|  |  |
| --- | --- |
| |  | | --- | | Kode/Nama Rumpun Ilmu\*\* : 458 / Teknik Informatika  Bidang Fokus : Teknologi Informasi dan Komunikasi |       **LAPORAN**  **PENELITIAN DOSEN UPI “YPTK” PADANG**  logo_upi          **ANALISA TINGKAT PEREKONOMIAN MASYARAKAT PESISIR PANTAI MENGGUNAKAN ALGORITMA DATA MINING**  **TIM PENGUSUL**  **NUGRAHA RAHMANSYAH, S.KOM, M.KOM / 1031038901 (Ketua )**  **SHARY ARMONITHA LUSINIA, S.KOM, M.KOM / 1021128801 (Anggota 1)**      **UNIVERSITAS PUTRA INDONESIA “YPTK” PADANG**  **SEPTEMBER 2018** |

**RINGKASAN**

Kenagarian Lakitan merupakan salah satu dari sekian banyak nagari yang tersebar di Kabupaten Pesisir Selatan. Kenagarian Lakitan ini sebahagian besar berada di pesisir pantai sehingga sebahagian besar masyarakat nagari ini berprofesi sebagai nelayan. Selain berprofesi sebagai nelayan ada juga masyarakat yang bekerja sebagai, pegawai, pedagang, buruh tani, buruh perkebunan, buruh bangunan dan profesi lainnya. Berdasarkan kondisi perekonomian saat ini, masyarakat kenagarian lengayang hanya dikategorikan menjadi dua kelompok, yaitu miskin dan tidak miskin. Penelitian ini dilakukan untuk lebih menganalisa kondisi perekonomian masyarakat, sehingga kita dapat mengetahui kondisi rill tingkat perekonomian dan kita juga dapat mengetahui korelasi atau hubungannya dengan jenis pekerjaan atau profesi masyarakat. Hasil dari penelitian ini akan memberikan manfaat kepada pihak terkait, baik pemerintah ataupun pihak swasta untuk merencanakan peningkatan perekonomian masyarakat khususnya masyarakat kenagarian lakitan, dan masyarakat Kabupaten Pesisir Selatan pada umumnya. Analisa terhadap permasaalahan ini menggunakan konsep data mining dengan algoritma C.45.

**Kata Kunci : *Perekonomian, Data Mining, Algoritma C4.5***

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang Masalah**

Data mining, sering disebut juga sebagai *Knowledge Discovery in Database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data-data yang berukuran besar **(Santoso, 2007)**. Dan keluaran dari Data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition sekarang jarang digunakan karena sudah termasuk bagian dari Data mining. KDD yaitu keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, intreprestasi dan visualisasi dari pola-pola sejumlah kumpulan data.

Pada penelitian (Rina dan Yuwono 2014) membahas mengenai penggunaan Algoritma C.45 dalam menganalisa pola prilaku nasabah dalam pengkreditan dimana dari hasil uji coba sebanyak 20 data uji dihasilkan 13 data benar dan 7 data salah, dengan tingkat akurasi 65%. Dan pada penelitian (Heri dan Sudiyatno 2014) yang membahas tentang prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan, dan prestasi masa lalu dimana dari hasil penelitian tersebut rata-rata keberhasilan dalam melakukan klasifikasi data mencapai akurasi di 95,7%. Hal ini menunjukkan bahwa algoritma ini memiliki performa yang handal dalam melakukan klasifikasi. Dan pada penelitian (Muflikhati, dkk 2010) yang membahas mengenai kondisi sosial ekonomi dan tingkat kesejahteraan keluarga, studi kasus di wilayah pesisir Jawa Barat. Hasil penelitian ini menunjukkan bahwa tidak terdapat perbedaan yang signifikan dalam hal umur ayah dan ibu, total aset, dan pengeluaran per kapita pada keluarga nelayan dan bukan nelayan. Akan tetapi, keluarga nelayan memiliki jumlah anggota keluarga dan pendapatan per kapita yang lebih besar secara signifikan daripada keluarga bukan nelayan. Sebaliknya, pendidikan ayah dan pendidikan ibu pada keluarga nelayan lebih rendah dibandingkan dengan pendidikan ayah dan ibu pada keluarga bukan nelayan.

Jika kesejahteraan keluarga diukur hanya dari aspek ekonomi, maka keluarga nelayan lebih sejahtera dibandingkan dengan keluarga bukan nelayan. Sebaliknya, jika kesejahteraan diukur dengan berbagai dimensi kehidupan, maka keluarga nelayan lebih rendah tingkat kesejahteraannya.

Seiring dengan perkembangan zaman yang semakin pesat, tentu kebutuhan manusia semakin bertambah. Hal ini diakibatkan oleh pertambahan penduduk yang sangat cepat yang menimbulkan aneka permasalahan yang serius bagi kesejahteraan umat manusia. Pertumbuhan penduduk yang begitu pesat dewasa ini sedikit banyak dipengaruhi oleh transisi cepat yang melanda kecenderungan kependudukan. Seiring perkembangan zaman, tentu kebutuhan terhadap manusia bertambah oleh karena itu ekonomi secara terus-menerus mengalami pertumbuhan dan perubahan.

Perubahan yang secara umum terjadi pada perekonomian yang dialami seperti pengangguran , kesempatan kerja, hasil produksi,dan sebagainya. Jika hal ini ditangani dengan tepat maka suatu negara mengalami keadaan ekonomi yang stabil, mempengaruhi kesejahteraan kehidupan penduduk yang ada. Pemerintahpun juga telah mencoba mengatasi permasalahan ekonomi dengan berbagai macam alternative solusi. Namun permasalahan ekonomi masih dapat ditemukan di berbagai daerah.

Kenagarian Lakitan merupakan salah satu dari sekian banyak nagari yang tersebear di Kabupaten Pesisir Selatan. Kenagarian Lakitan ini sebahagian besar berada di pesisir pantai sehingga sebahagian besar masyarakat nagari ini berprofesi sebagai nelayan. Selain berprofesi sebagai nelayan ada juga masyarakat yang bekerja seabagai, pegawai, pedagang, buruh tani, buruh perkebunan, buruh bangunan dan profesi lainnya. Berdasarkan kondisi perekonomian saat ini, masyarakat kenagarian lengayang hanya dikategorikan menjadi dua kelompok, yaitu miskin dan tidak miskin. Penelitian ini dilakukan untuk lebih menganalisa kondisi perekonomian masyarakat, sehingga kita dapat mengetahui kondisi rill tingkat perekonomian dan kita juga dapat mengetahui korelasi atau hubungannya dengan jenis pekerjaan atau profesi masyarakat. Hasil dari penelitian ini akan memberikan manfaat kepada pihak terkait, baik pemerintah ataupun pihak swasta untuk merencanakan peningkatan perekonomian masyarakat khususnya masyarakat kenagarian lakitan, dan masyarakat Kabupaten Pesisir Selatan pada umumnya. Analisa terhadap permasaalahan ini menggunakan konsep data mining dengan algoritma C.45

* 1. **Perumusan Masalah**

Adapun rumusan masalah yang ada pada penelian ini adalah :

1. Bagaimana kondisi perekonomian masyarakat pesisir pantai di Kenagarian Lakitan saat ini?
2. Bagaimana cara menganalisa hubungan pekerjaan dengan perekonomian masyarakat pesisir?
3. Bagaimana cara menganalisa tingkat perekonomian masyarakat pesisir menggunakan algoritma C4.5?
4. Bagaimana cara merancang pohon keputusan (*decision tree*) untuk mengidentifikasi tingkat perekonomian masyarakat pesisir?
5. Bagaimana cara membantu pihak kabupaten, kecamatan maupun kelurahan/desa dalam memberikan solusi terhadap masyarakat yang memiliki tingkat perekonomian yang rendah?

**1.3 Batasan Masalah**

Dari permasalahan yang telah diuraikan maka batasan-batasan dalam penelitian ini sebagai berikut:

1. Data yang diproses berupa data kategori dan tidak ada data *missing value.*
2. Analisa data secara manual dipilih sebanyak 30 sampai 50 sampel data dan data set yang diuji sebanyak 100 sampai 150 data.
3. Dalam tahap implementasi, peneliti menggunakan perangkat lunak Rapid Miner 5.3

**BAB II**

**LANDASAN TEORI**

* 1. **Tingkat Ekonomi**
     1. **Pengertian Ekonomi Secara Umum**

Di dalam struktur sosial kemasyarakatan banyak terdapat ukuran-ukuran di dalam pelapisan-pelapisan yang terjadi di dalam masyarakat tersebut yang lebih dikenal dengan istilah stratifikasi sosial diantaranya adalah pelapisan yang terjadi karena kekayaan seseorang yang lebih dikenal dengan sebutan tingkat ekonomi. Sebelum beranjak lebih jauh untuk memahami hal tersebut perlu untuk menelaah kembali pengertian dari ekonomi itu sendiri sebagai arti dasar pembentukan tingkatan atau pelapisan yang terjadi di dalam struktur sosial kemasyarakatan tersebut. Ekonomi sendiri adalah sebuah cabang ilmu sosial yang berobjek pada individu dan masyarakat, secara etimologis dapat diartikan ekonomi terdiri dari dua suku kata bahasa Yunani yaitu *oikos* dan *nomos* yang berarti tata laksana rumah tangga (Rosyidi, 2009:5). Dapat dilihat dari namanya maka pada saat pertama kali diperkenalkan ekonomi sendiri mempunyai ruang lingkup kajian dan permasalahan yang sangat terbatas yaitu hanya pada tata laksana rumah tangga dan hanya pada permasalahan mencukupi kebutuhan rumah tangga saja. Untuk melihat defenisi ekonomi secara utuh Rosyidi (2009:7) mendefinisikannya sebagai berikut :

“ilmu ekonomi adalah salah satu cabang ilmu pengetahuan yang berdaya upaya untuk memberikan pengetahuan dan pengertian tentang gejala-gejala masyarakat yang timbul karna perbuatan manusia dalam usahanya untuk memenuhi kebutuhannya atau untuk mencapai kemakmuran”

Dari definisi di atas dapat dikatakan bahwa ekonomi secara umum mengkaji mengenai pemenuhan kebutuhan manusia dan kemakmuran manusia, dua hal pokok dari permasalahan ekonomi tersebut yaitu kebutuhan dan pencapaian kemakmuran merupakan salah satu dasar di dalam pelapisan sosial di dalam masyarakat bila dihubungkan dengan permasalahan mikro tingkat ekonomi masyarakat, dengan kata lain semakin makmur seseorang dan semakin mampu untuk memenuhi kebutuhannya dengan berbagai tingkatannya maka semakin tinggi pula tingkat ekonomi seseorang di dalam struktur sosial kemasyarakatan

* + 1. **Faktor Faktor Tingkat Ekonomi**

Ada beberapa faktor yang dapat menentukan tinggi rendahnya sosial ekonomi orang tua di masyarakat, diantaranya tingkat pendidikan, jenis pekerjaan, tingkat pendapatan, kondisi lingkungan tempat tingal, pemilikan kekayaan, dan partisipasi dalam aktivitas kelompok dari komunitasnya. Dalam hal ini uraiannya dibatasi hanya 4 faktor yang menentukan yaitu tingkat pendidikan, pendapatan, dan kepemilikan kekayaan, dan jenis pekerjaan

1. Tingkat Pendidikan

Menurut UU RI No. 20 Tahun 2003 pasal 1, pada dasarnya jenjang pendidikan adalah tahapan pendidikan yang ditetapkan berdasarkan tingkat perkembangan peserta didik, tujuan yang akan dicapai, dan kemampuan yang dikembangkan. Pendidikan menurut UU No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didika secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara. Pendidikan adalah aktivitas dan usaha untuk meningkatkan kepribadian dengan jalan membina potensi-potensi pribadinya, yaitu rokhani (pikir, cipta, rasa, dan hati nurani) serta jasmani (panca indera dan keterampilan-keterampilan).

Menurut UU RI No. 20 Tahun 2003 pasal 3 Pendidikan bertujuan untuk “Mencerdaskan kehidupan bangsa dan mengembangkan manusia Indonesia seutuhnya, yaitu manusia yang beriman dan bertaqwa kepada Tuhan Yang Maha Esa dan berbudi pekerti luhur, memiliki pengetahuan dan ketrampilan, kesehatan jasmani dan rohani, kepribadian yang mantap dan bertanggung jawab kemasyarakatan dan kebangsaan”. Untuk mencapai tujuan tersebut, pendidikan diselenggarakan melalui jalur pendidikan sekolah (pendidikan formal) dan jalur pendidikan luar sekolah (pendidikan non formal). Jalur pendidikan sekolah (pendidikan formal) terdapat jenjang pendidikan sekolah, jenjang pendidikan sekolah pada dasarnya terdiri dari pendidikan prasekolah, pendidikan dasar, pendidikan menengah, dan pendidikan tinggi

1. Tingkat Pendapatan

Pendapatan adalah jumlah semua pendapatan kepala keluarga maupun anggota keluarga lainnya yang diwujudkan dalam bentuk uang dan barang. Menurut Sumardi dalam Yerikho (2007) mengemukakan bahwa pendapatan yang diterima oleh penduduk akan dipengaruhi oleh tingkat pendidikan yang dimilikinya. Dengan pendidikan yang tinggi mereka akan dapat memperoleh kesempatan yang lebih luas untuk mendapatkan pekerjaan yang lebih baik disertai pendapatan yang lebih besar. Sedangkan bagi penduduk yang berpendidikan rendah akan mendapat pekerjaan dengan pendapatan yang kecil.

Dalam penelitian ini yang dimaksud dengan pendapatan orang tua adalah penghasilan yang di terima orang tua dalam bentuk uang dari hasil kerja baik secara formal maupun informal . Berdasarkan penggolongannya, BPS membedakan pendapatan penduduk menjadi 4 golongan yaitu :

1. Golongan pendapatan sangat tinggi adalah jika pendapatan rata-rata lebih dari Rp. 3.500.000,00 per bulan
2. Golongan pendapatan tinggi adalah jika pendapatan rata-rata antara Rp. 2.500.000,00 s/d Rp. 3.500.000,00 per bulan
3. Golongan pendapatan sedang adalah jika pendapatan rata-rata dibawh antara Rp. 1.500.000 s/d Rp. 2.500.000,00 per bulan
4. Golongan pendapatan rendah adalah jika pendapatan rata-rata Rp.1.500.000,00 per bulan kebawah.

Dari keterangan diatas dapat dikatakan bahwa pendapatan juga sangat berpengaruh terhadap tingkat ekonomi seseorang. Apabila seseorang mempunyai pendapatan yang tinggi, maka dapat dikatakan bahwa tingkat ekonominya tinggi juga. Disamping memiliki penghasilan pokok setiap Keluarga biasanya memiliki penghasilan lain yang meliputi penghasilan tambahan dan penghasilan insidentil.

1. Kepemilikan Kekayaan dan Fasilitas

Pemilikan kekayaan atau fasilitas adalah kekayaan dalam bentuk barang-barang dimana masih bermanfaat dalam menunjang kehidupan ekonominya. Fasilitas atau kekayaan itu antara lain:

1. Barang-barang berharga

Menurut Abdulsyani (1994), bahwa pemilikan kekayaan yang bernilai ekonomis dalam berbagai bentuk dan ukuran seperti perhiasan, televisi, kulkas dan lain-lain dapat menunjukkan adanya pelapisan dalam masyarakat.

1. Jenis-jenis kendaraan pribadi.

Kendaraan pribadi dapat digunakan sebagai alat ukur tinggi rendahnya tingkat sosial ekonomi orang tua. Misalnya: orang yang mempunyai mobil akan merasa lebih tinggi tingkat taraf ekonominya dari pada orang yang mempunyai sepeda motor.

1. Jenis Pekerjaan

Pekerjaan akan menentukan status sosial ekonomi karena dari bekerja segala kebutuhan akan dapat terpenuhi. Pekerjaaan tidak hanya mempunyai nilai ekonomi namun usaha manusia untuk mendapatkan kepuasan dan mendapatkan imbalan atau upah, berupa barang dan jasa akan terpenuhi kebutuhan hidupnya. Pekerjaan seseorang akan mempengaruhi kemampuan ekonominya, untuk itu bekerja merupakan suatu keharusan bagi setiap individu sebab dalam bekerja mengandung dua segi, kepuasan jasmani dan terpenuhinya kebutuhan hidup.

Menurut Manginsihi (2013: 15), pekerjaan adalah kegiatan yang dilakukan oleh orang tua siswa untuk mencari nafkah. Pekerjaan yang ditekuni oleh setiap rang berbeda-beda, perbedaan itu akan menyebabkan perbedaan tingkat penghasilanyang rendah sampai padatingkat penghasilan yang tinggi, tergantung pada pekerjaan yang ditekuninya. Contoh pekerjaan berstatus sosioekonomi rendah adalah pekerja pabrik, buruh manual, penerima dana kesejahteraan, dan pekerja pemeliharaan. Santrock (2007: 282)

Jadi untuk menentukan status sosial ekonomi yang dilihat dari pekerjaan, maka jenis pekerjaan dapat diberi batasan sebagai berikut:

1. Pekerjaan yang berstatus tinggi, yaitu tenaga ahli teknik dan ahli jenis, pemimpin ketatalaksanaan dalam suatu instansi baik pemerintah maupun swasta, tenaga administrasi tata usaha.
2. Pekerjaan yang berstatus sedang, yaitu pekerjaan di bidang penjualan dan jasa.
3. Pekerjaan yang berstatus rendah, yaitu petani dan operator alat angkut/bengkel
   1. **Data Mining** 
      1. **Definisi Data Mining**

*Data mining* adalah sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis dari sekumpulan data yang berjumlah banyak, data mining sering dianggap sebagai bagian dari *Knowledge Discovery in Database* (KDD) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data, proses KDD secara garis besar dapat dijelaskan sebagai berikut (Sunjana, 2010) :

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

1. *Pre-processing/ Cleaning*

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Selain itu dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

1. *Transformation*

*Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

1. *Interpretation/ Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap

* + 1. **Pengelompokan Data Mining**

Menurut Larose dalam bukunya yang berjudul ”Discovering Knowledge in Data: An Introduction to Data Mining”, data mining dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, (Gunadi dan Indra Sensuse, 2012) yaitu

1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

1. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data (record) lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

1. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

1. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

1. Pengklasteran (Clusterring)

Pengklasteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan record dalam klaster yang lain. Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

1. Asosiasi

Tugas asosiasi dalam data mining adalah untuk menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah market basket analysis atau analisis keranjang belanja.

* + 1. **Tahap-Tahap Data Mining**

Data mining seharusnya dipahami sebagai suatu proses, yang memiliki tahapan-tahapantertentu dan juga ada umpan balik dari setiap tahapan ke tahapan sebelumnya. Pada umumnyaproses data mining berjalan interaktif karena tidak jarang hasil data mining pada awalnya tidaksesuai dengan harapan analisnya sehingga perlu dilakukan desain ulang prosesnya.Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap yangdiilustrasikan di Gambar di bawah ini. Tahap-tahap tersebut.bersifat interaktif di mana pemakai terlibatlangsung atau dengan perantaraan knowledge base (Syafiq Adriansyah, 2010).

1. Pembersihan data

Digunakan untuk membuang data yang tidak konsisten dan noise

1. Intergrasi Data

Data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Hasil integrasi data sering diwujudkan dalam sebuah data warehouse karena dengan data warehouse, data dikonsolidasikan dengan struktur khusus yang efisien. Selain itu data warehouse juga memungkinkan tipe analisa seperti OLAP.

1. Transformasi data

Transformasi dan pemilihan data ini untuk menentukan kualitas dari hasil data mining, sehingga data diubah menjadi bentuk sesuai untuk di-Mining.

1. Aplikasi Teknik Data Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai.

1. Evaluasi pola yang ditemukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

1. Presentasi Pengetahuan

Presentasi pola yang ditemukan untuk menghasilkan aksi tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat.

* 1. **Pohon Keputusan**

Pohon keputusan atau dikenal dengan decision tree adalah salah satu metode klasifikasi yang menggunakan representasi suatu struktur pohon yang berisi alternatife-alternatif untuk pemecahan suatu masalah. Pohon ini juga menunjukkan factor-faktor yang mempengaruhi hasil alternative dari keputusan tersebut.

Decision tree menggunakan struktur hierarki untuk pembelajaran *supervised*. Proses dari *decision* *tree* di mulai dari *root* *node* hingga *leaf* *node* yang dilakukan secara rekursif. Pada *decision* *tree* terdiri dari tiga bagian (Ariadni dan Arieshanti) yaitu :

1. Root node

Node ini merupakan node yang terletak paling atas dari suatu pohon.

1. Internal node

Node ini merupakan node percabangan, hanya terdapat satu input serta mempunyai minimal dua output.

1. Leaf node

Node ini merupakan node akhir, hanya memiliki input, dan tidak memiliki output.

* 1. **Algoritma C4.5**

Sebelum membahas algoritma C4.5 perlu dijelaskan terlebih dahulu algoritma ID3 karena C4.5 adalah ekstensi dari algoritma *decision-tree* ID3. Algoritma ID3/C4.5 ini secara rekursif membuat sebuah *decision tree* berdasarkan training data yang telah disiapkan. Algoritma ini mempunyai inputan berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenaranya. Sedangkan samples merupakan *field- field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010).

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam Rumus 1 (Defiyanti, Pardede).

Di mana :

*S* : Himpunan kasus

*A* : Atribut

*n* : Jumlah partisi atribut *A*

|*Si*| : Jumlah kasus pada partisi ke *i*

|*S*| : Jumlah kasus dalam *S*

Sedangkan perhitungan nilai entropy dapat dilihat pada rumus 2 berikut(Hardikhar, Shrivastava, dan choundhary, 2012):

Di mana :

*S* : Himpunan Kasus

*A* : Fitur

*n* : Jumlah partisi *S*

*pi* : Proporsi dari *S* *i* terhadap *S*

* 1. **Information Gain**

*Information* *gain* adalah salah satu attribute *selection measure* yang digunakan untuk memilih test atribute tiap node pada tree. Atribut dengan information gain tertinggi dipilih sebagai test atribut dari suatu node.. Ada 2 kasus berbeda pada saat penghitungan Information Gain, pertama untuk kasus penghitungan atribut tanpa *missing value* dan kedua, penghitungan atribut dengan *missing value* (Sujana, 2010)*.*

1. Penghitungan Information Gain tanpa missing value

Misalkan S berisi *s* data samples. Anggap atribut untuk class memiliki *m* nilai yang berbeda, Ci (untuk i = 1, …,I). anggap s*i* menjadi jumlah samples S pada class Ci. Maka besar information-nya dapat dihitung dengan :

Di mana adalah probalitas dari sampel yang mempunyai class

1. Penghitungan Information Gain dengan missing value

Untuk atribut dengan missing value penghitungan information gain-nya diselesaikan dengan Gain Ratio. Sebelum menghitung gain ratio terlebih dahulu dihitung I ( s1, s2,…,sm ) dan E(A).

Di mana perhitungan dan *E(A)* hanya dilakukan pada atribut yang ada nilai nya.

* 1. **Metode Heuristik**

Silver et al. (1980) mendefinisikan sebuah metode *heuristic* sebagai prosedur untuk pemecahan masalah matematika didefinisikan dengan baik oleh *intuitif* pendekatan di mana struktur. Masalah dapat ditafsirkan dan dieksploitasi cerdas untuk mendapatkan solusi yang masuk akal.

Menurut perak et al. (1980), ada beberapa kemungkinan alasan untuk menggunakan *heuristic* metode solusi. Ini termasuk:

1. Matematika masalahnya tentu seperti alam yang analitik (bentuk tertutup)

atau iteratif solusi prosedur tidak diketahui.

1. Meskipun prosedur solusi analitik atau berulang-ulang yang tepat yang mungkin ada, itu mungkin produk mahal untuk menggunakan atau mungkin tidak realistis dalam persyaratan data mereka.
2. Metode heuristik, Desain, mungkin sederhana untuk pembuat keputusan untuk memahami, sehingga sangat meningkatkan kemungkinan implementasi.
3. Untuk masalah yang didefinisikan dengan baik yang dapat diselesaikan secara optimal *heuristic* metode dapat digunakan untuk keperluan belajar, misalnya, untuk mengembangkan perasaan intuitif sebagai variabel apa penting. (Ini erat paralel salah satu alasan utama untuk menggunakan metode simulasi dalam penelitian operasional).
4. Heuristic dapat digunakan sebagai bagian dari prosedur yang berulang-ulang yang menjamin menemukan solusi yang optimal. Ada dua kemungkinan yang berbeda:
5. Untuk dengan mudah mendapatkan solusi layak awal.
6. Untuk membuat keputusan di langkah perantara prosedur solusi tepat.
7. Heuristik dapat digunakan untuk memberikan solusi awal seperti 'baik'. Dalam pencacahan implisit pendekatan pemecahan solusi awal yang baik dapat memberikan terikat yang secara drastis mengurangi upaya komputasi.

**BAB III**

**TUJUAN DAN MANFAT PENELITIAN**

* 1. **Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah:

1. Menganalisa kelayakan kredit menggunakan algoritma C4.5
2. Merancang atribut-atribut atau prediktor variable untuk manganalisa dengan menggunakan algoritma C4.5
3. Membangun descision tree pada algritma C4.5 secara manual.
4. Menguji algoritma C4.5 dengan menggunakan software.
   1. **Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah:

1. Membantu pihak kelurahan atau dinas terkait di Kenagarian Lakitan dalam mengambil keputusan untuk menentukan masyarakat penerima bantuan.
2. Sebagai referensi bagi penelitian selanjutnya yang berkaitan dengan penggunaan teknik data mining

**BAB IV**

**METODE PENELITIAN**

* 1. **Kerangka Penelitian**

Dalam menganalisa data, kita harus memikirkan tahapan atau langkah-langkah kerja. Dalam hal ini, penulis memaparkan beberapa tahap dalam menganalisa data yang dapat dilihat pada Gambar 3.1 yang merupakan kerangka penelitian.

Pengumpulan Data

Analisa Masalah

Implementasi Menggunakan software

Pemilihan Variabel Data

Menentukan Tujuan Penelitian

Mempelajari Literatur

Pengujian Data Secara Manual

**Gambar 4.1 Kerangka Penelitian**

Gambar di atas merupakan acuan penulis sebagai kerangka pemikiran dalam menganalisa data.

1. Analisa Masalah

Sebelum kita melakukan penelitian, terlebih dahulu kita menentukan masalah tingkat perekonomian masyarakat pesisir pantai di Kenagarian Lakitan.

1. Menentukan Tujuan Penelitian

Sebelum kita melakukan penelitian, kita perlu menetukan untuk apa penelitian ini dilakukan, dan apa kegunaan penelitian ini, sehingga kita bisa mengetahui manfaat dari penelitian yang kita lakukan.

1. Mempelajari Literatur

Untukmengetahui langkah-langkah dan teknik yang digunakan dalam menganalisa penelitian ini, kita mencari sumber-sumber pengetahuan yang berkaitan dengan penelitian yang akan kita lakukan, baik berupa jurnal, seminar nasional dan internasional, buku dan lain sebagainya.

1. Pengumpulan Data

Dalam pengumpulan data ini, digunakan beberapa metode yang mendukung antara lain :

1. Interview yaitu melakukan penelusuran untuk mendapatkan data dan informasi melalui tanya jawab dan wawancara dengan pihak orang yang berkopenten terhadap permasalahan yang diteliti.

2. Studi Pustaka yaitu pengumpulan data dan penelusuran informasi dengan cara membaca dan mempelajari buku-buku dan jurnal-jurnal penelitian yang berkaitan dan menunjang, baik dalam menganalisa data dan informasi maupun pemecahan masalah secara keseluruhan.

1. Pemilihan Variabel Data

Setelah kita melakukan analisa masalah, selanjutnya kita harus menentukan variable input data. Pada *algoritma* *C4.5* akan dilakukan perancangan *decision tree,* dengan memilih atribut pada *field- field* data sebagai akar, membuat cabang untuk masing-masing nilai. Membagi kasus dalam cabang, dan mengulangi proses untuk masing-masing cabang samapi semua kasus pada cabang memiliki kelas yang sama.

1. Pengujian Data Secara Manual

Pengujian data anggota kredit ini menggambarkan faktor-faktor atau model keterhubungan data anggota kredit. Pada *algoritma C4.5* inputan berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan samples merupakan *field- field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data.

1. Implementasi Menggunakan *Software*

Di dalam penganalisaan data nasabah kredit ini penulis menggunakan *software* Rapid Miner versi 5.3. Di mana software Rapid Miner memiliki pilihan *decision tree* untuk keputusan algoritma *C4.5.*

**BAB V**

**HASIL DAN LUARAN YANG DICAPAI**

Pada bab ini di laporkan hasil dan luaran yang telah dicapai dalam rangkaian kegiatan penelitian yang dilaksanakan pada bulan Februari 2018 sampai dengn Semptember 2018 di Kenagarian Lakitan.

* 1. **Hasil Yang Dicapai**
     1. Data uji coba

Berikut data yang telah di peroleh dari hasil penelusuran ke Kenagarian Lakitan, Kabupaten Pesisir Selatan

**Tabel 5.1. Data Uji Coba**



* + 1. Penetapan Intrumen

Dari data yang telah diperoleh, maka akan ditentukan variabel data yang menjadi variabel penentu keputusan yaitu, masyarakat yang menerima bantuan berupa menerima dan tidak menerima. Data yang menjadi variabel penentu dalam pembentukan pohon keputusan pada penelitian ini penghasilan, pengeluaran, pekerjaan, jumlah tanggungan, status kepemilikan rumah,produksi pertanian, produsi perternakan, asset tanah, kondisi rumah, dan asset lainnya. Pemilihan variabel penentu tersebut dengan pertimbangan bahwa variabel-variabel yang dipilih merupakan variabel yang paling mempengaruhi keputusan penerimaan bantuan.

* 1. **Luaran yang dicapai**

Adapun luaran yang telah dicapai sebagai berikut :

1. Artikel ilmiah berupa jurnal
2. Prosiding berupa draf
3. Bahan ajar materi Data Mining Algoritma C4.5

**BAB VI**

**KESIMPULAN DAN SARAN**

* 1. **Kesimpulan**

Dari uraian pada bab-bab yang sudah dibahas sebelumnya dapat ditarik kesimpulan :

1. Metode pohon keputusan (*decision tree*) yang diproses dengan *software Rapidminer* dapat mengidentifikasi tingkat perekonomian dengan baik.
2. Pemilihan variabel (atribut kondisi dan atribut keputusan) yang akan digunakan dalam menentukan sebuah klasifikasi juga sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan.
3. Sistem yang dibangun dapat membantu dalam mengklasifikasikan kriteria tingkat perekonomian masyarakat pesisir pantai dalam menerima bantuan. Sehingga dapat membantu pihak kenagarian dalam pengambilan keputusan.
4. Algoritma *C4.5* dianggap sebagai algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan (*decision tree*) maupun dalam aturan rule *If – Then* sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.
   1. **Saran**

Adapun beberapa saran dari penulis untuk pengembangan tesis ini adalah:

1. Dalam menggunakan algoritma C4.5 untuk melakukan klasifikasi, harus dilakukan pemilihan variabel yang tepat agar hasil dari pohon keputusan lebih akurat atau terperinci
2. Penelitian ini dapat dilanjutkan dengan menambahkan variabel-variabel yang memiliki hubungan dengan masalah pemberian bantuan berdasarkan tingkat perekonomian agar nantinya tingkat pemberian keputusan yang lebih baik dapat tercapai.
3. Penelitian ini dapat dilanjutkan dengan menggunakan metode-metode data mining lainnya untuk mendapatkan perbandingan metode yang lebih baik.

**DAFTAR PUSTAKA**

Ariatni, Ratih., dan Arieshanti Isye. *“Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur Yang Tidak Pasti”* 1-7.

Bin Ismail, Zuhaimy. (2008) *“Pembangunan Kaedah Heuristik Berasaskan Algoritma Genetika Untuk Menyelesaikan Masalah Penjalanan Kendaraan”* Vote No. 74285.

Defiyanti, Sofi., dan Crispina Pardede, D.L. *“Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam-Mail”*

Dewi Indah Sari, Rina., Yuwono Sindunata. (2014) *“Peneraan Data Mining Untuk Analisa Pola Perilaku Nasabah Dalam Pengkreditan Menggunakan Metode C.45 Studi Kasus Pada KSU InsanKamil Demak”* Vol. 8 No.2.

Gunadi, Goldie., dan Indra Sensue, Dana. (2012) *“Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Prodek Buku Dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth) : Studi Kasus Percetakan PT. Gramedia”* Vol. 4 No. 1, 118-132.

Hardikar, Surbini., Shrivastava, Ankur dan Choudhary, Vijay. (2012) *“Comparison Between ID3 and C4.5 in Contrast to IDS”* Vol. 02(7), 659-667.

Kusrini, dan Taufiq Luthfi, Emha. *“Algoritma C4.5”*

Kresna, Sanjaya Dicky (2013) *“Pengaruh Tingkat Ekonomi Terhadap Prestasi Belajar Siswa SMA Negeri 1 Garum Kabupaten Blitar Tahun 2012/2013”*

Prasetyo, Eko. (2012) *“Data Mining Konsep dan Aplikasi menggunakan MATLAB”* Andi, Yogyakarta

Ridwan, Mujib., Suyono, Hadi., dan Sarosa, M. (2013) *“Penerapan Data Mining Untuk Evaluasi Kinerja Akademin mahasiswa Menggunakan Algoritma Naïve Bayes Classifier”* Vol. 7 No. 1, 59-64.

Santosa, Budi. (2007) *“Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis”* Graha Ilmu, Yogyakarta

Sunjana. (2010) *“Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree”* 24-29.

Sunjana. (2010) *“Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5”* 31-34.

Susanto, Heri., Sudiyatno. (2014) *“Data Mining Untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan, dan Prestasi Masa Lalu”* Vol. 4 No 2, 222-231.

Syafiq Ardiyansyah, Ghiyats. (2010) *“Tugas Data Mining dan Knowledge Discovery in Database”* 1-5.