

IJCIT

(Indonesian Journal on Computer and Information Technology)

Journal Homepage: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>

Penerapan Decision Tree Untuk Klasifikasi Tingkat Pendapatan

Evy Priyanti

Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika

Jakarta, Indonesia

e-mail: evy.evp@bsi.ac.id

ABSTRAK

Semua orang pasti menginginkan pendapatan yang besar. Banyak faktor yang membuat pendapatan setiap orang berbeda dari mulai faktor Usia, jenis pekerjaan, status Pendidikan, pekerjaan, ikatan pernikahan, warna kulit, jenis kelamin, keuntungan, durasi waktu kerja, asal negara. Penelitian ini bertujuan untuk menentukan faktor penentu dalam meraih pendapatan yang besar menggunakan metode decision tree. Decision tree ini memiliki banyak kelebihan diantaranya sangat mudah diimplementasikan dan disajikan dalam bentuk diagram. Dengan mengklasifikasikan faktor-faktor penentu pendapatan menggunakan decision tree maka akan jelas terlihat faktor penentu yang dapat membuat seseorang memiliki pendapatan lebih besar. Penelitian ini menghasilkan data dengan pendapatan terbesar didapatkan pada atribut relasi. Seseorang dengan relasi yang luas dapat memiliki peluang pendapatan yang lebih besar yaitu sebesar 75.12%. Selain itu faktor usia dan lama bekerja juga menjadi faktor terbesar lainnya. Dengan kata lain memperluas jaringan komunikasi dan rekan kerja yang luas akan berdampak positif bagi penambahan nilai pendapatan.

Kata kunci: Pohon Keputusan, Pendapatan, Data Mining

ABSTRACTS

Everyone wants a big income. There are many factors that make everyone's income different, starting from age, type of work, education status, occupation, marriage ties, skin color, gender, benefits, duration of work, country of origin. This research aims to determine the determining factors in achieving a large. This decision tree has many advantages including very easy to implement and presented in the form of a diagram. By classifying income determinants using the decision tree method. It will be clear that the determining factors that can make a person have a greater income. This research resulted in data with the largest income obtained on relationship attributes. A person with extensive relationships can have a greater income opportunity, which is 75.12%, besides age and length of work are also the other biggest factors. In other words, expanding the network of communication and extensive co-workers will have a positive impact on adding value to income.

Keywords: Decision Tree, Income, Data Mining

1. PENDAHULUAN

Pendapatan yang besar menjadi impian setiap orang. Banyak faktor yang menentukan seseorang mendapatkan pendapatan yang besar. Dari banyak faktor, tentunya penting untuk mengetahui faktor terbesar yang menentukan pendapatan. Hal ini penting, agar setiap orang memperhatikan dan mempersiapkan sejak dini faktor penentu ini.

Pada penelitian sebelumnya yang dilakukan di Kabupaten Minahasa, tiga faktor penentu pendapatan rumah tangga digunakan dalam penelitian yaitu faktor Pendidikan, jumlah anggota keluarga dan pengalaman kerja menggunakan regresi dengan pertumbuhan meningkat sebesar 5,66% dan mendapatkan hasil berupa 15,75% datang dari faktor Pendidikan (Jeine Silvane Rungkat, 2020). Namun pada



penelitian ini atribut yang diteliti masih terlalu sedikit sehingga kurang relevan jika digunakan pada wilayah lain diluar kabupaten Minahasa.

Hasil pengumpulan data pendapatan yang didapatkan oleh Ronny Kohavi dan Barry Becker pada sensus Beureu terdiri dari Usia, kelas kerja, pemerintah negara bagian, ekonomi, pendidikan, nomor pendidikan, status perkawinan, pekerjaan, hubungan, ras, jenis kelamin, keuntungan modal, kerugian modal, jam per minggu, dan negara asal adalah beberapa di antaranya. atribut yang dapat ditemukan dalam data (Becker, n.d.).

Beberapa atribut menunjukkan hubungan yang kuat antara pendapatan dan relasi tingkat, dengan implikasi bahwa seseorang dengan banyak relasi akan memiliki lebih banyak peluang untuk meningkatkan pendapatan selama satu tahun. Selain itu juga keterikatan atau pengaruh keluarga juga tidak kalah penting jika dibandingkan dengan relasi, dikatakan bahwa hubungan yang erat dikeluarga akan membuat kehidupan yang lebih baik (Alya Iklima Hidayah, 2021). Seseorang yang memiliki pendidikan yang lebih tinggi memiliki peluang untuk mendapatkan penghasilan yang lebih tinggi. Juga lama bekerja, semakin lama seorang pekerja semakin besar pula penghasilan yang didaparkannya dan ini membuktikan bahwa seseorang dengan pengalaman yang lebih banyak dapat menghasilkan pendapatan yang lebih banyak pula. Pendapatan juga sangat berpengaruh erat dengan pengeluaran dalam rumah tangga, karena kedua hubungan ini menjadikan seseorang berupaya meningkatkan pendapatan dan menekan pengeluaran untuk mencapai kemakmuran (Heny Ratnaningtyas, 2021).

Klasifikasi merupakan algoritma yang menggunakan data dengan target/class/label berupa berupa nominal atau kategorikal (Wahono, 2020). Klasifikasi pendapatan menggunakan algoritma Decision Tree disebut juga dengan pohon keputusan yang merupakan model dari klasifikasi (Mawadatul Maulidah, 2020). Valentin Mihov dalam penelitiannya yang berjudul Adult Income Dataset Analysis using IPython pada tahun 2015 memisahkan antara data yang berhubungan dengan data yang tidak berhubungan dengan Learning Model Logistic regression. Topiwalla juga melakukan penelitian pada tahun 2013 dengan judul Machine Learning pada kumpulan data UCI Dewasa menggunakan beberapa algoritma klasifikasi dan meningkatkan akurasi menggunakan peningkatan gradien yang parah.

Decision tree merupakan metode klasifikasi yang sangat populer saat ini dimana berbagai klasifikasi data dapat dilakukan seperti studi mesin, penyakit medis, pengklasifikasian teks dan gambar serta berbagai macam bidang analisis lainnya (Bahzad Taha Jijo, 2021).

Decision tree merupakan salah satu algoritma yang dapat mengklasifikasikan data. Beberapa kelebihan dari decision tree adalah mudah dan visualisasi sehingga mempermudah dalam penyajian data (Gupta, 2017).

Konsep data dalam decision tree adalah :

- a. Data disimpan dalam format tabel dengan atribut dan catatan.
- b. Atribut mengacu pada parameter yang digunakan sebagai kriteria untuk membangun sebuah pohon. Misalnya, saat menentukan pemain tenis utama, perhatikan cuaca, angin, dan suhu. Atribut pertama adalah atribut yang berisi data per item yang dirujuk oleh atribut target.
- c. Atribut memiliki satu set properti yang diberi nama setelah instance. Decision tree yang dievaluasi memungkinkan interaksi dengan model grafis yang kuat dan mampu mewakili struktur label yang kompleks.

Decision tree dapat dilatih secara efisien dan Bersama-sama menggunakan fungsi yang memungkinkan parameter menunjukkan secara eksperimental struktur label yang kompleks dan menghasilkan model yang terbaik.

Decision tree juga digunakan untuk menganalisis apa saja penyebab kesalahan yang terjadi dan pentingnya variabel yang terlibat dalam proses supaya hasil yang didapatkan bisa lebih maksimal (Giuseppina Pappalardo, 2021). Karakteristik negatif meliputi tidak pernah menikah, memiliki anak, pembantu rumah tangga, dan perceraian, sedangkan karakteristik positif meliputi capital gain, menikah-sipil-pasangan, usia, jam kerja per minggu, dan eksekutif-manajerial. Terakhir, karakteristik negatif akan diabaikan agar proses analisis dapat berjalan lebih lancar.

Operator Decision Tree adalah operator bersarang yang memiliki subproses. Subproses harus memiliki pengajaran yang diharapkan menghasilkan model pohon. Representasi data dalam Decision Tree memiliki keunggulan dibandingkan dengan pendekatan lain yang bermakna dan mudah diinterpretasikan. Tujuannya adalah untuk membuat model klasifikasi yang memprediksi nilai label berdasarkan beberapa atribut input. Setiap pohon simpul interior pohon dengan salah satu

atribut input yang sesuai. Setiap simpul daun mewakili nilai label yang diberikan nilai atribut input yang diwakili oleh jalur dari akar ke daun. Deskripsi ini dapat dengan mudah dipahami dengan mempelajari Contoh Proses dari operator Pohon Keputusan (RapidMiner, 2022).

Decision Tree pada RapidMiner terbagi menjadi 2 bagian yaitu bagian input dan Output. Untuk input adalah data latih yang berbentuk table data, data inilah yang nantinya akan mengirimkan output dan sebagai operator pada proses nantinya. Sedangkan Output terdiri dari model yang akan menghasilkan pohon keputusan dan klasifikasi dari data yang diolah (RapidMiner, 2022).

Pada penelitian, peneliti bermaksud menggunakan pohon keputusan dimana Pohon keputusan dapat menggabungkan eksplorasi data dan pemodelan, yang merupakan Langkah pertama yang sangat baik dalam proses pemodelan, bahkan ketika digunakan sebagai model akhir untuk teknik lainnya (Rizmayanti, 2021).

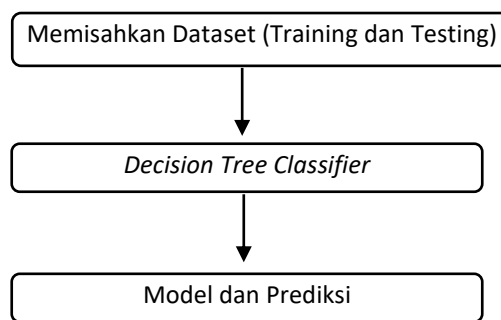
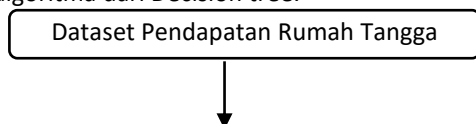
2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder dengan metode penelitian survei oleh Becker pada sensus Beureu (Becker, n.d.), yang menghasilkan beberapa data pendapatan untuk beberapa negara di dunia dengan besaran pendapatan rumah tangga yang lebih besar (Becker, n.d.).

faktor Usia, jenis pekerjaan, status Pendidikan, pekerjaan, ikatan pernikahan, warna kulit, jenis kelamin, keuntungan, durasi waktu kerja, asal negara adalah beberapa di antaranya. atribut yang mungkin ditemukan dalam data. Kumpulan data yang telah dikumpulkan akan diuji dengan algoritma pohon keputusan dimana pengaruh terbesar dalam pendapatan akan didapatkan.

Pohon keputusan berguna untuk eksplorasi data karena mereka mengungkap koneksi tersembunyi antara sejumlah besar variabel input dan variabel tujuan tunggal.

Pohon keputusan ini akan membentuk hirarki yang akan memperlihatkan atribut mana yang akan mengklasifikasikan berdasarkan struktur tertinggi sebagai model prediksi dalam pengambilan keputusan. Gambar 1 menunjukkan algoritma dari Decision tree.



Gambar 1. Langkah-Langkah Penelitian

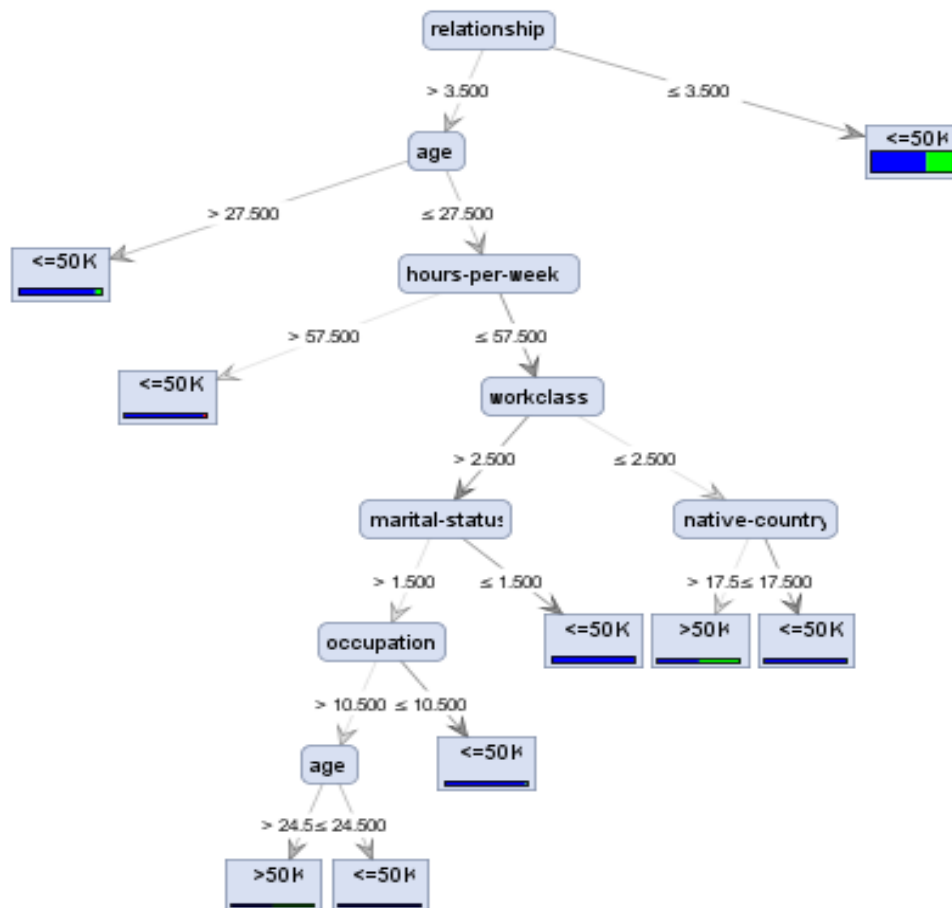
3. HASIL DAN PEMBAHASAN

Berdasarkan dataset yang didapatkan dari sensus Beureu, atribut yang diteliti terdiri dari:

- Usia: terdiri dari 17 tahun hingga 90 tahun.
- Jenis Pekerja: pegawai swasta, wiraswasta yang bukan dari perusahaan, pemerintah federal, pemerintah daerah, pemerintah pusat, pekerja tanpa bayaran, dan tidak pernah bekerja.
- Status Pendidikan: HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th.
- Status Pernikahan terdiri dari: Menikah-Civ-spouse=menikah dengan warga sipil, Bercerai=bercerai, Belum menikah=punya pasangan tapi belum resmi menikah, Terpisah=terpisah, Janda=janda atau duda yang belum menikah, Married-spouse-absent=pasangan ideal, Married-AF-spouse=menikah dengan angkatan, Keluarga juga terdiri dari dua bagian, keluarga lengkap dan keluarga tidak lengkap.
- Pekerjaan: teknisi, pengrajin, Layanan lainnya, Bagian Penjualan, Manajer eksekutif, Spesialis, Bagian Kebersihan, Inspeksi pengoperasian mesin, Administrasi administrasi, Pelaut, Pemandahan transportasi, Layanan rumah pribadi, Layanan pelindung, Angkatan Bersenjata
- Ikatan: istri, anak kandung, suami, bukan bagian dari keluarga inti, relasi lainnya, tidak menikah
- Warna Kulit: Putih, Asia-Pasifik-Pulau, Amerika-India-Eskimo, Lainnya, dan Hitam adalah kelompok ras.
- jenis kelamin: wanita, pria.
- Keuntungan yang biasa didapatkan dalam uang bentuk: bernilai dari 1 hingga 99.999.
- durasi waktu kerja: Jam per minggu bernilai antara 1 dan 99.

k. Asal negara: Amerika Serikat, Kamboja, Inggris, Puerto Riko, Kanada, Jerman, Outlying-US (Guam-USVI-etc), India, Jepang, Yunani, Selatan, Cina, Kuba, Iran, Honduras, Filipina, Italia, Polandia, Jamaika, Vietnam, Meksiko, Portugal, Irlandia, Prancis, Republik Dominika, Laos, Ekuador, Taiwan, Haiti,

Kolombia, Hongaria, Guatemala, Nikaragua, Skotlandia, Thailand, Yugoslavia, El Salvador, Trinidad&Tobago, Peru,
 Hasil yang didapatkan dengan menggunakan Decision Tree pada klasifikasi Pendapatan dapat dilihat pada gambar 1.



Gambar 1. Decision Tree Pendapatan

Pada Gambar 1 jelas terlihat bahwa relationship memiliki nilai tertinggi dimana membuktikan bahwa seseorang dengan jumlah relasi yang luas akan mendapatkan peluang pendapatan yang lebih besar pula selain itu usia juga sangat menentukan besarnya pendapatan seseorang.

Pada gambar 2 Jelas menunjukkan bahwa Decision Tree dapat mengklasifikasikan tingkat

pendapatan berdasarkan 15 atribut yang ada, dan menghasilkan nilai akurasi sebesar 75.12%. hal ini sangatlah penting dimana Ketika sebuah keputusan diambil haruslah mempertimbangkan banyak factor dan resiko yang akan dihadapi. Dengan kematangan dan keputusan yang tepat maka setiap Langkah yang diambil akan meminimalisir resiko atau kerugian dikemudian hari.

Table View Plot View

accuracy: 75.12% +/- 1.62% (mikro: 75.12%)

	true <=50K	true >50K	true >=50	class precision
pred. <=50K	4843	1600	1	75.16%
pred. >50K	3	1	0	25.00%
pred. >=50	0	0	0	0.00%
class recall	99.94%	0.06%	0.00%	

Gambar 2. Nilai akurasi

Gambar 3 menunjukkan performance vector dimana nilai akurasi dan confusion matrik dengan nilai akurasi sebesar 75.12% dan kappa sebesar -0.000+/-0.0003. Bentuk dari node pada

Decision Tree jelas terlihat pada gambar 4 dimana setiap atribut terlihat keterikatan pada setiap node.

PerformanceVector

PerformanceVector:
 accuracy: 75.12% +/- 1.62% (mikro: 75.12%)
 ConfusionMatrix:
 True: <=50K >50K >=50
 <=50K: 4843 1600 1
 >50K: 3 1 0
 >=50: 0 0 0
 kappa: -0.000 +/- 0.003 (mikro: 0.000)
 ConfusionMatrix:
 True: <=50K >50K >=50
 <=50K: 4843 1600 1
 >50K: 3 1 0
 >=50: 0 0 0

Gambar 3. Performance Vector

Tree

```

relationship > 3.500
| age > 27.500: <=50K { <=50K=876, >50K=62, >=50=0}
| age <= 27.500
| | hours-per-week > 57.500: <=50K { <=50K=19, >50K=0, >=50=1}
| | hours-per-week <= 57.500
| | | workclass > 2.500
| | | | marital-status > 1.500
| | | | | occupation > 10.500
| | | | | age > 24.500: >50K { <=50K=1, >50K=1, >=50=0}
| | | | | age <= 24.500: <=50K { <=50K=2, >50K=0, >=50=0}
| | | | | occupation <= 10.500: <=50K { <=50K=58, >50K=1, >=50=0}
| | | | | marital-status <= 1.500: <=50K { <=50K=742, >50K=0, >=50=0}
| | | | workclass <= 2.500
| | | | | native-country > 17.500: >50K { <=50K=1, >50K=1, >=50=0}
| | | | | native-country <= 17.500: <=50K { <=50K=51, >50K=1, >=50=0}
relationship <= 3.500: <=50K { <=50K=3096, >50K=1535, >=50=0}
    
```

Gambar 4. Decision Tree dengan node

4. KESIMPULAN

Berdasarkan hasil klasifikasi dataset pendapatan yang didapatkan dengan menggunakan Decision Tree atau pohon keputusan, maka dihasilkan nilai akurasi sebesar 75.12% dimana faktor terbesar berada pada relasi dan dilanjutkan dengan usia dan lama bekerja. Dengan demikian semakin luas relasi akan semakin besar pula peluang untuk bisa menambah pendapatan, begitu pula dengan usia dan lama bekerja, dimana seseorang dengan keahlian khusus karena sudah lama menggeluti suatu bidang akan mendapatkan penghasilan yang berbeda dibandingkan dengan seseorang yang baru terjun dibidang tersebut. Untuk penelitian selanjutnya perlu dilakukan pengembangan supaya nilai akurasi dan pola yang lebih jelas terlihat.

5. REFERENSI

- Alya Iklima Hidayah, I. S. (2021). Pengaruh Relasi dalam Keluarga terhadap Subjective Well-Being Anak dan Remaja di Masa Pandemi Covid 19. *Prosiding Psikologi*, 272-277.
- Bahzad Taha Jijo, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal Of Applied Science And Technology Trends*, 20-28.
- Becker, R. K. (n.d.). *UCI Dataset*. Retrieved from UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/adult>
- Giuseppina Pappalardo, A. D. (2021). Decision Tree Method to Analyze the Performance of Lane Support Systems. *Sustainability*, 1-13.
- Gupta, P. (2017, May 18). *Decision Trees in Machine Learning*. Retrieved from <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>
- Heny Ratnaningtyas, N. A. (2021). Pengaruh Pendapatan Rumah Tangga dan Pengeluaran Rumah Tangga Terhadap Stabilitas Keuangan Rumah Tangga Pada Pelaku Wirausaha Di Objek Wisata Danau Cipondoh. *Jurnal Ekonomi Pembangunan*, 35-45.
- Jeine Silvane Rungkat, P. K. (2020). Pengaruh Pendidikan, Jumlah Anggota Keluarga Dan Pengalaman Kerja Terhadap Pendapatan Rumah Tangga Di Kabupaten Minahasa. *Jurnal Pembangunan Ekonomi dan Keuangan Daerah*, 1-15.
- Mawadatul Maulidah, W. G. (2020). Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku. *Jurnal Ilmiah Ekonomi Dan Bisnis*, 89.
- RapidMiner. (2022). *Decision Tree*. Retrieved from RapidMiner Studio: https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/decision_tree_multiway.html
- Rizmayanti. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus SMK Multicomp Depok). *Swabumi*, 9-18.
- Wahono, R. S. (2020). *RSW*. Retrieved from Data Mining: <https://romisatriawahono.net/lecture/dm/romi-dm-aug2020.pdf>