

PERANCANGAN APLIKASI DATA MINING UNTUK KLASIFIKASI KELUARGA MISKIN MENGGUNAKAN METODE DECISION TREE

Enggi Fernanda¹, Dewi Suranti², Aji Sudarsono³

Universitas Dehasen, Bengkulu

e-mail: ¹enggifernanda0601@gmail.com, ²dewisuranti@unived.ac.id,

³ajisudarsono@unived.ac.id

Abstract: *The data mining application for classifying poor families was created using the Visual Basic.Net programming language. To classify poor families, the application has applied the Decision Tree Method by looking at the results of the Gain and Entropy values for each branch. Where the branches used consist of work, property tax, floors, walls, water sources, PLN power and income. With this application, it is hoped that it can help the Kerkap District, North Bengkulu Regency in analyzing and identifying poor families who are entitled to receive assistance. Based on the results of the tests that have been carried out, the functionality of the data mining application for classifying poor families is good and smooth as it should be, and is able to provide information on the results of poor family classification analysis using the decision tree approach.*

Keywords: Data Mining; Decission Tree, C4.5; Poor Families.

Abstrak: Aplikasi data mining untuk klasifikasi keluarga miskin dibuat menggunakan bahasa pemrograman Visual Basic.Net. Untuk melakukan klasifikasi keluarga miskin, pada aplikasi telah diterapkan Metode Decision Tree dengan melihat hasil nilai Gain dan Entropy pada masing-masing cabang. Dimana cabang yang digunakan terdiri pekerjaan, Pajak Bumi Bangunan, lantai, dinding, sumber air, daya PLN dan penghasilan. Dengan adanya aplikasi ini, diharapkan dapat membantu pihak Kecamatan Kerkap Kabupaten Bengkulu Utara dalam menganalisa dan mengetahui keluarga miskin yang berhak untuk mendapatkan bantuan. Berdasarkan hasil pengujian yang telah dilakukan, fungsional dari aplikasi data mining untuk klasifikasi keluarga miskin dengan baik dan lancar sebagaimana mestinya, dan mampu memberikan informasi hasil analisis klasifikasi keluarga miskin melalui pendekatan decision tree.

Kata kunci: Data Mining; Decission Tree, C4.5; Keluarga Miskin

PENDAHULUAN

Masalah kemiskinan merupakan salah satu permasalahan yang mendasar yang menjadi perhatian bersama bagi pemerintah. Kemiskinan memiliki arti yang lebih dalam karena berkaitan dengan ketidak-mampuan untuk mencapai aspek di luar pendapatan (non-income factors) seperti akses kebutuhan minimum, kesehatan, pendidikan, air bersih. Kemiskinan menjadi faktor utama dari sejumlah kejadian sosial, politik serta ekonomi. Kejadian dan fenomena ini terjadi pada sebagian masyarakat di wilayah Bengkulu Utara tepatnya pada

Kecamatan Kerkap. Berdasarkan data statistik yang diperoleh pada website resmi BPS (Badan Pusat Statistik) Bengkulu Utara jumlah penduduk miskin pada tahun 2021 berjumlah 35.550 jiwa.

Sulitnya melakukan atau menentukan klasifikasi keluarga miskin masih menjadi perdebatan dikalangan masyarakat. Oleh karena itu penulis mencoba melakukan penelitian dalam mengklasifikasi keluarga miskin dengan kriteria pekerjaan, Pajak Bumi Bangunan, lantai, dinding, sumber air, daya PLN dan penghasilan dengan memanfaatkan teknik data mining.

Data mining merupakan teknologi yang menggabungkan metode analisis

tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar. Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat dan tersimpan di dalam database yang besar. Salah satu algoritma pada data mining adalah Decision Tree C.45. Algoritma ini dipilih karena memiliki kelebihan, yaitu mudah dimengerti, fleksibel dan menarik karena dapat divisualisasikan dalam bentuk gambar. Dengan menggunakan algoritma Decision Tree C4.5 diharapkan dapat memberikan hasil suatu model pohon keputusan yang dapat mengklasifikas keluarga miskin serta dapat membantu dalam mengambil menentukan keputusan bagi pemerintah setempat.

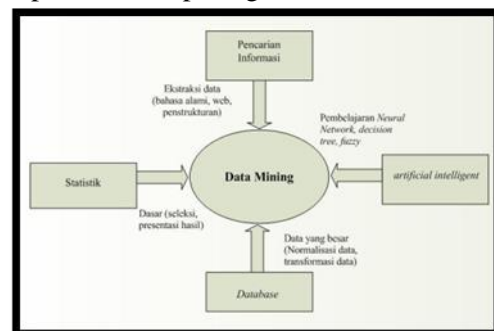
Data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani jumlah data yang sangat besar, dimensi data yang tinggi, dan data yang heterogen dan berbeda sifat. Data mining dibagi menjadi beberapa kelompok yaitu, deskripsi, estimasi, prediksi, klasifikasi, pengklasteran, dan asosiasi (Fajrin & Algifanri, 2018).

Data Mining atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya (Sunia, Budi, & Jusia, 2019).

Data mining sering disebut dengan istilah Knowledge Discovery in Database (KDD). Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data

berukuran besar (Nasution, Windarto, & M.Fauzan, 2020)

Berdasarkan kedua pengertian tersebut di atas, dapat disimpulkan bahwa data mining merupakan suatu proses yang digunakan untuk menganalisis data agar dapat mengekstraksi pengetahuan secara otomatis berdasarkan jumlah data yang sangat besar, dimensi data yang tinggi dan data yang heterogen dan berbeda sifat seperti terlihat pada gambar 1 berikut .:



Gambar 1. Bidang Cakup Data Mining

Tahapan data mining terdiri dari :

Pembersihan data (data cleaning)

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang 4 duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga 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.

Integrasi data (data integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik.

Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis

yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

Transformasi data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

Proses mining

Sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.

Evaluasi pola (pattern evaluation)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based 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

Presentasi pengetahuan (Presentation)

Merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa

membantu mengkomunikasikan hasil data mining.

Data mining pada dasarnya terdiri dari beberapa fungsi atau komponen yang beroperasi secara berurutan yang mana tiap tahapnya akan memberikan output untuk tahap berikutnya. Terdapat enam fungsi utama dari data mining yang dapat dijabarkan sebagai berikut :

Classification, tahap menemukan model yang membagi data menjadi beberapa kelas Regression, tahap dimana data dipetakan kedalam variable prediksi nyata Clustering, tahap untuk mengidentifikasi kumpulan kategori atau kelompok yang mendeskripsikan data. Dependency Modelling, tahap untuk menemukan model yang mendeskripsikan hubungan atau ketergantungan antar variable dari data. Deviation Detection, tahap menemukan perubahan signifikan pada data. Summarization, tahap menemukan deskripsi akhir dari sekumpulan data.

Teknik dalam bidang data mining terus berkembang seiring dengan penelitian-penelitian yang terus dilakukan pada bidang data mining, sebagai berikut.

Classification

Clustering

Prediction

Association Rule

Neural Networks

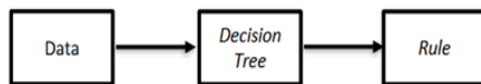
Decision Trees

Nearest Neighbor Method

Decision Tree

Decision tree adalah sebuah diagram alir yang mirip dengan struktur pohon, dimana setiap internal node menotasikan atribut yang diuji, setiap cabangnya merepresentasikan hasil dari atribut tes tersebut dan leaf node merepresentasikan kelas-kelas tertentu atau distribusi dari kelas-kelas. Istilah Decision Tree adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan untuk memprediksi data yang belum memiliki kelas data tertentu (Pribadi, Poningsih, & Tambunan, 2020).

Decision tree merupakan teknik model prediksi yang dapat digunakan untuk klasifikasi dan prediksi tugas. Decision tree menggunakan teknik “membagi dan menaklukkan” untuk membagi ruang pencarian masalah menjadi himpunan masalah. Proses pada decision tree adalah mengubah bentuk data tabel menjadi sebuah model tree. Model tree akan menghasilkan rule dan disederhanakan (Bahri & Lubis, 2020)



Gambar 2. Konsep Pohon Keputusan

Algoritma C.45

Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru (Pribadi, Poningsih, & Tambunan, 2020).

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu (Eska, 2016).

Cara algoritma C4.5 untuk membangun pohon keputusan yaitu :

Memilih atribut yang akan digunakan sebagai akar Membuat sebuah cabang untuk setiap nilai Mengulang proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 : Menyiapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah

di kelompokkan ke dalam kelas tertentu Setelah itu tentukan akar dari pohon. Pilih akar dari atribut, cara adalah dengan menghitung nilai gain dari semua atribut, yang menjadi akar pertama adalah nilai gain yang paling. Sebelum menentukan nilai gain, terlebih dahulu hitung nilai entropy. Untuk menentukan nilai entropy gunakan rumus:

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i$$

Dimana :

S = Himpunan Kasus

n = Jumlah Partisi S

pi = Proporsi Si terhadap S

Setelah itu menentukan nilai gain menggunakan rumus :

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \cdot \text{Entropy}(S_i) \right)$$

Dimana :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Selanjutnya mengulangi langkah ke-2 sampai semua record terpartisi secara sempurna. Proses partisi pohon keputusan akan berhenti pada saat :

Semua record dalam simpul N mendapat kelas yang sama Tidak ada atribut di dalam record yang dipartisi lagi Tidak ada record di dalam cabang yang kosong

METODE

Metode pengumpulan data yang penulis gunakan pada penelitian ini adalah sebagai berikut :

1. Observasi Metode pengumpulan data yang dilakukan dengan cara mengadakan pengamatan atau peninjauan secara langsung ke kantor Camat Kerkap Kabupaten Bengkulu

- Utara dengan tujuan mendapatkan informasi atau data yang diperlukan untuk penelitian ini.
2. Wawancara Melakukan tanya jawab langsung kepada Ibu Indriani selaku sekretaris kecamatan Kerkap Kabupaten Bengkulu Utara terkait dengan permasalahan kemiskinan yang ada di Kecamatan Kerkap
 3. Studi Pustaka Mempelajari dan mencari data yang berasal dari buku, jurnal dan referensi yang diperoleh dari toko buku dan perpustakaan Universitas Dehasen Bengkulu yang berhubungan dengan masalah yang ditulis.

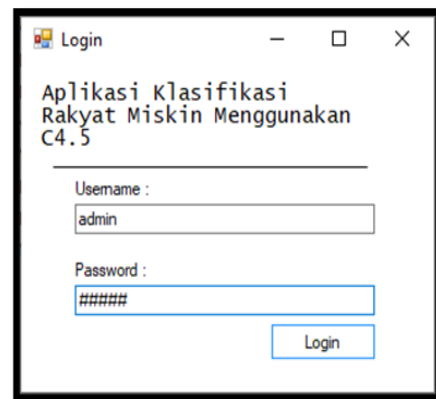
HASIL DAN PEMBAHASAN

Hasil

Aplikasi data mining untuk klasifikasi keluarga miskin dibuat menggunakan bahasa pemrograman Visual Basic .Net. Untuk melakukan klasifikasi keluarga miskin, pada aplikasi telah diterapkan Metode Decision Tree C4.5 dengan melihat hasil nilai Gain dan Entropy pada masing-masing kriteria. Dimana kriteria yang digunakan terdiri pekerjaan, Pajak Bumi Bangunan, lantai, dinding, sumber air, daya PLN dan penghasilan. Dengan adanya aplikasi ini, diharapkan dapat membantu dalam melakukan klasifikasi keluarga miskin berdasarkan data kriteria yang digunakan.

Pembahasan

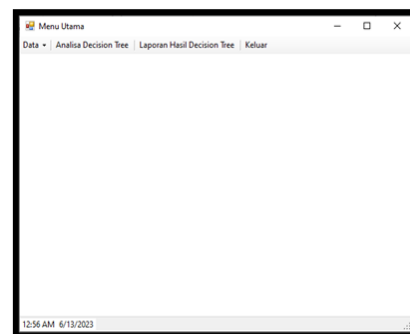
Pembahasan aplikasi merupakan tahap bagaimana aplikasi untuk dijalankan berdasarkan desain yang telah dibuat dan dirancang pada tahap sebelumnya kemudian dimasukkan ke bahasa pemrograman yang digunakan. Pada aplikasi klasifikasi keluarga miskin menggunakan algoritma C4.5 yang telah dibuat terdapat beberapa interface (antarmuka) yaitu:
Login Aplikasi



Gambar 3. Interface Login Aplikasi

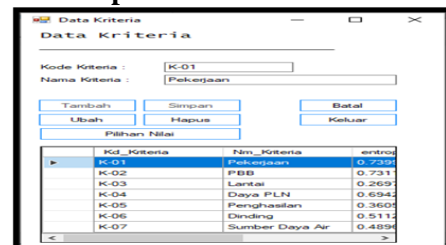
Interface Menu Utama

Menu utama merupakan form menu yang memiliki beberapa sub menu yang dapat diakses oleh admin untuk melakukan pengolahan data pada Aplikasi data mining, yaitu data, analisis decision tree, dan laporan hasil decision tree yang memiliki fungsi yang berbeda-beda. Adapun interface menu utama terlihat pada gambar 4 berikut :



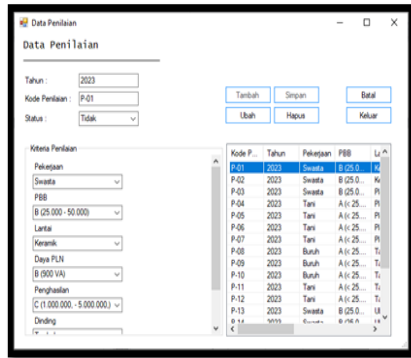
Gambar 4. Interface Menu Utama

Interface Input Data Kriteria



Gambar 5. Interface Input Data Kriteria

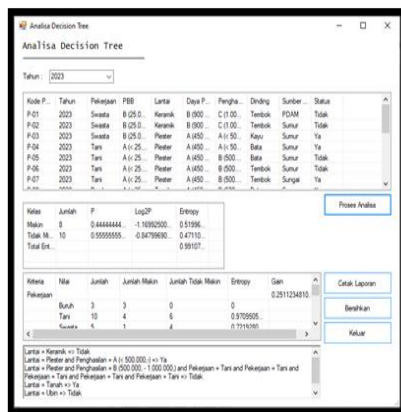
Interface Penilaian



Gambar 6. Interface Penilaian

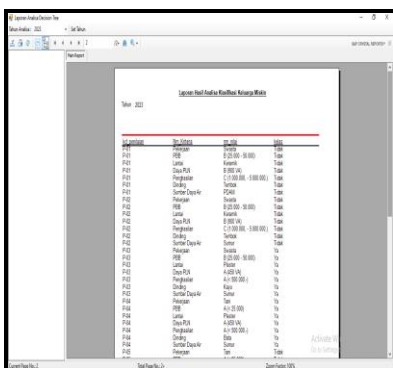
Gambar 6 di atas menampilkan proses penilaian terhadap data yang telah diujicobakan.

Interface Decision Tree Metode C.45



Gambar 7. Interface Decision Tree

Laporan Hasil Decision Tree



Gambar 8. Tampilan Laporan Hasil Decision Tree

SIMPULAN

Klasifikasi keluarga miskin pada aplikasi telah diterapkan Metode Decision Tree dengan melihat hasil nilai Gain dan Entropy pada masing-masing kriteria. Dimana cabang yang digunakan terdiri dari 7 yaitu pekerjaan, Pajak Bumi Bangunan, lantai, dinding, sumber air, daya PLN dan penghasilan. digunakan terdiri dari 7 yaitu pekerjaan, Pajak Bumi Bangunan, lantai, dinding, sumber air, daya PLN dan penghasilan. Berdasarkan hasil pengujian yang telah dilakukan, fungsional dari aplikasi data mining untuk klasifikasi keluarga miskin berjalan dengan baik dan lancar sebagaimana mestinya, dan mampu memberikan informasi hasil analisis keluarga miskin melalui pendekatan decision tree.

DAFTAR PUSTAKA

Bahri, S., & Lubis, A. (2020). Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier League. *Jurnal Sintaksis: Pendidikan Guru Sekolah Dasar, IPA, IPS dan Bahasa Inggris*, 63-70.

Budiarto, E. (2017). Pembuatan Aplikasi Web Berbasis SMS Sebagai Media Penyalur Informasi dan Komunikasi Antara Sekolah Dengan Orang Tua Siswa. *Jurnal Ilmiah Pendidikan Teknik Kejuruan (JIPTEK)*, Vol.X No.1

Eska, J. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5. *Jurnal Teknologi dan Sistem Informasi (Jurteks)* , Vol.2 No.2 ISSN : 2407-1811.

Fajrin, A. A., & Algifanri, M. (2018). Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma FP-Growth Pada Data Transaksi Penjualan Spare Part Motor. *Kumpulan Jurnal Ilmu*

- Komputer (KLIK), Vol.05 No.01. ISSN:2406-7857.
- Firman, A. (2019). Analisis dan Perancangan Sistem Informasi. Surabaya: Penerbit Qiara Media.
- Lasminiasih. (2016). Perancangan Sistem Informasi Kredit Mikro Mahasiswa Berbasis Web. Jurnal Sistem Informasi (JSI) Vol.8 No.1 April 2016 ISSN : 2085-1588.
- Lubis, A. (2016). Basis Data Dasar Untuk Mahasiswa Ilmu Komputer. Yogyakarta: Deepublish.
- Nasution, I., Windarto, A., & M.Fauzan. (2020). Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. Building of Informatics, Technology and Science (BITS), 76-83.
- Pribadi, E., Poningsih, & Tambunan, H. (2020). Analisa Tingkat Kepuasan Masyarakat Terhadap Pelayanan Pengadilan Agama Pematangsiantar Menggunakan Algoritma C4.5. BRAHMANA: Jurnal Penerapan Kecerdasan Buatan, 33-40
- R.H Sianipar. (2017). Visual Basic.Net Untuk Programmer. Yogyakarta: Andi Offset.
- Santoso, & Nurmalina, R. (2017). Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut). Jurnal Integrasi, Vol.9 No.1 April 2017 e-ISSN : 2548-9828.
- Sunia, D., Budi, K., & Jusia, P. (2019). Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means. Jurnal Ilmiah Mahasiswa Teknik Informatika , 121-134.
- Suprpto, U. (2021). Pemodelan Perangkat Lunak (C3) Kompetensi Keahlian : Rekayasa Perangkat Lunak Untuk SMK/MAK Kelas XI. Jakarta: Grasindo.