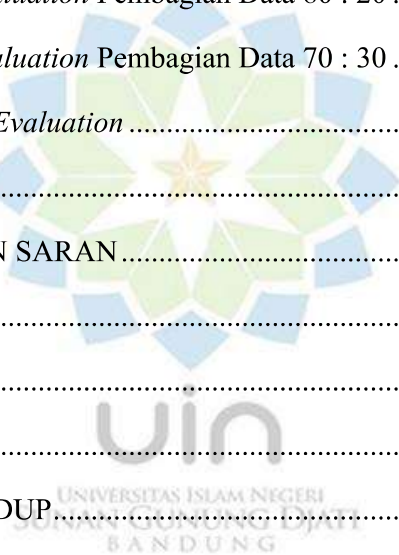


## DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
LEMBAR PENGESAHAN .....	ii
LEMBAR PERNYATAAN KARYA.....	iii
LEMBAR PERSEMBAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiv
DAFTAR ISTILAH .....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
1.6 Kerangka Pemikiran .....	6
1.7 Metodologi Penelitian.....	7
1.8 Sistematika Penulisan .....	9
BAB II KAJIAN LITERATUR .....	11
2.1 <i>The State Of The Art</i> .....	11
2.2 Landasan Teori .....	17
2.2.1 Berlian.....	17
2.2.2 Data Mining .....	19

2.2.3	Machine Learning .....	19
2.2.4	Regresi .....	20
2.2.5	Adaptive Boosting <i>Regressor</i> .....	21
2.2.6	<i>Random Forest Regressor</i> .....	23
2.2.7	K-Nearest Neighbor (KNN) <i>Regressor</i> .....	24
2.2.8	Mean Squared <i>Error</i> (MSE) .....	26
2.2.9	Root Mean Square <i>Error</i> (RMSE).....	27
2.2.10	R-Squared (R <sup>2</sup> -Score) .....	28
2.2.11	CRISP-DM.....	29
2.2.12	Bahasa Pemrograman <i>Python</i> .....	31
BAB III METODOLOGI PENELITIAN.....		32
3.1	<i>Business Understanding</i> .....	33
3.2	<i>Data Understanding</i> .....	33
3.3	<i>Data Preparation</i> .....	35
3.4	<i>Modeling</i> .....	42
3.5	<i>Evaluation</i> .....	43
BAB IV HASIL DAN PEMBAHASAN.....		45
4.1	Hasil <i>Business Understanding</i> .....	45
4.2	Hasil <i>Data Understanding</i> .....	46
4.2.1	<i>Import Package</i> dan <i>Dataset</i> .....	46
4.2.2	Deskripsi Variabel .....	47
4.2.3	Deskripsi Statistik pada <i>Dataset</i> .....	48
4.2.4	<i>Univariate Analysis</i> .....	49
4.2.5	<i>Correlation Matrix</i> .....	53
4.3	Hasil <i>Data Preparation</i> .....	54
4.3.1	Menangani <i>Missing Value</i> dan <i>Outliers</i> .....	55

4.3.2	Reduksi Dimensi.....	59
4.3.3	<i>Splitting Data</i> .....	62
4.3.4	Standarisasi .....	63
4.4	Hasil <i>Modeling</i> .....	65
4.4.1	Deklarasi Algoritma.....	66
4.4.2	Durasi Pelatihan Model .....	69
4.5	Hasil <i>Evaluation</i> .....	71
4.5.1	Skenario <i>Evaluation</i> Pembagian Data 90 : 10 .....	72
4.5.2	Skenario <i>Evaluation</i> Pembagian Data 80 : 20 .....	74
4.5.3	Skenario <i>Evaluation</i> Pembagian Data 70 : 30 .....	76
4.5.4	Hasil Akhir <i>Evaluation</i> .....	78
4.6	Pembahasan .....	79
BAB V SIMPULAN DAN SARAN.....		82
5.1	Simpulan .....	82
5.2	Saran .....	83
DAFTAR PUSTAKA .....		84
DAFTAR RIWAYAT HIDUP.....		91



## DAFTAR GAMBAR

Gambar 1.1 Kerangka Pemikiran.....	6
Gambar 1.2 Alur CRISP-DM.....	7
Gambar 2.1 Adaboost <i>Regressor</i> .....	21
Gambar 2.2 <i>Random Forest Regressor</i> .....	23
Gambar 2.3 <i>K-Nearest Neighbors Regressor</i> .....	25
Gambar 3.1 <i>Flowchart</i> Metodologi Penelitian .....	32
Gambar 3.2 Contoh tabel <i>dataset</i> .....	34
Gambar 3.3 Visualisasi data dengan <i>Scatter Plot</i> .....	35
Gambar 3.4 Metode IQR.....	37
Gambar 3.5 Contoh <i>Encoding Data</i> .....	38
Gambar 3.6 Contoh Implementasi Metode PCA .....	39
Gambar 3.7 Skenario Pembagian Data <i>Train-Test</i> .....	40
Gambar 3.8 Skenario <i>Modeling</i> .....	42
Gambar 3.9 Skenario <i>Evaluation</i> .....	43
Gambar 4.1 Import Package dan Dataset.....	46
Gambar 4.2 Ilustrasi Variabel .....	47
Gambar 4.3 Deskripsi Statistik .....	48
Gambar 4.4 Pembagian Kelas Pada Variabel .....	49
Gambar 4.5 Proses Analisis Variabel <i>Cut</i> .....	49
Gambar 4.6 Proses Analisis Variabel <i>Color</i> .....	50
Gambar 4.7 Proses Analisis Variabel <i>Clarity</i> .....	51
Gambar 4.8 Analisis Variabel Numerik.....	52
Gambar 4.10 <i>Correlation Matrix</i> .....	53
Gambar 4.11 <i>Drop Kolom Depth</i> .....	54
Gambar 4.12 Memeriksa <i>Missing Value</i> .....	55

Gambar 4.13 Tampilan Baris dengan <i>Missing Value</i> .....	55
Gambar 4.14 <i>Drop</i> baris dengan <i>Missing Value</i> .....	56
Gambar 4.15 Tampilan <i>Boxplot</i> Variabel Numerik.....	56
Gambar 4.16 Penerapan IQR .....	57
Gambar 4.17 Penerapan <i>Encoding</i> pada Variabel Kategori.....	58
Gambar 4.18 Pengecekan Korelasi Variabel X, Y, dan Z .....	60
Gambar 4.18 Implementasi PCA .....	60
Gambar 4.19 Proporsi informasi dari setiap PC.....	61
Gambar 4.20 Tampilan <i>Dataset</i> Setelah proses PCA .....	61
Gambar 4.21 <i>Splitting Data</i> .....	62
Gambar 4.22 Standarisasi menggunakan <i>StandardScaler</i> .....	63
Gambar 4.23 Variabel Numerik <i>X_Train</i> setelah Standarisasi.....	64
Gambar 4.24 <i>Import</i> Model dan Metrik.....	65
Gambar 4.25 Kode Program <i>Random Forest Regressor</i> .....	66
Gambar 4.26 Kode Program <i>K-Nearest Neighbors Regressor</i> .....	67
Gambar 4.27 Kode Program Adaptive Boosting <i>Regressor</i> .....	68
Gambar 4.28 Standarisasi Data Uji ( <i>Test</i> ) .....	71

## DAFTAR TABEL

Tabel 2.1 Referensi Penelitian .....	11
Tabel 3.1 Deskripsi Variabel .....	34
Tabel 4.1 Deskripsi Variabel .....	47
Tabel 4.2 Variabel <i>Clarity</i> .....	51
Tabel 4.3 Pembagian Skenario <i>Splitting Data</i> .....	62
Tabel 4.4 Skenario <i>Splitting Data</i> .....	63
Tabel 4.5 Spesifikasi Perangkat .....	65
Tabel 4.6 Skenario <i>Random Forest Regressor</i> .....	66
Tabel 4.7 Skenario <i>K-Nearest Neighbors Regressor</i> .....	67
Tabel 4.8 Skenario Adaptive Boosting <i>Regressor</i> .....	68
Tabel 4.9 Durasi Pelatihan Model dengan <i>Splitting 90 : 10</i> .....	69
Tabel 4.10 Durasi Pelatihan Model dengan <i>Splitting 80 : 20</i> .....	69
Tabel 4.11 Durasi Pelatihan Model dengan <i>Splitting 70 : 30</i> .....	70
Tabel 4.12 Nilai MSE dan RMSE dengan <i>Splitting 90 : 10</i> .....	72
Tabel 4.13 Nilai <i>R-Squared</i> dan Prediksi dengan <i>Splitting 90 : 10</i> .....	73
Tabel 4.14 Nilai MSE dan RMSE dengan <i>Splitting 80 : 20</i> .....	74
Tabel 4.15 Nilai <i>R-Squared</i> dan Prediksi dengan <i>Splitting 80 : 20</i> .....	75
Tabel 4.16 Nilai MSE dan RMSE dengan <i>Splitting 70 : 30</i> .....	76
Tabel 4.17 Nilai <i>R-Squared</i> dan Prediksi dengan <i>Splitting 70 : 30</i> .....	77
Tabel 4.18 Algoritma Terbaik.....	78
Tabel 4.19 Rata Rata Akurasi dan Prediksi Gabungan.....	79
Tabel 4.20 Rata Rata <i>Error</i> Gabungan .....	80

## DAFTAR ISTILAH

**Overfitting** : Suatu kondisi di mana model terlalu kompleks dan terlalu cocok dengan data pelatihan sehingga tidak dapat digeneralisasi dengan baik ke data baru.

**Underfitting** : Suatu kondisi di mana model terlalu sederhana sehingga tidak dapat menangkap pola yang ada di dalam data.

**Outliers** : Data yang jauh berbeda dari data lainnya dalam kumpulan data.

**Scripting** : Proses menulis kode Python yang dapat dijalankan secara langsung tanpa perlu dikompilasi terlebih dahulu.

**Library** : Kumpulan modul dan fungsi yang dapat digunakan untuk memperluas fungsionalitas Python.

**Package** : *Package* pada Python adalah cara untuk mengorganisir modul-modul yang terkait menjadi satu kesatuan yang lebih besar. *Package* ini biasanya terdiri dari beberapa *file* modul Python dan *sub-package* yang terorganisir dalam struktur direktori tertentu.

**Mohs** : Skala pengukuran kekerasan mineral.

**Abrasive** : Material kasar yang digunakan untuk mengasah atau membersihkan permukaan benda.

**Lazy Learning** : Suatu pendekatan pembelajaran mesin di mana model tidak melakukan proses pembelajaran pada saat pelatihan, melainkan hanya saat pengujian.

**Regressor** : Model pembelajaran mesin yang digunakan untuk memprediksi nilai kontinu dari suatu variabel target.

**Karat/Carat** : Satuan pengukuran berat berlian yang setara dengan 0,2 gram.