

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
HALAMAN MOTTO DAN PERSEMBAHAN.....	iv
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xix
INTISARI.....	xxii
ABSTRACT.....	xxiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	7
BAB III LANDASAN TEORI.....	23
3.1 Gaokao.....	23
3.2 Model Kompetisi.....	23
3.3 Seleksi Fitur.....	28
3.3.1 <i>Filtering Method</i> .....	28
3.3.2 <i>Analysis of Variance (ANOVA)</i> .....	28
3.3.3 Metode Kruskal-Wallis.....	29
3.3.4 <i>Wrapper Method</i> .....	30
3.3.5 <i>Embedded Method</i> .....	31
3.4 <i>Gradient Boosted Decision Tree</i> .....	32
3.5 <i>Linear Regression</i> .....	33

3.6	<i>Ordinary Least Squares</i> .....	34
3.7	<i>Multi Layer Perceptron</i> .....	35
3.8	<i>Long Short Term Memory (LSTM)</i> .....	36
3.9	<i>Clustering</i> .....	38
3.9.1	<i>K-Means Clustering</i> .....	38
3.9.2	<i>K-Means++</i> .....	39
3.9.3	<i>Elbow Method</i> .....	40
3.10	<i>Gradient Descent</i> .....	41
3.11	<i>Regularisasi</i> .....	42
3.12	<i>ETL (Extract, Transform, Load)</i> .....	42
3.13	<i>Frequency Encoding</i> .....	43
3.14	<i>Precision</i> .....	43
3.15	<i>Mean Square Error (MSE)</i> .....	43
3.16	<i>Mean Absolute Error (MAE)</i> .....	44
BAB IV ANALISIS DAN PERANCANGAN .....		45
4.1	<i>Deskripsi Umum Penelitian</i> .....	45
4.2	<i>Alat dan Bahan</i> .....	48
4.3	<i>Pengumpulan Data</i> .....	49
4.4	<i>Prapemrosesan Data</i> .....	51
4.4.1.	<i>Konversi Data</i> .....	52
4.4.2.	<i>Penerjemahan ke Bahasa Indonesia</i> .....	53
4.5	<i>Pemilihan Data</i> .....	53
4.5.1.	<i>Ekstraksi Data (Extract)</i> .....	54
4.5.2.	<i>Transformasi Data (Transform)</i> .....	54
4.5.3.	<i>Pemuatan Data Hasil (Load)</i> .....	55
4.6	<i>Pembagian Data Training dan Testing</i> .....	59
4.7	<i>Pembangunan Model Preliminary Prediction</i> .....	60
4.7.1	<i>Gradient Boosted Decision Tree (GBDT)</i> .....	60
4.7.2	<i>Linear Regression</i> .....	62

4.7.3	<i>Multi Layer Perceptron (MLP)</i> .....	67
4.7.4	<i>Long Short Term Memory (LSTM)</i> .....	72
4.8	Pembuatan Model Kompetisi.....	78
4.9	Evaluasi Model.....	84
BAB V IMPLEMENTASI.....		88
5.1	Gambaran Umum.....	88
5.2	Implementasi Proses ETL dalam ETL.ipynb.....	89
5.2.1	Meng- <i>import</i> Library yang Digunakan.....	89
5.2.2	Melakukan Persiapan Data.....	89
5.2.3	Normalisasi Nama Kolom.....	91
5.2.4	Membuat <i>Key</i> yang Menggabungkan Nama Program Studi dan Nama Universitas.....	94
5.2.5	Melakukan Penggabungan Data.....	95
5.2.6	Menyimpan Data Hasil Penggabungan.....	96
5.2.7	Eksplorasi Data Gabungan.....	96
5.2.8	Konversi dan Normalisasi Data.....	98
5.2.9	Membuat Peringkat.....	99
5.2.10	Ekstraksi Fitur Baru.....	101
5.2.11	Pembagian <i>Stable Set</i> dan <i>Unstable Set</i> .....	102
5.2.12	Melakukan Seleksi Data.....	103
5.3	Eksplorasi Fitur.....	104
5.3.1	Analisis Korelasi.....	104
5.3.2	Analisis Variansi menggunakan ANOVA.....	107
5.3.3	Eksplorasi Variansi.....	108
5.4	Implementasi Prediksi.....	108
5.4.1	Prediksi menggunakan <i>Gradient Boosted Decision Tree</i> .....	109
5.4.2	Prediksi menggunakan <i>Linear Regression</i> .....	115
5.4.3	Prediksi menggunakan <i>Multi Layer Perceptron</i> .....	127
5.4.4	Prediksi menggunakan <i>Long Short Term Memory (LSTM)</i> ...	137

5.5	Implementasi Model Kompetisi .....	144
BAB VI HASIL DAN PEMBAHASAN .....		154
6.1	Eksplorasi Data .....	154
6.1.1	Sebaran Data Jenis Perguruan Tinggi .....	154
6.1.2	Sebaran Data Perguruan Tinggi .....	155
6.1.3	Sebaran Data Fakultas (Departemen).....	155
6.1.4	Sebaran Data Program Studi .....	156
6.2	Eksplorasi Fitur .....	157
6.2.1	Analisis Korelasi Secara Grafis .....	157
6.2.2	Analisis <i>Heatmap</i> .....	160
6.2.3	Analisis Korelasi Secara Matematis.....	162
6.2.4	Analisis Variansi Menggunakan ANOVA.....	163
6.2.5	Eksplorasi Varian .....	164
6.3	Hasil Prediksi .....	165
6.3.1	Prediksi Menggunakan <i>Gradient Boosted Decision Tree</i> (GBDT) 165	
6.3.2	Prediksi Menggunakan <i>Linear Regression</i> .....	167
6.3.3	Prediksi Menggunakan <i>Multi Layer Perceptron</i> .....	176
6.3.4	Prediksi Menggunakan <i>Long Short Term Memory</i> (LSTM)...	188
6.4	Model Kompetisi.....	193
6.4.1	Model Kompetisi pada <i>Gradient Boosted Decision Tree</i> .....	193
6.4.2	Model Kompetisi pada <i>Linear Regression</i> .....	198
6.4.3	Model Kompetisi pada <i>Multi Layer Perceptron</i> .....	202
6.4.4	Model Kompetisi pada <i>Long Short Term Memory</i> .....	208
6.5	Pembahasan Hasil .....	210
6.6	Pengaruh Regularisasi pada Model <i>Multi Layer Prerceptron</i> dan <i>Long Short Term Memory</i> .....	213
6.7	Pengaruh Metode Penentuan <i>Centroid</i> pada Model Kompetisi dengan <i>K-Means Clustering</i> .....	214

6.8	Pengaruh Model Kompetisi.....	214
BAB VII KESIMPULAN DAN SARAN .....		216
7.1	Kesimpulan .....	216
7.2	Saran.....	217
DAFTAR PUSTAKA .....		218