



**ANALISA PERBANDINGAN ANTARA LINEAR
REGRESSION DAN *RANDOM FOREST* REGRESSION
DALAM PREDIKSI PENERIMAAN MAHASISWA BARU DI
STKIP NU KABUPATEN TEGAL TAHUN 2025-2026**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar

Sarjana Pendidikan

Oleh

SOFANIA MA'RIFATUL JANAHA

NIM. 2102000080

**PROGRAM STUDI PENDIDIKAN INFORMATIKA
SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN
NAHDLATUL ULAMA KABUPATEN TEGAL**

2025

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

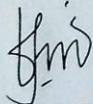
ANALISA PERBANDINGAN ANTARA LINEAR REGRESSION DAN
RANDOM FOREST REGRESSION DALAM PREDIKSI PENERIMAAN
MAHASISWA BARU DI STKIP NU KABUPATEN TEGAL TAHUN 2025-
2026

Yang Disiapkan Dan Disusun Oleh :

SOFANIA MA'RIFATUL JANAH
NIM 2102000080

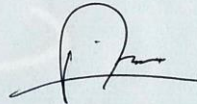
Diperiksa dan disetujui oleh :

Pembimbing I



Firman, M.Kom
NIDN. 0617087707

Pembimbing II



Rizal Nurzuli, M.Kom
NIDN.0618058706

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

ANALISA PERBANDINGAN ANTARA LINEAR REGRESSION DAN
RANDOM FOREST REGRESSION DALAM PREDIKSI PENERIMAAN
MAHASISWA BARU DI STKIP NU KABUPATEN TEGAL TAHUN 2025-

2026

Oleh:

SOFANIA MA'RIFATUL JANAH

NIM 2102000080

Telah dipertahankan di depan Tim Penguji

Pada Tanggal 19 September 2025

Dinyatakan telah memenuhi syarat

Tegal, 19 September 2025

Pembimbing I

Firman, M.Kom
NIDN. 0617087707

Penguji I

Rito Cipto Sigitta Hariyono. M.Kom
NIDN. 0619128301

Pembimbing II

Riza Nurzuli, M.Kom.
NIDN. 0618058706

Penguji II

Fadilah Falah Svifa, M.Pd
NIDN. 0609029101

Mengetahui,

Ketua STKIP Nahdlatul Ulama Kabupaten Tegal

Dr. H. Ahmad Taufiq, M.Pd.I
NIDN. 2112126901

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Sofania Ma'rifatul Janah
NIM : 2102000080
Program Studi : Pendidikan Informatika
Judul Skripsi :
**ANALISA PERBANDINGAN ANTARA LINEAR
REGRESSION DAN RANDOM FOREST
REGRESSION DALAM PREDIKSI
PENERIMAAN MAHASISWA BARU DI STKIP
NU KABUPATEN TEGAL TAHUN 2025-2026**

Dengan penuh tanggung jawab, saya menyatakan bahwa karya tulis ilmiah ini merupakan hasil pemikiran dan penyusunan saya sendiri, tanpa menyalin, memalsukan, ataupun merekayasa isi, baik sebagian maupun secara keseluruhan. Jika di kemudian hari terbukti bahwa terdapat unsur penjiplakan, pemalsuan, atau manipulasi data dalam naskah ini, maka saya siap menerima segala konsekuensi dan sanksi sesuai aturan dan ketentuan yang berlaku.

Tegal, 19 September 2025

Yang membuat Pernyataan



Sofania Ma'rifatul Janah
NIM. 2102000080

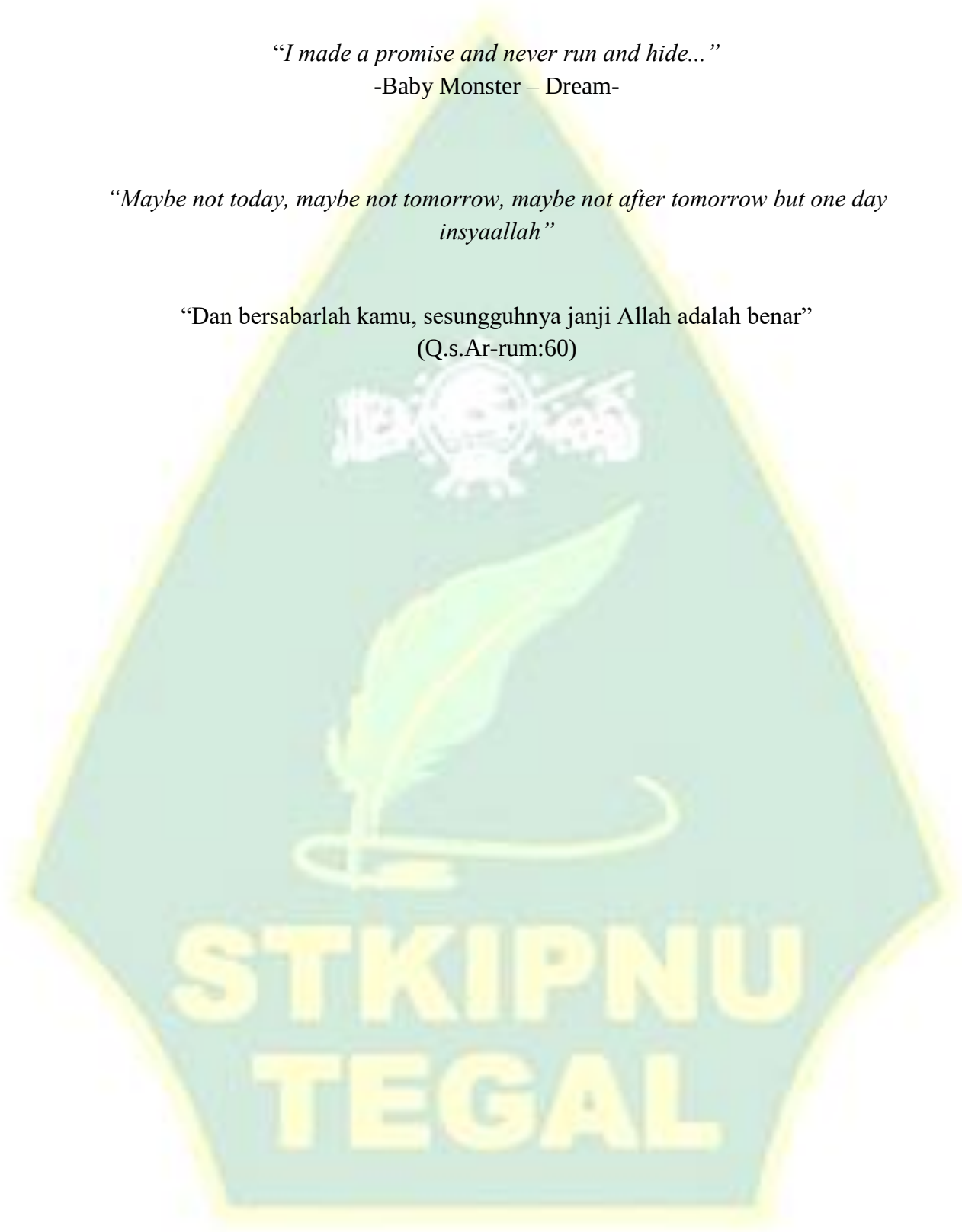
MOTTO

“I made a promise and never run and hide...”

-Baby Monster – Dream-

“Maybe not today, maybe not tomorrow, maybe not after tomorrow but one day insyaallah”

“Dan bersabarlah kamu, sesungguhnya janji Allah adalah benar”
(Q.s.Ar-rum:60)



ABSTRAK

Sofania Ma'rifatul Janah.2025. Analisa Perbandingan Antara Linear Regression dan *Random Forest* Regression Dalam Prediksi Penerimaan Mahasiswa Baru di STKIPNU Kabupaten Tegal Tahun 2025-2026. Skripsi. Program Studi Pendidikan Informatika. STKIP Nahdlatul Ulama Kab.Tegal.Pembimbing : (I) Firman, M.Kom, (II) Rizal Nurzuli M.Kom,.

Kata kunci: Prediksi Mahasiswa Baru, *Machine Learning*, *Linear Regression*, *Random Forest*

STKIPNU Kabupaten Tegal, perguruan tinggi swasta di Kecamatan Slawi yang berdiri pada 28 Januari 2016, mengalami penurunan jumlah mahasiswa baru dalam dua tahun terakhir. Kondisi ini menimbulkan ketidakpastian yang dapat memengaruhi stabilitas operasional kampus. Penelitian ini bertujuan memprediksi jumlah mahasiswa baru dengan metode *Machine Learning* menggunakan algoritma *Linear Regression* dan *Random Forest*. Proses evaluasi kinerja model dilakukan menggunakan metrik Root Mean Squared Error (RMSE) dan Mean Absolute Percentage Error (MAPE). Hasil menunjukkan *Linear Regression* memiliki RMSE sekitar 31–32 dan MAPE sekitar 39–40%, yang menunjukkan kesalahan prediksi moderat hingga tinggi. Sebaliknya, *Random Forest* menghasilkan RMSE sekitar 4–5 dan MAPE di bawah 7%, yang menandakan akurasi sangat baik. Dengan demikian, *Random Forest* terbukti jauh lebih akurat dibandingkan *Linear Regression* dalam memprediksi jumlah penerimaan mahasiswa baru. Hasil penelitian ini dapat menjadi dasar bagi STKIPNU Kabupaten Tegal dalam perencanaan strategis dan pengambilan keputusan.

ABSTRACT

Sofania Ma'rifatul Janah.2025. *Comparative Analysis Between Linear Regression and Random Forest Regression in Predicting New Student Admissions at STKIPNU Tegal Regency in 2025-2026*. Skripsi. Program Studi Pendidikan Informatika. STKIP Nahdlatul Ulama Kab.Tegal. Pembimbing: (I) Firman, M.Kom, (II) Rizal Nurzuli M.Kom,.

Keywords: *New Student Prediction, Machine Learning, Linear Regression, Random Forest*

STKIPNU Kabupaten Tegal, a private university located in Slawi District and established on January 28, 2016, has experienced a decline in the number of new students over the past two years. This condition creates uncertainty that can affect the operational stability of the campus. This study aims to predict the number of new students using Machine Learning methods with the Linear Regression and Random Forest algorithms. Model performance was evaluated using the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) metrics. The results show that Linear Regression produced an RMSE of around 31–32 and a MAPE of about 39–40%, indicating a moderate to high prediction error. In contrast, Random Forest achieved an RMSE of approximately 4–5 and a MAPE below 7%, demonstrating very high accuracy. Thus, Random Forest proved to be far more accurate than Linear Regression in predicting the number of new student admissions. These findings can serve as a basis for STKIPNU Kabupaten Tegal in strategic planning and decision-making.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“ANALISA PERBANDINGAN ANTARA LINEAR REGRESSION DAN *RANDOM FOREST* REGRESSION DALAM PREDIKSI PENERIMAAN MAHASISWA BARU DI STKIP NU KABUPATEN TEGAL TAHUN 2025-2026”**. Penulisan skripsi ini dilakukan sebagai salah satu syarat untuk menyelesaikan program sarjana di STKIP NU Kabupaten Tegal.

Tidak dapat dipungkiri bahwa dalam proses penyusunan skripsi ini, penulis menghadapi berbagai tantangan, baik dari segi waktu, data, maupun pemahaman terhadap materi. Namun, berkat doa, dorongan, dan bantuan dari berbagai pihak, tantangan tersebut dapat dilalui dengan baik hingga skripsi ini berhasil diselesaikan. Oleh karena itu, dengan penuh rasa hormat dan penghargaan, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Bapak Dr. Akhmad Taufiq, M.Pd.I. selaku ketua sekolah tinggi keguruan dan ilmu pendidikan nahdlatul ulama (STKIP NU) kabupaten Tegal yang telah memberikan fasilitas kepada kami untuk menyelesaikan studi kami.
2. Bapak Firman, M. Kom., selaku Dosen Pembimbing I yang telah dengan sabar membimbing, memberikan waktu, energi, arahan, serta masukan yang sangat berarti dalam menyempurnakan skripsi ini. Nasihat dan dorongan Bapak menjadi bekal berharga dalam proses penulisan maupun pengembangan pemikiran akademik penulis.
3. Bapak Rizal Nurzuli, M.Kom., selaku Dosen Pembimbing II yang tak kenal lelah memberikan masukan secara detail dan objektif. Ketelitian dan dedikasi Bapak dalam mengarahkan penulis menjadikan proses ini berjalan lebih terarah dan menghasilkan karya yang lebih baik.
4. Kedua orang tua penulis, Bapak M.Solehudin dan Ibu Alhikmah, yang dengan penuh cinta dan keikhlasan telah membesarkan, membimbing, serta

mendoakan penulis setiap saat. Terima kasih atas kasih sayang, pengorbanan, dan doa yang tidak pernah putus. Segala pencapaian ini tidak akan berarti tanpa kehadiran dan restu dari beliau berdua. Skripsi ini penulis persembahkan sebagai bentuk penghormatan dan rasa terima kasih yang tak terhingga.

5. Seluruh dosen dan staf STKIP NU Kabupaten Tegal yang telah membekali penulis dengan ilmu, pengalaman, serta inspirasi selama menempuh studi.
6. Rekan-rekan mahasiswa, sahabat seperjuangan, yang telah berbagi ilmu, berdiskusi, saling membantu, dan memberikan semangat hingga akhirnya penulis mampu menyelesaikan tugas akhir ini. Kebersamaan dan perjuangan kita akan selalu menjadi kenangan yang membekas.
7. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam proses penulisan skripsi ini. Setiap bentuk bantuan, sekecil apa pun, sangat berarti bagi penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, baik dari segi isi maupun penyajiannya. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan dan penyempurnaan di masa yang akan datang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat, baik sebagai referensi ilmiah maupun sebagai kontribusi kecil dalam pengembangan ilmu pengetahuan, khususnya di bidang data science dan analisis prediktif.

Tegal, 22 Mei 2025

Penyusun

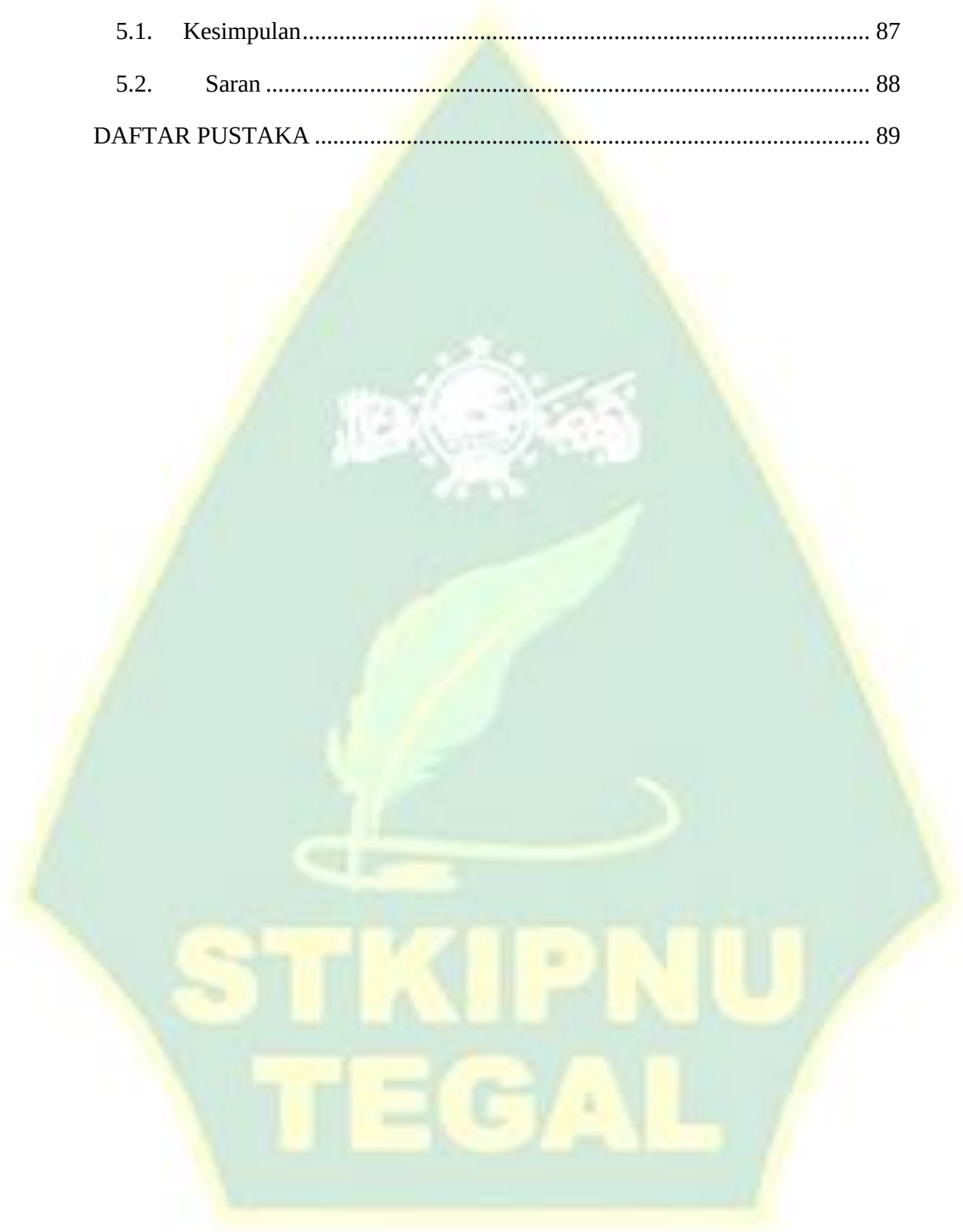
Sofania Ma'rifatul Janah

DAFTAR ISI

HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN KEASLIAN SKRIPSI.....	ii
MOTTO	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Rumusan Masalah	4
1.5 Tujuan.....	5
1.6 Manfaat.....	5
1.7 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	7
2.1 Prediksi Penerimaan Mahasiswa Baru	7
2.1.1 Mahasiswa Baru	7

2.1.2	Konsep Prediksi	8
2.1.3	Prediksi Penerimaan Mahasiswa Baru	11
2.2	Regresi Linear	18
2.3	<i>Random Forest Regression</i>	21
2.4	Penelitian Relevan	23
2.5	Kerangka Berpikir	32
2.6	Hipotesis Penelitian	35
BAB III METODE PENELITIAN.....		36
3.1.	Pendekatan Penelitian.....	36
3.2.	Metode Penelitian.....	37
3.3.	Instrument Penelitian.....	38
3.4.	Tahapan Penelitian	43
3.4.1.	Pengumpulan Data	43
3.4.2.	Pengolahan Data Awal.....	43
3.5.	Tempat Penelitian.....	63
BAB IV HASIL DAN PEMBAHASAN		65
4.1	Atribut Data	65
4.2	Data Split.....	69
4.3	Linear regression	70
4.3.1.	Linaer regression menggunakan data split 30:70.....	70
4.3.2.	Linaer regression menggunakan data split 80:20.....	77
4.4	Radom Forest	80
4.4.1.	<i>Random Forest Regression</i> menggunakan data split 70:30	80
4.4.2.	<i>Random Forest Regression</i> menggunakan data split 80:20	83
4.5	Hasil.....	85

BAB V KESIMPULAN DAN SARAN.....	87
5.1. Kesimpulan.....	87
5.2. Saran	88
DAFTAR PUSTAKA	89



DAFTAR TABEL

Tabel 2. 1 State of The Art.....	28
Tabel 3. 1 Data primer	39
Tabel 3. 2 Data Historis	40
Tabel 3. 3 Beasiswa.....	45
Tabel 3. 4 Promosi	45
Tabel 3. 5 Sarana dan Prasarana	45
Tabel 3. 6 Nilai Min dan Max.....	46
Tabel 3. 7 Biaya	46
Tabel 3. 8 Jumlah mahasiswa per tahun.....	46
Tabel 3. 9 Estimasi Jarak	47
Tabel 3. 10 Dataset Linear Regression	48
Tabel 3. 11 perhitungan nilai $X_n Y$	49
Tabel 3. 12 perhitungan nilai $X_n X_n$	49
Tabel 3. 13 perhitungan $X_n Y$	50
Tabel 3. 14 Matriks A	51
Tabel 3. 15 Matriks A1	52
Tabel 3. 16 Matriks A2	52
Tabel 3. 17 Matriks A3	53
Tabel 3. 18 Matriks A4	53
Tabel 3. 19 Matriks A5	54
Tabel 3. 20 Matriks A6	54
Tabel 3. 21 Matriks A7	55
Tabel 3. 22 Nilai Determinan.....	55
Tabel 3. 23 Nilai β	56
Tabel 3. 24 Dataset Random Forest	56
Tabel 3. 25 penambahan nomor hasil bootstrapping	57
Tabel 3. 26 perhitungan squares Error	58
Tabel 3. 27 menghitung MSE pada atribut Beasiswa = 0	59

Tabel 3. 28 menghitung MSE pada atribut Beasiswa = 1	59
Tabel 3. 29 menghitung MSE pada Tabel Promosi = 0	60
Tabel 3. 30 menghitung MSE pada Tabel Promosi = 1	60
Tabel 3. 31 menghitung MSE dari atribut Promosi = 1	61
Tabel 3. 32 menghitung MSE dari Atribut Sarpras = 0	61
Tabel 3. 33 menghitung MSE pada atribut Biaya<0.16.....	62
Tabel 3. 34 nilai weighted Average MSE	63
Tabel 4. 1 Atribut	66
Tabel 4. 2 Perbandingan Kinerja.....	85



DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berpikir	33
Gambar 4. 1 tampilan halaman kerja	71
Gambar 4. 2 Hasil	72
Gambar 4. 3 nilai RMSE	75
Gambar 4. 4 nilai MAPE	75
Gambar 4. 5 Hasil prediksi dengan ratio 80:20	77
Gambar 4. 6 nilai RMSE	79
Gambar 4. 7 Random Forest Regression data split 70:30	80
Gambar 4. 8 nilai RMSE	81
Gambar 4. 9 nilai MAPE	82
Gambar 4. 10 Random Forest Regression data split 80:20	83
Gambar 4. 11 hasil RMSE ratio 80:20	84
Gambar 4. 12 Nilai persentase MAPE	84

DAFTAR LAMPIRAN

Lampiran 1 Surat Permohonan Penelitian.....	95
Lampiran 2 Surat Selesai Penelitian	96
Lampiran 3 Surat Keterangan Perpustakaan	97
Lampiran 4 Surat Keterangan Turnitin	98
Lampiran 5 Hasil Turnitin.....	99
Lampiran 6 Surat Keterangan Selesai Revisi.....	100



BAB I

PENDAHULUAN

1.1 Latar Belakang

Institusi perguruan tinggi mempunyai peran strategis dalam membentuk sumber daya manusia yang berkualitas serta mendukung kemajuan suatu bangsa. Selain menjadi tempat untuk belajar dan mengembangkan keterampilan, perguruan tinggi juga menjadi pusat penelitian dan pengabdian kepada masyarakat. Endrawati dkk (2023) menegaskan bahwa perguruan tinggi dituntut untuk terus berinovasi dalam era revolusi industri 4.0 karena perkembangan teknologi menuntut kemampuan adaptasi yang cepat, baik dalam layanan pendidikan, kurikulum, maupun tata kelola institusi.

STKIPNU Kabupaten Tegal merupakan perguruan tinggi swasta yang berdiri pada 28 Januari 2016 dan berlokasi di Kecamatan Slawi. Dalam beberapa tahun pertama, STKIPNU menunjukkan perkembangan positif. Data Badan Pusat Statistik Kabupaten Tegal tahun 2019–2023 mencatat peningkatan jumlah mahasiswa baru sejak kampus ini berdiri. Dikutip dari Rahmadani dkk (2023) dalam hal perguruan tinggi swasta, jumlah mahasiswa baru memiliki pengaruh langsung terhadap keberlanjutan operasional kampus karena sebagian besar pendapatan berasal dari biaya pendidikan mahasiswa.

Namun, dalam dua tahun terakhir STKIPNU mengalami penurunan jumlah mahasiswa baru yang cukup signifikan (Hasil wawancara dengan Bagian Akademik STKIPNU, 2025). Berbagai faktor dapat memengaruhi keputusan calon mahasiswa untuk mendaftar, seperti lokasi kampus, biaya pendidikan, program studi, kondisi fasilitas, kegiatan kemahasiswaan, serta status akreditasi (Madyatmadja dkk., 2021). Penurunan ini menjadi tantangan serius karena dapat berdampak pada stabilitas operasional kampus.

Di sisi lain, pemerintah berupaya memperluas akses pendidikan tinggi melalui Program KIP Kuliah yang bertujuan meringankan biaya kuliah mahasiswa kurang mampu (Amadi dkk., 2023). STKIPNU juga menyediakan fasilitas beasiswa bagi pendaftar. Namun hingga kini belum tersedia metode untuk memprediksi jumlah pendaftar yang memanfaatkan program tersebut, padahal informasi ini sangat penting untuk perencanaan kapasitas sarana, beban dosen, dan pembiayaan kampus (Fauzy & Pramaditya, 2024).

Seiring berkembangnya teknologi, pemanfaatan *Machine Learning* menjadi solusi potensial untuk melakukan prediksi berbasis data. *Machine Learning* mampu mempelajari pola dari data historis dan menghasilkan estimasi yang akurat. Beberapa algoritma yang sering digunakan adalah *Linear Regression* (LR) dan *Random Forest* (RF) (Roihan dkk., 2020). Penelitian Roihan dkk (2020) menunjukkan bahwa *Random Forest* dapat menghasilkan MAPE 6,22% yang dikategorikan sangat akurat. Sementara

itu, (Sebayang, 2022) menggunakan *Linear Regression* untuk memprediksi jumlah mahasiswa baru dan memperoleh tingkat akurasi mencapai 96,556%.

Meskipun demikian, sebagian besar penelitian sebelumnya hanya menggunakan data akademik seperti tahun masuk dan jumlah pendaftar. Belum banyak penelitian yang melibatkan faktor non-akademik seperti beasiswa, akreditasi, fasilitas kampus, dan kegiatan kemahasiswaan, padahal faktor tersebut sangat memengaruhi minat calon mahasiswa (Luqman, Mia, & Firman, 2023)

Berdasarkan kondisi tersebut, penelitian ini bertujuan membandingkan kinerja algoritma *Linear Regression* dan *Random Forest* dengan memanfaatkan data akademik dan non-akademik untuk memprediksi jumlah mahasiswa baru di STKIPNU Kabupaten Tegal. Hasil penelitian ini diharapkan dapat menjadi dasar bagi kampus dalam pengambilan keputusan strategis, perencanaan promosi, serta pengelolaan fasilitas kampus secara lebih efektif.

1.2 Identifikasi masalah

Dengan memperhatikan latar belakang permasalahan yang sudah dijabarkan, kemudian diidentifikasi beberapa permasalahan utama yang menjadi dasar dilakukannya penelitian ini:

1. Penurunan jumlah mahasiswa baru dalam dua tahun terakhir di STKIPNU Kab.Tegal yang berdampak kepada stabilitas operasional dan pendanaan institusi.
2. Strategi sosialisasi yang telah diimplementasikan oleh pihak institusi belum menunjukkan efektivitas yang optimal, sebagaimana ditunjukkan oleh tren penurunan jumlah pendaftar meskipun kegiatan promosi telah dilakukan secara rutin ke sekolah-sekolah di wilayah sekitar kampus.

1.3 Batasan Masalah

Berdasarkan permasalahan latar belakang diatas, batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data historis dari tahun 2020-2024 dengan data tambahan seperti ketersediaan bantuan beasiswa, sarana dan prasarana Kampus, dan Status akreditasi.
2. Penelitian ini hanya membandingkan dan mengimplementasikan dua algoritma *Machine Learning* yaitu *Random Forest* serta linear regression.

1.4 Rumusan Masalah

Merujuk pada latar belakang tersebut, permasalahan yang dapat dirumuskan mencakup:

Algoritma mana yang memberikan hasil paling akurat dalam meramalkan jumlah mahasiswa baru di STKIPNU Kab.tegal?

1.5 Tujuan

Merujuk pada perumusan masalah, tujuan studi ini yaitu:

1. Untuk menerapkan algoritma *Linear Regression* dan *Random Forest* dalam memprediksi jumlah mahasiswa baru di STKIPNU Kab.Tegal.
2. Untuk mengetahui tingkat keakuratan yang diperoleh dari perbandingan antara algoritma *Linear Regression* dan *Random Forest*.

1.6 Manfaat

Berdasarkan tujuan penelitian diatas, beberapa manfaat yang akan didapati dari studi ini sebagai berikut:

1. Penelitian ini harapannya mampu membantu pihak kampus dalam merancang strategi peningkatan jumlah mahasiswa baru melalui prediksi yang lebih akurat, sehingga dapat menjaga keberlangsungan operasional kampus secara optimal.
2. Memberikan wawasan terkait efektivitas dari masing-masing metode dalam meramalkan jumlah mahasiswa baru.

1.7 Sistematika Penulisan

Sistematika pembahasan dibuat agar mempermudah pembaca dalam memahami isi penelitian. Secara garis besar sistematika pembahasan dalam penelitian ini adalah :

BAB I PENDAHULUAN

pada bagian ini berisikan latar belakang, identifikasi masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika dalam penulisan skripsi.

BAB II LANDASAN TEORI

Pada bagian ini berisikan semua hal yang berkaitan dengan penelitian terkait, teori-teori serta metode yang digunakan untuk mendukung penelitian tugas akhir.

BAB III METODE PENELITIAN

pada bagian ini mencakup pembahasan mengenai data, metode, serta langkahlangkah penelitian, serta memberikan gambaran umum tentang sistem yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Dalam bagian ini membahas laporan hasil dan data pendukung dari penelitian.

BAB V KESIMPULAN DAN SARAN

Dalam bagian ini menjelaskan bagaimana penelitian yang diharapkan dapat membantu pengembangan ilmu dengan adanya kesimpulan dan saran dari penelitian ini

BAB VI PENUTUP

Dalam bagian ini membahas hasil dan rekomendasi untuk penelitian mendatang.

BAB II

LANDASAN TEORI

2.1 Prediksi Penerimaan Mahasiswa Baru

2.1.1 Mahasiswa Baru

Mahasiswa merupakan individu yang sedang menjalani proses pendidikan di jenjang perguruan tinggi dan telah resmi terdaftar sebagai peserta didik. Secara umum, mahasiswa berada pada tahap perkembangan rentang usia remaja akhir hingga dewasa awal, yakni sekitar 18 hingga 24 tahun (Jannah, 2024). Menurut Siswoyo (2007) dalam Asqia dan Wahyuni (2024), juga menyatakan bahwa mahasiswa adalah seseorang yang tengah mengikuti pendidikan di perguruan tinggi, baik di institusi negeri, swasta, maupun lembaga setara lainnya.

Menurut Hurlock (2015) dalam Marlinang dan Sarajar (2025), mahasiswa baru berada pada tahap perkembangan dewasa awal, yang ditandai dengan peningkatan kemandirian, pembentukan identitas, dan pengambilan keputusan yang lebih kompleks, termasuk keputusan memilih perguruan tinggi. Selain itu, Apriyanti dkk (2023) menyebutkan bahwa fase ini terkait dengan pencarian jati diri akademik dan profesional. Karakteristik perkembangan ini memengaruhi pola pikir calon mahasiswa dalam menentukan pilihan studi, program pendidikan, maupun institusi perguruan tinggi yang diminati.

Penelitian terbaru oleh Tamba dan Sitanggang (2024) menunjukkan bahwa keputusan mahasiswa baru memilih perguruan tinggi dipengaruhi oleh lima faktor utama: reputasi institusi, fasilitas kampus, biaya pendidikan, akreditasi program studi, serta rekomendasi dari keluarga dan teman. Hal ini sejalan dengan temuan Makhrisa dan Pradikto (2025) yang menyatakan bahwa lingkungan sosial terutama peran orang tua dan guru berkontribusi besar dalam membentuk preferensi calon mahasiswa sebelum mendaftar.

Selain itu, mahasiswa baru juga dipengaruhi oleh perkembangan era digital. Penelitian Oktavian dkk (2025) menunjukkan bahwa mayoritas calon mahasiswa mencari informasi perguruan tinggi melalui media sosial dan website resmi kampus. Hal ini menegaskan bahwa karakteristik mahasiswa baru saat ini lebih digital-oriented sehingga strategi promosi kampus harus menyesuaikan pola tersebut. Dengan demikian, pemahaman tentang mahasiswa baru tidak hanya terbatas pada aspek sosiodemografis, tetapi juga perilaku pencarian informasi mereka di era digital.

2.1.2 Konsep Prediksi

Prediksi adalah kegiatan menilai suatu keadaan atau peristiwa yang mungkin terjadi di masa depan, berdasarkan data dan informasi yang tersedia dimasa kini ataupun tahun sebelumnya, supaya bisa menghindari atau meminimalisir kesalahan yang sama seperti tahun sebelumnya. Prediksi digunakan untuk menilai kondisi dimasa

mendatang (W. Handoko, 2021). Prediksi tidak selalu sesuai dengan kenyataan (Nurmantika, 2022). Prediksi dapat bersifat ilmiah atau hanya berdasarkan pendapat pribadi. Misalnya, perkiraan cuaca dibuat dari data terbaru dan pengamatan satelit, begitu pula prediksi gempa atau letusan gunung. Sebaliknya, prediksi hasil pertandingan sepak bola atau olahraga biasanya didasarkan pada opini dan sudut pandang masing-masing orang.

Dalam teori peramalan menurut Makridakis (1998) dalam Andelawati (2021), prediksi dilakukan melalui dua pendekatan: kuantitatif, yang menggunakan data historis dan model matematis seperti regresi, time series, dan machine learning; serta kualitatif, yang menggunakan pendapat ahli, survei, atau intuisi. Keduanya dapat diterapkan dalam konteks pendidikan, terutama dalam memproyeksikan jumlah mahasiswa baru, kebutuhan sarana, atau tren minat program studi. Pendekatan kuantitatif memberikan hasil yang lebih stabil jika tersedia data historis yang memadai, sementara pendekatan kualitatif digunakan ketika data terbatas.

Dalam konteks manajemen pendidikan, prediksi memegang peran strategis untuk perencanaan kapasitas (ruang kelas, tenaga pendidik), pengelolaan anggaran, perencanaan pemasaran, serta kebijakan beasiswa. Penelitian Handoko dkk (2025) menunjukkan bahwa prediksi jumlah mahasiswa baru sangat membantu perguruan tinggi dalam menentukan alokasi anggaran operasional dan promosi. Sementara itu,

studi Martanto dkk (2019) menegaskan bahwa prediksi berbasis machine learning mampu meningkatkan akurasi perencanaan penerimaan mahasiswa hingga 73%.

Analisis prediktif merupakan proses pemanfaatan data masa lalu untuk memperkirakan hasil di masa depan (Sasongko, 2023). Dalam hal penerimaan mahasiswa baru, analisis ini dapat membantu perguruan tinggi memperkirakan jumlah pendaftar dan mahasiswa yang akan diterima. Perkiraan tersebut penting untuk merancang strategi operasional yang tepat, seperti penentuan kapasitas ruang kelas, pembagian anggaran, hingga penyediaan tenaga kerja.

Tidak seperti analisis deskriptif yang sekadar menggambarkan kondisi masa lalu, analisis prediktif memberikan gambaran strategis dengan mengidentifikasi faktor-faktor yang memengaruhi penerimaan mahasiswa. Melalui metode seperti regresi linear dan *Random Forest*, perguruan tinggi dapat memperkirakan dampak perubahan variabel terhadap jumlah pendaftar, sehingga mendukung pengambilan keputusan yang lebih akurat.

Analisis dan prediksi penerimaan mahasiswa baru menjadi langkah strategis yang harus dilakukan oleh setiap perguruan tinggi, termasuk STKIP NU Kabupaten Tegal. Dengan memahami faktor-faktor yang memengaruhi penerimaan mahasiswa dan memanfaatkan model prediktif seperti regresi linear dan *Random Forest*, kampus dapat

meningkatkan efisiensi dalam perencanaan dan pengelolaan sumber daya. Dalam jangka panjang, hal ini akan membantu institusi mencapai visi dan misinya serta beradaptasi dengan dinamika dunia pendidikan yang terus berubah.

2.1.3 Prediksi Penerimaan Mahasiswa Baru

Azizah (2011) dikutip dari Wijaya dkk (2022) menjelaskan bahwa Penerimaan mahasiswa baru merupakan tahapan penting di lembaga pendidikan, baik sekolah maupun perguruan tinggi, yang bertujuan untuk menyeleksi calon mahasiswa berdasarkan kriteria tertentu yang telah ditetapkan oleh institusi pendidikan tersebut.

Prediksi penerimaan mahasiswa baru adalah suatu proses untuk memprediksi kemungkinan diterimanya calon mahasiswa ke perguruan tinggi berdasarkan data dan informasi yang tersedia. Proses ini sering kali melibatkan penggunaan metode statistik, algoritma pembelajaran mesin, atau teknik lain yang dapat menganalisis variabel-variabel yang relevan untuk menghasilkan prediksi yang akurat. Tujuan dari prediksi penerimaan mahasiswa baru adalah untuk membantu pihak perguruan tinggi dalam membuat keputusan yang lebih objektif dan efisien terkait penerimaan mahasiswa baru, baik dalam hal kuota yang tersedia maupun kualitas calon mahasiswa yang diterima (Luthfi Mawardi dkk., 2024).

Dalam proses penerimaan mahasiswa baru, terdapat berbagai faktor yang memengaruhi keputusan seseorang untuk mendaftar dan memilih perguruan tinggi. Tidak hanya terbatas pada aspek akademik, faktor psikologis, sosial, dan lingkungan juga turut berperan penting dalam menentukan pilihan yang ditentukan. Diantara pendekatan yang bisa diimplementasikan dalam memahami dinamika ini secara lebih mendalam yaitu teori Krumboltz. Teori ini menjelaskan bahwa keputusan karier, termasuk pemilihan institusi pendidikan tinggi, dipengaruhi oleh empat elemen utama: pengalaman belajar, pengaruh lingkungan, faktor genetik atau karakteristik pribadi, serta kemampuan individu dalam memproses informasi dan memanfaatkan peluang yang tersedia (Sari dkk., 2021). Dalam proses penerimaan mahasiswa baru, latar belakang akademik mencerminkan pengalaman belajar, kondisi sosial ekonomi keluarga merepresentasikan pengaruh lingkungan, sedangkan respons calon mahasiswa terhadap promosi kampus dan informasi beasiswa menunjukkan kemampuan mereka dalam mengelola informasi.

Secara keseluruhan, penerimaan mahasiswa baru dipengaruhi oleh kombinasi faktor internal dan eksternal yang saling berkaitan (Budiyanto dkk., 2025). Adapun Faktor internal meliputi:

a. Produk

Produk disini meliputi Sistem dan Konsep Pembelajaran, Status Akreditasi dan Tenaga Pengajar (Dosen) yang professional dan

citra kampus. Lembaga pendidikan dengan citra merek (branding) yang kuat memiliki kemampuan untuk menarik perhatian calon peserta didik dan orang tua, sekaligus membangun kepercayaan terhadap mutu pendidikan yang diselenggarakan (Hemsley-brown & Oplatka dalam (Ngisti, Vidia, Reka, & Ahmad, 2024).

b. Harga

Dalam poin ini meliputi biaya pendidikan. Biaya pendidikan menjadi pertimbangan penting bagi calon mahasiswa dalam menentukan pilihan institusi pendidikan tinggi, dengan memperhatikan kemampuan keuangan pribadi maupun dukungan dari orang tua untuk mencukupi kebutuhan selama masa studi (Mustofa & Khotib, 2025). Perguruan Tinggi Swasta (PTS) umumnya menetapkan biaya kuliah yang lebih tinggi karena dikelola oleh yayasan atau perusahaan, sehingga tidak memperoleh bantuan dana dari pemerintah. Meskipun Perguruan Tinggi Swasta (PTS) bukan lembaga pemerintah, mereka tetap berpeluang memperoleh bantuan berupa subsidi atau insentif tertentu sesuai dengan ketentuan yang berlaku (Nurpratama dkk., 2024). Dalam Undang-Undang No. 20 Tahun 2003 mengenai Sistem Pendidikan Nasional Pasal 12 Ayat 1, dijelaskan bahwa peserta didik memiliki hak menerima beasiswa apabila memiliki prestasi akademik dan berasal dari keluarga kurang mampu, serta mendapatkan bantuan pembiayaan pendidikan bagi yang tidak mampu.

Adapun beasiswa itu sendiri dikutip dari Noviyanti dan Dermawan (2022), diklasifikasikan ke dalam dua jenis yaitu beasiswa penuh (*full scholarship*) dan beasiswa parsial (*partial scholarship*). Beasiswa penuh mencakup seluruh kebutuhan biaya studi yang sepenuhnya ditanggung oleh penyedia beasiswa. Sementara itu, beasiswa parsial hanya menutupi sebagian kebutuhan, biasanya terbatas pada biaya hidup, sedangkan sisanya tetap menjadi tanggungan pribadi penerima.

Berdasarkan bentuknya, beasiswa dapat diklasifikasikan ke dalam tiga jenis utama yang masing-masing memiliki karakteristik dan tujuan yang berbeda, yaitu:

1. Beasiswa Prestasi Akademik, dibagikan untuk mahasiswa dengan hasil akademik yang tinggi, misalnya berdasarkan Indeks Prestasi Kumulatif (IPK) yang unggul.
2. Beasiswa Bantuan, Ditujukan bagi mahasiswa dari keluarga dengan kondisi ekonomi kurang mampu namun memiliki prestasi. Persyaratannya biasanya mencakup data penghasilan orang tua, pengeluaran rutin, dan kebutuhan hidup lainnya.
3. Beasiswa Atletik, Beasiswa ini ditujukan bagi mahasiswa yang mempunyai ketrampilan luar biasa pada segi olahraga. Institusi pendidikan tinggi sering merekrut atlet berprestasi dan memberikan pembiayaan pendidikan penuh sebagai bentuk apresiasi atas prestasi mereka di bidang olahraga. Jenis

beasiswa ini biasanya tidak dapat dilamar secara umum, melainkan ditawarkan secara khusus kepada mereka yang memiliki kualifikasi atletik yang menonjol.

Sedangkan berdasarkan sumber pendanaannya, beasiswa dibedakan menjadi dua yaitu:

1. Beasiswa dari Pemerintah Indonesia, Pemerintah menyediakan berbagai program bantuan pendidikan untuk jenjang perguruan tinggi, seperti Bidikmisi, Peningkatan Prestasi Akademik (PPA), dan Lembaga Pengelola Dana Pendidikan (LPDP).
2. Beasiswa dari LSM, Swasta, atau Korporasi, Biasanya diberikan sebagai unsur dari program tanggung jawab sosial perusahaan (CSR). Contohnya termasuk beasiswa dari bank swasta, Sampoerna Foundation, dan berbagai organisasi nasional maupun internasional lainnya.

c. Sarana dan Prasarana

Sarana pendidikan mencakup alat dan perlengkapan yang digunakan secara langsung dalam proses belajar mengajar, seperti ruang kelas, meja, kursi, serta media pembelajaran. Sementara itu, prasarana meliputi fasilitas pendukung yang tidak digunakan secara langsung, seperti gedung, lapangan olahraga, dan jaringan listrik (Simorangkir dkk., 2022). Keberadaan sarana prasarana yang cukup sangat krusial untuk mendukung kelancaran kegiatan pendidikan. Tanpa dukungan fasilitas yang memadai, proses

pembelajaran dapat mengalami hambatan serius, bahkan berisiko tidak berjalan sebagaimana mestinya.

d. Promosi

Menurut Lupiyoadi (2008) dikutip dalam Hamidah & Nurhuda (2022), promosi adalah komponen krusial dalam bauran pemasaran yang harus dijalankan oleh perusahaan guna memperkenalkan dan memasarkan produk maupun jasa yang ditawarkannya. Dalam melakukan promosi macamnya bisa promosi langsung atau promosi dengan memanfaatkan media social. Hikmah dkk (2023) mengungkapkan promosi media sosial merupakan faktor krusial pada strategi pemasaran perguruan tinggi dalam menarik minat calon mahasiswa.

Menurut Gunelius (2011:10) dalam Pratiwi (2020), pemasaran melalui media sosial merupakan pendekatan promosi yang memanfaatkan platform digital sosial untuk menyebarkan informasi. Jika digunakan secara aktif, dengan jaringan yang kuat dan konten yang relevan, media sosial mampu meningkatkan ketertarikan sekaligus memperluas penyebaran informasi. Sejumlah penelitian, seperti yang dilakukan Febyola dan Widyanesti (2024) serta Sagynbekova (2020), menunjukkan bahwa media sosial memiliki pengaruh yang signifikan terhadap penyebaran informasi dari mulut ke mulut secara digital (*electronic word of mouth*) di kalangan pelajar.

Sementara itu, salah satu faktor eksternal yang mempengaruhi penerimaan mahasiswa baru diantaranya adalah pengaruh keluarga. Keluarga memegang peranan esensial dalam membentuk kerangka berpikir dan arah hidup individu, termasuk dalam pengambilan keputusan terkait jenjang pendidikan. Dukungan dari lingkungan keluarga dapat menjadi katalisator dalam meningkatkan dorongan internal anak untuk melanjutkan pendidikan ke tingkat yang lebih tinggi. Menurut Slameto (2013) dikutip dalam Vita dkk (2023), variabel seperti pola pengasuhan, atmosfer domestik, kapasitas ekonomi, serta literasi orang tua terhadap kebutuhan perkembangan anak, turut berkontribusi terhadap etos belajar peserta didik.

Orang tua dengan latar belakang pendidikan yang lebih tinggi cenderung memiliki kecenderungan lebih besar dalam mendorong anaknya untuk menempuh pendidikan lanjutan. Oleh sebab itu, keterlibatan orang tua dalam proses pendidikan anak menjadi aspek yang tidak dapat dikesampingkan. Selain itu, tingkat pemahaman keluarga terhadap urgensi pendidikan turut memengaruhi kecenderungan individu dalam mengambil keputusan untuk melanjutkan studi ke perguruan tinggi. Salah satu determinan yang paling dominan dalam konteks ini adalah aspek finansial, yang kerap menjadi penghambat akses pendidikan tinggi, yang mana seseorang dengan keluarga yang berpenghasilan rendah akan membuat seseorang kehilangan minat serta motivasi untuk melanjutkan pendidikan.

2.2 Regresi Linear

Regresi Linear merupakan teknik statistik yang digunakan untuk memodelkan hubungan antara dua variabel, yakni variabel independen (X) dan variabel dependen (Y), dengan asumsi bahwa hubungan tersebut bersifat linier. Metode ini kerap dimanfaatkan untuk melakukan prediksi atau estimasi nilai variabel dependen berdasarkan nilai variabel independen yang tersedia. Montgomery dalam bukunya *Introduction to Linear Regression Analysis* mendefinisikan regresi linear sebagai "metode statistik yang memodelkan hubungan linier antara variabel dependen dan satu atau lebih variabel independen." Sementara itu, Kutner dalam *Applied Linear Statistical Models* menyatakan bahwa regresi linear adalah "alat statistik yang digunakan untuk mengkaji hubungan antara variabel dependen dan satu atau lebih variabel independen"(Mulayawan, 2022).

Analisis hubungan antara data numerik dengan data numerik juga dikenal sebagai uji korelasi atau regresi linear (Vandiah, 2022). Regresi linear bertujuan untuk memprediksi nilai variabel dependen berdasarkan satu atau lebih variabel independen dengan menemukan garis regresi terbaik (best fit) yang merepresentasikan data tersebut. Pada regresi linear sederhana, hanya terdapat satu variabel independen, sedangkan pada regresi linear berganda, lebih dari satu variabel independen digunakan untuk memprediksi variabel dependen (Fauziyah, 2023).

Regresi linear sederhana memodelkan hubungan linear diantara satu variable bebas (x) dan variabel terikat (Y). Persamaan umum untuk regresi linear sederhana dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 x + e$$

Keterangan:

- Y = Variabel Dependen yang ingin diprediksi
- X = Variabel Independen yang digunakan
- β_0 = Intercept
- β_1 = Koefisien Regresi
- e = Error Term

Pada regresi linear berganda, ada lebih dari satu variabel independen yang diimplementasikan dalam memperkirakan variabel dependen. Contohnya, prediksi penjualan bisa dipengaruhi oleh harga, promosi, lokasi, dan banyak faktor lainnya, yang semuanya dapat dimodelkan menggunakan regresi linear berganda (Yuliara, 2022). Adapun persamaan regresi linear berganda dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N + e$$

Keterangan:

- Y = Variabel Dependen yang ingin diprediksi
- X_1, X_2 = Variabel Independen yang digunakan
- β_1, β_2 = Koefisien Regresi
- e = Error Term

Dalam persamaan ini β_0 disebut intersep dan β_1 disebut slope atau kemiringan. Kedua parameter ini yang menentukan garis lurus terbaik yang meminimalkan jumlah error antara data dan garis prediksi. Dalam teori matematika Intersep adalah Titik di mana garis memotong sumbu Y ketika X bernilai 0 pada diagram kartesius. Sedangkan dalam statistik, intersep adalah nilai rata-rata variabel Y saat X bernilai 0. Artinya, jika X tidak memberikan pengaruh, maka Y bernilai sama dengan intersep. Intersep berfungsi sebagai konstanta dasar dalam model regresi, namun tidak selalu perlu diinterpretasikan, terutama jika nilai X dalam data tidak mendekati atau sama dengan 0 karena bisa jadi tidak memiliki makna khusus. Slope/kemiringan yaitu seberapa besar perubahan rata-rata Y untuk setiap kenaikan satu satuan X (Sabrina dkk., 2023).

Setelah model regresi linear dibangun, perlu dilakukan evaluasi untuk menilai seberapa baik model tersebut dalam memprediksi nilai Y. Beberapa ukuran yang digunakan untuk evaluasi adalah (Maulana dkk., 2024):

- a. Koefisien Determinasi (R^2): R^2 mengukur proporsi variansi dalam variabel dependen yang dapat dijelaskan oleh variabel independen. Nilai R^2 berkisar antara 0 dan 1. Semakin dekat R^2 ke 1, semakin baik model dalam menjelaskan variansi data.
- b. Uji Signifikansi F (ANOVA): Uji F digunakan untuk menguji apakah model regresi secara keseluruhan signifikan dalam menjelaskan variansi data. Jika nilai p dari uji F kecil (misalnya < 0.05), maka model regresi dianggap signifikan.

- c. Uji Signifikansi Koefisien Regresi (t-test): Uji t digunakan untuk menguji apakah setiap koefisien regresi ($\beta_1, \beta_2, \dots, \beta_n$) signifikan dalam memprediksi variabel dependen. Jika nilai p untuk koefisien regresi lebih kecil dari 0.05, maka koefisien tersebut dianggap signifikan.

2.3 *Random Forest Regression*

Random Forest (RF) adalah diantara teknik *ensemble learning* yang berbasis pada struktur pohon keputusan, dikembangkan dari konsep dasar *Decision Tree*. Tidak seperti *Decision Tree* yang hanya membangun satu pohon klasifikasi atau regresi, *Random Forest* membentuk sejumlah pohon secara paralel, kemudian menggabungkan hasil prediksi dari seluruh pohon tersebut untuk memperoleh output akhir yang lebih stabil dan akurat. Breiman (2001) dikutip dari Billah (2024) mendeskripsikan *Random Forest* sebagai kumpulan dari beberapa pohon prediktor di mana setiap pohon bergantung pada vektor acak yang dipilih secara independen dan memiliki distribusi yang seragam. Dalam proses pembentukan pohon, pada setiap simpul (node) dilakukan pemilihan acak terhadap sejumlah variabel independen untuk menentukan pemisah terbaik, sehingga model memiliki keragaman antar pohon dan tidak terlalu tergantung pada fitur tertentu (Nur et al., 2023).

Metode ini dikembangkan dari pendekatan CART (*Classification and Regression Tree*), dengan mengintegrasikan dua teknik utama, yaitu *bootstrap aggregating* atau bagging pengambilan sampel acak berulang dari data asli dengan pengembalian dan pemilihan acak terhadap subset fitur saat membangun setiap pohon. Kombinasi ini membuat *Random Forest* mampu

menangani data berdimensi tinggi dan lebih tahan terhadap overfitting dibandingkan dengan model pohon tunggal, sehingga meningkatkan keandalan dan kinerja model secara keseluruhan.

Di balik pendekatannya yang sederhana, *Random Forest* menawarkan sejumlah keunggulan yang menjadikannya sangat efektif dalam berbagai situasi. Pertama, memberikan akurasi lebih baik bahkan menurut Breiman (2001) “Akurasinya setara dengan Adaboost dan terkadang lebih baik.” Serta “Relatif tahan terhadap outlier dan noise.” Sehingga mempunyai output yang stabil pada dataset yang kompleks.

Kedua, *Random Forest* bisa digunakan pada data dengan banyak fitur, baik yang berupa angka maupun kategori. Menurut Johnson dkk (2024) RF sangat baik dalam mengolah data yang besar dan rumit, serta mampu menemukan pola yang tidak bisa ditangkap oleh model linear sederhana. RF juga tidak memerlukan asumsi khusus tentang bentuk hubungan antar data, sehingga bisa digunakan dalam berbagai situasi. Ketiga, Memberikan estimasi internal yang berguna terhadap kesalahan, kekuatan, korelasi, dan pentingnya variabel, serta mudah untuk di paralelkan.

Erkamim dkk (2023) menyebutkan beberapa parameter dalam *Random Forest Regression* dapat disesuaikan untuk mengoptimalkan kinerja model. Salah satunya adalah sebagai berikut:

- a. *n_estimators*, yang menentukan jumlah pohon keputusan yang akan dibangun dalam hutan, di mana semakin banyak pohon yang digunakan,

semakin baik hasil prediksi yang dicapai, meskipun dengan biaya komputasi yang lebih tinggi.

- b. `max_depth` mengatur kedalaman maksimum dari pohon keputusan, menentukan berapa banyak tingkat pembelahan yang diperbolehkan dalam satu pohon.
- c. `min_samples_split` mengatur jumlah minimum sampel yang dibutuhkan untuk membagi simpul pohon,
- d. `min_samples_leaf` mengontrol jumlah minimum sampel yang diperlukan di daun pohon keputusan.
- e. `max_features` membatasi jumlah maksimum fitur yang dipertimbangkan untuk setiap pembelahan simpul, yang mempengaruhi kompleksitas model dan kinerja *Random Forest*.
- f. `bootstrap` menentukan apakah teknik bootstrap (pemilihan acak dengan pengembalian) digunakan dalam membangun setiap pohon keputusan.

2.4 Penelitian Relevan

1. Penelitian milik Nurani dkk (2023), yang berjudul “Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma”. mengaplikasikan dua teknik regresi, yakni decision tree dan regresi linear berganda, untuk mengestimasi Body Mass Index (BMI) berdasarkan variabel-variabel lain pada dataset penyakit asma. Evaluasi keakuratan dilakukan menggunakan MAPE (Mean Absolute Percentage Error). Temuan penelitian menunjukkan bahwa regresi linear berganda menghasilkan performa superior dibandingkan dengan metode

decision tree berdasarkan nilai MAPE, yaitu berturut-turut sebesar 12,737%; 12,76%; 12,89%; dan 12,99% untuk proporsi data uji 10%, 20%, 30%, dan 40%. Sebaliknya, metode decision tree mencatat nilai MAPE yang sedikit lebih tinggi, yakni 12,758%; 12,79%; 12,92%; dan 13,13%. Pola hasil yang serupa juga teridentifikasi pada metrik evaluasi lain seperti MAE dan RMSE, sementara nilai R^2 menunjukkan kecenderungan sebaliknya.

2. Penelitian milik Sejati dkk (2022), dengan judul “Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan *Random Forest* untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur”. mengimplementasikan tiga algoritma klasifikasi, yaitu Naive Bayes, K-Nearest Neighbor, dan *Random Forest*. Dataset yang digunakan mencakup data penerimaan mahasiswa baru periode 2014–2019 dengan tujuan mengklasifikasikan status calon mahasiswa diterima atau mundur. Penelitian ini menggunakan 19.603 data pelatihan dan 4.901 data pengujian. Hasil analisis memperlihatkan bahwa algoritma *Random Forest* menduduki performa tertinggi dengan akurasi sebesar 73,61%, mengungguli K-Nearest Neighbor yang memperoleh 72,08% serta Naive Bayes yang mencapai 70,47%.
3. Penelitian yang dilakukan oleh Mawardi dkk (2024), dalam penelitiannya yang berjudul “Prediksi Jumlah Penerimaan Mahasiswa Baru dengan Metode *Linear Regression* dan *Exponential Smoothing*”. menggunakan pendekatan kuantitatif dengan metode komparatif. Landasan teori yang

digunakan meliputi Data Mining, Linear Regression, serta MAPE (Mean Absolute Percentage Error). Fokus utama penelitian ini adalah membandingkan ketepatan metode *Linear Regression* dan Exponential Smoothing dalam memproyeksikan jumlah penerimaan mahasiswa baru serta menentukan metode dengan prediksi paling akurat berdasarkan nilai MAPE. Hasil penelitian mengindikasikan bahwa *Linear Regression* memperoleh nilai MAPE 0,99%, yang menunjukkan tingkat ketepatan prediksi sangat tinggi. Sebaliknya, Exponential Smoothing mencatat MAPE 5,07%, yang meskipun masih tergolong akurat, namun kurang optimal dibanding Linear Regression. Oleh karenanya, *Linear Regression* direkomendasikan sebagai metode prediksi yang lebih representatif terhadap data aktual.

4. Penelitian milik Bagaskara dkk (2023), dalam penelitiannya yang berjudul “Prediksi Penerimaan Mahasiswa Baru Universitas Singaperbangsa Karawang dengan Naive Bayes”. mengembangkan sistem pendukung keputusan berbasis algoritma Naive Bayes untuk memprediksi penerimaan mahasiswa baru yang telah lolos seleksi di universitas tersebut. Variabel yang dianalisis meliputi asal sekolah calon mahasiswa, nilai akhir ujian, daya tampung universitas, probabilitas diterima, dan profil kandidat. Evaluasi menggunakan teknik data mining dengan algoritma Naive Bayes dilakukan terhadap 816 data dalam dataset. Pengujian dengan data SNBT berhasil mengklasifikasikan 4 data dengan tingkat akurasi prediksi sebesar

97,79%, menandakan keberhasilan yang sangat tinggi dalam prediksi penerimaan mahasiswa baru.

5. Penelitian oleh Susanti dkk (2022) yang berjudul “Sistem Pendukung Keputusan untuk Penentuan Lokasi Wilayah Promosi Penjarangan Mahasiswa Baru dengan Metode TOPSIS” mengembangkan sistem pendukung keputusan untuk menentukan prioritas wilayah promosi PMB di IST AKPRIND Yogyakarta. Penelitian ini menggunakan metode TOPSIS dengan beberapa kriteria penilaian, antara lain jumlah mahasiswa dari setiap kabupaten/kota dalam tiga tahun sebelumnya, jumlah sekolah, jumlah peserta didik SMA/SMK, serta jumlah perguruan tinggi pesaing di setiap provinsi. Hasil perhitungan TOPSIS menunjukkan bahwa Kabupaten Bantul memperoleh nilai preferensi tertinggi, yaitu 1,00, diikuti oleh Kota Yogyakarta, Kabupaten Sleman, Kabupaten Cilacap, dan Kabupaten Blora. Urutan prioritas tersebut didominasi oleh wilayah di Daerah Istimewa Yogyakarta dan Jawa Tengah, sedangkan wilayah alternatif di luar kedua provinsi tersebut yang direkomendasikan adalah Provinsi Nusa Tenggara Timur. Temuan ini menghasilkan rekomendasi strategis bagi institusi dalam menentukan area promosi yang paling potensial.
6. Penelitian oleh Aprilia dkk (2025) dengan judul “Komparasi Algoritma Naive Bayes dan Decision Tree pada Klasifikasi Penerimaan Peserta Didik Baru” melakukan studi komparatif untuk membandingkan performa dua algoritma klasifikasi, yaitu Naïve Bayes dan Decision Tree, dalam

memprediksi status penerimaan peserta didik baru. Penelitian ini menilai kinerja kedua algoritma menggunakan metrik akurasi, presisi, recall, dan f1-score. Hasil penelitian menunjukkan bahwa Decision Tree memberikan performa terbaik dengan akurasi 98,77%, presisi 99,19%, recall 97,56%, dan f1-score 98,34%. Sementara itu, algoritma Naïve Bayes mencatat akurasi lebih rendah, yakni 94,44%, dengan presisi 96,54%, recall 89,02%, dan f1-score 92,04%. Temuan ini menegaskan bahwa Decision Tree lebih unggul dalam mengklasifikasikan penerimaan peserta didik baru dibandingkan Naïve Bayes pada dataset yang digunakan.

Berdasarkan tinjauan pustaka yang ada, penelitian ini memiliki novelty atau kebaruan yang cukup jelas dibandingkan dengan penelitian sebelumnya. Pertama, dari sisi metodologi, penelitian ini tidak hanya menggunakan satu algoritma konvensional, melainkan mengombinasikan regresi linear berganda dan *Random Forest* regression sebagai algoritma pembanding. Pendekatan ini memberikan ruang untuk mengevaluasi keunggulan masing-masing metode, baik dari segi interpretabilitas model linear maupun kekuatan prediktif algoritma berbasis ensemble learning. Kedua, objek penelitian dalam skripsi ini lebih luas karena tidak hanya mengandalkan data historis pendaftar, tetapi juga memperhitungkan variabel-variabel lain yang relevan dalam konteks penerimaan mahasiswa baru, seperti biaya pendidikan, sarana-prasarana, serta program beasiswa.

Kebaruan penelitian ini terletak pada integrasi antara pendekatan algoritma yang berbeda dengan cakupan variabel yang lebih komprehensif,

sehingga menghasilkan analisis prediksi yang lebih representatif terhadap kondisi nyata di lapangan. Dengan demikian, penelitian ini tidak hanya mereplikasi studi terdahulu, tetapi juga memberikan kontribusi baru berupa pemodelan prediksi yang mampu menangkap kompleksitas faktor-faktor penentu penerimaan mahasiswa baru. Hasil penelitian diharapkan dapat menjadi acuan bagi pengambil kebijakan di institusi pendidikan tinggi untuk merumuskan strategi penerimaan mahasiswa yang lebih tepat sasaran dan berbasis data.

Tabel 2. 1 *State of The Art*

No	Nama	Judul	Tahun	Persamaan	Perbedaan	Hasil
1.	Alfida Tegar Nurani, Adi Setiawa,B ambang Susanto	Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma	2023	Menggunakan <i>Random Forest</i> (dalam perbandingan dengan Decision Tree) untuk prediksi	Fokus pada prediksi BMI mahasiswa dengan data asma, tidak terkait langsung dengan prediksi penerimaan mahasiswa baru.	Regresi Linear Berganda menunjukkan kinerja yang lebih baik daripada Decision Tree dalam dataset ini

No	Nama	Judul	Tahun	Persamaan	Perbedaan	Hasil
2.	Sejati, MunawarP illiang, Akbar	Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan <i>Random Forest</i> Untuk Prediksi Calon Mahasiswa Yang Diterima atau Mundur	2022	Menggunak an <i>Random Forest</i> (dalam perbanding an dengan Naïve Bayes, KNN) untuk prediksi	Pada jurnal ini membahas masalah klasifikasi atau mengkategor ikan calon mahasiswa ke dalam dua kelas diskrit: "diterima" atau "mundur".	<i>Random Forest</i> menunjukkan kinerja terbaik dengan akurasi 73,61%.
3.	Alfaruq, Mawardi, Triyono	Prediksi Jumlah Penerimaan Mahasiswa Baru dengan Metode <i>Linear Regression</i> dan	2024	Sama-sama menggunak an algoritma <i>Linear Regression</i> dalam melakukan prediksi.	Perbedaan variabel yang dibahas, pada jurnal ini menggunaka n tahun ajaran saja.	<i>Linear Regression</i> memberikan hasil prediksi yang lebih mendekati jumlah penerimaan mahasiswa

No	Nama	Judul	Tahun	Persamaan	Perbedaan	Hasil
		Exponential Smoothing			Serta penggunaan algoritma <i>Linear Regression</i> disini menggunakan <i>Linear Regression</i> sederhana.	baru yang sebenarnya dengan nilai MAPE yang lebih kecil dibandingkan Exponential Smoothing
4.	Bagaskara , Rochmawati, Adha, Trifena	Prediksi Penerimaan Mahasiswa Baru Universitas Singaperbangsa Karawang dengan Naive Bayes	2023	Sama-sama membahas terkait penerimaan mahasiswa baru	Penggunaan algortima dalam melakukan prediksi	Naive Bayes menghasilkan akurasi prediksi mencapai 97,79%,
5.	Susanti dkk	Sistem Pendukung Keputusan Untuk	(2022)	Sama-sama membahas terkait penerimaan	Perbedaan variabel penggunaan algoritma	Hasil analisis menunjukkan Kabupaten Bantul

No	Nama	Judul	Tahun	Persamaan	Perbedaan	Hasil
		Penentuan Lokasi Wilayah Promosi Penjaringan Mahasiswa Baru Dengan Metode TOPSIS		mahasiswa baru	yang diterapkan	memiliki nilai preferensi tertinggi, yaitu 1,00.
6.	Aprilia dkk	Komparasi Algoritma Naive Bayes Dan Decision Tree Pada Klasifikasi Penerimaan Peserta Didik Baru	(2025)	Sama-sama menggunakan 2 algoritma untuk studi komparasi	Perbedaan jenis algoritma yang digunakan	Algoritma Decision Tree menunjukkan kinerja tertinggi dengan akurasi 98,77%, sedangkan Naïve Bayes berada di bawahnya dengan akurasi 94,44%.

2.5 Kerangka Berpikir

Kerangka berpikir, yang juga dikenal sebagai kerangka konseptual, merupakan suatu representasi sistematis yang menggambarkan hubungan antara landasan teori dengan variabel-variabel utama yang telah diidentifikasi sebagai faktor penting dalam suatu penelitian. Kerangka ini disusun untuk memberikan alur logis dalam memahami permasalahan penelitian, sekaligus menjadi dasar dalam merumuskan hipotesis dan menentukan arah analisis yang akan dilakukan. Melalui kerangka berpikir, peneliti dapat menjelaskan bagaimana konsep-konsep teoretis saling berkaitan serta memengaruhi fenomena yang menjadi fokus kajian, sehingga proses penelitian dapat berjalan secara terstruktur dan sistematis (Syahputri dkk., 2023).

Kerangka berpikir dalam penelitian ini bertujuan untuk membandingkan dua metode prediksi, yaitu regresi linear dan Random Forest regression, dalam memprediksi penerimaan mahasiswa baru di STKIPNU Kabupaten Tegal. Melalui kerangka berpikir ini, diharapkan dapat diperoleh gambaran yang jelas mengenai perbedaan kinerja kedua metode prediksi dalam menghasilkan estimasi jumlah mahasiswa baru yang akurat. Adapun kerangka berpikir penelitian ini digambarkan dengan skema berikut.

Gambar 2. 1 Kerangka Berpikir

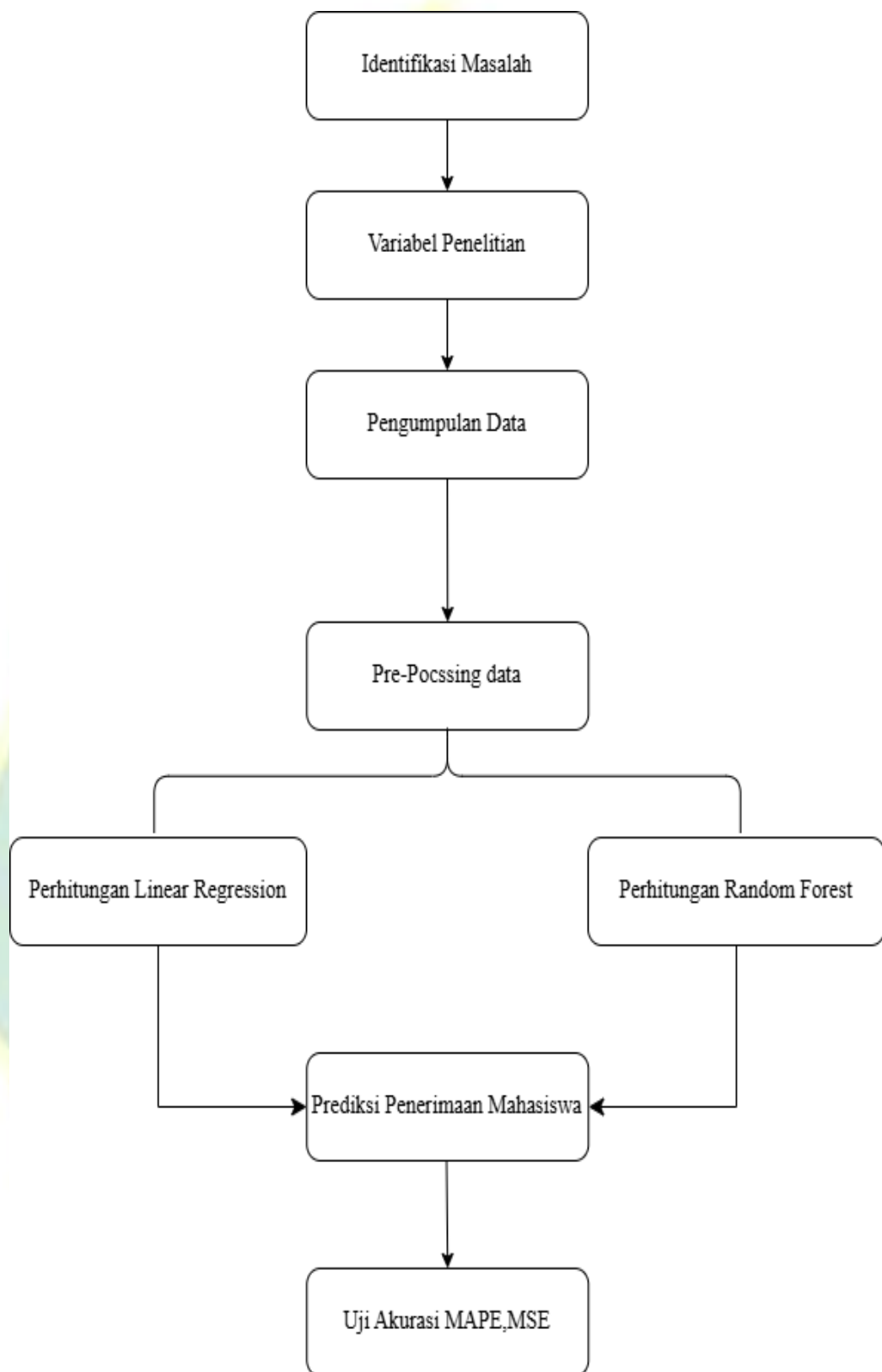


Diagram tersebut menggambarkan langkah-langkah sistematis dalam penelitian ini, dimulai dari Identifikasi Masalah, yaitu munculnya fluktuasi jumlah mahasiswa baru yang diterima di STKIP NU Kabupaten Tegal, sehingga perlunya model prediksi yang akurat untuk perencanaan kampus. Selanjutnya, penelitian menentukan Variabel Penelitian, yang mencakup variabel dependen (jumlah mahasiswa baru) dan variabel independen (seperti biaya pendidikan, fasilitas, beasiswa, dan pilihan program studi). Setelah variabel ditetapkan, tahap berikutnya adalah Pengumpulan Data, di mana data historis jumlah pendaftar dan data pendukung variabel tersebut dikumpulkan.

Data yang terkumpul kemudian melalui tahap Pre-processing, yakni persiapan data untuk analisis, meliputi pembersihan data, penanganan nilai hilang, normalisasi, dan encoding variabel kategorikal. Setelah data siap, penelitian dilanjutkan dengan dua jalur analisis: Perhitungan Regresi Linear dan Perhitungan *Random Forest*, di mana masing-masing metode digunakan untuk membangun model prediksi penerimaan mahasiswa baru.

Hasil dari kedua model kemudian digabungkan dalam tahap Prediksi Penerimaan Mahasiswa, yang bermuara pada Uji Akurasi menggunakan metrik evaluasi seperti MAPE (*Mean Absolute Percentage Error*) dan MSE (*Mean Squared Error*). Dengan menggunakan dua pendekatan estimasi tersebut, penelitian melakukan perbandingan kinerja model, hingga akhirnya dapat mengambil keputusan secara objektif mengenai metode mana yang lebih akurat dan tepat digunakan untuk memprediksi jumlah mahasiswa baru di masa mendatang.

2.6 Hipotesis Penelitian

Menurut Sugiyono (2021) hipotesis merupakan asumsi sementara yang diajukan sebagai jawaban awal terhadap pertanyaan penelitian yang dirumuskan dalam bentuk kalimat interogatif, yang nantinya akan diuji kebenarannya melalui proses penelitian.

Dalam penelitian, hipotesis biasanya ditulis dalam bentuk simbol dan terdiri dari dua jenis, yaitu hipotesis nol (H_0) dan hipotesis alternatif (H_1 atau H_a). Keduanya selalu dibuat berpasangan. Artinya, jika H_0 ditolak, maka H_1 harus diterima, dan sebaliknya. Dengan begitu, peneliti bisa membuat keputusan yang jelas tentang hasil analisisnya (Setyawan, 2021).

Adapun rumusan hipotesis yang diajukan dalam penelitian ini adalah:

- a. Hipotesis Utama (H_0): Tidak ada perbedaan yang signifikan antara hasil prediksi dengan metode Regresi Linear dan *Random Forest Regression* dalam memprediksi penerimaan mahasiswa baru di STKIP NU Kab. Tegal.
- b. Hipotesis Alternatif (H_1): Terdapat perbedaan yang signifikan antara hasil prediksi dengan metode Regresi Linear dan *Random Forest Regression* dalam memprediksi penerimaan mahasiswa baru di STKIP NU Kab. Tegal, dengan *Random Forest Regression* menunjukkan akurasi yang lebih baik dibandingkan Regresi Linear.

BAB III

METODE PENELITIAN

3.1. Pendekatan Penelitian

Dalam kajian ini, peneliti mengadopsi paradigma kuantitatif. Berdasarkan pandangan (Creswell & Creswell, 2018) pendekatan kuantitatif mengedepankan sifat penelitian yang bersifat pre-determined, di mana analisis dan interpretasi data dilakukan melalui teknik statistik. Peneliti yang menerapkan pendekatan ini bertujuan untuk menguji suatu teori dengan cara mendekomposisi hipotesis-hipotesis yang terperinci, kemudian mengumpulkan data empiris guna memverifikasi atau menolak hipotesis tersebut.

Selain itu, menurut Sugiyono (sugiyono, 2012) penelitian kuantitatif dapat dipahami sebagai suatu metodologi yang berakar pada filsafat positivisme. Metode ini digunakan untuk meneliti populasi atau sampel tertentu dengan pengumpulan data yang dilakukan melalui instrumen penelitian yang terstruktur, serta analisis data yang bersifat numerik/statistik. Tujuan utamanya adalah untuk melakukan pengujian terhadap hipotesis yang telah dirumuskan sebelumnya.

Pendekatan ini memungkinkan peneliti untuk menggambarkan hubungan antar variabel secara jelas dan sistematis, serta menghasilkan kesimpulan yang dapat diuji validitasnya. Mengacu pada latar belakang serta rumusan masalah yang telah dijelaskan, penelitian ini menerapkan pendekatan kuantitatif guna melakukan prediksi terhadap jumlah

penerimaan mahasiswa baru di STKIPNU Kabupaten Tegal. Dengan pendekatan ini, diharapkan hasil analisis dapat memberikan gambaran yang akurat dan dapat dijadikan dasar dalam pengambilan keputusan strategis oleh pihak institusi pendidikan terkait.

3.2. Metode Penelitian

Metodologi penelitian wajib ditetapkan oleh peneliti sebelum pelaksanaan studi guna memberikan kerangka acuan, arahan, serta pedoman yang sistematis selama proses penelitian berlangsung. Cresweel (2010) menyatakan bahwa “metode penelitian adalah prosedur atau teknik yang digunakan untuk memperoleh solusi atas berbagai persoalan penelitian”. Sedangkan Sugiyono (2012) menjelaskan bahwa “metode penelitian pada hakikatnya merupakan pendekatan ilmiah dalam memperoleh data dengan maksud dan tujuan tertentu”. Dengan demikian, metode penelitian berfungsi sebagai panduan bagi penulis agar proses penelitian dapat diarahkan secara tepat sehingga capaian tujuan dapat terwujud.

Penelitian ini mengaplikasikan pendekatan komparatif. Menurut Sugiyono, penelitian komparatif merupakan kajian yang membandingkan keberadaan satu atau lebih variabel pada dua kelompok sampel atau lebih yang diambil pada waktu berbeda (Abdhul, 2022). Selain itu, Nazir dalam buku karya Ibrahim dkk. mengemukakan bahwa penelitian komparatif adalah jenis penelitian deskriptif yang bertujuan untuk menemukan jawaban fundamental mengenai hubungan sebab-akibat dengan

menganalisis faktor-faktor penyebab munculnya suatu fenomena (Ibrahim, et al., 2018).

Secara lebih operasional, penelitian komparatif merupakan suatu investigasi yang terstruktur di mana peneliti tidak melakukan manipulasi terhadap variabel bebas karena variabel tersebut telah ada secara faktual atau memang tidak dapat diubah. Dengan kata lain, peneliti hanya mengumpulkan dan menganalisis data yang telah tersedia di lapangan tanpa melakukan intervensi terhadap data tersebut (Emzir, 2013).

Jadi, metode komparatif adalah jenis metode dengan membandingkan dua hal atau lebih sehingga ditemukan gambaran yang lebih rinci dari berbagai sumber. Selain itu, dapat ditemukan perbedaan serta persamaan dari hal yang dibandingkan.

3.3. Instrument Penelitian

3.3.1 Sumber Data

Penelitian tidak dapat dipisahkan dari keberadaan data sebagai bahan mentah utama dalam perolehan informasi yang valid dan akurat. Dalam studi ini, peneliti memanfaatkan data kuantitatif yang bersumber dari sumber primer maupun sumber sekunder. Menurut (Arikunto, 2010) “sumber data penelitian merupakan entitas atau subjek tempat data dapat diakses atau diperoleh.” Adapun klasifikasi sumber data dalam penelitian ini adalah sebagai berikut.

a. Data Primer

Data primer merupakan data yang diperoleh secara langsung melalui pengamatan atau pengukuran empiris terhadap subjek atau pelaku yang terlibat secara langsung dalam objek penelitian. Data ini dikumpulkan dan diolah secara mandiri oleh peneliti sehingga memiliki karakteristik sebagai data asli atau data orisinal yang terbaru (Siyoto & Sodik, 2015).

Data primer dalam penelitian ini yakni Informasi yang diperoleh melalui wawancara langsung atau komunikasi dengan wakil ketua 3 di STKIPNU Kabupaten Tegal.

Data primer dalam penelitian ini, secara terperinci dijelaskan pada tabel berikut:

Tabel 3. 1 Data primer

No.	Jenis Data Primer	Sumber Data
1.	Jumlah Dosen	Waket 3
2.	Visi dan misi institusi	Waket 3
3.	Tata letak institusi	Waket 3

b. Data Sekunder

Data sekunder merupakan data yang sudah tersedia sebelumnya dan diperoleh dari sumber lain. merupakan data yang diperoleh peneliti dari berbagai sumber terdahulu (peneliti sebagai tangan kedua) (Dr.Sandu Siyoto & M.Ali sodik, 2015, p. 68).

Adapun dalam penelitian ini menggunakan data sekunder berupa Data historis penerimaan mahasiswa baru dari arsip resmi STKIPNU Kabupaten Tegal, termasuk jumlah pendaftar, dan jumlah yang diterima.

Data sekunder dalam penelitian ini, secara terperinci dijelaskan pada tabel berikut.

Tabel 3. 2 Data Historis

No.	Nama	Prodi	Beasiswa	Promosi	Sarpras	Biaya
1	Neisya Widayani	pe	Ya	Ya	Tidak	4800000
2	Agus Ramdani	pi	Ya	Ya	Baik	2200000
3	Nur Zulfa Khaerunnisa	pi	Ya	Ya	Baik	4100000
4	Neina Febrihaf Sari	bk	Ya	Ya	Tidak	2400000
5	Rina Rahmawati	pe	Tidak	Tidak	Baik	7150000
6	Aris Setiawan	pi	Ya	Ya	Baik	4100000
7	Dani Rohidi	bk	Ya	Ya	Baik	3700000
8	Saskiya Azka Safitri S.H	pi	Ya	Tidak	Baik	2500000
9	Zidan Nurrohim	pi	Ya	Ya	Baik	1500000
10	Rizqi Maulida	pe	Tidak	Ya	Baik	7150000
...	...	...	...	...	...	...

3.3.2 Teknik Pengumpulan Data

a. Data Time Series

Data time series adalah data yang terdiri dari hasil pengamatan yang dicatat secara berurutan berdasarkan waktu dengan selang waktu yang konsisten. Jenis data ini umum digunakan dalam kehidupan sehari-hari karena biasanya dikumpulkan dalam interval waktu tertentu, seperti setiap hari, minggu, atau bulan (Al'afi dkk., 2020). Menurut Supranto (1987), data time series atau data berkala adalah data yang dikumpulkan secara berurutan berdasarkan waktu untuk memantau perkembangan suatu aktivitas. Data berkala juga berguna dalam mengidentifikasi pola variabel yang dapat mendukung pengambilan keputusan dan perencanaan masa depan.

Umumnya, pola dalam data time series terdiri dari empat komponen utama, yaitu:

- 1) Tren jangka panjang atau sekuler, yaitu arah umum pergerakan data dalam periode waktu yang luas.
- 2) Gerakan siklis (cyclical movements), Gerakan siklis, berupa perubahan naik turun dalam jangka panjang yang mungkin terjadi secara berulang.
- 3) Pola musiman (seasonal movement), yakni variasi berulang yang muncul secara teratur setiap tahun akibat faktor musiman atau kebiasaan tahunan.

- 4) Gerakan acak (irregular or random movements), yaitu perubahan tak terduga yang timbul karena kejadian seperti bencana, krisis, atau peristiwa sosial-politik lainnya.

b. Angket

Menurut Darwin dkk (Ititi dkk., 2022), angket merupakan cara pengumpulan data dengan menyediakan daftar pertanyaan atau pernyataan dalam bentuk kuesioner untuk diisi oleh responden sesuai dengan kebutuhan dari masing-masing variable penelitian. Angket yang dimaksud dalam penelitian ini adalah daftar pertanyaan yang diberikan kepada responden dengan model pembelajaran untuk mengetahui tanggapan responden tentang hasil belajar selama proses pelaksanaan penggunaan model pembelajaran diterapkan.

c. Wawancara

Wawancara merupakan salah satu metode yang digunakan dalam pengumpulan data penelitian melalui interaksi dua arah antara peneliti dan responden. Teknik ini melibatkan percakapan langsung (tatap muka) di mana peneliti mengajukan pertanyaan yang telah disiapkan sebelumnya untuk menggali informasi terkait objek penelitian. Dalam penelitian ini, digunakan jenis wawancara semiterstruktur, yaitu wawancara yang memungkinkan peneliti mengembangkan

pertanyaan secara fleksibel namun tetap berpedoman pada kerangka wawancara yang telah dirancang. Menurut Sugiyono (2012), wawancara semiterstruktur termasuk dalam kategori *in-depth interview*, yang memungkinkan peneliti mengeksplorasi informasi lebih dalam dibandingkan wawancara terstruktur.

3.4. Tahapan Penelitian

3.4.1. Pengumpulan Data

Proses pengumpulan data diperlukan untuk mendapatkan data yang relevan terkait dengan penelitian yang dilakukan. Dalam penelitian ini penulis menggunakan data berupa data historis pada penerimaan mahasiswa baru ditahun-tahun sebelumnya untuk memprediksi penerimaan mahasiswa baru ditahun depan. Data tersebut dilakukan penulis melalui wawancara kepada pihak intitusi, yaitu ibu ayu siska tri mayasari M.Pd selaku wakil ketua 3 dari STKIP NU Kab.Tegal. Data yang yang didapatkan berupa data historis pada tahun 2018-2024.

3.4.2. Pengolahan Data Awal

Penelitian ini menganalisis data penerimaan mahasiswa baru dengan mengadopsi kerangka Knowledge Discovery in Databases (KDD). KDD merupakan metode pencarian informasi dari data dalam jumlah yang besar dan mampu menjadi sebuah solusi. KDD merupakan terminology computer science yang mengacu pada proses dalam melakukan pencarian informasi yang

berguna pada dataset. Di dalam KDD, proses data mining dilakukan sebagai bentuk pendekatan keseluruhan terhadap penemuan pengetahuan sebagai langkah spesifik.

a. Data *Preprocessing*

1) Data *Cleaning*

Data yang telah dikumpulkan melalui kuisioner dan data historis akan diperiksa untuk memastikan tidak ada nilai yang tidak relevan atau kosong (*missing values*). Dalam hal ini menghapus atribut-atribut yang tidak mempengaruhi suatu prediksi yaitu berupa atribut NIK, Tempat tanggal lahir, NISN, No.Hp, dan Nama Orang tua Siswa.

2) Data Labeling

Data labeling dilakukan untuk memberikan penanda atau kategori pada data yang akan digunakan sebagai target variabel (label) dalam proses pelatihan model.

3) Data Transformation

Tahap ini berfungsi untuk melakukan pengubahan struktur atau nilai-nilai dalam dataset guna meningkatkan kualitas atau model *Machine Learning* agar data lebih mudah dipahami.

4) Data splitting

Pada tahap ini dibagi menjadi dua subset yaitu data pelatihan yang diterapkan dalam melatih model *Linear Regression* serta

Random Forest sedangkan data pengujian diterapkan dalam mengevaluasi kinerja model.

b. Penerapan data

Tabel 3. 3 Beasiswa

Status Beasiswa	Nilai Numerik
Menerima	1
Tidak Menerima	0

Tabel 3. 4 Promosi

Jumlah promosi pertahun	Nilai
2017	10
2018	18
2019	24
2020	24
2021	24
2022	35
2023	35
2024	35

Tabel 3. 5 Sarana dan Prasarana

Sarana Prasarana	Nilai Numerik
Sangat lengkap	5
Lengka	4
Cukup Lengkap	3

Kurang lengkap	2
Tidak lengkap	1

Tabel 3. 6 Nilai Min dan Max

Biaya	Nilai Numerik
Min	Rp 200,000
Max	Rp 5,800,000
Diference	Rp 5,600,000

Tabel 3. 7 Biaya

Nilai	Normalisasi
Rp 700,000	0.09
Rp 700,000	0.09
Rp 700,000	0.09
Rp 700,000	0.09
Rp 700,000	0.09
Rp 4,500,000	0.77
Rp 700,000	0.09
Rp 200,000	0.00
Rp 4,500,000	0.77
Rp 200,000	0.00
...	...

Tabel 3. 8 Jumlah mahasiswa per tahun

Tahun	Jumlah mahasiswa
2017	54

Tahun	Jumlah mahasiswa
2018	54
2019	110
2020	150
2021	86
2022	27
2023	24
2024	63

Tabel 3. 9 Estimasi Jarak

Estimasi Jarak
15
40
1
20
11
53
20
18

c. Metode/Model yang digunakan

Pada tahap ini akan diimplementasikan algoritma regresi linear dan *Random Forest* untuk melakukan prediksi terhadap penerimaan mahasiswa baru. Data yang sudah dibersihkan dan disiapkan sebelumnya akan digunakan untuk melatih kedua algoritma agar dapat mengenali pola-pola yang ada. Metode

Random Forest akan membentuk sejumlah pohon keputusan untuk mengetahui faktor mana yang paling memengaruhi keputusan calon mahasiswa dalam mendaftar. Sedangkan regresi linear digunakan untuk melihat seberapa kuat hubungan antara faktor-faktor internal dengan jumlah penerimaan mahasiswa baru.

1) Regresi Linear

Setelah melewati tahap pre-processing seperti pembersihan hingga transformasi nilai, maka menghasilkan data sebagai berikut:

Tabel 3. 10 Dataset Linear Regression

Y	X1	X2	X3	X4	X5	X6
54	7	1	3	15	10	0,09
54	7	1	4	40	10	0,09
54	7	1	4	1	10	0,09
54	7	1	5	20	10	0,09
54	7	1	4	11	10	0,09
54	7	0	4	53	10	0,77
54	7	1	3	20	10	0,09
110	8	1	2	18	18	0,00
110	8	0	2	4	18	0,77

Lalu mencari nilai X_1Y , X_2Y , X_3Y , X_4Y sebagai berikut:

Tabel 3. 11 perhitungan nilai $X_n Y$

$x_1 y$	$x_2 y$	$x_3 y$	$x_4 y$	$x_5 y$	$x_6 y$
378	54	162	810	540	4,82
378	54	216	2160	540	4,82
378	54	216	54	540	4,82
378	54	270	1080	540	4,82
378	54	216	594	540	4,82
378	0	216	2862	540	41,46
378	54	162	1080	540	4,82
880	110	220	1980	1980	0,00
880	0	220	440	1980	84,46
880	110	220	2310	1980	0,00

Setelah itu mencari nilai $X_1 X_2$ dst...

Tabel 3. 12 perhitungan nilai $X_n X_n$

$x_1 x_2$	$x_1 x_3$	$x_1 x_4$	$x_1 x_5$	$x_1 x_6$	$x_2 x_3$	$x_2 x_4$	$x_2 x_5$	$x_2 x_6$	$x_3 x_4$	$x_3 x_5$	$x_3 x_6$	$x_4 x_5$	$x_4 x_6$	$x_5 x_6$
7	21	105	70	0.63	3	15	10	0.09	45	30	0.27	150	1.34	0.89
7	28	280	70	0.63	4	40	10	0.09	160	40	0.36	400	3.57	0.89
7	28	7	70	0.63	4	1	10	0.09	4	40	0.36	10	0.09	0.89
7	35	140	70	0.63	5	20	10	0.09	100	50	0.45	200	1.79	0.89
7	28	77	70	0.63	4	11	10	0.09	44	40	0.36	110	0.98	0.89
0	28	371	70	5.38	0	0	0	0.00	212	40	3.07	530	40.70	7.68
7	21	140	70	0.63	3	20	10	0.09	60	30	0.27	200	1.79	0.89
8	16	144	144	0.00	2	18	18	0.00	36	36	0.00	324	0.00	0.00
0	16	32	144	6.14	0	0	0	0.00	8	36	1.54	72	3.07	13.82
8	16	168	144	0.00	2	21	18	0.00	42	36	0.00	378	0.00	0.00

Selanjutnya menghitung nilai X_1^2 dst...

Tabel 3. 13 perhitungan $X_n Y$

X_1^2	X_2^2	X_3^2	X_4^2	X_5^2	X_6^2
49	1	9	225	100	0.01
49	1	16	1600	100	0.01
49	1	16	1	100	0.01
49	1	25	400	100	0.01
49	1	16	121	100	0.01
49	0	16	2809	100	0.59
49	1	9	400	100	0.01
64	1	4	324	324	0.00
64	0	4	16	324	0.59
64	1	4	441	324	0.00

Setelah menghitung nilai $X_1 \dots X_4$, maka langkah berikutnya adalah membuat matriks dari hasil perkalian X dan Y dengan pola sebagai berikut:

$$A = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 & \sum x_4 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \sum x_1 x_3 & \sum x_1 x_4 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \sum x_2 x_3 & \sum x_2 x_4 \\ \sum x_3 & \sum x_1 x_3 & \sum x_2 x_3 & \sum x_3^2 & \sum x_3 x_4 \\ \sum x_4 & \sum x_1 x_4 & \sum x_2 x_4 & \sum x_3 x_4 & \sum x_4^2 \end{bmatrix}$$

$$H = \begin{bmatrix} \sum Y \\ \sum x_1Y \\ \sum x_2Y \\ \sum x_3Y \\ \sum x_4Y \end{bmatrix}$$

Sehingga jika diterapkan pada dataset yang ada
maka sebagai berikut:

(Matriks A)

Tabel 3. 14 Matriks A

100	1608	45	237	1594	2441	65.48	8906
1608	27724	638	3699	24840	41282	1163.96	137100
45	638	45	111	621	1076	18.52	3877
237	3699	111	631	3766	5511	146.70	21009
1594	24840	621	3766	45060	38038	1063.45	142451
2441	41282	1076	5511	38038	64467	1755.89	206646
65.48	1163.96	18.52	146.70	1063.45	1755.89	56.43	5217.392857

(Matriks A1)

Tabel 3. 15 Matriks A1

8906	1608	45	237	1594	2441	65.48
137100	27724	638	3699	24840	41282	1163.96
3877	638	45	111	621	1076	18.52
21009	3699	111	631	3766	5511	146.70
142451	24840	621	3766	45060	38038	1063.45
206646	41282	1076	5511	38038	64467	1755.89
5217.39	1163.96	18.52	146.70	1063.45	1755.89	56.43

(Matriks A2)

Tabel 3. 16 Matriks A2

100	8906	45	237	1594	407	65.48
1608	137100	638	3699	24840	6421	1163.96
45	3877	45	111	621	193	18.52
237	21009	111	631	3766	983	146.70
1594	142451	621	3766	45060	6524	1063.45
2441	206646	193	983	6524	1745	261.05
65.48	5217.39	18.52	146.70	1063.45	261.05	56.43

(Matriks A3)

Tabel 3. 17 Matriks A3

100	1608	8906	237	1594	407	65.48
1608	27724	137100	3699	24840	6421	1163.96
45	638	3877	111	621	193	18.52
237	3699	21009	631	3766	983	146.70
1594	24840	142451	3766	45060	6524	1063.45
2441	6421	206646	983	6524	1745	261.05
65.48	1163.96	5217.39	146.70	1063.45	261.05	56.43

(Matriks A4)

Tabel 3. 18 Matriks A4

100	1608	45	8906	1594	407	65.48
1608	27724	638	137100	24840	6421	1163.96
45	638	45	3877	621	193	18.52
237	3699	111	21009	3766	983	146.70
1594	24840	621	142451	45060	6524	1063.45
2441	6421	193	206646	6524	1745	261.05
65.48	1163.96	18.52	5217.39	1063.45	261.05	56.43

(Matriks A5)

Tabel 3. 19 Matriks A5

100	1608	45	237.00	8906	407	65.48
1608	27724	638	3699.00	137100	6421	1163.96
45	638	45	111.00	3877	193	18.52
237	3699	111	631.00	21009	983	146.70
1594	24840	621	3766.00	142451	6524	1063.45
2441	6421	193	983.00	206646	1745	261.05
65.48	1163.96	18.52	146.70	5217.39	261.05	56.43

(Matriks A6)

Tabel 3. 20 Matriks A6

100	1608	45	237.00	1594	8906	65.48
1608	27724	638	3699.00	24840	137100	1163.96
45	638	45	111.00	621	3877	18.52
237	3699	111	631.00	3766	21009	146.70
1594	24840	621	3766.00	45060	142451	1063.45
2441	41282	1076	5511.00	38038	206646	261.05
65.48	1163.96	18.52	146.70	1063.45	5217.39	56.43

(Matriks A7)

Tabel 3. 21 Matriks A7

100	1608	45	237.00	1594	2441	8906
1608	27724	638	3699.00	24840	41282	137100
45	638	45	111.00	621	1076	3877
237	3699	111	631.00	3766	5511	21009
1594	24840	621	3766.00	45060	38038	142451
2441	41282	1076	5511.00	38038	64467	206646
65.48	1163.96	18.52	146.70	1063.45	1755.89	5217.39

Dari masing-masing matriks ini kemudian dicari lagi nilai determinannya yang mana hasil-nya sebagai berikut:

Tabel 3. 22 Nilai Determinan

Determinan (A)	3.64475E+16
Determinan (A1)	7.57511E+18
Determinan (A2)	-1.28522E+17
Determinan (A3)	1.04721E+19
Determinan (A4)	3.1821E+18
Determinan (A5)	3.59705E+16
Determinan (A6)	-2.51228E+18
Determinan (A7)	-2.61011E+18

Setelah diketahui nilai determinan-nya barulah menghitung nilai $\beta_1 \dots \beta_n$ dengan rumus sebagai berikut:

$$\frac{\text{Det}(A_1 \dots A_n)}{\text{Det}(A)}$$

Yang setelah diterapkan mneghasilan nilai sebagai berikut:

Tabel 3. 23 Nilai β

β_1	β_2	β_3	β_4	β_5	β_6	β_7
207.84	-0.017	-81.482	0.3039	0.0113	-69.8	1.04

Setelah nilai $b_1 \dots b_7$ ditemukan maka nilai tersebut sudah bisa diterapkan pada persamaan regresi linear, sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_N X_N + e$$

d. Random Forest

Didapatkan data dari hasil pre-poscressing sebagai berikut:

Tabel 3. 24 Dataset Random Forest

No.	Y	X1	X2	X3	X4	X5	X6
1	54	7	1	3	15	10	0.09
2	54	7	1	4	40	10	0.09
3	54	7	1	4	1	10	0.09
4	54	7	1	5	20	10	0.09
5	54	7	1	4	11	10	0.09
6	54	7	0	4	53	10	0.77
7	54	7	1	3	20	10	0.09

No.	Y	X1	X2	X3	X4	X5	X6
8	110	8	1	2	18	18	0.00
9	110	8	0	2	4	18	0.77
10	110	8	1	2	21	18	0.00

Dari data set ini dilakukan bootstrapping /nomor random dari 1-100 menghasilkan dataset sebagai berikut:

Tabel 3. 25 penambahan nomor hasil bootstrapping

No.Acak	x1	X2	X3	X4	X5	X6	Y
34	19	0	2	7	24	0.86	86
66	19	1	4	2	24	0.14	86
94	18	1	3	10	35	0.96	63
88	18	1	2	20	35	0.96	24
72	19	0	2	10	24	0.86	86
88	18	1	2	20	35	0.96	24
56	19	0	2	17	24	0.86	86
12	8	1	2	3	18	0.00	110
6	7	0	4	53	10	0.77	54
83	19	1	3	17	35	0.95	27

Setelah ditentukan tabel bootstrapping maka tentukan juga nilai squared error dengan rumus:

$$\text{Squared error} = \left(\frac{\text{jumlah pendaftar}}{\text{rata-rata jumlah pendaftar}} \right)^2$$

Tabel 3. 26 perhitungan squares Error

squared error
2596.92
2596.92
781.76
121.88
2596.92
121.88
2596.92
5619.00
359.48
64.64

Setelah squared error dihitung maka perhitungan MSE awal sudah bisa dihitung dengan cara:

$$\text{MSE awal} = \frac{\text{hasil penjumlahan Squared Error}}{\text{jumlah dataset}}$$

Karena disini menggunakan 10 data set maka pembaginya yaitu 10 sehingga menghasilkan MSE awal 519.84. MSE ini digunakan untuk mengukur seberapa baik split mengurangi kesalahan prediksi.

Mengevaluasi masing-masing atribut untuk split terbaik.

- a. Kelompok (dosen < 12)

Tabel 3. 27 menghitung MSE pada atribut Beasiswa = 0

x1	X2	X3	X4	X5	X6	Y
8	1	2	3	18	0.00	110
7	0	4	53	10	0.77	54
9	0	1	45	18	0.09	150
9	1	3	30	18	0.09	150
7	1	4	1	10	0.09	54
9	0	1	45	18	0.09	150

Ditemukan jumlah data (n_0) = 23, nilai rata-rata jumlah

mahasiswa ($\hat{y}$) = $\left(\frac{110+54+150+150+54+150+n}{23} \right) = 106.77$, maka

$$MSE0 = \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 1909.63.$$

Kelompok (dosen > 12)

Tabel 3. 28 menghitung MSE pada atribut Beasiswa = 1

x1	X2	X3	X4	X5	X6	Y
19	0	2	7	24	0.86	86
19	1	4	2	24	0.14	86
18	1	3	10	35	0.96	63
18	1	2	20	35	0.96	24
19	0	2	10	24	0.86	86
18	1	2	20	35	0.96	24

Ditemukan jumlah data (n_0)= 76, nilai rata-rata jumlah

mahasiswa ($\hat{y}$) = $\left(\frac{86+86+63+n}{76}\right) = 77.28$, maka MSE0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 1040.25.$$

b. Kelompok (beasiswa = 0)

Tabel 3. 29 menghitung MSE pada Tabel Promosi = 0

x1	X2	X3	X4	X5	X6	Y
19	0	2	7	24	0.86	86
19	0	2	10	24	0.86	86
19	0	2	17	24	0.86	86
7	0	4	53	10	0.77	54
18	0	3	13	35	0.96	63
15	0	2	31	24	0.86	150

Ditemukan jumlah data (n_0)= 60, nilai rata-rata jumlah

mahasiswa ($\hat{y}$) = $\left(\frac{86+86+86+54+63+n}{60}\right) = 89.45$, maka MSE0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 857.55.$$

Kelompok (beasiswa = 1)

Tabel 3. 30 menghitung MSE pada Tabel Promosi = 1

x1	X2	X3	X4	X5	X6	Y
19	1	4	2	24	0.14	86
18	1	3	10	35	0.96	63
18	1	2	20	35	0.96	24
18	1	2	20	35	0.96	24
8	1	2	3	18	0.00	110
19	1	3	17	35	0.95	27

Ditemukan jumlah data (n_0)= 40, nilai rata-rata jumlah

mahasiswa ($\hat{y}$) = $\left(\frac{86+63+24+24+110+n}{40}\right) = 74.075$, maka MSE0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 2013.52.$$

c. Kelompok (sarpras<4)

Tabel 3. 31 menghitung MSE dari atribut Promosi = 1

X1	X2	X3	X4	X5	X6	Y
19	0	2	7	24	0.86	86
18	1	3	10	35	0.96	63
18	1	2	20	35	0.96	24
19	0	2	10	24	0.86	86
18	1	2	20	35	0.96	24
19	0	2	17	24	0.86	86

Ditemukan jumlah data (n_0)= 88, nilai rata-rata jumlah

mahasiswa ($\hat{y}$) = $\left(\frac{86+23+24+n}{88}\right) = 85.56$, maka MSE0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 1387.2.$$

d. Kelompok (jarak<20)

Tabel 3. 32 menghitung MSE dari Atribut Sarpras = 0

X1	X2	X3	X4	X5	X6	Y
19	0	2	7	24	0.86	86
19	1	4	2	24	0.14	86
18	1	3	10	35	0.96	63
18	1	2	20	35	0.96	24
19	0	2	10	24	0.86	86
18	1	2	20	35	0.96	24

Ditemukan jumlah data (n_0)= 82, nilai rata-rata jumlah mahasiswa ($\hat{y}$) = $\left(\frac{86+86+63+n}{82}\right)=82.85$, maka MSE_0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 1145.91$$

e. Biaya <0.45

Tabel 3. 33 menghitung MSE pada atribut Biaya<0.16

X1	X2	X3	X4	X5	X6	Y
19	1	4	2	24	0.14	86
8	1	2	3	18	0.00	110
9	0	1	45	18	0.09	150
19	1	4	2	24	0.14	86
9	1	3	30	18	0.09	150
7	1	4	40	10	0.09	54
19	1	3	8	24	0.14	86

Ditemukan jumlah data (n_0)= 22, nilai rata-rata jumlah mahasiswa ($\hat{y}$) = $\left(\frac{86+110+150}{22}\right)= 108.76$, maka MSE_0

$$= \frac{\text{jumlah squared error}}{\text{jumlah data } (n_0)} = 1279.28.$$

Ketika semua MSE telah ditentukan nilainya maka selanjutnya menentukan nilai weighted Average MSE yang mana bisa dihitung menggunakan rumus:

$$\text{weighted Average MSE} = \left(\frac{n_1}{N} \text{MSE}_1\right) + \left(\frac{n_2}{N} \text{MSE}_2\right)$$

dimana:

n_i = jumlah data di kelompok-i(hasil split)

N = Total jumlah data di node induk (sebelum di split)

MSE_i = *Mean Squared Error* dari kelompok ke – i

Jadi dari masing-masing dataset menghasilkan weighted Average sebagai berikut

Tabel 3. 34 nilai weighted Average MSE

weighted average MSE	Beasiswa	Dosen	Sarpras	Jarak	Biaya
	1319.94	1210.71	1290.26	1375.76	366.72

Setelah menghitung weighted Average MSE untuk setiap atribut, maka berdasarkan tabel diatas bandingkan hasilnya dan gunakan nilai weighted Average MSE terkecil, karena menunjukan split paling efektif dalam mengurangi variasi data.

3.5. Tempat Penelitian

Lokasi dalam penelitian ini dilaksanakan di STKIPNU Kab.Tegal yang terletak di Jl. Jend. A. Yani No. 21 Procot Slawi, Kab. Tegal, sebuah perguruan tinggi yang berfokus pada pengembangan pendidikan di wilayah Kabupaten Tegal. Penelitian dilakukan dengan menggunakan data historis penerimaan mahasiswa baru yang dimiliki oleh institusi ini, termasuk data terkait jumlah pendaftar, jumlah mahasiswa diterima, program studi, dan tahun ajaran tertentu.

Objek penelitian ini adalah metode prediksi penerimaan mahasiswa baru di STKIPNU Kabupaten Tegal menggunakan regresi linear dan *Random Forest*. Variabel bebasnya meliputi faktor-faktor seperti jumlah pendaftar, jumlah dosen, beasiswa, sarana dan prasana, jarak, jumlah promosi pertahun , dan Biaya (variabel X), sedangkan variabel terikatnya adalah jumlah mahasiswa baru yang diterima (variabel Y).



BAB IV

HASIL DAN PEMBAHASAN

4.1 Atribut Data

Setelah melalui tahap kajian literatur yang dilakukan secara mendalam dan terstruktur, langkah berikutnya dalam proses penelitian ini adalah melaksanakan pengolahan data secara sistematis dan terencana sehingga data yang diperoleh dapat dianalisis serta diinterpretasikan secara lebih komprehensif untuk menjawab rumusan masalah sekaligus menguji hipotesis penelitian yang telah dirumuskan sebelumnya. Dalam pelaksanaannya, teknik pengumpulan data dilakukan melalui metode dokumentasi, di mana peneliti memperoleh beragam informasi penting yang meliputi jumlah dosen pengajar, jumlah kegiatan promosi yang dilaksanakan setiap tahun, estimasi jarak antara rumah mahasiswa dengan perguruan tinggi, serta jumlah mahasiswa baru pada masing-masing program studi setiap tahunnya. Pada tahap ini peneliti juga menghimpun informasi yang lebih rinci terkait komponen biaya pendidikan, yang mencakup biaya pendaftaran, Sumbangan Pengembangan Institusi (SPI), Sumbangan Pembinaan Pendidikan (SPP), biaya kemahasiswaan, hingga biaya Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS).

Selain metode dokumentasi, peneliti juga mengumpulkan data tambahan melalui penyebaran kuesioner untuk menilai kelengkapan sarana dan prasarana STKIP NU Kabupaten Tegal berdasarkan pengalaman

mahasiswa. Kuesioner ini diberikan kepada mahasiswa angkatan 2017–2024 guna memperoleh gambaran yang akurat dan representatif. Data yang terkumpul kemudian diolah sebagai dasar prediksi jumlah calon mahasiswa baru di masa mendatang, sehingga hasil analisis dapat mendukung pengambilan keputusan strategis dan penyusunan kebijakan pengembangan kampus.

Adapun data yang dimaksud disajikan secara lengkap pada bagian berikut.

Tabel 4. 1 Atribut

Y	X1	X2	X3	X4	X5	X6	
54	7	1	3	15	5	Rp	700,000
54	7	1	4	40	5	Rp	700,000
54	7	1	4	1	5	Rp	700,000
54	7	1	5	20	5	Rp	700,000
54	7	1	4	11	5	Rp	700,000
54	7	0	4	53	5	Rp	4,500,000
54	7	1	3	20	4	Rp	700,000
110	8	1	2	18	4	Rp	200,000
110	8	0	2	4	4	Rp	4,500,000
110	8	1	2	21	4	Rp	200,000
...	...	...	...	...	...	...	
63	18	1	2	8	3	Rp	5,800,000

Y	X1	X2	X3	X4	X5	X6	
63	18	1	1	3	5	Rp	5,800,000
Y	X1	X2	X3	X4	X5	X6	
63	18	0	1	17	3	Rp	5,600,000
63	18	1	2	20	5	Rp	5,600,000
63	18	0	3	13	4	Rp	5,600,000
63	18	1	3	10	5	Rp	5,600,000
63	18	0	1	10	4	Rp	5,600,000
63	18	1	1	20	3	Rp	5,600,000
63	18	1	2	8	4	Rp	5,800,000
63	18	0	2	10	5	Rp	5,800,000
63	18	1	2	21	5	Rp	5,600,000
63	18	1	3	6	5	Rp	5,600,000

Pada tabel diatas Y adalah variabel dependen atau variabel yang ingin di prediksi berupa jumlah calon mahasiswa yang akan datang, sehingga pada atribut data diatas berupa jumlah mahasiswa dari tahun-tahun sebelumnya. Sedangkan x ialah variabel independen atau factor-faktor yang akan digunkana untuk menjelaskan prediksi dari Y.

Adapun variabel X1 merupakan jumlah dosen. Peneliti memperoleh daftar lengkap seluruh dosen dari sumber resmi, namun karena kebutuhan penelitian berfokus pada jumlah dosen per tahun, data tersebut kemudian dikelompokkan kembali berdasarkan tahun. Proses

pengelompokan ini dilakukan dengan mengacu pada informasi yang tercantum di Pangkalan Data Pendidikan Tinggi (PDDikti).

X2 merupakan status mahasiswa apakah ia menerima beasiswa atau tidak. Peneliti mendapatkannya dari staf administrasi di perguruan tinggi ini. Data ini diperoleh dalam satu berkas yang sama dengan data jarak dan nama dosen.

X3 merupakan atribut sarana dan prasana. Adapun data ini peneliti dapatkan dari sebar angket pada mahasiswa STKIPNU dari 2017-2024. Hal ini dilakukan karena pengalaman tentang sarana dan prasarana di tiap tahunnya pasti akan ada perbedaan. Sebar angket ini menggunakan satuan skala likert sangat lengkap 5, lengkap 4, cukup lengkap 3, kurang lengkap 2, tidak lengkap 1.

X4 merupakan variabel jarak, yaitu jarak tempat tinggal calon mahasiswa ke kampus. Data mengenai jarak ini peneliti peroleh melalui staf bagian administrasi perguruan tinggi, yang menyediakan informasi alamat domisili mahasiswa. Selanjutnya, peneliti melakukan pengolahan data dengan menyesuaikan alamat tersebut sehingga dapat dihitung dan dikelompokkan berdasarkan jarak tempuh ke kampus sebagai variabel penelitian.

X5 merupakan jumlah kegiatan promosi. Data ini mencakup frekuensi promosi yang diselenggarakan STKIP NU, baik secara luring misalnya kunjungan ke sekolah-sekolah menengah atas maupun kegiatan

sosialisasi tatap muka maupun secara daring, seperti publikasi melalui media sosial, situs web resmi, dan saluran media massa lainnya.

X6 merupakan variabel yang merepresentasikan atribut biaya pendidikan. Nilai pada variabel ini diperoleh dengan menjumlahkan seluruh komponen biaya yang wajib dibayarkan mahasiswa, meliputi biaya pendaftaran, Sumbangan Pengembangan Institusi (SPI), Sumbangan Pembinaan Pendidikan (SPP), biaya kemahasiswaan, serta biaya ujian tengah semester (UTS) dan ujian akhir semester (UAS).

4.2 Data Split

Proses berikutnya setelah pengumpulan data dan data melalui tahap pra-olah adalah membagi dataset menjadi dua yaitu data train atau train dataset dan data test atau test dataset. Proses ini merupakan bagian penting dalam penelitian ini yaitu mencari rasio data split terbaik untuk model time series pada prediksi penerimaan mahasiswa baru. Beberapa *rule of thumb* atau referensi dalam membagi data pertama berdasarkan professor Andrew Ng mengatakan *rule of thumb* dalam pembagian data split yaitu 70:30 menjadi mungkin terjadi pada data 100,1000 atau 10000, namun pada jumlah data 1000000 menjadi tidak sesuai dengan rasio 70:30 maka dipakailah split data 99:1. Selain itu terdapat *rule of thumb* lain dalam pembagian data yaitu 80:20 atau dikenal dengan *pareto principle* yang sering dipakai dalam matematika, ekonomi dan komputer (Adinugroho, 2022).

4.3 Linear regression

Setelah seluruh data penelitian terkumpul dan melalui tahap pra-olah, analisis prediktif menggunakan metode *Linear Regression* dilakukan untuk mengetahui hubungan linier antara variabel independen seperti jumlah dosen, jumlah kegiatan promosi, dan faktor pendukung lain dengan jumlah mahasiswa baru yang diterima setiap tahun, sehingga hasilnya tidak hanya menggambarkan pola data masa lalu tetapi juga memberikan perkiraan yang dapat dipertanggungjawabkan secara ilmiah.

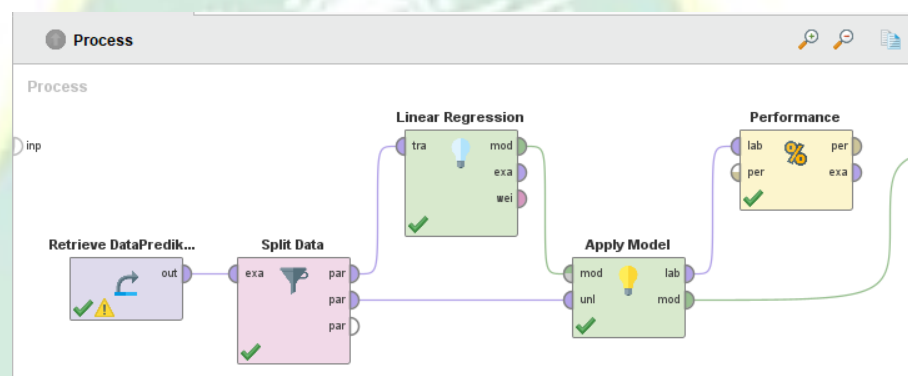
Untuk memastikan proses perhitungan berjalan konsisten dan transparan, peneliti memanfaatkan perangkat lunak Altair AI Studio Educational 2025. Aplikasi ini menyediakan lingkungan pemodelan *Machine Learning* berbasis alur kerja (workflow), sehingga setiap tahapan mulai dari pemanggilan data, pembagian dataset, pelatihan model, hingga evaluasi performa dapat diatur secara jelas dan mudah direproduksi. Dengan dukungan perangkat tersebut, hasil perhitungan *Linear Regression* dapat ditampilkan secara rinci sekaligus memudahkan proses analisis pada bab ini.

4.3.1. Linear regression menggunakan data split 30:70

Gambar berikut memperlihatkan tampilan halaman kerja (*workspace*) yang telah disusun oleh peneliti berdasarkan data penelitian yang telah dikumpulkan dan selanjutnya melalui tahap pra-olah data. Pada tahap ini, data telah dibersihkan dari nilai yang tidak

lengkap, duplikasi, maupun inkonsistensi, sehingga siap untuk digunakan dalam proses analisis dan pemodelan. *Workspace* tersebut menjadi pusat pengolahan data, di mana seluruh variabel penelitian ditata secara sistematis untuk mendukung penerapan metode analisis dan memastikan proses penelitian berjalan secara terstruktur dan akurat.

Gambar 4. 1 tampilan halaman kerja



Alur proses dimulai dari operator Retrieve Data, yaitu tahap pengambilan (retrieve) dataset bernama 'DataPrediksiPenerimaanMahasiswa' yang tersimpan di local repository. Dataset tersebut berisi data historis jumlah mahasiswa baru per tahun beserta variabel-variabel prediktor seperti jumlah dosen, jumlah kegiatan promosi, dan variabel lain yang telah ditentukan. Setelah data diambil, langkah berikutnya adalah Split Data. Pada tahap ini, data dibagi menjadi dua bagian, yakni data latih (training set) dan data uji (testing set). Rasio pembagian yang digunakan adalah 70 % untuk pelatihan dan 30 % untuk pengujian, sehingga model dapat

belajar dari mayoritas data namun tetap memiliki porsi yang memadai untuk evaluasi performa.

Operator *Linear Regression* kemudian digunakan untuk melatih model regresi linier berdasarkan data latih. Algoritma ini mencari hubungan linier antara variabel-variabel independen (misalnya jumlah dosen, kegiatan promosi, dan variabel lain yang relevan) dengan variabel dependen berupa jumlah mahasiswa baru. Hasil pelatihan model selanjutnya disalurkan ke operator *Apply Model*, yang berfungsi menerapkan model yang telah dibangun pada data uji agar diperoleh nilai prediksi. Tahap terakhir adalah operator *Performance (Regression)* yang bertugas mengevaluasi tingkat akurasi model. Di sini dihitung metrik-metrik regresi seperti Root Mean Square Error (RMSE).

Gambar 4. 2 Hasil

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
tahun	3.591	4.317	0.198	0.678	0.832	0.409	
X1	1.199	1.586	0.151	0.935	0.756	0.452	
X2	-40.816	10.740	-0.634	0.994	-3.800	0.000	****
X3	-4.086	5.002	-0.096	0.973	-0.817	0.417	
X5	-0.524	0.831	-0.113	0.796	-0.630	0.531	
X6	-77.326	20.766	-0.878	0.938	-3.724	0.000	****
(Intercept)	-7101.457	8697.801	?	?	-0.816	0.417	

Berdasarkan hasil analisis regresi linear yang diperoleh melalui pengolahan data, terbentuk sebuah model dengan beberapa variabel independen dan satu konstanta. Persamaan regresi yang dihasilkan menunjukkan bahwa nilai intersep (konstanta) sebesar $-7101,456$ dengan p-value 0,417. Nilai intersep ini tidak signifikan secara statistik

pada taraf kepercayaan 95%, sehingga dapat diartikan bahwa titik awal prediksi yang diberikan model ini kurang dapat dijadikan dasar yang kuat.

Meskipun demikian, konstanta tetap berfungsi sebagai nilai dasar perhitungan prediksi ketika seluruh variabel independen berada pada nilai nol, meskipun kontribusinya dalam model ini tidak terbukti signifikan.

Adapun variabel independen yang diuji dalam model ini antara lain: variabel Tahun, yang memiliki koefisien 3,590 dengan p-value 0,678, menunjukkan bahwa setiap peningkatan satu tahun hanya sedikit menaikkan jumlah mahasiswa baru, namun pengaruhnya tidak signifikan. Variabel X1 (jumlah dosen) memiliki koefisien 1,199 dengan p-value 0,756, mengindikasikan hubungan positif tetapi tidak signifikan secara statistik, sehingga tambahan jumlah dosen belum terbukti memberikan dampak langsung pada peningkatan jumlah mahasiswa baru. Variabel X2 (status beasiswa) justru menunjukkan hasil yang signifikan dengan koefisien $-40,816$ dan p-value 0,0003. Nilai ini menandakan adanya hubungan negatif yang kuat, artinya status mahasiswa penerima beasiswa berpotensi menurunkan jumlah penerimaan mahasiswa baru secara signifikan. Temuan ini dapat dijelaskan karena pemberian beasiswa mungkin lebih terkait dengan jumlah terbatas kuota atau kebijakan internal, sehingga tidak selalu berbanding lurus dengan peningkatan jumlah pendaftar.

Selanjutnya, variabel X3 (sarana dan prasarana) memiliki koefisien $-4,086$ dengan p-value $0,417$. Hal ini menunjukkan pengaruh negatif yang tidak signifikan, sehingga ketersediaan sarana dan prasarana belum memberikan kontribusi nyata terhadap jumlah mahasiswa baru dalam model ini. Variabel X5 (jumlah promosi) memiliki koefisien $-0,523$ dengan p-value $0,530$, menandakan bahwa meskipun kegiatan promosi dilakukan, pengaruhnya terhadap peningkatan jumlah mahasiswa baru tidak signifikan dan bahkan menunjukkan arah negatif. Hal ini mungkin disebabkan oleh kurang tepatnya strategi promosi atau karena faktor lain yang lebih dominan dalam memengaruhi keputusan calon mahasiswa.

Variabel terakhir yaitu X6 (atribut biaya) menunjukkan koefisien $-77,325$ dengan p-value $0,0004$. Hasil ini sangat signifikan secara statistik pada taraf kepercayaan 95% , dan menegaskan bahwa biaya perkuliahan memiliki pengaruh yang besar terhadap jumlah penerimaan mahasiswa baru. Dengan kata lain, semakin tinggi biaya yang ditetapkan, maka semakin berkurang jumlah calon mahasiswa yang berminat untuk mendaftar. Hal ini sejalan dengan teori ekonomi pendidikan yang menyebutkan bahwa faktor biaya merupakan salah satu pertimbangan utama dalam pengambilan keputusan pendidikan tinggi.

Secara keseluruhan, hasil ini mengindikasikan bahwa dari seluruh variabel yang dianalisis, hanya X2 (status beasiswa) dan X6

(atribut biaya) yang terbukti signifikan dalam memengaruhi jumlah penerimaan mahasiswa baru. Kedua variabel ini menjadi faktor dominan yang perlu diperhatikan lebih lanjut dalam kebijakan kampus. Sementara itu, variabel lain seperti jumlah dosen, sarana dan prasarana, serta jumlah promosi belum menunjukkan pengaruh berarti secara statistik, meskipun tetap dapat dipertimbangkan dalam penelitian lanjutan dengan cakupan data yang lebih besar atau pendekatan model yang lebih kompleks. Temuan ini memberikan gambaran awal bahwa faktor biaya dan kebijakan beasiswa merupakan aspek krusial yang paling menentukan dalam menarik minat mahasiswa baru, sedangkan variabel lain cenderung berperan sebagai faktor pendukung yang belum terukur secara signifikan pada penelitian ini.

Gambar 4. 3 nilai RMSE

root_mean_squared_error

root_mean_squared_error: 31.861 +/- 0.000

Gambar 4. 4 nilai MAPE

relative_error

relative_error: 39.96% +/- 46.11%

Berdasarkan gambar 4.3, diperoleh nilai root mean squared error (RMSE) sebesar $31,861 \pm 0,000$. Satuan nilai RMSE mengikuti satuan variabel dependen, yaitu jumlah mahasiswa. Artinya, secara

rata-rata prediksi jumlah penerimaan mahasiswa baru yang dihasilkan oleh model memiliki deviasi sekitar 31 mahasiswa dibandingkan dengan data aktual pada setiap pengamatan. Nilai ini menggambarkan besarnya kesalahan prediksi model: semakin kecil angka RMSE, semakin baik kemampuan model dalam mendekati nilai sebenarnya. Dengan RMSE sebesar 31,861, tingkat kesalahan prediksi tergolong cukup tinggi, menandakan bahwa model regresi linear masih belum mampu menggambarkan hubungan variabel independen dan dependen secara optimal.

Selain itu, pada gambar 4.4, hasil evaluasi juga menunjukkan nilai mean absolute percentage error (MAPE) atau *relative error* sebesar $39,96\% \pm 46,11\%$. Nilai MAPE mengukur rata-rata kesalahan prediksi dalam bentuk persentase terhadap nilai aktual, sehingga memberikan gambaran proporsional seberapa jauh prediksi menyimpang dari data sebenarnya. Secara umum, MAPE di bawah 10% dianggap sangat baik, 10–20% baik, 20–50% moderat, dan di atas 50% tergolong buruk. Dengan MAPE mendekati 40%, model ini berada pada kategori kesalahan moderat ke tinggi, yang berarti perbedaan antara nilai prediksi dan nilai aktual relatif besar dan dapat memengaruhi keakuratan hasil.

4.3.2. Linaer regression menggunakan data split 80:20

Pengujian algoritma *Linear Regression* dengan menggunakan data split 80:20 yang mana 80 adalah data traning dan 20 adalah data testing. Hal ini dilakukan untuk menilai apakah tingkat kesalahan dapat ditekan lebih rendah.

Gambar 4. 5 Hasil prediksi dengan ratio 80:20

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
tahun	2.769	3.995	0.149	0.672	0.693	0.490	
X1	1.138	1.518	0.136	0.937	0.750	0.456	
X2	-41.592	10.437	-0.620	0.990	-3.985	0.000	****
X3	-4.218	4.756	-0.094	0.985	-0.887	0.378	
X6	-83.722	19.052	-0.892	0.913	-4.394	0.000	****
(Intercept)	-5446.944	8050.531	?	?	-0.677	0.501	

Berdasarkan hasil analisis regresi linear terbaru, diperoleh model dengan empat variabel independen, yaitu jumlah dosen (X1), status mahasiswa apakah menerima beasiswa atau tidak (X2), kondisi sarana dan prasarana (X3), serta atribut biaya (X6). Variabel dependen yang diprediksi adalah jumlah penerimaan mahasiswa baru. Model ini menghasilkan nilai intersep sebesar $-5.446,94$ dengan p-value 0,501. Artinya, ketika seluruh variabel independen bernilai nol, prediksi jumlah mahasiswa baru berada di sekitar -5.447 mahasiswa. Nilai intersep yang negatif dan tidak signifikan ini menunjukkan bahwa konstanta model hanya berfungsi sebagai titik awal matematis dan tidak memiliki makna praktis yang kuat dalam konteks penerimaan mahasiswa.

Untuk masing-masing variabel independen, X1 (jumlah dosen) memiliki koefisien 1,138 dengan p-value 0,456. Nilai positif ini mengindikasikan bahwa setiap penambahan satu dosen diperkirakan dapat menambah sekitar satu mahasiswa baru, meskipun pengaruhnya tidak signifikan secara statistik. Variabel X2 (status mahasiswa apakah menerima beasiswa atau tidak) menunjukkan koefisien -41,59 dengan p-value 0,00016. Hasil ini menandakan bahwa semakin tinggi proporsi mahasiswa penerima beasiswa, jumlah mahasiswa baru cenderung menurun sekitar 41 orang. Temuan ini menunjukkan bahwa kebijakan atau distribusi beasiswa berperan penting dalam pola pendaftaran mahasiswa, dan variabel ini terbukti signifikan memengaruhi jumlah penerimaan mahasiswa baru.

Selanjutnya, X3 (kondisi sarana dan prasarana) memiliki koefisien -4,22 dengan p-value 0,378, yang menunjukkan hubungan negatif yang lemah dan tidak signifikan dengan jumlah mahasiswa baru. Sementara itu, X6 (atribut biaya) memiliki koefisien -83,72 dengan p-value 0,000036, menunjukkan pengaruh yang sangat signifikan. Nilai ini mengindikasikan bahwa semakin tinggi atribut biaya, seperti besaran biaya kuliah atau biaya tambahan lainnya, jumlah mahasiswa baru diprediksi menurun sekitar 84 orang. Secara keseluruhan, temuan ini menegaskan bahwa status beasiswa (X2) dan atribut biaya (X6) merupakan faktor dominan yang secara signifikan memengaruhi fluktuasi jumlah penerimaan mahasiswa baru, sedangkan

jumlah dosen (X1) dan kondisi sarana prasarana (X3) tidak memberikan pengaruh yang berarti dalam model regresi linear ini.

Gambar 4. 6 nilai RMSE

root_mean_squared_error

root_mean_squared_error: 16.746 +/- 0.000

Berdasarkan hasil evaluasi kinerja model regresi linear, diperoleh nilai root mean squared error (RMSE) sebesar $32,145 \pm 0,000$. Satuan nilai RMSE mengikuti satuan variabel dependen, yaitu jumlah mahasiswa. Angka ini menunjukkan bahwa rata-rata prediksi jumlah penerimaan mahasiswa baru yang dihasilkan model menyimpang sekitar 32 mahasiswa dari data aktual pada setiap pengamatan. Semakin kecil nilai RMSE, semakin baik kemampuan model dalam mendekati nilai sebenarnya. Sehingga RMSE sebesar 32,145 mengindikasikan tingkat kesalahan prediksi yang tergolong moderat hingga cukup besar. Hal ini menandakan bahwa meskipun model telah menangkap pola hubungan antara variabel independen jumlah dosen, status beasiswa, kondisi sarana prasarana, serta atribut biaya masih terdapat variasi lain yang belum sepenuhnya dijelaskan.

Selain itu, nilai relative error atau Mean Absolute Percentage Error (MAPE) yang diperoleh adalah $39,25\% \pm 44,29\%$. Persentase ini menunjukkan bahwa secara rata-rata, selisih antara prediksi model dan data aktual adalah sekitar 39% dari nilai sebenarnya, dengan tingkat

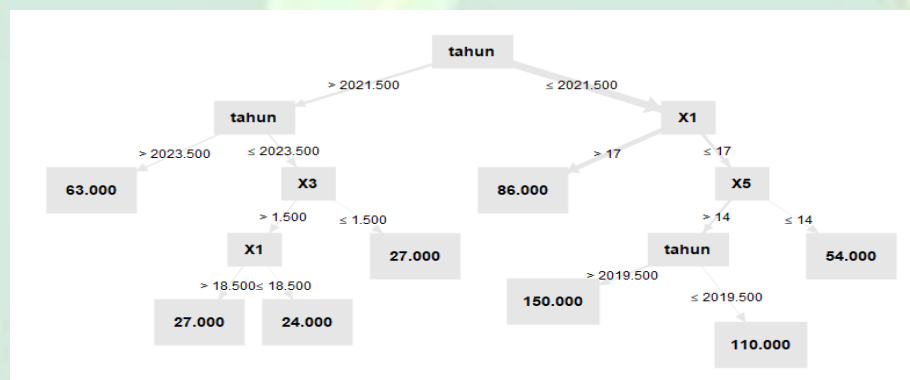
penyebaran kesalahan yang cukup tinggi karena simpangan bakunya mencapai 44,29%. MAPE yang mendekati 40% termasuk kategori kesalahan yang besar, sehingga interpretasi prediksi harus dilakukan dengan hati-hati.

4.4 Radom Forest

Setelah dilakukan olah data pada algoritma linear regression, maka selanjutnya menguji dataset menggunakan algortima *Random Forest* untuk mengetahui algoritma mana yang lebih akurat dalam menguji data set tersebut. Disini peneliti juga sama-sama akan menguji data dengan data split menggunakan ratio 70:30 dan 80:20.

4.4.1. *Random Forest* Regression menggunakan data split 70:30

Gambar 4. 7 Random Forest Regression data split 70:30



Proses dimulai dari akar pohon keputusan, di mana data dibagi menjadi dua kelompok besar berdasarkan tahun. Ini menunjukkan bahwa variabel tahun adalah prediktor paling signifikan dalam model ini.

Jalur 1: Data dengan tahun > 2021.500 . Data yang termasuk dalam kelompok ini dianggap memiliki karakteristik yang berbeda, sehingga diarahkan ke jalur keputusan terpisah. Di sini, model kembali membagi data berdasarkan tahun, menunjukkan adanya pola yang lebih spesifik pada rentang tahun ini.

Jika variabel tahun lebih besar dari 2023.500, model memprediksi nilai 63.000. Angka ini merupakan nilai rata-rata dari 9 sampel data yang jatuh ke dalam kategori ini, menjadikannya prediksi akhir untuk kelompok tersebut.

Jika tahun kurang dari atau sama dengan 2023.500, model akan menggunakan variabel X3 sebagai prediktor berikutnya. Apabila nilai X3 lebih besar dari 1.500, model akan memeriksa X1. Di sini, terjadi dua kemungkinan: jika X1 lebih dari 18.500, nilai prediksi adalah 27.000 (berdasarkan 6 sampel), tetapi jika X1 kurang dari atau sama dengan 18.500, nilai prediksi menjadi 24.000 (berdasarkan 2 sampel). Namun, jika nilai X3 sendiri kurang dari atau sama dengan 1.500, model tidak perlu memeriksa variabel lain dan langsung memberikan prediksi 27.000 (berdasarkan 2 sampel).

Gambar 4. 8 nilai RMSE

root_mean_squared_error

root_mean_squared_error: 4.876 +/- 0.000

Gambar 4. 9 nilai MAPE

relative_error

`relative_error: 6.74% +/- 11.82%`

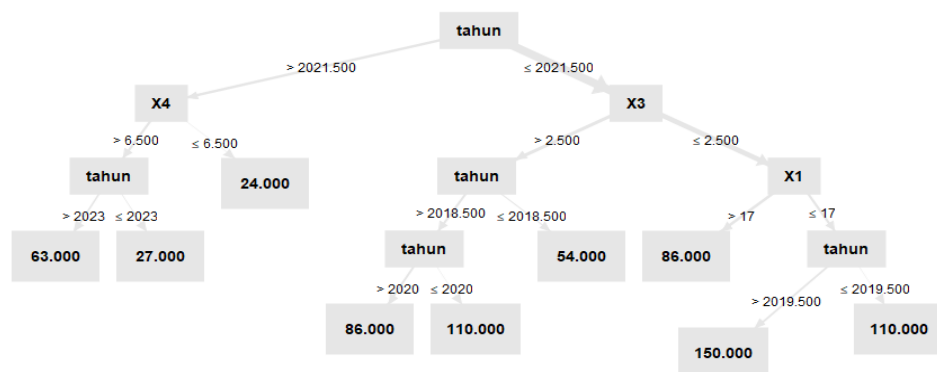
Pada gambar 4.8, menunjukan nilai Root Mean Squared Error (RMSE) sebesar 4.876. Angka ini menunjukkan bahwa, secara rata-rata, kesalahan prediksi model sekitar 4.876 unit dari nilai aktual. Nilai RMSE ini sangat sensitif terhadap kesalahan prediksi yang besar. Artinya, model yang menghasilkan beberapa kesalahan ekstrem akan memiliki nilai RMSE yang jauh lebih tinggi. Hal ini penting untuk dicatat, karena nilai 4.876 menunjukkan bahwa meskipun model yang telah dibuat memiliki akurasi yang baik secara umum, ada beberapa kasus di mana prediksi ini cukup jauh dari nilai sebenarnya. Namun, secara keseluruhan, nilai yang relatif rendah ini menegaskan bahwa model yang telah peneliti buat memiliki kemampuan prediksi yang andal.

Berdasarkan gambar 4.9, metrik evaluasi yang ditampilkan adalah Relative Error dengan nilai 6.74%. Nilai 6.74% menunjukkan bahwa model prediksi yang telah dikembangkan memiliki tingkat akurasi yang memadai. Angka ini menandakan bahwa secara keseluruhan, rata-rata kesalahan prediksi model yang telah peneliti bangun adalah relatif kecil. Oleh karena itu, nilai ini menjadi bukti

kuat yang mendukung kesimpulan bahwa model yang peneliti bangun efisien dan dapat diandalkan untuk melakukan prediksi.

4.4.2. *Random Forest Regression* menggunakan data split 80:20

Gambar 4. 10 Random Forest Regression data split 80:20



Jika variabel tahun memiliki nilai di atas 2021.500, model akan melanjutkan analisis dengan melihat variabel X4. Apabila X4 lebih dari 6.500, model akan memeriksa tahun lagi. Di sini, terjadi dua kemungkinan: jika tahun lebih dari 2023, prediksi akhirnya adalah 63.000; namun, jika tahun kurang dari atau sama dengan 2023, prediksinya adalah 27.000. Jika variabel X4 memiliki nilai kurang dari atau sama dengan 6.500, model akan langsung memberikan prediksi sebesar 24.000 tanpa perlu menganalisis variabel lainnya.

Jika nilai X3 kurang dari atau sama dengan 2.500, model akan menggunakan variabel X1 untuk membuat prediksi. Apabila nilai X1 lebih besar dari 17, prediksi akhirnya adalah 86.000. Namun, jika X1 kurang dari atau sama dengan 17, model akan memeriksa kembali variabel tahun. Di sini, jika tahun lebih besar dari 2019.500, prediksi

akhirnya adalah 150.000, sedangkan jika tahun kurang dari atau sama dengan 2019.500, prediksi yang dihasilkan adalah 110.000.

Gambar 4. 11 hasil RMSE ratio 80:20

root_mean_squared_error

`root_mean_squared_error: 4.149 +/- 0.000`

Gambar 4. 12 Nilai persentase MAPE

relative_error

`relative_error: 4.90% +/- 7.25%`

Dari hasil evaluasi yang peneliti lakukan pada gambar 4.11, model regresi linier yang telah kami kembangkan menunjukkan nilai Root Mean Squared Error (RMSE) sebesar 4.149. Nilai 4.149 yang peneliti peroleh mengindikasikan bahwa, secara rata-rata, model yang peneliti buat memiliki kesalahan prediksi sekitar 4.149 unit. Metrik ini sangat berguna karena memberikan bobot lebih besar pada kesalahan yang signifikan, sehingga sangat sensitif terhadap adanya *outlier* atau prediksi yang jauh dari nilai sebenarnya. Oleh karena itu, nilai RMSE ini menunjukkan bahwa, meskipun model yang peneliti buat ini sudah cukup andal, masih ada celah kecil antara prediksi yang dihasilkan dengan kondisi data yang sebenarnya.

Dari hasil yang diperoleh, nilai *Relative Error* adalah 4.90%. Nilai ini menunjukkan bahwa, secara rata-rata, kesalahan prediksi model kami adalah sekitar 4.90% dari nilai aktual. Kesalahan sebesar

4.90% menunjukkan bahwa model kami memiliki tingkat akurasi yang memadai dan dapat diandalkan untuk membuat prediksi.

4.5 Hasil

Penelitian ini dirancang untuk membandingkan kinerja dua algoritma *Machine Learning*, yaitu *Linear Regression* dan *Random Forest*, dalam memprediksi jumlah penerimaan mahasiswa baru. Peneliti menggunakan dua skema pembagian data, yakni rasio 70:30 dan 80:20, untuk melatih dan menguji kedua model. Evaluasi kinerja model didasarkan pada metrik Root Mean Squared Error (RMSE) dan Mean Absolute Percentage Error (MAPE). Hasil ini digunakan untuk menguji hipotesis utama (H_0) dan hipotesis alternatif (H_1).

Untuk membandingkan kinerja kedua model, kami menggunakan metrik evaluasi Root Mean Squared Error (RMSE) dan Mean Absolute Percentage Error (MAPE) dengan dua rasio pembagian data yang berbeda, yaitu 70:30 dan 80:20.

Tabel 4. 2 Perbandingan Kinerja

Rasio Data Split	Metode	RMSE	MAPE (%)
70:30	Regresi Linier	31,861	39,96
70:30	<i>Random Forest</i>	4,876	6,74
80:20	Regresi Linier	32,145	39,25
80:20	<i>Random Forest</i>	4,149	4,90

Dari tabel di atas, terlihat perbedaan yang sangat mencolok antara kedua metode. Model Regresi Linier menunjukkan nilai RMSE yang tinggi, yaitu sekitar 31 hingga 32 mahasiswa, dan nilai MAPE yang berada di kisaran 39-40%. Nilai ini mengindikasikan bahwa rata-rata kesalahan prediksi model Regresi Linier tergolong besar (kategori moderat-tinggi), sehingga prediksinya kurang andal.

Sebaliknya, model *Random Forest Regression* menghasilkan nilai RMSE yang jauh lebih rendah, yaitu di kisaran 4 hingga 5 mahasiswa, dan nilai MAPE yang sangat rendah, di bawah 7%. Nilai MAPE di bawah 10% secara umum dianggap menunjukkan akurasi yang sangat baik. Hal ini membuktikan bahwa kesalahan prediksi model *Random Forest* jauh lebih kecil dibandingkan Regresi Linier, sehingga kemampuannya dalam memprediksi jumlah mahasiswa baru sangat andal.

Dengan demikian, hasil pengujian ini menolak Hipotesis Utama (H_0) yang menyatakan tidak ada perbedaan signifikan antara kedua metode. Sebaliknya, hasil ini menerima Hipotesis Alternatif (H_1), yang menyatakan bahwa terdapat perbedaan signifikan, dan secara jelas menunjukkan bahwa *Random Forest Regression* memberikan akurasi yang jauh lebih baik daripada Regresi Linier.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Implementasi algoritma Linear Regression dan Random Forest dilakukan secara sistematis mulai dari pengumpulan dan pra-pengolahan data hingga pemodelan dan evaluasi menggunakan data historis penerimaan mahasiswa baru yang diperoleh melalui dokumentasi dan kuesioner dengan variabel jumlah dosen (X1), status beasiswa (X2), sarana prasarana (X3), jarak tempat tinggal (X4), jumlah promosi (X5), dan biaya pendidikan (X6), yang kemudian dibersihkan serta dibagi menjadi data latih dan uji dengan rasio 70:30 dan 80:20. Linear Regression digunakan untuk memodelkan hubungan linier antara variabel independen dan jumlah mahasiswa baru (Y) menggunakan Altair AI Studio, dengan hasil menunjukkan bahwa hanya status beasiswa (X2) dan biaya pendidikan (X6) yang signifikan, sedangkan Random Forest membangun kumpulan pohon keputusan menggunakan teknik bagging dan pemilihan fitur acak sehingga mampu menghasilkan prediksi yang lebih stabil.

Hasil evaluasi menunjukkan bahwa Random Forest memiliki nilai RMSE dan MAPE yang lebih rendah dibandingkan Linear Regression, dengan kinerja terbaik pada rasio 80:20 dan nilai RMSE sebesar 4,149, yang menegaskan keunggulannya dalam menangkap pola non-linier dan menjadikannya algoritma paling efektif untuk memprediksi jumlah penerimaan mahasiswa baru pada kasus ini.

5.2. Saran

Berdasarkan temuan penelitian ini, beberapa saran yang dapat diajukan untuk penelitian selanjutnya adalah:

- a. Melakukan pengujian dengan algoritma *Machine Learning* lain seperti *Support Vector Regression* (SVR) atau *Neural Networks* untuk membandingkan lebih banyak metode prediksi.
- b. Mencari variabel-variabel independen tambahan yang berpotensi memiliki pengaruh signifikan, seperti tren demografi, popularitas program studi, atau faktor ekonomi, untuk meningkatkan akurasi model.
- c. Menggunakan data dalam jumlah yang lebih besar dan dalam rentang waktu yang lebih panjang untuk memperkuat validitas model dan mengurangi variasi kesalahan prediksi.
- d. Mengembangkan model prediksi yang sudah ada menjadi sebuah prototipe aplikasi web atau *website*. Dengan mengimplementasikan model ke dalam *website*, pihak manajemen atau pembuat kebijakan dapat memasukkan data prediktor (misalnya, jumlah dosen, biaya, atau promosi) secara langsung dan mendapatkan hasil prediksi jumlah mahasiswa baru secara *real-time*. Hal ini akan mengubah model prediksi yang bersifat teoritis menjadi alat bantu pengambilan keputusan yang praktis, interaktif, dan mudah diakses oleh pihak yang membutuhkan.

DAFTAR PUSTAKA

- Abdhul, Y. (2022). Penelitian Komparatif: Pengertian, Jenis dan Contoh. *Deepublish Store (blog)*.
- Adinugroho, R. (2022). *Perbandingan Rasio Split Data Training Dan Data Testing Menggunakan Metode Lstm*.
- Al'afi, A. M., Widiart, W., Kurniasari, D., & Usman, M. (2020). Peramalan Data Time Series Seasonal Menggunakan Metode Analisis Spektral. *Jurnal Siger Matematika*, 1(1), 10–15. <https://doi.org/10.23960/jsm.v1i1.2484>
- Amadi, A. S. M., Hasan, S., Rifanto, N. A., Wildan, M., Afifah, N. Q., & Nisak, N. M. (2023). Upaya Pemerintah dalam Menjamin Hak Pendidikan untuk Seluruh Masyarakat di Indonesia: Sebuah Fakta yang Signifikan. *Educatio*, 18(1), 161–171. <https://doi.org/10.29408/edc.v18i1.14798>
- Andelawati, D. (2021). *Sistem Prediksi Penjualan Makanan dan Minuman pada Kafe dengan Menggunakan Metode Triple Exponential Smoothing*. Universitas Multimedia Nusantara.
- Aprilia, K., Khaira, U., & Hutabarat, B. F. (2025). Komparasi Algoritma Naive Bayes Dan Decision Tree Pada Klasifikasi Penerimaan Peserta Didik Baru. *JURNAL SISTEM INFORMASI DAN TEKNIK KOMPUTER*, 10(2), 547–553.
- Apriyanti, Y., Darmansyah, R., Hayati, Kurnia, L., Zebua, R., Ramli, A., Mawadah, Mamlu'ah, A., & Barokah, A. (2023). *Pengantar Ilmu Manajemen Pendidikan: Teori dan praktek mengelola Lembaga Pendidikan Era Industri 4.0 & Soceity 5.0*.
- Arikunto, S. (2010). *Prosedur Penelitian : Suatu Pendekatan Praktik*. Rineka Cipta.
- Asqia, N., & Wahyuni, S. (2024). ANALISIS KEAKTIFAN BERORGANISASI TERHADAP IPK MAHASISWA PAUD INSTITUT AGAMA ISLAM NEGERI PAREPARE. *Jurnal Tawadhu*, 8(2), 140–152.
- Billah, I. S. M. (2024). *Prediksi urban heat island menggunakan metode random forest regression: Studi kasus Kota Malang*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Budiyanto, F., Waluyo, S. E. Y., Fitriyani, Z. A., Ridha, M., & Ashriana, A. N. (2025). OPTIMASI STRATEGI OPERASIONAL DAN PEMASARAN PERGURUAN TINGGI MELALUI PEMODELAN SISTEM DINAMIS DAN ANALISIS STATISTIK PREFERENSI MAHASISWA BARU DI WILAYAH MOJOKERTO. *CENDEKIA : Jurnal Ilmu Pengetahuan*, 5(3), 1006–1018.

- Cahya Bagaskara, G. W., Rochmawati, M., Adha, I. A., & Trifena, M. (2023). Prediksi Penerimaan Mahasiswa Baru Universitas Singaperbangsa Karawang dengan Naive Bayes. *Journal Sensi*, 9(2), 198–208. <https://doi.org/10.33050/sensi.v9i2.2913>
- Creswell, J. W., & Creswell, J. D. (2018). Research Design Kualitatif, Kualitatif And Mixed Method Approach. In *Writing Center Talk over Time* (5 ed.). SAGE Publications Ltd. <https://doi.org/10.4324/9780429469237-3>
- Emzir. (2013). *Metodologi Penelitian Pendidikan : Kuantitatif dan Kualitatif*. Rajawali Pers.
- Endrawati, T., Sabirin, S., Deluma, R. Y., Imelda, D. Q., & Suluh, S. I. (2023). Human resources for the development of higher education and academic culture of achievement and continuous learning. *International research journal of management, IT and social sciences*, 10(1), 25–34. <https://doi.org/10.21744/irjmis.v10n1.2265>
- Erkamim, M., Zidni, M., & Widarti, E. (2023). *Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM*. 02, 127–134.
- Fauziyah, R. N. (2023). *Analisis Data Menggunakan Uji Korelasi dan Uji Regresi Linier di Bidang Kesehatan Masyarakat dan Klinis*.
- Fauzy, I., & Pramaditya, H. (2024). Peramalan Penerimaan Mahasiswa Baru Dengan Perbandingan Metode Regresi Linier Sederhana dan Single Eksponential Smoothing. *Jurnal Riset Mahasiswa Bidang Teknologi Informasi*, 6(2), 182–193.
- Febyola, A., & Widyanesti, S. (2024). The influence of social media usage and electronic word-of-mouth (eWOM) on purchase intention for local fashion products on Tiktok social media. *Keynesia : International Journal of Economy and Business*, 3, 94–103. <https://doi.org/10.55904/keynesia.v3i2.1237>
- Handoko, W. (2021). Prediksi Jumlah Penerimaan Mahasiswa Baru Dengan Metode Single Exponential Smoothing (Studi Kasus: Amik Royal Kisaran). *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2), 125–132. <https://doi.org/10.33330/jurteks.v5i2.356>
- Handoko, Y., Satria, E., & Nasrulloh, A. (2025). Algoritma Regresi Linier Dalam Prediksi Jumlah Pendaftar Program Pendidikan Di Lingkungan Pesantren (Studi Kasus : Yayasan Al – Mustofa Tambakbaya). *Jurnal Algoritma*, 22(1), 557–569. <https://doi.org/10.33364/algoritma/v.22-1.1802>
- Hikmah, Nuril Asmawan, Yudhi Ferdi Andri Putra, S. A. A. K. (2023). Analisis Pengaruh Strategi Promosi Terhadap Keputusan Mahasiswa Dalam Memilih

Perguruan Tinggi. *REMB.*

- Ititi, F., Sulastri, A., Muspita, Z., & Sururuddin, M. (2022). Pengaruh Model Pembelajaran Eth (Everyone Is A Teacher Here) dengan bantuan Ice Breaking Untuk Meningkatkan Minat Dan Hasil Belajar Matematika Siswa Kelas 4 MI Husnul Abror. *Jurnal Pendidikan dan Konseling*, 4(4), 15.
- Jannah, M. (2024). *PSIKOLOGI PERKEMBANGAN I (Usia Prakelahiran-Remaja)*. repository.ar-raniry.ac.id. [https://repository.ar-raniry.ac.id/id/eprint/38802/1/DIKTAT KULIAH PSI PERKEMBANGAN I.pdf](https://repository.ar-raniry.ac.id/id/eprint/38802/1/DIKTAT%20KULIAH%20PSI%20PERKEMBANGAN%20I.pdf)
- Johnson, E. A., Inyangetoh, J. A., Rahmon, H. A., Jimoh, T. G., Dan, E. E., & Esang, M. O. (2024). An Intelligent Analytic Framework for Predicting Students Academic Performance Using Multiple Linear Regression and Random Forest. *European Journal of Computer Science and Information Technology*, 12(3), 56–70. <https://doi.org/10.37745/ejcsit.2013/vol12n35670>
- Khana Wijaya, Rishi Suprianto, & Endi Istiawan. (2022). Implementasi Framework Bootstrap Dalam Perancangan Sistem Penerimaan Mahasiswa Baru Pada Sekolah Tinggi Ilmu Tarbiyah Al-Qur'an Al-Ittifaqiah Berbasis Web. *Jurnal Masda*, 1(1), 1–13. <https://doi.org/10.58328/jm.v1i1.42>
- Luthfi Mawardi, Umar Alfaruq, & Gandung Triyono. (2024). Prediksi Jumlah Penerimaan Mahasiswa Baru Dengan Metode Linear Regression dan Exponential Smoothing. *Indonesian Journal of Computer Science*, 13(1), 1154–1166. <https://doi.org/10.33022/ijcs.v13i1.3649>
- Madyatmadja, E. D., Kusumawati, L., Jamil, S. P., Kusumawardhana, W., Informasi, S., & Nusantara, U. B. (2021). Infotech: journal of technology information. *Raden Ario Damar*, 7(1), 55–62.
- Makhrisa, R., & Pradikto, S. (2025). Analisis Peran Lingkungan Sosial Terhadap Minat Peserta Didik dalam Memilih Pendidikan Tinggi. *Jurnal Kajian dan Penelitian Umum*, 3(1).
- Marlinang, S., & Sarajar, D. K. (2025). HUBUNGAN POLA ASUH DEMOKRATIS DENGAN KONTROL DIRI MAHASISWA. *Journal of Innovation Research and Knowledge*, 4(12), 8935–8944.
- Martanto, M., Ali, I., & Mulyawan, M. (2019). Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Machine Learning dengan Teknik Deep Learning. *Jurnal Informatika: Jurnal Pengembangan IT*, 4, 191–194. <https://doi.org/10.30591/jpit.v4i2-2.1877>
- Maulana, T., Astuti, R., & Muhammad Basysyar, F. (2024). Implementasi Algoritma Regresi Linear Untuk Memprediksi Pendapatan Pt Pln

Berdasarkan Penggunaan Per Kelompok Pelanggan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3196–3202. <https://doi.org/10.36040/jati.v7i6.8083>

Mulayawan, R. (2022). *Linear Regression*. Rifqi Mulayawa.com.

Mustofa, & Khotib. (2025). Analisis Faktor-Faktor yang Mempengaruhi Minat Mahasiswa Memilih Program Studi Manajemen Zakat dan Wakaf Fakultas Ekonomi dan Bisnis Islam UIN Sunan Ampel Surabaya. *Al Buhuts*, 21(1), 76–97.

Noviyanti, R., & Dermawan, D. A. (2022). Studi Literatur Pengaruh Beasiswa Terhadap Prestasi Belajar Mahasiswa. *IT-Edu: Jurnal Information Technology and Education*, 7(2), 58–66. <https://doi.org/10.26740/it-edu.v7i2.49145>

Nur Hamidah, & Nurhuda. (2022). Pengaruh Promosi Terhadap Pengambilan Keputusan Peserta Didik Sma Kelas Xii Ips Di Kecamatan Bungaraya Kabupaten Siak Dalam Memilih Universitas Islam Riau. *Peka*, 10(1), 1–11. [https://doi.org/10.25299/peka.2022.vol10\(1\).11066](https://doi.org/10.25299/peka.2022.vol10(1).11066)

Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>

Nurmantika, J. (2022). Prediksi Penerimaan Mahasiswa Baru Pascasarjana dengan Menggunakan Model Simulasi Monte Carlo. *Jurnal Sistim Informasi dan Teknologi*, 3, 287–291. <https://doi.org/10.37034/jsisfotek.v3i4.126>

Nurpratama, M., Yudianto, A., & Firdaus, T. (2024). PENGEMBANGAN YAYASAN PERGURUAN TINGGI SWASTA MEMBENTUK BADAN LAYANAN KEGIATAN USAHA. *Jurnal Investasi*, 10(1), 1–7.

Oktavian, B., Safitri, A., & Andini, M. (2025). Pemanfaatan Media Sosial sebagai Sarana Menjaring Calon Mahasiswa dalam Memilih Perguruan Tinggi. *Jurnal Ilmiah Ekonomi Dan Manajemen*, 3(5), 440–448.

Pratiwi, A. (2020). Peran Media Sosial Dalam Meningkatkan Penjualan Online Saat Pandemi Covid-19. *Jurnal Ilmiah Satyagraha*, 3, 73–81. <https://doi.org/10.47532/jis.v3i2.179>

Rahmadhani, A., Sihabudin Sahid, D. S., & Lulu Widyasari, Y. D. (2023). Implementasi SEM-Multiple Linear Regression dalam Prediksi Jumlah Pendaftaran Mahasiswa Baru di Perguruan Tinggi XYZ. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(2), 150–162. <https://doi.org/10.25077/teknosi.v9i2.2023.150-162>

- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sabrina, F. A., Nandita, W. V., & Maharani, D. D. (2023). Uji Asumsi Klasik untuk Menghindari Pelanggaran Asumsi Klasik pada Regresi Linier Ordinary Least Squares (OLS) dalam Ekonometrika. *Jurnal Ilmiah Manajemen Ekonomi Dan Akuntansi*, 1(1), 195–203.
- Sari, A. K., Yusuf, A. M., Iswari, M., & Afdal, A. (2021). Analisis Teori Karir Krumboltz: Literature Review. *Jurnal Ilmiah Bimbingan Konseling Undiksha*, 12(1), 116–121. <https://doi.org/10.23887/jjbk.v12i1.33429>
- Sebayang, W. B. (2022). Adolescent Childbirth with Asphyxia Neonatorum. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 7(2), 669–672. <https://doi.org/10.30604/jika.v7i2.1507>
- Sejati, P., Munawar, M., Pilliang, M., & Akbar, H. (2022). Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(7). <https://doi.org/10.25126/jtiik.2022976737>
- Setyawan, D. A. (2021). Hipotesis Dan Variabel Penelitian. In *Tahta Media Group* (1 ed.). CV. Tahta Media Group. [https://poltekkes-solo.ac.id/cni-content/uploads/modules/attachments/20210902143751-Buku Modul Hipotesis dan Variabel Penelitian.pdf](https://poltekkes-solo.ac.id/cni-content/uploads/modules/attachments/20210902143751-Buku%20Modul%20Hipotesis%20dan%20Variabel%20Penelitian.pdf)
- Simorangkir, D. B. P., Sihombing, S., & Simatupang, L. F. (2022). Pengaruh Sarana Prasarana Dan Lingkungan Kampus Terhadap Pengambilan Keputusan Mahasiswa Memilih Kuliah Di Universitas Hkbp Nomensen Pematangsiatar Tahun Ajaran 2018/2019. *Jurnal Pendidikan Indonesia: Teori, Penelitian, dan Inovasi*, 3(4). <https://doi.org/10.59818/jpi.v3i4.377>
- Siyoto, S., & Sodik, M. A. (2015). *DASAR METODOLOGI PENELITIAN* (Ayup (ed.); 1 ed.). Literasi Media Publishing.
- Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Susanti, E., Ariyana, R. Y., Wibowo, S. M., & Sya'bani, D. R. (2022). Sistem Pendukung Keputusan Untuk Penentuan Lokasi Wilayah Promosi Penjarangan Mahasiswa Baru Dengan Metode TOPSIS. *Techno.com*, 21(4), 765–777.
- Syahputri, A. Z., Fallenia, F. Della, & Syafitri, R. (2023). Kerangka berfikir penelitian kuantitatif. *Tarbiyah: Jurnal Ilmu Pendidikan dan Pengajaran*, 2(1), 160–166.

- Tamba, D., & Sitanggang, A. (2024). FAKTOR-FAKTOR YANG MEMPENGARUHI KEPUTUSAN MAHASISWA BARU TA 2023/2024 MEMILIH PROGRAM STUDI MANAJEMEN PADA FAKULTAS EKONOMI DAN BISNIS UNIVERSITAS KATOLIK SANTO THOMAS MEDAN. *Jurnal Manajemen dan Bisnis*, 24(1), 145–152.
- Vandiah, H. (2022). *PENGARUH BRAND IMAGE, BRAND TRUST DAN CELEBRITY ENDORSER TERHADAP MINAT BELI PRODUK SKINCARE SCARLETT (STUDI KASUS MAHASISWA STEI 2018 YANG BERBELANJA DI E-COMMERCE)*. Sekolah Tinggi Ilmu Ekonomi Indonesia Jakarta.
- Vita, D., Nainggolan, Y., & Utara, U. S. (2023). *Pengaruh Lingkungan Keluarga Terhadap Keinginan Siswa / I Kelas Xii Ipa Sma Negeri 1 Panombeian Panei*. 2(8), 2559–2566.
- Yuliara, I. M. (2022). Regresi Linier Sederhana dan Berganda. In *Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam - Universitas Udayana* (Nomor July).



**STKIPNU
TEGAL**

Lampiran 1 Surat Permohonan Penelitian

 **SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN
NAHDLATUL 'ULAMA KABUPATEN TEGAL
(STKIPNU)**
SK. BAN-PT Nomor 1290/SK/BAN-PT/AK/PT/VII/2024
Kampus : Jl. Jend. A. Yani No. 21 Prorot Slawi Kode Pos 52412 Telp. (0283) 491576 E-mail : stkipnutegalkab@gmail.com

Nomor : 015/I/SPm-STKIPNU/VII/2025
Lamp. : -
Perihal : Surat Permohonan Penelitian

Yth. Ketua STKIP NU Kabupaten Tegal
Di
Tempat

Assalamualaikum Wr. Wb.

Salam Silaturahmi teriring doa kami sampaikan semoga Bapak/ ibu senantiasa dalam lindungan Allah SWT, serta dimudahkan dalam melaksanakan aktivitas sehari-hari.

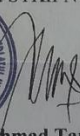
Sehubungan dengan adanya Tugas Akhir Skripsi Program Studi Pendidikan Informatika Sekolah Tinggi Keguruan dan Ilmu Pendidikan Nahdaltul Ulama Kabupaten Tegal, maka kami mengajukan izin penelitian serta kerjasama dari Ketua STKIP NU Kabupaten Tegal Yang Bapak/ Ibu pimpin untuk kiranya memberikan izin penelitian kepada mahasiswa kami untuk melaksanakan kegiatan tersebut dibawah ini :

Nama : Sofania Ma'rifatul Janah
NIM : 2102000080
Program Studi : Pendidikan Informatika

Demikian permohonan ini disampaikan atas perhatian dan kerjasamanya, disampaikan terimakasih.

Wassalamualaikum Wr. Wb.

Ditetapkan : di Slawi
Pada Tanggal : 14 Juli 2025
Ketua STKIPNU Kabupaten Tegal

 
Dr. Ahmad Taufiq, M.Pd.
NIDN. 2112126901

Lampiran 2 Surat Selesai Penelitian



**SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN
NAHDLATUL 'ULAMA KABUPATEN TEGAL
(STKIPNU)**

SK. BAN-PT Nomor 1290/SK/BAN-PT/Ak/PT/VII/2024

Kampus : Jl. Jend. A. Yani No. 21 Prokot Slawi Kode Pos 52412 Telp. (0283) 491576 E-mail : stkipnutegaikab@gmail.com

Nomor : 010/Z/P-STKIPNU/XII/2025
Lamp. : -
Perihal : Surat Pemberitahuan

Yth. Ketua STKIP NU Kabupaten Tegal
Di
Tempat

Assalamualaikum Wr. Wb.

Dengan hormat,

Bersama ini kami sampaikan bahwa penelitian dalam rangka penyusunan skripsi mahasiswa yang dilaksanakan di lingkungan STKIP NU Kabupaten Tegal telah selesai dilaksanakan.

Adapun identitas mahasiswa peneliti adalah sebagai berikut:

Nama : Sofania Ma'rifatul Janah
NIM : 2102000080
Program Studi : Pendidikan Informatika
Judul Penelitian :

"Analisa Perbandingan Antara Linear Regression dan Random Forest Regression Dalam Prediksi Penerimaan Mahasiswa Baru di STKIP NU Kabupaten Tegal Tahun 2025-2026"

Penelitian tersebut telah dilaksanakan pada:

Waktu Pelaksanaan : 20-23 Agustus 2025
Tempat : STKIP NU Kabupaten Tegal

Demikian surat pemberitahuan selesai penelitian ini kami sampaikan sebagai bentuk laporan. Atas perhatian dan kerja sama yang telah diberikan, kami ucapkan terima kasih.

Wassalamualaikum Wr. Wb.

Ditetapkan : di Slawi
Pada Tanggal : 23 Desember 2025
Ketua STKIPNU Kabupaten Tegal



Dr. H. Akhmad Taufiq, M.Pd. I
NIDN. 2112126901

Lampiran 3 Surat Keterangan Perpustakaan

 **SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN
NAHDLATUL 'ULAMA KABUPATEN TEGAL
(STKIPNU)**
SK. BAN PT Nomor 124/SK/BAN-PT/Ak.B/PT/II/2024
Kampus : Jl. Jend. A. Yani No. 21 Procot Slawi Kode Pos 52412 Telp. (0283) 491576 E-mail : stkipnutegalkab@gmail.com

SURAT KETERANGAN PERPUSTAKAAN
No. 510/Pust/STKIPNU/VIII/2025

Saya Yang Bertandatangan Dibawah ini :

Nama : DIAN PURNAMASARI, S.I.Pust
Jabatan : Kepala Perpustakaan
NIPY : 1709.890722.017

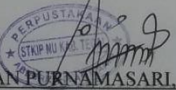
Menyatakan bahwa Mahasiswa dibawah ini :

Nama : SOFANIA MA'RIFATUL JANAH
NIM : 2102000080
Prodi : Pendidikan Informatika
Semester : VIII A (Delapan)

Telah menyelesaikan administrasi tidak ada tanggungan pinjaman buku, menyelesaikan laporan PPL, KKL dan KKN pada tahun 2024/2025, serta telah mengumpulkan di Perpustakaan Abdurrahman Wahid STKIP NU Kabupaten Tegal.

Slawi, 18 September 2025

Petugas Perpustakaan
Abdurrahman Wahid STKIP NU Kab. Tegal


DIAN PURNAMASARI, S.I.Pust
NIPY. 1709.890722.017

TEGAL

Lampiran 4 Surat Keterangan Turnitin

 **SEKOLAH TINGGI KEGURUAN DAN ILMU PENDIDIKAN
NAHDLATUL 'ULAMA KABUPATEN TEGAL
(STKIPNU)**
SK. BAN PT Nomor 124/SK/BAN-PT/Ak.B/PT/II/2024
Kampus : Jl. Jend. A. Yani No. 21 Pracot Slawi Kode Pos 52412 Telp. (0283) 491576 E-mail : stkipnutegalkab@gmail.com

SURAT KETERANGAN TURNITIN
No. 511/Pust/Turnitin/STKIPNU/VIII/2025

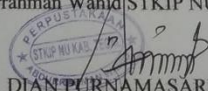
Sehubungan dengan kewajiban **Tes Turniti** dengan maksimal 30% sebagai salah satu kelengkapan persyaratan administrasi bagi mahasiswa tingkat akhir, dengan ini unit sistem Informasi dan Jurnal menerangkan bahwa :

Nama : SOFANIA MA'RIFATUL JANAH
NIM : 2102000080
Prodi : Pendidikan Informatika
Semester : VIII A(Delapan)
Tanggal Cek Turniti : 18 September 2025
Status : Lulus dengan Prosentasi 19%

Surat ini dibuat berdasarkan keadaan yang sebenar-benarnya, dan hasilnya terlampirkan.
Demikian surat ini untuk bisa dipergunakan sebagaimana mestinya.

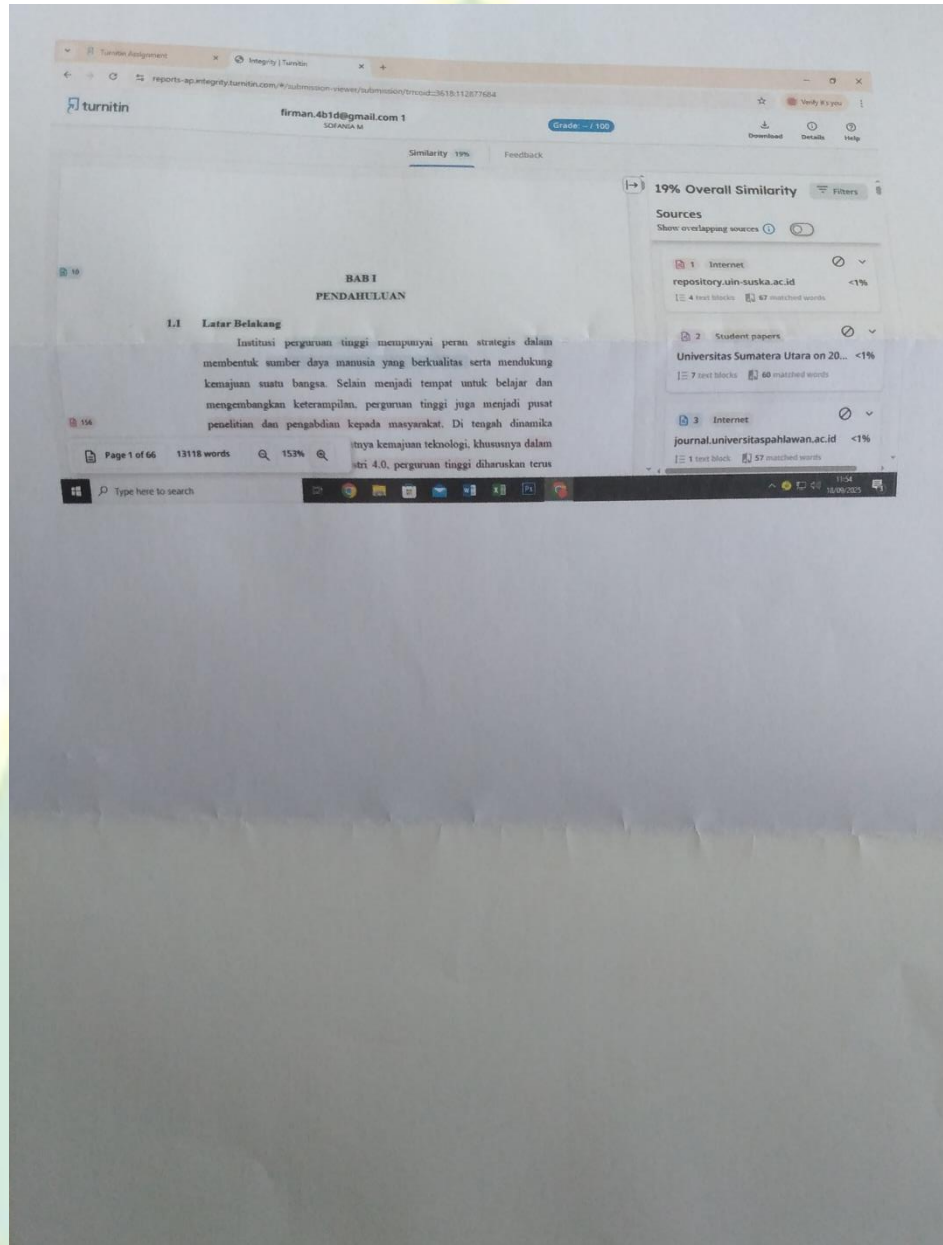
Slawi, 18 September 2025

Petugas Perpustakaan
Abdurrahman Wahid STKIP NU Kab. Tegal


DIAN PURNAMASARI, S.I.Pust
NIPY. 1709.890722.017

TEGAL

Lampiran 5 Hasil Turnitin



Lampiran 6 Surat Keterangan Selesai Revisi

100

Lampiran 6 Surat Keterangan Selesai Revisi

SURAT PERNYATAAN SELESAI REVISI

Saya yang bertanda tangan dibawah ini :

Nama : Sofania Ma'rifatul Janah

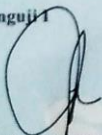
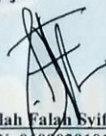
NIM : 2102000080

Program Studi : Pendidikan Informatika

Tanggal Sidang : 19 September 2025

Menegaskan bahwa mahasiswa tersebut telah menyelesaikan Bimbingan Revisi Skripsi. Demikian surat pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Tegal, 20 Desember 2025

Penguji I  <u>Rito Cipto Sigitta Hariyono, M.Kom</u> NIDN. 0610128301	Penguji II  <u>Fadilah Falah Svifa, M.Pd</u> NIDN. 0609029101
--	---

TEGAL