

**KOMPUTASI ALGORITMA *DECISION TREE* DAN *RANDOM FOREST* PADA KLASIFIKASI BEASISWA SATU DESA DUA SARJANA KABUPATEN BOJONEGORO**

Skripsi

Disusun sebagai salah satu syarat  
Untuk memperoleh gelar Sejana Statistika  
Program Studi Statistika



Oleh

Anang Ma'ruf  
2520190056

**PROGRAM STUDI STATISTIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS NAHDLATUL ULAMA SUNAN GIRI  
2023**

# PERNYATAAN

## PERNYATAAN

Saya menyatakan bahwa skripsi ini masih dalam batas toleransi plagiat dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Bojonegoro, 28 Agustus 2023



Anang Maruf

NIM. 2520190056

# HALAMAN PERSETUJUAN

## HALAMAN PERSETUJUAN

Nama : Anang Ma'ruf

NIM : 2520190056

Judul : Komputasi Algoritma Decision Tree dan Random Forest pada Klasifikasi  
Basiswa Satu Desa Dua Sarjana Kabupaten Bojonegoro.

Telah disetujui dan dinyatakan memenuhi syarat untuk diajukan dalam ujian  
skripsi.

Bojonegoro, 11 Agustus 2023

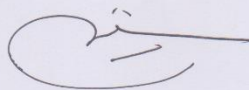
Pembimbing I



Nur Mahmudah, M.Stat

NIDN: 0715039201

Pembimbing II



Fetrika Anggraini, M. pd

NIDN: 0718038803

# HALAMAN PENGESAHAN

## HALAMAN PENGESAHAN

Nama : Anang Ma'ruf

NIM : 2520190056

Judul : Komputasi Algoritma *Decision Tree* dan *Random Forest* Pada Klasifikasi  
Beasiswa Satu Desa Dua Sarjana Kabupaten Bojonegoro

Telah dipertahankan dihadapan penguji pada tanggal 01 September 2023

Dewan Penguji

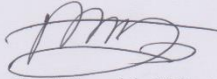
Penguji I



Alif Yuanita Kartini, M.Si  
NIDN. 0721048606

Tim Pembimbing

Pembimbing I



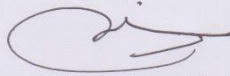
Nur Mahmudah, M.Stat  
NIDN. 0715039201

Penguji II



Dr. Nurul Huda, M.HI  
NIDN. 2114067801

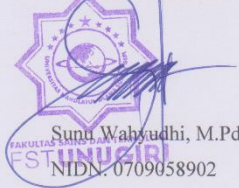
Pembimbing II



Fetrika Anggraini, M.Pd  
NIDN. 0721048606

Mengetahui,

Dekan Fakultas Sains dan Teknologi



Sunu Wahyudhi, M.Pd  
NIDN. 0709058902

Mengetahui,

Ketua Program Studi



Nita Cahyani, M.Stat  
NIDN. 0704038906

## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

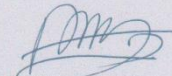
Nama : Anang Ma'ruf  
NIM : 2520190056  
Judul skripsi : Komputasi Algoritma Decision Tree dan Random Forest pada  
Klasifikasi Beasiswa Satu Desa Dua Sarjana Kabupaten  
Bojonegoro

Telah diujikan dalam Seminar Proposal pada tanggal 4 Mei 2023

Telah diujikan dalam Seminar Proposal pada tanggal 4 Mei 2023

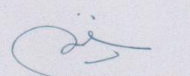
Pembimbing 1

Pembimbing 2



Nur Mahmudah, M.Stat.


NIDN. 0715039201



Fetrika Anggraini, M.Pd.

NIDN. 0718039803

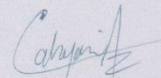
Penguji



Alif Yuanita Kartini, M.Si.

NIDN. 0721048606

Mengetahui,  
Ketua Program Studi Statistika



Nita Cahyani, M.Stat.

NIDN. 0704038906

## MOTTO

*“Ketakutanmu Membatasi Kemampuanmu.”*



# UNUGIRI

## KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT karena atas limpahan rahmat, dan ridhonya sehingga dapat menyelesaikan skripsi yang berjudul “*Komputasi Algoritma Decision Tree, dan Random forest pada klasifikasi beasiswa satu desa dua sarjana Kabupaten Bojonegoro*” dengan tepat waktu dan lancar. Skripsi ini disusun untuk memenuhi persyaratan akademik untuk memperoleh gelar Sarjana pada Program Studi S1 Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.

Kelancaran dalam pengerjaan dan penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak serta doa dan dukungan dari orang tua. Dengan ketulusan hati, penulis mengucapkan terimakasih kepada:

1. Bapak K. M. Jauharul Ma’arif, M.Pd.I. selaku Rektor Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
2. Bapak Sunu Wahyudi, M.Pd. Selaku Dekan Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
3. Ibu Nita Cahyani, M.Stat. selaku Ketua Program Studi Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
4. Ibu Nur Mahmudah, M.Stat. selaku Dosen pembimbing satu yang sudah memberikan arahan serta bimbingan hingga penulis dapat menyelesaikan Skripsi dengan baik.
5. Ibu Fetrika Anggraini, M.Pd. selaku Dosen pembimbing kedua yang sudah memberikan arahan dan selalu menyempatkan waktu untuk bimbingan.
6. Seluruh Dosen Program Studi Statistika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro yang telah memberikan ilmunya.
7. Orang Tua serta keluarga yang selalu mendukung agar skripsi ini dapat segera terselesaikan.

Semoga Allah SWT melimpahkan nikmat dan karunia kepada semua pihak yang telah memberikan bantuan, perhatian serta dukungan kepada penulis dalam menyelesaikan skripsi ini.

Bojonegoro, .....2023

Penulis

Anang Ma’ruf

## ABSTRACT

There is a scholarship organized by the Bojonegoro regency government which makes many residents interested in getting the scholarship. Not only from the lower middle class, but people with high economic status also participate in receiving the scholarship. The number of criteria selected in determining the awarding of scholarships resulted in management difficulties in making a decision and the time needed to be longer. Classification is a process of finding a model or function that describes and distinguishes a class of data or a concept that has the goal of being able to use the model resulting from the classification to make predictions for class objects where the class label is unknown. This research is an attempt to implement the Decision Tree Algorithm and Random Forest as a decision support for the classification of student candidates for one village two undergraduate scholarships. Build a decision support system application that is able to classify students as prospective scholarship recipients using the Decision Tree and Random Forest algorithms. In this study the data to be used is secondary data sourced from the Bojonegoro District Education Office. The data used in this thesis is the One Village Two Bachelor Scholarship Data which was obtained from the Bojonegoro district Education Office in 2022 through submissions made by prospective One Village Two Bachelor Scholarship recipients domiciled in Bojonegoro district. Information was obtained that Random forest obtained more results better than the Decision Tree with an accuracy value of 80%, a Sensitivity of 90% and a Specificity of 65%.4. This study uses the One Village Two Undergraduate Scholarship dataset. Model performance measurements of the Decision Tree and Random Forest algorithms show that in general the values of the Accuracy, Recall or Sensitivity, and Specificity matrices for both algorithms are quite good. The Random Forest algorithm produces a higher Accuracy value than the Decision Tree algorithm, which is 80%. And this algorithm produces Recall 90% for the positive class and 65% for the negative class. And it is known that the most influential variable is the 2nd semester Study Result Card (KHS).

Keyword : Scholarship, Classification, Decision Tree, Random Forest

# UNUGIRI



## ABSTRACT

Adanya beasiswa yang diselenggarakan oleh pemerintah kabupaten Bojonegoro yang membuat banyak warga yang berminat untuk mendapatkan beasiswa tersebut. Bukan hanya dari kalangan menengah kebawah akan tetapi warga yang perekonomiannya tinggi pun ikut serta dalam menerima beasiswa tersebut. Banyaknya kriteria yang diseleksi dalam menentukan keputusan pemberian beasiswa mengakibatkan pihak manajemen kesulitan dalam mengambil sebuah keputusan dan waktu yang dibutuhkan menjadi lebih lama. Klasifikasi merupakan sebuah proses untuk menemukan suatu model atau fungsi yang menggambarkan dan membedakan sebuah kelas data atau konsep yang mempunyai tujuan dapat menggunakan model hasil klasifikasi tersebut untuk membuat prediksi kelas objek dimana kelas labelnya tidak diketahui. penelitian ini sebagai upaya untuk mengimplementasikan Algoritma Decision Tree dan Random Forest sebagai pendukung keputusan klasifikasi mahasiswa calon penerima beasiswa satu desa dua sarjana. Membangun aplikasi sistem pendukung keputusan yang mampu mengklasifikasikan mahasiswa calon penerima beasiswa dengan menggunakan algoritma Decision Tree dan Random Forest. Pada penelitian ini data yang akan digunakan ialah data sekunder yang bersumber dari Dinas Pendidikan kabupaten Bojonegoro. Data yang digunakan dalam skripsi ini merupakan Data Beasiswa Satu Desa Dua Sarjana yang di dapat dari oleh Dinas Pendidikan kabupaten Bojonegoro tahun 2022 melalui pengajuan yang dilakukan oleh calon penerima Beasiswa Satu Desa Dua Sarjana yang berdomisili di kabupaten Bojonegoro. diperoleh informasi bahwa Random forest memperoleh hasil lebih baik dari pada Decision Tree dengan nilai Akurasi sebesar 80%, Sensitivity sebesar 90% dan Specivity sebesar 65%.4. Penelitian ini menggunakan dataset Beasiswa Satu Desa Dua sarjana. Pengukuran performa model dari algoritma Decision Tree dan Random Forest menunjukkan bahwa secara umum nilai dari matriks Accuracy, Recall atau Sensitivity, dan Spesificity pada kedua algoritma cukup baik. Algoritma Random Forest menghasilkan nilai Accuracy yang lebih tinggi daripada algoritma Decision Tree, yaitu sebesar 80%. Dan algoritma ini menghasilkan Recall 90% untuk kelas positif dan 65% untuk kelas negatif. Dan diketahui bahwa variabel yang paling berpengaruh yaitu Kartu Hasil Studi (KHS) semester 2.

Kata Kunci : Beasiswa, Klafisikasi, Decision Tree, Random Forest

## DAFTAR ISI

|  |             |
|--|-------------|
| <b>PERNYATAAN</b> .....                              | <b>iii</b>  |
| <b>HALAMAN PERSETUJUAN</b> .....                     | <b>iv</b>   |
| <b>HALAMAN PENGESAHAN</b> .....                      | <b>v</b>    |
| <b>HALAMAN PENGESAHAN</b> .....                      | <b>vi</b>   |
| <b>MOTTO</b> .....                                   | <b>vii</b>  |
| <b>KATA PENGANTAR</b> .....                          | <b>viii</b> |
| <b>ABSTRACT</b> .....                                | <b>ix</b>   |
| <b>DAFTAR ISI</b> .....                              | <b>xi</b>   |
| <b>DAFTAR TABEL</b> .....                            | <b>xii</b>  |
| <b>DAFTAR GAMBAR</b> .....                           | <b>xiii</b> |
| <b>BAB 1. PENDAHULUAN</b> .....                      | <b>1</b>    |
| 1.1 Latar Belakang .....                             | 1           |
| 1.2 Rumusan Masalah .....                            | 5           |
| 1.3 Tujuan Penelitian .....                          | 6           |
| 1.4 Manfaat Penelitian .....                         | 6           |
| 1.5 Batasan Masalah.....                             | 7           |
| <b>BAB 2. TINJAUAN PUSTAKA DAN DASAR TEORI</b> ..... | <b>8</b>    |
| 2.1 Tinjauan Pustaka .....                           | 8           |
| 2.2 Dasar Teori.....                                 | 16          |
| 2.2.1 Statistika Deskriptif .....                    | 16          |
| 2.2.2 Pemetaan .....                                 | 19          |
| 2.2.3 Data Mining.....                               | 20          |
| 2.2.5 Klasifikasi.....                               | 20          |
| 2.2.6 Decision Tree .....                            | 21          |
| 2.2.7 Random Forest .....                            | 22          |
| 2.2.8 Program Beasiswa Satu Desa Dua Sarjana .....   | 23          |
| 2.2.9 Confusion Matrix .....                         | 25          |
| <b>BAB 3. METODE PENELITIAN</b> .....                | <b>28</b>   |
| 3.1 Sumber Data.....                                 | 28          |
| 3.2 Variabel Penelitian .....                        | 28          |
| 3.2 Langkah Analisis.....                            | 31          |
| <b>BAB 4. HASIL DAN PEMBAHASAN</b> .....             | <b>34</b>   |
| 4.1 Statistika Deskriptif.....                       | 34          |
| 4.2 Pemetaan .....                                   | 37          |
| 4.3 Pre Processing .....                             | 38          |
| 4.4 Decision Tree .....                              | 39          |
| 4.5 Random Forest .....                              | 42          |
| 4.6 Perbandingan Nilai Klasifikasi .....             | 44          |
| <b>BAB 5. KESIMPULAN</b> .....                       | <b>45</b>   |
| 5.1 Kesimpulan .....                                 | 45          |
| 5.2 Saran.....                                       | 48          |
| <b>DAFTAR PUSTAKA</b> .....                          | <b>49</b>   |

## DAFTAR TABEL

|   |    |
|---|----|
| 2.1 Penelitian Terdahulu .....                    | 8  |
| 2.2 Confusion Matrix .....                        | 25 |
| 3.1 Variabel Penelitian .....                     | 28 |
| 4.1 Statistika Deskriptif.....                    | 34 |
| 4.2 Crosstab Pendaftar Berdasarkan Domisili ..... | 35 |
| 4.3 Crosstab Berdasarkan Akreditasi .....         | 37 |
| 4.4 Tipe Data.....                                | 38 |
| 4.5 Data Testing dan Data Training .....          | 39 |
| 4.6 Model Decision Tree.....                      | 39 |
| 4.7 Hasil Prediksi .....                          | 40 |
| 4.8 Confusion Matrix Decision Tree.....           | 41 |
| 4.9 Variabel Paling Berpengaruh .....             | 42 |
| 4.10 Model Random Forest.....                     | 42 |
| 4.11 Hasil Prediksi Random Forest.....            | 42 |
| 4.12 Confusion Matrix Random Forest.....          | 43 |
| 4.13 Perbandingan Nilai.....                      | 44 |



**UNUGIRI**

## DAFTAR GAMBAR

|                                   |    |
|-----------------------------------|----|
| 3.1 Diagram Alir Penelitian ..... | 32 |
| 4.1 Pemetaan .....                | 37 |
| 4.2 Visualisasi Data.....         | 40 |



# UNUGIRI