

Penerapan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar di Kecamatan Juai

Arif Wicaksono Gegadang Putro^{1*}, Tedy Setiadi²

^{1,2}Program Studi Teknik Informatika, Universitas Ahmad Dahlan

arif2000018229@webmail.uad.ac.id, tedy.setiadi@tif.uad.ac.id

Abstract - A statement from a school indicating that a student has successfully completed an educational program is known as graduation. Using the C4.5 decision tree algorithm, a model was developed for this study to predict the graduation status of Juai District students. The average is used to build the prediction model using grades from semester 7 to semester 12, or grades 4 to 6. The attributes used to predict graduation are the average values of Religious Education, Citizenship Education, Mathematics, Science, and Social Sciences, as well as Cultural Arts and Skills and Banjar Language and Literature. The obtained prediction model is a decision tree with three decision leaves. This prediction model has an excellent accuracy of 96.67%.

Keywords: *Predicting passing status, decision tree algorithm C4.5*

Abstrak - Pernyataan dari sekolah yang menunjukkan bahwa seorang siswa telah berhasil menyelesaikan program pendidikan dikenal sebagai kelulusan. Menggunakan algoritma pohon keputusan C4.5, sebuah model dikembangkan untuk penelitian ini untuk memprediksi status kelulusan siswa Kecamatan Juai. Rata-rata digunakan untuk membangun model prediksi dengan menggunakan nilai dari semester 7 sampai semester 12, atau nilai kelas 4 sampai kelas 6. Atribut yang digunakan untuk memprediksi kelulusan adalah nilai rata-rata Pendidikan Agama, Pendidikan Kewarganegaraan, Matematika, IPA, dan IPS, Seni Budaya dan Keterampilan serta Bahasa dan Sastra Banjar. Model prediksi yang diperoleh berupa pohon keputusan dengan tiga daun keputusan. Model prediksi ini memiliki akurasi yang sangat baik sebesar 96,67%.

Kata kunci: *Memprediksi status kelulusan, algoritma pohon keputusan C4.5*

I. PENDAHULUAN

Di Indonesia, pendidikan sekolah dasar (SD) biasanya berlangsung selama enam tahun. Seorang siswa dapat dianggap lulus tepat waktu jika berhasil menyelesaikan sekolah dasar dalam waktu enam tahun. Jika tidak, maka siswa dianggap tidak lulus tepat waktu. Siswa dituntut untuk belajar, dan dalam hal ini mereka harus belajar lebih giat lagi untuk mencapai nilai yang baik.

Kelulusan merupakan hasil dari suatu keputusan. Siswa dinyatakan lulus atau tidaknya merupakan sebuah proses dalam pengambilan keputusan. Seperti halnya dilakukan di SD kecamatan Juai. Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang pengambilan keputusan tidak cukup hanya mengandalkan data operasional saja, tetapi diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada, seperti informasi untuk mendukung pengambilan keputusan yang terkait dengan prediksi kelulusan siswa.

Berdasarkan perbandingan antar komponen, metode Decision Tree dapat digunakan untuk mengidentifikasi komponen utama. Tentu saja, ada sejumlah faktor yang harus diperhitungkan saat menentukan tingkat kelulusan. Faktor pendukung tingkat kelulusan dibandingkan satu sama lain. Komponen-komponen ini dikategorikan berdasarkan persentase masing-masing kategori yang mendapat dukungan paling banyak untuk memudahkan perbandingan.

II. METODOLOGI

2.1 Data Mining

Terdapat berbagai pendapat dari banyak peneliti tentang definisi data mining, diantaranya sebagai berikut

- a. Data Mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan(Suntoro, 2019).

- b. Data Mining adalah sebuah proses menemukan sesuatu yang bermakna dengan memilah data melalui repository dengan bantuan teknologi sosialisasi pola, statistik, serta matematika (Larose, 2006).
- c. Data Mining adalah sebuah aktivitas Analisa data yang memiliki jumlah besar demi menemukan pola(pattern) dan aturan(rule) yang berguna.
- d. Data Mining adalah proses yang menggunakan Teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

2.2 Tahap-Tahap Data Mining

- a. Pembersihan Data (Cleaning)
Pembersihan Data merupakan proses yang digunakan untuk membuang data yang tidak konsisten dan bersifat noise dari data yang terdapat di berbagai basis data yang mungkin berbeda konsisten dan bersifat noise dari data yang terdapat di berbagai basis data yang mungkin berbeda format maupun platform yang kemudian diintegrasikan dalam satu database datawarehouse.
- b. Seleksi Data
Data yang ada pada database atau datawarehouse seringkali tidak semuanya dipakai. Hanya data yang sesuai kebutuhan saja yang dianalisa dan akan diambil dari database atau datawarehouse
- c. Transformasi Data
Data diubah lalu digabungkan ke dalam format yang sesuai untuk dilakukan proses dalam data mining
- d. Proses Mining
Proses Mining merupakan proses utama saat metode diterapkan untuk menemukan informasi atau pengetahuan berharga dari data.
- e. Evaluasi Pola
Untuk mengidentifikasi pola-pola menarik ke dalam basis pengetahuan yang ditemukan
- f. Presentasi Pengetahuan.
Yaitu visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

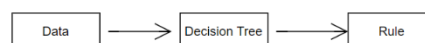
2.3 Klasifikasi

Merupakan prosedur untuk menemukan model atau fungsi yang mendeskripsikan suatu kelas data atau konsep dengan tujuan memodelkan dan memprediksi apakah atribut data tersebut nominal atau numerik dengan label nominal. Model berdasarkan data sebelumnya digunakan dalam klasifikasi. Estimasi kelas objek yang tidak diketahui adalah tujuan dari klasifikasi. Memprediksi kelas target untuk setiap data baru adalah tujuan lain.

2.4 Decision Tree

Decision Tree adalah 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 (dunham, 2003).

Proses pada decision tree adalah mengubah bentuk data tabel menjadi sebuah model tree. Model tree akan menghasilkan rule dan disederhanakan (Basuki & Syarif, 2003). Konsep pohon keputusan seperti gambar berikut.



Data dalam decision tree biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteris dalam pembentukan tree.

Decision tree merupakan salah satu Teknik klasifikasi terhadap objek atau record. Teknik ini terdiri dari kumpulan decision node, dan dihubungkan oleh cabang, bergerak ke bawah dari root node sampai berakhir di leaf node (Yusuf W, 2007).

2.5 Algoritma C4.5

Dalam algoritma C4.5 (Larose, 2005), terdapat beberapa tahap dalam membuat sebuah pohon keputusan, yaitu :

1. Mempersiapkan data training. Data training disebut data masa lalu atau data yang biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas – kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan dipilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung lah dahulu nilai entropy nya. Untuk menghitung nilai entropy digunakan rumus:

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

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

pi = Proporsi Si terhadap S

3. Menghitung nilai Gain menggunakan persamaan 2.

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

Keterangan:

S = Himpunan kasus

A = Fitur

n = Jumlah partisi atribut A

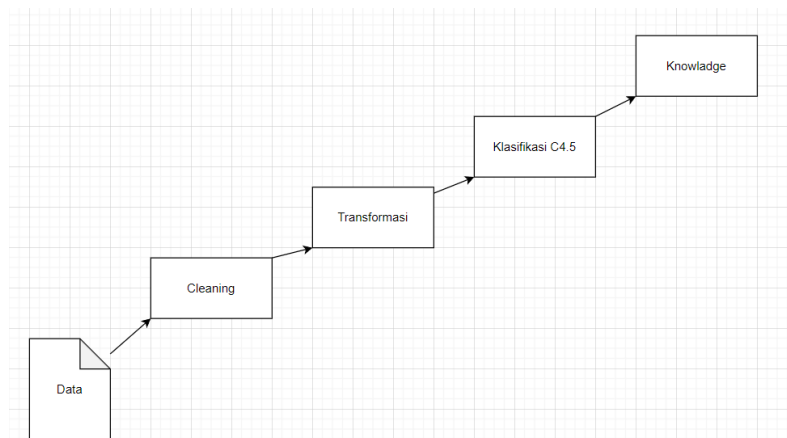
|Si| = Proporsi Si terhadap S

|S| = Jumlah kasus dalam S

4. Agar semua record terpartisi, ulangi Langkah 2 dan Langkah 3
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua record dalam simpul N mendapatkan kelas yang sama
 - b. Di dalam record tidak ada lagi atribut yang dipartisi
 Tidak ada record di dalam cabang yang kosong.

III. HASIL DAN PEMBAHASAN

Implementasi sistem dilakukan sesuai dengan tahap-tahap data mining yang telah ditentukan, diantaranya sebagai berikut. <https://github.com/arfwcksn21/Klasifikasi-Decision-Tree-untuk-Memprediksi-Kelulusan-Siswa-Sekolah-Dasar.git>



1. Data diperoleh dari kecamatan Juai yang berada di Kabupaten Balangan, Kalimantan Selatan.
2. Proses clening terhadap data training siswa sekolah dasar di kecamatan juai.
3. Transformasi yang dilakukan terhadap data yang diperoleh yaitu mengubah atribut Rerata_Keseluruhan dari yang bertipe numerik menjadi tipe kategori Lulus dan Tidak Lulus.
4. Setelah dilakukan tahapan preprocessing, selanjutnya dilakukan proses klasifikasi dengan menggunakan algoirtma C4.5 untuk memprediksi kelulusan siswa sekolah dasar.
5. Hasil yang didapatkan yaitu berupa pohon keputusan.

3.1 Import Library dan data

Dari data set yang diperoleh kolom sebagai atribut dan sebagai label//class. Fitur dalam data yang diperoleh adalah NISN, mean_PA, mean_PKn, mean_Matematika, mean_IPA, mean_IPS, mean_SBK, mean_Penjaskes, mean_Bataqu, dan mean_BSB, Sedangkan label pada data adalah kategori Lulus dan Tidak Lulus. Pada proses import data ini akan menjadi proses awal data mining klasifikasi untuk memprediksi kelulusan siswa sekolah dasar di kecamatan juai, dimana data training yang akan digunakan berbentuk exceli dengan record 224 dan 12 atribut.

```

import pandas as pd

df = pd.read_excel('SD_kec_juai.xlsx')
df

#import library yang akan digunakan
from IPython.display import Image
from sklearn.impute import SimpleImputer
from sklearn import tree
from subprocess import check_call
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
import numpy as np
  
```

3.2 Cleaning

Setelah dilakukan proses import data, kemudian akan dilakukan proses clening terhadap data training siswa sekolah dasar di kecamatan juai, variabel yang tidak memiliki informasi lengkap (IS NOT NULL) akan di-cleaning. Hasil cleaning dapat dilihat pada gambar dibawah.

```
df.Jumlah.fillna(0, inplace=True)
df.Rerata_Keseluruhan.fillna(0, inplace=True)

d_train = df[['RISN', 'mean_Pk', 'mean_PKn', 'mean_Matematika', 'mean_IPA', 'mean_IPS', 'mean_SBK', 'mean_Penjaskes', 'mean_Batagi', 'mean_BSE']]
d_train.fillna(0, inplace=True)

[mean_IPA, mean_IPS, mean_SBK, mean_Penjaskes, mean_Batagi, mean_BSE, 'Jumlah', 'Rerata_Keseluruhan']].reset_index(drop=True)
```

3.3 Transformasi

Transformasi yang dilakukan yaitu mengubah atribut Rerata_Keseluruhan dari yang bertipe numerik menjadi tipe kategori Lulus dan Tidak Lulus. hasil setelah dilakukan proses transformasi dapat dilihat pada gambar dibawah.

```
for index, row in d_train.iterrows():
    if row['Rerata_Keseluruhan'] >= 70:
        d_train.loc[index, 'Status'] = 'Lulus'
    else:
        d_train.loc[index, 'Status'] = 'Tidak Lulus'
d_train

y = d_train['Status'].values
y
```

IV. Klasifikasi Decision Tree (C4.5)

Setelah dilakukan tahapan preprocessing, selanjutnya dilakukan proses klasifikasi dengan menggunakan algoirtma C4.5 untuk memprediksi kelulusan siswa sekolah dasar. Proses klasifikasi dapat dilihat pada gambar dibawah.

```
columns = ['mean_Pk', 'mean_PKn', 'mean_Matematika', 'mean_IPA', 'mean_IPS', 'mean_SBK', 'mean_Penjaskes', 'mean_BSE']
features = d_train[list(columns)].values
features

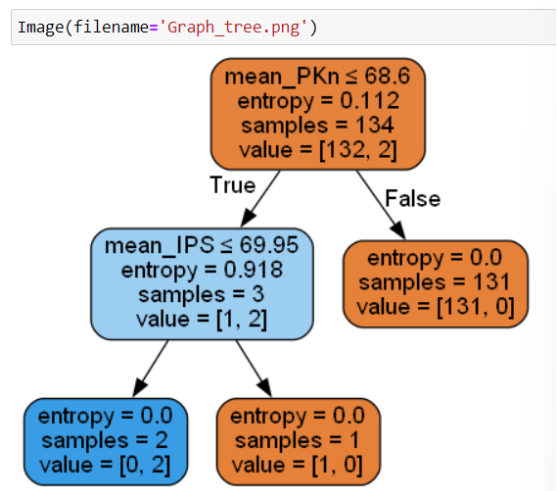
imputer = SimpleImputer(missing_values=np.NaN, strategy='mean', fill_value=None, verbose=0, copy=True)
x = imputer.fit_transform(features)
x

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.4)

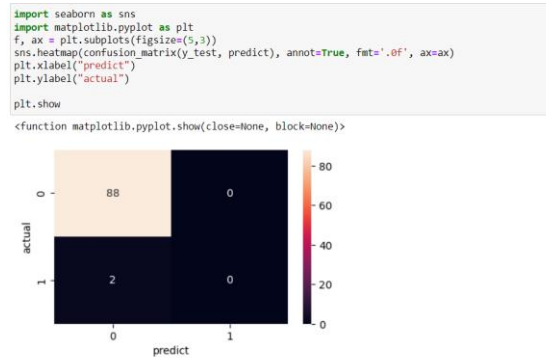
clf = tree.DecisionTreeClassifier(criterion='entropy')
clf = clf.fit(x_train, y_train)
predict = clf.predict(x_test)

with open("Graph_tree.dot", "w") as f:
    f = tree.export_graphviz(clf, out_file=f, feature_names = columns, filled=True, rounded=True, special_characters = True)
check_call(['dot', '-Tpng', 'Graph_tree.dot', '-o', 'Graph_tree.png'])
0
```

Menampilkan Decision Tree



Menampilkan matplotlib



Melakukan report klasifikasi dan akurasi skor

```
from sklearn.metrics import classification_report
print(classification_report(y_test, predict))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Lulus | 0.98 | 1.00 | 0.99 | 88 |
| Tidak Lulus | 0.00 | 0.00 | 0.00 | 2 |
| accuracy | | | 0.98 | 90 |
| macro avg | 0.49 | 0.50 | 0.49 | 90 |
| weighted avg | 0.96 | 0.98 | 0.97 | 90 |

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test, predict)
```

0.9666666666666667

IV. KESIMPULAN

Berdasarkan hasil desain implementasi dan pengujian dapat disimpulkan bahwa algoritma C4.5 digunakan untuk memprediksi kelulusan siswa sekolah dasar di Kecamatan Juai dengan membangun sistem menggunakan algoritma C4.5 setelah menganalisis persyaratan untuk data. dan menguji berbagai sampel data untuk mengetahui keakuratan sistem dalam membuat prediksi.

Berdasarkan hasil pengujian diperoleh hasil yang baik dengan akurasi sebesar 96,67% untuk data tersebut.

V. DAFTAR PUSTAKA

- [1]. Mulia.I, Muanas. 2021, *Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 dan Software Weka. Vol.5 No 1. ISSN 2580 – 5339.* Institut Bisnis dan Informatika Kesatuan.
- [2]. Rohmawan.E.P. 2018, *Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network. Vol.20 No.1.* Universitas Dehasan Bengkulu.
- [3]. Pambudi.R.H, Setiawan.D.B, Indriati. 2018, *Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. Vol.2 No.7.* Universitas Brawijaya.
- [4]. Nugroho.R.G. 2016, *Rangkuman.* Diakses 20 Januari 2023.
- [5]. Rionaldi. 2014, *Klasifikasi Data Mining Menggunakan Algoritma Decision Tree.* Universitas Brawijaya.
- [6]. Bahri.S, Lubis.A. 2020, *Metode Klasifikasi Decision Tree Untuk Memprediksi Juara English Premier Laugue. Vol.2, No.1, April 2020. e-ISSN: 2715-6176/p-ISSN: 2715-5536.* Universitas Pembangunan Panca Budi.

- [7]. Ryandhikafm. 2017, *Klasifikasi Menggunakan Metode Decision Tree dan Naïve Bayes*. Diakses 20 Januari 2023.
- [8]. Setiadi.T, Jamaludin. 2018, *Penerapan Klasifikasi Bayes Untuk Memprediksi Jenis Latihan Siswa Pancak Silat(Studi Kasus Pencak Silat PSHT)*. Vol.7, No.1, Juli 2018. ISSN 2549-8037, e-ISSN 2549-8045. Univeristas Ahmad Dahlan.
- [9]. Lastrini.E. 2020, *Penerapan Metode Algoritma C4.5 Untuk Prediksi Mahasiswa Aktif di STIKOM Binaniaga Bogor*. STIKOM Binaraga Bogor
- [10]. Adhiva.J. 2021, *Sistem Informasi Pemanfaatan Industri dan Konversasi Turunan Kelapa di Kabupaten Indragiri Hilir*. Universitas Islam Negri Sultan Kasim Riau.
- [11]. Nassrullah.A.H. 2018, *Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out*.ILKOM Jurnal Ilmiah 10(2) 244-250.
- [12]. Sadewo.M.G. Windarto.A.P. Dimanik.I.S. Hartama.D. 2018, *Penerapan C4.5 Untuk Memprediksi Kepuasan Pasien Terhadap Sistem Informasi Manajemen Rumah Sakit*.Pemantangsiantar.