

BAB I

PENDAHALUAN

1.1 Latar Belakang

Pendapatan adalah penghasilan total dari suatu usaha atau aktivitas tertentu, baik dalam bentuk uang atau barang. Pendapatan biasanya dihitung setiap tahun, bulan, minggu atau bahkan hari (Ramadhan et al., 2023).

Penelitian ini menggunakan data sensus pendapatan di Amerika Serikat dibandingkan negara lain karena Amerika mencakup berbagai aspek demografis, ekonomi, sosial, dan geografis yang lengkap dan mudah diakses untuk penelitian. Oleh karena itu, data dari Amerika Serikat tersedia cakupan yang terperinci, yang memungkinkan generalisasi yang lebih baik dalam penelitian. Selain itu, data ini tersedia secara publik seperti Kaggle sehingga memudahkan penelitian. Tetapi, data yang digunakan dalam penelitian berasal dari data mentah dan harus melalui tahap preprocessing seperti memperbaiki data yang hilang. Dengan tahap ini, hasil analisis menjadi dapat diandalkan.

Penelitian ini memprediksi pendapatan dengan menggunakan dataset. Prediksi adalah gagasan tentang apa yang akan terjadi di masa depan (Saraswathi et al., 2023). Prediksi pendapatan memungkinkan individu dapat menerima pendapatan yang diinginkan individu berdasarkan kualifikasi dan pekerjaannya (Prof. D. M. Lothe, Prakash Tiwari, Nikhil Patil, Sanjana Patil, 2023). Penelitian ini bertujuan memprediksi pendapatan tahunan individu berbagai faktor seperti pendidikan, pengalaman kerja, usia dan lainnya untuk mengetahui pendapatan tahunan individu berada di atas atau di bawah \$50.000, yang merupakan indikator umum dalam menentukan tingkat kesejahteraan di Amerika Serikat.

Decision Tree menjadi salah satu algoritma yang bagus untuk mengukur performa menggunakan confusion matrix (Jahan & Shahariar, 2020). Dalam salah satu penelitian menggunakan *Decision Tree C 4.5* menunjukkan data pengujian dengan algoritma ini menghasilkan pengukuran terbaik dengan rata-rata akurasi 96,45%, presisi 96,90%, serta recall 95,38% (Haidar et al., 2020). Pada penelitian

lain setelah data di bagi menjadi data latih dan data uji algoritma *Decision Tree* menunjukkan akurasi terbesar diantara algoritma Regresi Linear dan Random Forest (M. Saraswathi, J. Akhila, dan K. Sireesha. 2023). Oleh sebab itu, *Decision Tree* dipilih untuk sistem prediksi pendapatan.

Dengan demikian, peneliti menggunakan algoritma *Decision Tree C 4.5* untuk menganalisa akurasi dan dapat digunakan untuk sistem prediksi pendapatan.

1.2 Rumusan Masalah

Dari latar belakang yang telah dibuat, rumusan masalah adalah:

1. Bagaimana cara menggunakan algoritma *Decision Tree C 4.5* untuk mengembangkan sistem yang dapat memprediksi pendapatan dengan membaginya ke dalam dua kategori: yang dibawah \$50.000 atau yang diatas \$50.000?
2. Bagaimana tingkat akurasi algoritma *Decision Tree C 4.5* dalam memprediksi pendapatan?

1.3 Tujuan Penelitian

1. Membuat sistem prediksi pendapatan yang akan dibagi dua kategori berupa dibawah 50.000\$ atau diatas 50.000\$ menggunakan algoritma *Decision Tree C 4.5*.
2. Menganalisis karakteristik tingkat akurasi algoritma *Decision Tree C 4.5*.

1.4 Manfaat Penelitian

Dari penelitian sistem prediksi ini selain memberikan estimasi pendapatan tahunan diharapkan sistem prediksi pendapatan pertahun yang dihasilkan dapat bermanfaat bagi individu untuk membuat keputusan yang lebih baik seperti pengembangan karier. Penelitian ini juga memberikan pemahaman terkait algoritma *Decision Tree C 4.5* yang diimplementasikan untuk prediksi pendapatan tahunan.

1.5 Batasan Masalah

1. Hasil prediksi berdasarkan algoritma *Decicion Tree C 4.5*.
2. Terbatas pada dataset dummy situs *kaggle.com*. Datanya tentang pendapatan Amerika Serikat dengan menggunakan mata uang USD. Atribut data berdasarkan dataset mentah yang dibatasi pada atribut: age, education level, job title, years of experience.
3. Output penelitian dari pengujian adalah sistem prediksi pendapatan menggunakan *Google Colaboratory* dan menguji nilai akurasi prediksi pendapatan berdasarkan algoritma *Decicion Tree C 4.5*.

1.6 Sistematika Penulisan

Sistematika penulisan proposal tugas Metodologi Penelitian ini, sebagai berikut:

BAB I. PENDAHULUAN

Pada Bab Pendahuluan dibahas mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada Bab Kajian Literatur akan dibahas dasar-dasar teori yang akan digunakan dalam penelitian, seperti machine learning, data mining, *Decicion Tree C 4.5*

BAB III. METODOLOGI PENELITIAN

Pada Bab Metodologi Penelitian akan dibahas mengenai tahapan pengujian untuk penelitian prediksi pendapatan *Decicion Tree C 4.5*