi ini akan mempengaruhi penentuan persediaan ban di gudang. Metode yang digunakan dalam penelitian ini adalah C4.5. Data yang diolah adalah data transaksi penjualan ban pada CV. Roda Inti Mas. Hasil yang didapatkan pada penelitian ini adalah dapat mengatur jumlah persediaan dengan baik, sehingga tidak terjadi kekurangan atau kelebihan stok. Sehingga penelitian ini sangat membantu dalam pengaturan optimalisasi transaksi ban.

Kata Kunci: Penjualan; Permintaan; Transaksi; Persediaan; C4.5

1. Pendahuluan

Penjualan adalah transaksi antara dua pihak yaitu penjual dengan pembeli. Ketentuannya harus menguntungkan kedua belah pihak. Tingkat penjualan dapat dipengaruhi oleh kepuasan pembeli [1]. Salah satunya, penjual menggunakan kegiatan promosi sehingga menarik minat pembeli untuk membeli barang dan jasa yang ditawarkan, misalnya berupa pemberian *voucher* atau *discount* disetiap transaksinya. Promosi yang dilakukan *saler* bertujuan untuk menambah dan memperluas area pemasaran [2].

Proses transaksi pembelian barang oleh pembeli dan transaksi penjualan barang oleh penjual setiap harinya akan didokumentasikan dalam bentuk *record* data transaksi penjualan. Pada *record* data penjualan ada informasi yang bisa dikelola guna strategi pemasaran dan penjualan, di samping itu berguna sebagai informasi pola penjualan suatu produk periode tertentu [3]. Berdasarkan penelitian sebelumnya yang dilakukan, metodologi yang diusulkan terdiri dari fase seperti *preprocessing*, pemilihan atribut, klasifikasi berdasarkan pohon keputusan dan evaluasi kinerja. Sehingga hasil yang didapatkan berupa penemuan asosiasi baru dan dependensi dari atribut [4]. Kemudian metode *Data Mining* dipilih karena dapat menghasilkan model dan kriteria yang mudah diinterpretasikan oleh klasifikasi pelatihan data dan pengujian data dengan algoritma genetik. Sehingga penggunaan metode C4.5 dapat memberikan keputusan siapa yang berhak dan tidak berhak menerima Bantuan Langsung Masyarakat (BLSM) sesuai dengan kriteria yang ada [5]. Sementara algoritma C4.5 juga digunakan untuk menentukan tingkat penerimaan kembali rumah sakit pasien diabetes. Dataset yang digunakan dalam penelitian ini diambil dari UCI *Machine Learning Repository*, dan hasil implementasi algoritma C4.5 menghasilkan akurasi 74,5% [6]. Selanjutnya dengan algoritma C4.5 dapat menargetkan iklan yang dikirim pada pengguna, dengan hasil 0,6 sehingga memiliki korelasi yang baik dengan minat pengguna [7]. Penerapan lain dengan metode *Data Mining*, memungkinkan pengolahan multidimensi, sehingga menunjukkan hasil terbaik karena beroperasinya arus informasi saat membuat pohon keputusan, menggunakan *single-pass* menipis untuk menghindari pelatihan ulang, dan memotong cabang untuk meningkatkan model dibangun [8]. Melalui perbandingan lima algoritma klasifikasi *Data*

Mining (Gaussian Naïve Bayes, *Multinomial Naive Bayes*, *Bernoulli Naive Bayes*, *Decision Tree* dan SVM) untuk menemukan algoritma terbaik untuk kasus menentukan kelas UKT, hasil yang didapatkan bahwa algoritma klasifikasi *Data Mining* cocok untuk diterapkan dalam kasus penentuan kelas UKT siswa, dengan keputusan hasil prediksi Pohon menghasilkan nilai akurasi rata-rata 0,814 atau 81,4% [9]. *Data Mining* memainkan penting peran di memprediksi prestasi akademik siswa yang berfokus pada identifikasi lambat, rata-rata dan cepat pelajar antara siswa dan menampilkan prediksi menggunakan algoritma berdasarkan klasifikasi, dengan akurasi 95,454% dan karena itu *RandomTree* terbukti menjadi algoritma *classifier* berpotensi efektif dan efisien [10].

Pada penelitian ini diajukan sebuah metode dalam menggali informasi data transaksi penjualan menggunakan algoritma C4.5 untuk mengetahui dan menganalisis jumlah persediaan barang dalam transaksi penjualan. Data atau informasi yang ada selanjutnya akan diklasifikasikan untuk menemukan model yang menggambarkan kelas data yang nantinya akan diubah ke dalam bentuk dataset.

2. Metode Penelitian

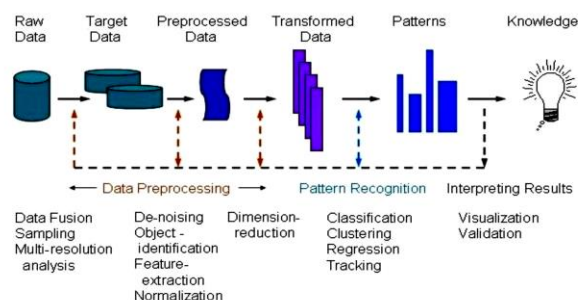
Metode penelitian merupakan sebuah usaha dalam menemukan suatu hal yang baru berdasarkan metode yang ilmiah, dalam penelitian terdapat tiga elemen penting yaitu sasaran, usaha untuk mencapai sasaran atau tujuan, serta adanya metode ilmiah. Tahapan yang dilakukan dalam penelitian ini meliputi identifikasi masalah, menganalisa masalah, studi literatur, mengumpulkan dan klasifikasi proses algoritma C4.5.

2.1. Analisa Data

Analisis data dilakukan dengan mengambil data barang berupa data penjualan ban dalam 3 tahun terakhir, disamping itu data persediaan ban juga dijadikan sebagai data analisa jumlah persediaan ban untuk masa yang akan datang sesuai klasifikasi dan merek ban yang ada.

2.2. Data Mining

Data Mining dapat didefinisikan sebagai ekstraksi informasi yang berarti dari sejumlah besar data. Ini telah digunakan dalam banyak aplikasi seperti data pendidikan data, *web*, *text mining* [4]. Jenis utama dari data mining meliputi klasifikasi, *cluster*, dan *association rule* [11].



Gambar 1. Tahapan proses *Data Mining*

Dalam proses *Data Mining* terdapat satu proses yang disebut klasifikasi. Klasifikasi adalah proses menemukan model yang menggambarkan kelas data. Dalam proses klasifikasi, data akan dimuat dalam bentuk *decision tree* (pohon keputusan).

2.3. Decision Tree

Decision Tree adalah pohon keputusan *classifier* yang mengklasifikasikan data ke dalam label kelas yang telah ditentukan. Sebuah pohon keputusan adalah struktur sederhana di mana setiap node non-terminal merupakan tes atau keputusan dan dianggap item data [4].

2.4. Klasifikasi

Klasifikasi sangat berperan penting dalam proses algoritma C4.5. Di mana algoritma C4.5 mampu menghasilkan pohon atau aturan model yang mudah diinterpretasikan dan diubah untuk struktur aturan *Query Language* (SQL), memiliki tingkat akurasi yang dapat diterima, dapat menangani diskrit dan numerik atribut, dan efisien dalam menangani atribut tipe diskrit [12].

2.5. Algoritma C4.5

Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- a. Pilih sebuah atribut sebagai *root*.
- b. Membuat cabang untuk setiap nilai.
- c. Perhatikan kasus di cabang.
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama [13].

Akar dari pohon keputusan algoritma C4.5 adalah atribut data dengan nilai *gain* tertinggi.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (1)$$

Di mana:

S : Himpunan kasus
A : Atribut
n : Jumlah partisi atribut A
|S_i| : Jumlah kasus pada partisi ke i
|S| : Jumlah kasus dalam S

Untuk mendapatkan nilai *gain*, terlebih dahulu hitung nilai entropy atribut data.

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

Di mana:

S : Himpunan Kasus
n : Jumlah partisi S
p_i: Proporsi dari S_i terhadap S

3. Hasil dan Pembahasan

Hasil yang akan dicapai dalam penelitian ini adalah mengetahui jumlah persediaan ban yang harus disediakan oleh CV. RODA INTI MAS dalam memenuhi permintaan dan penjualan ban dengan menggunakan metode klasifikasi algoritma C4.5.

3.1. Implementasi Algoritma C4.5

Proses implementasi algoritma C4.5 dalam proses *Data Mining* melalui beberapa tahapan berikut:

1. Transformasi Data
2. Mencari Nilai Entropy Total
3. Mencari Nilai Entropy Semua Atribut
4. Mencari Nilai Gain Setiap Atribut
5. Menentukan Node Akar
6. Mencari Node Cabang

Data yang akan diolah menggunakan algoritma C4.5 adalah data transaksi penjualan ban CV. RODA INTI MAS, dan jumlah data yang akan diolah adalah sebanyak 30 buah data transaksi penjualan selama 3 tahun.

Dari Tabel 1 didapatkan hasil transformasi data seperti pada Tabel 2, dengan atribut kategori, satuan, pembeli, dan persediaan.

Tabel 1. Contoh data transaksi penjualan ban

No	Jenis Ban	Kategori Ban	Satuan	Pembeli	Jumlah Jual	Persediaan
1	JK JET TUFDX 1000-20	JK	SET	PT. TRANSCO UTAMA	577	577
2	BL APL XT7 900- 20	APL	SET	ARU PADATU	20	20
3	MA 326 1000-20	APL	SET	ASIONG	70	70
4	APL TXS 1000-20	APL	SET	KONTAN	400	400
5	APL ENDRC CD	APL	SET	PT. SARANA AGRO W	176	176
-	-	-	-	-	-	-
30	SM95 750 KOS	SM	BUAH	CV.BINTANG S. PRIMA	16	16

Tabel 2. Transformasi data

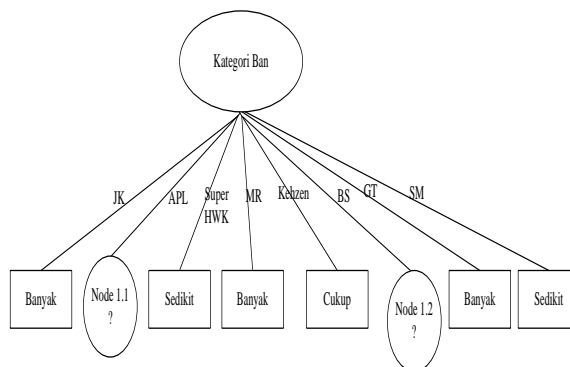
Kategori	Satuan	Pembeli	Persediaan
JK	SET	Bukan Langganan	Banyak
APL	SET	Bukan Langganan	Sedikit
APL	SET	Bukan Langganan	Cukup
APL	SET	Langganan	Banyak
APL	SET	Bukan Langganan	Banyak
APL	SET	Bukan Langganan	Banyak
APL	SET	Bukan Langganan	Sedikit
APL	SET	Bukan Langganan	Cukup
SUPER HWK	SET	Langganan	Sedikit
SUPER HWK	SET	Bukan Langganan	Sedikit
MR	SET	Bukan Langganan	Banyak
SUPER HWK	SET	Bukan Langganan	Sedikit
APL	SET	Bukan Langganan	Banyak
APL	SET	Bukan Langganan	Banyak
APL	SET	Langganan	Banyak
APL	SET	Bukan Langganan	Sedikit
KEHIZEN	SET	Langganan	Cukup
BS	SET	Bukan Langganan	Sedikit
BS	BUAH	Langganan	Sedikit
BS	BUAH	Bukan Langganan	Sedikit
BS	BUAH	Langganan	Cukup
BS	BUAH	Bukan Langganan	Banyak
APL	BUAH	Bukan Langganan	Sedikit
APL	BUAH	Bukan Langganan	Sedikit
APL	BUAH	Langganan	Sedikit
APL	BUAH	Bukan Langganan	Sedikit
APL	BUAH	Langganan	Sedikit
GT	BUAH	Bukan Langganan	Banyak
GT	BUAH	Bukan Langganan	Banyak
SM	BUAH	Bukan Langganan	Sedikit

Dari transformasi data pada Tabel 2, didapatkan hasil perhitungan entropi dan *gain* pada masing-masing atribut data seperti yang terlihat pada Tabel 3.

Tabel 3. Hasil perhitungan entropy total dan gain tertinggi

Perhitungan Node 1						
Jumlah Kasus		Sedikit	Cukup	Banyak	Entropy	Gain
Total	30	15	4	11	1,4183	
Kategori Ban						0,4401
JK	1	0	0	1	0,0000	
APL	16	8	2	6	1,4056	
SUPER HWK	3	3	0	0	0,0000	
MR	1	0	0	1	0,0000	
KEHZEN	1	0	1	0	0,0000	
BS	5	3	1	1	1,3709	
GT	2	0	0	2	0,0000	
SM	1	1	0	0	0,0000	
Satuan						0,0544
Set	18	7	3	8	1,4806	
Buah	12	8	1	3	1,1887	
Pembeli						0,0341
Langganan	8	4	2	2	1,5	
Bukan Langganan	22	11	2	9	1,3420	

Dari Tabel 3, didapatkan bentuk pohon keputusan sementara seperti pada Gambar 2.



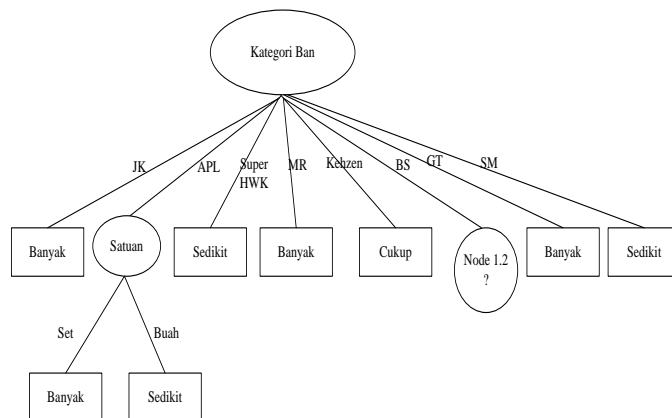
Gambar 2. Pohon keputusan node 1

Dari gambar 2, dapat dilihat jika pada atribut kategori ban “APL” dan “BS” masih terdapat cabang yang harus dicari. Pencarian cabang pada node 1.1 dapat dilihat pada Tabel 4.

Tabel 4. Hasil perhitungan entropy total dan gain tertinggi untuk atribut kategori ban “APL”

Node		Jumlah Kasus	Sedikit	Cukup	Banyak	Entropy	Gain
1.1	Kategori Ban- APL	16	8	2	6	1,4056	
	Satuan						0,4188
1.1	Set	11	3	2	6	1,4353	
	Buah	5	5	0	0	0,0000	
1.1	Pembeli						0,3112
	Langganan	4	2	0	2	0,0000	
1.1	Bukan Langganan	12	6	2	4	1,4591	

Dari Tabel 4 didapatkan lagi cabang baru untuk atribut kategori ban “APL” yaitu satuan.



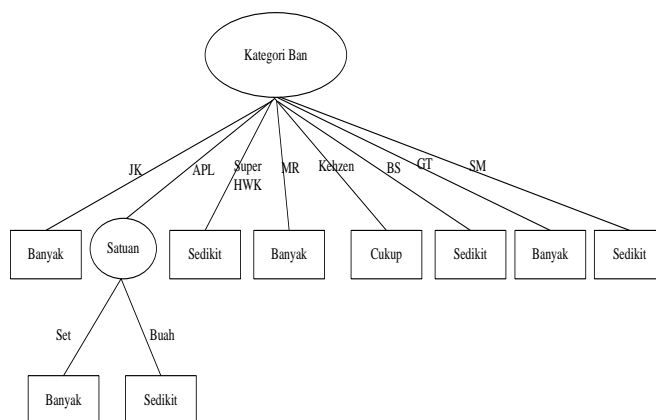
Gambar 3. Pohon keputusan node 1.1

Selanjutnya untuk kategori ban “BS” dapat dilihat pada Tabel 5.

Tabel 5. hasil perhitungan *entropy* total dan *gain* tertinggi untuk atribut kategori ban “BS”

Node	Jumlah Kasus	Sedikit	Cukup	Banyak	Entropy	Gain
Kategori Ban- BS	5	3	1	1	1,3709	
Satuan						0,1709
Set	1	1	0	0	0,0000	
Buah	4	2	1	1	1,5	
Pembeli						0,0000
Langganan	2	1	1	0	0,0000	
Bukan Langganan	3	2	0	1	0,0000	

Dari Tabel 5, didapatkan bentuk pohon keputusan terakhir yang dihasilkan dari pengolahan data transaksi penjualan dan persediaan ban sebanyak 30 buah data.

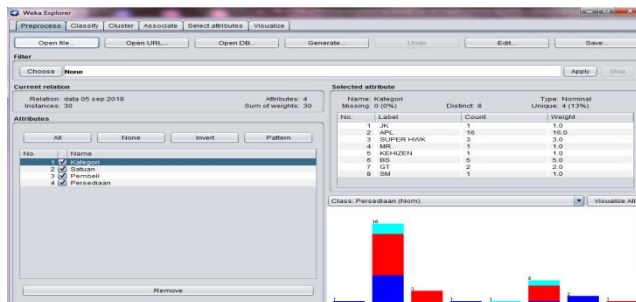


Gambar 4. Pohon keputusan node terakhir

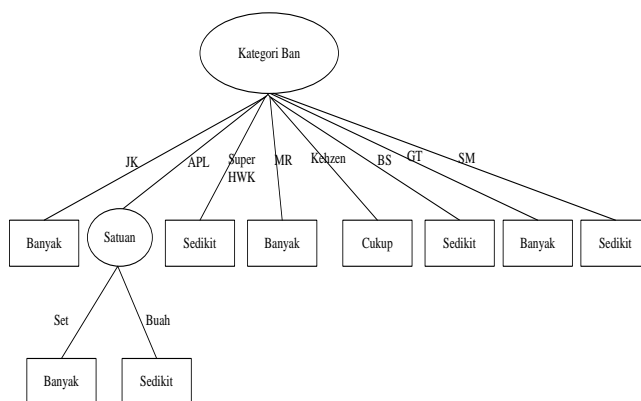
3.2. Implementasi *Software Weka*

Weka merupakan sebuah *software Data Mining* yang berupa *machine learning*. *Weka* memiliki fasilitas komprehensif dalam menganalisa data yaitu perangkat *preprocessing* data, algoritma pembelajaran, serta metode evaluasi. *Software Weka* memiliki *interface* berbasis

GUI. Pada *Weka* data *input explorer* tersedia dalam beberapa format data, di antaranya CSV (*comma separated value*), dan *binary*.



Gambar 5. Open file.CSV pada aplikasi *Weka*



Gambar 6. *Decision tree* pada *Weka*

3.3. Tes Data

Hasil yang diperoleh dari beberapa kali pengujian data transaksi penjualan dan persediaan ban dengan jumlah data yang berbeda dan lebih banyak didapatkan hasil tes data 30 buah data dengan perhitungan manual dan *Weka*, hasil yang didapatkan sama. Kemudian dilakukan lagi pengujian pada data dengan jumlah data 60, 90, dan 100 buah data dan hasil yang diperoleh pada pohon keputusan yang terbentuk juga sama.

Tabel 6. Hasil tes data

Jumlah Data Tes	Correctly Classified Instances		Incorrectly Classified Instances	
	Jumlah Data	Akurasi (%)	Jumlah Data	Akurasi (%)
60	46	76,67	14	23,33
90	67	74,44	23	25,56
100	75	75	25	25

Dari pengujian dan tes data yang dilakukan, hasil pohon keputusan yang didapatkan selalu memiliki akar dengan atribut Kategori, dan memiliki cabang satuan. Dari pohon keputusan pada gambar 7, dapat dibuat *rule-rule* pada analisis *Data Mining* transaksi penjualan ban menggunakan algoritma C4.5 sebagai berikut:

1. *If* kategori *is* JK *Then* persediaan *is* banyak.
2. *If* kategori *is* APL *and* satuan *is* set *Then* persediaan *is* banyak.
3. *If* kategori *is* APL *and* satuan *is* buah *Then* persediaan *is* sedikit.
4. *If* kategori *is* super HWK *Then* persediaan *is* sedikit.
5. *If* kategori *is* MR *Then* persediaan *is* banyak.
6. *If* kategori *is* KEHIZEN *Then* persediaan *is* cukup.
7. *If* kategori *is* BS *Then* persediaan *is* sedikit.

8. *If kategori is GT Then persediaan is banyak.*
9. *If kategori is SM Then persediaan is sedikit.*

4. Kesimpulan

Berdasarkan hasil pengujian dan implementasi *Data Mining* menggunakan algoritma C4.5 dalam melakukan pengklasifikasian data penjualan Ban di CV. RODA INTI MAS, maka didapatkan kesimpulan bahwa proses pengujian pada data tes, dilakukan sebanyak 3 kali pengujian yaitu terhadap 60 data, 90 data, dan 100 data, dengan tingkat akurasi antara 74% – 76%, dengan atribut akar (*root*) adalah kategori, dan atribut yang menjadi cabang adalah satuan, sedangkan atribut yang menjadi ranting adalah banyak dan sedikit.

Referensi

- [1] G. Srinath, G. Jason, M. Murali K, and S. Shrihari. "Assessing sales contest effectiveness: the role of salesperson and sales district characteristics". *Marketing Letter*, Springer, vol. 27, no.3, pp. 589-602, 2015.
- [2] Casadei D, Serra G, Tani K. "Implementation of a Direct Control Algorithm for Induction Motors Based on Discrete Space Vector Modulation". *IEEE Transactions on Power Electronics*, vol.15, no. 4, pp. 769-777, 2007.
- [3] Van Heerde H. J., Neslin S. A. "Sales Promotion Models". *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science, Springer, vol. 254, 2017.
- [4] Natsuki S., Katsutoshi Y. "The influence of sales areas and bargain sales on customer behavior in a grocery store". *Neural Computing and Application*, vol. 26, no. 2, pp. 355-361, Springer-Verlag London, 2014.
- [5] Shanmugarajeshwari V., Lawrance R. "An Assay of Teachers' Attainment using Decision Tree Based Classification Techniques". *International Conference on circuits Power and Computing Technologies [ICCPCT]*, 2017.
- [6] Gustian D., Yulitasari S., Hundayani D., Muslih, Nunik. "Comparison of C4.5 Method Based Optimization Algorithm to Determine Eligibility of Beneficiaries of Direct Community Assistance (Case Study: Kelurahan Cicurug)", *Int. Conf. On Comp. Eng. and Desg.* 2017.
- [7] Tamin F., Satvika I N. M. "Implementation of C4.5 Algorithm to Determine Hospital Readmission Rate of Diabetes Patient". *4th International Conference on New Media Studies*, 2017.
- [8] Xuefei W., Yan S. "Design and Implementation of Targeting Advertising System based on C4.5 Algorithm". *4th International Conference on Computer Science and Network Technology (ICCSNT)*, 2015.
- [9] Kravets Alla G, Skorobogatchenko Dmitry A, Skorobogatchenko, Salnikova Natalia A, Orudjev Nazim Y, Poplavskaya Olga V. "The Traffic Safety Management System in Urban Conditions Based on the C4.5 Algorithm". *Moscow Workshop on Electronic and Networking Technologies (MWENT)*, 2018.
- [10] Yasir Utomo M. N., Permanasari A. E., Tungadi Eddy, Syamsuddin I. "Determining Single Tuition Fee of Higher Education in Indonesia a Comparative Analysis of Data Mining Classification Algorithms". *International Conference on New Media Studies*, 2017.
- [11] Mhetre V., Nagar M. "Classification based data mining algorithms to predict slow, average and fast learners in educational system using Weka". *International Conference on Computing Methodologies and Communication (ICCMC)*, 2017.
- [12] Condrobimo Raharto A., Bahtiar S. A., Trisetyarso A. "Data Mining Technique with Cluster Anaysis Use K-Means Algorithm for LQ45 Index On Indonesia Stock Exchange". *International Conference on Information and Communications Technology (ICOIACT)*, 2018.
- [13] Wang U., Yan S., "Design and Implementation of Targeting Advertising System based on C4.5 Algorithm". *4th International Conference on Computer Science and Network Technology (ICCSNT)*, 2015.
- [14] Kravets Alla G, Skorobogatchenko Dmitry A, Skorobogatchenko, Salnikova Natalia A, Orudjev Nazim Y, dan Poplavskaya Olga V. "The Traffic Safety Management System in Urban Conditions Based on the C4.5 Algorithm". *Moscow Workshop on Electronic and Networking Technologies (MWENT)*, 2018.