

**KLASIFIKASI PENERIMAAN MUSYRIF DAN MUSYRIFAH
PUSAT MA'HAD AL-JAMI'AH MELALUI METODE
*BOOTSTRAP AGGREGATING CLASSIFICATION ANALYSIS
REGRESSION TREES***

SKRIPSI

**OLEH
AFIDATUL MASBAKHAH
NIM. 18610055**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**KLASIFIKASI PENERIMAAN MUSYRIF DAN MUSYRIFAH
PUSAT MA'HAD AL-JAMI'AH MELALUI METODE
*BOOTSTRAP AGGREGATING CLASSIFICATION ANALYSIS
REGRESSION TREES***

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Afidatul Masbakhah
NIM. 18610055**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022**

**KLASIFIKASI PENERIMAAN MUSYRIF DAN MUSYRIFAH
PUSAT MA'HAD AL-JAMI'AH MELALUI METODE
BOOTSTRAP AGGREGATING CLASSIFICATION ANALYSIS
REGRESSION TREES**

SKRIPSI

**Oleh
Afidatul Masbakhah
NIM. 18610055**

Telah Disetujui Untuk Diuji

Tanggal, 16 Juni 2022

Dosen Pembimbing I



Angga Dwi Mulyanto, M.Si
NIP. 19890813 201903 1 012

Dosen Pembimbing II



Ari Kusumastuti, M.Pd., M.Si
NIP. 19770521 200501 2 004

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, S.Pd., M.Sc
NIP. 19741129 200012 2 005

**KLASIFIKASI PENERIMAAN MUSYRIF DAN MUSYRIFAH
PUSAT MA'HAD AL-JAM'AH MELALUI METODE
BOOTSTRAP AGGREGATING CLASSIFICATION ANALYSIS
REGRESSION TREES**

SKRIPSI

**Oleh
Afidatul Masbakhah
NIM. 18610055**

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal, 21 Juni 2022

Ketua Penguji : Fachrur Rozi, M.Si

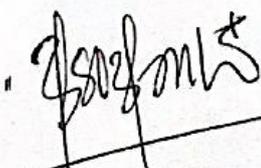
Anggota Penguji 1 : Ria Dhea Layla Nur Karisma, M.Si

Anggota Penguji 2 : Angga Dwi Mulyanto, M.Si

Anggota Penguji 3 : Ari Kusumastuti, M.Pd., M.Si



Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, S.Pd., M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Afidatul Masbakhah

NIM : 18610055

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Klasifikasi Penerimaan Musyrif dan Musyrifah Pusat Ma'had Al-Jami'ah Melalui Metode *Bootstrap Aggregating Classification Analysis Regression Trees*

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri. Bukan merupakan pengambilan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cupikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 21 Juni 2022

Yang membuat pernyataan,



Afidatul Masbakhah
NIM.18610055

MOTO

حَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

“Sebaik-baik manusia adalah yang bermanfaat bagi manusia yang lain”

PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

Kedua orang tua penulis Sabih dan Ruhanah, juga keluarga penulis. Serta sahabat Banser dan Apani, Satu Angkatan Aksioma'18, Galaxy Family, keluarga Majesty, Musyrifah ABA'12 lantai kayangan, dan seluruh pihak yang selalu mendoakan, memberi semangat, nasihat dan kasih sayang yang tak terhingga. Sehingga menjadikan alasan bagi penulis untuk selalu semangat dalam berproses.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji bagi Allah atas rahmat, taufik serta hidayah-Nya, sehingga penulis mampu menyelesaikan penyusunan skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama proses penyusunan skripsi ini, penulis mendapatkan banyak bimbingan, arahan serta dukungan dari berbagai pihak. Sebagai bentuk terimakasih, penulis mengucapkan terimakasih yang sebesar-besarnya dan penghargaan yang setinggi-tingginya yang ditujukan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Angga Dwi Mulyanto, M.Si, selaku dosen pembimbing I yang telah memberikan banyak bimbingan, arahan, dukungan serta perbaikan demi kebaikan penyusunan skripsi.
5. Ari Kusumastuti, M.Pd., M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan, dukungan serta perbaikan kepada penulis.
6. Fachrur Rozi, M. Si, selaku Ketua Penguji dalam Ujian Skripsi.
7. Ria Dhea Layla Nur Karisma, M.Si, selaku Penguji I dalam Ujian Skripsi.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim
9. Orang tua yang selalu memanjatkan do'a dan memberikan semangat yang tiada henti dan seluruh keluarga.
10. Seluruh mahasiswa angkatan 2018 yang selalu berjuang bersama sampai mencapai titik ini.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca maupun bagi penulis, serta dapat dijadikan sebagai penambah wawasan ilmu matematika terutama dalam bidang matematika statistika.

Malang, 21 Juni 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT.....	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	7
1.5 Batasan Masalah	7
BAB II KAJIAN TEORI	8
2.1 <i>Data Training</i> dan <i>Data Testing</i>	8
2.2 <i>Classification Analysis Regression Trees</i> (CART)	8
2.2.1 Proses Membentuk Pohon Klasifikasi	11
2.2.2 Pemangkasan Pohon Klasifikasi.....	14
2.2.3 Penentuan Pohon Klasifikasi Optimal	15
2.3 <i>Bootstrap Aggregating</i> (Bagging).....	16
2.4 Mengukur Tingkat Akurasi Ketepatan Klasifikasi	18
2.5 Penentuan <i>Variable Importance</i>	20
2.6 Musyrif Musyrifah.....	21
2.7 Penerimaan Musyrif Musyrifah.....	22
2.8 Kajian Al-Qur'an Tentang <i>Bootstrap Aggregating Classification Analysis Regression Trees</i>	23
2.9 Kajian Penerimaan Musyrif Musyrifah dengan Metode <i>Bootstrap Aggregating Classification Analysis Regression Trees</i>	25
BAB III METODE PENELITIAN	27
3.1 Jenis Penelitian	27
3.2 Data dan Sumber Data.....	27
3.3 Teknik Analisis Data	27
BAB IV HASIL DAN PEMBAHASAN.....	31
4.1 Analisis Data.....	31
4.2 Analisis CART	37

4.2.1 Penentuan Presentase Pembagian Data <i>Training</i> dan Data <i>Testing</i>	37
4.2.2 Pohon Klasifikasi.....	40
4.2.3 Tingkat Akurasi Ketepatan Klasifikasi CART.....	42
4.2.4 Variabel Penting Pada CART.....	43
4.3 Analisis <i>Bagging</i> CART.....	45
4.3.1 Tingkat Akurasi Ketepatan Klasifikasi <i>Bagging</i> CART	45
4.3.2 Pohon Klasifikasi Setelah <i>Bagging</i> CART.....	48
4.4 Variabel Penting Pada Klasifikasi <i>Bagging</i> CART.....	49
BAB V PENUTUP	52
5.1 Kesimpulan.....	52
5.2 Saran	53
DAFTAR PUSTAKA	54
LAMPIRAN	56
RIWAYAT HIDUP	72

DAFTAR TABEL

Tabel 2.1	Perhitungan APER	19
Tabel 3.1	Variabel Prediktor	29
Tabel 4.1	Deskriptif Indikator Nilai Ujian Tes Tulis	33
Tabel 4.2	Deskriptif Indikator <i>Interview</i> Bahasa Arab	35
Tabel 4.3	Deskriptif Indikator <i>Interview</i> Bahasa Inggris.....	35
Tabel 4.4	Deskriptif Indikator <i>Interview</i> Komitmen.....	36
Tabel 4.5	Ketepatan Klasifikasi Pada Presentase Data.....	38
Tabel 4.6	Ketepatan Klasifikasi Pada Presentase Data.....	38
Tabel 4.7	Ketepatan Klasifikasi Pada Presentase Data.....	38
Tabel 4.8	Ketepatan Klasifikasi Pada Presentase Data.....	39
Tabel 4.9	Ketepatan Klasifikasi Pada Presentase Data.....	39
Tabel 4.10	Hasil Ketepatan Klasifikasi Pohon Klasifikasi	42
Tabel 4.11	Nilai Ketepatan Klasifikasi CART	43
Tabel 4.12	<i>Variable Importance</i> Pada Klasifikasi CART	44
Tabel 4.13	Hasil Ketepatan Klasifikasi <i>Bagging</i> CART	45
Tabel 4.14	Nilai Akurasi Ketepatan Klasifikasi <i>Bagging</i> CART	47
Tabel 4.15	Tingkat Kepentingan Variabel	51

DAFTAR GAMBAR

Gambar 2.1	Struktur <i>Classification Trees</i>	11
Gambar 2.2	Ilustrasi <i>Bagging</i> CART.....	18
Gambar 4.1	Diagram Kategori Hasil Penerimaan Musyrif Musyrifah.....	31
Gambar 4.2	Diagram Diterima dan Tidak Diterima Nilai IPK.....	32
Gambar 4.3	Diagram Diterima dan Tidak Diterima Ujian Baca AL-Qur'an	34
Gambar 4.4	Pohon Klasifikasi Penerimaan Musyrif Musyrifah.....	40
Gambar 4.5	Bar Plot <i>Variable Importance</i> Pada CART.....	44
Gambar 4.6	Kurva ROC.....	47
Gambar 4.7	Pohon Klasifikasi <i>Bagging</i> CART	48
Gambar 4.8	Bar Plot <i>Variable Importance</i>	50

DAFTAR SIMBOL

$p(j_0 t)$: proporsi kelas j pada simpul t
$N_j(t)$: banyaknya kelas j yang diamati pada <i>terminal node</i> t
$N(t)$: jumlah total pengamatan dalam <i>terminal node</i>
$R_\alpha(T)$: <i>resubstitution</i> suatu pohon T pada kompleksitas α
α	: parameter <i>cost complexity</i> bagi penambahan satu simpul akhir pohon T
$R(T)$: <i>resubstitution estimate</i> (asumsi pengganti)
$R^{ts}(T_t)$: total proporsi kesalahan <i>test sample estimate</i>
N_2	: jumlah pengamatan data <i>training</i>
$X(.)$: bernilai 0 jika pernyataan salah dan bernilai 1 jika pertanyaan benar
n_{11}	: banyak observasi Y kelas 1 tepat klasifikasi dengan observasi Y kelas 1
n_{12}	: banyak observasi Y kelas 1 tepat klasifikasi dengan observasi Y kelas 2
n_{21}	: banyak observasi Y kelas 2 tepat klasifikasi dengan observasi Y kelas 1
n_{22}	: banyak observasi Y kelas 2 tepat klasifikasi dengan observasi Y kelas 2
n_{1a}	: banyak subjek observasi Y kelas 1
n_{2a}	: banyak subjek observasi Y kelas 2
n_{1b}	: banyak prediksi Y kelas 1
n_{2b}	: banyak prediksi Y kelas 2
N	: banyak observasi/total prediksi

DAFTAR LAMPIRAN

- Lampiran 1 Data Penerimaan Musyrif dan Musyrifah Pusat Ma'had Al-Jami'ah
- Lampiran 2 Statistik Deskriptif Variabel Prediktor
- Lampiran 3 Kombinasi Pembagian Data Training Dan Testing Pada Presentase Data
- Lampiran 4 Pembentukan Pohon Klasifikasi
- Lampiran 5 *Variable Importance* pada CART
- Lampiran 6 Proses *Bagging* CART
- Lampiran 6 Hasil *Bagging* CART
- Lampiran 7 Penentuan *Variable Importance Bagging* CART

ABSTRAK

Masbakhah, Afidatul. 2022. **Klasifikasi Penerimaan Musyrif dan Musyrifah Pusat Ma'had Al-Jami'ah Melalui Metode *Bootstrap Aggregating Classification Analysis Regression Trees***. Skripsi Program Studi Matematika Fakultas Sains dan teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (1) Angga Dwi Mulyanto, M.Si (II) Ari Kusumastuti, M.Pd., M.Si.

Kata kunci: Penerimaan Musyrif Musyrifah, Analisis Tingkat Akurasi, *Variable Importance*, CART, *Bagging* CART, Ketepatan Klasifikasi.

Penelitian ini bertujuan untuk mengetahui tingkat akurasi klasifikasi dan variabel yang memiliki tingkat kepentingan tertinggi pada klasifikasi penerimaan musyrif dan musyrifah. Analisis ini menggunakan variabel respon diterima dan tidak diterima. Variabel prediktornya yaitu IPK, ujian tulis, interview bahasa arab dan inggris, serta interview komitmen. Data diambil dari hasil seleksi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Malang. Metode yang digunakan adalah *Bagging* CART dengan menghitung nilai akurasi menggunakan APER, sensitivity, dan specificity. Tingkat kepentingan variabel pada pohon klasifikasi dihitung dengan menggunakan varians dari data. Data diolah setelah dilakukan analisis deskriptif. Setelah itu pohon klasifikasi didapatkan dengan analisis CART. Kemudian dilakukan *bagging* untuk memperbaiki tingkat akurasi dari pengklasifikasian. Hasil dari penelitian ini menunjukkan bahwa tingkat akurasi yang diperoleh melalui metode *bagging* CART sebesar 0,9607 atau 96,07% dengan sensitivity sebesar 0,98 atau 98% dan specificity sebesar 0,9423 atau 94,23% serta nilai APER-nya sebesar 3,13%. Setelah dilakukan 150 kali *resampling* *bagging* CART mampu meningkatkan akurasi sebesar 1.02%. *Variable importance* yang mempengaruhi pohon klasifikasi yaitu nilai interview komitmen dengan presentase sebesar 42,14%. Nilai tingkat akurasi setelah dilakukan *bagging* CART total tergolong tinggi, sehingga dapat dikatakan bahwa pohon klasifikasi sesuai dan layak digunakan untuk mengklasifikasikan data baru.

ABSTRACT

Masbakhah, Afidatul. 2022. **Classification of Musyrif and Musyrifah Reqruiement of Pusat Ma'had Al-Jami'ah Using Bootstrap Aggregating Classification Analysis Regression Trees Method.** Thesis the Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors (I) Angga Dwi Mulyanto, M.Si. (II) Ari Kusumastuti, M.Pd., M.Si.

Keywords: Acceptance Musyrif Musyrifah, Accuracy Level Analysis, Variable Importance, CART, Bagging CART, Accuracy of Classification.

This study aims to determine the level of accuracy of classification and variables that have the highest level of importance in the classification of receipts musyrif and musyrifah. This analysis uses accepted and not accepted response variables. The predictor variables are GPA, written exams, Arabic and English interviews, and commitment interviews. The data was taken from the results of the selection of musyrif and musyrifah admissions at the Ma'had Al Jami'ah Center UIN Malang. The method used is CART Bagging by calculating the accuracy value using APER, sensitivity, and specificity. The level of importance of variables in the classification tree is calculated using the variance of the data. The data is processed after descriptive analysis. After that, the classification tree was obtained by CART analysis. Then bagging is carried out to improve the accuracy level of the classification. The results of this study showed that the accuracy rate obtained through the CART bagging method was 0.96 07 or 96.0 7% with a sensitivity of 0.98 or 98% and a specificity of 0.9423 or 94.23% and an APER value of 3.13%. After 150 resampling bagging CART was able to increase accuracy by 1.02%. The variable importance that affects the classification tree is the value of the commitment interview with a percentage of 42.14%. The accuracy level value of the seteis that the total CART bagging is relatively high, so it can be said that the classification tree is appropriate and feasible to use to classify new data.

مستخلص البحث

المسبحة، أفعدة. ٢٠٢٢. تصنيف المقبولين المشريف والمشرفة بمعهد الجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج من خلال أشجار الانحدار. البحث الجامعي، قسم الرياضيات، كلية العلوم والتكنولوجيا الجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف (١) أنجا دوي مولياتنو، الماجستير (٢) أري كوسوماستوتي، الماجستير.

الكلمات الأساسية: القبول، المشريف والمشرفة. مستوى الدقة، الأهمية المتغيرة، التعبئة، *Bagging CART*، *CART*، دقة التصنيف.

تهدف هذه الدراسة إلى تحديد مستوى دقة التصنيف والمتغيرات التي لها أعلى مستوى من الأهمية في تصنيف إيصالات المشرف والمشرفة. يستخدم هذا التحليل متغيرات الاستجابة المقبولة وغير المقبولة. المتغيرات التنبؤية هي المعدل التراكمي، والامتحانات الكتابية، والمقابلات باللغتين العربية والإنجليزية، ومقابلات الالتزام. تم أخذ البيانات من نتائج اختبار قبول المسيريف والمسيريفة في مركز معاذ الجامعة UIN Malang. الطريقة المستخدمة هي تعبئة *CART* عن طريق حساب قيمة الدقة باستخدام *APER* والحساسية والخصوصية. يتم حساب مستوى أهمية المتغيرات في شجرة التصنيف باستخدام تباين البيانات. تتم معالجة البيانات بعد التحليل الوصفي. بعد ذلك، تم الحصول على شجرة التصنيف عن طريق تحليل *CART*. ثم يتم التعبئة لتحسين مستوى دقة التصنيف. أظهرت نتائج هذه الدراسة أن معدل الدقة الذي تم الحصول عليه من خلال طريقة تعبئة *CART* كان ٠,٩٦٠٧ أو ٩٦,٠٧٪ مع حساسية ٠,٩٨ أو ٩٨٪ وخصوصية ٠,٩٤٢٣ أو ٩٤,٢٣٪ وقيمة *APER* ٣,١٣٪. بعد ١٥٠ إعادة أخذ العينات التعبئة والتعليق *CART* كانت قادرة على زيادة الدقة بنسبة ١,٠٢٪. الأهمية المتغيرة التي تؤثر على شجرة التصنيف هي قيمة مقابلات الالتزام بنسبة ٤٢,١٤٪. قيمة مستوى الدقة بعد تعبئة إجمالي سلة التسوق مرتفعة نسبياً، لذلك يمكن القول أن شجرة التصنيف مناسبة ومجدية للاستخدام لتصنيف البيانات الجديدة.

BAB I PENDAHULUAN

1.1 Latar Belakang

Musyrif dan musyrifah merupakan mahasiswa Universitas Islam Negeri Maulana Malik Ibrahim Malang yang sedang menempuh Pendidikan Sarjana minimal semester tiga. Musyrif dan musyrifah sangat membantu dalam proses berkembangnya mahasiswa baru selama berada di ma'had. Selain menjadi pendamping kegiatan wajib ma'had, musyrif musyrifah juga menjadi pendamping kamar dari setiap mahasiswa baru. Mereka mendampingi mulai dari awal mahasantri berkegiatan di pagi hari sampai dengan malam hari. Mereka menemani dan memberikan arahan sesuai kebutuhan mahasiswa baru saat bertempat di ma'had (Pusat Ma'had Al Jami'ah, 2011).

Musyrif dan musyrifah juga mendampingi mahasiswa baru dalam keterikatan emosional. Harapannya bisa menumbuhkan jiwa ulul albab pada diri mahasiswa baru. Penjelasan mengenai jiwa ulul albab sebagaimana yang tercantum pada al-Qur'an surat Ali Imron ayat 190-191 (Qur'an Kemenag, 2016) di bawah ini:

إِنَّ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ وَاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِّأُولِي الْأَلْبَابِ ۝ الَّذِينَ يَذْكُرُونَ اللَّهَ قِيَامًا وَتُعُودًا وَعَلَىٰ جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمَوَاتِ وَالْأَرْضِ رَبَّنَا مَا خَلَقْتَ هَذَا بَاطِلًا سُبْحَانَكَ فَقِنَا عَذَابَ النَّارِ

Artinya:

Sesungguhnya dalam penciptaan langit dan bumi, dan pergantian malam dan siang terdapat tanda-tanda (kebesaran Allah) bagi orang yang berakal. (yaitu) orang-orang yang mengingat Allah sambil berdiri, duduk atau dalam keadaan berbaring, dan mereka memikirkan tentang penciptaan langit dan bumi (seraya berkata), “Ya Tuhan kami, tidaklah Engkau menciptakan semua ini sia-sia; Mahasuci Engkau, lindungilah kami dari azab neraka”.

Tidak semua mahasiswa aktif semester tiga bisa menjadi musyrif musyrifah. Namun, untuk memilih mahasiswa dan mahasiswi tersebut diatur dalam surat pengumuman yang telah ditetapkan oleh Pusat Ma'had al Jami'ah. Aturan tersebut ditetapkan pada surat keluaran No: Un.03. Ma'had/KP.01.01/2/2012 di awal penerimaan seleksi musyrif dan musyrifah.

Ada beberapa ketentuan dalam penerimaan musyrif musyrifah yang ditetapkan oleh Pusat Mahad Al-Jami'ah. Adapun ketentuan menjadi musyrif dan musyrifah yaitu harus memiliki nilai Indeks Prestasi Kumulatif (IPK) minimal 3,25 untuk mahasiswa non-saintek dan 2,75 untuk mahasiswa saintek. Pada penerimaan ini juga dilakukan dengan seleksi ujian tulis, ujian lisan dengan materi baca al-Qur'an, tes *interview* Bahasa Arab, Bahasa Inggris dan komitmen (Pusat Ma'had Al Jami'ah, 2019).

Penerimaan musyrif musyrifah didasarkan oleh beberapa hal. Pertama, didasarkan oleh sistem seleksi penerimaan musyrif musyrifah. Kedua, didasarkan pada kriteria penerimaan musyrif musyrifah sebagai penentu apakah bisa diterima atau tidak yang disesuaikan dengan kriteria yang telah ditentukan. Ketiga, didasarkan pada kebijakan penerimaan musyrif musyrifah yaitu setiap musyrif musyrifah yang diterima, wajib memenuhi persyaratan yang telah ditentukan dengan adanya surat tugas.

Banyak Indikator yang dapat menentukan penerimaan seleksi musyrif dan musyrifah. Akan tetapi, selama ini masih belum ada wujud pengklasifikasian yang digunakan untuk mengklasifikasi seleksi penerimaan musyrif dan musyrifah. Maka dari beberapa indikator ketentuan yang ada pada seleksi penerimaan musyrif dan musyrifah, dapat dicari indikator yang dapat mempengaruhi penerimaan musyrif

dan musyrifah. Berdasarkan faktor-faktor tersebut dapat dikelompokkan melalui metode pengklasifikasian.

Penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah diklasifikasikan menjadi dua kategori yaitu diterima dan tidak diterima. Metode statistika nonparametik yang dapat digunakan menjadi sarana untuk mengelompokkan data yang disusun dengan sistematis salah satunya adalah pengklasifikasian. Metode yang digunakan untuk mengklasifikasikan penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah adalah *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*.

Metode *Classification Analysis Regression Trees (CART)* dilakukan untuk mendapatkan sebuah kelompok data yang memiliki akurasi baik sebagai pertanda dari suatu pengklasifikasian (Breiman, 1998). Metode ini dilakukan untuk mengetahui hubungan antara variabel respon dan variabel prediktor dengan pohon klasifikasi. Metode ini juga digunakan untuk penentuan bentuk pohon klasifikasi yang sesuai berdasarkan variabel yang digunakan. Adanya perubahan yang terletak pada data *training* akan berakibat pada kestabilan dan tingkat akurasi prediksi dari pohon klasifikasi (Breiman, 1998). Sehingga, adanya metode *bagging* CART dapat digunakan untuk membantu kestabilan dan keakurasian prediksi pada pohon klasifikasi. Melalui metode ini nantinya akan diperoleh variabel-variabel yang mempengaruhi penerimaan musyrif dan musyrifah sesuai kategori diterima dan tidak diterima. Variabel yang digunakan akan diketahui kriteria yang dibutuhkan variabel mana yang memiliki pengaruh besar terhadap penerimaan musyrif dan musyrifah setelah tingkat akurasi diketahui.

Metode *Bootstrap Aggregating Classification Analysis Regression Trees* (CART) sudah pernah dilakukan oleh I Ketut Putu Suniantara dan Muhammad Rusli. Mereka meneliti tentang ketepatan klasifikasi *bagging* CART pada klasifikasi ketidaktepatan waktu kelulusan mahasiswa STIKOM Bali. Tingkat akurasi dari klasifikasi dilihat dari nilai *accuracy*, *sensitivity* dan *specificity*. Hasil dari CART diperoleh nilai *accuracy* sebesar 86.18%, *sensitivity* sebesar 93.69% dan nilai *specificity* sebesar 80.26%. Sedangkan, pada metode *bagging* CART diperoleh nilai *accuracy* sebesar 90.56%, *sensitivity* sebesar 95.23% dan nilai *specificity* sebesar 86.78%. Berdasarkan dari kedua hasil nilai akurasi yang diperoleh dapat disimpulkan bahwa penerapan metode *bagging* pada CART dapat meningkatkan akurasi klasifikasi sebesar 4,38% pada akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa STIKOM Bali (Suniantara & Rusli, 2017).

Metode *Bootstrap Aggregating Classification Analysis Regression Trees* (CART) juga sudah pernah dilakukan oleh Winalia Agwil dkk. Mereka menganalisis mengenai klasifikasi ketetapan waktu lulus mahasiswa dari program studi S1 Matematika dengan menggunakan *bagging* CART. Tingkat akurasi dari klasifikasi dilihat dari nilai *accuracy*, *sensitivity* dan *specificity*. Hasil dari CART diperoleh nilai *accuracy* sebesar 82.1%, *sensitivity* sebesar 68.2% dan nilai *specificity* sebesar 91.2%. Sedangkan, pada metode *bagging* CART diperoleh nilai *accuracy* sebesar 85.7%, *sensitivity* sebesar 77.3% dan nilai *specificity* sebesar 91.2% (Agwil, dkk., 2020). Sehingga dapat disimpulkan bahwa penerapan metode *bagging* pada CART pada penelitian tersebut dapat meningkatkan performa klasifikasi.

Penelitian yang menggunakan metode *Bootstrap Aggregating Classification Analysis Regression Trees* (CART) juga pernah dilakukan oleh Ahmad Rusadi Arrahmi dkk. yang meneliti tentang klasifikasi masa studi mahasiswa. Hasil klasifikasi dari 49 data *testing* pada algoritma CART dengan teknik *bagging* ada 13 data (26,531%) untuk masa studi tepat waktu dan 36 data (73,469%) masuk pada klasifikasi tidak tepat waktu. Tingkat akurasi klasifikasi pada algoritma CART sebesar 79.592% sedangkan tingkat akurasi klasifikasi algoritma *bagging* CART diperoleh sebesar 81.633% (Arrahimi, dkk., 2019). Dapat dilihat tingkat akurasi dengan penerapan metode *bagging* pada algoritma CART dapat meningkatkan nilai akurasi klasifikasi.

Berdasarkan beberapa penelitian terdahulu di atas, dapat diketahui bahwa *bagging* CART mampu meningkatkan tingkat akurasi pada beberapa pengklasifikasian sehingga data yang diklasifikasi memiliki akurasi yang akurat. Pada suatu penelitian juga, pernah dilakukan adanya perbandingan perhitungan tingkat akurasi *bagging* dengan beberapa kali pengulangan. Salah satunya dengan 100 kali dan diperoleh hasil pengulangan sebanyak 100 kali dengan nilai keakurasian yang tinggi daripada banyaknya percobaan dengan metode yang lain (Huang, dkk., 2004). Salah satu kegunaan *bagging* juga untuk mengatasi ketidakstabilan pada pohon klasifikasi. Sehingga pada penelitian ini manfaat *bagging* terhadap seleksi musyrif dan musyrifah diharapkan mampu memberikan tingkat akurasi yang lebih akurat.

Berdasarkan uraian latar belakang di atas, memberikan inisiatif penulis untuk melakukan penelitian dengan judul “Klasifikasi Penerimaan Musyrif dan Musyrifah Pusat Ma’had Al-Jami’ah Melalui Metode *Bootstrap Aggregating*

Classification Analysis Regression Trees". Penelitian ini dilakukan untuk mendapatkan model klasifikasi terbaik dari proses *bootstrap aggregating* (bagging) CART.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu:

1. Bagaimana tingkat akurasi pada klasifikasi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang menggunakan metode *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*?
2. Bagaimana tingkat kepentingan variabel-variabel yang mempengaruhi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang menggunakan *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui tingkat akurasi pada klasifikasi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang menggunakan metode *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*.
2. Menganalisis tingkat kepentingan variabel-variabel yang mempengaruhi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang menggunakan *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*.

1.4 Manfaat Penelitian

Manfaat yang diinginkan dari penelitian ini adalah:

1. Pembaca dapat mengetahui perhitungan dan hasil tingkat akurasi pada pengklasifikasian penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang menggunakan metode *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees*.
2. Pembaca dapat mengetahui penerapan metode *bagging* CART pada penentuan tingkat kepentingan indikator yang mempengaruhi penerimaan musyrif dan musyrifah di Pusat Ma'had Al-Jami'ah UIN Maulana Malik Ibrahim Malang.

1.5 Batasan Masalah

Pada penelitian ini, ada batasan masalah untuk membatasi supaya masalah yang akan diteliti menjadi lebih spesifik dan lebih jelas, maka batasan masalahnya yaitu:

1. Data penelitian berupa data sekunder yang diperoleh dari hasil seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al-Jami'ah Tahun 2019/2020 dalam bentuk *sheet* tabel pada *Microsoft Excel*.
2. Penelitian ini menggunakan metode CART model *classification trees*.
3. Pengukuran tingkat akurasi menggunakan $1 - APER$ (untuk akurasi), *sensitivity*, dan *specificity*.
4. Perhitungan *software* menggunakan *RStudio* dengan *package rpart* dan *ipred*.

BAB II KAJIAN TEORI

2.1 Data Training dan Data Testing

Data *training* adalah populasi data yang mempunyai atribut label atau kelas yang digunakan untuk mengetahui karakteristik data sehingga memperoleh model data. Sedangkan data *testing* merupakan populasi data yang mempunyai label atau kelas yang digunakan untuk mengukur ketepatan model data pada klasifikasi data *testing* (Musu, dkk., 2021). Pada pembagian data untuk data *training* dan data *testing* harus dipertimbangkan secara khusus meliputi karakteristik data yang dimiliki dan komposisi atau presentasi dari pembagian jumlah data. Ketidaksesuaian dalam penentuan komposisi jumlah data yang dibagi untuk data *training* dan *testing* akan mempengaruhi tingkat akurasi pada klasifikasi (Nair, dkk., 2001).

Kombinasi data *training* dan *testing* biasanya menggunakan model kombinasi 90%:10%, 85%:15%, 80%:20% dan 75%:25%. Kombinasi terbaik dari kombinasi yang ditentukan diperoleh dari hasil ketepatan klasifikasi dari persentase terbaik pada data *training*. Perhitungan ketepatan data *training* dan *testing* menjadi proses terbentuknya pohon klasifikasi yang baik (Nair, dkk., 2001).

2.2 Classification Analysis Regression Trees (CART)

Classification Analysis Regression Trees (CART) merupakan salah satu metode statistika nonparametrik yang menggunakan teknik pohon klasifikasi. Tujuannya untuk memperoleh hasil klasifikasi suatu kelompok dengan hasil yang akurat dari penelitian. CART digunakan untuk mengetahui hubungan dari variabel respon (dependen) dengan variabel prediktor (independen) dalam bentuk pohon

klasifikasi (Lewis, 2000).

Pada metode CART terdapat 2 analisis, diantaranya *classification trees* dan *regression trees*. Apabila variabel respon yang diteliti berupa data berskala kategorik, maka akan dihasilkan *classification trees*. Sedangkan, apabila variabel respon yang diteliti berupa data berskala kontinu maka akan dihasilkan CART berupa *regression trees* atau pohon regresi (Lewis, 2000). Pada pembentukan pohon klasifikasi CART, analisis yang digunakan adalah model *classification trees*. Pada prosesnya, terdapat proses yang sering disebut dengan *binary recursive partitioning* (pemilihan rekursif biner).

Proses rekursif yaitu proses pemilahan dan pemecahan data dilakukan secara berkali-kali sehingga dapat memperoleh hasil pohon klasifikasi yang tersusun dari banyak *node*. Pemilahan ini dilakukan secara terus menerus hingga tidak ada lagi kesempatan melakukan proses pemilahan selanjutnya. Setiap *node* induk akan dilakukan pemilahan yang disebut dengan biner, sehingga akan diperoleh tepat dua *node* anak. Kemudian, *partitioning* merupakan proses pemecahan sampel menjadi bagian yang lebih kecil dari simpul sebelumnya (Lewis, 2000).

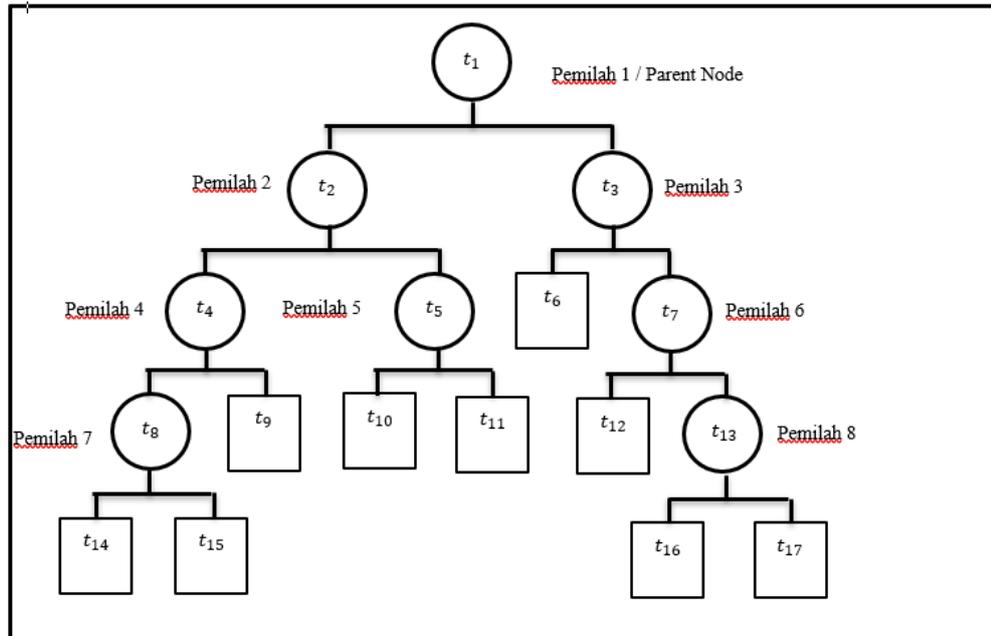
Keuntungan menggunakan metode CART pada penelitian, antara lain (Andriyashin, 2005):

1. CART adalah statistika nonparametrik, sehingga tidak butuh adanya asumsi distribusi khusus dan uji hipotesis yang spesifik.
2. Tidak perlu memilih variabel terlebih dahulu, maksudnya variabel-variabel pada sampel *training* CART akan secara langsung memilih model paling signifikan dan pemisahan yang terbaik.

3. Bentuk CART tersusun sesuai dengan kebutuhan sehingga dapat menghitung semua data dengan cepat dan tepat, dalam hal perhitungan CART sangat efisien.
4. Dapat mengatasi *outlier* dengan baik, pada umumnya algoritma pemisahan akan memisahkan outlier pada beberapa *node* tertentu.
5. Dapat mengatasi dataset dengan menggunakan struktur yang besar dan kompleks.
6. Mampu menganalisis dengan baik segala jenis campuran data, baik data kontinu atau kategorik.

Proses membentuk pohon klasifikasi pada proses CART ada empat tahap, diantaranya (Lewis, 2000):

1. Pemecahan *node* induk menjadi dua *node* anak dilakukan dengan berulang-ulang melalui aturan pemilahan tertentu.
2. Melakukan pemberhentian pada proses pembentukan pohon klasifikasi sehingga diperoleh pohon klasifikasi maksimal (T_{max}) atau pohon klasifikasi terakhir.
3. Proses pemangkasan pohon klasifikasi menjadi lebih kecil yang biasanya disebut proses *pruning trees*, dimana akan menghasilkan pohon klasifikasi optimal.
4. Menentukan tingkat akurasi pada ketepatan klasifikasi melalui hasil pohon klasifikasi yang paling optimal (Lewis, 2000).



Gambar 2.1 Struktur *Classification Trees*

(Sumber: Breiman, 1998)

Pohon klasifikasi yang terbentuk dari metode CART diawali dengan simpul awal atau *parent node* yaitu t_1 . Simpul dalam merupakan simpul anak dari *parent node* pohon klasifikasi yang memiliki anak cabang dibawahnya yaitu $t_2, t_3, t_4, t_5, t_7, t_8, t_{13}$. Sedangkan terminal node merupakan simpul akhir yang sudah tidak menghasilkan cabang atau *child node* kembali yang berbentuk kotak yaitu $t_6, t_9, t_{10}, t_{11}, t_{12}, t_{13}, t_{14}, t_{15}, t_{16}, t_{17}$.

2.2.1 Proses Membentuk Pohon Klasifikasi

Awal membentuk pohon klasifikasi yaitu menentukan variabel sebagai pemilah dan nilai dari variabel (nilai *threshold*) dijadikan pemilah pada tiap *node*. Pada prosesnya membutuhkan data *training* sampel L yang terdiri atas N pengamatan. Proses membentuk pohon klasifikasi terdapat tiga tahap, sebagaimana berikut (Breiman, 1998):

1. Penyaringan Pemilah

Pemilah pada setiap simpul disaring/dipilah untuk memperoleh

pemilah dengan simpul yang nilai variabelnya memiliki tingkat kehomogenan paling besar. Penyaringan pemilah dilakukan dengan mengolah data yang masih bersifat heterogen, sehingga pemilah dari hasil penyaringan dapat digunakan untuk pengurangan keheterogenan dan memaksimalkan ukuran kehomogenan pada setiap *node* induk. Proses pada setiap *node* induknya didasarkan pada kriteria pemilah terbaik (Breiman, 1998).

Pengukuran tingkat keheterogenan suatu kelas pada pohon klasifikasi diukur melalui *node* dengan metode *impurity measure* $i(t)$. Pada metode ini dapat ditemukan fungsi pemilah yang paling optimal yang berupa indeks gini. Teori ini banyak digunakan pada penelitian analisis pohon klasifikasi sebab akan selalu menyaring terlebih dahulu kelas yang memiliki anggota paling besar atau kelas yang paling penting dari *node*. Ukuran yang digunakan untuk mendefinisikan data *impurity* adalah varians. Misalkan 1 merupakan pengamatan yang dilakukan pada simpul t kelas ke- j dan 0 untuk lainnya. Contoh Estimasi varian untuk simpul t pada pengamatan sebagai berikut (Breiman, 1998).

$$p(j|t)(1 - p(j|t)) \quad (2.1)$$

Indeks gini pada simpul t kelas ke- j dapat didefinisikan sebagai:

$$\begin{aligned} i(t) &= \sum_{j=1}^k p(j|t)(1 - p(j|t)) \\ &= \sum_{j=1}^k p(j|t) - p^2(j|t) \\ &= \sum_{j=1}^k p(j|t) - \sum_{j=1}^k p^2(j|t) \\ &= 1 - \sum_{j=1}^k p^2(j|t) \end{aligned} \quad (2.2)$$

Pengevaluasian pemilahan s pada simpul t menggunakan pemilah

terbaik (*Goodness of split*). Misal terdapat pemilahan s yang memilah simpul t menjadi dua *node* tepat, yakni *node* kiri t_L dengan proporsi P_L dan *node* kanan t_R dengan proporsi P_R dengan diketahui $P_R = 1 - P_L$. Sehingga pemilah terbaik sebagai penurunan keheterogenan dapat didefinisikan penurunan keheterogenan seperti berikut (Andriyashin, 2005):

$$\Delta i(s, t) = i(t) - P_L \cdot i(t_L) - P_R \cdot i(t_R) \quad (2.3)$$

dengan $\Delta i(s, t)$ sebagai penurunan tingkat keheterogenan (*improvement*).

Pengembangan pohon klasifikasi dilakukan pada *node* utama t_1 dengan mencari s^* yang memberikan nilai penurunan keheterogenan tertinggi, yaitu:

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t) \quad (2.4)$$

Sehingga simpul t_1 dipilah menjadi dua, t_2 dan t_3 dengan menggunakan s^* . Dengan cara yang sama akan dilakukan pemilahan pemilahan pada t_2 dan t_3 yang dipisah dan seterusnya.

2. Penentuan *Terminal Node*

Pada tahap penentuan simpul terminal, simpul t akan dijadikan pemilah atau *terminal node* dengan catatan jika sudah tidak ada penurunan heterogenitas yang berarti. Kriteria dalam penentuan simpul t menjadi *terminal node* adalah apabila pada simpul t terdapat $n > 5$ dan pohon yang dibentuk akan berhenti apabila sudah mencapai level yang ditentukan pada kriteria pohon maksimal (Breiman, 1998).

3. Pemberian Label Kelas

Memberikan label kelas pada *terminal node* bertujuan untuk

mendapatkan informasi mengenai karakteristik atau sifat dan klasifikasi pada hasil pengamatan di setiap kelas variabel respon yang telah terbentuk. Penentuannya menggunakan aturan jumlah terbanyak dengan persamaan dibawah ini (Breiman, 1998):

$$p(j_0|t) = \max p(j|t) = \max \frac{N_j(t)}{N(t)} \quad (2.5)$$

dengan:

$p(j_0|t)$ = proporsi kelas j pada simpul t

$N_j(t)$ = banyaknya kelas j yang diamati pada *terminal node* t

$N(t)$ = jumlah total pengamatan dalam *terminal node*

j_0 adalah label kelas *terminal node* t . Diasumsikan kesalahan klasifikasi untuk besar setiap kelas adalah sama, sehingga pemberian label kelas j_0 akan memberikan nilai prediksi kesalahan pengelompokan paling kecil pada simpul t yakni sebesar $r(t) = 1 - p(j_0|t)$.

2.2.2 Pemangkasan Pohon Klasifikasi

Pohon optimal ditentukan dengan menggunakan salah satu metode yaitu pemangkasan (*pruning*) (Breiman, 1998). Pada tahapan ini, terjadi pemangkasan atau penghilangan bagian-bagian pohon maksimal yang kurang penting sehingga akan diperoleh pohon klasifikasi yang optimal. Proses pembentukan pohon klasifikasi yang menghasilkan pohon maksimal (T_{max}) akan mengakibatkan dua permasalahan, yaitu (Breiman, 1998):

1. Pohon maksimal dapat mengakibatkan data-data baru mengalami kecocokan nilai yang berlebihan (*overfitting*), meskipun T_{max} memiliki tingkat akurasi yang tinggi dengan nilai kesalahan klasifikasi terkecil. Banyaknya variabel independent yang digunakan untuk memprediksi data baru akan

mengakibatkan *overfitting*.

2. T_{max} akan sulit dipahami dan diinterpretasikan karena T_{max} yang *overfitting*.

Sehingga untuk mengatasi kedua permasalahan di atas, pemangkasan pohon dilakukan dengan memotong atau memangkas T_{max} menjadi pohon klasifikasi (T) yang ukurannya lebih optimal dan lebih kecil dengan menggunakan metode *cost complexity minimum*. Ukuran *cost complexity* dengan $a > 0$ menggunakan persamaan berikut:

$$R_{\alpha}(T) = R(T) + a|\tilde{T}| \quad (2.6)$$

dengan:

$R_{\alpha}(T)$ = *resubstitution* suatu pohon T pada kompleksitas α

α = parameter *cost complexity* bagi penambahan satu simpul akhir pohon T

$R(T)$ = *resubstitution estimate* (asumsi pengganti)

Proses pemangkasan pohon dilakukan dengan mengambil t_R dan t_L dari T_{max} yang dihasilkan oleh simpul induk t . pemangkasan dilakukan apabila persamaan $R(t) = R(t_L) + R(t_R)$ telah terpenuhi pada *node* anak dan *node* induk. Maka diperoleh pohon T_1 yang memenuhi kriteria persamaan $R(T_1) = R(T_{max})$. Proses ini dilakukan berulang-ulang hingga pemangkasan tidak terjadi lagi.

2.2.3 Penentuan Pohon Klasifikasi Optimal

Penentuan pohon optimal bisa ditentukan dari ukuran besar kecilnya pohon. Pohon klasifikasi yang ukurannya besar akan memberikan nilai prediksi pengganti minimum, sehingga pohon tersebut kebanyakan dipilih menjadi prediksi nilai respon. Akan tetapi, pohon yang ukurannya besar mengakibatkan *cost complexity* yang besar dikarenakan struktur data yang digunakan berbentuk kompleks. Sehingga, dibutuhkan pohon klasifikasi sederhana dengan nilai prediksi pengganti

terkecil. Jika pengganti terkecil adalah $R(t)$ yang bisa disebut dengan nilai estimasi terbaik, maka dapat dipilih pohon yang besar, karena memiliki nilai $R(t)$ yang kecil.

Metode yang digunakan untuk menentukan pohon optimal dengan data yang besar adalah *test sample estimate*. Pada proses ini, untuk tahapan yang pertama membagi data menjadi dua, data *training* (L_1) yang digunakan untuk menentukan pohon klasifikasi dan data *testing* (L_2) yang digunakan untuk menentukan nilai $R^{ts}(T_t)$. Dibawah ini persamaan *test sample estimate*.

$$R^{ts}(T_t) = \frac{1}{N_2} \sum_{(x_n, j_n) \in L_2} X(d(x_n) \neq j_n) \quad (2.7)$$

dengan:

$R^{ts}(T_t)$: total proporsi kesalahan *test sample estimate*

N_2 : jumlah pengamatan data *training*

$X(.)$: bernilai 0 jika pernyataan salah dan bernilai 1 jika pertanyaan benar

Tujuan dari penentuan pohon optimal ini untuk mengestimasi (menduga) proporsi kesalahan yang ada pada pohon klasifikasi, sehingga pohon optimal yang dipilih T_t dengan $R^{ts}(T_t) = \min_t R^{ts}(T_t)$.

2.3 *Bootstrap Aggregating (Bagging)*

Suatu metode yang dicetuskan yang digunakan untuk memperbaiki stabilitas dan kekuatan prediksi menggunakan cara reduksi variansi dari suatu predictor dinamakan *bagging* (Breiman, 1998). Beberapa versi yang terbentuk berasal dari sebuah himpunan data dengan replikasi *bootstrap*. Pada beberapa kasus pada suatu himpunan data asli ataupun data simulasi, *bagging* dapat meningkatkan akurasi. *Bagging* mampu meningkatkan akurasi terhadap perubahan signifikan yang disebabkan oleh perubahan himpunan data. Metode dasar *bagging* adalah

resampling bootstrap yang menghasilkan prediktor dengan banyak versi, dimana saat digabungkan seharusnya menghasilkan hasil yang lebih baik dibandingkan dengan prediktor tunggal yang diciptakan untuk menyelesaikan masalah yang sama.

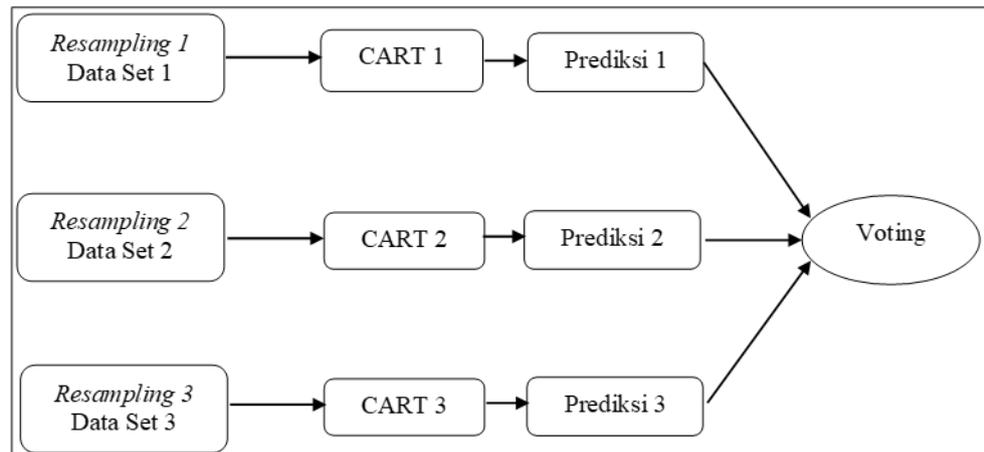
Terdapat dua tahapan *bagging* CART yaitu *bootstrap*, dari data yang ada diambil contoh (*resampling*) dan *aggregating* yaitu menggabungkan nilai-nilai estimasi menjadi satu nilai estimasi. Maka klasifikasi dengan *bagging* CART adalah sebagai berikut (Sartono & Syafitri, 2010):

1. Tahapan *Bootstrap*

- a. Mengambil contoh sebanyak n secara acak dengan pengembalian dari data *training*.
- b. Menyusun data terbaik dari data yang digunakan.
- c. Melakukan langkah a dan b sebanyak k kali sampai diperoleh k buah pohon acak.

2. Tahapan *Aggregating*

Estimasi gabungan dilakukan berdasarkan k pohon yang sudah terbentuk, pemilihan estimasi dapat menggunakan *majority vote*. Selain pengambilan suara terbanyak, penggabungan estimasi juga dapat dilakukan dengan menjumlahkan estimasi peluang pada masing-masing kelas yang diakibatkan pada setiap pohon estimasi (Rokach, 2008). Pada umumnya, ilustrasi model klasifikasi *bagging* CART seperti dibawah ini:



Gambar 2.2 Ilustrasi *Bagging* CART

2.4 Mengukur Tingkat Akurasi Ketepatan Klasifikasi

Ada empat masalah klasifikasi yang harus diperhatikan pada proses pengukuran ketepatan klasifikasi, yaitu (Lewis, 2000):

1. Variabel respon yang memiliki sifat kategorik.
2. Adanya variabel prediktor yang dapat mempengaruhi variabel respon.
3. Penggunaan data *training* untuk pembentukan pohon klasifikasi.
4. Adanya data *testing* untuk mengukur ketepatan klasifikasi.

Perhitungan akurasi klasifikasi dapat digunakan untuk mengukur ketepatan klasifikasi (Richard, dkk, 2007). Ukuran yang bisa digunakan untuk mengetahui ketepatan hasil klasifikasi adalah *Apparent Error Rate* (APER) atau total akurasi. Tipe perbandingan pada prosedur klasifikasi bisa dilihat berdasarkan klasifikasi yang salah, sehingga nilai dari kesalahan jumlah observasi dalam suatu pengklasifikasian fungsi klasifikasi disebut APER. Nilai APER adalah klasifikasi dari nilai proporsi sampel. Perhitungan APER lebih mudah menggunakan tabel silang antara sampel hasil observasi dengan sampel taksiran klasifikasi (Richard, dkk, 2007). Perhitungan APER dapat dilihat sebagai berikut:

Tabel 2.1 Perhitungan APER

Observasi Y	Prediksi		Total
	1	2	
1	n_{11}	n_{12}	n_{1a}
2	n_{21}	n_{22}	n_{2a}
total	n_{1b}	n_{2b}	n

Keterangan:

n_{11} : banyak observasi Y kelas 1 tepat klasifikasi dengan observasi Y kelas 1

n_{12} : banyak observasi Y kelas 1 tepat klasifikasi dengan observasi Y kelas 2

n_{21} : banyak observasi Y kelas 2 tepat klasifikasi dengan observasi Y kelas 1

n_{22} : banyak observasi Y kelas 2 tepat klasifikasi dengan observasi Y kelas 2

n_{1a} : banyak subjek observasi Y kelas 1

n_{2a} : banyak subjek observasi Y kelas 2

n_{1b} : banyak prediksi Y kelas 1

n_{2b} : banyak prediksi Y kelas 2

n : banyak observasi/total prediksi

Beberapa ukuran yang digunakan untuk mengukur ketepatan klasifikasi yaitu:

1. Formula APER (*Apparent Error Rate*) yang merupakan jumlah tingkat ketepatan klasifikasi seperti dibawah ini:

$$APER = \frac{\text{Jumlah prediksi salah}}{\text{Jumlah total prediksi}} \quad (2.8)$$

$$APER = \left(1 - \frac{n_{11} + n_{22}}{n}\right) \times 100\% \quad (2.9)$$

Ketepatan Klasifikasi = 100% - APER (%)

2. *Total accuracy rate* (1-APER) adalah jumlah tingkat ketepatan klasifikasi, formula 1-APER sebagai berikut:

$$1 - APER = \frac{n_{11} + n_{22}}{n} \times 100\% \quad (2.10)$$

3. *Sensitivity* menunjukkan tingkat akurasi pada sampel kelas i dengan formula dibawah ini:

$$Sensitivity = \frac{n_{11}}{n_1} \times 100\% \quad (2.11)$$

4. *Specificity* menunjukkan tingkat akurasi pada sampel kelas j dengan formula:

$$Specificity = \frac{n_{22}}{n_2} \times 100\% \quad (2.12)$$

2.5 Penentuan *Variable Importance*

Pada penerimaan musyrif dan musyrifah terdapat indikator-indikator atau variabel yang dijadikan pertimbangan diterima atau tidak diterima menjadi musyrif dan musyrifah. Namun, untuk variabel terpenting atau yang paling berpengaruh belum diketahui melihat semua variabel memiliki pengaruh dalam proses seleksi. Pada pengklasifikasian CART dan *bagging* CART, setelah terbentuk pohon klasifikasi akan terlihat varibel mana atau indikator yang memiliki pengaruh besar pada proses pengklasifikasian. Variabel penting bisa diketahui dengan urutan variabel pada ketentuan di bawah simpul induk. Variabel penting juga bisa diketahui dari *total of reduction* RSS (*Residual Sum of Square*) atau nilai varians dari data *overall* yang sudah dihasilkan dari pohon klasifikasi yang dibagi dengan data keseluruhan (Bobbitt, 2020).

Perhitungan *variable importance* dapat dihitung dengan formula:

$$VarImp = Var \times n \times 100\% \quad (2.14)$$

$$Var = \frac{1}{n} \sum_{i=0}^N (y_i - \bar{y})^2 \quad (2.15)$$

2.6 Musyrif Musyrifah

Pada buku murobbi murobbiah dan pengembangan musyrif dan musyrifah pusat ma'had al jami'ah, dari segi bahasa musyrif dan musyrifah berasal dari kata *asyrofa-yusyrifu-isyrofan* yang artinya memuliakan, membimbing, mengawasi, mengontrol mendekati dan memberi intruksi. Musyrif dan musyrifah adalah panggilan bagi pengurus yang ada di Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang. Musyrif dan musyrifah adalah mahasiswa semester 3, 5 dan 7 UIN Maulana Malik Ibrahim Malang yang telah lulus seleksi rekrutmen pada setiap tahunnya dengan kualifikasi yang sudah ditentukan pada saat seleksi.

Tugas utama dari musyrif dan musyrifah di Pusat Ma'had Al jami'ah adalah mendampingi mahasiswa baru atau biasa dikenal dengan mahasantri dalam melakukan kegiatan-kegiatan ma'had. Secara spesifiknya, tugas musyrif dan musyrifah dapat mengacu pada program-program yang ada di ma'had dan sesuai dengan tujuannya pula (Pusat Ma'had Al Jami'ah, 2011). Adapun tugas musyrif dan musyrifah diantaranya:

1. Pendampingan dan memonitoring mahasantri dalam bidang spiritual.
2. Pendampingan dan memonitoring mahasantri dalam bidang akademik yang ada di ma'had, yang meliputi:
 - a. Ubudiyah
 - b. Kebahasaan
 - c. Keta'liman (ta'lim afkar dan ta'lim al-Qur'an)
 - d. Kerumahtanggaan
 - e. Kesantrian
 - f. Keamanan

3. Memonitoring dan mengevaluasi program divisi masing-masing.

2.7 Penerimaan Musyrif Musyrifah

Hal yang harus dilakukan sebelum mahasiswa dapat dinyatakan menjadi musyrif dan musyrifah yaitu melalui proses rekrutmen (penerimaan) musyrif musyrifah. Proses ini dilakukan sesuai dengan surat pengumuman seleksi penerimaan No: Un.03. Ma'had/KP.01.01/2/2012, yang diterbitkan oleh Ma'had Sunan Ampel al-Aly UIN Maulana Malik Ibrahim Malang. Terdapat beberapa syarat yang harus dipenuhi oleh peserta calon musyrif musyrifah, diantaranya harus memiliki nilai indeks prestasi kumulatif minimal 3,25 bagi mahasiswa non-saintek dan minimal 2,75 bagi mahasiswa saintek. Penerimaan musyrif dan musyrifah dilaksanakan dengan proses seleksi dengan indikator (Pusat Ma'had Al Jami'ah, 2019):

1. Nilai Indeks Prestasi Kumulatif
2. Nilai Ujian tulis
3. Nilai Ujian lisan baca al-Qur'an, didapatkan dengan membaca al-Qur'an disertai penjelasan mengenai hukum bacaan tajwid yang ada pada bacaan al-Qur'an yang dibaca.
4. Nilai *Interview* Bahasa Arab
5. Nilai *Interview* Bahasa Inggris
6. Nilai *Interview* Komitmen, didapatkan dengan menjawab pertanyaan-pertanyaan seputar pengalaman dan kesanggupan yang diajukan oleh pengasuh sebagai penguji, yang nantinya menjadi ujung tombak tolak ukur penerimaan musyrif dan musyrifah.

Berdasarkan kualifikasi syarat dapat diterimanya mahasiswa UIN Maulana Malik Ibrahim Malang sebagai musyrif dan musyrifah, para calon musyrif musyrifah juga harus memiliki sikap dan perilaku yang baik sehingga dapat menjadi contoh yang baik bagi mahasantri baik dalam bidang spiritual, akademik maupun moral.

2.8 Kajian Al-Qur'an Tentang *Bootstrap Aggregating Classification*

Analysis Regression Trees

Classification analysis regression trees atau disebut CART adalah metode klasifikasi dengan algoritma *recursive biner* dengan pemilahan pada suatu kelompok untuk diperoleh suatu pengklasifikasian. Tujuannya untuk menemukan karakteristik dari suatu populasi dengan hasil yang akurat.

Bootstrap aggregating (bagging) merupakan salah satu dari metode esamble yang digunakan untuk meningkatkan tingkat akurasi pada prediksi pada suatu ketepatan klasifikasi, dimana dengan melakukan pengulangan berkali-kali untuk mendapatkan kesalahan yang tidak terlalu besar.

Firman Allah yang menjelaskan tentang klasifikasi dan pengulangannya ada pada surat Ar-Rahman ayat 62-71 yang berbunyi:

وَمِنْ دُونِهِمَا جَنَّتَانِ . فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ . مُدْهَمَمَتَانِ . فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ . فِيهِمَا عَيْنَانِ
 نَضَّاخَتَانِ . فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ . فِيهِمَا فَاكِهَةٌ وَنَخْلٌ وَرُمَّانٌ . فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ
 فِيهِنَّ حَيْرَاتٌ حِسَانٌ . فَبِأَيِّ آلَاءِ رَبِّكُمَا تُكَذِّبَانِ

Artinya:

Dan selain dari dua surga itu ada dua surga lagi. Maka nikmat Tuhanmu yang manakah yang kamu dustakan? Kedua surga itu (kelihatan) hijau tua warnanya. Maka nikmat Tuhanmu yang manakah yang kamu dustakan? Di dalam keduanya (surga itu) ada dua buah mata air yang memancar. Maka nikmat Tuhanmu yang manakah yang kamu dustakan? Di dalam kedua surga itu ada buah-buahan, kurma dan delima. Maka nikmat Tuhanmu yang manakah yang kamu dustakan? Di dalam surga-surga itu ada bidadari-bidadari yang baik dan jelita. Maka nikmat Tuhanmu yang manakah yang kamu dustakan? (Q. S. Ar-Rahman 62-71).

Pada ayat ini dijelaskan bahwasanya siapa saja yang bertaqwa menghadap Tuhannya, baginya ada dua surga. Surga untuk manusia dan untuk jin. Selain dua surga tersebut, masih ada dua surga lagi untuk orang-orang mukim yang kurang derajatnya dari orang-orang mukmin yang dimasukkan pada surga yang pertama. (Qur'an Kemenag, 2016).

Menurut tafsir Ibnu Katsir ada dua surga selain dari kedua surga yang disebutkan di ayat-ayat sebelumnya. Akan tetapi, tingkatannya, keutamaannya dan kedudukannya masih dibawah surga yang sebelumnya. Satu surga dengan semua wadah dan segala sesuatunya dari emas dan surga yang lain dengan wadah dan segala sesuatunya dari perak. Surga yang pertama untuk kaum muqorrabin dan surga yang kedua untuk ashabul yamin. Abu Musa mengatakan bahwa surga dari emas untuk kaum muqorrabin sedangkan surga dari perak untuk kaum ashabul yamin. Selain dari dua surga tersebut ada lagi dua surga yang tingkatannya berada dibawah kedua surga sebelumnya. Begitupun seterusnya (Learn Quran, 2018).

Menurut Ibnu Zaid, pembagian tingkatan surga ditentukan dalam hal keutamaannya. Gambaran tentang kedua surga tersebut mengandung pengertian surga mana yang lebih dipentingkan. Ini menunjukkan bahwa sesuatu kemuliaan dan ketinggian yang diprioritaskan melebihi yang tidak mulia (Learn Quran, 2018).

Berdasarkan ayat dan penjelasan di atas, dapat diketahui bahwasanya Allah SWT. mengelompokkan atau mengklasifikasikan hamba-hambanya di akhirat nanti sesuai dengan amal apa yang dikerjakan di dunia untuk ditempatkan diantara surga-surga. Jika amal baik akan ditempatkan di surga bagi pemilik amal baik, begitu sebaliknya. Pada surat Ar-Rahman ini sangat banyak penekanan yang diulang beberapa kali, karena pada surat ini memberitahukan dan mengajak seluruh

manusia untuk selalu memperbanyak bersyukur kepada Allah atas nikmat-nikmat yang telah diberikan kepada mereka.

2.9 Kajian Penerimaan Musyrif Musyrifah dengan Metode *Bootstrap Aggregating Classification Analysis Regression Trees*

Musyrif dan musyrifah adalah mahasiswa semester 3, 5 dan 7 UIN Maulana Malik Ibrahim Malang yang telah lulus seleksi rekrutmen pada setiap tahunnya dengan kualifikasi yang sudah ditentukan pada saat seleksi. Penerimaan musyrif dan musyrifah didasari beberapa hal, sesuai yang sudah diatur pada surat keluaran No: Un.03. Ma'had/KP.01.01/2/2012. Ada beberapa indikator yang digunakan sebagai persyaratan diterima atau tidaknya menjadi musyrif dan musyrifah.

Beberapa indikator tersebut akan diklasifikasikan sehingga dapat diperoleh hasil klasifikasi diterima dan tidak diterima pendaftar sebagai musyrif dan musyrifah. Sehingga metode yang dapat digunakan untuk pengklasifikasian penerimaan musyrif dan musyrifah adalah *bootstrap aggregating (bagging) classification analysis regression trees*. Metode ini digunakan untuk membantu kestabilan dan keakuratan prediksi pada pohon klasifikasi yang sesuai dengan variabel berpengaruh.

Setelah diperoleh data penerimaan musyrif dan musyrifah, data tersebut diolah dengan menetapkan variabel respon dan variabel prediktor. Sebelum membagi data untuk proses pohon klasifikasi, data dianalisis menggunakan statistika deskriptif untuk mengetahui nilai-nilai dari variabel yang memungkinkan untuk pengklasifikasian. Kemudian pembagian data menjadi dua bagian, yaitu data *training* dan *testing* dengan beberapa perbandingan.

Proses pengklasifikasian dilakukan dengan CART sampai pohon klasifikasi

terbentuk dengan diperoleh nilai akurasi, *sensitivity* dan *specificity*. Setelah pohon klasifikasi terbentuk melalui proses CART, untuk memperbaiki prediksi yang lebih baik dilakukan proses *bagging* yang melibatkan variabel respon dengan kategori diterima dan tidak diterima. Proses *bagging* juga melibatkan variabel prediktor yang berskala berskala rasio yaitu ujian tulis (x_2), *interview* Bahasa Arab (x_7), *interview* Bahasa Inggris (x_5) dan *interview* komitmen (x_6).

Setelah prediksi dari *bagging* muncul, pengukuran akurasi ketepatan pohon klasifikasi dapat dilakukan dengan menggunakan $1 - APER$ dengan diperoleh nilai akurasi, *sensitivity* dan *specificity*. Dari proses *bagging* CART nanti akan diketahui variabel apa saja yang sangat berpengaruh pada klasifikasi penerimaan musyrif dan musyrifah.

BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan studi literatur dan pendekatan deskriptif kuantitatif. Studi literatur adalah tahapan yang mengakumulasikan teori-teori referensi sebagai panduan peneliti dalam memecahkan permasalahan yang ada pada penelitian. Pendekatan studi literatur yaitu pendekatan dengan mempelajari serta mengkaji buku-buku, jurnal serta artikel yang dibutuhkan yang berkaitan dengan penelitian sebagai sumber referensi.

Sedangkan, metode deskriptif kuantitatif yaitu langkah analisis data yang meneliti mengenai populasi dan sampel tertentu dengan proses analisis menggunakan ilmu statistika yang bersifat kuantitatif. Penelitian kuantitatif lebih ditekankan pada analisis data numerik, kesimpulan yang menjelaskan paparan objek dan hubungan signifikan antara variabel yang diteliti.

3.2 Data dan Sumber Data

Data pada penelitian ini berupa data sekunder. Pengambilan datanya dilakukan pada hari Senin tanggal 03 Januari 2022 di kantor Pusat Ma'had Al Jami'ah. Data yang diperoleh merupakan hasil seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah pada tahun 2019/2020 dengan jumlah 511 data yang juga bersumber dari website pusat ma'had al jami'ah tahun 2019 (Pusat Ma'had Al Jami'ah, 2019).

3.3 Teknik Analisis Data

Penelitian ini menggunakan metode *Bootstrap Aggregating (Bagging)* *Classification Regression Trees* (CART) dengan bantuan *software R* dan *minitab*

19 dengan tahapan sebagai berikut.

1. Mengumpulkan data sekunder.
2. Menetapkan variabel respon dan variabel prediktor yang digunakan pada penelitian, seperti berikut:

- a. Variabel Respon

Variabel respon (Y) terdiri dari diterima dan tidak diterima pada seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah. Variabel respon menggunakan skala nominal sebagai berikut:

Y=1, untuk yang diterima seleksi

Y=2, untuk yang tidak diterima seleksi

Kategori tersebut ditentukan berdasarkan akumulasi nilai yang diperoleh dari setiap tes pada seleksi penerimaan musyrif dan musyrifah.

- b. Variabel Prediktor

Pada penelitian ini menggunakan variabel prediktor yang dianggap berpengaruh pada penentuan seleksi penerimaan musyrif musyrifah Pusat Ma'had Al Jami'ah. Variabel-variabel tersebut adalah sebagai berikut:

Tabel 3.1 Variabel Prediktor

No.	Variabel		Skala	Keterangan
1.	Nilai Indeks Prestasi Kumulatif	x_1	Nominal	1: ≥ 2.75 bagi saintek dan ≥ 3.00 bagi non saintek 2: ≤ 2.75 bagi saintek dan ≤ 3.00 bagi non saintek
2.	Nilai Ujian Tulis	x_2	Rasio	1-100
3.	Nilai Ujian Baca Al-Qur'an	x_3	Nominal	1: Tartil 2: Tidak Tartil
4.	Nilai <i>Interview</i> Bahasa Arab	x_4	Rasio	1-100
5.	Nilai <i>Interview</i> Bahasa Inggris	x_5	Rasio	1-100
6.	<i>Interview</i> Komitmen	x_6	Rasio	1-10

3. Melakukan analisis deskriptif pada data sekunder.
4. Pembagian data menjadi dua bagian, data *training* dan data *testing* dengan mencari kombinasi terbaik dari ketepatan hasil perbandingan 90%:10%, 85%:15%, 80%:20% dan 75%:25%.
5. Mengklasifikasikan data dengan model pohon CART.
 - a. Memilih pemilah dari variabel predictor
 - b. Menentukan *terminal node*
 - c. Memberikan tanda label kelas
 - d. Pemangkasan pohon
 - e. Pembentukan pohon optimal
6. Prediksi data menggunakan data *testing* dengan pohon CART
7. Proses *Bootstrap Aggregating (Bagging)*
 - a. *Resampling* sebanyak 100 sampai 150 kali.
 - b. Memperoleh hasil ketepatan klasifikasi *bagging* CART
 - c. Memperoleh nilai akurasi *bagging* CART
 - d. Ilustrasi *bootstrap aggregating*

Pada proses *bagging* ada ilustrasi *bagging* akan melibatkan variabel respon dengan kategori diterima dan tidak diterima serta melibatkan variabel prediktor berskala rasio yaitu Ujian Tulis (x_2), *Interview* Bahasa Arab (x_4), *Interview* Bahasa Inggris (x_5) dan *Interview* Komitmen (x_6).

8. Pengukuran akurasi ketepatan pohon klasifikasi *bagging* CART dengan APER *accuracy rate*, *sensitivity* dan *specificity*.

$$APER = \left(1 - \frac{n_{11} + n_{22}}{n} \right) \times 100\%$$

$$Accuracy = (1 - APER)$$

$$Sensitivity = \frac{n_{11}}{n_1} \times 100\%$$

$$Specificity = \frac{n_{22}}{n_2} \times 100\%$$

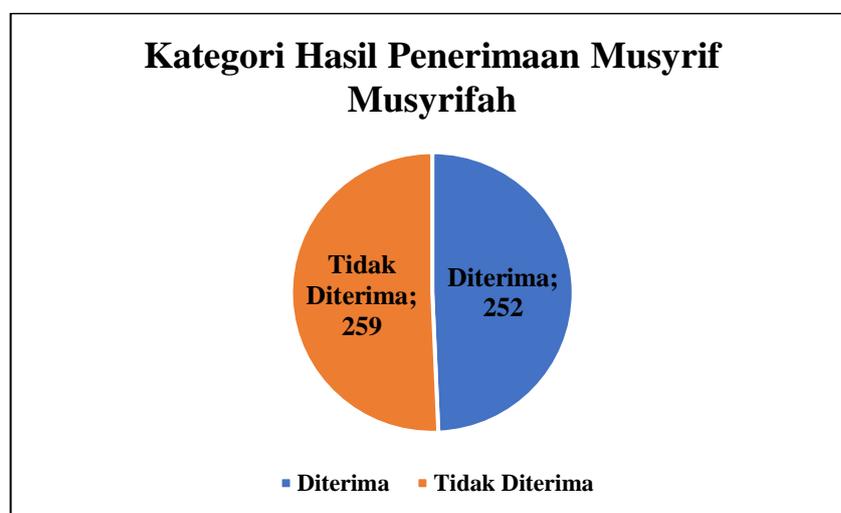
9. Menentukan variabel yang berpengaruh dari interpretasi hasil dari pohon akhir *bagging* CART.

BAB IV HASIL DAN PEMBAHASAN

4.1 Analisis Data

Berdasarkan data yang didapat pada penelitian ini, jumlah banyaknya calon musyrif dan musyrifah di Pusat Ma'had Al jami'ah masa abdi 2019/2020 sebanyak 511 orang. Pada penerimaan musyrif dan musyrifah, dari jumlah pendaftar sebanyak 252 orang dinyatakan telah diterima dan sebanyak 259 orang dinyatakan tidak diterima. Diterima dan tidak diterimanya calon musyrif dan musyrifah menjadi variabel respon pada penelitian ini.

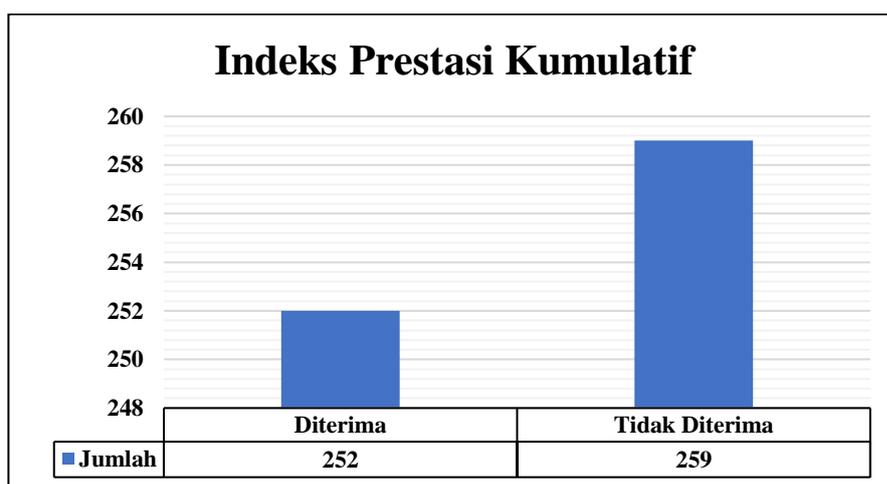
Terdapat juga beberapa indikator yang mempengaruhi penentuan diterima dan tidak diterimanya calon musyrif dan musyrifah di Pusat Ma'had Al Jami'ah, indikator tersebut diantaranya adalah nilai indeks prestasi kumulatif, nilai ujian tulis, nilai ujian baca al-Qur'an, nilai *interview* Bahasa Arab, nilai *interview* Bahasa Inggris dan nilai *interview* komitmen. Indikator-indikator tersebut menjadi variabel prediktor pada penelitian ini dan dapat digunakan untuk menentukan klasifikasi diterima dan tidak diterima pada penerimaan musyrif dan musyrifah.



Gambar 4.1 Diagram Kategori Hasil Penerimaan Musyrif Musyrifah

Perbandingan banyaknya pendaftar calon musyrif dan musyrifah yang diterima dan tidak diterima berdasarkan hasil penerimaan digambarkan dengan gambar diagram lingkaran seperti Gambar 4.1.

Berdasarkan Gambar 4.1 diagram lingkaran hasil penerimaan musyrif dan musyrifah diperoleh informasi bahwa dari 511 calon musyrif dan musyrifah, sebanyak 252 diterima menjadi musyrif musyrifah dan sebanyak 259 tidak diterima menjadi musyrif musyrifah. Indikator pertama yang mempengaruhi seleksi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah yaitu nilai indeks prestasi kumulatif dari setiap calon musyrif dan musyrifah. Didapatkan klasifikasi penerimaan musyrif dan musyrifah berdasarkan IPK digambarkan dengan diagram batang seperti Gambar 4.2.



Gambar 4.2 Diagram Diterima dan Tidak Diterima Nilai IPK

Berdasarkan diagram batang pada Gambar 4.2 diperoleh bahwa dari banyaknya calon musyrif dan musyrifah yang nilai indeks prestasi kumulatif sudah memenuhi kriteria terdapat 252 orang yang diterima menjadi musyrif dan musyrifah, sedangkan yang tidak diterima sebanyak 259 orang. Indikator yang kedua yaitu nilai ujian tulis yang datanya merupakan data berskala rasio dari 1

hingga 100. Berikut adalah deskriptif karakteristik dari nilai ujian tulis berdasarkan diterima atau tidaknya hasil seleksi penerimaan seperti pada Tabel 4.1.

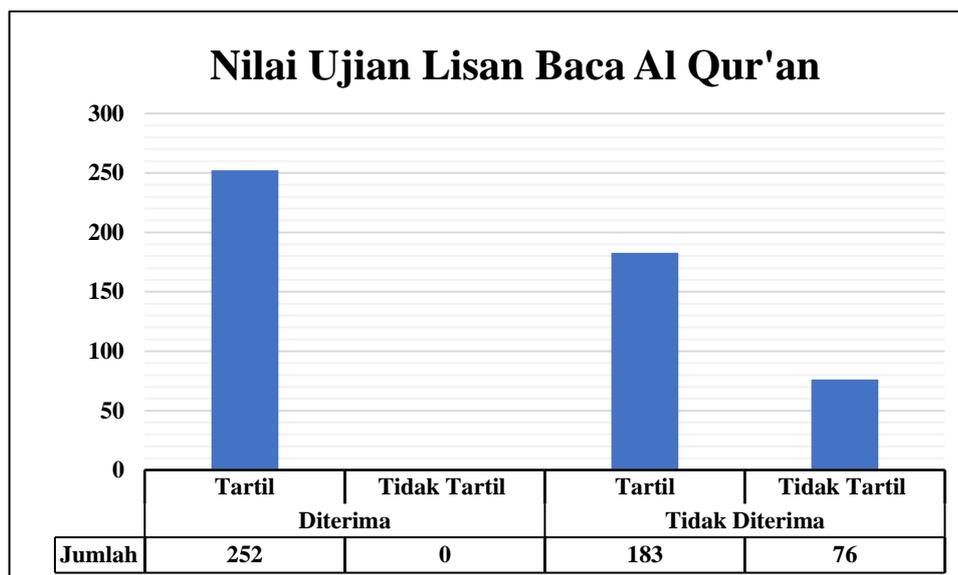
Tabel 4.1 Deskriptif Indikator Nilai Ujian Tes Tulis

Variabel	Mean	St Dev	Minimum	Maksimum
Diterima	88,562	3,224	80,000	96,000
Tidak Diterima	82,565	3,534	78,000	94,000

Berdasarkan Tabel 4.1 dapat dilihat bahwa dari indikator nilai ujian tes tulis pada seleksi penerimaan musyrif dan musyrifah, rata-rata ujian tes tulis yang diterima sebesar 88,562 dengan standart deviasi 3,224. Diketahui juga nilai ujian tes tulis dengan rentang nilai antara 80,000 sampai dengan 96,000 memiliki peluang untuk bisa diterima, akan tetapi dengan nilai maksimum juga belum diterima kecuali didukung dengan indikator-indikator lain yang ada pada seleksi penerimaan.

Pada indikator nilai ujian tes tulis yang tidak diterima ditunjukkan dengan rata-rata sebesar 82,565 dengan standar deviasi 3,534. Nilai ujian tes tulis dengan rentang 78,000 sampai dengan 94,000 dinyatakan tidak diterima padahal nilai 94,000 dapat masuk pada kategori diterima, hal ini disebabkan oleh indikator lainnya yang tidak sesuai dengan kriteria.

Indikator selanjutnya yaitu nilai ujian lisan baca al-Qur'an, dengan kategori tartil dan tidak tartil. Pada ujian lisan al-Qur'an, dapat diketahui untuk kategori tartil berpeluang tidak diterima dengan dipengaruhi indikator yang lain. Berikut adalah deskriptif karakteristik dari nilai ujian tulis berdasarkan diterima atau tidaknya hasil seleksi penerimaan seperti pada Gambar 4.3.



Gambar 4.3 Perbandingan Diterima dan Tidak Diterima Ujian Baca Al-Qur'an

Gambar 4.3 menjelaskan bahwa perbandingan indikator ketiga yaitu nilai ujian lisan baca al-Qur'an. Berdasarkan gambar tersebut diperoleh informasi bahwa kategori tartil pada ujian lisan baca al-Qur'an terdapat 252 orang yang diterima. Sedangkan pada indikator tidak diterima terdapat 183 kategori tartil dan 76 kategori tidak tartil yang tidak diterima menjadi musyrif dan musyrifah. Hal ini menunjukkan bahwa setiap calon musyrif dan musyrifah yang dapat membaca al-Qur'an dengan tartil pada kategori tidak diterima memiliki peluang diterima sebesar 0,71 dari dan berpeluang tidak diterima sebesar 0,29 yang diakibatkan oleh indikator lain sehingga mempengaruhi hasil seleksi penerimaan musyrif dan musyrifah.

Indikator selanjutnya yaitu nilai *interview* Bahasa Arab dengan datanya berupa skala rasio dari 1 sampai dengan 100. Adapun deskriptif karakteristik dari indikator nilai *interview* Bahasa Arab seperti pada Tabel 4.2.

Tabel 4.2 Deskriptif Indikator *Interview* Bahasa Arab

Variabel	Mean	St Dev	Minimum	Maximum
Diterima	86,618	4,159	80,000	95,000
Tidak Diterima	79,240	3,247	75,000	95,000

Berdasarkan Tabel 4.2 diperoleh nilai rata-rata *interview* Bahasa Arab yang diterima pada seleksi penerimaan sebesar 86,618 dengan standar deviasi 4,159 dari keseluruhan calon musyrif dan musyrifah. Pada tabel tersebut diketahui juga minimum nilai *interview* Bahasa Arab yang harus didapatkan supaya bisa diterima adalah 80,000 dan maksimum nilai *interview* Bahasa Arab dari keseluruhan calon musyrif dan musyrifah adalah 95,000 dengan catatan didukung oleh indikator yang lain.

Calon musyrif dan musyrifah yang tidak diterima bisa dilihat dari nilai *interview* Bahasa Arab, jika nilai yang diperoleh masuk dalam rentang 75,000 sampai dengan 95,000 maka ada peluang tidak diterima. Dikarenakan rata-rata nilai *interview* Bahasa Arab sebesar 79,240 dengan standar deviasi 3,247. Sehingga calon musyrif dan musyrifah yang ingin berpeluang menjadi musyrif dan musyrifah harus mendapatkan nilai *interview* Bahasa Arab yang masuk dalam rentang minimum dan maksimum pada kategori diterima.

Indikator selanjutnya adalah nilai *interview* Bahasa Inggris dengan skala rasio 1 sampai dengan 100. Deskriptif karakteristik nilai *interview* Bahasa Inggris pada seleksi penerimaan ditunjukkan oleh Tabel 4.3.

Tabel 4.3 Deskriptif Indikator Nilai *Interview* Bahasa Inggris

Variabel	Mean	St Dev	Minimum	Maximum
Diterima	86,940	2,867	75,000	95,000
Tidak Diterima	76,740	3,546	70,000	90,000

Tabel 4.3 menjelaskan bahwa rata-rata nilai dari indikator *interview* Bahasa Inggris yang diterima pada seleksi penerimaan sebesar 86,940 dengan standar deviasi 2,867 dari keseluruhan calon musyrif dan musyrifah. Pada tabel tersebut diketahui juga minimum nilai *interview* Bahasa Inggris yang harus didapatkan supaya bisa diterima adalah 75,000 dan maksimum nilai *interview* Bahasa Inggris dari keseluruhan calon musyrif dan musyrifah adalah 95,000 dengan faktor pendukung dari indikator lain.

Berdasarkan Tabel 4.3 juga, calon musyrif dan musyrifah yang tidak diterima bisa dilihat dari nilai *interview* Bahasa Inggris, jika nilai yang diperoleh masuk dalam rentang 70,000 sampai dengan 90,000 maka berpeluang tidak diterima. Rata-rata nilai *interview* Bahasa Inggris sebesar 76,740 dengan standar deviasi 3,546. Sehingga calon musyrif dan musyrifah harus mendapatkan nilai *interview* Bahasa Inggris yang masuk dalam rentang minimum dan maksimum pada kategori diterima dengan syarat pula didukung dengan indikator yang lain.

Indikator yang terakhir adalah nilai *interview* komitmen dengan skala rasio 1 hingga 10. Deskriptif karakteristik dari nilai *interview* komitmen ditunjukkan oleh Tabel 4.4.

Tabel 4.4 Deskriptif Indikator Nilai *Interview* Komitmen

Variabel	Mean	St Dev	Minimum	Maximum
Diterima	8,474	0,546	5,000	9,000
Tidak Diterima	6,527	0,899	5,000	9,000

Tabel 4.4 menunjukkan bahwa pada indikator nilai *interview* komitmen dikategorikan diterima jika calon musyrif dan musyrifah mendapatkan nilai dalam rentang nilai minimum 5 dan nilai maksimum sebesar 9. Dikarenakan pada indikator nilai *interview* komitmen diperoleh nilai rata-rata sebesar 8,474 dengan

standar deviasi 0,546 maka calon musyrif dan musyrifah yang ingin diterima dianjurkan untuk mendapatkan nilai *interview* komitmen masuk dalam rentang nilai maksimum dan minimum pada kategori diterima.

Calon musyrif dan musyrifah dikategorikan tidak diterima jika mendapatkan nilai *interview* komitmen dengan rata-rata sebesar 6,527 dengan standar deviasi 0,899, atau bisa juga dikarenakan nilai masuk dalam rentang nilai minimum dan nilai maksimum pada kategori tidak diterima. Rentang nilai minimum dan maksimum pada kategori diterima dan tidak diterima diketahui sama, hal ini disebabkan adanya indikator lain yang mempengaruhi tidak diterima atau diterimanya calon musyrif dan musyrifah pada proses seleksi penerimaan.

4.2 Analisis CART

Analisis *Classification Analysis Regression Trees* (CART) digunakan untuk mengetahui klasifikasi penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah yang dibagi dalam dua kategori yaitu diterima dan tidak diterima. Pada penelitian ini tahap pertama yaitu melakukan pembagian data menjadi dua, yaitu data *training* dan data *testing* dengan beberapa macam presentase terhadap 511 data penerimaan musyrif dan musyrifah pada tahun 2019/2020 di Pusat Ma'had Al Jami'ah.

4.2.1 Penentuan Presentase Pembagian Data *Training* dan Data *Testing*

Pada penelitian ini, pembagian presentase data *training* dan data *testing* menggunakan 4 macam kombinasi presentase, yaitu: (90%,10%), (85%,15%), (80%,20%), dan (75%,25%). Selanjutnya membentuk pohon optimal di setiap pembagian persentase data yang ada, kemudian dihitung ketepatan klasifikasi pada tiap-tiap klasifikasi pembagian persentase data. Setelah diperoleh hasil dari

ketepatan klasifikasi, dilanjutkan dengan perbandingan, kombinasi dari persentase akan dipilih jika ketepatan klasifikasinya mempunyai nilai data *training* yang paling baik. Data *training* digunakan pada proses membentuk pohon klasifikasi sedangkan data *testing* digunakan untuk melakukan pengukuran kelayakan atau kebaikan model.

Pada pemilihan kombinasi terbaik data *training*, dilakukan 9 kali percobaan dengan melihat ketepatan klasifikasi pada persentase data seperti:

Tabel 4.5 Ketepatan Klasifikasi Pada Presentase Data

No	Presentase Data (%)		Percobaan 1		Percobaan 2	
	Training	Testing	Training	Testing	Training	Testing
1	90	10	0,50391645	0,49608355	0,501305483	0,498694517
2	85	15	0,50855746	0,49144254	0,506112469	0,493887531
3	80	20	0,51612903	0,48387097	0,495391705	0,504608295
4	75	25	0,50869565	0,49130435	0,493478261	0,506521739

Tabel 4.6 Ketepatan Klasifikasi Pada Presentase Data

No	Presentase Data (%)		Percobaan 3		Percobaan 4	
	Training	Testing	Training	Testing	Training	Testing
1	90	10	0,496083551	0,503916449	0,483028721	0,516997128
2	85	15	0,49633252	0,50366748	0,488997555	0,511002445
3	80	20	0,497995853	0,502304147	0,491783410	0,509216590
4	75	25	0,497826087	0,502173913	0,489130435	0,510869565

Tabel 4.7 Ketepatan Klasifikasi Pada Presentase Data

No	Presentase Data (%)		Percobaan 5		Percobaan 6	
	Training	Testing	Training	Testing	Training	Testing
1	90	10	0,485639690	0,51436031	0,49347258	0,50652742
2	85	15	0,493887530	0,50611247	0,498777506	0,501222494
3	80	20	0,497695853	0,502304147	0,5	0,5
4	75	25	0,5	0,5	0,5	0,5

Tabel 4.8 Ketepatan Klasifikasi Pada Presentase Data

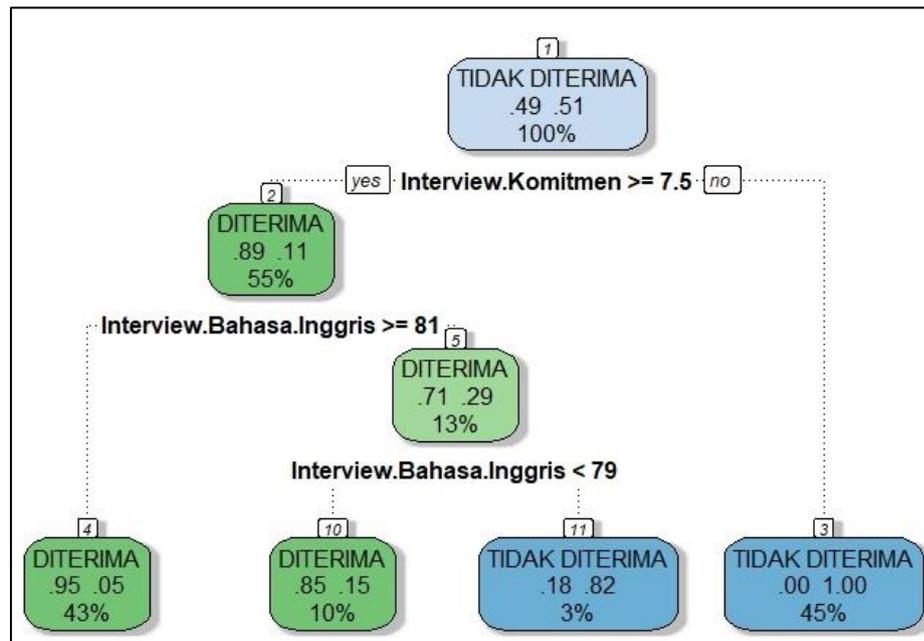
No	Presentase Data (%)		Percobaan 7		Percobaan 8	
	Training	Testing	Training	Testing	Training	Testing
1	90	10	0.486956522	0.513043478	0.480434783	0.519565217
2	85	15	0.486175120	0.51382488	0.486175115	0.513824885
3	80	20	0.493887531	0.506112469	0.486552567	0.513447433
4	75	25	0.490861619	0.509138381	0.483028720	0.51697128

Tabel 4.9 Ketepatan Klasifikasi Pada Presentase Data

No	Presentase Data (%)		Percobaan 9	
	Training	Testing	Training	Testing
1	90	10	0.493478261	0.506521739
2	85	15	0.495391705	0.504608295
3	80	20	0.498777506	0.501222494
4	75	25	0.496083551	0.503916449

Berdasarkan Tabel 4.5 sampai Tabel 4.9 dapat diketahui bahwasanya dari empat kombinasi presentase data, setelah dilakukan 9 kali percobaan didapatkan satu kombinasi terbaik. Hasil dari kombinasi data tersebut yang dibandingkan pertama adalah nilai dari ketepatan klasifikasi data *training* karena dari data *training* dapat dilihat kelayakan dan ketepatan pohon klasifikasi. Pada 9 kali percobaan, data *training* yang paling tinggi ketepatannya muncul sebanyak 7 kali yaitu pada kombinasi 80% data *training* dan 20% data *testing*. Sehingga kombinasi terbaik yang digunakan untuk pembagian data *training* dan data *testing* adalah 80% dan 20%. Sehingga pada analisis CART presentase data yang digunakan adalah 80% data *training* dan 20% data *testing*.

4.2.2 Pohon Klasifikasi



Gambar 4.4 Pohon Klasifikasi Penerimaan Musyrif Musyriah

Gambar 4.4 diperoleh pohon klasifikasi dengan simpul induk (t_1) yang hasilnya tidak diterima dengan indikator *interview* komitmen, karena data yang menyatakan tidak diterima lebih banyak daripada pohon yang menyatakan diterima. Diketahui bahwa variabel yang penting berpengaruh pada proses pengklasifikasian dari CART adalah nilai *interview* komitmen dan nilai *interview* Bahasa Inggris. Pada pohon klasifikasi Gambar 4.4, diketahui ada dua simpul dalam yaitu t_2 , t_5 , dan empat simpul terminal yaitu t_3 , t_4 , t_{10} dan t_{11} . Berdasarkan pohon di atas, dapat diketahui bahwa:

1. *Node* induk berisi kategori tidak diterima dengan peluang 100%. Jika nilai *interview* komitmen lebih dari sama dengan 7,5 maka peluang tidak diterima sebesar 0,49 dan jika nilai *interview* komitmen kurang dari sama dengan 7,5 maka peluang tidak diterima sebesar 0,51.

2. Pada simpul dalam t_2 dengan kategori diterima berpeluang sebesar 55%.
Jika nilai *interview* Bahasa Inggris lebih dari sama dengan 81, maka peluang diterima sebesar 0,89. Sedangkan jika nilai *interview* Bahasa Inggris kurang dari sama dengan 81, maka peluang diterima sebesar 0,11.
3. Pada simpul dalam t_5 dengan kategori diterima berpeluang sebesar 10%.
Jika nilai *interview* Bahasa Inggris lebih dari sama dengan 79, maka peluang diterima sebesar 0,71. Sedangkan jika nilai *interview* Bahasa Inggris kurang dari sama dengan 79, maka peluang diterima sebesar 0,29.
4. Pada simpul terminal t_4 dengan kategori diterima berpeluang sebesar 43%.
Jika nilai *interview* Bahasa Inggris lebih dari sama dengan 81, maka peluang diterima sebesar 0,95. Sedangkan jika nilai *interview* Bahasa Inggris kurang dari sama dengan 81, maka peluang diterima sebesar 0,05.
5. Pada simpul terminal t_{10} dengan kategori diterima berpeluang sebesar 13%.
Jika nilai *interview* Bahasa Inggris lebih dari sama dengan 79, maka peluang diterima sebesar 0,85. Sedangkan jika nilai *interview* Bahasa Inggris kurang dari sama dengan 79, maka peluang diterima sebesar 0,15.
6. Pada simpul terminal t_3 dengan kategori tidak diterima berpeluang sebesar 45%.
Jika nilai *interview* komitmen lebih dari sama dengan 7,5, maka peluang tidak diterima sebesar 0,00. Jika nilai *interview* komitmen kurang dari 7,5, maka peluang tidak diterima sebesar 1,00.
7. Pada simpul terminal t_{11} dengan kategori tidak diterima berpeluang sebesar 3%.
Jika nilai *interview* Bahasa Inggris kurang dari 79, maka peluang tidak diterima sebesar 0,82. Jika nilai *interview* Bahasa Inggris lebih dari 79, maka peluang tidak diterima sebesar 0,18.

4.2.3 Tingkat Akurasi Ketepatan Klasifikasi CART

Ketepatan pohon klasifikasi yang telah terbentuk pada penerimaan musyrif dan musyrifah untuk menentukan diterima dan tidak diterimanya calon musyrif dan musyrifah dapat dilihat dari nilai *Apparent Error Rate* (APER). Juga dapat dilihat dari tingkat *sensitivity* untuk melihat tingkat akurasi model penentuan diterimanya musyrif dan musyrifah dan *specificity* untuk melihat tingkat akurasi model penentuan tidak diterimanya musyrif dan musyrifah.

Berdasarkan pohon klasifikasi CART pada Gambar 4.4, diperoleh perhitungan tingkat akurasi model menggunakan data *testing*. Berikut ini perhitungan tingkat akurasi ketepatan klasifikasinya dapat dilihat pada Tabel 4.10.

Tabel 4.10 Hasil Ketepatan Klasifikasi Pohon Klasifikasi

Observasi	Prediksi		Total
	Diterima	Tidak Diterima	
Diterima	49	4	53
Tidak Diterima	1	48	49
Total	50	52	102

Tabel 4.10 diketahui bahwa pada pohon klasifikasi CART dengan total jumlah data sampel sebanyak 102 orang, data yang diprediksi diterima ada sebanyak 49 orang sampel yang diterima dan 1 orang sampel lainnya disklasifikasi. Sedangkan pada data yang diprediksi tidak diterima ada 48 orang dikategorikan tidak diterima dan 4 orang lainnya yang disklasifikasi. Perhitungan tingkat akurasi menggunakan $1 - APER$, *sensitivity*, dan *specificity* pada pohon klasifikasi penerimaan musyrif dan musyrifah yaitu:

$$\begin{aligned}
 APER &= \left(1 - \frac{n_{11} + n_{22}}{N}\right) \times 100\% \\
 &= \left(1 - \frac{49 + 48}{102}\right) \times 100\% \\
 &= (1 - 0,9509) \times 100\% \\
 &= 0,0495 \times 100\% = 4,95\%
 \end{aligned}$$

$$Accuracy = (1 - APER) \times 100\% = 0.9505 \times 100\% = 95.05\%$$

$$Sensitivity = \frac{n_{11}}{N_1} \times 100\% = \frac{49}{50} \times 100\% = 0,98 \times 100\% = 98\%$$

$$Specificity = \frac{n_{22}}{N_2} \times 100\% = \frac{48}{52} \times 100\% = 0,9231 \times 100\% = 92,31\%$$

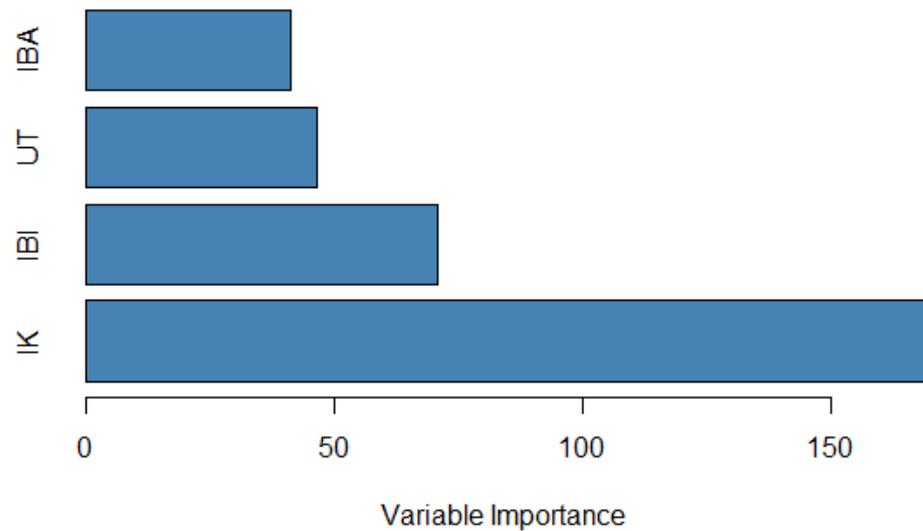
Berdasarkan perhitungan di atas diperoleh nilai *APER* sebesar 4,91% dengan nilai akurasi sebesar 0,9509 atau 95,09% dengan nilai *sensitivity* sebesar 0,98 atau 98% dan nilai *specificity* sebesar 0,9231 atau 92,31%. Tingkat akurasi tersebut menunjukkan bahwa CART sudah baik dalam mengklasifikasikan seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al-Jami'ah. Proses selanjutnya akan digunakan metode *bagging*.

Tabel 4.11 Nilai Ketepatan Klasifikasi CART

APER	Akurasi	<i>Sensitivity</i>	<i>Specificity</i>
4,91%	95.09%	98%	92,31%

4.2.4 Variabel Penting Pada CART

Berdasarkan proses klasifikasi CART pada seleksi penerimaan musyrif dan musyrifah diketahui variabel prediktor sebagai *variable importance*. Tingkat pengaruh variabel pada pohon klasifikasi dapat dilihat melalui bar plot pada gambar 4.5 dibawah ini.



Gambar 4.5 Bar Plot *Variable Importance* Pada CART

Perhitungan persentase tingkat pengaruh variabel prediktor pada klasifikasi penerimaan musyrif dan musyrifah sebagai berikut:

$$VarImp = var \times N \times 100\%$$

1. *Interview* Komitmen

$$VarImp = 0,3321 \times 100\% = 33,21\%$$

2. *Interview* Bahasa Inggris

$$VarImp = 0,1383 \times 100\% = 13,83\%$$

3. Ujian Tulis

$$VarImp = 0,0907 \times 100\% = 9,07\%$$

4. *Interview* Bahasa Arab

$$VarImp = 0,0809 \times 100\% = 8,09\%$$

Tabel 4.12 *Variable Importance* Pada Klasifikasi CART

Variabel	Overall	Tingkat Kepentingan (%)
Interview Komitmen	169.68145	33,21%
Interview Bahasa Inggris	70.66685	13,83%
Ujian Tulis	46.28004	9,07%
Interview Bahasa Arab	41.33051	8,09%

Sehingga dari perhitungan di atas, diketahui bahwa indikator *interview* komitmen menjadi variabel yang sangat berpengaruh, karena tingkat kepentingannya sebesar 33,21% dari keseluruhan nilai pada data seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al-Jami'ah.

4.3 Analisis *Bagging* CART

Metode *bootstrap aggregating* (*bagging*) telah dilakukan untuk meningkatkan akurasi pada CART yang sudah diketahui sebelumnya. *Resampling* yang digunakan pada *bagging* telah dilakukan sebanyak 150 kali dan dihasilkan variabel prediktor penting pada seleksi penerimaan musyrif dan musyrifah.

4.3.1 Tingkat Akurasi Ketepatan Klasifikasi *Bagging* CART

Ketepatan pohon klasifikasi yang telah terbentuk pada penerimaan musyrif dan musyrifah untuk menentukan diterima dan tidak diterimanya calon musyrif dan musyrifah dapat dilihat dari nilai *Apparent Error Rate* (APER). Juga dapat dilihat dari tingkat *sensitivity* untuk melihat tingkat akurasi model penentuan diterimanya musyrif dan musyrifah dan *specificity* untuk melihat tingkat akurasi model penentuan tidak diterimanya musyrif dan musyrifah.

Setelah dilakukan *bagging* dengan *resampling* sebanyak 150 kali, yang dianggap memiliki tingkat akurasi yang tinggi diperoleh prediksi pada klasifikasi *bagging* menggunakan data *testing*. Berikut ini perhitungan tingkat akurasi ketepatan klasifikasinya dapat dilihat pada tabel 4.13.

Tabel 4.13 Hasil Ketepatan Klasifikasi *Bagging* CART

Observasi	Prediksi		Total
	Diterima	Tidak Diterima	
Diterima	49	3	52
Tidak Diterima	1	49	50
Total	50	52	102

Berdasarkan Tabel 4.13, diketahui bahwa pada analisis *bootstrap aggregating* CART diperoleh hasil prediksi pada kategori diterima sebanyak 49 data yang dapat diklasifikasikan diterima dan 1 data yang disklasifikasi atau tidak diterima. Sedangkan pada data kategori tidak diterima terdapat sebanyak 49 data yang diklasifikasikan tidak diterima dan 3 lainnya disklasifikasi. Berikut adalah hasil ketepatan klasifikasi *bootstrap aggregating* CART.

Perhitungan tingkat akurasi *bagging* menggunakan $1 - APER$, *sensitivity*, dan *specificity* pada pohon klasifikasi penerimaan musyrif dan musyrifah yaitu sebagai berikut:

$$\begin{aligned} APER &= \left(1 - \frac{n_{11} + n_{22}}{n}\right) \times 100\% \\ &= \left(1 - \frac{49 + 49}{102}\right) \times 100\% \\ &= (1 - 0,9607) \times 100\% \\ &= 0,0393 \times 100\% = 3,93\% \end{aligned}$$

$$Accuracy = 1 - APER = 1 - 3,93\% = 96,07\%$$

$$Sensitivity = \frac{n_{11}}{N_1} \times 100\% = \frac{49}{50} \times 100\% = 0,98 \times 100\% = 98\%$$

$$Specificity = \frac{n_{22}}{N_2} \times 100\% = \frac{49}{52} \times 100\% = 0,9423 \times 100\% = 94,23\%$$

Berdasarkan perhitungan tingkat akurasi menggunakan $1 - APER$ pada pohon klasifikasi penerimaan musyrif dan musyrifah diperoleh nilai akurasi Pada tabel 4.7 dapat diketahui bahwa dari model *bagging* CART diperoleh nilai *APER* sebesar 3,93% sehingga tingkat akurasi menggunakan $1 - APER$ sebesar 0,9607 atau 96,07%. Nilai *sensitivity* diperoleh sebesar 0,98 atau 98% dan nilai *specificity* sebesar 0,9423 atau 94,23% yang artinya perhitungan *bagging* lebih akurat untuk

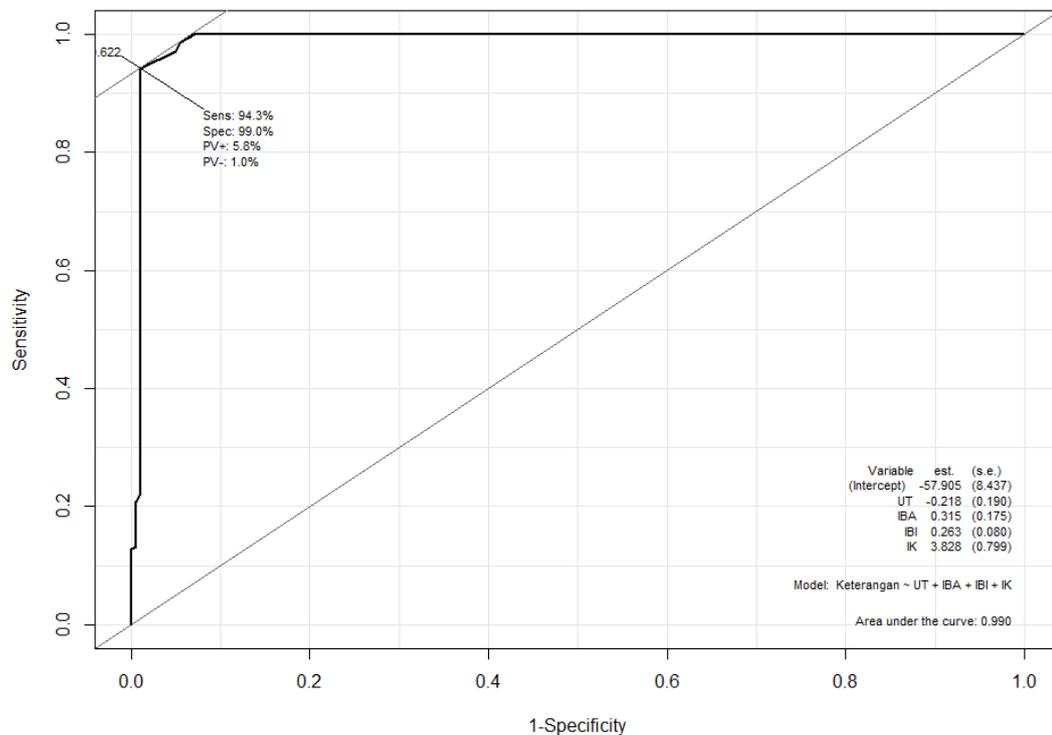
ketepatan klasifikasi seleksi penerimaan musyrif dan musrifah Pusat Ma'had Al-Jami'ah UIN Malang.

Tabel 4.14 Nilai Akurasi Ketepatan Klasifikasi *Bagging* CART

APER	Akurasi	Sensitivity	Specificity
3,93%	96,07%	98%	94,23 %

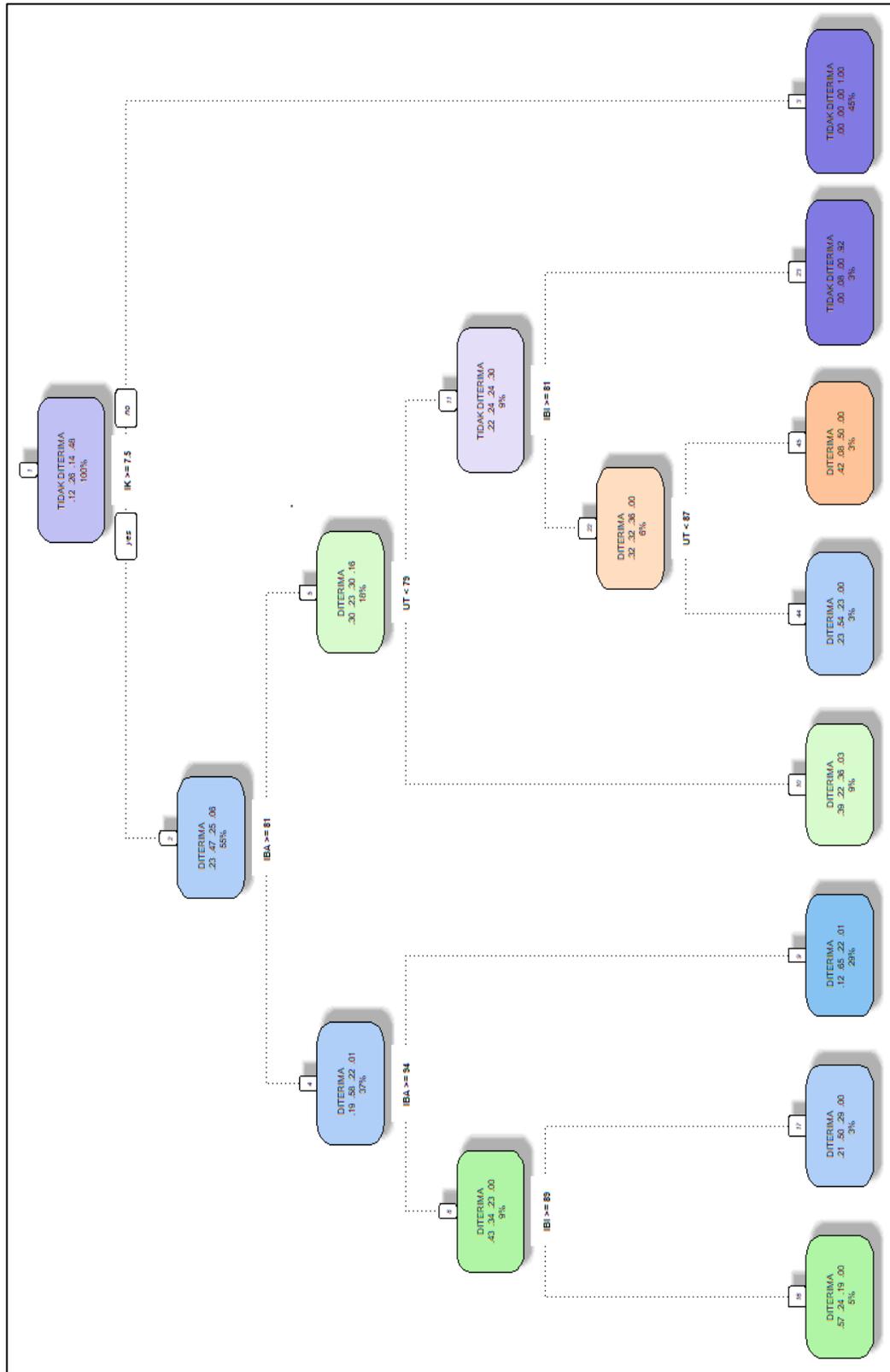
Dapat dilihat penggunaan replikasi *bagging* CART pada data penerimaan musyrif dan musyrifah di Pusat Ma'had Al Jami'ah dapat meningkatkan ketepatan akurasi sebesar 1,02% dibanding metode CART.

Perbandingan tingkat akurasi dari CART dan *bagging* CART dapat dilihat melalui kurva ROC di bawah ini.



Gambar 4.6 Kurva ROC

4.3.2 Pohon Klasifikasi Setelah *Bagging* CART

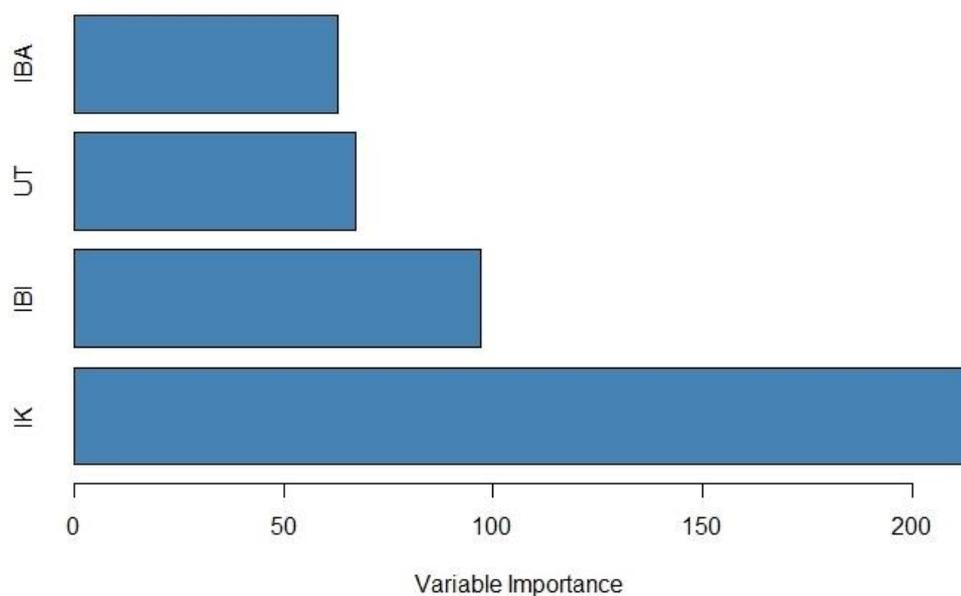


Gambar 4.7 Pohon Klasifikasi *Bagging* CART Penerimaan Musyrif Musyriyah

Gambar 4.7 diperoleh pohon klasifikasi dari *bagging* CART dengan simpul induk (t_1) yang hasilnya tidak diterima dengan indikator *interview* komitmen, karena data yang menyatakan tidak diterima lebih banyak daripada pohon yang menyatakan diterima juga sama seperti pohon klasifikasi pada CART. Pada pohon klasifikasi *bagging* Gambar 4.6, diketahui untuk simpul dalamnya lebih rinci daripada pohon CART dengan kriteria empat variabel. Pada pohon klasifikasi Gambar 4.6, diketahui ada 6 simpul dalam yaitu $t_2, t_4, t_5, t_8, t_{11}, t_{22}$ dan 8 simpul terminal yaitu $t_3, t_9, t_{10}, t_{17}, t_{18}, t_{23}, t_{44},$ dan t_{45} . Pada pohon *bagging*, urutan variabel yang berpengaruh pada proses pengklasifikasian adalah nilai *interview* komitmen, nilai *interview* Bahasa Arab, nilai *interview* Bahasa Inggris dan ujian tulis. Sehingga dapat diketahui bahwa pengklasifikasian berubah pada CART setelah dilakukan *bagging* lebih memberikan hasil yang lebih rinci.

4.4 Variabel Penting Pada Klasifikasi *Bagging* CART

Berdasarkan proses klasifikasi *bagging* CART pada seleksi penerimaan musyrif dan musyrifah diperoleh tingkat kepentingan variabel prediktor sebagai indikator yang penting sehingga dapat mempengaruhi hasil seleksi penerimaan musyrif dan musyrifah. Tingkat pengaruh variabel setelah *bagging* dengan nilai *overall* dapat dilihat melalui bar plot pada Gambar 4.8.



Gambar 4.8 Bar Plot *Variabel Importance Bagging CART*

Perhitungan persentase tingkat pengaruh variabel prediktor pada klasifikasi penerimaan musyrif dan musyrifah sebagai berikut.

$$VarImp = var \times N \times 100\%$$

1. *Interview* Komitmen

$$VarImp = 0,4214 \times 100\% = 42,14\%$$

2. *Interview* Bahasa Inggris

$$VarImp = 0,1952 \times 100\% = 19,52\%$$

3. Ujian Tulis

$$VarImp = 0,1336 \times 100\% = 13,36\%$$

4. *Interview* Bahasa Arab

$$VarImp = 0,1238 \times 100\% = 12,38\%$$

Tabel 4.15 Tingkat Kepentingan Variabel *Bagging* CART

Variabel	Overall	Tingkat Kepentingan(%)
<i>Interview</i> Komitmen	215.34065	42,14%
<i>Interview</i> Bahasa Inggris	99.74735	19,52%
Ujian Tulis	68.26003	13,36%
<i>Interview</i> Bahasa Arab	63.28216	12.38%

Tabel 4.15 dan perhitungan di atas menunjukkan bahwa indikator *interview* komitmen menjadi variabel yang sangat berpengaruh setelah dilakukan *bagging*, karena tingkat kepentingannya sebesar 42,14% dari keseluruhan nilai pada data seleksi penerimaan musyrif dan musyrifah Pusat Ma'had Al-Jami'ah. Sehingga tidak ada perubahan pada tingkat kepentingan variabel pada pohon klasifikasi setelah dilakukan *bagging*, hanya saja tingkat kepentingannya naik sebesar 9,02% dibanding CART.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan klasifikasi penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah diperoleh kesimpulan sebagai berikut:

1. Tingkat akurasi ketepatan klasifikasi pada pengklasifikasian penerimaan musyrif dan musyrifah Pusat Ma'had Al Jami'ah UIN Maulana Malik Ibrahim Malang dengan menggunakan metode *Bootstrap Aggregating (Bagging) Classification Analysis Regression Trees* sebesar 0,9607, atau 96,07% dengan *sensitivity* sebesar 0,98 atau 98% dan diperoleh nilai *specificity* sebesar 0,9423 atau 94,23% serta nilai APER-nya sebesar 3,93%. *Bagging* CART terbukti dapat meningkatkan tingkat akurasi CART pada data penerimaan musyrif dan musyrifah sebesar 1,02%. Nilai tingkat akurasi total tergolong tinggi, sehingga dapat dikatakan bahwa pohon klasifikasi sesuai dan layak digunakan untuk mengklasifikasikan data baru.
2. Berdasarkan analisis *bagging* CART yang telah dilakukan 150 kali *resampling*. Terdapat *variable importance* yang mempengaruhi pohon klasifikasi diantaranya yaitu nilai *interview* komitmen merupakan variabel yang memberikan pengaruh terbesar dengan persentase tingkat kepentingan 42,14% dari data keseluruhan. Kedua, yaitu nilai *interview* Bahasa Inggris dengan persentase 19,52%, nilai ujian tulis dengan persentase 13,36% dan yang memiliki pengaruh terkecil dalam pengklasifikasian yaitu ujian nilai *interview* Bahasa Arab dengan persentase sebesar 12,38% dari keseluruhan data.

5.2 Saran

Saran bagi peneliti selanjutnya yang menggunakan metode dan penerapan yang sama dapat memberikan kombinasi baru untuk pembagian data *training* dan data *testing*. Pada penerapan metode *bagging* CART dapat menambahkan *resampling* yang lebih banyak agar didapatkan variasi yang lebih banyak dan hasil yang berbeda dari penelitian sebelumnya. Selain itu, peneliti dapat mengklasifikasikan pohon klasifikasi dari data penerimaan beberapa tahun sebelumnya agar didapatkan karakteristik yang lebih akurat dari klasifikasi, karena untuk kasus penerimaan musyrif dan musyrifah selalu berubah ubah setiap tahunnya dan nilai dari setiap calon musyrif dan musyrifah berbeda-beda yang mungkin akan mengakibatkan hasil dari pohon klasifikasi juga berbeda.

DAFTAR PUSTAKA

- Agwil, W., Fransiska, H., & Hidayati, N. (2020). Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging CART. *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika*, 6(1), 155-166.
- Andriyashin, A. (2005, Maret 24). Financial Applications of Classification and Regression Trees. *CASE - Center of Applied Statistics and Economics*, p. 7.
- Arrahimi, A. R., Ihsan, M. K., Kartini, D., Faisal, M. R., & Indriani, F. (2019). Teknik Bagging Dan Boosting Pada Algoritma CART untuk Klasifikasi Masa Studi Mahasiswa. *Jurnal Sains dan Informatika*, 6(2), 21-30.
- Bobbitt, Z. (2020, November 23). *How to Perform Bagging in R (Step-by-Step)*. Retrieved Juni 11, 2022, from Statistics Simplified Statology: <https://www.statology.org/bagging-in-r/>
- Breiman, J. H. (1998). *Classification and Regression Trees*. New York Washington, D.C: CRC Press Print.
- Huang, Y., Cai, J., Ji, L., & Li, Y. (2004). Classifying G-protein coupled receptors with bagging classification tree. *Computational Biology and Chemistry*, 28(1), 275-280.
- Learn Quran. (2018, Januari 16). *Learn Quran Tafsir*. Retrieved from Learn Quran Tafsir: <https://tafsir.learn-quran.co>.
- Lewis, R. J. (2000). *An Introduction to Classification and Regression Tree (CART) Analysis*. Fransisco, California: UCLA Medical Center.
- Musu, W., Ibrahim, A., & Heriadi. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5. *PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, X(1), 186-195.
- Nair, A., Kuban, B. D., Obuchowsk, N., & Vince, D. G. (2001). Assesing Spectral Algorithms To Predict Atherosclerotic Plaque Composition with Normalized and Raw Intravascular Ultrasound Data. *Ultrasound in Medicine and Biology*, 27(10), 1319-1331.
- Pusat Ma'had Al Jami'ah. (2011, November 30). *Profil Ma'had*. Retrieved Januari 03, 2022, from Pusat Ma'had Al Jami'ah: <https://msaa.uin-malang.ac.id/>
- Pusat Ma'had Al Jami'ah. (2019, Maret 29). *Rekrutmen Musyrif/ah, Ta'mir dan Santri Tahfidz 2019/2020*. Retrieved Januari 03, 2022, from Pusat Ma'had Al Jami'ah: <https://msaa.uin-malang.ac.id/>
- Qur'an Kemenag. (2016, Agustus 25). *Qur'an Kemenag*. Retrieved Januari 30, 2022, from Qur'an Kemenag: <https://quran.kemenag.go.id/>

- Richard, A., Wichern, & Johnson, D. W. (2007). *Applied Logistic Regression*. New Jersey: Pearson Prentice Hall.
- Rokach, L. (2008). *Ensemble Methods For Classifiers dalam Data Mining and Knowledge Discovery Handbook*. New York: Springer Science+Business Media.
- Sartono, B., & Syafitri, U. D. (2010). Metode Pohon Gabungan: Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal. *Forum Statistika dan Komputasi*, 15(1), 1-7.
- Suniantara, I. P., & Rusli, M. (2017). Ketepatan Klasifikasi Bagging CART Pada Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa STIKOM Bali. *Konferensi Nasional Sistem & Informatika 2017*, 237-240.

LAMPIRAN

Lampiran 1: Data Penerimaan Musyrif dan Musyrifah Pusat Ma'had Al-Jami'ah

Data Penelitian

No	Ujian Tulis	<i>Interview</i> Bahasa Arab	<i>Interview</i> Bahasa Inggris	<i>Interview</i> Komitmen	Keterangan
1	90	92	88	8	DITERIMA
2	88	88	90	8	TIDAK DITERIMA
3	70	75	72	8	TIDAK DITERIMA
4	86	88	78	8	DITERIMA
5	95	80	84	6	TIDAK DITERIMA
6	75	70	75	8	DITERIMA
7	80	85	85	7	TIDAK DITERIMA
8	90	96	92	8	DITERIMA
9	92	94	90	8	DITERIMA
10	63	80	75	8	DITERIMA
11	75	80	70	8	TIDAK DITERIMA
12	65	70	78	6	TIDAK DITERIMA
13	85	90	85	8	DITERIMA
14	98	100	96	8	DITERIMA
15	96	98	88	8	DITERIMA
16	60	75	70	9	DITERIMA
17	88	80	85	7	TIDAK DITERIMA
18	90	90	88	7	TIDAK DITERIMA
19	78	80	78	8	TIDAK DITERIMA
20	90	92	88	6	TIDAK DITERIMA
21	88	88	90	9	DITERIMA
22	70	75	72	9	DITERIMA
23	86	88	78	8	DITERIMA
24	95	80	84	8	DITERIMA
25	75	70	75	9	DITERIMA
26	80	85	85	7	TIDAK DITERIMA
27	90	96	92	8	DITERIMA
28	92	94	90	9	DITERIMA
29	63	80	75	8	DITERIMA

30	75	80	70	7	TIDAK DITERIMA
31	65	70	78	6	TIDAK DITERIMA
32	85	90	85	8	TIDAK DITERIMA
33	98	100	96	6	TIDAK DITERIMA
34	96	98	88	8	DITERIMA
35	60	75	70	9	DITERIMA
36	88	80	85	9	DITERIMA
37	90	90	88	7	TIDAK DITERIMA
38	78	80	78	9	DITERIMA
39	75	70	75	8	DITERIMA
40	80	85	85	8	DITERIMA
41	90	96	92	8	TIDAK DITERIMA
42	92	94	90	9	DITERIMA
43	63	80	75	8	DITERIMA
44	75	80	70	8	DITERIMA
45	65	70	78	8	DITERIMA
46	85	90	85	9	DITERIMA
47	98	100	96	9	DITERIMA
48	96	98	88	6	TIDAK DITERIMA
49	60	75	70	7	TIDAK DITERIMA
50	88	80	85	9	DITERIMA
51	90	90	88	7	TIDAK DITERIMA
52	78	80	78	8	DITERIMA
53	85	90	85	7	TIDAK DITERIMA
54	98	100	96	8	DITERIMA
55	96	98	88	7	TIDAK DITERIMA
56	60	75	70	8	DITERIMA
57	88	80	85	6	TIDAK DITERIMA
58	90	90	88	6	TIDAK DITERIMA
59	78	80	78	8	DITERIMA
60	75	70	75	9	DITERIMA
61	80	85	85	9	DITERIMA
62	90	96	92	8	DITERIMA
63	92	94	90	9	DITERIMA
64	63	80	75	8	DITERIMA
65	75	80	70	5	TIDAK DITERIMA
66	65	70	78	5	TIDAK DITERIMA

67	85	90	85	5	TIDAK DITERIMA
68	98	100	96	8	DITERIMA
69	96	98	88	8	DITERIMA
70	60	75	70	5	TIDAK DITERIMA
71	88	80	85	8	DITERIMA
72	90	90	88	7	TIDAK DITERIMA
73	78	80	78	8	DITERIMA
74	78	75	80	8	TIDAK DITERIMA
75	86	80	90	9	DITERIMA
76	84	80	75	7	TIDAK DITERIMA
77	90	90	84	9	DITERIMA
78	84	80	75	6	TIDAK DITERIMA
79	80	80	85	9	DITERIMA
80	82	80	87	8	DITERIMA
81	88	85	87	8	DITERIMA
82	78	75	80	7	TIDAK DITERIMA
83	86	85	85	8	DITERIMA
84	82	78	80	6	TIDAK DITERIMA
85	88	80	90	9	DITERIMA
86	86	80	75	7	TIDAK DITERIMA
87	88	85	78	6	TIDAK DITERIMA
88	80	80	70	7	TIDAK DITERIMA
89	92	94	88	9	DITERIMA
90	78	75	80	8	TIDAK DITERIMA
91	84	80	75	6	TIDAK DITERIMA
92	84	80	75	7	TIDAK DITERIMA
93	78	75	80	5	TIDAK DITERIMA
94	90	90	85	9	DITERIMA
95	82	78	80	6	TIDAK DITERIMA
96	86	80	75	7	TIDAK DITERIMA
97	88	85	78	5	TIDAK DITERIMA
98	80	80	70	6	TIDAK DITERIMA
99	78	75	80	7	TIDAK DITERIMA

100	84	80	75	7	TIDAK DITERIMA
101	88	90	87	8	DITERIMA
102	92	88	85	9	DITERIMA
103	84	80	75	7	TIDAK DITERIMA
104	78	75	80	6	TIDAK DITERIMA
105	88	90	87	8	DITERIMA
106	88	82	86	9	DITERIMA
107	92	92	88	9	DITERIMA
108	96	93	90	9	DITERIMA
109	90	88	94	9	DITERIMA
110	92	90	88	9	DITERIMA
111	90	88	90	9	DITERIMA
112	88	85	88	8	DITERIMA
113	90	88	85	9	DITERIMA
114	86	84	87	8	DITERIMA
115	86	84	88	9	DITERIMA
116	88	82	85	8	DITERIMA
117	90	88	90	9	DITERIMA
118	90	86	88	9	DITERIMA
119	92	92	90	9	DITERIMA
120	94	90	88	9	DITERIMA
121	90	88	85	8	DITERIMA
122	88	85	88	9	DITERIMA
123	86	85	87	8	DITERIMA
124	90	90	87	8	DITERIMA
125	86	85	87	9	DITERIMA
126	92	90	95	9	DITERIMA
127	88	86	90	8	DITERIMA
128	90	90	87	8	DITERIMA
129	82	78	80	6	TIDAK DITERIMA
130	90	88	86	8	DITERIMA
131	86	80	95	8	DITERIMA
132	86	82	85	8	DITERIMA
133	86	85	85	8	DITERIMA
134	82	80	82	8	DITERIMA
135	88	86	88	8	DITERIMA
136	90	86	90	8	DITERIMA
137	86	80	75	6	TIDAK DITERIMA
138	92	90	88	9	DITERIMA
139	90	88	92	9	DITERIMA
140	92	94	88	9	DITERIMA
.
511	94	85	87	8	DITERIMA

Lampiran 2: Statistik Deskriptif Variabel Prediktor

Statistika Deskriptif Nilai Ujian Tulis

Descriptive Statistics: Ujian Tulis

Variable	Keterangan	N	N*	Mean	SE Mean	StDev	Minimum	Q1
Ujian Tulis	Diterima	249	0	88,562	0,204	3,224	80,000	86,000
	Tidak Diterima	262	0	82,565	0,218	3,534	78,000	79,500

Variable	Keterangan	Median	Q3	Maximum
Ujian Tulis	Diterima	88,000	90,000	96,000
	Tidak Diterima	84,000	86,000	94,000

Statistika Deskriptif Nilai *Interview* Bahasa Arab

Descriptive Statistics: Interview Bahasa Arab

Variable	Keterangan	N	N*	Mean	SE Mean	StDev	Minimum
Interview Bahasa Arab	Diterima	249	0	86,618	0,264	4,159	80,000
	Tidak Diterima	262	0	79,240	0,201	3,247	75,000

Variable	Keterangan	Q1	Median	Q3	Maximum
Interview Bahasa Arab	Diterima	84,000	87,000	90,000	95,000
	Tidak Diterima	77,250	80,000	80,000	95,000

Statistika Deskriptif Nilai *Interview* Bahasa Inggris

Descriptive Statistics: Interview Bahasa Inggris

Variable	Keterangan	N	N*	Mean	SE Mean	StDev
Interview Bahasa Inggris	Diterima	249	0	86,940	0,182	2,867
	Tidak Diterima	262	0	76,740	0,219	3,546

Variable	Keterangan	Minimum	Q1	Median	Q3
Interview Bahasa Inggris	Diterima	75,000	85,000	87,000	88,000
	Tidak Diterima	70,000	75,000	78,000	80,000

Variable	Keterangan	Maximum
Interview Bahasa Inggris	Diterima	95,000
	Tidak Diterima	90,000

Statistika Deskriptif Nilai *Interview* Komitmen

Descriptive Statistics: Interview Komitmen

Variable	Keterangan	N	N*	Mean	SE Mean	StDev	Minimum
Interview Komitmen	Diterima	249	0	8,4739	0,0346	0,5465	5,0000
	Tidak Diterima	262	0	6,5267	0,0556	0,8998	5,0000

Variable	Keterangan	Q1	Median	Q3	Maximum
Interview Komitmen	Diterima	8,0000	8,0000	9,0000	9,0000
	Tidak Diterima	6,0000	7,0000	7,0000	9,0000

Lampiran 3: Kombinasi Pembagian Data Training Dan Testing Pada Presentase Data

Script dan Output Pembagian Data Training Dan Testing Pada Presentase Data

1. Presentase data 75% dan 25%

```
#membuat data training dan posting
#presentasse data 75% dan 25%
n=round(nrow(data)*0.75);n
set.seed(99191)
sampel=sample(1:nrow(data),n)
data.training=data[sampel,]
data.testing=data[-sampel,]
dim(data.training)
dim(data.testing)
set.seed(99191)
train = sample(1:nrow(data), 383)
data.train=data[train,]
data.test=data[-train,]

fit.tree = rpart(Keterangan ~ ., data=data.train, method = "class", cp=0.008)
fit.tree
n= 383

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 383 188 TIDAK DITERIMA (0.49086162 0.50913838)
2) Interview.Komitmen>=7.5 210 22 DITERIMA (0.89523810 0.10476190)
4) Interview.Bahasa.Arab>=79 185 12 DITERIMA (0.93513514 0.06486486) *
5) Interview.Bahasa.Arab< 79 25 10 DITERIMA (0.60000000 0.40000000)
10) Ujian.Tulis< 76.5 16 1 DITERIMA (0.93750000 0.06250000) *
11) Ujian.Tulis>=76.5 9 0 TIDAK DITERIMA (0.00000000 1.00000000) *
3) Interview.Komitmen< 7.5 173 0 TIDAK DITERIMA (0.00000000 1.00000000) *
> |
```

2. Presentase data 80% dan 20%

```
#membuat data training dan posting
n=round(nrow(data)*0.80);n
set.seed(99191)
sampel=sample(1:nrow(data),n)
data.training=data[sampel,]
data.testing=data[-sampel,]
dim(data.training)
dim(data.testing)
set.seed(99191)
train = sample(1:nrow(data), 409)
data.train=data[train,]
data.test=data[-train,]

fit.tree = rpart(Keterangan ~ ., data=data.train, method = "class", cp=0.008)
fit.tree
n= 409

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 409 202 TIDAK DITERIMA (0.49388753 0.50611247)
2) Interview.Komitmen>=7.5 226 24 DITERIMA (0.89380531 0.10619469)
4) Interview.Bahasa.Ingggris>=81 174 9 DITERIMA (0.94827586 0.05172414) *
5) Interview.Bahasa.Ingggris< 81 52 15 DITERIMA (0.71153846 0.28846154)
10) Interview.Bahasa.Ingggris< 79 41 6 DITERIMA (0.85365854 0.14634146) *
11) Interview.Bahasa.Ingggris>=79 11 2 TIDAK DITERIMA (0.18181818 0.81818182) *
3) Interview.Komitmen< 7.5 183 0 TIDAK DITERIMA (0.00000000 1.00000000) *
> |
```

3. Presentase data 85% dan 25%

```
#presentase data 85% dan 25%
n=round(nrow(data)*0.85);n
set.seed(99191)
sampel=sample(1:nrow(data),n)
data.training=data[sampel,]
data.testing=data[-sampel,]
dim(data.training)
dim(data.testing)
set.seed(99191)
train = sample(1:nrow(data), 434)
data.train=data[train,]
data.test=data[-train,]

fit.tree = rpart(Keterangan ~ ., data=data.train, method = "class", cp=0.008)
fit.tree

n= 434

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 434 213 TIDAK DITERIMA (0.49078341 0.50921659)
 2) Interview.Komitmen>=7.5 237 24 DITERIMA (0.89873418 0.10126582)
   4) Interview.Bahasa.Inggris>=81 185 9 DITERIMA (0.95135135 0.04864865) *
   5) Interview.Bahasa.Inggris< 81 52 15 DITERIMA (0.71153846 0.28846154)
     10) Interview.Bahasa.Inggris< 79 41 6 DITERIMA (0.85365854 0.14634146) *
     11) Interview.Bahasa.Inggris>=79 11 2 TIDAK DITERIMA (0.18181818 0.81818182) *
  3) Interview.Komitmen< 7.5 197 0 TIDAK DITERIMA (0.00000000 1.00000000) *
```

4. Presentase data 90% dan 10%

```
#presentase data 90% dan 10%
n=round(nrow(data)*0.90);n
set.seed(99191)
sampel=sample(1:nrow(data),n)
data.training=data[sampel,]
data.testing=data[-sampel,]
dim(data.training)
dim(data.testing)
set.seed(99191)
train = sample(1:nrow(data), 460)
data.train=data[train,]
data.test=data[-train,]

fit.tree = rpart(Keterangan ~ ., data=data.train, method = "class", cp=0.008)
fit.tree

n= 460

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 460 226 TIDAK DITERIMA (0.4913043 0.5086957)
 2) Interview.Komitmen>=7.5 252 26 DITERIMA (0.8968254 0.1031746) *
  3) Interview.Komitmen< 7.5 208 0 TIDAK DITERIMA (0.0000000 1.0000000) *
```

Lampiran 4: Pembentukan Pohon Klasifikasi

Script Pohon Klasifikasi

```
•  
#klasifikasi dengan decision tree  
fit=rpart(Keterangan~Ujian.Tulis+Interview.Bahasa.Arab+Interview.Bahasa.Inggri  
summary(fit)  
  
#plot decision tree  
library(rattle)  
fancyRpartPlot(fit)  
  
#prediksi  
prediksi=predict(fit, newdata = data.testing, type = "class")  
#matriksi  
table(prediksi, data.testing$Keterangan)  
akurasi=(49+48)/(102)  
akurasi  
sensitivity=49/50  
sensitivity  
specificity=48/52  
specificity
```

Output Prediksi CART

```
> table(prediksi, data.testing$Keterangan)  
  
prediksi      DITERIMA  TIDAK DITERIMA  
DITERIMA      49         4  
TIDAK DITERIMA 1         48  
> akurasi=(49+48)/(102)  
> akurasi  
[1] 0.9509804  
> sensitivity=49/50  
> sensitivity  
[1] 0.98  
> specificity=48/52  
> specificity  
[1] 0.9230769
```

Lampiran 5: *Variable Importance* pada CART

Script Menentukan *Variable Importance*

```
#hitung variabel penting
VI<-data.frame(var=names(data[-1]),imp=varImp(fit.tree))
print(VI)

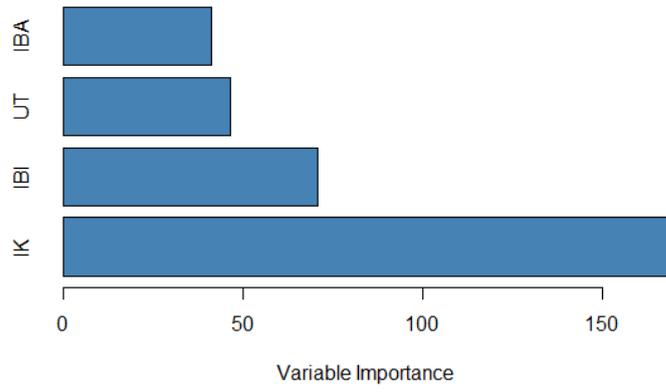
#sort variable importance descending
VI_plot<-VI[order(VI$Overall, decreasing = TRUE),]
VI_plot

#Visualize variable importance with horizontal bar plot
barplot(VI_plot$Overall,
        names.arg=rownames(VI_plot),
        horiz=TRUE,
        col='steelblue',
        xlab='Variable Importance')
```

Output *Variable Importance*

```
> #hitung variabel penting
> VI<-data.frame(var=names(data[-1]),imp=varImp(fit.tree))
> print(VI)
  var Overall
IBA   IBA  41.33051
IBI   IBI  70.66685
IK    IK  169.68145
UT Keterangan 46.28004
> #sort variable importance descending
> VI_plot<-VI[order(VI$Overall, decreasing = TRUE),]
> VI_plot
  var Overall
IK    IK  169.68145
IBI   IBI  70.66685
UT Keterangan 46.28004
IBA   IBA  41.33051
> #Visualize variable importance with horizontal bar plot
> barplot(VI_plot$Overall,
+         names.arg=rownames(VI_plot),
+         horiz=TRUE,
+         col='steelblue',
+         xlab='Variable Importance')
```

Bar Plot *Variable Importance*



Lampiran 6: Proses *Bagging* CART

Script *Bagging* CART

```
# Bagging CART
# load data
data=read.delim("Data.txt")
data1=as.factor(data$Keterangan)
data$Keterangan=data1
class(data$Keterangan)

#fit the bagged model
bag<-bagging(
  formula=Keterangan~.,
  data=data,
  nbagg=150,
  coob=TRUE,
  control=rpart.control(minsplit = 2, cp=0)
)

#Display fitted bagged Model
bag

#fit model
fit<-bagging(Keterangan~.,nbagg=150, data=data)
print(fit)
fit
```

Output *Bagging* CART

```
Bagging classification trees with 150 bootstrap replications

Call: bagging.data.frame(formula = Keterangan ~ ., data = data, nbagg = 150,
  coob = TRUE, control = rpart.control(minsplit = 2, cp = 0))

Out-of-bag estimate of misclassification error: 0.0528
```

Script Prediksi *Bagging* CART

```
# make predictions
predictions <- predict(fit, data[,1:4])
predictions
```

Output Prediksi Bagging CART

```
> # make predictions
> predictions <- predict(fit, newdata=data.testing)
> predictions
 [1] DITERIMA      TIDAK DITERIMA TIDAK DITERIMA DITERIMA      DITERIMA      TIDAK DITERIMA
 [7] DITERIMA      TIDAK DITERIMA TIDAK DITERIMA DITERIMA      DITERIMA      TIDAK DITERIMA
[13] DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA      TIDAK DITERIMA
[19] DITERIMA      TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA
[25] DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA
[31] DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA
[37] DITERIMA      DITERIMA      TIDAK DITERIMA TIDAK DITERIMA DITERIMA      DITERIMA
[43] TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA DITERIMA      TIDAK DITERIMA
[49] TIDAK DITERIMA TIDAK DITERIMA DITERIMA      DITERIMA      TIDAK DITERIMA TIDAK DITERIMA
[55] DITERIMA      TIDAK DITERIMA DITERIMA      TIDAK DITERIMA DITERIMA      TIDAK DITERIMA
[61] TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA DITERIMA      TIDAK DITERIMA DITERIMA
[67] DITERIMA      TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA
[73] DITERIMA      TIDAK DITERIMA DITERIMA      TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA
[79] TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA TIDAK DITERIMA
[85] TIDAK DITERIMA TIDAK DITERIMA DITERIMA      TIDAK DITERIMA DITERIMA      DITERIMA
[91] TIDAK DITERIMA DITERIMA      TIDAK DITERIMA DITERIMA      DITERIMA      DITERIMA
[97] DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA      DITERIMA
Levels: DITERIMA TIDAK DITERIMA
```

Lampiran 7: Hasil *Bagging* CART

Output Prediksi *Bagging* CART

No	Keterangan
1	DITERIMA
2	DITERIMA
3	TIDAK DITERIMA
4	DITERIMA
5	TIDAK DITERIMA
6	DITERIMA
7	TIDAK DITERIMA
8	DITERIMA
9	DITERIMA
10	DITERIMA
11	DITERIMA
12	TIDAK DITERIMA
13	DITERIMA
14	DITERIMA
15	DITERIMA
16	DITERIMA
17	TIDAK DITERIMA
18	TIDAK DITERIMA
19	DITERIMA
20	TIDAK DITERIMA
21	DITERIMA
22	DITERIMA
23	DITERIMA
24	DITERIMA
25	DITERIMA
26	TIDAK DITERIMA
27	DITERIMA
28	DITERIMA
29	DITERIMA
30	TIDAK DITERIMA
31	TIDAK DITERIMA
32	DITERIMA
33	TIDAK DITERIMA
34	DITERIMA
35	DITERIMA
36	DITERIMA
37	TIDAK DITERIMA
38	DITERIMA
39	DITERIMA
40	DITERIMA
41	DITERIMA
42	DITERIMA

43	DITERIMA
44	DITERIMA
45	DITERIMA
46	DITERIMA
47	DITERIMA
48	TIDAK DITERIMA
49	TIDAK DITERIMA
50	DITERIMA
51	TIDAK DITERIMA
52	DITERIMA
53	TIDAK DITERIMA
54	DITERIMA
55	TIDAK DITERIMA
56	DITERIMA
57	TIDAK DITERIMA
58	TIDAK DITERIMA
59	DITERIMA
60	DITERIMA
61	DITERIMA
62	DITERIMA
63	DITERIMA
64	DITERIMA
65	TIDAK DITERIMA
66	TIDAK DITERIMA
67	TIDAK DITERIMA
68	DITERIMA
69	DITERIMA
70	TIDAK DITERIMA
71	DITERIMA
72	TIDAK DITERIMA
73	DITERIMA
74	TIDAK DITERIMA
75	DITERIMA
76	TIDAK DITERIMA
77	DITERIMA
78	TIDAK DITERIMA
79	DITERIMA
80	DITERIMA
81	DITERIMA
82	TIDAK DITERIMA
83	DITERIMA
84	TIDAK DITERIMA
85	DITERIMA

86	TIDAK DITERIMA
87	TIDAK DITERIMA
88	TIDAK DITERIMA
89	DITERIMA
90	TIDAK DITERIMA
91	TIDAK DITERIMA
92	TIDAK DITERIMA
93	TIDAK DITERIMA
94	TIDAK DITERIMA
95	DITERIMA
96	TIDAK DITERIMA
97	TIDAK DITERIMA
98	TIDAK DITERIMA
99	TIDAK DITERIMA
100	TIDAK DITERIMA
101	DITERIMA
102	DITERIMA

Script Tabel Prediksi Bagging CART

```
# summarize accuracy
table(predictions, data.testing$Keterangan)
akurasi=(49+49)/(102)
akurasi
sensitivity=49/50
sensitivity
specificity=49/52
specificity
```

Output Tabel Prediksi Bagging CART

```
> # summarize accuracy
> table(predictions, data.testing$Keterangan)

predictions      DITERIMA  TIDAK DITERIMA
  DITERIMA          49           3
  TIDAK DITERIMA    1           49
> akurasi=(49+49)/(102)
> akurasi
[1] 0.9607843
> sensitivity=49/50
> sensitivity
[1] 0.98
> specificity=49/52
> specificity
[1] 0.9423077
~ |
```

Lampiran 8: Penentuan *Variable Importance* Bagging CART

Script Penentuan *Variable Importance*

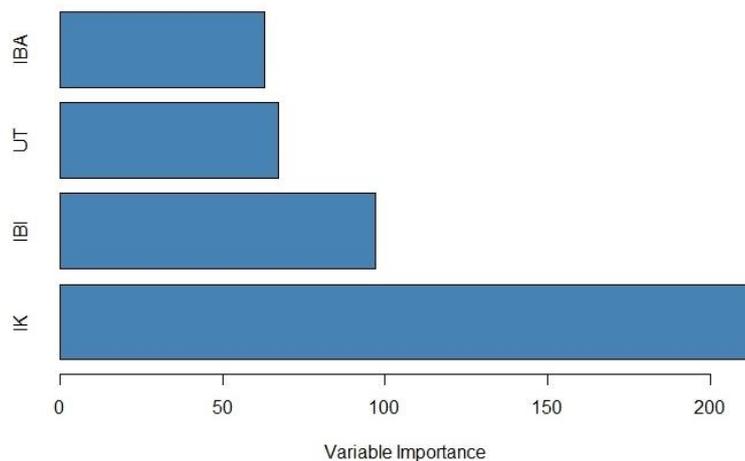
```
#hitung variabel penting
VI<-data.frame(var=names(data[-1]),imp=varImp(bag))
print(VI)

#sort variable importance descending
VI_plot<-VI[order(VI$Overall, decreasing = TRUE),]

#visualize variable importance with horizontal bar plot
barplot(VI_plot$Overall,
        names.arg=rownames(VI_plot),
        horiz=TRUE,
        col='steelblue',
        xlab='Variable Importance')
```

Output *Variable Importance*

```
> VI<-data.frame(var=names(data[-1]),imp=varImp(bag))
> print(VI)
  var Overall
IBA   IBA  63.28216
IBI   IBI  99.74735
IK    IK  215.34065
UT Keterangan 68.26003
> #sort variable importance descending
> VI_plot<-VI[order(VI$Overall, decreasing = TRUE),]
> VI_plot
  var Overall
IK    IK  215.34065
IBI   IBI  99.74735
UT Keterangan 68.26003
IBA   IBA  63.28216
```



RIWAYAT HIDUP



Penulis Bernama Afidatul Masbakhah, lahir di Kabupaten Gresik pada tanggal 29 Juni 2000, biasa dipanggil Afida. Bertempat tinggal di desa Tiremenggala, kecamatan Dukun, kabupaten Gresik. Anak satu-satunya dari pasangan suami istri Sabih dan Ruhanah. Pendidikan formal yang telah ditempuh diawali dengan pendidikan dasar di MI Al-Islami Tiremenggala dan lulus pada tahun 2012. Setelah itu melanjutkan pendidikan di MTs Kanjeng Sepuh Sidayu dan lulus pada tahun 2015. Kemudian melanjutkan pendidikan di jenjang SMA di MAN 3 Jombang dan lulus pada tahun 2018. Setelah itu melanjutkan pendidikan perguruan tinggi di Universitas Islam Negeri Maulana Malik Ibrahim Malang Program Studi Matematika. Selama menempuh pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang, ia aktif dalam berbagai organisasi, diantaranya menjadi pengurus HMJ “Integral” Matematika selama dua periode, Musyrifah Pusat Ma’had Al-Jami’ah selama tiga tahun, Senat Mahasiswa Fakultas Sains dan Teknologi di Komisi C (Bidang Kemahasiswaan dan Keorganisasian), Pengurus Rayon *Pencerahan* Galileo dua periode, pengurus Dewan Mahasiswa Universitas Islam Negeri Maulana Malik Ibrahim Malang dan juga dua tahun tergabung dalam anggota Generasi Bank Indonesia. Pembaca dapat menghubungi penulis melalui email: afidah29600@gmail.com.



BUKTI KONSULTASI SKRIPSI

Nama : Afidatul Masbakhah
NIM : 18610055
Fakultas/Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Klasifikasi Penerimaan Musyrif Musyrifah Pusat Ma'had Al Jami'ah Melalui Metode *Bootstrap Aggregating Classification Analysis Regression Trees*
Pembimbing I : Angga Dwi Mulyanto, M. Si
Pembimbing II : Ari Kusumastuti, M. Pd., M. Si.

No	Tanggal	Hal	Tanda Tangan
1.	18 Januari 2022	Judul (Pembimbing 1)	1.
2.	24 Januari 2022	Revisi Bab I, II (Pembimbing 1)	2.
3.	02 Februari 2022	Revisi Bab I, II, III (Pembimbing 1)	3.
4.	04 Maret 2022	Analisis Data di Software (Pembimbing 1)	4.
5.	14 Maret 2022	Revisi Analisis Data di Software (Pembimbing 1)	5.
6.	16 Maret 2022	Revisi Bab II (Pembimbing 2)	6.
7.	16 Maret 2022	Acc Bab I, II, III (Pembimbing 1)	7.
8.	18 Maret 2022	Acc Bab I, II (Pembimbing 2)	8.
9.	26 April 2022	Revisi Bab IV, V (Pembimbing 1)	9.
10.	28 April 2022	Revisi Bab II (Pembimbing 2)	10.
11.	23 April 2022	ACC Bab IV, V (Pembimbing 1)	11.
12.	24 April 2022	ACC Bab II, IV, V (Pembimbing 2)	12.
13.	14 Juni 2022	ACC Revisi Seminar Hasil (Pembimbing 1)	13.
14.	15 Juni 2022	ACC Revisi Seminar Hasil (Pembimbing 2)	14.
15.	22 Juni 2022	ACC Keseluruhan (Pembimbing 1)	15.

Malang, 24 Juni 2022
Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005