

**ANALISIS PERAMALAN MAHASISWA BARU MENGGUNAKAN
ALGORITMA *RANDOM FOREST* DAN *LINEAR REGRESSION***

TESIS

Untuk memenuhi sebagian persyaratan memperoleh gelar
Magister Teknologi Informasi

Disusun Oleh :

Muhammad Nurdin

227064518023



**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA**

UNIVERSITAS NASIONAL

JAKARTA

2024

LEMBAR PERSETUJUAN REVIEW AKHIR

Tesis dengan judul :

ANALISIS PERAMALAN MAHASISWA BARU MENGUNAKAN ALGORITMA RANDOM FOREST DAN LINEAR REGRESSION

Dibuat untuk melengkapi salah satu persyaratan menjadi Magister Komputer pada Program Studi Magister Teknologi Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tesis ini diujikan pada Sidang Akhir Semester Genap 2023-2024 pada tanggal 24 Agustus Tahun 2024

Jakarta, 30 Agustus 2024

Dosen Pembimbing


Dr. Fauziah, S.Kom., MMSI.
NIDN. 0304107503

Ketua Program Studi



Ir. Asrul Sani, S.T., M.T., M.Kom., Ph.D
NIDN. 0303067003

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Nurdin
NPM : 227064518023
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Magister Teknologi Informasi
Tanggal Sidang : 24 Agustus 2024

JUDUL DALAM BAHASA INDONESIA :

ANALISIS PERAMALAN MAHASISWA BARU MENGGUNAKAN ALGORITMA
RANDOM FOREST DAN LINEAR REGRESSION

JUDUL DALAM BAHASA INGGRIS :

ANALYSIS FORCASTING STUDENT USING RANDOM FOREST AND LINEAR
REGRESSION ALGORITHMS

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 3/9/2024	TGL : 3/9/2024	TGL :
 Dr. Fauziah, S.Kom, M.MSI	 	

PERNYATAAN ORISINALITAS

Saya bersaksi bahwa tesis ini sesungguhnya adalah karya ilmiah yang saya buat berdasarkan penelitian saya pribadi dengan arahan dan bimbingan oleh dosen pembimbing. Karya ilmiah ini tidak pernah diterbitkan oleh orang lain sebelumnya, terkecuali sesuai apa yang saya kutip dan yang tercantum pada daftar pustaka.

Jika memang terbukti bahwa naskah tesis ini mengandung unsur plagiarisme, saya bersedia untuk menerima pembatalan tesis dan pencabutan gelar akademik Magister yang telah saya peroleh dan saya juga siap bertanggung-jawab sesuai aturan yang berlaku.



KATA PENGANTAR

Proses penyusunan tesis ini berhasil diselesaikan berkat dukungan, bantuan, yang bermanfaat dan berharga dari berbagai pihak. Saya mengakui bahwa laporan ini telah mendapat banyak bimbingan dari berbagai individu dalam bentuk saran, kritik, dan bantuan lainnya. Oleh karena itu, saya bersyukur kepada Allah SWT yang telah memberikan nikmat dan karunia serta mengucapkan terima kasih yang besar kepada Ibu Dr. Fauziah, S.Kom., MMSI selaku pembimbing yang telah memberikan ilmu kepada saya sehingga dapat menyelesaikan tesis ini. Selanjutnya tidak lupa saya menghaturkan ucapan terima kasih untuk :

1. Bapak Dr. Agung Triayudi, S.Kom., M.Kom. Selaku Dekan Fakultas Teknologi Komunikasi dan Informatika,
2. Bapak Ir. Asrul Sani, ST, MT, M.Kom, Ph.D, selaku Ketua Program Studi Magister Teknologi Komunikasi dan Informatika,
3. Bapak dan Ibu Dosen Fakultas Teknologi Komunikasi dan Informatika beserta Staf Tata Usaha FTKI yang telah membantu perkuliahan saya,
4. Ibu ku tersayang, Istriku tercinta dan anak-anakku, Farhan, Fathia dan Fauziah .

Penelitian ini bertujuan untuk meramalkan jumlah mahasiswa baru di Universitas dengan mengolah data jumlah mahasiswa baru dari tahun 2017 hingga 2023 untuk memperkirakan jumlah mahasiswa baru dalam 5 tahun ke depan. Metode yang digunakan adalah *Random Forest* dan *Linear Regression*, yang telah terbukti efektif sebagai dasar studi analitik dalam meramalkan data time series. Saya sadar bahwa penulisan ini jauh dari sempurna dan mempunyai keterbatasan. Kritik dan saran konstruktif dari pembaca untuk meningkatkan dan mengembangkan penelitian ini di masa depan.

Jakarta, 17 Juli 2024


Muhammad Nurdin

ABSTRAK

Perencanaan jumlah mahasiswa baru (PMB) merupakan aspek krusial bagi perguruan tinggi karena memiliki dampak signifikan terhadap perencanaan sumber daya manusia dan alokasi anggaran. Namun, proses prediksi jumlah mahasiswa baru yang tepat menghadapi berbagai tantangan akibat pengaruh faktor eksternal seperti kondisi ekonomi, kebijakan pemerintah, dan tingkat persaingan di antara lembaga pendidikan. Penelitian ini bertujuan untuk melakukan analisis peramalan penerimaan mahasiswa baru di Universitas Nasional dengan menggunakan algoritma Random Forest (RF) dan Linear Regression (LR). Data yang digunakan dalam penelitian ini adalah data historis pembelian formulir mahasiswa dari tahun 2017 hingga 2023. Pada periode tersebut, jumlah pembeli formulir tercatat sebanyak 37.078, dengan rincian jumlah pembeli formulir per tahun sebagai berikut: 5.178 pada tahun 2017, 6.733 pada tahun 2018, 5.978 pada tahun 2019, 5.042 pada tahun 2020, 5.184 pada tahun 2021, 5.360 pada tahun 2022, dan 3.603 pada tahun 2023. Metode peramalan, yaitu RF dan LR, diterapkan untuk membandingkan performa prediksi yang dihasilkan. Hasil analisis menunjukkan bahwa model RF memberikan performa yang lebih baik dibandingkan dengan LR, hal ini ditunjukkan oleh nilai MAE yang lebih rendah pada RF (202.614) dibandingkan dengan LR (503.233), yang mengindikasikan kesalahan mutlak rata-rata yang lebih kecil pada model Random Forest. Selain itu, MSE dan RMSE juga lebih rendah pada RF (69.764,003 dan 264.178) dibandingkan dengan LR (268.276,644 dan 518.023), yang menunjukkan bahwa RF lebih efektif dalam menangani outlier serta memberikan bobot yang lebih rendah pada kesalahan besar. Nilai MAPE yang lebih rendah pada RF (4.109%) dibandingkan dengan LR (10.647%) menunjukkan tingkat kesalahan relatif yang lebih kecil. Sementara itu, nilai MAD pada RF (187.781) yang lebih rendah juga menunjukkan konsistensi prediksi yang lebih baik dibandingkan dengan LR (467.999). Secara keseluruhan, model RF memberikan hasil prediksi yang lebih akurat dan konsisten, sehingga dapat disimpulkan bahwa RF merupakan model yang lebih unggul dalam memprediksi jumlah mahasiswa baru dibandingkan dengan LR.

Kata kunci : Prediksi Mahasiswa Baru, Random Forest (RF), Linear Regression (LR), Data Pembelian Formulir, Metrik Evaluasi (MAE, MSE, RMSE, MAPE, MAD), Akurasi dan Konsistensi

ABSTRACT

The planning of new student admissions (PMB) is a crucial aspect for higher education institutions as it significantly impacts human resource planning and budget allocation. However, accurately predicting the number of new students faces various challenges due to the influence of external factors such as economic conditions, government policies, and the level of competition among educational institutions. This study aims to analyse the forecasting of new student admissions at Universitas Nasional using the Random Forest (RF) and Linear Regression (LR) algorithms. The data used in this research is historical data on student form purchases from 2017 to 2023. During this period, the number of form purchases recorded was 37,078, with the following breakdown per year: 5,178 in 2017, 6,733 in 2018, 5,978 in 2019, 5,042 in 2020, 5,184 in 2021, 5,360 in 2022, and 3,603 in 2023. The forecasting methods, RF and LR, were applied to compare the predictive performance of the models. The analysis results indicate that the RF model provides better performance compared to LR, as evidenced by a lower MAE in RF (202.614) compared to LR (503.233), indicating a smaller average absolute error in the Random Forest model. Additionally, the MSE and RMSE are also lower in RF (69,764.003 and 264.178) compared to LR (268,276.644 and 518.023), showing that RF is more effective in handling outliers and giving lower weights to large errors. The lower MAPE in RF (4.109%) compared to LR (10.647%) indicates a smaller relative error rate. Meanwhile, the lower MAD in RF (187.781) also shows better prediction consistency compared to LR (467.999). Overall, the RF model provides more accurate and consistent prediction results, leading to the conclusion that RF is a superior model in predicting the number of new students compared to LR.

Keywords: Keywords: New Student Admission Prediction, Random Forest (RF), Linear Regression (LR), Form Purchase Data, Evaluation Metrics (MAE, MSE, RMSE, MAPE, MAD), Accuracy and Consistency

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
KATA PENGANTAR	iii
ABSTRAK	iv
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Manfaat.....	4
1.5. Batasan Masalah.....	5
1.6. Sistematika Penulisan.....	5
BAB II LANDASAN KEPUSTAKAAN	7
2.1. Studi Literatur.....	7
2.2. Landasan Teori.....	11
BAB III METODE PENELITIAN	13
3.1. Tahapan Penelitian	13
3.2. Pengumpulan Data.....	16
3.3. Preprocessing Data.....	20
3.4. Pelatihan Model.....	21
3.5. Evaluasi Model.....	21
3.6. Analisa Hasil	22
3.7. Evaluasi Hasil.....	23
3.8. Forecasting/ Proses Peramalan	23
3.9. Tahapan Akhir Proses Perhitungan	24
BAB IV IMPLEMENTASI	25
4.1 Data Set	25
4.2. Prosesing Data.....	26

4.3. Dinamisasi Dataset Pola Berjenjang:	31
4.4. Observasi Utama	32
BAB V ANALISA DAN EVALUASI HASIL.....	34
5.1. Analisa dan Evaluasi Hasil Prediksi.....	34
5.2. Analisa dan Evaluasi Hasil Performa	37
BAB VI PENUTUP	41
6.1. Simpulan.....	41
6.2. Saran.....	41
DAFTAR PUSTAKA	43



DAFTAR GAMBAR

Gambar 2. 1. Gambar Skematik Random Forest	12
Gambar 3. 1. Tahapan Penelitian	13
Gambar 3. 2. Data Set Transaksi Pembelian Formulir.....	16
Gambar 3. 3. Data set pembelian formulir tahun 2017	17
Gambar 3. 4. Data set pembelian formulir tahun 2018.....	17
Gambar 3. 5. Data set pembelian formulir tahun 2019	18
Gambar 3. 6. Data set pembelian formulir tahun 2020	18
Gambar 3. 7. Data set pembelian formulir tahun 2021	19
Gambar 3. 8. Data set pembelian formulir tahun 2022	19
Gambar 3. 9. Data set pembelian formulir tahun 2023.....	20
Gambar 4. 1. Dataset Jumlah Pembeli Formulir Tahun 2017 sampai 2023	25
Gambar 4. 3. Hasil Pengujian RF dan LR.....	26
Gambar 4. 4. Gambar Prediksi Pembelian Formulir Tahun 2024-2028	27
Gambar 4. 5. Grafik Perbandingan MAE RF dan LR.....	28
Gambar 4. 6. Grafik Perbandingan MSE RF dan LR	29
Gambar 4. 7. Grafik Perbandingan RMSE RF dan LR.....	29
Gambar 4. 8. Grafik Perbandingan MAPE RF dan LR.....	30
Gambar 5 1. Pengolahan Peramalan RF dan LR Menggunakan Google Colabs.....	37
Gambar 5 2. Pengolahan Metrik RF dan LR Menggunakan Google Colabs.....	39

DAFTAR TABEL

Tabel 2. 1. Studi Literatur	7
Tabel 3. 1. Data Set Transaksi Pembelian Formulir	16
Tabel 4.1. Prediksi Jumlah Pendaftar.....	27
Tabel 4. 2. Analisis Metrik Dataset.....	28
Tabel 4. 3. Analisis Data Set yang digunakan	30

