

## PENERAPAN ALGORITMA CART KLASIFIKASI SOSIAL EKONOMI MASYARAKAT KELURAHAN AMESSANGENG

**Irmayani**

irnia2104@gmail.com

Universitas Cokroaminoto Palopo

### ABSTRAK

Data mining meliputi beberapa metode untuk membantu pengambilan keputusan salah satunya adalah metode klasifikasi. Metode klasifikasi meliputi beberapa cara salah satunya adalah *Algoritma Clasificassion And Regression Trees*. Model klasifikasi pada CART menggunakan cara *decision tree* pada pembentukan keputusan. Algoritma CART ini merupakan algoritma yang sederhana namun memperoleh data yang akurat sehingga dapat digunakan untuk membantu lembaga-lembaga pemerintahan yang membutuhkan. Data yang digunakan adalah data social ekonomi masyarakat kelurahan Amessangeng. Hasil dari penelitian ini membentuk *decision tree* yang dapat membantu pengambilan keputusan terhadap kebijakan social ekonomi di Kelurahan Amessangeng

**Kata Kunci:** Algoritma, CART, Data Mining, Ekonomi, Klasifikasi, Sosial.

### 1. Pendahuluan

Teknologi internet telah mengubah dunia, dengan banyaknya data yang tersedia sehingga lebih memudahkan untuk mengaksesnya. Data digunakan sebagai acuan untuk melihat kondisi yang pernah terjadi bahkan dapat memprediksi kejadian yang akan terjadi melalui histori data tersebut. Dengan bantuan data mining kejadian-kejadian dapat diprediksi jauh sebelum waktu kejadian. Selain itu juga dapat membantu pengambilan keputusan untuk mengantisipasi sebelum terjadinya suatu kejadian.

Data mining meliputi beberapa metode untuk membantu pengambilan keputusan salah satunya adalah metode klasifikasi. Metode klasifikasi meliputi beberapa cara salah satunya adalah *Algoritma Clasificassion And Regression Trees*. Model klasifikasi pada CART menggunakan cara *decision tree* pada pembentukan keputusan. Algoritma CART ini merupakan algoritma yang sederhana namun memperoleh data yang akurat sehingga dapat digunakan untuk membantu

lembaga-lembaga pemerintahan yang membutuhkan.

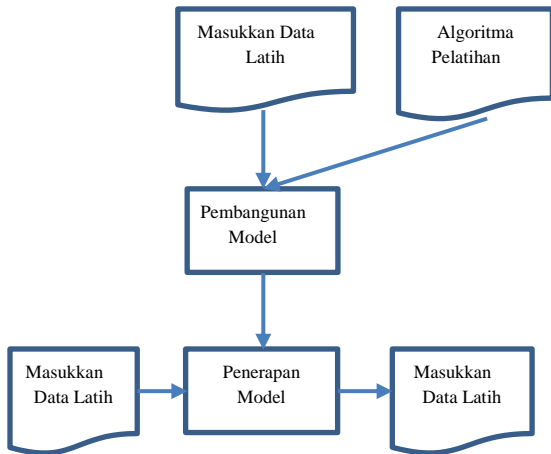
Pemanfaatan metode klasifikasi CART digunakan pada penelitian ini dengan menggunakan data social ekonomi kelurahan Amessangeng Kota Palopo. Dengan mengambil sampel pada masyarakat dan menggunakan variabel-variabel predictor yang dapat digunakan menghasilkan suatu kesimpulan demi menghasilkan informasi yang akurat yang dapat digunakan oleh pihak yang membutuhkan terkhusus lembaga-lembaga pemerintahan seperti BAPPEDA.

Penelitian yang menggunakan CART mampu dapat data yang akurat sehingga mampu mengklasifikasikan data social ekonomi masyarakat kelurahan Amessangeng pada penentuan status ekonomi masyarakat tersebut sehingga membantu penyaluran bantuan tepat .

### 2. Landasan Teori

Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (target)  $f$  yang memetakan tiap himpunan atribut  $x$

ke satu dari label kelas  $y$  yang didefinisikan sebelumnya. Fungsi target disebut juga model klasifikasi [1]

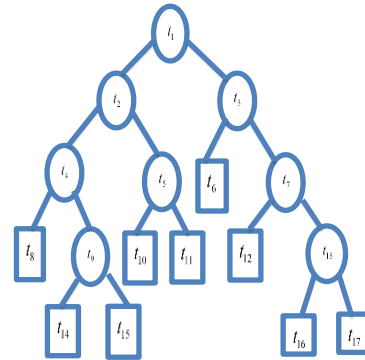


Gambar 1. Blok Diagram Model Klasifikasi

Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, di mana suatu model yang menerima masukan, kemudian melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. [2].

Kerangka kerja yang ditunjukkan meliputi dua langkah proses, yaitu induksi dan deduksi. Induksi merupakan langkah untuk membangun model klasifikasi dari data latih yang diberikan, disebut juga proses pelatihan, sedangkan deduksi merupakan langkah untuk menerapkan model pada data uji sehingga kelas yang sesungguhnya dari data uji yang disebut juga prediksi [2].

*Algoritma Classification And Regression Trees* merupakan salah satu metode atau algoritma dari teknik pohon keputusan (decision tree). Metode yang dikembangkan oleh Leo Breiman, dkk merupakan teknik klasifikasi dengan menggunakan algoritma penyekatan rekursif secara biner (). Ilustrasi dari struktur pohon klasifikasi ditunjukkan pada gambar di bawah ini [3].



Gambar 2. Struktur Pohon Klasifikasi

Algoritma CART melalui tiga tahapan antara lain

1. Penyusunan calon cabang (candidate split). Penyusunan ini dilakukan terhadap seluruh variable predictor secara lengkap (exhaustive). Daftar yang berisi calon cabang disebut daftar calon cabang mutakhir.
2. Penilaian kinerja keseluruhan calon cabang yang ada pada daftar calon cabang mutakhir. Kinerja dari setiap calon cabang akan diukur melalui ukuran yang disebut kesesuaian (goodness). Kesesuaian dari calon cabang  $s$  pada noktah keputusan  $t$  dilambangkan dengan:

$$\phi(s|t) = 2P_L P_R \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|$$

Bila didefinisikan besarnya :

$$Q(s|t) = \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|$$

Maka selanjutnya ditemukan persamaan:

$$\phi(s|t) = 2P_L P_R Q(s|t)$$

Menentukan calon cabang yang akan dijadikan cabang. Hal ini dilakukan dengan memilih nilai kesesuaian  $\phi(s|t)$  terbesar. Setelah itu gambarkanlah percabangan. Jika tidak ada lagi noktah maka pelaksanaan Algoritma CART akan diberhentikan. Namun, jika masih terdapat noktah keputusan, pelaksanaan algoritma tetap dilanjutkan ke langkah kedua dengan terlebih dahulu membuang calon cabang yang telah berhasil menjadi cabang sehingga daftar

calon cabang mutakhir yang baru (Susanto.,Suryadi. 2009)

**Metode Penelitian**

Metode penelitian yang digunakan analisis kuantitatif dengan tahap penelitian identifikasi masalah, pengambilan sampel, pembentukan algoritma CART dari data /sampel dan penyusunan laporan.

Prosedur penelitian pada Klasifikasi Algoritma and Classification Regression Trees dilakukan melalui langkah-langkah sebagai berikut:

1. Menyusun calon cabang (candidate split). Penyusunan ini dilakukan terhadap seluruh *variable predictor* secara lengkap (exhaustive).
2. Menentukan calon cabang kiri dan calon cabang kanan terhadap candidate split yang telah ditentukan melalui tabel calon cabang mutakhir.

3. Menentukan nilai melalui persamaan

$$P_L = \frac{\text{jumlah catatan pada calon cabang kiri } t_L}{\text{jumlah catatan pada data latihan}}$$

4. Menentukan nilai melalui persamaan

$$P_R = \frac{\text{jumlah catatan pada calon cabang kanan } t_R}{\text{jumlah catatan pada data latihan}}$$

5. Menentukan nilai  $P(j|t_L)$  melalui persamaan berdasarkan variable predictor yang memiliki Aset tinggi, sedang, rendah dan tidak ada.

$$P(j|t_L) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kiri } t_L}{\text{jumlah catatan pada data latihan}}$$

6. Menentukan nilai  $P(j|t_R)$  melalui persamaan berdasarkan variable predictor yang memiliki Aset tinggi, sedang, rendah dan tidak ada.

$$P(j|t_R) = \frac{\text{jumlah catatan berkategori } j \text{ pada calon cabang kanan } t_R}{\text{jumlah catatan pada data latihan}}$$

7. Menentukan nilai  $Q(s|t)$

$$Q(s|t) = \sum_{j=1}^{\text{jumlah kategori}} |P(j|t_L) - P(j|t_R)|$$

8. Menentukan nilai kesesuaian (goodness) melalui persamaan

$$\phi(s|t) = 2P_L P_R Q(s|t)$$

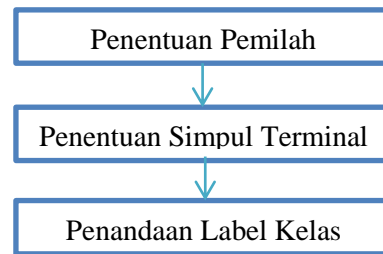
9. Membandingkan nilai goodness (kesesuaian) pada setiap calon cabang.

10. Menentukan dan memilih nilai kesesuaian (goodness) tertinggi kemudian menetapkan sebagai noktah dasar pada decision tree.

11. Membentuk calon cabang baru dengan tidak mengikutkan calon cabang yang telah dipilih.

12. Kembali menghitung nilai variable – variabel pembentuk nilai keputusan melalui beberapa iterasi hingga membentuk noktah terminasi sehingga tidak membentuk noktah keputusan.

Bagan pembentukan *decision tree* pada CART



Gambar 3. Bagan alur penelitian

**3. Hasil dan Pembahasan**

**3.1 Menentukan Calon Cabang.**

Penyusunan calon cabang dilakukan terhadap seluruh variable predictor secara lengkap (exhaustive). Pada kasus ini yang menjadi variable predictor seluruhnya adalah pekerjaan, status tempat tinggal, pendapatan dan kepemilikan harta benda. Pembuatan calon cabang mengikuti kaidah

Algoritma CART, yaitu adanya niktah keputusan yang selalu bercabang biner atau bercabang dua. Calon cabang untuk variable predictor pekerjaan adalah sebagai berikut:

1. Pekerjaan= Wiraswasta, dan Pekerjaan = (Pedagang, Pensiunan, Petani, Buruh, Karyawan)
2. Pekerjaan= Pedagang, dan Pekerjaan = (Pedagang, Wiraswasta, Pensiunan, Petani, Buruh, Karyawan)
3. Pekerjaan= Pensiunan, dan Pekerjaan = (Pedagang, Wiraswasta, Petani, Buruh, Karyawan)

4. Pekerjaan= Petani, dan Pekerjaan = (Pedagang, Pensiunan,Wiraswasta, Buruh, Karyawan)
5. Pekerjaan= Buruh, dan Pekerjaan = (Pedagang, Pensiunan, Petani,Wiraswasta, Karyawan)
6. Pekerjaan= Karyawan, dan Pekerjaan = (Pedagang, Pensiunan, Petani,Buruh, Wiraswasta).

Calon cabang untuk variable predictor status tempat tinggal adalah sebagai berikut:

1. Status tempat tinggal = Milik Sendiri, dan Status Rumah tinggal=(Numpang, Sewa)
2. Status tempat tinggal = Numpang, dan Status Rumah tinggal= (Milik Sendiri, Sewa)
3. Status tempat tinggal = Sewa, dan Status Rumah tinggal= (Numpang,Milik Sendiri).

Calon cabang untuk variable predictor Pendapatan adalah sebagai berikut:

1. Pendapatan = 1.500.000-, dan pendapatan > 1.500.000-
2. Pendapatan = 2.500.000-, dan pendapatan > 2.500.000-
3. Pendapatan = 3.500.000-, dan pendapatan > 3.500.000-

Calon cabang untuk variable predictor kepemilikan aset adalah sebagai berikut:

1. Aset = tinggi, dan aset = (sedang, rendah, tidak ada)
2. Aset = sedang, dan aset = (tinggi, rendah, tidak ada)
3. Aset = rendah, dan aset = (sedang, tinggi, tidak ada)
4. Aset = tidak ada, dan aset = (sedang, rendah, tinggi)

Berdasarkan aturan Algoritma CART yang setiap titik keputusan bercabang biner, maka calon cabang akan diberi nama calon cabang kiri dan calon cabang kanan. Selengkapnya calon cabang disajikan pada tabel 1.

**Tabel 1. Daftar Calon Cabang Mutakhir**

No	Calon Cabang Kiri	Calon Cabang Kanan
1	Pekerjaan= Wiraswasta	Pekerjaan = (Pedagang, Pensiunan, Petani,

	Buruh, Karyawan)
2	Pekerjaan = (Wiraswasta, Pensiunan, Petani, Buruh, Karyawan
	Pekerjaan = (Pedagang, Wiraswasta, Petani, Buruh, Karyawan)
	Pekerjaan = (Pedagang, Pensiunan, Wiraswasta, Buruh, Karyawan)
	Pekerjaan = (Pedagang, Pensiunan, Petani, Wiraswasta, Buruh, Karyawan)
	Pekerjaan = (Pedagang, Pensiunan, Petani, Wiraswasta, Buruh, Karyawan)
	Pekerjaan = (Pedagang, Pensiunan, Petani, Buruh, Wiraswasta)
	Pekerjaan = (Pedagang, Pensiunan, Petani, Buruh, Wiraswasta)
	Status Rumah tinggal= (Numpang, Sewa)
	Status Rumah tinggal= (Milik Sendiri, Sewa)
	Status Rumah tinggal= (Numpang, Sewa) Keputusan <=2.500.000
	Pendapatan > 1.500.000-,
	Pendapatan > 2.500.000-,
	Pendapatan > 3.500.000-,

### 3.2 Menilai Keseluruhan Calon Cabang

Langkah kedua adalah mneilai kinerja keseluruhan calon cabang keseluruhan pada daftar calon cabang mutakhir. Kinerja dari setiap calon cabang akan diukur melalui ukuran yang disebut kesesuaian (goodness). Hasil perhitungan nilai kesesuaian bagi tiap calon cabang masih terdapat dalam daftar calon cabang mutakhir.

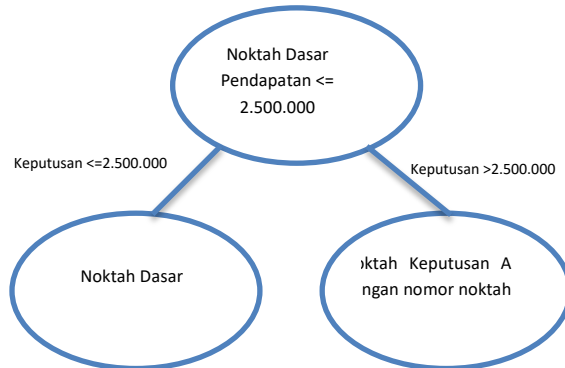
Tabel 2. Nilai Tiap Calon Cabang Mutakhir

No	$P_L$	$P_R$	$2P_L P_R$	$Q(s t)$	$\phi(s t)$
1	0.14814	0.851852	0.25240	0.70652	0.178326
2	0.16666	0.8333	0.27777	0.84444	0.23456
3	0.44444	0.5555	0.49382	0.216	0.1069
4	0.14814	0.8518	0.2524	0.7608	0.19204
5	0.05555	0.9444	0.10493	0.74509	0.07818
6	0.0370	0.962963	0.07133	1.76923	0.126200
7	0.6296	0.3703	0.4663923	0.16470	0.07681
8	0.18518	0.8148	0.301783	0.61818	0.186556
9	0.18518	0.8148	0.30178	0.42727	0.128943
10	0.51851	0.4814	0.499314	0.58791	0.29208
11	0.72222	0.2777	0.4012345	0.74871	0.300411
12	0.851852	0.1481	0.252400	0.96739	0.244170

### 3.3 Menentukan Calon Cabang Mutakhir

Tahap selanjutnya adalah menentukan calon cabang manakah yang akan dijadikan sebagai cabang muthakhir. Pemilihan cabang muthakhir ini ditentukan oleh nilai kesesuaian (goodness) terbesar yang dimiliki

oleh tiap calon cabang. Pada table 4 terdapat nilai kesesuaian semua calon cabang, namun yang memiliki kesesuaian terbesar adalah nilai kesesuaian (goodness) pada calon cabang No. 11 dengan nilai goodness sebesar 0.30. maka dari calon cabang no. 11 terpilih menjadi cabang muthakhir. Pemilihan cabang mutakhir digambarkan sebagai berikut:



Gambar 4. Pohon Keputusan Klasifikasi Data Sosial Ekonomi Masyarakat Kelurahan Amessangeng

Cabang kanan merupakan cabang kanan yang pendapatannya > 2.500.000,- yang memiliki asset rendah, sedang dan tinggi. Maka dari itu cabang kanan ini dijadikan sebagai noktah keputusan. Karena cabang kanan merupakan noktah keputusan pertama maka disebut noktah keputusan A. noktah keputusan akan berlanjut hingga mencapai noktah dasar.

### 4. Kesimpulan dan Saran

Klasifikasi data menggunakan algoritma klasifikasi And Regression Trees melalui beberapa tahap. Tahapan pertama hingga akhir bertujuan untuk mendapatkan nilai goodness (kesesuaian) terbesar hingga terpilih suatu noktah dasar pembentuk pohon keputusan. Iterasi yang terdapat pada tahap tersebut dilakukan secara berulang hingga membentuk noktah dasar dan menghasilkan semua noktah tereminasi hingga tidak membentuk cabang lagi.

**Daftar Pustaka**

- [1] Lewis, M.D .,Roger, J. (2000). An Introduction to Classification and Regression Trees (CART) Analysis. Annual Meeting of Society For Academic Emergency. California, UCLA Medical Center
- [2] Prasetyo, E. 2012. Data Mining Konsep dan Aplikasi menggunakan Matlab. Andi. Yogyakarta
- [3] Ritno, E.S.Y. 2019. Implementasi Algoritma CART pada keluarga pada Desadagang Kelambir Tg Morawa. Jurnal Riset Komputer . Vol.6 No.3 Juni 2019. ISSN 2407389.
- [4] Susanto, S., Suryadi, D.2013.Pengantar Data Mining.Andi. Yogyakarta.