

## Penerapan Algoritma C4.5 Dalam Prediksi Penyewa Sepeda Saifudin

Program Studi Manajemen Informatika, AMIK BSI Yogyakarta  
saifudin.sfn@bsi.ac.id

**Abstract** - At the time of machine-paced technology like this, a bike needs only at certain times each day do not need a bike, with this condition it appears the business opportunity to rent out bicycles. In business there are problems, namely a bicycle rental bicycle rental business still has not run optimally visible from shelter conditions containing bicycles are still not widely used by the community. The aim of this study was to test the accuracy of the algorithm C4.5 based methodology CRISP-DM. By using the C4.5 algorithm proved its accuracy in classifying the tenant bicycle-related predictions bike rental.

**Keywords:** Data Mining, C4.5 algorithm, rental, bicycles.

**Abstrak** - Pada saat teknologi mesin yang serba seperti ini, sepeda kebutuhan hanya pada waktu tertentu setiap hari tidak perlu sepeda, dengan kondisi ini muncul peluang bisnis untuk menyewakan sepeda. Dalam bisnis ada masalah, yaitu sewa sepeda bisnis sewa sepeda masih belum berjalan optimal terlihat dari kondisi tempat tinggal yang mengandung sepeda masih belum banyak digunakan oleh masyarakat. Tujuan dari penelitian ini adalah untuk menguji keakuratan metodologi algoritma C4.5 berdasarkan CRISP-DM. Dengan menggunakan algoritma C4.5 terbukti akurasi dalam mengklasifikasikan penyewa prediksi sepeda terkait penyewaan sepeda.

**Kata kunci:** Data Mining, algoritma C4.5, sewa, sepeda.

### A. PENDAHULUAN

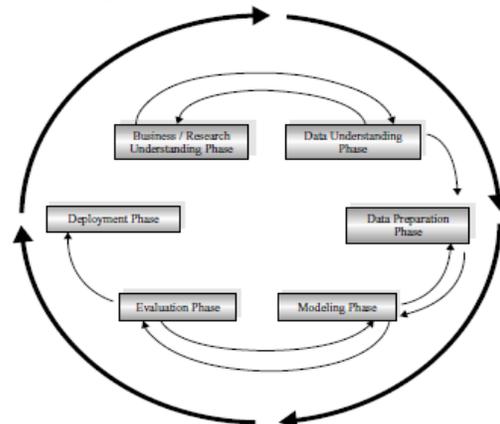
Pada masa teknologi serba mesin seperti sekarang ini, masyarakat atau orang-orang disekitar kita sudah dimudahkan dengan peralatan yang menggunakan mesin yang menggunakan bahan bakar minyak atau menggunakan daya listrik untuk menggerakkan peralatan yang kita gunakan sehari-hari. Dengan banyaknya kendaraan bermesin dengan menggunakan bahan bakar minyak dan menggunakan sumber daya listrik peralatan manual yang menggunakan tenaga manusia mulai ditinggalkan dan disebagian besar dirumah peralatan tersebut sudah tidak ada atau tidak dibeli lagi, peralatan tersebut adalah sepeda. Kebutuhan akan sepeda sekarang ini sudah tergantikan oleh kendaraan bermotor seperti mobil dan sepeda motor. Kebutuhan sepeda hanya pada saat-saat tertentu saja tidak setiap hari membutuhkan sepeda, dengan adanya kondisi seperti ini maka muncul peluang usaha untuk menyewakan sepeda.

### B. TINJAUAN PUSTAKA

Usaha penyewaan sepeda masih belum berjalan secara maksimal terlihat dari kondisi *shelter* yang berisi sepeda masih belum banyak dimanfaatkan oleh masyarakat disebabkan karena kurangnya promosi yang dapat meningkatkan jumlah pengguna (Badriyah, 2014).

*Data Mining* merupakan sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang terdiri dari keseluruhan proses, preprosesing data,

pembentukan model, model evaluasi, dan tahap akhir penyebaran model (Larose, 2005).



Gambar1. *Data Mining* sebagai proses CRISP-DM

Untuk mengembangkan aplikasi berdasarkan model yang dibuat, digunakan *Rapid Miner*. Tahap berikutnya yaitu implementasi, pada tahap ini objek implementasi dilakukan pengujian dari data yang didapatkan, teknik *sampling* menggunakan *purposive sampling*, dan desain ekperimennya digunakan proses *Cross Industry Standard Process for Data Mining* (CRISP-DM). Untuk mengukur tingkat keakurasian menggunakan *Confusion Matrix* dan kurva *Receiver Operating Characteristic* (ROC), setelah dilakukan pengukuran akan dihasilkan algoritma yang terbukti akurat dalam hal klasifikasi, rule dari algoritma yang terbukti akurat.

### C. METODE PENELITIAN

Jenis penelitian yang dilakukan merupakan penelitian ekperimen. Dalam metode ini, peneliti

memanipulasi paling sedikit satu variabel, mengontrol variabel lain yang relevan, dan mengobservasi pengaruhnya terhadap variabel terikat. Manipulasi variabel bebas inilah yang merupakan salah satu karakteristik yang membedakan penelitian eksperimental dari penelitian-penelitian lain.

Tabel 1. Atribut dan Nilai Kategori

No	Atribut	Nilai	Kategori
1.	Tanggal	date	
2.	Musim		
		Semi	
		Dingin	
		Panas	
		Gugur	
3.	Bulan		
		Januari	
		Februari	
		Maret	
		April	
		Mei	
		Juni	
		Juli	
		Agustus	
		September	
		Oktober	
		November	
		Desember	
4.	Hari		
		Senin	
		Selasa	
		Rabu	
		Kamis	
		Jumat	
		Sabtu	
		Minggu	
5.	Hari Kerja		
		Kerja	
		Libur	
6.	Cuaca		
		Cerah	
		Berawan	
		Gerimis	
7.	Suhu Normal		
8.	Suhu terasa		
9.	Kelembaban		
10.	Kecepatan Angin		
11.	Kelas		
		Sewa	
		tidak	

Mengingat jumlah populasi besar, hingga tidak mungkin dilakukan pemilihan sampel dengan carapengundian. Dalam keadaan populasi yang demikian, digunakanlah teknik *sampling* random sistematis (*Systematic Random Sampling*). Persyaratan yang harus dipenuhi agar teknik *sampling* ini dapat digunakan, yakni tersedianya kerangka *sampling* (ukuran populasinya diketahui dengan pasti), dan populasinya mempunyai pola

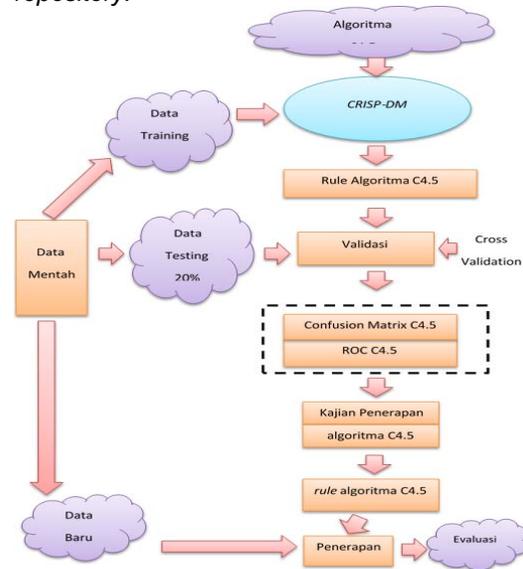
beraturan yang memungkinkan untuk diberikan nomor urut serta bersifat homogen. Interval sampel atau juga disebut *sampling rasio* diperoleh dengan cara membagi ukuran populasi dengan ukuran sampel yang dikehendaki (N/n). Misalnya, dari populasi (N) berukuran 731, akan diambil sampel (n) berukuran 146, maka interval *sampling*nya adalah  $731/146 = 5$  atau  $= 5$ . Maka unsur *sampling* pertama adalah satu anelementer yang bernomors.

Dalam penelitian ini data diperoleh dengan cara mendokumentasikan data yang berada di UCI *Machine Learning Repository*, dalam penelitian ini data dibedakan menjadi dua yaitu:

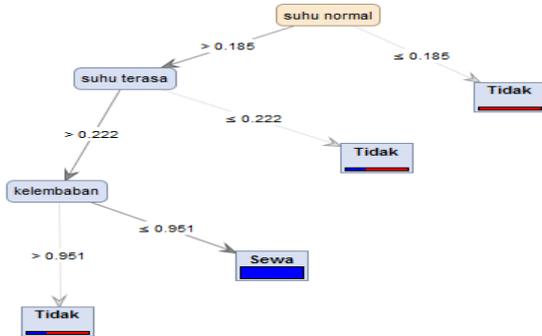
1. Data sekunder merupakan data yang diambil dari *data Laboratory of Artificial Intelligence and Decision Support (LIAAD)* di Portugal.
2. Data primer didapatkan dari hasil komputasi dengan menggunakan algoritma C4.5.

#### D. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini menguji keakuratan prediksi penyewaan sepeda dengan menggunakan algoritma C4.5. Data yang dianalisa adalah *Laboratory of Artificial Intelligence and Decision Support (LIAAD)* di Portugal yang diambil dari *UCI machine learning repository*.



Gambar2. Kerangka Pendekatan Penelitian



Gambar3. Tree hasil perhitungan dengan algoritma C4.5

Data training digunakan untuk menentukan apakah seorang pelanggan akan menyewa sepeda “sewa” atau “tidak”. Berikut akan dibahas prediksi apakah pelanggan akan “sewa” atau “tidak”, menggunakan metode klasifikasi dengan menghitung nilai Entropy dan nilai Gain. Dari hasil perhitungan *entropy* dan *gain* yang didapat pada Tabel 4.1, terlihat bahwa atribut *duration* mempunyai nilai gain tertinggi yaitu 0,28784. Oleh karena itu *duration* merupakan simpul akar pada pohon keputusan.

Tabel2. Nilai Entropy dan Gain

Node		JML Kasus (S)	Sewa (S1)	Tidak (S2)	Entropy	Gain
1	Total	584	551	33	0,313	
	Musim					0,089643426
	Semi	144	115	29	0,725	
	Dingin	143	139	4	0,184	
	Panas	146	146	0	0,000	
	Gugur	151	151	0	0,000	
	Bulan					0,119880437
	Januari	48	32	16	0,918	
	Februari	47	35	12	0,820	
	Maret	49	48	1	0,144	
	April	48	48	0	0,000	
	Mei	49	49	0	0,000	
	Juni	48	48	0	0,000	
	Juli	50	50	0	0,000	
	Agustus	50	50	0	0,000	
	September	48	48	0	0,000	
	Oktober	50	47	3	0,327	
	November	48	47	1	0,146	
	Desember	49	49	0	0,000	
	Hari					0,066873134
	Senin	84	79	5	0,326	
	Selasa	83	76	7	0,417	
	Rabu	83	76	7	0,417	
	Kamis	83	77	6	0,374	
	Jumat	83	80	3	0,224	
	Sabtu	84	81	3	0,222	
	Minggu	84	82	2	0,162	
	Hari Kerja					0,006293895
	Kerja	398	370	28	0,367	
	Libur	186	181	5	0,179	
	Cuaca					0,00913417
	Berawan	196	181	15	0,390	
	Cerah	374	359	15	0,243	
	Gerimis	14	11	3	0,750	

Pohon keputusan pada gambar 3 untuk menginterpretasikan pohon keputusan menjadi aturan-aturan (*rule*), aturan-aturan (*rule*) ini untuk memudahkan dalam penerapannya, aturan-aturan (*rule*) tersebut :

1. R1: IF *suhu normal* <= 0.185 THEN *class*= tidak
2. R2: IF *suhu normal* >= 0.185 AND *Suhu terasa* <= 0.222 AND THEN *class*= tidak
3. R3: IF *suhu normal* >= 0.185 AND *Suhu terasa* <= 0.222 AND *kelembaban* <= 0.951 THEN *class*= sewa
4. R4: IF *suhu normal* >= 0.185 AND *Suhu terasa* <= 0.222 AND *kelembaban* > 0.951 THEN *class*= tidak

accuracy: 94.69%			
	true Sewa	true Tidak	class precision
pred. Sewa	537	17	96.93%
pred. Tidak	14	16	53.33%
class recall	97.46%	48.48%	

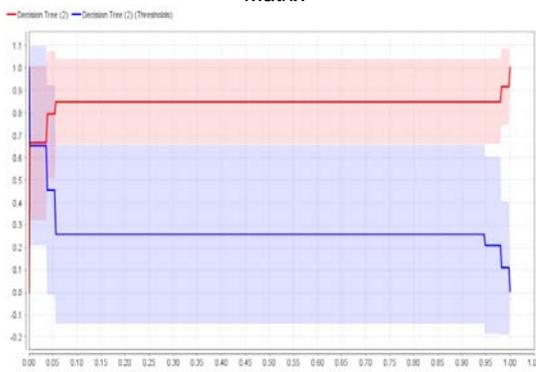
Gambar 4. Tingkat Akurasi

**PerformanceVector**

```

PerformanceVector:
accuracy: 94.69%
ConfusionMatrix:
True:   Sewa   Tidak
Sewa:  537    17
Tidak:  14    16
    
```

Gambar 5. Hasil Pengukuran dengan *Confusion Matrix*



Gambar 6. Kurva ROC

Dari hasil pengujian dievaluasi tingkat akurasi menggunakan dua model yaitu menggunakan *confusion matrix* dan *Receiver Operating Characteristic (ROC)*. Setelah diujicoba dengan metode *cross validation*, didapatkan hasil pengukuran terhadap data training yaitu *accuracy* = 94.69%, *precision* = 97.46% dan *recall* = 48.48%.

Dari hasil evaluasi baik secara *confusion matrix* maupun kurva ROC ternyata terbukti akurat. Dengan hasil ini, maka kemungkinan terjadinya salah sasaran dalam promosi kepelanggan dapat dikurangi dengan demikian algoritma C4.5 dapat memberikan solusi untuk permasalahan penentuan strategi promosi untuk penyewaan sepeda.

### E. KESIMPULAN

Dari hasil penelitian terbukti bahwa algoritma C4.5 akurat untuk menganalisa prediksi penyewaan sepeda yang dilakukan oleh pengusaha penyewaan sepeda. Hal ini dibuktikan dengan tingkat akurasi dari hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa pelanggan yang menyewa sepeda atau "sewa" dan yang pelanggan yang tidak menyewa sepeda atau "tidak" dengan tingkat keakuratan 94.69%.

Agar penelitian ini bisa ditingkatkan nilai akurasi dengan penerapan algoritma C4.5, berikut adalah saran-saran yang diusulkan:

1. Hasil penelitian ini diharapkan dapat diterapkan pada usaha-usaha penyewaan sepeda untuk meningkatkan akurasi dari promosi yang dilakukan kepada pelanggan penyewa sepeda.
2. Untuk mendukung pengambilan keputusan dan pengembangan sistem informasi manajemen, model ini dapat menerapkan sistem yang menggunakan perangkat lunak.
3. Disertai dengan pembuatan *Standard Operational Procedure (SOP)*, dilakukan sosialisasi bagi seluruh karyawan, pelatihan bagi *end-user*, dan berikutnya penerapan dilapangan supaya hasil yang didapatkan maksimal.
4. Penelitian ini diharapkan dapat dikembangkan lagi setiap dua tahun, dikarenakan perkembangan teknologi sekarang ini cepat sekali.

### DAFTAR PUSTAKA

- [1] Badriyah. S (2014). *Perancangan Aplikasi Mobile "Gowes" Sebagai Media Penunjang Program Bike Sharing Bandung*. Bandung: FIK, Telkom University.
- [2] <https://archive.ics.uci.edu/ml/machine-learning-databases/00275/Bike-Sharing-Dataset.zip>
- [3] Cheng, Xu, Pei dan Liu (2010), Hierarchical distributed data classification in wireless sensor networks: Canada 2010
- [4] Han dan Kamber (2007), *Data Mining Concepts and Techniques*, San Fransisco: Morgan Kauffman.
- [5] Konda, P. (2009) *The use of CRISP-DM methodology for data mining in banking*, Slovenia: Universitas Ljubljana.
- [6] Larose, D. T.(2005), *Discovering Knowledge In Data*, New Jersey: John Wiley & Sons, Inc.
- [7] Moro, S. and Laureano, RMS (2011) Using Data Mining For Bank Direct Marketing: An Application Of The CRISP-DM Methodology, Portugal: ESM
- [8] Rocha, B. C. and Junior, R. T. S (2010) Identifying Bank Frauds Using CRISP-DM And Decision Trees, Brasil: IJCSIT
- [9] Triwahyuniati, N (2008) Pelaksanaan Analisis Pemberian Kredit Di PT Bank Huga Cabang Semarang, Semarang: Universitas Diponegoro.
- [10] Wei dan You, (2011) *C4.5 Classifier for Solving the Problem of Water Resources Engineering*, Taiwan:Toko University.