Volume 7 No.1 Januari 2015 ISSN : 2085 – 1669 e-ISSN : 2460 – 0288 Website : jurnal.ftumj.ac.id/index.php/jurtek Email : jurnalteknologi@ftumj.ac.id



PENERAPAN DATA MINING DENGAN METODE KALSIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI

Popy Meilina^{1,*}

¹Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta Jl. Cempaka Putih Tengah 27 Jakarta 10510 *E-mail: popy.meilina@gmail.com

Diterima: 31 Agustus 2014

Direvisi: 23 September 2014

Disetujui: 11 Nopember 2014

ABSTRAK

Konversi Bahan Bakar Minyak (BBM) ke Bahan Bakar Gas (BBG) yang dilakukan pemerintah bermula atas melimpahnya gas di bumi Indonesia. kabupaten Cilacap melakukan survey terhadap masyarakat untuk menentukan masyarakat mana yang lebih diprioritaskan untuk mendapatkan bantuan pemberian kompor gas beserta tabungnya. Berdasarkan survey pada tahun 2010 di Kabupaten Cilacap Kecamatan Kroya terdapat 17 kelurahan dengan 28.131 data dengan 31 field atau kolom. Data diolah dengan beberapa tahapan yaitu : pengecekan data, integrasi data, target data, preprocessing, visualisasi, proses data mining, dan pengetahuan. Pada proses data mining dilakukan proses decision tree, regresi, dan k means untuk clustering yang ditampilkan dalam bentuk scatter plot. Berdasarkan decision tree maka diperoleh data kelurahan yang menjadi prioritas untuk diberi bantuan yaitu : Kelurahan Bajing Kulon, Kedawung, Pekuncen, dan Pesanggarahan. Atribut yang digunakan untuk proses regresi yaitu per kelurahan dengan variabel x yaitu pendidikan atau pekerjaan dan variabel y yaitu penghasilan.

Kata kunci: data mining, decision tree, regresi

ABSTRACK

A conversion from oil fuel to gas fuel that done by Indonesian Government are based on the overwhelming gas in Indonesia. Cilacap District has done a survey to people that live there in order to determine which people that would be consider to be prioritized to get subsidiary gas stove and gas tube. Based on survey in 2010 at Kroya an administrative area on Cilacap District, that have 17 villages with 28.131 data and 13 field aply. Data was processed with several stages, which is: data checking, data integration, data target, preprocessing, visualize and data mining process. In data mining process, there were comparison of three methods that are Decision Tree, Regression, and K-Means Clustering which visualize in a form of Scatter Plot. Results with decision tree method gain that subsidiary priorities are following villages; Bajing Kulon, Kedawung, Pekuncen and Pesanggrahan. While on Regression method using education and job status for x variable and income for y variable.

Keywords : data mining, decision tree, regression

PENDAHULUAN

Indonesia dengan kekayaan gas alam yang sangat besar, seharusnya pengadaan enegi gas dapat dimaksimalkan untuk memenuhi kebutuhan dalam negeri. Namun masyarakat masih sedikit yang memanfaatkan kekayaan ini. Masyarakat masih memandang Bahan Bakar Minyak (BBM) sebagai energi yang utama dalam kegiatan sehari-harinya.

Jumlah subsidi untuk BBM yang sangat besar, dan kondisi Indonesia yang sudah menjadi net importir minyak menjadi beberapa alasan bagi Indonesia untuk segera melakukan program konversi BBM ke Bahan Bakar Gas (BBG).

Diawali dengan keinginan pemerintah untuk mengkonversi Bahan Bakar Minyak (BBM) ke Bahan Bakar Gas (BBG), sehingga BUMN (Badan Usaha Milik Negara) khususnya di kabupaten Cilacap pada tahun 2009 berusaha untuk memaksimalkan keinginan pemerintah yang kemudian akan memberikan secara cuma-cuma kompor gas dan tabung gas untuk golongan yang tidak mampu, dan kemudian BUMN pun melakukan survey secara menyeluruh untuk mendapatkan data yang valid kelompok masyarakat mana diprioritaskan pertama vang untuk mendapatkan bantuan tersebut.

Berdasarkan data survev vang dilakukan **BUMN** Peneliti bermaksud menggunakannya untuk mendapatkan pengetahuan sehingga bisa dimanfaatkan oleh berbagai pihak untuk kepentingan tertentu khususnya dalam hal ini pemerintah terkait. Analisa yang dilakukan menggunakan Data Mining berdasarkan dari karakteristik data yang diperoleh seperti Pekerjaan, Pendidikan, Penghasilan dan atau Pengeluaran sehingga hasil yang didapat diharapkan lebih tepat sasaran.

Data Mining (Penambangan data) merupakan bidang ilmu multidisplin, menggambarkan area-area kerja yang termasuk didalamnya adalah teknologi basis data, pembelajaran mesin, statistik, pengenalan pola, pengambilan informasi, jaringan saraf tiruan, sistem berbasis pengetahuan, kecerdasan buatan. komputasi kinerja-tinggi, dan visualisasi data [Han & Kamber, 2006].

Ada banyak metodologi Data Mining, salah satu yang populer adalah Pohon Keputusan (*Decision Tree*). Pohon keputusan merupakan salah satu metode klasifikasi yang sangat menarik yang melibatkan konstruksi pohon keputusan yang terdiri dari node keputusan yang di hubungkan dengan cabangcabang dari simpul akar sampai ke node daun (akhir). Pada node keputusan attribut akan diuji, dan setiap hasil akan menghasilkan cabang. Setiap cabang akan diarahkan ke node lain atau ke node akhir untuk menghasilkan suatu keputusan (Larose, 2005).

Menurut Maimon (2005), pohon keputusan adalah salah satu metode klasifikasi yang dinyatakan sebagai partisi rekursif. Pohon keputusan terdiri dari node yang membentuk pohon yang berakar, semua node memiliki satu masukan. Node yang keluar disebut node tes. Node yang lain disebut node keputusan atau sering disebut node daun. Setiap simpul internal membagi dua atau lebih sub-ruang sesuai dengan kategori attribut dan akan dipartisi sesuai dengan nilai kategori kasus. Kasus-kasus tersebut membentuk pohon keputusan yang menghasilkan problem solving.

Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifiksi banyak digunakan untuk memprediksi kelas pada suatu label tertentu, vaitu dengan mengklasifikasi data (membangun model) berdasarkan training set nilai-nilai (label kelas) dalam dan atribut mengklasifikasikan tertentu dan menggunakannya dalam mengklasifikasikan data yang baru. Pohon keputusan biasanya digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Pohon keputusan dimulai dengan sebuah root node (titik awal) yang digunakan oleh user untuk mengambil tindakan. Berdasarkan node root ini, user memecahkan leaf node sesuai dengan algoritma decision tree. Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen) dengan satu atau lebih variabel bebas (independen). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda [Xin Yan dan Xiao Gang Su, 2009).

Berikut di bawah ini adalah tampilan data yang Peneliti gunakan.

No. No. <th></th> <th>1</th> <th>and and and and and and and and and and</th> <th>A.U.T</th> <th></th> <th>_</th> <th>-</th> <th>- 14-14</th> <th>-</th> <th>-</th> <th>and that I had</th> <th>0</th>		1	and	A.U.T		_	-	- 14-14	-	-	and that I had	0
B 1 1 0	And Annual States	0	-	8.0.7		_						
			-	A.U.T.	100000					1.64.11	m - +-	
Description Control	Allen (Monda) Male (Monda) Male (Monda) Male (Monda) Male (Monda) Male (Monda)		The	and the sector		A R A R A R A		- 6.0			Second Marco	
Bitter film Control Description Description <thdescrip< th=""> Descrip Descri</thdescrip<>	Allen (Mapalia Males (Mapalia) Billes (Mapalia Males (Mapalia	100	where a		R 104 10.0	44.89		-	de martinia		ranger balant	-
Billard Films Open Adda Jonage Diffe	Millio Travalla 1998: Strackarts 1998: Strackarts	100		10.44	(Bolt based of	140-05	All ranks	1	4444	1.00	Diff. Stand in Standy	
Control Aligner Control Al	anto invers	-	854	100-00 M	CROID DALINING	10840	Anoner	-1-	Annale	inter 1	Diff. Training the Diff. Among States	
Instruction	With Imparts		and a second	internet bit	1.00 million dan series	145-01	(Artor)	1.1	Approximation of	1.4	printed block	
STATUTE Statute Annue 1 Burger Dire Dire <thdire< th=""> Dire</thdire<>		10.00	996	Version 21	united balances	1000	1 Auro	1	1,000	100	and furial failures	
District Vision Open Adda I allower District Vision Open Adda District Vision Open Adda District Vision Open Adda District Vision Open Adda District Vision Open District Vision Open District Vision Open District Vision Open District Vision Distrin Vision District Vision Distri	10100 0010413	10.000	Nise .	(accessed)	older beiming	104-0	Distant .	X	Annala	1.00	prohesi kiloo	
Information Information <thinformation< th=""> <thinformation< th=""></thinformation<></thinformation<>	series contents	oprilie-	i inter	Photos in the	part of the	129-0	direct .	- X.	Approxim.	100	Distance without	
Instruction	min Invala	-	104	100.000.00	+ (0.0) ()++ (u)(u)g	104040	1111-0-0-	1	10100	1.4	the local is then	
Interpretation Composition Composition <thcomposition< th=""></thcomposition<>	annu cinutti	10156-0	1494	10.00	CONTRACTOR INCOME.	1.0910	I I I I I I I I I I I I I I I I I I I	1	Anne	Arei	CONTRACT IN CASE	
Self-particular size Self-part	Aller Imulti	3146.	1144	Disting 10	Autolo, KY 6	10010	Carlight	1	44446	104	per la serie de	
Bit Park Jose Bit Apple Jose Bit Appl	anter consta	20160	404	100.00 10	umpletered	108.00	1740.564		44440	ing i	and the st is Dated	
Shiftyer 10.00 Tep Avenue 1 Balance Scott 2020 ER 0	William Distinguisti	Miller.	154	ALL MARKED	24.000(4)(0)		Canter .	1	Antes	1.00	District in Cares	
Distruction operating	MINE THREE	Million.	834	interne al	Southern Ballin		1000.040	1	Annie	144	DRIVER NOW	
Simple Vites res res <t< td=""><td>ation constraints</td><td>10.044</td><td>104</td><td></td><td>utility (antisety)</td><td>1049-01</td><td>- 64-000</td><td>1</td><td>Anna</td><td>100</td><td>the heat is here.</td><td></td></t<>	ation constraints	10.044	104		utility (antisety)	1049-01	- 64-000	1	Anna	100	the heat is here.	
Operator Loss (e.g., style) 1 Style(e) Description Description <thdescription< th=""> Description <thdescription< th=""> Description</thdescription<></thdescription<>	STREET, L'ADARD	10.64	938		LINES DO ANNUAL	1000	LEGH	1	deployee.	1.00	COLUMN TO INCOME	
Image: Name Construction Construction </td <td>Million College Ta</td> <td>10.00</td> <td>10.000</td> <td>Date of the local division of the local divi</td> <td>designs also in</td> <td></td> <td>- Sections</td> <td></td> <td>Avenue</td> <td>1.00</td> <td>(the local is close)</td> <td></td>	Million College Ta	10.00	10.000	Date of the local division of the local divi	designs also in		- Sections		Avenue	1.00	(the local is close)	
Different Value operating and a state i. The base OPE II and the base of a state ii. The base of a state State of the base of the b	attin consint	per très	1476.0	Concernant II.	fariate million	Laterat	(farmer	1	44444	6.40	DRAME TO DAY	
Sill berg in Anno 1 Name	alities contacts	10.00	The		cold of the little of	a dela rest	- kees layleast	1.1	4444	100	(10) Rowall & Linear	
Distance force one of the second of	at the particular	10.45		mum M. M.	Apres Bills		cheads.	1	Anne	inger .	DRIVING BURN	
Default (Name	00460 (004ar7)	10.000	1041	ALC: N. W. W.	off-the location of	10040	ther		Apenalar .	14	petroni kilog	
(1) And 2 hour pay Annual Y have (1) Annual (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	million mitropille	10.04	14/64		LONDED DOLARING	LITTLE.	And province	1	44446	114	Different Robbell	
Interview Tex Address I Desc Desc <thdesc< th=""> <thdesc< th=""> Desc <</thdesc<></thdesc<>	Hitter Tetradia	1046	1000	10.00 3.0-	OPERATING.	10645	ileast -	X	Annie	-	its, hereit 'n Haan	
Implementations two two <thtwo< th=""> two <thtwo< th=""> <</thtwo<></thtwo<>	tenso conserta	101104	19454	rector21.	LINE COLORING	1000	1284	1	Antes	100	ATC NEW YORK	
Operation District Status St	inter timala	101040	the state	100 MI	idit (beliding)	100.00) the latest	1	101468	-	the local latitude	
Off-back king type Advance 1 Owner Output Owner	ALTER VALUE	10.00	324	No. of Acres	20.000.0002		10th Mat	1	Antel	Aires	part what is place	
District Table Sey Anton I (See 1992 - 1992) Long and S Stor S (Start Table See Sey Anton I (See 1997) Start See Start Table See See See See See See See See See S	Mines Minesta	10110-0	104	inclusion in .	united balance	1000	Children		April 10	Ange .	251 for at labour	
Differentiation and Annual I classifier and Annual	mile invate	100.044	10.000	100.00	LOCAL DIAMONG	100.0	(basi	- k	1,12,000	Strail	the here's time.	
detroictions are seven i close the the the seven is even in	With Ellipsing	W/Re-	1034	Number of Control of C	4/1/		Territeter	1	dantes.	100	Diff Databal de Garage	
	MINU XIMINT	MINN.	19.84	SHOW M.	LOP-PLICE ADDRESS	1000	Liberario	1	A114/0	100	Distance In Column	
C. Personal & Later Annual Annual C. Contra State Contra	and control ()		1400	- M.M. N	cannot the internet	1 de la	Lines.	- X.	Antes	100	part of the lines.	
	and the second s	-										

Gambar 1. Tampilan sumber data (data mentah) hasil survey Pihak BUMN pada Kabupaten Cilacap Kecamatan Kroya

Seperti terlihat pada gambar 1 terdapat beberapa field yang dijadikan pedoman bagi pihak BUMN yaitu: Kategori, kabupaten, kecamatan, kelurahan, dusun, rt, rw, nama, no id, no ktp, no kk, alamat, no sk, jenis kelamin, pendidikan, pekerjaan, pengeluaran, jumlah keluarga, masak, liter, dan lain-lain. Data yang ada yaitu 28.131 survey data dengan 31 field atau kolom.

Dibawah ini adalah field-field yang diambil dari data kependudukan Kecamatan Kroya yang ada pada data MS. Excell.

A 4 *1 4	
Atribut	Keterangan
Kategori	Didalamnya mencakup data yang terdiri dari : Rumah
	Tangga tetap, Usaha Kecil Menengah (UKM), Rumah
	Tangga Musiman
Kelurahan	Mencakup 17 macam Kelurahan yang ada pada
	Kecamatan Kroya
Jenis Kelamin	Merupakan jenis kelamin penduduk yang bersangkutan
Pendidikan	Merupakan jenjang pendidikan akhir yang diselesaikan
	oleh penduduk yang bersangkutan
Pekerjaan	Jenis pekerjaan yang dilakoni oleh setiap penduduknya
Pengeluaran	Seberapa banyak pengeluaran yang dikeluarkan oleh
	setiap keluarga perbulannya
Jumlah keluarga	Banyaknya jumlah anggota setiap keluarga
Masak	Dengan menggunakan apa setiap penduduknya memasak
Liter	Jumlah liter yang dihabiskan oleh penduduknya jika
	memasak dengan menggunakan kompor minyak
Rupiah	Jumlah pengeluaran untuk minyak tanah yang harus
_	dibeli oleh penduduk yang memasak dengan
	menggunakan kompor minyak.

n . . .

METODE

Sumber data untuk penelitian merupakan data hasil survey yang dilakukan oleh BUMN Kabupaten Cilacap Kecamatan Kroya pada tahun 2009. Data yang digunakan dalam penelitian ini sebanyak 28.131 data pada Kecamatan Kroya yang terdiri dari 17 kelurahan. Langkah-langkah yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Pengecekan Data (Proses Pembersihan Data)

Dalam Penelitian kali ini dicari hubungan beberapa atribut dari data yang ada. Karena tidak semua tabel yang digunakan maka perlu dilakukan pembersihan data yang akan diolah benar-benar relevan dengan yang dibutuhkan. mengapa data perlu di bersihkan sebelum diproses ? Hal ini terjadi karena biasanya data yang akan digunakan belum baik, penyebabnya antara lain :

- Incomplete : kekurangan nilai-nilai pada atribut atau atribut tertentu lainnya.

- Noisy : berisi kesalahan atau nilai-nilai outlier yang menyimpang yang tidak sesuai dengan data yang lainnya
- Inconsisten : ketidakcocokan dalam penggunaan kode atau nama. Disini kualitas data yang baik didasarkan oleh keputusan yang baik dan data warehouse memerlukan integrasi kualitas data yang konsisten.

Pada tahap ini proses Pembersihan data menggunakan dua tools yaitu, Microsoft Excell dan Orange.

2. Proses Pengintegrasian Data

Integrasi data merupakan penggabungan data dari berbagai database ke dalamsatu database baru. Disini proses mengubah data atau konversi data dalam numerik. bentuk Sama seperti tahap sebelumnya, pada tahap ini proses pengintegrasian data menggunakan dua tools yang saling mendukung satu sama lain, yaitu Microsoft Excell dan Orange.

3. Target Data

Tahap ini merupakan penjelasan mengenai data yang siap digunakan dalam proses data mining, dimana target data menjadi input untuk proses-proses selanjutnya.

4. Preprocess

Sebelum diproses data mining sering kali diperlukan *preprocessing*. *Datapreprocessing* menerangkan tipe-tipe proses vang melaksanakan data mentah untuk mempersiapkan proses prosedur yang lainnya. Tujuannya preprocessing dalam data mining adalah mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif untuk kebutuhan pemakai, dengan indikator sebagai berikut :

- Mendapatkan hasil yang lebih akurat
- Pengurangan waktu komputasi untuk *large scale problem*
- Membuat nilai data menjadi lebih kecil tanpa merubah informasi yang dikandungnya.
- Terdapat beberapa metode yang digunakan untuk *preprocessing* seperti :
 - a. *Sampling*, menyeleksi subset representatif dari populasi data yang besar.
 - b. *Diskretisasi*, Bagian dari data reduksi tetapi memiliki arti penting tersendiri, terutama untuk data numerik.
 - c. Remove missing
 - d. Continuize
 - e. Impute
 - f. Feature selection
- 5. Visualisasi

Tahap selanjutnya adalah visualisasi data, yaitu melihat data dengan menggunakan penjabaran secara grafis seperti Histogram, distribution diagram, diagram titik atau penjabaran secara lokasi seperti mean, median, modus, kuartil dan persentil.

6. Proses Data Mining

Tahap ini merupakan tahap pembentukan *decision tree*, regresi, dan klasifikasi K-means. Proses yang dilakukan pada *decision tree* menggunakan *software orange* dan proses untuk mencari regresi serta K-Means menggunakan *software* matlab.

7. Pengetahuan

Diharapkan setelah melalui setiap tahapan maka akan dihasilkan sebuah pengetahuan yang dapat dimanfaatkan untuk kepentingan pemerintah dan masyarakat.



Gambar 2. Metodologi Penelitian

HASIL DAN PEMBAHASAN

Penelitian yang dilakukan melalui beberapa proses yaitu :

• PreProcessing

PreProcessing yang dilakukan yaitu proses pembersihan data mengalami tiga tahap pembersihan yaitu: *Incomplete, Noisy* dan *Inconsisten.* Berikut dibawah ini penjelasan dan prosesnya.

1. Incomplete

Pada tahap ini, penulis membersihkan data berdasarkan data yang tidak lengkap atau data yang tidak terisi. Kesimpulan dari tahap ini adalah pembersihan data, dalam artian Poppy Meilina : Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Decision Tree Dan Regresi Jurnal Teknologi~7~(1) pp11-20 © 2015

bahwa jika *attribute* data tersebut kosong atau tidak terdapat nilai didalamnya maka *attribute* data tersebut akan dihapus. Hal tersebut di atas dikarenakan penulis tidak memiliki data pendukung untuk mengisi *attribute* data yang kosong.

2. Noisy dan Inkonsistensi Data

Merupakan suatu data yang memiliki kelainan, hal ini dikarenakan karena kesalahan operator dalam memasukkan data kedalam database, permasalahan pada pentransmisian data, keterbatasan teknologi, atau tidak dilakukannya penyeragaman data, seperti data wilayah, jenis kelamin, dan lain sebagainya.

Proses pembersihan data ini menggunakan suatu tools data mining yaitu Orange.

Pengubahan Data menjadi Numeric

Pada langkah selanjutnya data yang sudah dibersihkan akan diubah value nya menjadi bentuk numeric, hal ini dimaksudkan agar pengolahan data pada MatLab dapat lebih mudah dan cepat, karena MatLab hanya memproses data-data numeric.

> Kategori

Pada attribute kategori data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada attribute kategori.

Tabel 2 Pengintegrasian data Kategori

Data awal	Konversi data
	numeric
Rumah Tangga	1.00
Tetap	
UKM	2.00
Rumah Tangga	3.00
Musiman	

➢ Kabupaten

Sama seperti halnya pada langkah pertama maka data atau attribute kabupaten Cilacap akan diubah menjadi bentuk numeric. Seperti terlihat pada tabel dibawah.

Tabel 3. Transfe	ormasi data Kabupaten
Data awal	Konversi data
	numeric
Cilacap	1

➢ Kecamatan

Pengubahan pun berlanjut pada data atau attribute Kecamatan, hal yang dilakukan pun

sama seperti pada langkah-langkah sebelumnya.

Tabel 4. 7	Fransformasi data Kecamatan
Data awal	Konversi data numerik
Kroya	102

Disini penulis mengubah data kecamatan Kroya menjadi bentuk numeric "102" untuk memudahkan dalam pengenalan pada 17 kelurahan yang ada pada kecamatan tersebut. yang akan ditampilkan pada tabel dibawah ini.

Tabel 5. Transformasi data Kelurahan

Data awal Konversi	
Kecamatan	data numerik
Ayamalas	10201
Bajing	10202
Bajing Kulon	10203
Buntu	10204
Gentasari	10205
Karangmangu	10206
Karang Turi	10207
Kedawung	10208
Kroya	10209
Mergawati	10210
Mujur	10211
Mujur Lor	10212
Pekuncen	10213
Pesanggrahan	10214
Pucung Kidul	10215
Pucung Lor	10216
Sikampuh	10217

Pendidikan

Pada attribute Pendidikan data akan diubah menjadi numeric dari 1-5, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pendidikan.

Tabel 6. Transformasi data Pendidikan		
Data awal	Konversi data	
	numerik	
SD	1	
SLTP	2	
SLTA	3	
Universitas	4	
Lainnya	5	

Pekerjaan

Pada attribute Pekerjaan data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pekerjaan.

Tabel 7. Transformasi data Pekerjaan

ISSN: 2085-1669	9
e-ISSN : 2460 – 028	8

Data awal	Konversi data numeric
PETANI	1
SWASTA	2
Lainnya	3
PNS	4
TNI/POLRI	5

Pengeluaran

Pada attribute Pengeluaran data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pengeluaran.

Tabel 8. Transformasi	data Pengeluaran
-----------------------	------------------

Data awal	Konversi data
	numeric
500 Ribu	1
500 Ribu – 1	2
Juta	
1 Juta – 1,5	3
Juta	
>1,5 juta	4

> Masak

Pada attribute Masak data akan diubah menjadi numeric dari 1-3, berikut dibawah ini tabel pengintegrasian data pada *attribute* Pekerjaan.

Tabel	9.	Transformasi	data Masak	
-------	----	--------------	------------	--

Data awal	Konversi data					
	numerik					
Minyak Tanah	1					
Kayu Bakar	2					
Lainnya	3					

Transformasi Data

Tahapan Transformasi data adalah pengubahan format data tersimpan menjadi bentuk standar format file yang sesuai dengan aplikasi yang akan digunakan. Pada penelitian kali ini, penulis menggunakan aplikasi MatLab untuk melakukan proses data mining dengan Proses Aturan Prediksi menggunakan Regresi Linear, aturan Klasifikasi menggunakan Decision Tree. Pengubahan atau Transformasi data ke dalam format file yang sesuai adalah mengubah format file data sebelumnya yang merupakan bentuk file Tab Delimited yaitu Text atau .txt menjadi bentuk .dat.

• Proses Data Mining

Pada tahapan ini Peneliti mencoba membuat proses aturan klasifikasi pada data mining menggunakan decision *tree*. Algoritma klasifikasi akan menggunakan data tersebut untuk menghasilkan pengetahuan. Pengetahuan yang hendak dihasilkan dalam adalah klasifikasi untuk menggolongkan bagaimana ketentuan seorang warga mendapatkan suatu bantuan, baik itu bantuan secara materiil ataupun bantuan yang bersifat warga pelatihan agar tersebut dapat meningkatkan kesejahteraan hidupnya. Pada proses ini, Peneliti mengambil attribute kelurahan, pendidikan, pekerjaan dan pengeluaran untuk menentukan siapa warga vang memiliki ketentuan tersebut diatas berhak menerima bantuan. Berikut ini merupakan tahapan awal sampai akhir proses klasifikasi dengan decision tree dengan menggunakan Orange.

1. Input Data

Langkah awal proses ini adalah memasukan data yang telah dianalisa dan diproses dengan metode Cleansing (Pembersihan Data). Data tersebut merupakan data valid yang dibutuhkan untuk proses klasifikasi dengan menggunakan Orange. Berikut dibawah ini adalah langkahlangkahnya.

a. Pilih Widget File pada tabulasi Data di Orange, seperti terlihat pada gambar dibawah ini.



Gambar 3. Tampilan Widget File pada Orange

Setelah Widget File berada pada kanvas, maka selanjutnya.

b. klik dua kali pada widget File tersebut untuk melihat propertiesnya dan memilih file data yang terdapat pada direktori tersimpan, terlihat pada gambar dibawah ini. Poppy Meilina : Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Decision Tree Dan Regresi Jurnal Teknologi $~7~(1)~{\rm pp}~11-20$ © 2015



Gambar 4. Tampilan Properties Widget File pada Orange

2. Lihat Data

Untuk melihat data yang telah dipilih pada langkah pertama, dapat dilakukan dengan meletakkan widget Data Table di kanvas yang terdapat pada tabulasi yang sama dengan widget File yaitu tabulasi Data.



Gambar 5. Tampilan Widget Data Table pada Orange.

Setelah Widget Data Table diletakkan pada kanvas maka hubungkanlah dengan widget File, hal ini dilakukan agar data yang dipilih pada langkah sebelumnya dapat terlihat pada widget Data Table. Seperti gambar dibawah ini.

ALC: No. of Lot.						-	of the second dist					
# C	-	-	100									
Contraction of the local division of the loc			inget.	10.000	-	-		-	1494		. 10,000	
	10.	1		18	18.	108	A	1		1.1	4 100	Research 1
mus.	14	1		4	36	108	1	1		1	1.11	A
ACCURATION.	0.	1		1	38	100	¥	1		¥	1.000	ACCESS 10
and the set of the set of the	1.4	1		18	*	100		1	1	1	1	C
	14				*	108.1	+	h	1	1		CONTRACTOR OF
-	1.				16	118				1.	-	A.C
P CITCLE :	10.				jii	14	h .	1	1	1.	1	C
And The Contemport	1.	3		1	10	-	1	1.1	-)	1	1	B
100 (D.W.	1	1		1	10	18	1.	F	1	1		6000 E
and the first of	1.0	1		1		100	3.	1.5	1	4.5	4	C
	14				-	14	a		3	1.	1	A
	1	э.				100	1			1		Contraction of the local distance of the loc
and a state of the	1				18	108	1.	1		1	1	C
# (AL)	1	а,			16	108	3	1.	- A	1		6. C
	10.				10	148	1.	1	1	1.	1	A 1000
÷		2			14	100		1		1	4	1000 C
areas and	10	1		10	* :	100	1.	1	1	1	1	1000 E
Summer do Talay	1.	1				10.0	1	1	1	1	1	CONTRACTOR OF
	1.	5				108	4.1	1		1	1	ACCOUNTS 0
		1			10.	14	A	1.		1	1	C
	1.	З.		1	10.1	-	1	1.	1	1	1	1 mm
	1	1		1	10		*	1	1	1	1	C
	1.0	31				100	÷	£1.7	· · · ·	1.	1	
	1.	1				100	1		3	40.5	10	A
	10	1				100			1	1	1	6
	1.	51		1	18	108		1	15	1.	1	Contraction of Contra
	1.	1			16	108	1	1.	-	1	1	C
-	a.	1		-	14	100	1	1	1	1	1	A CONTRACTOR
	(hite	-									-	

Gambar 6. Tampilan Properties Widget Data Table pada Orange.

3. Proses Klasifikasi

Untuk melakukan Proses Klasifikasi pada data yang ada, dapat dipilih *widget Classification Tree* pada *Tabulasi Classify* dan letakkan pada kanvas *Orange* seperti terlihat pada gambar dibawah. Dan kemudian hubungkan dengan *Widget File*.



Gambar 7. Tampilan *Widget Classification Tree* pada *Orange*.

4. View Klasifikasi

Untuk melihat hasil yang didapatkan pada langkah sebelumnya, bisa dilakukan view dengan widget Classification Tree Viewer dan letakkan di kanvas, kemudian hubungkan dengan widget Classification Tree



Gambar 8. Tampilan Widget Classification Tree Viewer pada Orange.

Langkah selanjutnya adalah meng-klik dua kali pada *widget Classification Tree Viewer* tersebut untuk melihat output yang ada. output tersebut ditampilkan pada gambar dibawah ini.

and other thanks	Restricted Inc.	that Strike (fee Deleted (er-	
Contractor 1		A LOS ON DO DO DO	
	and the second sec	- The second second	
	A MARKED AND	a time last proteiner	275
10000-0000-000			6.1
Color State	a second data de	THE REPORT	145
test data	Manual Street	the late water	
and an other states of the second	A second citation	and the latest	
	A Design of the local division of the local	a test and advect	
in the second second	Sector Date:	1 10 10 10 10 10 10 10 10 10 10 10 10 10	
	distant in the second	100 100 State	
	a maker shared	time the effect.	
	10.00 - 20.00	10 10 m	
-	a province of the local division of	in the late laters.	
	Company of the local division of the local d	A DR DR DAWN	
	Design Will Dr.	1 HE IN TAILS	
-	· · · · · · · · · · · · · · · · · · ·		
	· March 1000 000	1 100 FR 101 F	
	* Max mo. (1983)	in the lac main	
	1 March 1 (1974)	1 (08. 10) M(10)	
	terms distant.	1 (08 10 hot)	
	- Manage with the	a series inc. The series	
	4.00.00-0278	i the set press.	
	11 Mill Allowed 1-270		
	200000 - 1001	1 104 10 HOLD	
	100001	1 100 17 Man	
	- statistics (1 100 17 BUD -	
	#*******	1. (B) (c) 100	
	- all, Mappi 178	- 100 . Doi: 100.000	
	ad (75.00-000)	1 THE . ML (2 HOL)	
	4.000000011	100 2 200	
	1.00000000000	10 F 10	
	politic 11	 (a) (b) (c) (b) (c) (c) 	
	114.44	1 108 10. MAR	
	- BRIDE-198-88	1. 100 AL (March 1)	
	101000-09	104 TO 100	
	ALC: 10.000 (MILL)		
	and a second sec	1 102 - 10 - Mile -	
	produced and	100 D 100	
	- NUMBER OF STREET	1 M. M. M. M. L.	

Gambar 9. Tampilan *Properties Widget Classification Tree Viewer* pada *Orange*.

5. View Pohon Klasifikasi / Decision Tree Berikut ini adalah tampilan hasil yang berbentuk suatu pohon klasifikasi dengan menggunakan widget classification tree graph yang dihubungkan dengan widget classification tree yang ada di kanvas, seperti terlihat di bawah ini.



Gambar 10. Tampilan *Properties Widget Classification Tree Viewer* pada *Orange*.

Output dengan menggunakan Classification *Tree* graph dapat terlihat pada gambar dibawah ini.

• Proses Aplikasi

Prediksi pada Klasifikasi dengan Rumus Regresi linear untuk menentukan mana kelurahan yang memiliki prioritas untuk diberikan bantuan dari 17 kelurahan yang ada di kecamatan kroya.

1. Membuka Aplikasi berbentuk *.fig pada yang ditunjukkan lingkaran merah pada gambar di bawah ini :



Gambar 11. Tampilan awal MatLab dengan lingkaran merah untuk membuka file Figure yang telah dibuat.

2. Kemudian akan muncul tabel dialog seperti di bawah, pilih *Open Exiting GUI*. Akan tampil daftar file *.fig yang telah kita buat. Pilih file *TAela.fig*.



Gambar 12. Tampilan dialogue box open file figure

3. kemudian jalankan aplikasi dengan mengklik icon Run Figure, seperti yang ditampilkan pada gambar di bawah yang ditunjukkan oleh likngkaran merah.

Poppy Meilina : Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Decision Tree Dan Regresi Jurnal Teknologi~7~(1) pp11-20 © 2015



Gambar 13. Tampilan Aplikasi data Mining dengan lingkaran merah untuk menjalankan Program

4. Akan muncul *table dialog* deperti gambar di bawah, kemudian klik **Lihat Data**

	PROSES RE	RESI DAN K-MEANS	
Name and Party	(m)		>
- Ferdilitat	Press for	-	
			Charlen ing A Alexand
			-
(manufic states)			. Marce

Gambar 14. Tampilan Aplikasi Data Mining untuk melihat data

5. Pilihlah file yang berbentuk *.dat yang ingin diproses. Dan klik *open*.



Gambar 15. Tampilan Dialogue Box Open File

6. Muncullah tabel dialog *Import Wizard* di bawah ini, klik *next*.

least	Column Sep	entropy -										
OB	anni Di	iere d'A	-	Tel ()		Number of text header lives						
-	e of Critices	Adarperi	Second State	ATLAN.	-tertle	iei -						
Butegoos .		habup	kabupatani		autamats #	data tantana						
2	1.6	1.17	102/01		1.075		1 1	1.1				
Ε.		1.14	58201	1.1		11	1		100	367 -		
E	1.1	104	10001	1		111	- 2		240			
Ε.	14	111	11211	10		1.5	- 1		346	- 14		
Ε.	12	114	11271	10		141	1	1	202	182		
Ε.	1.1	114	11211	-		1	1	1	104	262		
L	1.0	114	11211	14		81	1	1	192	262		
	1.1	1.16	11201	1		27	1	. 1	260	255		
2	1.1	111	11211	12			1		100	381		
E C	1	2.68	13231	1			1		346	352		
1	1.1	122	18291	2		38	1	1	346	382		
	- 12	128	11272	2		31	- A	8	182	362		
	- N -	1.14	1.1211	4		111	I		101	367 -		
4 (141		- M-			(2)		

Gambar 16. Tampilan *Dialogue Box Import Wizard*

7. Pilih file yang ingin di import, kemudian klik *finish*.



Gambar 17. Tampilan *Dialogue Box Import Wizard*

Untuk melakukan Proses Regresi terlebih 8. dahulu pilihlah variable x dan variable y yang ingin diproses, untuk gambar di bawah variable x yang dipilih adalah pendidikan dan variable adalah у pengeluaran, sehingga akan muncul persamaan garis linear seperti yang ditunjukkan oleh lingkaran merah di bawah.



Gambar 18. Tampilan Proses Regresi

9. Proses Clustering K-means yang dilakukan yaitu menentukan secara random letak tiga buah k dan menentukan anggota yang termasuk kedalamnya. Di bawah ini adalah tampilan scatter diagram untuk k-means.



Gambar 19. Tampilan Scatter Diagram

Penarikan Informasi

Berdasarkan hasil yang didapat pada 3 langkah data mining diatas, maka dapat ditarik beberapa kesimpulan atau informasi yang dapat dijadikan pengetahuan untuk diambil suatu keputusan, informasi tersebut dapat dijabarkan sebagai berikut.

1. KlasifikasidenganPohonKeputusan/Decision Tree.

Berdasarkan proses software Orange dengan menggunakan widget Classification *Tree* Viewer dan Classification *Tree* Graph bahwa keputusan yang dapat diambil oleh kepala kecamatan tentang kelurahan yang akan diberikan bantuan dengan skala prioritas adalah Kelurahan Bajing Kulon, Kedawung, Pekuncen, Pesanggarahan dan seterusnya.

2. Klasifikasi dengan Regresi Linear

Untuk mendapatkan informasi yang dibutuhkan dalam pengambilan keputusan dengan menggunakan Regresi Linear maka data yang ada yaitu data Pada Kecamatan Kroya dengan 17 kelurahannya harus di pecah menjadi data per kelurahan saja. Data per Kelurahan inilah nantinya akan dihitung dengan regresi linear pada Aplikasi Data Mining yang Peneliti rancang, yang kemudian hasil dari proses tersebut satu per satu Kelurahan yang ada di Kecamatan Kroya akan dibandingan secara keseluruhan. Hasil dari regresi linear ini adalah mencari nilai dari Variabel Y terkecil di kelurahan yang ada.

KESIMPULAN

Dapat disimpulkan bahwa dengan menggunakan *decision tree* prioritas kelurahan yang dapat diberi bantuan yaitu: Kelurahan Bajing Kulon, Kedawung, Pekuncen, dan Pesanggarahan.

Proses pada regresi berdasarkan per kelurahan dengan atribut untuk variabel x adalah pekerjaan atau pendidikan dan atribut untuk variabel y adalah penghasilan.

Terlepas dari penjelasan pada paragraph di atas, Peneliti sadari masih banyak kekurangan dalam penelitian ini, hal ini dikarenakan karena variable data yang terlalu luas dan proses pembersihan data yang memakan waktu lama, sehingga pemilihan proses data mining dengan penggunaan algoritma yang diterapkan mungkin memiliki kelemahan dalam keakuratan hasil data yang ada, tetapi hasil yang didapatkan tidak mengurangi keakuratan algoritma atau rumus yang diterapkan.

DAFTAR PUSTAKA

- Han, Jiawei and Kamber, Micheline (2006), Penambangan data: Concepts and Techniques. Elsevier
- Xin Yan, & Xiao Gang Su, (2009), *Linear Regression Analysis*. London: World Scientific Publishing Co. Pte. Ltd., Covent Garden
- Larose, & Daniel T. (2005). *Discovering knowledge in data: an introduction to data mining*, John Wiley and Sons, USA
- Lior Rokach, & Oded Maimon, (2005), *Data Mining With Decision Tree*. World Scientific Publishing Co. Pte. Ltd, Series in Machine Perception Artificial Intelligence Volume 69