

p-ISSN: 2962-4738 e-ISSN: 2962-4584
Vol. 2 No. 7 Juli 2023

**PENERAPAN ALGORITMA DECISION TREE UNTUK SELEKSI PENERIMA
BEASISWA (STUDI KASUS: SMPN 1 SOREANG)**

Putri Andriani Hartanto

Universitas Pendidikan Indonesia, Purwakarta, Indonesia

Email: putriandri@upi.edu

Abstrak

SMP Negeri 1 Soreang merupakan salah satu SMP Negeri favorit di daerah Soreang Kabupaten Bandung. Untuk meningkatkan kualitas sekolah dilakukan melalui penyebaran beasiswa. Penentuan beasiswa ditentukan dengan menggunakan penilaian yang memiliki kriteria tertentu. Dengan memanfaatkan teknik klasifikasi salah satunya model algoritma Decision Tree dipilih untuk dapat membantu dalam memberikan keputusan siapa yang layak untuk mendapatkan beasiswa agar tepat sasaran. Algoritma Decision Tree adalah metode pengklasifikasian untuk memanfaatkan data agar menemukan hubungan antara variabel target dan variabel yang diinputkan. Data berisi nilai siswa sebanyak 237 orang. Dari proses modelling didapatkan nilai akurasi sebesar 92%. Sehingga dapat disimpulkan bahwa algoritma ini bisa membantu dalam memberikan keputusan seleksi bagi penerima beasiswa jenjang Sekolah Menengah Pertama (SMP). Algoritma Decision Tree dipilih karena memiliki tingkat akurasi yang tinggi dibandingkan dengan algoritma klasifikasi yang lainnya. Selain itu, algoritma Decision Tree digunakan untuk mengambil sebuah keputusan dari informasi yang didapatkan.

Kata Kunci: Algoritma, Decision Tree, Beasiswa.

Abstract

SMP Negeri 1 Soreang is one of the favorite public junior high schools in the Soreang area, Bandung Regency. To improve the quality of schools, it is done through the distribution of scholarships. Determination of scholarships is determined by using an assessment that has certain criteria. By utilizing classification techniques, one of which is the Decision Tree algorithm model, which is chosen to be able to assist in making decisions about who is eligible to get a scholarship so that it is right on target. Decision Tree Algorithm is a classification method to utilize data in order to find the relationship between the target variable and the input variable. The data contains student scores of 237 people. From the modeling process, the accuracy value is 92%. So it can be concluded that this algorithm can assist in providing selection decisions for junior high school (SMP) scholarship recipients. The Decision Tree algorithm was chosen because it has a high level of accuracy compared to other classification algorithms. In addition, the Decision Tree algorithm is used to make a decision from the information obtained.

Keywords: Algorithm, Decision Tree, Scholarship.

PENDAHULUAN

SMP Negeri 1 Soreang adalah Sekolah Menengah Pertama Negeri yang berlokasi di Soreang Kabupaten Bandung. Didirikan pada tahun 1960, hingga sekarang SMP Negeri 1 Soreang terus menjadi sekolah favorit pilihan. Dengan adanya titel sebagai sekolah favorit, maka SMP Negeri 1 Soreang terus meningkatkan kualitas sekolah melalui penyebaran beasiswa. Beasiswa diberikan kepada pelajar sebagai tunjangan untuk bantuan biaya belajar [1]. Beasiswa diperuntukkan bagi

siswa yang memiliki latar belakang ekonomi yang kurang mampu agar mendapat kesempatan menjalankan pendidikan. Penentuan beasiswa di sekolah ini menggunakan penilaian yang memiliki kriteria yang perlu dilakukan analisis terhadap kriteria tersebut dengan detail [2].

Dengan menggunakan metode klasifikasi yaitu menilai objek data yang dimasukkan ke beberapa kelas tertentu dari sejumlah kelas yang tersedia [3]. Untuk menunjang hal tersebut maka model algoritma Decision Tree ini dipilih untuk dapat membantu dalam memberikan keputusan siapa yang layak untuk mendapatkan beasiswa agar tepat sasaran sehingga harus melewati seleksi dengan proses yang mudah dan cepat. Adapun metode yang diterapkan dalam proses seleksi ini dengan pemberian kriteria pada nilai siswa untuk menjadi syarat konversi nilai yang sudah ditentukan sebelumnya yang kemudian dianalisis untuk mendapatkan keputusan akhir.

Penggunaan kriteria memanfaatkan data yang berasal dari tahun sebelumnya sebagai acuan untuk alternatif dalam seleksi penerimaan beasiswa jenjang Sekolah Menengah Pertama (SMP).

Pada penelitian kali ini terdapat proses seleksi atribut untuk mengidentifikasi kriteria atribut tersebut. Terdapat 2 ukuran dalam proses seleksi atribut yakni Kriteria Entropy dan Gini Index. Penelitian sebelumnya membahas mengenai “Sistem Pendukung Keputusan Seleksi Penerima Beasiswa dengan Metode Decision Tree” dengan studi kasus pada jenjang Sekolah Dasar oleh Rhamadianoor, boy abidin R, oleh karena itu peneliti mencoba untuk menggunakan algoritma Decision Tree untuk seleksi beasiswa pada jenjang Sekolah Menengah Pertama (SMP) berdasarkan dengan kriteria yang sudah ditentukan agar hasil yang didapatkan dapat efektif dan efisien.

Algoritma Decision Tree dipilih karena memiliki tingkat akurasi yang tinggi dibandingkan dengan algoritma yang klasifikasi yang lainnya. Selain itu, algoritma Decision Tree digunakan untuk mengambil sebuah keputusan dari informasi yang didapatkan.

METODE PENELITIAN

Penelitian ini menggunakan tahapan pada AI Project Cycle yang terdiri dari: (1) Problem Scoping, (2) Data Acquisition atau pengumpulan data, (3) Modeling, (4) Evaluasi dan Validasi Model.

A. Problem Scoping

Masalah merupakan suatu hal yang terkait dengan sebuah proses dan pekerjaan. Masalah membutuhkan solusi untuk dapat dipecahkan. Problem scoping mengacu pada faktor yang mempengaruhi masalah. The 4Ws dari Problem Scoping diantaranya:

1) Who?

Mengacu kepada siapa yang menghadapi masalah, siapa pemangku kepentingan dalam masalah tersebut, serta siapa yang akan terkena dampak akibat masalah tersebut. Pada studi kasus kali ini, yang menghadapi masalah adalah sekolah, karena sekolah yang melakukan seleksi terhadap penerimaan beasiswa, pemangku kepentingan dari masalah ini yaitu guru serta orang tua, sedangkan yang akan terkena dampak dari masalah ini yaitu siswa.

2) What?

Mengacu kepada apa masalah yang terjadi. Pada tahap ini perlu untuk menentukan sifat masalah yang tepat. Masalah utama yaitu sering terjadi salah sasaran bagi penerima beasiswa dalam penyaluran dananya. Faktor masalah ini disebabkan karena riset yang tidak begitu informatif serta penentuan kriteria yang kurang mendetail.

3) Where?

Mengacu kepada situasi atau lokasi masalah terjadi. Permasalahan ini muncul di sektor pendidikan dan tidak dapat dipungkiri terjadi pada setiap jenjang sekolah.

4) Why?

Mengacu kepada alasan mengapa masalah ini perlu diselesaikan. Masalah ini penting dibahas karena akan berdampak pada kurang maksimalnya penyaluran dana beasiswa. Sehingga dilakukanlah pemodelan menggunakan algoritma *Decision tree* untuk mendapatkan nilai akurasi yang tinggi dalam mengambil sebuah keputusan.

B. *Data Acquisition* atau pengumpulan data

Data Acquisition atau pengumpulan data dibutuhkan untuk menyelesaikan sebuah masalah. Tahap ini melewati serangkaian proses dimulai dari pengambilan, pengumpulan, dan validasi data agar data yang dihasilkan sesuai dengan keperluan dalam proyek kecerdasan buatan ini. Pada data akuisisi kali ini, data diambil dari sekolah yang berisi nilai beberapa mata pelajaran. Dataset ini menjadi komponen yang paling penting dalam penelitian ini. Data berisi nilai siswa sebanyak 237 orang.

C. Modeling

Model yang digunakan dalam penelitian ini menggunakan model algoritma *Decision tree*, dimana algoritma ini dijadikan sebagai *classifier* atau metode klasifikasi untuk mengelompokkan data siswa yang diterima maupun yang tidak berdasarkan parameter yang diberikan.

Pemodelan menggunakan algoritma *Decision Tree* pada penelitian ini menggunakan kriteria gini indeks dan kriteria entropi. Kedua model ini digunakan untuk memprediksi keputusan dari inputan nilai mata pelajaran siswa untuk diberikan keputusan apakah dapat lolos atau tidaknya beasiswa tersebut.

D. Evaluasi dan Validasi Model

Evaluasi dilakukan untuk mengukur program yang telah dilaksanakan apakah dapat mencapai hasil sesuai dengan yang diharapkan. Evaluasi menerapkan teknik Confusion Matrix. Confusion Matrix merupakan suatu tabel yang menyajikan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [8].

Dari perolehan data *training* dan data *testing* akan memberikan keputusan melalui confusion matrix. Confusion matrix akan memberikan performa klasifikasi berdasarkan benar atau salah pada objek. Confusion matrix berisi informasi yang aktual serta prediksi pada sistem klasifikasi.

Tabel 1. Confusion Matrix

	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data nilai dari beberapa mata pelajaran. Data terdiri dari 237 siswa, dinamakan terdapat atribut dalam data tersebut yaitu, 'Nama Siswa', 'Kelas', 'No Induk', 'Nilai B.Indo', 'Nilai MTK', 'Nilai B.Ing', 'Nilai IPA', dan 'Jumlah Nilai'.

Data akan melewati proses data training dan data testing dimana akan dilakukan preprocessing data dan juga pelabelan menggunakan model algoritma *Decision Tree*. Untuk jumlah nilai diklasifikasikan pada tabel berikut, yaitu:

Tabel 2. Tabel Klasifikasi Jumlah Nilai

No	Jumlah Nilai	Klasifikasi
1	< 28	Tidak Lolos
2	>= 28	Lolos

Atribut yang dilibatkan pada *Split Validation* yaitu Nilai B.Indo, Nilai MTK, Nilai B.Ing, Nilai IPA, serta Jumlah Nilai, untuk atribut Nama dan No Induk tidak digunakan dalam proses *Split Validation*.

Project Artificial Intelligence kali ini menggunakan model algoritma *Decision Tree* (pohon keputusan) dalam studi kasus seleksi beasiswa pada tingkat Sekolah Menengah Pertama (SMP). Algoritma *Decision Tree* menggunakan struktur seperti pohon yang digunakan untuk memecahkan masalah tertentu yang digunakan untuk klasifikasi dan regresi. Pada *Decision Tree* kali ini menggunakan model dengan kriteria gini indeks dan kriteria entropi. Dataset berasal dari sekolah.

Pada implementasi *Decision Tree* terdapat proses seleksi atribut untuk mengidentifikasi atribut. Terdapat 2 ukuran dalam proses seleksi atribut yaitu *Information Gain* dan *Gini Index*. *Information Gain* menghitung perbedaan sebelum pemisahan dan rata-rata entropi setelah pemisahan nilai atribut yang diberikan. Sedangkan *Gini Index* melakukan proses dengan variabel target kategori “Sukses” atau “Gagal” atau melakukan pemisahan biner.

Sebelum masuk ke tahap pemodelan dilakukan terlebih dahulu proses pengklasifikasian. Pada proses klasifikasi, ditentukan terlebih dahulu variabel kelas untuk dapat menentukan seleksi penerima beasiswa pada studi kasus ini. Langkah-langkah dalam model algoritma *Decision Tree* kali ini yaitu:

1) *Import Library*

Mengimport library dilakukan untuk memanggil beberapa fungsi yang dibutuhkan misalnya untuk membuat tabel, membentuk objek N-Dimensional array, visualisasi data, menentukan hubungan antara dua variabel, serta menampilkan output dalam bentuk grafik.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

2) *Import Dataset*

Mengimport dataset untuk menggabungkan data, dataset memiliki 237 baris dan 8 kolom yang berisi 'Nama Siswa', 'Kelas', 'No Induk', 'Nilai B.Indo', 'Nilai MTK', 'Nilai B.Ing', 'Nilai IPA', dan 'Jumlah Nilai'.

```
data = '/content/DATASET_NILAI_SISWA.csv'

df = pd.read_csv(data, header=None)

df.head()
```

Gambar 2. Dataset

3) Data Preprocessing

	Nama Siswa	Kelas	No Induk	Nilai B.Indo	Nilai MTK	Nilai B.Ing	Nilai IPA	Jumlah Nilai
0	AGHNIA QALBI HASANAH	9A	9A2	72	82.5	66	72	292.5
1	AHMAD ADWA FARHANI	9A	9A3	68	80.0	62	78	288.0
2	AIDIL FIRMANSAH	9A	9A4	66	55.0	64	48	233.0
3	ANDINI DESWITA	9A	9A5	66	85.0	70	66	287.0
4	ARIEFAL APRILIAN HAMDANI	9A	9A6	66	77.5	56	52	251.5

Data *Preprocessing* dilakukan untuk menampilkan informasi data yang akan di analisis.

```
1 df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 237 entries, 0 to 236
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Nama Siswa      237 non-null   object
1   Kelas           237 non-null   object
2   No Induk        237 non-null   object
3   Nilai B.Indo    237 non-null   int64
4   Nilai MTK       237 non-null   float64
5   Nilai B.Ing     237 non-null   int64
6   Nilai IPA       237 non-null   int64
7   Jumlah Nilai    237 non-null   float64
dtypes: float64(2), int64(3), object(3)
memory usage: 14.9+ KB
```

Gambar 3. Tabel hasil *preprocessing*

Lalu dilakukan pengecekan *missing values*.

```
1 # check missing values in variables
2
3 df.isnull().sum()

Nama Siswa      0
Kelas           0
No Induk         0
Nilai B.Indo    0
Nilai MTK        0
Nilai B.Ing     0
Nilai IPA        0
Jumlah Nilai     0
dtype: int64
```

Gambar 4. *Missing Value*

Membuat parameter dengan kolom 'Status', dengan memberikan method np.where(), yang akan mengembalikan elemen yang dipilih dari nilai x dan y tergantung pada nilai array yang dihasilkan oleh kondisi tersebut.

```
df['Status'] = np.where(df['Jumlah Nilai']>=280, 'Lolos', 'Tidak Lolos')
df.head()
```

	Nama Siswa	Kelas	No Induk	Nilai B.Indo	Nilai MTK	Nilai B.Ing	Nilai IPA	Jumlah Nilai	Status
0	AGHНИЯ QALBI HASANAH	9A	9A2	72	82.5	66	72	292.5	Lolos
1	AHMAD ADWA FARHANI	9A	9A3	68	80.0	62	78	288.0	Lolos
2	AIDIL FIRMANSАH	9A	9A4	66	55.0	64	48	233.0	Tidak Lolos
3	ANDINI DESWITA	9A	9A5	66	85.0	70	66	287.0	Lolos
4	ARIEFAL APRILIAN HAMDANI	9A	9A6	66	77.5	56	52	251.5	Tidak Lolos

Gambar 5. Dataset dilengkapi dengan parameter

4) Modeling

Modeling dilakukan dengan membagi data menjadi data training dan data testing. Data training dilakukan untuk mengklasifikasi data yang belum ditentukan [9].

```
# split X and y into training and testing sets

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.33,
random_state = 42)
```

Melakukan klasifikasi *Decision Tree* dengan menggunakan kriteria *Gini Index* dan Kriteria Entropi serta membandingkan akurasi *train set* dan *test set* untuk memeriksa *overfitting*.

a. Gini Index

```
y_pred_gini = clf_gini.predict(X_test)

y_pred_train_gini = clf_gini.predict(X_train)

y_pred_train_gini
```

b. Kriteria Entropi

```
y_pred_en = clf_en.predict(X_test)

y_pred_train_en = clf_en.predict(X_train)
```

```
y_pred_train_en
```

Apabila skor *training set* dan *test set* sebanding maka tidak terjadi *overfitting*. Didapatkan pada model dengan kriteria *Gini Index*, nilai akurasi training set sebesar 0.9684 dan nilai akurasi test set sebesar 0.9367. Sedangkan pada model dengan kriteria Entropy, nilai akurasi Training Set sebesar 0.9620 dan nilai akurasi test set sebesar 0.9241.

a. *Gini Index*

```
1 # print the scores on training and test set
2
3 print('Training set score: {:.4f}'.format(clf_gini.score(X_train, y_train)))
4
5 print('Test set score: {:.4f}'.format(clf_gini.score(X_test, y_test)))

Training set score: 0.9684
Test set score: 0.9367
```

Gambar 6. Nilai akurasi *Gini Index*

b. Kriteria Entropi

```
1 # print the scores on training and test set
2
3 print('Training set score: {:.4f}'.format(clf_en.score(X_train, y_train)))
4
5 print('Test set score: {:.4f}'.format(clf_en.score(X_test, y_test)))

Training set score: 0.9620
Test set score: 0.9241
```

Gambar 7. Nilai akurasi Kriteria Entropi

5) *Evaluation*

Melakukan evaluasi dan validasi model dengan menggunakan confusion matrix, didapatkan hasil dari confusion matrix adalah sebagai berikut.

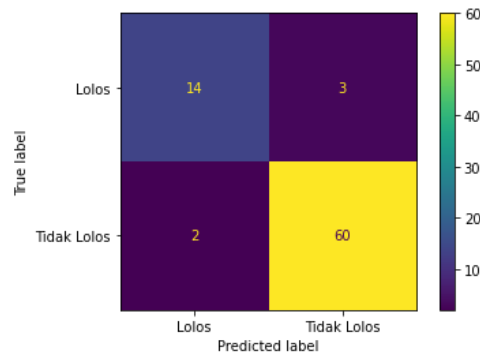
```
1 # Print the Confusion Matrix and slice it into four pieces
2
3 from sklearn.metrics import confusion_matrix
4
5 cm = confusion_matrix(y_test, y_pred_en)
6
7 print('Confusion matrix\n\n', cm)

Confusion matrix

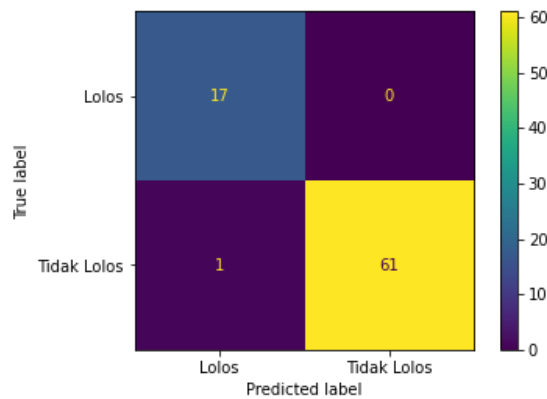
[[14  3]
 [ 3 59]]
```

Gambar 8. Confusion Matrix

Lalu confusion matrix divisualisasikan untuk memudahkan dalam mengevaluasi



Gambar 9. Confusion Matrix dengan model Gini Index



Gambar 10. Confusion Matrix dengan model Kriteria Entropi

Setelah confusion matrix divisualisasikan, dibuat tabel *Classification report* untuk mengevaluasi kinerja model klasifikasi yang menampilkan *recall*, *f1*, dan skor dukungan untuk model dan didapat sebagai berikut.

Tabel 3. *Classification Report*

	<i>Precision</i>	<i>Recall</i>	<i>f1-score</i>	<i>Support</i>
Lolos	0.82	0.82	0.82	17
Tidak Lolos	0.95	0.95	0.95	62
<i>Accuracy</i>			0.92	79
<i>Macro avg</i>	0.89	0.89	0.89	79
<i>Weighted avg</i>	0.92	0.92	0.92	79

KESIMPULAN

Setelah dilakukan proses modeling menggunakan algoritma Decision Tree dapat disimpulkan bahwa algoritma ini bisa membantu dalam memberikan keputusan seleksi bagi penerima beasiswa jenjang Sekolah Menengah Pertama (SMP). Algoritma Decision Tree ini dibuat untuk memprediksi akurasi untuk keputusan seleksi penerimaan beasiswa. Model yang digunakan yaitu, Gini Index dan Kriteria Entropi. Ini dibuktikan pada model dengan kriteria Gini Index, nilai akurasi training set sebesar 0.9684 dan nilai akurasi test set sebesar 0.9367. Sedangkan pada model dengan kriteria Entropy, nilai akurasi training Set sebesar 0.9620 dan nilai akurasi test set sebesar 0.9241.

Atribut yang berpengaruh terhadap seleksi penerimaan beasiswa yaitu jumlah nilai serta status klasifikasinya. Hasil penelitian ini dapat dijadikan pertimbangan untuk menerapkan metode dalam seleksi penerimaan beasiswa di SMP Negeri 1 Soreang. Untuk penelitian selanjutnya, dapat dilakukan dengan dataset yang diolah serta pemilihan atribut lain agar tingkat akurasi yang dihasilkan semakin tinggi. Penelitian selanjutnya dapat menerapkan algoritma yang lain seperti clustering agar dapat dijadikan perbandingan

BIBLIOGRAFI

- [1] E. Setiawan, "Kamus Besar Bahasa Indonesia (KBBI)," Kemdikbud (Pusat Bahasa), 2012 - 2014. [Online]. Available: <http://kbbi.web.id/>. [Accessed 17 June 2022].
- [2] Rhamadianoor, R. (2017) Sistem Pendukung Keputusan Seleksi Penerima Beasiswa Dengan Metode Decision Tree, Progresif: Jurnal Ilmiah Komputer, vol.12, no.1.
- [3] E. Prasetyo, Data Mining - Konsep dan Aplikasi Menggunakan Matlab, Yogyakarta: Penerbit Andi, 2012.
- [4] M. Mohri, A. Rostamizadeh and A. Talwalkar, Foundations of Machine Learning, MIT Press, 2012.
- [5] J. Han and M. Kamber, Data Mining : Concepts and Techniques 2nd Edition, San Francisco: Morgan Kaufmann Publishers, 2006.
- [6] Raharja, Yoso Putra. "Rancang Bangun Sistem Rekomendasi Beasiswa Menggunakan Algoritma Klasifikasi C4. 5 Pada Universitas Dian Nuswantoro." Universitas Dian Nuswantoro, Semarang, skripsi (2014).
- [7] Rustariyuni, S. D. (2014). Pengaruh gini ratio, pengeluaran non makanan per kapita, belanja daerah dan laju pertumbuhan ekonomi pada indeks pembangunan manusia kabupaten/kota di provinsi bali periode 2004-2012. Jurnal Piramida, 10(1), 45-55.
- [8] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. J-SAKTI (Jurnal Sains Komputer dan Informatika), 5(2), 697-711
- [9] Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. Jurnal Khatulistiwa Informatika, 7(1).



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.