

PERANCANGAN DATA WAREHOUSE DAN PENERAPAN ALGORITMA ID3 PADA ANALISA POTENSI DAN HAMBATAN PERKEMBANGAN DESA BERDASARKAN POLA HIDUP MASYARAKAT

Sehman¹

Teknik Informatika, Fakultas Teknik Universitas Maarif Hasyim Latif, Sidoarjo, Indonesia
Jl. Raya Ngelom Megare No.30, Ngelom, Taman, Kabupaten Sidoarjo, Jawa Timur
085265238725/ 61257
sehman@dosen.umaha.ac.id

ABSTRACT

The lack of understanding of the importance of data will result in a large impact loss because with the existence of an important sense of the data, an agency either private or public company will easily make decisions. A data warehouse is an option that must be considered for the village in processing and storing large classified data such as data from a village. In addition to the application of the research data warehouse, it also uses the ID3 algorithm which is useful as a theory to analyze the potentials and obstacles that affect the development of a village based on the data of the lifestyle of the local community. Data on population, education, health, conditions of infrastructure, village finance, village assets, community organizations, community economic activities and daily activities of the village community will be extracted and stored by building a data warehouse, data is sought from data warehouse which will be used as dataset pattern community life so that it is easily analyzed and can be used as training data to make decisions by applying the ID3 algorithm. The ID3 algorithm that has been applied to the community lifestyle dataset will produce a rule that has an influential potential class, and the potential to hamper village development. The accuracy of the system to be made has a hypothesis of 95%. because the rule generated by id3 will be compared with the rule collected based on respondents from the previous data, which is data on urban residents who have advanced..

Keywords: Data, Village, Id3, Pattern, Rule, Warehouse

1. PENDAHULUAN

Peraturan Undang-Undang 6 Tahun 2014 tentang Desa mengamanatkan pemerintah daerah wajib mengembangkan sistem informasi desa dan pembangunan kawasan perdesaan. Ditambah lagi dengan makin besarnya dana yang masuk ke desa, sekitar Rp100 triliun untuk > 74.000 desa tiap tahun. Dengan demikian setiap desa harus sudah memiliki tata kelola data dan informasi yang lebih baik dengan memanfaatkan keuntungan dari teknologi informasi, dalam rangka melakukan perencanaan, akuntabilitas, transparansi, dan pelayanan publik oleh pemerintah desa. Sistem pengelolaan data yang yang dapat dimanfaatkan adalah data warehouse.

Data warehouse adalah suatu paradigma baru dilingkungan pengambilan keputusan strategi. data warehouse bukan suatu produk tetapi suatu lingkungan dimana user dapat menemukan informasi strategi .

Selama ini data yang berada di desa hanyalah sebatas data administrasi seperti keuangan dan kepengurusan surat –surat, sedangkan data pendidikan, kesehatan, kondisi, sarana, prasarana dasar, kemiskinan pembangunan, perencanaan, aset desa, organisasi kemasyarakatan dan Kegiatan ekonomi masyarakat seperti petani, peternak, pedagang tidak begitu diperhatikan hal ini terlihat dengan tidak adanya data yang akurat yang tersimpan pada kantor kepala desa. Disamping permasalahan pengambilan keputusan yang diterapkan oleh

pemerintahan desa dilakukan hanya dengan rapat perangkat desa dengan tidak adanya sistem pendukung keputusan berdasarkan data yang akurat sehingga akan mempengaruhi kurang tepatnya keputusan tersebut.

Dengan penerapan Data Warehouse pada data penduduk Desa yang memiliki beraneka ragam dimensi tersebut akan terjaga dan tersimpan disuatu sistem aplikasi. Berdasarkan data warehouse akan ditemukan pula pola hidup masyarakat yang terlihat dari kegiatan sehari-hari dan dimana pola hidup masyarakat tersebut memiliki potensi yang mempengaruhi perkembangan desa. Perancangan data warehouse dan penerapan algoritma ID3 akan menggali data kemudian mencari pola yang akan dijadikan dataset agar dapat diolah sehingga menghasilkan berupa data potensi dan hambatan yang dapat mempengaruhi perkembangan desa berdasarkan cara atau pola hidup masyarakat setempat.

Hasil Penelitian Arief Dwi Hartadi; Oktalia Juwita, 2012. perancangan data warehouse dan penerapan teknik clustering spatial pada wesel: studi kasus pt xxx. Penelitian ini bertujuan untuk mengembangkan Data warehouse bagi PT XXX. Data warehouse ini bisa menampilkan informasi yang berguna bagi perusahaan, tanpa mengganggu sistem operasional. Data warehouse dikembangkan dengan menggunakan Oracle 11g. Selain itu, penelitian juga bertujuan untuk menggali informasi dan pengetahuan dengan basis data spasial. Penggalan data ini dilakukan menggunakan aplikasi WEKA dengan membandingkan

algoritma clustering DBSCAN, K-Means, dan EM. Penggalian informasi dari data spasial ini mampu menghasilkan peta wilayah yang menunjukkan tingkat penjualan produk-produk PT XXX, sehingga perusahaan dapat lebih fokus dalam memasarkan produk yang dipercaya lebih unggul di suatu wilayah tertentu. Hasil penelitian menunjukkan bahwa Data warehouse yang dikembangkan sudah berhasil menampilkan yang diperlukan oleh manajemen yaitu pada jumlah stok ritel, nilai penjualan ritel, dan nilai penerimaan wesel.

Hail Penelitian Halim Agung, 2017. Implementasi Star scheme dalam perangan aplikasi data warehouse administrasi desa wisata Kuningan. Pada penelitian ini didapatkan 46 tabel pada database dan 6 skema star untuk tabel fakta yaitu skema penduduk, skema pekerjaan, skema pendidikan, skema pengusaha, skema peternakan dan skema prekonomian yang dibangun dari beberapa tabel dimensi. Tujuan dari aplikasi ini adalah menganalisis dan membuat data warehouse aplikasi database untuk memudahkan pembuatan aplikasi administrasi didesa wisata cibuntu lebih teratur. Hasil aplikasi ini berguna untuk operasi dalam admininistrasi desa ,dewa desa untuk mengolah data yang ada dan mencetak laporan.

Jauhari Maulani, 2016 tentang “klasifikasi kesejahteraan rumah tangga dengan penerapan algoritma decision tree seleksi atribut genetik algoritma”. Penelitian ini menerapkan algoritma Decision Tree C4.5 dengan Seleksi Atribut GA (Genetik Algoritma) untuk klasifikasi status kesejahteraan rumah tangga. pada data kemiskinan dalam menentukan status kesejahteraan rumah tangga miskin kedalam 2 (dua) kategori RTS (Rumah Tangga Sasaran) yaitu RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin) dan RTM (Rumah Tangga Miskin). Setelah dilakukan pengujian yaitu algoritma C4.5 sendiri sudah memiliki akurasi yang cukup baik yaitu sebesar 82.08 % dengan proses seleksi atribut oleh algoritma genetika, model yang terbentuk dapat ditingkatkan lagi menjadi 82.98% dalam mengklasifikasikan status kesejahteraan rumah tangga miskin kedalam 2 (dua) kategori RTS (Rumah Tangga Sasaran) yaitu RTSM (Rumah Tangga Sangat Miskin/Fakir Miskin) dan RTM (Rumah Tangga Miskin).

Berdasarkan uraian latar belakang maka dapat dirumuskan masalah dalam penelitian ini adalah : 1) Bagaimana membuat data warehouse dengan data pendudukan, pendidikan, kesehatan, Kondisi sarana dan prasarana dasar, Kemiskinan dan kerentanan, Pembangunan Perencanaan, Keuangan desa, Aset desa, Organisasi kemasyarakatan dan Kegiatan ekonomi masyarakat? 2) Bagaimana dalam menentukan variabel pola hidup masyarakat dengan menyeleksi dari data warehouse dataset? 3) Bagaimana membangun sistem untuk mengidentifikasi pola prilaku masyarakat yang sangat berpengaruh terhadap perkembangan desa? 4) Bagaimana membangun sistem dengan merancang data warehouse dan menerapkan algoritma ID3 pada pola hidup masyrakat untuk menemukan rule yang paling berpengaruh terhadap perkembangan desa ? 5) Bagaimana mengolah dan menguji sistem yang telah dibuat sehingga menghasilkan akurasi yang baik?

Adapun tujuan dari penelitian ini adalah : 1) Sistem mampu menganalisa potensi dan hambatan perkembangan desa maka diharapkan desa gunung kembar dapat lebih berkembang.. 2). Menerapkan teknologi informasi pada desa gunung kembar yang hingga saat ini belum pernah disentuh oleh penelitian pada bidang terkait.

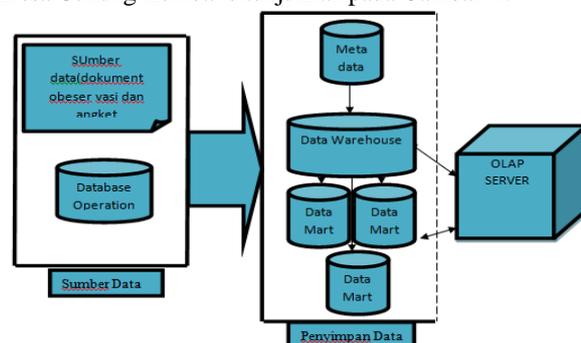
Sedangkan manfaat dalam penelitian ini adalah : 1) Hasil penelitian ini dapat mempermudah pemerintah desa untuk lebih mengalokasikan dana kepada masyarakat dikarenakan adanya sistem yang dapat memberikan gambaran pola masyarakat yang memiliki potensi dalam perkembangan desa. 2) Hasil penelitian ini dapat mempermudah pemerintah desa untuk lebih waspada dan meminimalisir pola hidup masyarakat dikarenakan adanya sistem yang dapat memberikan gambaran pola masyarakat yang memiliki hambatan dalam perkembangan desa. 3) Manfaat dari penelitian juga dapat membantu masyarakat desa gunung kembar untuk lebih memonitoring kebijakan yang diambil oleh pemerintah desa. 4) Mudahnya pemrintah dalam mengakses dan menganalisa data dengan adanya datawarehouse dan sistem yang terkomputerisasi.

2. METODE

Metode yang digunakan dalam penelitian ini adalah dengan mengumpulkan informasi aktual secara rinci yang melukiskan gejala yang ada, mengidentifikasi masalah atau memeriksa kondisi dan praktek-praktek yang berlaku, membuat perbandingan atau evaluasi dan menentukan apa yang dilakukan orang lain dalam menghadapi masalah yang sama dan belajar dari pengalaman mereka untuk menetapkan rencana dan keputusan pada waktu yang akan datang. Peneliti juga melukiskan secara sistematis fakta atau karakteristik populasi tertentu atau bidang tertentu yang menyangkut pada obyek dan subyek penelitian, dalam hal ini bidang secara aktual dan cermat. Metode ini menitik beratkan pada observasi dan suasana alamiah.

1) Pembuatan Data Warehouse dan Penerapan ID3

Adapun tahapan dalam pembuatan data Warehouse Desa Gunung Kembar ditunjukkan pada Gambar 1.



Gambar 1. Tahapan pembuatan data warehouse

Design data warehouse di bagi menjadi 2 bagian yaitu perancangan logika dan perancangan fisik.

a. Perancangan Logika

1. Mengumpulkan data pendudukan, pendidikan, kesehatan, Kondisi sarana dan prasarana dasar, Kemiskinan dan kerentanan, Pembangunan Perencanaan, Keuangan desa, Aset desa, Organisasi

kemasyarakatan dan Kegiatan ekonomi masyarakat, dan pola hidup masyarakat. Sumber data bersal dari banyak sumber seperti dari hasil observasi, angket, dokumen, dan file yang tersimpan pada komputer kantor desa.

2. Mendefinisikan masing-masing attribute
3. Membuat relasi

b. Perancangan Fisik

1. Merubah dari logical design menjadi objek-objek database
2. Tablespace, Table, integrity constraint dan aturan yang dianggap perlu.

2) **Extraction, Transformation and Loading (ETL)**

Adalah proses pengekstrakan data dari sumber data yang kemudian dimasukkan ke dalam data warehouse dan dilakukan secara periodik untuk kebutuhan bisnis dengan analisa data yang akurat. Tahapan ETL pengekstrakan data dari sumber data yang kemudian dimasukkan ke dalam data warehouse. Menggunakan Data Transformation Service (DTS) dari MS SQL Server.

3) **Creat Dimensi**

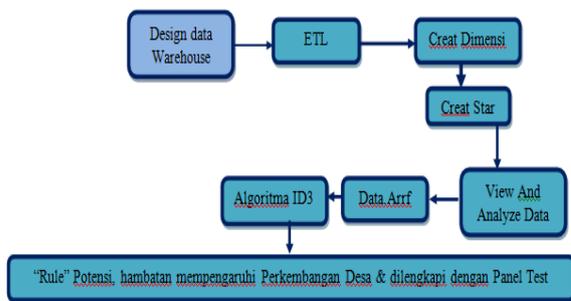
Dimension adalah sebuah struktur yang terbentuk dari satu atau lebih hirarki yang mengkategorisasi data. Dimensi terbentuk dari satu atau lebih tabel. Setiap kolomnya merepresentasikan level pada hierarchy. Adapun alur pembuatan dimensi adalah menentukan nama dimensi, memilih skema, memilih tipe dimensi, memilih tabel, memilih level, menentukan hirarki dan menentikan relasi.

4) **Membuat Cube.**

Adalah representasi kumpulan data dalam multi-dimensi dalam penelitian ini menggunakan skema star. View and Analyze DATA

Data sudah siap berada pada data warehouse dan siap untuk dijadikan data dalam pengambilan keputusan.

Berdasarkan dari pemaparan diatas maka dibuat diagram pembangunan data warehouse dan penerapan algoritma id3 seperti yagn ditunjukkan pada gambar 2.

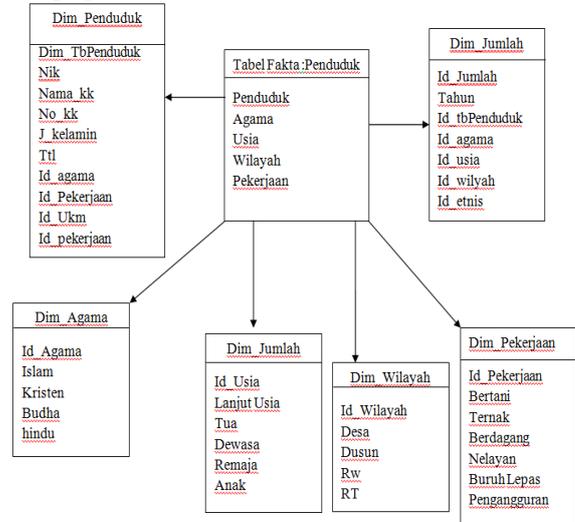


Gambar 2. Pembangunan data warehouse dan id3

Gambar 2 terlihat bahwa setelah data warehouse selesai dibangun maka akan diterapkan algoritma ID3 sebagai analisa data untuk pengambilan keputusan dengan cara merubah ekstensi data menjadi .arff. Data arff ini berisi data pola kehidupan masyarakat yang telah dipilih pada data warehouse kemudian data tersebut akan dijadikan sebagai

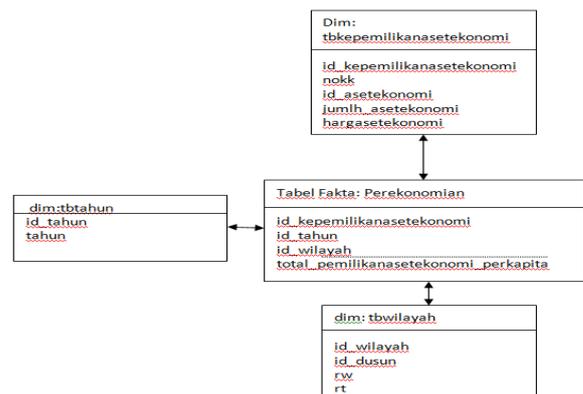
data training untuk kemudian dihitung dengan algoritma ID3 sehingga menemukan rule baru yaitu potensi dan hambatan yang mempengaruhi perkembangan desa.

Setelah data fakta ditetapkan maka akan dicari data dimensi dan akan diimplementasikan dengan menggunakan Star Schema (skema bintang) seperti yang ditunjukkan pada beberapa gambar dibawah ini.



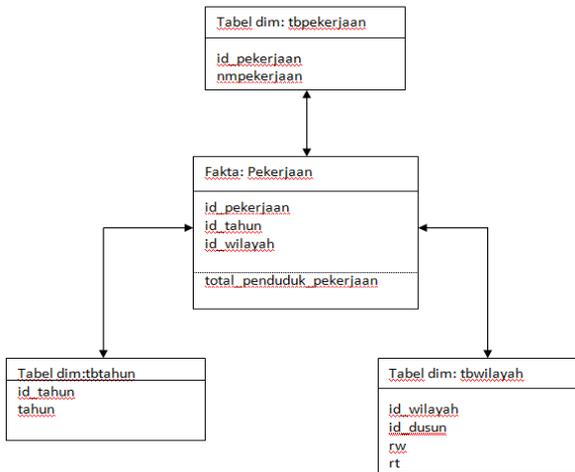
Gambar 3. Penduduk

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 3 merupakan gambaran dari kejadian nyata tentang penduduk desa gunung kembar seperti kepemilikan nik, nama,wilayah, jenis kelamin, tempat tanggal lahir, agama, pekerjaan dan usia.



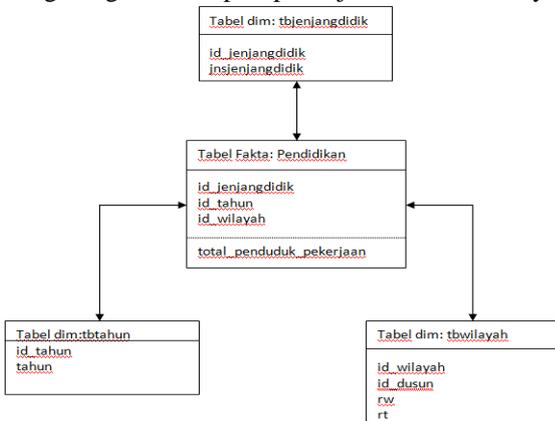
Gambar 4. perekonomian

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 4 merupakan gambaran dari kejadian nyata tentang penduduk desa gunung kembar seperti kepemilikan aset ekonomi, tahun dan wilayah.



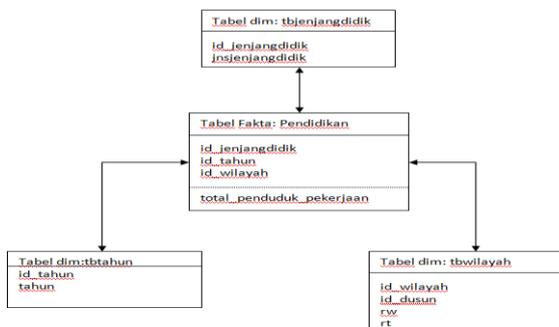
Gambar 5. Kegiatan

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 5 merupakan gambaran dari kejadian nyata tentang pekerjaan desa gunung kembar seperti pekerjaan, tahun dan wilayah.



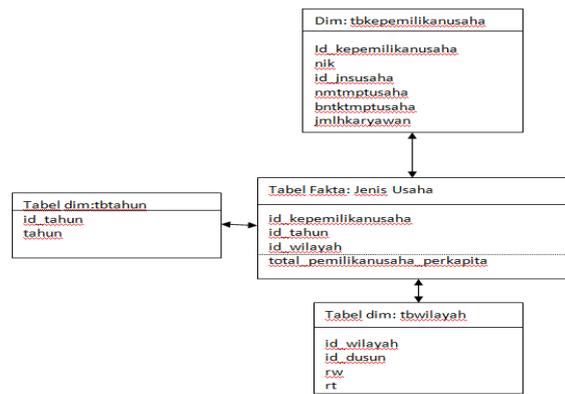
Gambar 6. Pekerjaan

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 6 merupakan gambaran dari kejadian nyata tentang pekerjaan desa gunung kembar seperti pekerjaan, tahun dan wilayah.



Gambar 7. Tabel Fakta Pendidikan

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 7 merupakan gambaran dari kejadian nyata tentang pendidikan didesa gunung kembar seperti jenjang pendidikan, tahun, dan wilayah.

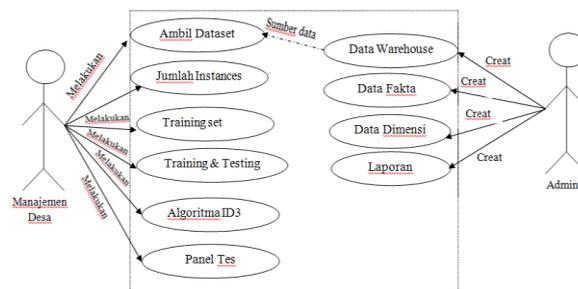


Gambar 8. Jenis Usaha

Model yang digambarkan pada pemodelan data warehouse seperti yang ditunjukkan pada Gambar 8 merupakan gambaran dari kejadian nyata tentang jenis usaha desa gunung kembar seperti kepemilikan usaha, tahun dan wilayah.

5) Desain Sistem

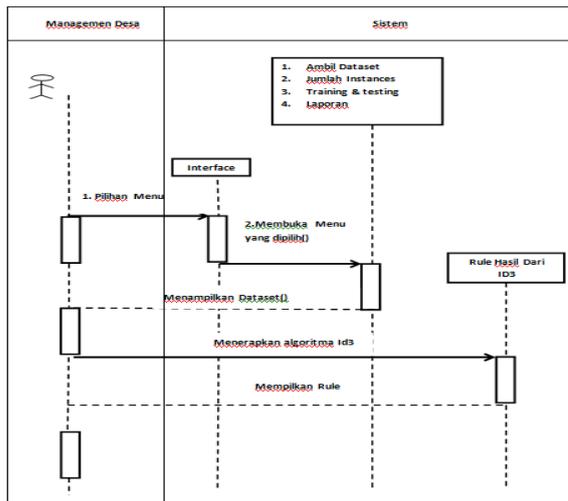
Pada sistem yang diusulkan bahwa seorang admin bertugas untuk memasukkan atau membuat data ke data warehouse sedangkan manajemen desa bertugas sebagai penganalisa dalam pengambilan keputusan. Adapun fitur yang akan disediakan dalam sistem yang diusulkan adalah fitur ambil dataset, jumlah instances, training set, training dan testing, panel tes dan fitur untuk data warehouse itu sendiri hanya memuat data fakta, data dimensi dan laporan yang disimpan menjadi satu yaitu data warehouse.



Gambar 9. Usecase Diagram System

Pada activity diagram tersebut digambarkan alur sistem/ prose dari sistem yang akan dijalankan berdasarkan fungsi dari fitur masing-masing yaitu sebagai contoh fitur penerapan Algoritma Id3. Sistem menampilkan user interface design dengan tampilan pilihan menu kemudian user memilih menu menambahkan dataset kemudian sistem mengeksekusi perintah tersebut dengan menampilkan lokasi dataset yang berextensi.arff dan user memilih dataset pola hidup masyarakat, selanjutnya sistem akan menampilkan jumlah instance dari dataset tersebut. Namun user juga bias merubah jumlah instances dari dataset yang ingin dijadikan data training dengan cara klik fitur training set maka sistem akan menampilkan pilihan berapa berapa persen(%) jumlah dataset yang ingin dijadikan data training. Setelah pemilihan instances selesai langkah selanjutnya adalah menerapkan

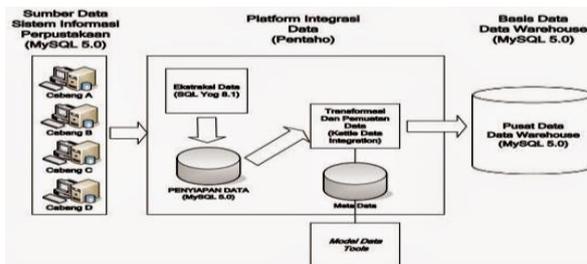
algoritma ID3 dengan klik button Algoritma ID3 maka sistem akan menampilkan hasil berupa Rule.



Gambar 10. Diagram activity analis algoritma id3

3. PEMBAHASAN

Proses penyimpanan data yang besar akan dmenggunakan pentaho yaitu perangkat lunak bisnis inteligent yang berjalan di platform Java dan mampu membangun Data Warehouse. Seperti yang ditunjukkan pada Gambar 10.



Gambar 10. Pentaho

Dalam proses pengambilan data dari penduduk desa setempat menggunakan proses angket dan observasi kemudian data tersebut disimpan menggunakan file csv di microsoft excel kemudian diupload ke library weka sehingga ekstensi berubah menjadi Arrf setelah itu lalu diupload ke program yang telah dibuat menggunakan netbean. Setelah proses menggunakan netbean selesai maka akan ditemukan potensi dari penduduk yang dapat menghambat atau menunjang kemajuan dari desa tersebut. Adapun proses yang dihasilkan id3 dijelaskan sebagai berikut. Adapun uji coba yang dilakukan ialah

1. Pembagian Data set

Dataset akan dibagi menjadi 2 subset yaitu data training, dan data testing Dataset original dengan jumlah 23.814 dengan record 2.646 dan 9 field

2. Training Set

Dari dataset original tersebut akan diambil 80% untuk data training dengan jumlah instances 2.116, data testing 20% dengan jumlah 529 instances. fungsi dilakukannya training data ialah untuk membentuk model dan melakukan pelatihan terhadap data tersebut kemudian dari data training

akan diuji model dengan menggunakan data testing. Seperti yang ditunjukkan pada gambar 11.

```
Time taken to build model: 0.04 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      1823      86.1531 %
Incorrectly Classified Instances    293       13.8469 %
Kappa statistic                    0.8478
Mean absolute error                 0.0313
Root mean squared error            0.1251
Relative absolute error             18.9383 %
Root relative squared error        43.5182 %
Total Number of Instances          2116
```

Gambar 11. Hasil Testing dan Training

Dari Gambar 11 tersebut ditunjukkan bahwa dari 80% memiliki jumlah 2116 instance namun setelah dilakukan proses training dengan menggunakan algoritma id3 maka ditemukan 1.823 klasifikasi kebenaran instances dengan akurasi 86, 1531 % dan kesalahan klasifikasai instance berjumlah 293 instances dengan akurasi 13.8469 %. Dari hasil proses yang telah dilakukan dapat disimpulkan hasil dari training tes cukup baik dan kemudian dilakukan percobaan pada tahap data testing. Tree yang dihasilkan pada dataset proses pembelajaran dengan training data 80 % Pada prediksi nilai cukup diketahui bahwa pendidikan merupakan node awal dengan demikian dapat dipastikan bahwa nilai entropy tertinggi pada prediksi nilai baik ialah pendidikan dan gain tertinggi ialah pekerjaan. dari hasil tree yang ditampilkan akan menghasilkan rule klasifikasi. Adapun akurasi yang dihasilkan dari tree berdasarkan nilai prediksi dapat dilihat pada Tabel 1.

Tabel 1. Akurasi

	Data training	Data testing	Akurasi
Baik	80%	20%	79.46 %
Cukup	80%	20%	76.24 %
Sangat baik	80%	20%	76.64 %
Sangat Baik			
Baik, cukup,	80%	20%	79.06 %

Dari tabel 1 tersebut dapat disimpulkan bahwa hasil akurasi rata rata ialah 78% dan membuktikan cukup baik.

K-FOLD/CROSS-VALIDATION

K-Fold/Cross-Validation berguna untuk menghindari overlapping pada data testing dalam tahapan ini memiliki 2 proses yaitu: membagi data menjadi K-subset yang berukuran sama dan menggunakan setiap k - subset untuk data testing dan sisanya sebagai data training. Cross Validation Method fold = 5 dengan

penjelasan berikut:

data terdiri dari 2.646 instances/record dibagi ke dalam 6 blok dengan jumlah instance yang sama yaitu: blok : a, b, c, d dan e, masing-masing terdiri dari 20 instances kualitas kombinasi parameter tertentu diuji dengan cara tahapan:

step 1: training memakai a,b,c,d,f testing memakai a akurasi a.

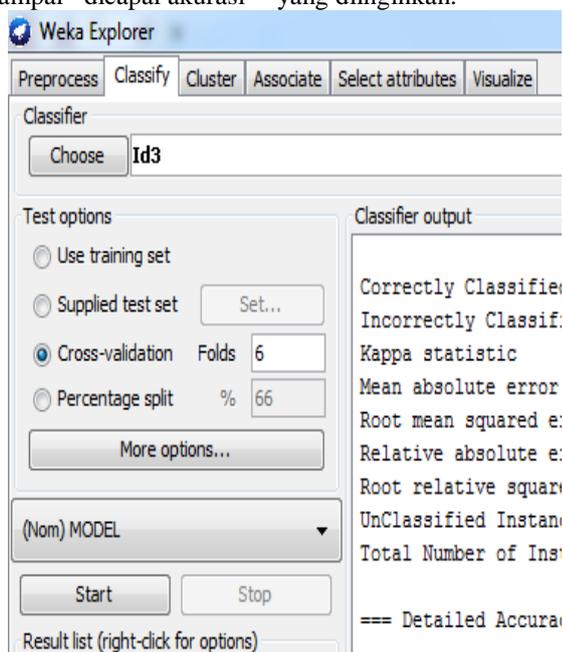
step 2: training memakai a,b,c,e ,f testing memakai d akurasi b

step 3: training memakai a,b, d,e,f testing memakai c akurasi c

step 4: training memakai a, c,d,e ,f testing memakai b akurasi d

step 5: training memakai b,c,d,e,f testing memakai a akurasi e

step 6: training memakai b, c, d, e, f esting memakai f. akurasi f rata-rata akurasi: $(a+b+c+d+e+f)/6$ mencerminkan kualitas parameter yang dipilih mengubah parameter model, dan mengulangi dari no.2 sampai dicapai akurasi yang diinginkan.



Gambar 12. Proses algoritma id3 di weka

Adapun dari proses yang dihasilkan oleh weka ialah *correctly classified instances 2118 dengan akurasi 80.0454 %*, *incorrectly Classified instances 508 dan akurasi 19.1988 %*, *kappa statistic 0.7872*, *mean absolute error 0.0383*, *root mean squared error 0.1513*, *relative absolute error 23.3653 %*, *root relative squared error 52.8366 %*, *unclassified instances 20 0.7559 %*, *total Number of instances 2646*.

4. KESIMPULAN

Berdasarkan dari hasil penelitian maka Jumlah data training dan testing yang digunakan mempengaruhi nilai akurasi pada setiap percobaan. Metode decision tree dapat melakukan klasifikasi jika

dataset bebas dari gangguan/noise dan tahapan yang membutuhkan waktu lama ialah fase preprocessing. Berdasarkan uji coba 80% untuk training set dan 20% untuk testing set dengan perhitungan ID3 menghasilkan akurasi 80%. Pendidikan merupakan atribut yang memiliki nilai entropy tertinggi dengan demikian Variable yang paling berpengaruh terhadap potensi kemajuan desa adalah pendidikan sedangkan yang paling menghambat ialah status pekerjaan dari warga desa setempat.

PUSTAKA

- Arief dwi hartadi; oktalia juwita.(2012). perancangan data warehouse dan penerapan teknik clustering spatial pada wesel. ComTech 199-212.
- Anuja priyama, abhijeeta, rahul guptaa, anju ratheeb, and saurabh srivastavab. (2013). comparative analysis of decision tree classification algorithms. International prees corporation. 334-337.
- Halim Agung. (2017). Implementasi Star Schema Dala Perancangan Aplikasi Data Warehouse Administrasi Desa Wisata Cibuntu. Jakarta.
- Julce Adiana Sidettea, Eko Sedyonob , Oky Dwi Nurhayatic. (2014). Pendekatan Metode Pohon Keputusan dengan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS.
- Permen desa pdt trans nomor 2 tahun 2016 tentang indeks desa membangun. <https://jogloabang.com/pustaka/permendesap dttrans-nomor-2-tahun-2016-tentang-indeks-desa-membangun>.
- R.Fiati. dan N.Latifah. 2013. Sistem Pendukung Keputusan untuk Menentukan Kelayakan Desa Mandiri Menggunakan FMADM Prosiding Seminar Nasional Ilmu Komputer FMIPA Unnes. 137-141
- Safroni, L.M., (2012). Manajemen dan Reformasi Pelayanan Publik dalam Konteks Birokrasi Indonesia, AM Publisng, Yogyakarta
- Satria. 2017. Kerangka Kerja untuk Mengupayakan Satu Sistem Informasi Desa yang Terintegrasi. KOMPAK Working Paper Series No.1
- Wahyudin;. (2009). Metode Iterative Dichotomizer 3 (ID3) 8 Untuk Penyeleksian Pemerimaan Mahasiswa Baru. Jurnal Pendidikan Teknologi Informasi dan Komunikasi, 5-15