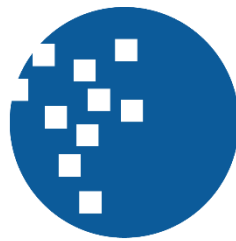


**IMPLEMENTASI MODEL SEGMENTASI PELANGGAN
BERDASARKAN TRANSAKSI PEMBELIAN PRODUK
PERUSAHAAN *FASHION XYZ* MENGGUNAKAN
ALGORITMA *CLUSTERING K-MEANS* DAN *DBSCAN***



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

William

00000042609

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**IMPLEMENTASI MODEL SEGMENTASI PELANGGAN
BERDASARKAN TRANSAKSI PEMBELIAN PRODUK
PERUSAHAAN *FASHION XYZ* MENGGUNAKAN
ALGORITMA *CLUSTERING K-MEANS* DAN *DBSCAN***



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom.)

William

00000042609

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA**

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

i

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : William

Nomor Induk Mahasiswa : 00000042609

Program studi : Sistem Informasi

Skripsi dengan judul:

"Implementasi Model Segmentasi Pelanggan Berdasarkan Transaksi Pembelian Produk Perusahaan *Fashion XYZ* Menggunakan Algoritma *Clustering K-Means* dan *DBSCAN*" merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk skripsi yang telah saya tempuh.

Tangerang, 13 Mei 2024



William

HALAMAN PENGESAHAN

Skripsi dengan judul

“Implementasi Model Segmentasi Pelanggan Berdasarkan Transaksi Pembelian Produk Perusahaan *Fashion XYZ* Menggunakan Algoritma *Clustering K-Means* dan *DBSCAN*”

Oleh

Nama : William
NIM : 00000042609
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Selasa, 28 Mei 2024

Pukul 10.00 s.d 12.00 dan dinyatakan

LULUS

Dengan susunan pengujian sebagai berikut.

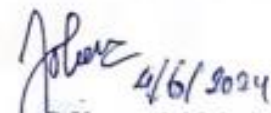
Ketua Sidang

Wella, S.Kom., M.M.S.I.
0305119101

 5/6/24

Penguji

Johan Setiawan, S.Kom., M.M., M.B.A.
0327106402

 4/6/2024

Pembimbing

Monika Evelin Johan, S.Kom.,
M.M.S.I.
0327059501

 5/6/24

Ketua Program Studi Sistem Informasi

Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001

 5/6/24

iii

LEMBAR PERSETUJUAN PUBLIKASI

KARYA ILMIAH MAHASISWA

Yang bertanda tangan dibawah ini:

Nama : William
Nomor Induk Mahasiswa : 00000042609
Program Studi : Sistem Informasi
Jenjang : S2/ S1 / D3
Judul Karya Ilmiah :

**IMPLEMENTASI MODEL SEGMENTASI PELANGGAN
BERDASARKAN TRANSAKSI PEMBELIAN PRODUK PERUSAHAAN
FASHION XYZ MENGGUNAKAN ALGORITMA CLUSTERING K-MEANS
DAN DBSCAN**

Menyatakan dengan sesungguhnya bahwa saya bersedia:

Memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Saya tidak bersedia, dikarenakan:

Dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 13 Mei 2024



(William)

* Jika tidak bisa membuktikan LoA jurnal/HKI selama 6 bulan kedepan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji syukur kepada Tuhan YME atas berkat dan karunia-Nya, penulis bisa menyelesaikan laporan skripsi yang berjudul “Implementasi Model Segmentasi Pelanggan Berdasarkan Transaksi Pembelian Produk Perusahaan *Fashion XYZ* Menggunakan Algoritma *Clustering K-Means* dan *DBSCAN*” dengan tepat waktu. Pembuatan skripsi ini diajukan sebagai salah satu prasyarat kelulusan untuk mencapai gelar Strata 1 pada Program Studi Sistem Informasi, Fakultas Teknik dan Informatika di Universitas Multimedia Nusantara.

Penulis ingin mengucapkan terima kasih terhadap seluruh keluarga Universitas Multimedia Nusantara untuk kesempatannya selama menimba ilmu 4 tahun ini. Selain itu, penulis juga ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Monika Evelin Johan, S.Kom., M.M.S.I., sebagai Pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan, dan motivasi atas terselesainya skripsi ini.
5. Bapak Johan Setiawan, S.Kom., MM., M.B.A., sebagai Pembimbing dan dosen mata kuliah IS-660 Information Systems Research Methodology, serta koordinator kelas *Masterclass* yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya skripsi ini.
6. Keluarga dan kerabat saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan skripsi ini.

Penulis menyadari bahwa laporan skripsi ini masih jauh dari kata sempurna dan tidak luput dari kesalahan, oleh karena itu penulis terbuka terhadap kritik dan saran

dari pihak lain. Semoga skripsi ini dapat bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi bagi para pembaca.

Tangerang, 13 Mei 2024



William



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

IMPLEMENTASI MODEL SEGMENTASI PELANGGAN BERDASARKAN TRANSAKSI PEMBELIAN PRODUK PERUSAHAAN *FASHION XYZ* MENGGUNAKAN ALGORITMA *CLUSTERING K-MEANS* DAN *DBSCAN*

William

ABSTRAK

Penggunaan *platform online* seperti *marketplace* semakin berkembang dan memiliki jumlah pengguna yang banyak di Indonesia. Industri *fashion* muncul sebagai salah satu bidang yang terus berkembang. Melihat keberagaman preferensi ini, diperlukan pendekatan yang optimal untuk menyusun strategi pemasaran yang efektif. Salah satu solusi yang dapat ditempuh adalah dengan memahami lebih mendalam perilaku konsumen melalui segmentasi pelanggan. Dengan pemahaman yang lebih mendalam terkait perilaku konsumen, perusahaan XYZ yang bergerak dalam industri *fashion* dapat meningkatkan efektivitas strategi pemasaran serta mampu merespons kebutuhan konsumen dengan lebih akurat agar meraih peningkatan penjualan yang signifikan.

Pembentukan model segmentasi pelanggan dalam penelitian ini dilakukan dengan menggunakan metode *clustering* dengan memanfaatkan algoritma *machine learning*. Penerapan algoritma *K-Means* dan *DBSCAN* menjadi metode utama yang dilakukan dalam penelitian ini dengan menggunakan acuan *Framework Data Mining* dalam melakukan proses pengolahan data. Hasil dari model segmentasi yang dibuat kemudian akan dilakukan metode evaluasi untuk mengukur seberapa baik hasil dari *clustering* yang dilakukan.

Hasil model segmentasi pelanggan dengan menggunakan algoritma *clustering K-Means* dan *DBSCAN* dibandingkan untuk menentukan algoritma mana yang menghasilkan hasil kluster terbaik. Hasil kluster terbaik yang digunakan untuk melakukan segmentasi pelanggan didapat pada algoritma *K-Means* dengan matriks *Silhouette Score* sebesar 0.549 dan skor *Davies-Bouldin Index* sebesar 0.593 dengan jumlah kluster k optimal = 7.

Kata kunci: *Clustering, DBSCAN, Fashion, K-Means, Segmentasi pelanggan*

**IMPLEMENTATION OF CUSTOMER SEGMENTATION
MODEL BASED ON XYZ FASHION COMPANY PRODUCT
PURCHASE TRANSACTIONS USING K-MEANS
CLUSTERING ALGORITHM AND DBSCAN**

William

ABSTRACT (English)

The use of online platforms such as marketplaces is growing and has a large number of users in Indonesia. The fashion industry is emerging as one of the growing fields. Seeing this diversity of preferences, an optimal approach is needed to develop an effective marketing strategy. One solution that can be taken is to understand consumer behavior more deeply through customer segmentation. With a deeper understanding of consumer behavior, XYZ company engaged in the fashion industry can improve the effectiveness of marketing strategies and be able to respond to consumer needs more accurately in order to achieve a significant increase in sales.

The formation of customer segmentation models in this study was carried out using clustering methods by utilizing machine learning algorithms. The application of the K-Means and DBSCAN algorithms is the main method carried out in this research using the Data Mining Framework reference in conducting data processing. The results of the segmentation model created will then be carried out an evaluation method to measure how good the results of the clustering are.

The results of the customer segmentation model using the K-Means and DBSCAN clustering algorithms are compared to determine which algorithm produces the best cluster results. The best cluster results used to segment customers are obtained in the K-Means algorithm with a Silhouette Score matrix of 0.549 and a Davies-Bouldin Index score of 0.593 with the optimal number of clusters $k = 7$.

Keywords: *Clustering, Customer Segmentation, DBSCAN, Fashion, K-Means*

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR	v
ABSTRAK	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	7
1.4.1 Tujuan Penelitian	7
1.4.2 Manfaat Penelitian	7
1.5 Sistematika Penulisan	8
BAB II LANDASAN TEORI	10
2.1 Penelitian Terdahulu	10
2.2 Tinjauan Teori	17
2.2.1 E-Commerce	17
2.2.2 Marketplace	18
2.2.3 Fashion	19
2.2.4 Segmentasi Pelanggan	20
2.3 Framework, Algoritma, dan Metode Evaluasi	22
2.3.1 Framework	22
2.3.2 Metode & Algoritma	25
2.3.2.1 K-Means Clustering	25

2.3.2.2	<i>DBSCAN</i>	27
2.3.2.3	<i>Standard Scaler</i>	28
2.3.2.4	<i>PCA Reduce Dimensionality</i>	29
2.3.3	Metode Evaluasi	30
2.3.3.1	<i>Silhouette Score</i>	30
2.3.3.2	<i>Davies-Bouldin Index (DBI)</i>	31
2.4	Tools yang digunakan	32
2.4.1	Jupyter Notebook	32
2.4.2	Visual Studio Code	33
2.4.3	Python	33
2.4.4	JavaScript	34
2.4.5	FastAPI	35
BAB III	METODOLOGI PENELITIAN	36
3.1	Gambaran Umum Objek Penelitian	36
3.2	Metode Penelitian	37
3.2.1	Alur Penelitian	38
3.2.2	Metode Data Mining	42
3.3	Teknik Pengumpulan Data	45
3.3.1	Populasi dan Sampel	45
3.3.2	Periode Pengambilan Data	46
3.4	Teknik Analisis Data	46
BAB IV	ANALISIS DAN HASIL PENELITIAN	49
4.1	<i>Business Understanding</i>	49
4.2	<i>Data Understanding</i>	51
4.2.1	<i>Data Cleaning</i>	60
4.2.2	<i>Exploratory Data Analysis</i>	62
4.3	<i>Data Preprocessing</i>	68
4.3.1	<i>Manipulate Data + Remove NaN Value</i>	68
4.3.2	<i>Label Encoding</i>	71
4.3.3	<i>Matrix Correlation (Heatmap)</i>	72
4.3.4	<i>Standard Scaler</i>	74
4.3.5	<i>Reduce Dimensionality (PCA Method)</i>	75

4.4	Modeling	76
4.4.1	<i>Spesifikasi</i>	76
4.4.2	<i>Mencari nilai Cluster K Optimal</i>	77
4.4.3	<i>K-Means Clustering</i>	79
4.4.4	<i>DBSCAN</i>	81
4.4.5	<i>Segmentasi</i>	84
4.5	Evaluation	96
4.5.1	<i>Silhouette Score</i>	97
4.5.2	<i>Davies-Bouldin Index</i>	99
4.6	Deployment	101
4.7	Analisis dan Pembahasan	109
BAB V	SIMPULAN DAN SARAN	113
5.1	Simpulan	113
5.2	Saran	113
DAFTAR PUSTAKA		115
LAMPIRAN		122

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan Penelitian Terdahulu	10
Tabel 3. 1 Tabel perbandingan framework	37
Tabel 3. 2 Tabel perbandingan algoritma	47
Tabel 3. 3 Tabel perbandingan tools yang akan digunakan	47
Tabel 4. 1 Jumlah Data Setiap Bulan	52
Tabel 4. 2 Deskripsi Kolom Dataframe	55
Tabel 4. 3 Segmentasi + Strategi Cluster 0	93
Tabel 4. 4 Segmentasi + Strategi Cluster 1	94
Tabel 4. 5 Segmentasi + Strategi Cluster 2	94
Tabel 4. 6 Segmentasi + Strategi Cluster 3	94
Tabel 4. 7 Segmentasi + Strategi Cluster 4	95
Tabel 4. 8 Segmentasi + Strategi Cluster 5	95
Tabel 4. 9 Segmentasi + Strategi Cluster 6	96
Tabel 4. 10 Tabel Hasil Skor Silhouette Score K-Means	98
Tabel 4. 11 Tabel Hasil Skor Silhouette Score DBSCAN	98
Tabel 4. 12 Tabel Hasil Skor DBI K-Means	100
Tabel 4. 13 Tabel Hasil Skor DBI DBSCAN	100
Tabel 4. 14 Tabel Perbandingan Metode Evaluasi dengan Penelitian Terdahulu	110

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR GAMBAR

Gambar 1. 1 Grafik Jumlah Traffic Marketplace di Indonesia	2
Gambar 2. 1 Sistem Arsitektur untuk melakukan Customer Segmentation.....	21
Gambar 2. 2 Framework CRISP-DM.....	22
Gambar 2. 3 DBSCAN Algorithm Point	27
Gambar 2. 4 Orthogonal Projection PCA	29
Gambar 3. 1 Alur Penelitian.....	39
Gambar 3. 2 Alur Proses Pengolahan Data (Data Understanding & EDA).....	40
Gambar 3. 3 Alur Proses Pengolahan Data (Data Preprocessing)	40
Gambar 3. 4 Alur Proses Pengolahan Data (Clustering).....	41
Gambar 3. 5 Alur Framework CRISP-DM	42
Gambar 4. 1 Gambaran Perusahaan XYZ.....	49
Gambar 4. 2 Zoom Meeting Dengan Owner.....	50
Gambar 4. 3 Tampilan Dashboard Shopee Seller Center	51
Gambar 4. 4 Tampilan Menu Bar "Pesanan Saya"	52
Gambar 4. 5 Dataset penjualan periode Agustus 2023 - Maret 2024	53
Gambar 4. 6 Dataset Final yang akan digunakan.....	53
Gambar 4. 7 Melakukan Import Library yang Akan Digunakan	54
Gambar 4. 8 Membaca Dataset	54
Gambar 4. 9 Sampel Dataset yang digunakan	55
Gambar 4. 10 Jumlah Rows dan Columns Dataset.....	55
Gambar 4. 11 Syntax Untuk mengetahui 5 baris teratas dataframe.....	57
Gambar 4. 12 Hasil dari syntax dataset_final.head()	57
Gambar 4. 13 Menampilkan Informasi tentang dataset	58
Gambar 4. 14 Menampilkan Statistik Deskriptif Kolom Numerik	59
Gambar 4. 15 Syntax Menghapus Kolom yang Rancu	60
Gambar 4. 16 Syntax mengisi nilai null dengan label baru	61
Gambar 4. 17 Output dari Gambar 4.16.....	61
Gambar 4. 18 Syntax mempersingkat isi kolom "Alasan Pembatalan".....	62
Gambar 4. 19 Hasil Visualisasi Jumlah Transaksi (Produk > 1 dan 1)	62
Gambar 4. 20 Visualisasi Bar Chart Alasan Pembatalan.....	63
Gambar 4. 21 Visualisasi Produk Terlaris	64
Gambar 4. 22 Visualisasi Word Cloud Produk Terlaris	65
Gambar 4. 23 Visualisasi Frekuensi Metode Pembayaran yang Dipilih	66
Gambar 4. 24 Visualisasi Top 10 Provinsi Berdasarkan Jumlah Pembelian	67
Gambar 4. 25 Visualisasi Top 10 Provinsi Berdasarkan Jumlah Pembelian	68
Gambar 4. 26 Print untuk mengetahui tipe data dari variabel.....	69
Gambar 4. 27 Melakukan konversi kolom object tertentu menjadi numerik.....	69
Gambar 4. 28 Output dari syntax gambar 4.27	70
Gambar 4. 29 Syntax Menghilangkan NaN Value.....	70
Gambar 4. 30 Menyimpan Data Frame Final Untuk digunakan di Algoritma Lain	71
Gambar 4. 31 Syntax Label Encoding Pada Kolom Kategorikal.....	72
Gambar 4. 32 Kolom Numerik Dataset Setelah di Encode.....	72

Gambar 4. 33 Syntax Kode Matrix Correlation (Heatmap).....	73
Gambar 4. 34 Heatmap Matrix Correlation	73
Gambar 4. 35 Syntax Kode Standard Scaler.....	74
Gambar 4. 36 Hasil Dataframe yang sudah distandarisasi.....	74
Gambar 4. 37 Syntax Kode PCA	76
Gambar 4. 38 Hasil Dataframe setelah implementasi PCA	76
Gambar 4. 39 Spesifikasi PC Yang Digunakan	76
Gambar 4. 40 Syntax Mencari Nilai Klaster K Optimal (1)	77
Gambar 4. 41 Gambar 4. 41 Syntax Mencari Nilai Klaster K Optimal (2)	78
Gambar 4. 42 Syntax Mencari Nilai Klaster K Optimal (3)	78
Gambar 4. 43 Jumlah Klaster K optimal K-Means Clustering	79
Gambar 4. 44 Syntax Algo K-Means	80
Gambar 4. 45 Hasil Clustering dengan menggunakan PCA dengan jumlah k=7 .	80
Gambar 4. 46 Syntax DBSCAN	81
Gambar 4. 47 Syntax DBSCAN (2).....	82
Gambar 4. 48 Visualisasi Clustering DBSCAN	82
Gambar 4. 49 Hasil Matriks Epsilon dan Min Pts terbaik	83
Gambar 4. 50 Hasil Matriks Epsilon dan Min Pts terbaik (2).....	84
Gambar 4. 51 Syntax untuk membuat menambahkan kolom "Cluster Label" Terhadap dataset.....	85
Gambar 4. 52 Hasil dataset setelah menambahkan kolom 'Cluster Label' (1).....	85
Gambar 4. 53 Hasil dataset setelah menambahkan kolom 'Cluster Label' (2).....	86
Gambar 4. 54 Hasil dataset setelah menambahkan kolom 'Cluster Label' (3).....	86
Gambar 4. 55 Hasil dataset setelah menambahkan kolom 'Cluster Label' (4).....	87
Gambar 4. 56 Frekuensi dari setiap cluster	87
Gambar 4. 57 Membuat Function untuk melakukan visualisasi karakteristik setiap cluster	88
Gambar 4. 58 Visualisasi Top 7 Frekuensi produk terlaris di masing-masing cluster.....	89
Gambar 4. 59 Visualisasi Top 7 Frekuensi total harga produk di masing-masing cluster.....	90
Gambar 4. 60 Visualisasi Top 7 Frekuensi demografis customer (provinsi) di masing-masing cluster.....	91
Gambar 4. 61 Visualisasi Top 7 Frekuensi metode pembayaran yang dipilih di masing-masing cluster.....	92
Gambar 4. 62 Visualisasi Silhouette Score K-Means Clustering	97
Gambar 4. 63 Visualisasi Silhouette Score DBSCAN.....	97
Gambar 4. 64 Visualisasi Indeks DBI K-Means Clustering	99
Gambar 4. 65 Visualisasi Indeks DBI DBSCAN	100
Gambar 4. 66 Syntax Kode main.py sebagai back-end	102
Gambar 4. 67 Kode Endpoint Backend POST untuk mengirim file input dari user untuk diproses (API Cluster).....	103
Gambar 4. 68 Kode Endpoint Backend POST untuk mengirim file input dari user untuk diproses(API dataviz).....	104
Gambar 4. 69 Fungsi akhir dari API /dataviz untuk melakukan visualisasi	105
Gambar 4. 70 Syntax kode front-end website (typescript)	105

Gambar 4. 71 Deklarasi State dan Melakukan Fetch Data dari API.....	106
Gambar 4. 72 Tampilan Website Sebelum user melakukan upload data.....	107
Gambar 4. 73 Tampilan Website Sebelum user melakukan upload data dan klik button proses	107
Gambar 4. 74 Tampilan Website Sebelum user melakukan upload data dan klik button proses (2).....	108
Gambar 4. 75 Tampilan Website Sebelum user melakukan upload data dan klik button proses (3).....	108

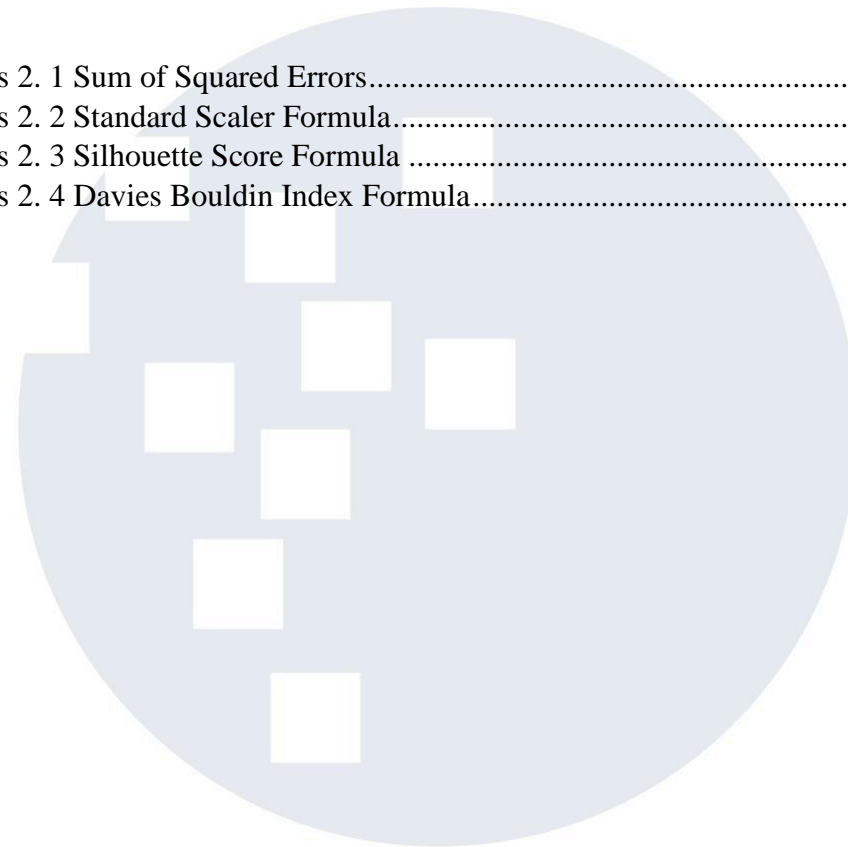


UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2. 1 Sum of Squared Errors.....	26
Rumus 2. 2 Standard Scaler Formula.....	28
Rumus 2. 3 Silhouette Score Formula	30
Rumus 2. 4 Davies Bouldin Index Formula.....	31

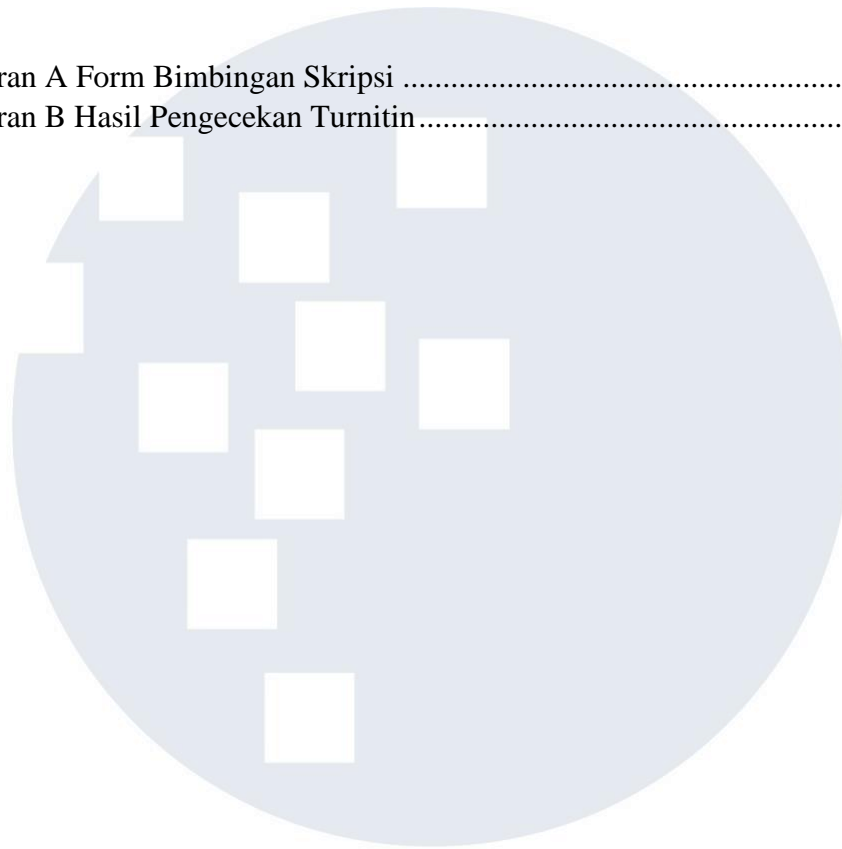


UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Form Bimbingan Skripsi	122
Lampiran B Hasil Pengecekan Turnitin.....	123



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA