

DAFTAR ISI

HALAMAN JUDUL	
FORMULIR PERNYATAAN KEASLIAN DAN PERSETUJUAN	
UNGGAH TUGAS AKHIR	
PERSETUJUAN DOSEN PEMBIMBING SKRIPSI	
PERSETUJUAN TIM PENGUJI SKRIPSI	
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
 BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	4
1.5.1 Manfaat Teoretis	5
1.5.2 Manfaat Praktis	5
1.6 Sistematika Penulisan	5
 BAB II LANDASAN TEORI	
2.1 Segmentasi Pelanggan	7
2.1.1 Segmentasi Demografi	7
2.1.2 Segmentasi Geografi	8
2.1.3 Segmentasi Perilaku	8
2.1.4 Segmentasi Psikografis	8
2.2 Pemasaran	8
2.3 Machine Learning	9
2.3.1 Supervised Learning	9
2.3.2 Unsupervised Learning	9
2.3.3 Semi-Supervised Learning	9
2.3.4 Reinforcement Learning	10
2.4 Decision Tree	10
2.5 Random Forest	11
2.6 Extreme Gradient Boosting (XGBoost)	12
2.6.1 Gradient Boosting	12
2.7 Support Vector Machine (SVM)	13
2.8 Hierarchical Clustering	15
2.8.1 Perhitungan dissimilarity antar Data	15
2.8.1.1 Euclidean Distance	16

2.8.1.2	<i>Manhattan Distance</i>	16
2.8.2	Perhitungan <i>dissimilarity</i> antar Klaster	16
2.8.2.1	<i>Agglomerative Clustering</i>	17
2.8.2.2	<i>Divisive Clustering</i>	18
2.9	Metrik Evaluasi	18
2.9.1	<i>Accuracy</i>	18
2.9.2	<i>Precision</i>	19
2.9.3	<i>Recall</i>	19
2.9.4	<i>F1-Score</i>	19
2.10	Tinjauan Pustaka	20
 BAB III METODE PENELITIAN		
3.1	Data	21
3.2	Data <i>Pre-processing</i>	24
3.2.1	Pembersihan Data	24
3.2.2	<i>Encoding Data</i>	24
3.2.2.1	<i>Loyalty Label</i>	24
3.2.2.2	Tipe Bebas Ongkir	27
3.2.2.3	<i>IsWeekend</i>	27
3.2.2.4	Waktu	27
3.2.2.5	Kategori Penjualan	27
3.2.2.6	Kota	28
3.2.2.7	<i>Hour Difference</i>	28
3.3	Penentuan Variabel Dependen dan Variabel Independen	28
3.4	Pembangunan Model	28
3.4.1	Pembangunan Model <i>Decision Tree</i>	29
3.4.2	Pembangunan Model <i>Random Forest</i>	29
3.4.3	Pembangunan Model <i>XGBoost</i>	30
3.4.4	Pembangunan Model SVM	30
3.4.5	Pembangunan Model Hierarchical Clustering	30
3.5	Prediksi Data	31
3.6	Analisis Evaluasi Metrik	31
 BAB IV ANALISIS DAN PEMBAHASAN		
4.1	Data	32
4.1.1	<i>Encoding Data</i>	32
4.1.2	Transformasi Data	34
4.2	Hasil Klasifikasi Menggunakan Metode <i>Decision Tree</i>	36
4.3	Hasil Klasifikasi Menggunakan Metode <i>Random Forest</i>	39
4.4	Hasil Klasifikasi Menggunakan Metode <i>XGBoost</i>	40
4.5	Hasil Klasifikasi Menggunakan Metode <i>Support Vector Machine</i>	43
4.6	Hasil <i>Clustering</i> Menggunakan Metode Hierarchical Clustering	45
4.6.1	Visualisasi Setiap Variabel	55
4.7	Hasil Evaluasi Metode Klasifikasi dan <i>Clustering</i>	58

4.8 Analisis Hasil	59
4.9 Implikasi Manajerial	65

BAB V PENUTUP

5.1 Kesimpulan	66
5.2 Saran	67
5.2.1 Saran Teoretis	67
5.2.2 Saran Praktis	68

DAFTAR PUSTAKA



DAFTAR GAMBAR

Gambar 1.1	Ilustrasi Segmentasi Pelanggan	1
Gambar 3.1	Diagram Alir Langkah-Langkah Penelitian	22
Gambar 4.1	Histogram Variabel <i>Hour Difference</i>	34
Gambar 4.2	Histogram Variabel <i>Hour Difference</i> Setelah Transformasi	36
Gambar 4.3	Hasil <i>Decision Tree</i>	37
Gambar 4.4	<i>Random Forest Plot</i>	39
Gambar 4.5	<i>Feature Importance</i> berdasarkan <i>XGBoost</i>	40
Gambar 4.6	Visualisasi Pohon dari Metode <i>XGBoost</i>	41
Gambar 4.7	Nilai Aktual dari Variabel <i>Loyalty Label</i>	44
Gambar 4.8	Nilai Prediksi dari Variabel <i>Loyalty Label</i>	44
Gambar 4.9	Grafik ROC dari Metode SVM	45
Gambar 4.10	Jumlah Klaster Optimal dengan <i>Elbow Method</i>	46
Gambar 4.11	Dendrogram dari <i>Hierarchical Clustering</i>	46
Gambar 4.12	Dendrogram Hasil Pengelompokan	47
Gambar 4.13	Visualisasi Sub-klaster pada Klaster 1 - Bagian 1	47
Gambar 4.14	Visualisasi Sub-klaster pada Klaster 1 - Bagian 2	49
Gambar 4.15	Visualisasi Sub-klaster pada Klaster 2 - Bagian 1	51
Gambar 4.16	Visualisasi Sub-klaster pada Klaster 2 - Bagian 2	52
Gambar 4.17	Frekuensi Data secara <i>Agglomerative</i>	55
Gambar 4.18	Distribusi Variabel Tipe Bebas Ongkir	55
Gambar 4.19	Distribusi Variabel <i>IsWeekend</i>	56
Gambar 4.20	Distribusi Variabel Waktu	56
Gambar 4.21	Distribusi Variabel Kategori Penjualan	57
Gambar 4.22	Distribusi 15 Kota Teratas	57
Gambar 4.23	Distribusi Variabel <i>Hour Difference</i>	58

DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix</i> untuk Klasifikasi	18
Tabel 3.1	Data Penjualan per Transaksi - Bagian 1	25
Tabel 3.2	Data Penjualan per Transaksi - Bagian 2	25
Tabel 3.3	Data Penjualan per Transaksi - Bagian 3	25
Tabel 3.4	Data Penjualan per Transaksi - Bagian 4	26
Tabel 3.5	Data Penjualan per Transaksi - Bagian 5	26
Tabel 4.1	Hasil <i>Encoding</i> untuk Variabel <i>Loyalty Label</i>	32
Tabel 4.2	Hasil <i>Encoding</i> untuk Variabel Tipe Bebas Ongkir	33
Tabel 4.3	Hasil <i>Encoding</i> untuk Variabel <i>IsWeekend</i>	33
Tabel 4.4	<i>Dummy Variable</i> dari Variabel Waktu	33
Tabel 4.5	<i>Dummy Variable</i> dari Variabel Kategori Penjualan	35
Tabel 4.6	<i>Dummy Variable</i> dari Variabel Kota	35
Tabel 4.7	Hasil Uji <i>kernel</i> untuk Algoritma SVM	43
Tabel 4.8	Karakteristik Pelanggan Sub-Klaster pada Klaster 1 - Bagian 1	48
Tabel 4.9	Karakteristik Pelanggan Sub-Klaster pada Klaster 1 - Bagian 2	50
Tabel 4.10	Karakteristik Pelanggan Sub-Klaster pada Klaster 2 - Bagian 1	51
Tabel 4.11	Karakteristik Pelanggan Sub-Klaster pada Klaster 2 - Bagian 2	52
Tabel 4.12	<i>Confusion Matrix</i> dari Setiap Model	58
Tabel 4.13	Evaluasi Metrik dari Setiap Model	59
Tabel 4.14	<i>Variable Importance</i> dari Setiap Model	61
Tabel 4.15	Persentase Proporsi Variabel Tipe Bebas Ongkir	61
Tabel 4.16	Persentase Proporsi Variabel <i>IsWeekend</i>	62
Tabel 4.17	Persentase Proporsi Variabel Kategori Penjualan	62
Tabel 4.18	Persentase Proporsi Variabel Waktu	62
Tabel 4.19	Persentase Proporsi Variabel Kota	63
Tabel 4.20	Ringkasan Statistik Variabel <i>Hour Difference</i>	63