Jurnal Riset Informatika dan Teknologi Informasi (JRITI)
Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)
Vol. 3, No. 2, Desember 2025 – Maret 2026, Hal. 299 - 305
Web : https://ejurnal.jejaringppm.org/index.php/jriti

ISSN (Print) : 3024-8167
E-ISSN : 3109-8959

Prediksi Pendapatan di Atas $50.000 Berdasarkan Pendidikan Terakhir, Status Pernikahan, dan Profesi Pria di Amerika Serikat Menggunakan Metode Decision Tree

Rahmat Santoso[1]
Universitas Serang Raya[1]
Email : rahmatsantoso11222112@gmail.com

ABSTRAKSI
Pendapatan merupakan salah satu aspek penting kehidupan manusia. Pendapatan yang cukup dapat mempengaruhi kehidupan seseorang. Salah satu cara untuk mengetahui pendapatan seseorang adalah dengan melakukan Analisa terhadap faktor-faktor pendukung pendapatan tersebut. Individu dan keluarga dengan pendapatan yang lebih tinggi memiliki akses ke pendidikan, perawatan, kesehatan, dan kualitas hidup yang lebih baik. Dengan prediksi tersebut, pemerintah juga dapat diuntungkan untuk memberikan pajak ke orang yang tepat. Decision tree dapat dengan mudah menangani data yang terdiri dari variabel yang terdiri dari beberapa kategori seperti pendidikan terakhir, status pernikahan, profesi, dan jenis kelamin seseorang. Data yang digunakan pada penelitian kali ini diperoleh dari UC Irvine Machine Learning Repository Dataset yang bernama Adult. Dataset tersebut diekstrak oleh Barry Becker dari basis data sensus pada 1994. Dataset ini ditangani dengan perhitungan manual dengan mencari nilai tiap Entropy dan menggunakan aplikasi Orange Mining untuk melakukan verifikasi terhadap perhitungan manual, serta penggambaran Decision Tree yang tepat atau tidak.

Kata Kunci: *Decision Tree, Orange Mining, Pendapatan, Prediksi*

ABSTRACT
Income is one of the important aspects of human life. Sufficient income can affect a person's life. One way to determine a person's income is to analyze the factors that support this income. Individuals and families with higher incomes have access to better education, care, health, and quality of life. With such predictions, the government can also benefit from giving taxes to the right people. Decision trees can easily handle data consisting of variables consisting of several categories such as the last education, marital status, profession, and gender of a person. The data used in this research was obtained from the UC Irvine Machine Learning Repository Dataset called Adult. The dataset was extracted by Barry Becker from the census database in 1994. This dataset is handled by manual calculation by finding the value of each Entropy and using the Orange Mining application to verify the manual calculation, as well as the depiction of the correct Decision Tree or not.

Keywords: *Decision Tree, Orange Mining, Income, Predict*

Penulis Korespondensi
Rahmat Santoso

Tanggal Submit : 31/01/2025
Tanggal Diterima : 30/11/2025
Tanggal Terbit : 23/12/2025

I. PENDAHULUAN

Pendapatan merupakan salah satu aspek penting yang dimiliki oleh masyarakat. Pendapatan dapat menggambarkan bagaimana kualitas hidup seseorang. Jika pendapatan orang tersebut lebih dari $50.000 maka kualitas hidupnya tergolong baik. Jika pendapatan kurang dari $50.000, terdapat beberapa faktor yang mendukung kualitas hidupnya baik atau tidak.

Beberapa penelitian sebelumnya yang melakukan penelitian terhadap pendidikan terakhir dan pendapatan telah menunjukan korelasi yang kuat antara pendidikan yang semakin tinggi maka pendapatan juga akan meningkat. Penelitian lain juga memberikan

kesimpulan bahwa faktor-faktor pendukung pendapatan seperti pengalaman kerja di sebuah industri juga memainkan peran penting dalam menentukan tingkat pendapatan. Dari beberapa jurnal yang ada tentang topik ini, kita dapat mengembangkan temuan-temuan sebelumnya yang melakukan penyempurnaan pemahaman tentang hubungan yang kompleks antara pendidikan terakhir, profesi, dan pendapatan seseorang.

Profesi merupakan salah satu penunjang tinggi atau rendahnya pendapatan seseorang. Dengan profesi tertentu, seseorang dapat menghasilkan pendapatan yang lebih banyak dibanding profesi yang lain. Sebagai contoh, seorang manajer disebuah perusahaan swasta akan mendapatkan penghasilan yang lebih banyak dibanding karyawan biasa seperti operator produksi. Maka dari itu, hubungan antara profesi dan pendapatan seseorang dapat dibilang cukup kuat.

II. PENELITIAN YANG TERKAIT

Salah satu penelitian yang terkait dengan topik penelitian kali ini oleh [1] yang melakukan Analisis Faktor Tingkat Pendidikan, Jenis Kelamin, dan Status Perkawinan Terhadap Pendapatan di Indonesia Berdasarkan IFLS-5.

Pada penelitiannya, mereka menggunakan Tingkat Pendidikan, Jenis Kelamin, dan Status Perkawinan yang menyatakan berpengaruh signifikan terhadap pendapatan. Metode yang digunakan pada penelitian tersebut adalah Regresi Linear Berganda dengan Tingkat Pendidikan sebagai X_1, Jenis Kelamin sebagai X_2, dan Tingkat Pendapatan sebagai Y.

Berbeda dengan penelitian kali ini, peneliti menggunakan metode Decision Tree karena tiap variabel memiliki beberapa kategori bukan nilai angka. Peneliti juga menambahkan satu variabel baru yaitu Status Pernikahan dan penelitian kali ini dilakukan berdasarkan dataset yang ada dari sensus penduduk Amerika Serikat pada tahun 1994.

III. METODE PENELITIAN

Metode yang digunakan pada penelitian kali ini adalah metode kuatitatif dengan pendekatan Decision Tree. Pohon keputusan (Decision Tree) adalah salah satu metode yang cukup mudah untuk diinterpretasikan oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan Keputusan [2].

Rancangan penelitian ini terdiri dari beberapa tahapan. yaitu :

1. Pengumpulan data.

   Data yang digunakan pada penelitian kali ini adalah data yang diambil dari dataset Adult pada UC Irvine Machine Learning Repositories. Data yang terkumpul dalam dataset ini sebanyak 32.561 data Dimana data tersebut merupakah data sensus penduduk di Amerika Serikat yang diambil pada tahun 1994 oleh Barry Becker. Data tersebut berisi atribut age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, dan native-country.

2. Pre-Processing Data

   Pre-Processing Data yaitu proses membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan-kesalahan pada data. Selain itu pada proses ini juga dilakukan enrichment yaitu proses memperkaya data yang sudah ada dengan data lain yang relevan [3].

   Untuk mendapatkan data yang berkualitas, dilakukan beberapa teknik pre-processing data :

   a. Data Validation

      Dilakukan untuk mengidentifikasi dan menghapus data yang ganjil (outlier/noise) data yang tidak konsisten, dan data yang tidak lengkap (missing value) [4]. Dataset adult akan di identifikasi apakah ada data yang ganjil, data yang tidak konsisten dan data yang tidak lengkap.

   b. Data Cleaning

      Proses cleaning ini mencakup antara lain; membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data [5]. Dari 111 data yang dikumpulkan akan di Analisa apakah ada data yang tidak konsisten atau tidak relevan, sehingga akan mengganggu pola aturan dari algoritma yang akan dibentuk.

3. Penerapan Metode

   Decision Tree adalah sebuah struktur pohon, Dimana setiap simpul (node) pohon merepresentasikan atribut yang telah diuji. Setiap cabang merupakan pembagian hasil uji dan node daun (leaf) merepresentasikan kelompok kelas tertentu. Level node teratas dari sebuah decision tree adalah akar (root) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu. Pada umumnya, decision tree melakukan strategi pencarian secara top-down untuk solusinya. Pada proses klasifikasi, nilai atribut akan diuji dengan cara mempelajari jalur

dari node akar (root) sampai ke node akhir (leaf) baru kemudian kelas baru akan ditentukan [6].

Untuk pencarian tiap Quarter, rumus yang digunakan adalah :

$$Qn = -\frac{An}{Tn} \times log_2\left(\frac{An}{Tn}\right) - \frac{Bn}{Tn} \times log_2\left(\frac{Bn}{Tn}\right)$$

Qn = Quarter ke-n
An = Jumlah A
Bn = Jumlah B
Tn = Jumlah A + Jumlah B

Sedangkan untuk Entropi, rumus yang digunakan adalah :

$$Entropi = \frac{T_1}{TD} \times Q_1 + \cdots + \frac{Tn}{TD} \times Qn$$

T_1 = Total Data pada Quarter 1
TD = Total Data
Q_1 = Quarter 1
Tn = Total Data ke-n
Qn = Quarter ke-n

4. Evaluasi Hasil

Setelah hasil klasifikasi diperoleh, selanjutnya hasil di evaluasi menggunakan Cross Validation (confusion matrix) untuk melihat akurasi, presisi dan recall yang dihasilkan oleh model yang diusulkan.

IV. HASIL DAN PEMBAHASAN

Dimana, metode ini dapat dengan mudahnya memprediksi sesuatu berdasarkan variabel yang memiliki banyak kategori didalamnya, seperti pada contoh dataset Adult yang diambil dari UC Irvine Machine Learning Repositories yang dimana variabel Education yang memiliki 7 kategori didalamnya, namun pada penelitian kali ini hanya menggunakan 3 kategori yaitu Masters, Some-college, dan Prof-school.

Variabel yang lain seperti Marital Status yang memiliki 4 kategori yaitu Never-Married, Married-Civ-Spouse, Divorced, dan Widowed. Untuk variabel terakhir yaitu Relationship yang berisi Own-Child, Husband, Not-In-Family, Other-Relative, dan Unmarried. Income yang dihasilkan hanya lebih atau kurang dari $50.000 dari 111 data sensus pria yang ada di Amerika Serikat. Beberapa data tersebut seperti diperlihatkan pada Tabel 1.

Setelah Pre-Processing tersebut, proses selanjutnya pada metode Decision Tree adalah menentukan Entropi dari tiap variabel. Untuk memudahkan dalam mencari data, seseorang yang memikiki Income lebih dari $50.000 diwakilkan dengan angka 1, dan untuk Income yang kurang dari $50.000 diwakilkan dengan angka 0. Untuk entropi pertama yaitu Relationship dan Income dan didapatkan bahwa Jumlah Data dari Entropi Pertama diperlihatkan pada Tabel 2.

Tabel 1. Dataset Setelah Proses Pre-Processing Data

No.	Education	Marital Status	Relationship	Income
1.	Masters	Never-married	Own-child	<=50K
2.	Masters	Married-civ-spouse	Husband	<=50K
3.	Masters	Never-married	Not-in-family	<=50K
4.	Masters	Married-civ-spouse	Husband	>50K
5.	Masters	Married-civ-spouse	Husband	>50K
…	…	…	…	…
108.	Masters	Married-civ-spouse	Husband	>50K
108.	Masters	Married-civ-spouse	Husband	>50K
109.	Masters	Married-civ-spouse	Husband	>50K
110.	Masters	Married-civ-spouse	Husband	>50K
111.	Masters	Married-civ-spouse	Husband	>50K

Tabel 2. Etropi Pertama, Relationship dan Income

Relationship	Income	Jumlah
Husband	0	4
Husband	1	78
Not-in-family	0	17
Not-in-family	1	1
Own-child	0	5
Own-child	1	0
Other-relative	0	3
Other-relative	1	0
Unmarried	0	2
Unmarried	1	1
Total		111

Dari Tabel 2, didapatkan perhitungan untuk mencari Entropi Pertama berdasarkan tiap Relationship dan Income, yaitu :

$$Q1 = -\frac{4}{82} \times log_2\left(\frac{4}{82}\right) - \frac{78}{82} \times log_2\left(\frac{78}{82}\right) = 0.28119$$

$$Q2 = -\frac{17}{18} \times log_2\left(\frac{17}{18}\right) - \frac{1}{18} \times log_2\left(\frac{1}{18}\right) = 0.30954$$

$$Q3 = -\frac{5}{5} \times log_2\left(\frac{5}{5}\right) - \frac{0}{5} \times log_2\left(\frac{0}{5}\right) = 0$$

$$Q4 = -\frac{3}{3} \times log_2\left(\frac{3}{3}\right) - \frac{0}{3} \times log_2\left(\frac{0}{3}\right) = 0$$

$$Q5 = -\frac{2}{3} \times log_2\left(\frac{2}{3}\right) - \frac{1}{3} \times log_2\left(\frac{1}{3}\right) = 0.9183$$

Setelah didapatkan nilai tiap Quarter, maka Entropinya adalah :

$$\frac{82}{111} \times 0.28119 + \frac{18}{111} \times 0.30954 + \frac{5}{111} \times 0 + \frac{3}{111} \times 0 + \frac{3}{111} \times 0.9183$$

$$= 0.2827$$

Setelah didapatkan Entropinya, peneliti akan menentukan Leaf Node untuk Relationship kategori Unmarried, list data Relationship dengan kategori Unmarried seperti pada Tabel 3.

Tabel 3. Penentuan Leaf Node Relationship Kategori Unmarried

Education	Marital Status	Income
Masters	Divorced	0
Masters	Widowed	1
Some-college	Widowed	0

Penentuan Leaf Node untuk Relationship kategori Not-in-family dengan beberapa data sebagai berikut (lihat Tabel 4).

Tabel IV. Penentuan Leaf Node Relationship kategori Not-in-family

Education	Marital Status	Income
Masters	Never-married	0
Masters	Never-married	0
Masters	Never-married	0
Masters	Divorced	0
Masters	Never-married	0
…	…	…
Masters	Never-married	0
Masters	Never-married	0
Masters	Never-married	0
Masters	Never-married	0
Masters	Divorced	0

Dari Tabel 4 diatas, kita mencari nilai terkecil dari Entropi Education dan Marital Status. Dan setelah dilakukan perhitungan secara manual, didapatkan Education memiliki Entropi terkecil dengan nilai 0 dengan perhitungan seperti berikut :

$$Q1 = -\frac{16}{16} \times log_2\left(\frac{16}{16}\right) - \frac{0}{16} \times log_2\left(\frac{0}{16}\right) = 0$$

$$Q2 = -\frac{0}{1} \times log_2\left(\frac{0}{1}\right) - \frac{1}{1} \times log_2\left(\frac{1}{1}\right) = 0$$

$$Q3 = -\frac{1}{1} \times log_2\left(\frac{1}{1}\right) - \frac{0}{1} \times log_2\left(\frac{0}{1}\right) = 0$$

$$Entropi = \frac{16}{18} \times 0 + \frac{1}{18} \times 0 + \frac{1}{18} \times 0 = 0$$

Untuk Leaf Node terakhir dari kategori Relationship ada pada Husband dengan data sebagai berikut (lihat Tabel 5).

Tabel V. Penentuan Leaf Node Relationship kategori Husband

Education	Marital Status	Income
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1
Some-college	Married-civ-spouse	0
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1
…	…	…
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1
Masters	Married-civ-spouse	1

Dari Tabel 5 diatas, kita mencari nilai terkecil dari Entropi Education dan Marital Status. Dan setelah dilakukan perhitungan secara manual, didapatkan Education memiliki Entropi terkecil dengan nilai 0 dengan perhitungan seperti berikut :

$$Q1 = -\frac{0}{15} \times log_2\left(\frac{0}{15}\right) - \frac{15}{15} \times log_2\left(\frac{15}{15}\right) = 0$$

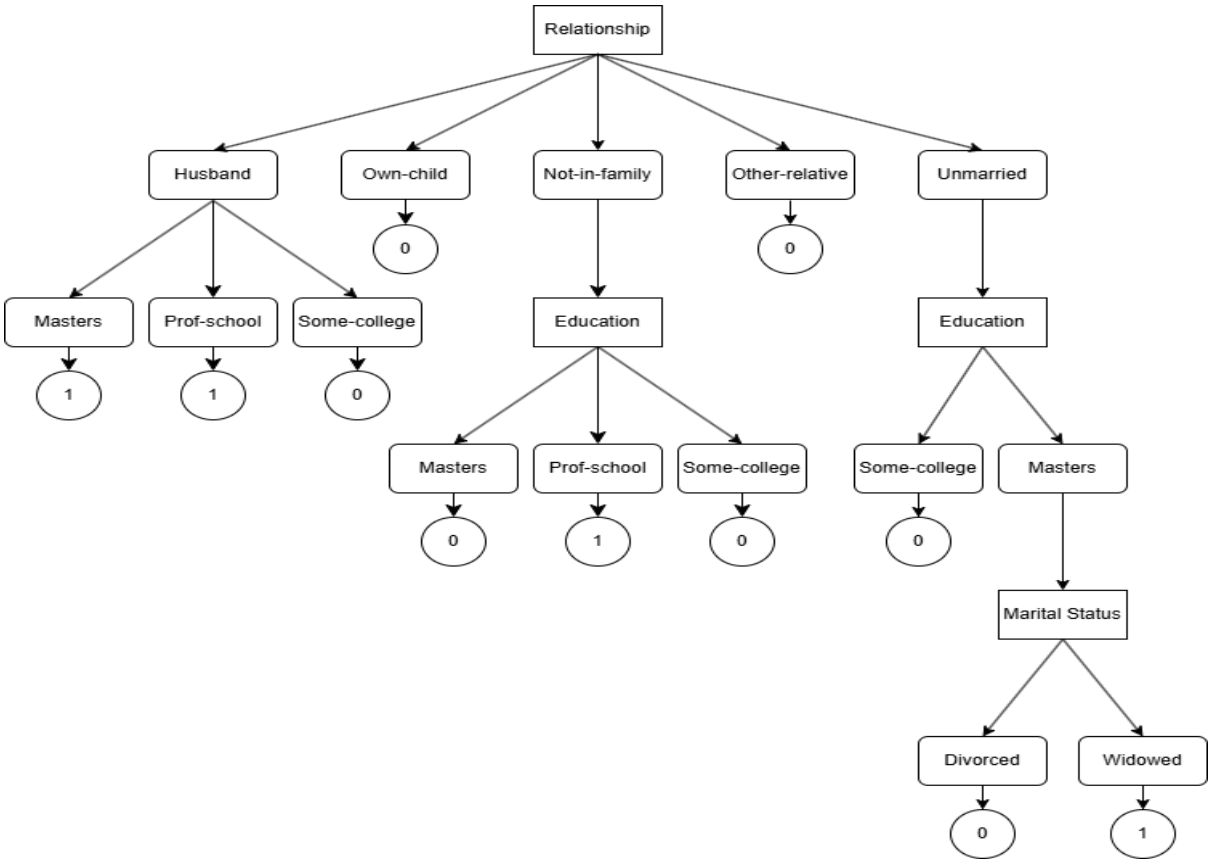
$$Q2 = -\frac{0}{1} \times log_2\left(\frac{0}{1}\right) - \frac{1}{1} \times log_2\left(\frac{1}{1}\right) = 0$$

$$Q3 = -\frac{2}{2} \times log_2\left(\frac{2}{2}\right) - \frac{0}{2} \times log_2\left(\frac{0}{2}\right) = 0$$

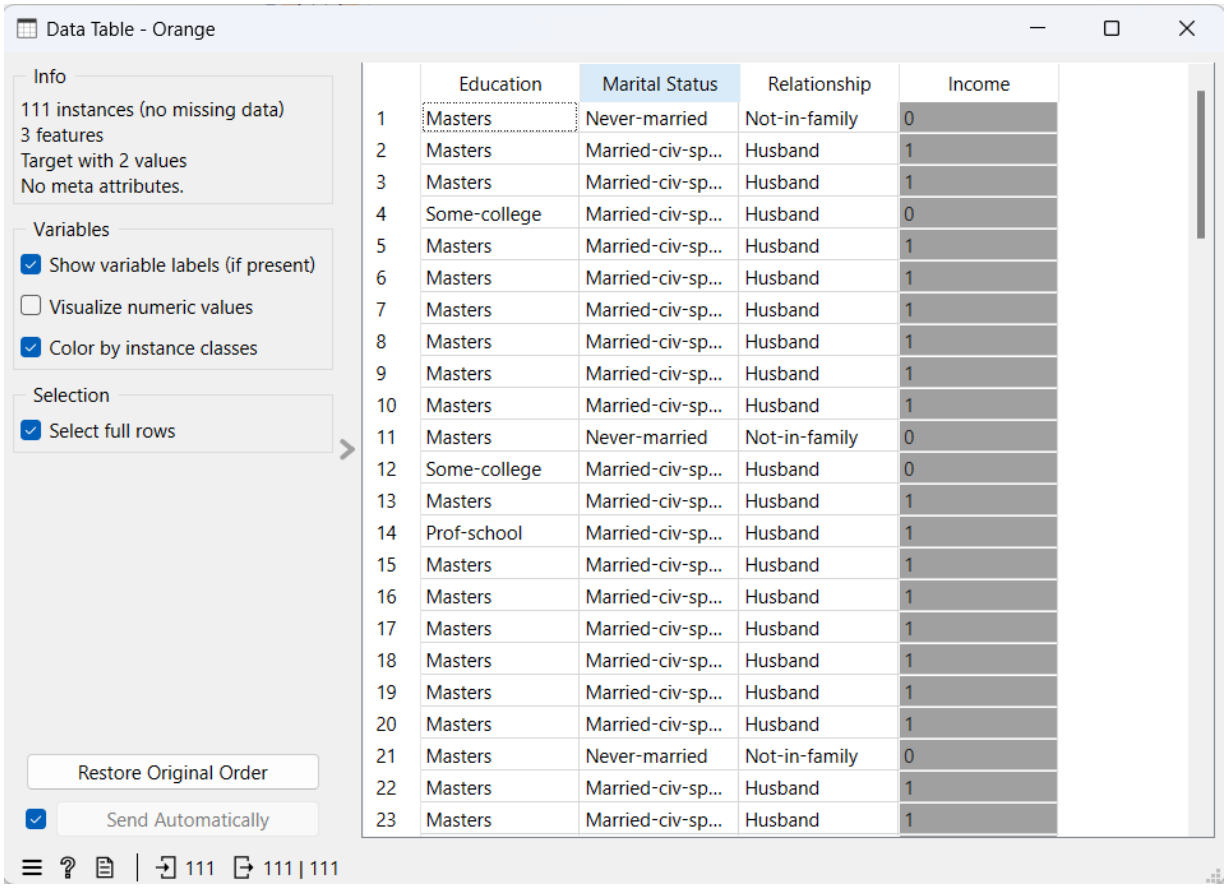
Dari 3 quarter diatas, maka Entropi yang dihasilkan dari Leaf Node Relationship kategori Husband adalah :

$$Entropi = \frac{15}{15} \times 0 + \frac{1}{1} \times 0 + \frac{2}{2} \times 0 = 0$$

Maka, Decision Tree yang dihasilkan dari perhitungan manual seperti diperlihatkan pada Gambar 1. Setelah didapatkan Decision Tree dengan perhitungan manual, kemudian dilakukan evaluasi menggunakan aplikasi Orange Mining. Langkah pertama yang dilakukan adalah memilih file yang akan digunakan pada aplikasi ini. Berikut adalah data yang tampil dengan menggunakan fitur Data Table pada Aplikasi Orange Mining.

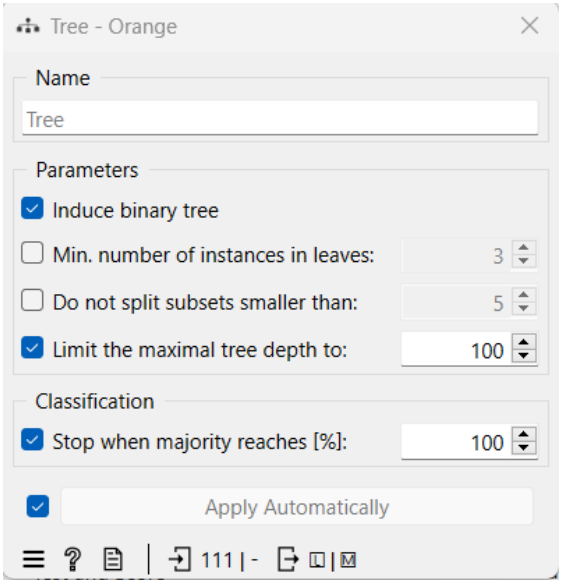


Gambar 1. Decision Tree

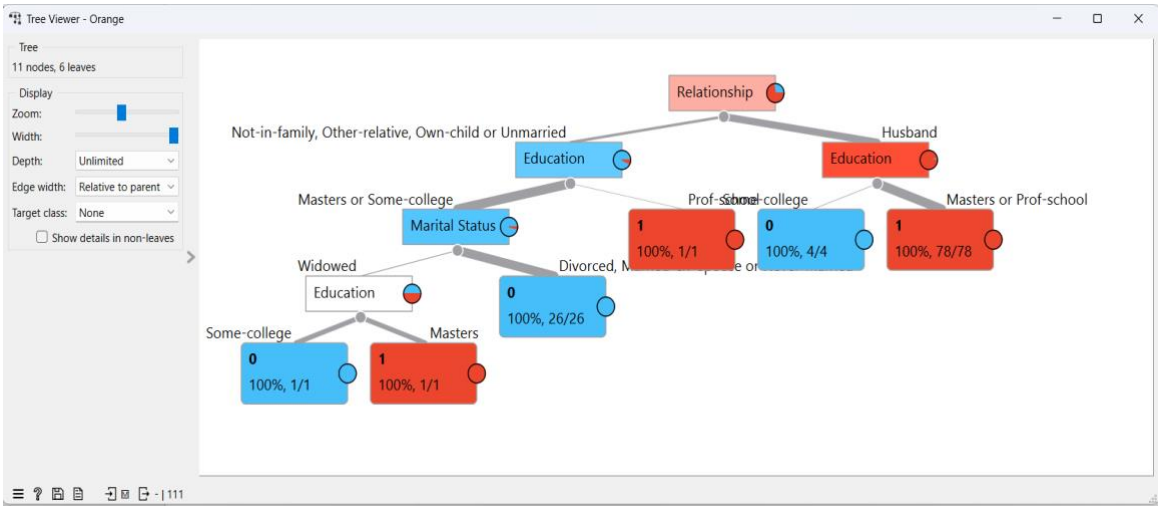


	Education	Marital Status	Relationship	Income
1	Masters	Never-married	Not-in-family	0
2	Masters	Married-civ-sp...	Husband	1
3	Masters	Married-civ-sp...	Husband	1
4	Some-college	Married-civ-sp...	Husband	0
5	Masters	Married-civ-sp...	Husband	1
6	Masters	Married-civ-sp...	Husband	1
7	Masters	Married-civ-sp...	Husband	1
8	Masters	Married-civ-sp...	Husband	1
9	Masters	Married-civ-sp...	Husband	1
10	Masters	Married-civ-sp...	Husband	1
11	Masters	Never-married	Not-in-family	0
12	Some-college	Married-civ-sp...	Husband	0
13	Masters	Married-civ-sp...	Husband	1
14	Prof-school	Married-civ-sp...	Husband	1
15	Masters	Married-civ-sp...	Husband	1
16	Masters	Married-civ-sp...	Husband	1
17	Masters	Married-civ-sp...	Husband	1
18	Masters	Married-civ-sp...	Husband	1
19	Masters	Married-civ-sp...	Husband	1
20	Masters	Married-civ-sp...	Husband	1
21	Masters	Never-married	Not-in-family	0
22	Masters	Married-civ-sp...	Husband	1
23	Masters	Married-civ-sp...	Husband	1

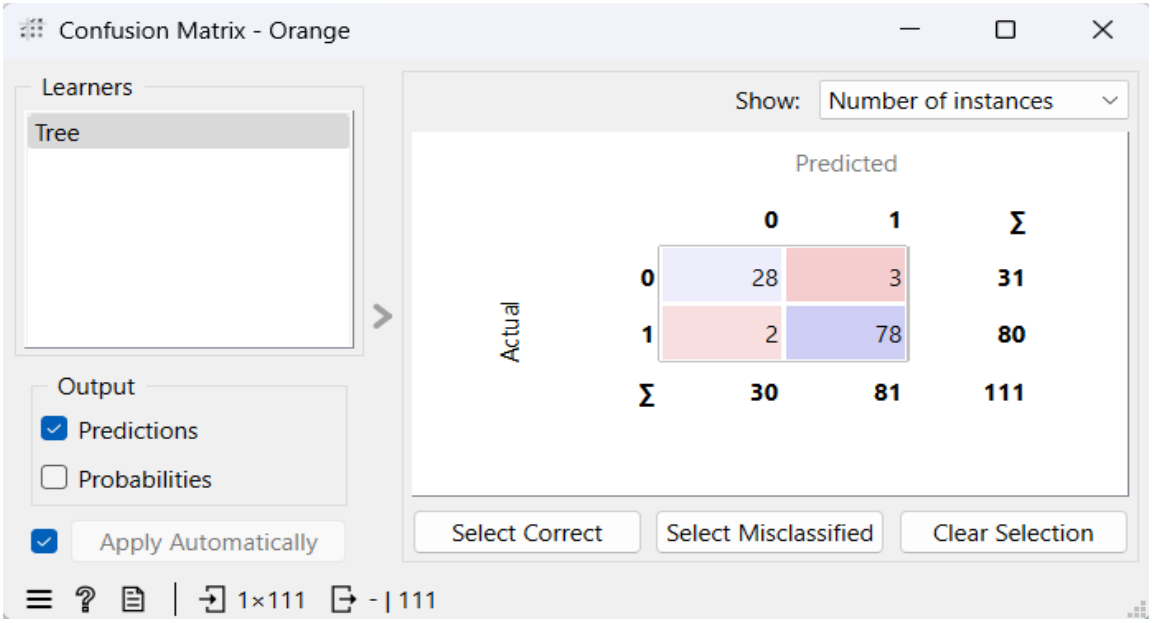
Gambar 2. Data Table pada Aplikasi Orange Mining



Gambar 3. Setting Fitur Tree pada Aplikasi Orange Mining



Gambar 4. Decision Tree dari Aplikasi Orange Mining



Gambar 5. Confusion Matrix dari Aplikasi Orange Mining

Setelah Data Table muncul dan sesuai, langkah selanjutnya adalah menggunakan fitur Tree pada aplikasi Orange Mining dengan custom setting untuk mendapatkan hasil yang maksimal, settingnya seperti Gambar 3.

Dari setting diatas, kita dapat langsung melihat Decision Tree yang dibuat secara otomatis dari Aplikasi Orange Mining dengan menggunakan fitur Tree Viewer, pohon yang terbentuk dari setting tersebut seperti diperlihatkan pada Gambar 4. Untuk pengujian dari data yang ada, didapatkan nilai Confusion Matrix seperti pada Gambar 5.

V. KESIMPULAN

Kesimpulan yang didapat pada penelitian kali ini adalah Pendapatan dapat terpengaruh secara signifikan karena Pendidikan Terakhir, Status Pernikahan, ataupun Hubungan yang terjalin saat ini. Perbedaan kategori dari 3 variabel tersebut memang mempengaruhi pendapatan, namun masih banyak faktor pendukung lain seperti Pengalaman Bekerja, Domisili, dan lain sebagainya.

Penggunaan Decision Tree pada penelitian kali ini sudah dirasa cukup, dari 111 data hanya 5 data yang kurang tepat dalam melakukan prediksi, sehingga prediksi dari penelitian ini cukup tepat hingga 95.5% namun karena masih terdapat kesalahan, mungkin dapat dilakukan penelitian dengan metode yang lain untuk prediksi Pendapatan kali ini.

UCAPAN TERIMA KASIH

Ucapan terima kasih saya haturkan kepada Barry Becker selaku pemilik dataset Adult yang berisi sensus penduduk Amerika Serikat tahun 1994 pada UC Irvine Machine Learning Repositories, Orang Tua yang selalu mendukung pendidikan yang sedang berjalan, Universitas Serang Raya dimana peneliti melakukan pendidikan, Bu Dentik Karyaningsih, M.Kom selaku dosen pada mata kuliah Data Mining, dan Firdan Maulana Muchtar sebagai teman yang mendukung dalam pembuatan jurnal ini.

DAFTAR PUSTAKA

[1] M. Akbariandhini and A. F. Prakoso, "Analisis faktor tingkat pendidikan, jenis kelamin, dan status perkawinan terhadap pendapatan di Indonesia berdasarkan IFLS-5," *JPEKA: Jurnal Pendidikan Ekonomi, Manajemen dan Keuangan*, vol. 4, no. 1, pp. 13-22, 2020.

[2] J. Eska, "Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C45," JURTEKSI (Jurnal Teknol. dan Sist. Informasi), vol. 2, pp. 9–13, 2016.

[3] D. F. Ristianti, "Komparasi Algoritma Klasifikasi pada Data Mining," vol. 1, no. 1, pp. 148–156, 2019.

[4] I. Carolina and R. Kresna, "Klasifikasi kelahiran prematur menggunakan algoritma c4.5," Semin. Nas. Teknol., pp. 668–672, 2018.

[5] Hariati, M. Wati, and B. Cahyono, "Penerapan Algoritma C4.5 Decision Tree pada Penentuan Penerima Program Bantuan Pemerintah Daerah Kabupaten Kutai Kartanegara," Jurti, vol. 2, no. 1, pp. 27–36, 2018.

[6] Y. Rosela, "IMPLEMENTASI KLASIFIKASI DECISION TREE MENGANALISA STATUS PENJUALAN BARANG MENGGUNAKAN C4 . 5 (Studi Kasus : Pt . Matahari Department Store Medan Mall)," J. Pelita Inform., vol. 18, no. 1, pp. 143–150, 2019.