

**EKSPLORASI POLA SEGMENTASI KONSUMEN *E-COMMERCE*
MENGUNAKAN INTEGRASI ANALISIS RFM DAN ALGORITMA
*CLUSTERING***

SKRIPSI

Oleh:

**NUR SA'DIAH NASUTION
NIM. 221121512**



**PROGRAM STUDI S-1 SISTEM INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

**EXPLORATION OF *E-COMMERCE* CUSTOMER SEGMENTATION
PATTERNS USING INTEGRATED RFM ANALYSIS AND
CLUSTERING ALGORITHMS**

UNDERGRADUATE THESIS

By:

**NUR SA'DIAH NASUTION
NIM. 221121512**



**UNDERGRADUATE PROGRAM OF S-1 INFORMATION SYSTEM
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

**EKSPLORASI POLA SEGMENTASI KONSUMEN *E-COMMERCE*
MENGUNAKAN INTEGRASI ANALISIS RFM DAN ALGORITMA
*CLUSTERING***

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Sistem Informasi

Oleh:

**NUR SA'DIAH NASUTION
NIM. 221121512**

Disetujui Oleh:

Dosen Pembimbing I,



Roni Yunis, S.Kom., M.T.

Dosen Pembimbing II,



Hanes, S.Kom., M.Kom.

Medan, 23 Juli 2026

Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-1 Sistem Informasi,



Tri Wulandari Ginting, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Sistem Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Nur Sa'diah Nasution

NIM : 221121512

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Eksplorasi Pola Segmentasi Pelanggan E-Commerce menggunakan Integrasi Analisis RFM dan Algoritma Clustering

Tempat Penelitian :

Alamat Tempat Penelitian :

No. Telp. Tempat Penelitian :

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 23 Juli 2026

Saya yang membuat pernyataan,



Nur Sa'diah Nasution

EKSPLORASI POLA SEGMENTASI KONSUMEN *E-COMMERCE* MENGGUNAKAN INTEGRASI ANALISIS RFM DAN ALGORITMA CLUSTERING

Abstrak

Segmentasi konsumen merupakan salah satu pendekatan yang digunakan untuk memahami karakteristik konsumen berdasarkan perilaku transaksi. Penelitian ini bertujuan untuk mengeksplorasi pola segmentasi konsumen e-commerce menggunakan integrasi analisis Recency, Frequency, Monetary (RFM) dengan algoritma clustering. Penelitian menggunakan metodologi Knowledge Discovery in Databases (KDD) yang meliputi tahapan selection, preprocessing, transformation, data mining, dan evaluation. Dataset yang digunakan berasal dari Kaggle dengan lebih dari 250.000 data transaksi konsumen e-commerce. Proses clustering dilakukan menggunakan algoritma K-Means, Gaussian Mixture Model (GMM), DBSCAN, dan Mean-Shift. Selanjutnya, hasil clustering dievaluasi menggunakan Silhouette Coefficient, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI), serta diuji stabilitasnya melalui Bootstrap Resampling. Hasil penelitian menunjukkan bahwa setiap algoritma menghasilkan karakteristik segmentasi yang berbeda. Pada dataset yang digunakan, K-Means memperoleh nilai evaluasi yang lebih tinggi dibandingkan algoritma lainnya dan menghasilkan tiga segmen konsumen, yaitu Champions, Potential Loyalists, dan Hibernating. Hasil penelitian ini memberikan gambaran mengenai penerapan analisis RFM dan algoritma clustering dalam segmentasi konsumen e-commerce.

Kata kunci: RFM, Segmentasi Konsumen, Clustering, E-Commerce

Abstract

Customer segmentation is one approach used to understand customer characteristics based on transaction behavior. This study aims to explore e-commerce customer segmentation patterns by integrating Recency, Frequency, Monetary (RFM) analysis and clustering algorithms. This study employs the Knowledge Discovery in Databases (KDD) methodology, which includes the stages of selection, preprocessing, transformation, data mining, and evaluation. The dataset used was sourced from Kaggle and contains more than 250,000 e-commerce customer transaction records. The clustering process was performed using the K-Means, Gaussian Mixture Model (GMM), DBSCAN, and Mean-Shift algorithms. Subsequently, the clustering results were evaluated using the Silhouette Coefficient, the Davies-Bouldin Index (DBI), and the Calinski-Harabasz Index (CHI), and their stability was tested via Bootstrap Resampling. The results show that each algorithm produces distinct segmentation characteristics. In the dataset used, K-Means achieved higher evaluation scores than the other algorithms and identified three customer segments: Champions, Potential Loyalists, and Hibernating. These findings provide insights into the application of RFM analysis and clustering algorithms in e-commerce customer segmentation.

Keywords: RFM, Customer Segmentation, Clustering, E-Commerce

KATA PENGANTAR

Ucapan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul Eksplorasi Pola Segmentasi Konsumen E-Commerce Menggunakan Integrasi Analisis RFM dan Algoritma Clustering dengan baik.

Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Sistem Informasi, Fakultas Informatika, Universitas Mikroskil. Penelitian ini bertujuan untuk mengeksplorasi pola segmentasi konsumen e-commerce melalui integrasi analisis RFM dengan beberapa algoritma clustering, sehingga diharapkan dapat memberikan gambaran mengenai karakteristik konsumen serta menjadi referensi dalam mendukung pengambilan keputusan berbasis data.

Ucapan terima kasih lainnya penulis sampaikan karena dalam proses penyusunan skripsi ini, penulis memperoleh banyak bantuan, bimbingan, dukungan, motivasi, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Roni Yunis, S.Kom., M.T., selaku Dosen Pembimbing I yang telah memberikan bimbingan, arahan, serta masukan kepada penulis selama proses penyusunan skripsi ini.
2. Bapak Hanes, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan bimbingan, saran, dan masukan kepada penulis selama proses penyusunan skripsi ini.
3. Bapak Hardy, S.Kom., M.SC., PH.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Ibu Tri Wulandari Ginting, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Sistem Informasi Fakultas Informatika Universitas Mikroskil Medan.
6. Kedua orang tua dan adik-adik tercinta yang selalu memberikan doa, dukungan, serta semangat kepada penulis selama masa perkuliahan hingga penyelesaian skripsi ini.
7. Penulis sendiri yang telah berupaya menyelesaikan penelitian dan penyusunan skripsi ini dengan penuh tanggung jawab.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun sebagai bahan perbaikan di masa mendatang.

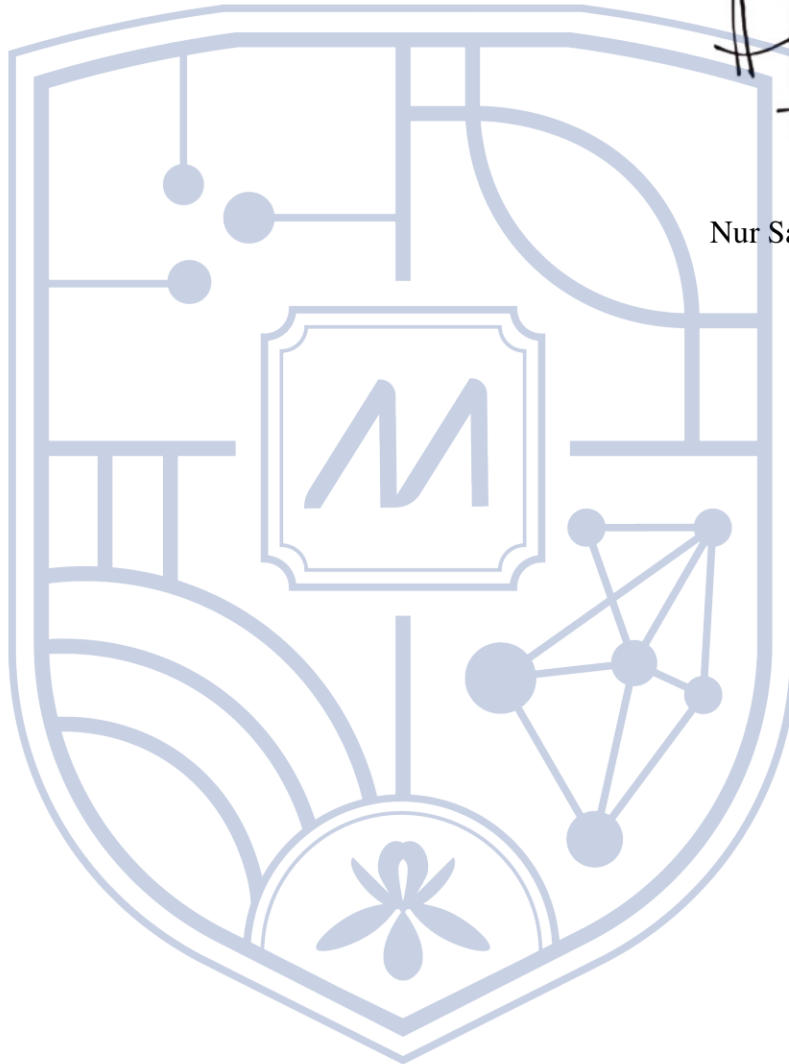
Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi referensi bagi penelitian selanjutnya, khususnya di bidang analisis data, data mining, dan segmentasi konsumen e-commerce.

Medan, 23 Juli 2026

Penulis,



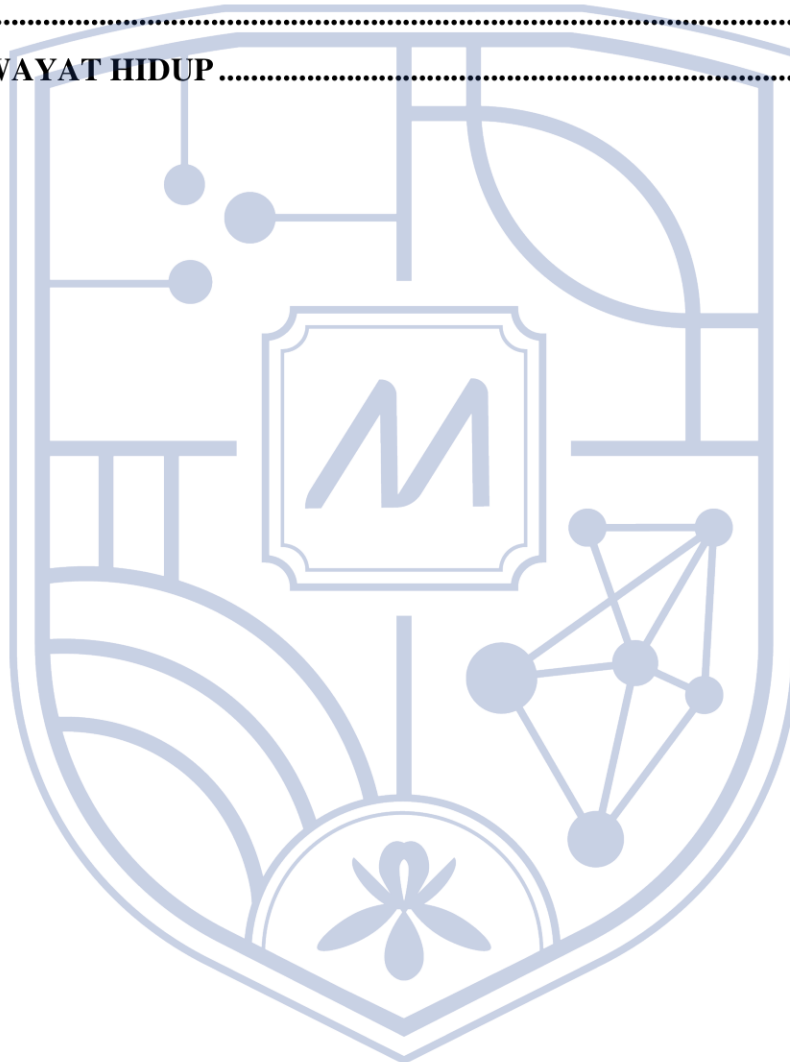
Nur Sa'diah Nasution



DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Manfaat.....	4
1.5 Ruang Lingkup.....	4
BAB II KAJIAN LITERATUR	5
2.1 Segmentasi Konsumen	5
2.2 <i>Knowledge Discovery in Database</i> (KDD).....	6
2.3 Model Analisis RFM (<i>Recency, Frequency, Monetary</i>).....	8
2.4 Algoritma <i>K-Means</i>	10
2.5 Algoritma DBSCAN	11
2.6 Algoritma <i>Gaussian Mixture Model</i> (GMM).....	13
2.7 Algoritma <i>Mean-Shift</i>	13
2.8 <i>Silhouette Score</i>	14
2.9 <i>Davies-Bouldin Index</i> (DBI)	15
2.10 <i>Calinski-Harabasz Index</i> (CHI)	16
2.11 Penelitian Terdahulu	17
BAB III METODOLOGI PENELITIAN	24
3.1 Pemilihan Data (<i>Data Selection</i>)	24
3.2 Pra-pemrosesan Data (<i>Preprocessing</i>)	25
3.3 Transformasi Data (<i>Transformation</i>)	46
3.4 Data Mining.....	54
3.5 Evaluasi (<i>Evaluation</i>).....	55

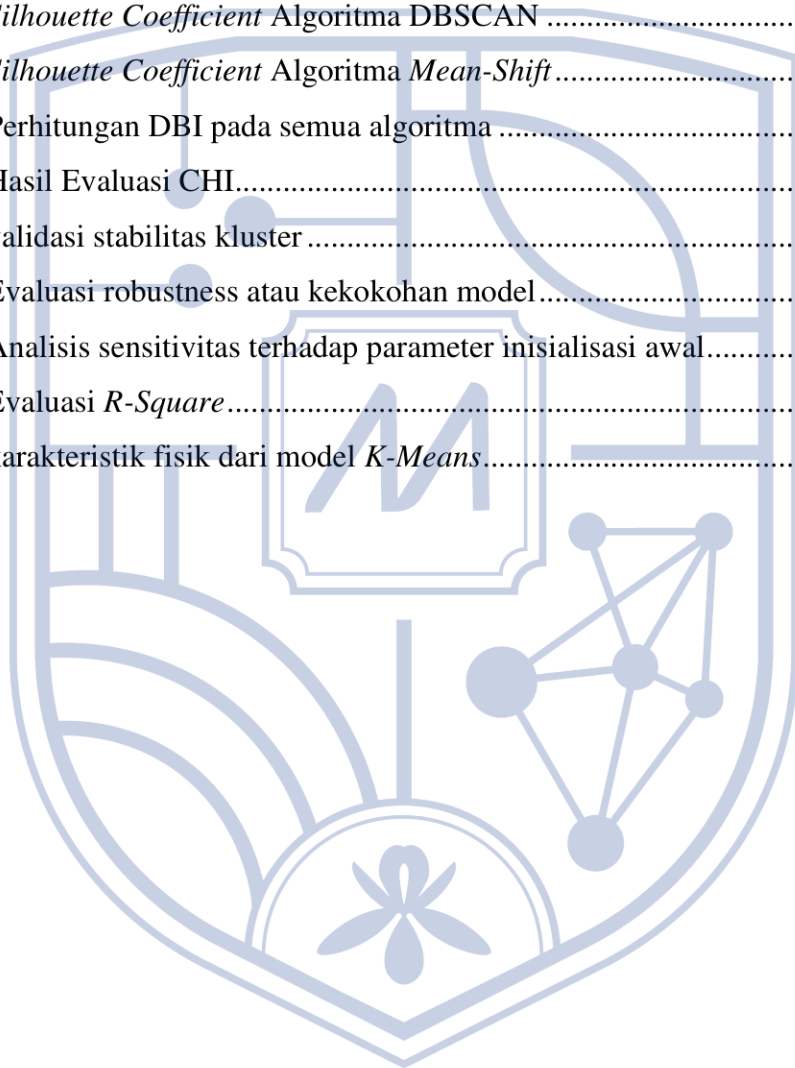
BAB IV HASIL DAN PEMBAHASAN	56
4.1 Hasil	56
4.2 Pembahasan.....	85
BAB V KESIMPULAN DAN SARAN.....	94
5.1 Kesimpulan.....	94
5.2 Saran.....	95
DAFTAR PUSTAKA.....	96
LAMPIRAN.....	101
DAFTAR RIWAYAT HIDUP	103



DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian.....	34
Gambar 3. 2 Pengecekan Struktur Data.....	36
Gambar 3. 3 Pengecekan Missing Value	37
Gambar 3. 4 Penanganan Missing Value	37
Gambar 3. 5 Pengecekan Duplikasi Data	38
Gambar 3. 6 Penyesuaian Tipe Data.....	39
Gambar 3. 7 Proses Pemeriksaan Variabel Duplikat	40
Gambar 3. 8 Penghapusan Variabel Duplikat.....	41
Gambar 3. 9 Analisis Statistik Deskriptif	42
Gambar 3. 10 Distribusi variabel numerik.....	43
Gambar 3. 11 Analisis Rentang Waktu Transaksi	45
Gambar 3. 12 Distribusi transaksi per tanggal.....	46
Gambar 3. 13 Outlier Detection.....	47
Gambar 3. 14 Visualisasi Outlier variabel Age	47
Gambar 3. 15 Visualisasi Outlier variabel Product Price	48
Gambar 3. 16 Visualisasi Outlier variabel Quantity	49
Gambar 3. 17 Visualisasi Outlier variabel Purchase Amount	50
Gambar 3. 18 Distribusi Variabel Age & TotalPurchaseAmount	51
Gambar 3. 19 Distribusi Variabel Kategorikal	52
Gambar 3. 20 Analisis Tren Waktu	53
Gambar 3. 21 Analisis Korelasi	55
Gambar 3. 22 Transformasi Data.....	57
Gambar 3. 23 Analisis Deskripsi data RFM.....	58
Gambar 3. 24 Distribusi Frekuensi Metrik RFM.....	60
Gambar 3. 25 Normalisasi Data.....	62
Gambar 3. 26 Distribusi Metrik RFM Setelah Normalisasi	63
Gambar 4. 1 Penentuan jumlah cluster dengan metode Elbow	66
Gambar 4. 2 Grafik Algoritma Berbasis Centroid	67
Gambar 4. 3 Penentuan parameter dengan Bayesian Information Criterion (BIC)	68
Gambar 4. 4 Ringkasan nilai BIC terbaik	69
Gambar 4. 5 K-Distance Plot untuk Penentuan Nilai Epsilon (ϵ) pada DBSCAN	70
Gambar 4. 6 Grafik Algoritma Berbasis Kepadatan (DBSCAN)	71

Gambar 4. 7 penentuan parameter optimal algoritma <i>Mean-Shift</i>	72
Gambar 4. 8 Hasil Visualisasi Algoritma <i>K-Means</i>	75
Gambar 4. 9 Hasil Visualisasi Algoritma GMM	77
Gambar 4. 10 Proses Implementasi Algoritma DBSCAN.....	79
Gambar 4. 11 Proses Implementasi Algoritma <i>Mean-Shift</i>	81
Gambar 4. 12 <i>Silhouette Coefficient</i> Algoritma <i>K-Means</i>	83
Gambar 4. 13 <i>Silhouette Coefficient</i> Algoritma GMM.....	84
Gambar 4. 14 <i>Silhouette Coefficient</i> Algoritma DBSCAN	85
Gambar 4. 15 <i>Silhouette Coefficient</i> Algoritma <i>Mean-Shift</i>	86
Gambar 4. 16 Perhitungan DBI pada semua algoritma	87
Gambar 4. 17 Hasil Evaluasi CHI.....	88
Gambar 4. 18 validasi stabilitas kluster	90
Gambar 4. 19 Evaluasi robustness atau kekokohan model.....	91
Gambar 4. 20 Analisis sensitivitas terhadap parameter inisialisasi awal.....	92
Gambar 4. 21 Evaluasi <i>R-Square</i>	93
Gambar 4. 22 karakteristik fisik dari model <i>K-Means</i>	94



DAFTAR TABEL

Tabel 2. 1 Literatur Review	25
Tabel 3. 1 Variabel Data Penelitian	35
Tabel 4. 1 Ringkasan Parameter Optimal Hasil Pengujian	73
Tabel 4. 2 Nilai Bobot Kontribusi (<i>Loading Scores</i>) Fitur RFM pada Komponen PCA.....	74
Tabel 4. 3 Rangkuman Hasil Evaluasi 4 Algoritma	89
Tabel 4. 4 Rentang Perbedaan 3 Cluster	95
Tabel 4. 5 Nilai Rata-Rata RFM dan Label Bisnis Kelompok Konsumen	99
Tabel 4. 6 Validasi Strategi dan Dampak Bisnis yang Diharapkan	102
Tabel 4. 7 Estimasi Nilai Transaksi Setiap Segmen	103
Tabel 4. 8 Simulasi Potensi Dampak Bisnis Strategi Segmentasi	104

