

Implementasi Data Mining untuk Memprediksi Jumlah Pendaftaran Calon Siswa/Siswi Baru Pada Tahun Berikutnya Di Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan Menggunakan Metode Algoritma Classification And Regression Tree (CART)

Khoirunnisa Nasution *, Hafizah**, Jufri Halim**

* Program Studi Sistem Informasi, STMIK Triguna Dharma

** Program Studi Sistem Informasi, STMIK Triguna Dharma

** Program Studi Sistem Informasi, STMIK Triguna Dharma

Article Info

Article history:

Received Jun 12th, 201x

Revised Aug 20th, 201x

Accepted Aug 26th, 201x

Keyword:

Data Mining

CART

Yayasan Perguruan Al-Hizra
Mis Alwashliyah Tanjung
Morawa Pekan

Prediksi Jumlah Siswa/Siswi
pada Tahun Berikutnya

ABSTRACT

Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan memiliki berbagai kendala dan permasalahan dalam jumlah siswa/siswi yang masuk dalam setiap tahunnya, sehingga menuai pro dan kontra untuk masyarakat khususnya orang tua dari siswa/siswi yang akan ditolak ataupun diterima dari yayasan tersebut. Oleh karena itu untuk menyelesaikan masalah ini maka dibangunlah sistem untuk memprediksi jumlah masuknya siswa/siswi untuk tahun berikutnya.

Data mining adalah bagian dari sistem informasi berbasis komputer untuk memperbaiki berbagai permasalahan yang ada. Salah satu metode yang digunakan dalam pemecahan masalah ini adalah metode Classification And Regression Tree (CART).

Hasil dari penelitian ini menunjukkan bahwa dengan metode CART ini akan menghasilkan prediksi yang benar atau tidak benar jumlah siswa/siswi yang akan masuk untuk tahun berikutnya berdasarkan perhitungan dari kriteria-kriteria yang sudah ditentukan dan dimasukkan ke dalam sebuah aplikasi.

Copyright © 2021 STMIK Triguna Dharma.

All rights reserved.

Corresponding Author: *First Author

Nama : Khoirunnisa Nasution

Program Studi : Sistem Informasi

Kampus : STMIK Triguna Dharma

Email : nstnisa148@gmail.com

1. PENDAHULUAN

Siswa siswi adalah sekelompok individu yang memiliki sifat berbeda antara siswa satu dengan siswa lainnya dalam segala hal, baik kecerdasan, emosi, kebiasaan belajar, dan kemampuan belajarnya [1].

Pada saat ini pendidikan sangat berpengaruh untuk semua kalangan dari yang muda hingga dewasa sekalipun, oleh karena itu setiap tahunnya jumlah pendaftaran calon siswa/siswi baru di setiap sekolah khususnya Sekolah dasar di Yayasan Perguruan Al-Hizra mis Alwashliyah Tanjung Morawa Pekan selalu meningkat. Dengan adanya peningkatan angka pendaftaran ini membuat seluruh guru yang ada di Yayasan

Perguruan Al-Hizra mis Alwashliyah ini menjadi kesulitan jika calon siswa/siswi yang mendaftar lebih dari kuota kelas yang ada.

Dilihat dari permasalahan diatas, maka ddengan adanya penelitian ini yang berjudul **“Implementasi Data Mining Untuk Memprediksi Jumlah Pendaftaran Calon Siswa/Siswi Pada Tahun Berikutnya Di Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan Menggunakan Metode Algoritma Classification And Regression Tree (CART).**

2. METODE PENELITIAN

Metode penelitian ini sangat bermanfaat agar dapat membantu penyelesaian kasus dan apa saja yang berhubungan erat dengan kasus yang akan diamati, adapun tahap-tahapannya adalah :

2.1 Data Collecting

Yaitu mengumpulkan semua data yang berhubungan dan akan diproses agar tercapai penyelesaian masalah dalam kasus tersebut. Adapun cara-caranya :

1. Metode Pengamatan (observasi)

Yaitu metode pengumpulan data dengan mengobservasi langsung dilapangan, serta mengamati masalah-masalah yang ada di Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan.

2. Metode Wawancara

Yaitu wawancara dengan staff/guru yang mengajar di Yayasan Perguruan Al-Hizra mis Alwashliyah Tanjung Morawa Pekan mengenai masalah-masalah serta solusi mengenai permasalahan dari kasus tersebut.

2.2 Studi Literatur

Studi literatur merupakan bantuan yang digunakan untuk penyelesaian dari seluruh kasus yang adadi deskripsi, dalam studi literatur ini meggunakan jurnal jurnal nasional dan buku buku untuk membantu penyelesaian masalah yang ada di Yayasan Perguruan Al-Hizra mis Alwashliyah Tanjung Morawa Pekan.

a. Data Mining

Data mining adalah kegiatan mengekstrasi atau mengembangkan suatu data serta menambang pengetahuan dari data yang berukuran/berjumlah besar, sehingga menjadi informasi yang berguna untuk pengembangan[2].

b. Pengertian Prediksi

Prediksi (forecasting) merupakan suatu alat bantu yang penting dalam perencanaan yang fektif dan efisien dalam bidang apapun sehingga keputusan yang diambil sangat signifikan.[3]

c. Algoritma Classification And Regression Tree (CART)

Cart dikembangkan oleh Leo Breimen, Jerome H, Freidman, Richard A, Olshen, dan Charles J. Stone. Metode cart ini merupakan suatu metodologi statistik untuk analisis klasifikasi, baik untuk variabel dependen berbentuk kategorik maupun kontinu sehingga menghasilkan pohon klasifikasi bila variabel dependennya kategorik dan pohon regresi bila variabel dependennya kontinu. Prinsip kerja dari analisis CART disebut binary recursie partitioning. Istilah “recursive” mengacu pada proses pemisahan simpul dilakukan, istilah “partitioning” mengacu pada data dipisah menjadi bagian-bagian atau partisi-partisi yang lebih kecil.[4]

Berikut ini adalah penyelesaian menggunakan algoritma CART, yaitu sebagai berikut ini :

1. Menentukan dan menyusun calon cabang terhadap seluruh variabel prediktor secara lengkap.

Misalnya : calon cabang untuk ariabel kategori jenis kelamin :

1. Laki-laki, Perempuan
2. Perempuan, Laki-laki

2. Menghitung nilai PL dan PR

$$PL = \frac{\text{jumlah catatan pada calon cabang kiri } tL}{\text{jumlah catatan pada data latihan}}$$

$$PR = \frac{\text{jumlah catatan pada calon cabang kanan } tR}{\text{jumlah catatan pada data latihan}}$$

3. Menghitung nilai $p(j|tL)$ dan $p(j|tR)$

$$P(j|tL) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kiri } tL}{\text{jumlah catatan nokta keputusan } t}$$

$$P(j|tR) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kanan } tR}{\text{jumlah catatan nokta keputusan } t}$$

4. Menghitung 2PLPR

Yaitu jumlah catatan pada calon kiri dan calon kanan
(2*PL*PR)

5. Menghitung $Q(s|t)$ dan $\emptyset(s|t)$

$$Q(s|t) = \sum |P(j|tL) - P(j|tR)|$$

$$\emptyset(s|t) = 2PLPR\emptyset(s|t)$$

6. Menentukan nilai literasi

Jika nilai literasi awal memiliki nilai yang besar pada nokta nilai dasar maka literasi akan menjadi bercabang sehingga muncullah nilai literasi 2, namun jika nilai literasi kecil maka literasi terhenti.

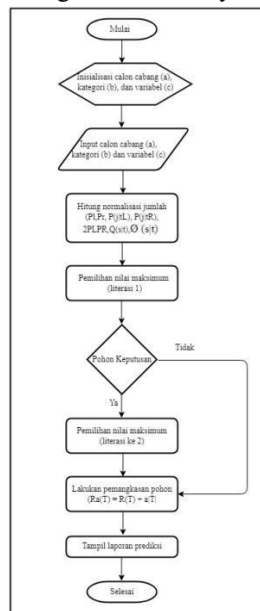
3. ANALISA DAN HASIL

3.1 ANALISA

Algoritma sistem merupakan penjelasan langkah-langkah penyelesaian masalah dalam perancangan data mining untuk memprediksi jumlah pendaftaran siswa/siswi untuk tahun selanjutnya dengan menggunakan metode algoritma CART. Hal ini dilakukan untuk menentukan keberhasilan yayasan al-hizra mis alwashliyah Tanjung Morawa Pekan dalam menghadapi persaingan dibidang pendidikan.

1. Flowchart dari Metode Penyelesaian Algoritma Cart

Berikut ini adalah *flowchart* dari metode Algoritma *CART* yaitu sebagai berikut :



Gambar 3.1 flowchat dari algoritma cart

2. Algoritma Cart

Berikut ini adalah penyelesaian menggunakan algoritma *Cart*, yaitu sebagai berikut ini :

Tabel 1 Data Siswa

No	Nama	Jenis Kelamin	Umur	Biaya	Status
1	Aditya Barus	Laki-laki	5	Rp.700.000	Diterima
2	Afifah Hasanah	Perempuan	6	Rp.700.000	Diterima

Keterangan :

Dari tabel diatas, dapat dijelaskan bahwa data yang diperoleh merupakan data hasil olahan pertahunnya, yaitu jumlah siswa yang mendaftar setiap tahunnya beserta identitasnya yang diperoleh dari Yayasan Al-Hizra mis Alwashliyah Tanjung Morawa Pekan. Pada tahap ini dilakukan pembersihan data primer menjadi data yang sudah siap untuk dihitung menggunakan algoritma cart.

Pada kasus ini, yang menjadi variabel prediktor adalah :

⇒ Calon cabang untuk variabel kategori jenis kelamin adalah :

Laki-laki, perempuan

Perempuan, laki-laki

⇒ Calon cabang untuk variabel prediktor adalah umur :

Umur : Cukup, Lebih

Umur : Lebih, Cukup

⇒ Calon cabang untuk variabel prediktor biaya adalah :

Biaya : Lumayan murah, Murah, Mahal

Biaya : Mahal, Murah, Lumayan murah

⇒ Calon cabang untuk variabel prediktor status adalah :

Status : Diterima, Tidak diterima

Status : Tidak diterima, Diterima

Tabel 2 Keterangan Kategori Data siswa

No	Nama Kategori	Jumlah
1	Laki-laki	57
2	Perempuan	45
3	Cukup	97
4	Lebih	5
5	Murah	40
6	Lumayan murah	40
7	Mahal	22
8	Diterima	102
9	Tidak diterima	0

Tabel 3 Keterangan Kategori Data biaya siswa pertahun

No	Nama Kategori	Nilai
1	Murah	Rp.700.000
2	Lumayan murah	Rp.802.000
3	Mahal	Rp.926.000

Tabel 4 Data Calon Cabang

Calon cabang	Calon cabang kiri (tL)	Calon cabang kanan (tR)
1	Jenis kelamin : laki-laki	Jenis kelamin : perempuan
2	Jenis kelamin : perempuan	Jenis kelamin : laki-laki
3	Umur : cukup	Umur : lebih
4	Umur : lebih	Umur : cukup
5	Biaya : lumayan murah	Biaya : mahal
6	Biaya : murah	Biaya : murah
7	Biaya : mahal	Biaya : lumayan murah
8	status : diterima	Status : tidak diterima
9	status : tidak diterima	Status : diterima

Hasil perhitungan nilai kesesuaian bagi tiap calon cabang yang masih terdapat dalam daftar calon cabang adalah, berikut ini :

Tabel 5 Hasil Konversi Data

Calon Cabang	PL	PR	Status	P(j tL)	P(j tR)	2PLPR	Q(s t)	Ø(s t)
1	0.558	0.441	Diterima Tidak diterima	1 0	1 0	0.492	0	0
2	0.441	0.558	Diterima Tidak diterima	0.739 0	1.266 0	0.492	-0.557	-0.274

Dari referensi metode CART yang telah dijelaskan, berikut adalah langkah-langkahnya :

1. Menghitung nilai PL dan PR

$$1. \quad PL = \frac{57}{102} = 0.558 \quad PR = \frac{45}{102} = 0.441$$

$$2. \quad PL = \frac{54}{102} = 0.441 \quad PR = \frac{57}{102} = 0.558$$

2. Menghitung nilai p(j|tL) dan p(j|tR)

Berikut adalah jumlah data hasil konversi j pada calon cabang tL, tR dan jumlah pada noktah keputusan t yaitu sebagai berikut :

$$1. \quad P(j|tL) = \frac{57}{57} = 1 \quad P(j|tR) = \frac{45}{45} = 1$$

$$2. \quad P(j|tL) = \frac{45}{57} = 0.739 \quad P(j|tR) = \frac{57}{45} = 1.266$$

3. Menghitung 2PLPR

Berikut adalah jumlah catatan pada calon kiri dan calon kanan yaitu :

$$1. \quad 2PLPR = 2 \times 0.558 \times 0.441 = 0.492$$

$$2. \quad 2PLPR = 2 \times 0.441 \times 0.558 = 0.492$$

4. Menghitung Q (s|t) dan Ø (s|t)

$$1. \quad Q(s|t) = \sum |P(j|tL) - P(j|tR)|$$

$$= |1 - 1| + |0 - 0|$$

$$= 0 + 0$$

$$= 0$$

$$2. \quad Q(s|t) = \sum |P(j|tL) - P(j|tR)|$$

$$= |0.709 - 1.266| + |0 - 0|$$

$$= (-0.557) + 0$$

$$= (-0.557)$$

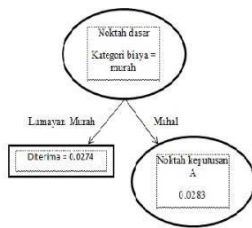
Berikut adalah nilai $\emptyset (s|t)$ yaitu :

1. $\emptyset (s|t) = 2PLPR\emptyset(s|t)$
 $= 2 \times 0.558 \times 0.441 \times 0 = 0$
 $= 0$
2. $\emptyset (s|t) = 2PLPR\emptyset(s|t)$
 $= 2 \times 0.441 \times 0.558 \times (-0.557)$
 $= 0.274$

Dari perhitungan nilai kesesuaian terhadap dalam cabang diatas dapat diambil analoginya sehingga dapat diperiksa perhitungan sejenis terhadap calon cabang yang lain.

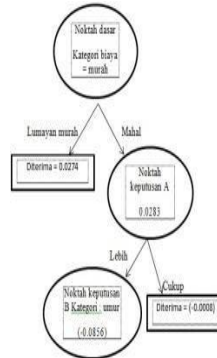
5. Menentukan cabang untuk pohon keputusan

a. Menentukan nilai literasi 1



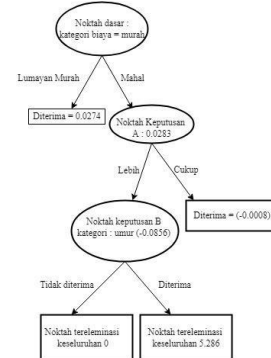
Gambar 1 literasi 1

b. Menentukan literasi 2



Gambar 2 literasi 2

c. Menentukan Literasi #



Gambar 3 literasi 3

Dari gambar 2 tampak :

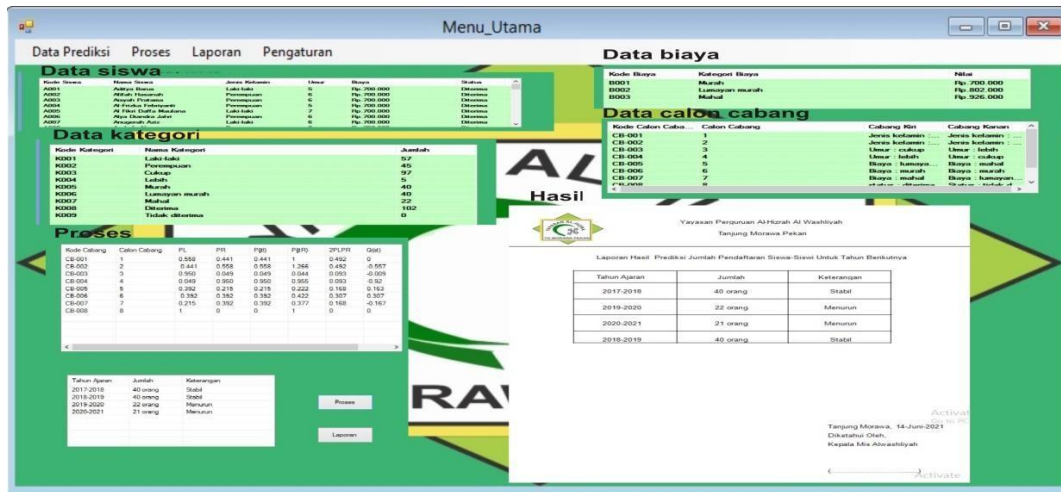
1. Pada noktah dasar , masih berhadapan dengan kategori biaya, yaitu biaya murah.
2. Calon cabang no 6 kini telah benar-benar menjadi cabang, adapun cabang kirinya yaitu cabang yang memenuhi syarat kategori = lumayan murah dengan status diterima dan berjumlah 0.0274 sehingga noktah terminasi akan dihasilkan.
3. Cabang kanannya, yaitu cabang memenuhi syarat kategori biaya = Mahal berjumlah 0.0283 sehingga dihasilkan noktah keputusan, karena merupakan noktah keputusan pertama, dapat disebut noktah keputusan A, noktah ini akan bercabang lebih lanjut karena bukan termasuk noktah terminasi.

3.2 HASIL

Pada tahun 2020-2021 jumlah siswa/siswi yang mendaftar diYayasan Perguruan Al-Hizra mis Alwashliyah Tanjung Morawa Pekan akan meningkat menjadi 32 orang dari tahun sebelumnya. Berdasarkan jumlah yang siswa yang diterima pada noktah terminalisasi akhir dari keseluruhan berjumlah 5.286.Keterangan jika 100 % = 40 orang dikarenakan fasilitas kelas hanya untuk 40 orang siswa/siswi maka $5.286/100 = 52.86$ maka dibulatkan menjadi 53 % setelah itu $53 \times 40/100 = 21.2$ atau 21 orang. Berikut ini adalah form data proses yaitu sebagai berikut :

Tabel 6 jumlah prediksi

Tahun ajaran	Jumlah	Keterangan
2017-2018	40 orang	Stabil
2018-2019	40 orang	Stabil
2019-2020	22 orang	Menurun
2020-2021	21 orang	Menurun



Gambar 3 Tampilan form proses CART



Gambar 4 Tampilan form laporan

4. KESIMPULAN

Berdasarkan analisa dari kasus yang terjadi maka diangkatlah tentang Implementasi Data Mining untuk Memprediksi Jumlah Siswa/siswi pada Tahun berikutnya di Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan Menggunakan Metode *Classification And Regression Tree (CART)*. Maka terdapat beberapa kesimpulan, yaitu :

1. Implementasi data mining dengan menggunakan metode *Classification And Regression Tree (CART)* dapat diterapkan dan digunakan untuk menentukan jumlah siswa/siswi yang akan masuk hingga tahun berikutnya khususnya di Yayasan Perguruan Al-Hizra Mis Alwashliyah Tanjung Morawa Pekan.
2. Pemodelan dari sistem yang digunakan dirancang dengan baik menggunakan UML dan *Flowchart*. UML terdiri dari *use Case Diagram*, *Activity Diagram* dan *Class Diagram*.




UCAPAN TERIMA KASIH

Puji syukur penulis ucapkan kepada Tuhan yang Maha Esa, karena rahmatnya lah peneliti dapat menyelesaikan skripsi ini dengan baik. Terima kasih yang sebesar-besarnya kepada Orang Tua tercinta, yang telah banyak memberi dukungan secara materi dan kasih sayang yang berlimpah.oleh karena itu, dengan segala krendahan hati diucapkan terima kasih sebesar-besarnya kepada Bapak Rudi Gunawan, SE., M.Si, selaku Ketua Sekolah Tinggi Manajemen Informatika dan Komputer Triguna Dharma Medan.Bapak Mukhlis Ramadhan, SE, M.Kom, selaku Wakil Ketua 1 Bidang Akademik STMIK Triguna Dharma.Bapak Puji Sari Ramadhan, S.Kom., M.Kom, selaku Ketua Program Studi Sistem Informasi STMIK Triguna Dharma.Ibu Hafizah S.Kom., M.Kom selaku Dosen Pembimbing I.Bapak Jufri Halim, SE., MM, selaku Dosen Pembimbing II.Bapak dan Ibu Dosen serta staff STMIK Triguna Dharma Medan yang telah memberikan bimbingan dan pengajaran selama menyelesaikan studi sebagaimana mestinya.

REFERENSI

- [1] S. Ali *et al.*, “PENGELOLAAN PENERIMAAN PESERTA DIDIK BARU (PPDB) PADA SD MUHAMMADIYAH PROGRAM KHUSUS BANYUDONO,” *ثثثثثث*, vol. ث ث ث ث ث ث ث ث, no. 2, p. ث ث ث ث ث ث ث ث, 2018, doi: 10.1051/mateconf/201712107005.
- [2] A. Fadli, “Konsep Data Minning,” *Konsep Data Min.*, pp. 1–9, 2003.
- [3] K. N. Ulfa and M. Syahrizal, “Perancangan Aplikasi Prediksi Jumlah Siswa Baru pada Yayasan Cerdas Murni menggunakan Exponential Smoothing,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 3, no. 6, pp. 59–64, 2016, doi: 10.1109/IPDPS.2009.5161034.
- [4] J. Gaussian, “Pembentukan Pohon Klasifikasi Biner Dengan Algoritma Cart (Classification and Regression Trees),” vol. 5, no. 10, pp. 583–592, 2016.

BIBLIOGRAFI PENULIS

	<p>Khoirunnisa Nasution Anak perempuan kelahiran Ngeri lama, 29 juni 1998, Anak Pertama dari 4 bersaudara ini merupakan mahasiswi STMIK TRIGUNA DHARMA yang sedang dalam proses menyelesaikan skripsi.</p>
	<p>Hafizah, S.Kom., M.Kom</p>
	<p>Jufri Halim, S.E., MM.</p>