



**ANALISIS SEGMENTASI DAN KLASIFIKASI PENJUALAN
PRODUK *MARKETPLACE* MENGGUNAKAN METODE
K-MEANS CLUSTERING DAN *K-NEAREST NEIGHBOR*
(STUDI KASUS: *SHOPEE*)**

LAPORAN TUGAS AKHIR

BERTA FIRANI
41821110020

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025



**ANALISIS SEGMENTASI DAN KLASIFIKASI PENJUALAN
PRODUK *MARKETPLACE* MENGGUNAKAN METODE
K-MEANS CLUSTERING DAN *K-NEAREST NEIGHBOR***

(STUDI KASUS: *SHOPEE*)

LAPORAN TUGAS AKHIR

BERTA FIRANI

41821110020

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Berta Firani

NIM : 41821110020

Program Studi : Sistem Informasi

Judul Laporan Skripsi : Analisis Segmentasi dan Klasifikasi Penjualan Produk

*Marketplace Menggunakan Metode K-Means Clustering
dan K-Nearest Neighbor (Studi Kasus: Shopee)*

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

MERCU BUANA

Jakarta, 30 Juni 2025



Berta Firani

LEMBAR PENGESAHAN

Nama : Berta Firani
NIM : 41821110020
Judul Tugas Akhir : Analisis Segmentasi dan Klasifikasi Penjualan Produk
Marketplace Menggunakan Metode K-Means Clustering dan K-Nearest Neighbor (Studi Kasus: Shopee)

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Juli 2025

Menyetujui,

Pembimbing : Suhendra, S.Kom, M.Kom

NIDN : 0308019002

Ketua Penguji : Sarwati Rahayu, ST, MMSI

NIDN : 0321127602

Penguji 1 : Sulis Sandiwarno, S.Kom, M.Kom

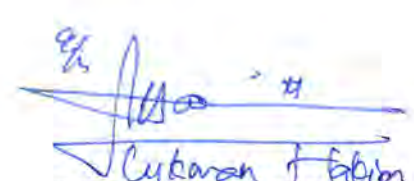
NIDN : 0302028803

Penguji 2 : Dr. Andi Nugroho, S.Kom., M.Kom

NIDN : 0305098303

Mengetahui,


Dr. Bambang Jokonowo, S.Si, M.T.I
Dekan Fakultas Ilmu Komputer


Ka.Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur kita panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan proposal penelitian yang berjudul "Analisis Segmentasi dan Klasifikasi Penjualan Produk *Marketplace* Menggunakan Metode *K-Means Clustering* dan *K-Nearest Neighbor* (Studi Kasus: *Shopee*)" ini dengan baik.

Penulis menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, tugas akhir ini tidak akan terselesaikan dengan baik. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis ingin menyampaikan penghargaan dan rasa terima kasih kepada:

1. Prof. Dr. Ir. Andi Adriansyah, M.Eng, selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S. Si, MTI, selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Lukman Hakim, S.T., MTI, selaku Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer.
4. Bapak Suhendra, S.Kom, M.Kom, selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Bapak/Ibu selaku Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Kedua orang tua dan orang terdekat yang selalu mendukung saya dalam proses studi

Akhir kata, saya berharap Allah SWT, Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 30 Juni 2025



Berta Firani

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Berta Firani

NIM : 41821110020

Program Studi : Sistem Informasi

Judul Laporan Skripsi : Analisis Segmentasi dan Klasifikasi Penjualan Produk

*Marketplace Menggunakan Metode K-Means Clustering
dan K-Nearest Neighbor (Studi Kasus: Shopee)*

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 30 Juni 2025



(Berta Firani)

ABSTRAK

Nama : Berta Firani
NIM : 41821110020
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Analisis Segmentasi dan Klasifikasi Penjualan Produk
Marketplace Menggunakan Metode *K-Means Clustering*
dan *K-Nearest Neighbor* (Studi Kasus: *Shopee*)

Kemajuan teknologi informasi telah mengubah pola transaksi jual beli dari metode tradisional ke arah bisnis daring melalui *platform marketplace* seperti *Shopee*. Namun, banyak penjual pemula menghadapi tantangan dalam menentukan produk yang potensial untuk dijual karena kurangnya wawasan berbasis data. Untuk mengatasi permasalahan ini, penelitian ini mengusulkan pendekatan yang mengombinasikan algoritma *K-Means Clustering* dan *K-Nearest Neighbor* (KNN) untuk segmentasi dan prediksi tingkat penjualan produk secara lebih terarah dan komprehensif.

Data produk di *Shopee* dikumpulkan melalui proses *web scraping*, dengan atribut meliputi nama produk, harga, kategori, kota, jumlah terjual, dan rating. Metode *K-Means* digunakan untuk mengelompokkan produk ke dalam beberapa kluster, kemudian diambil tiga kluster utama yang dilabeli sebagai "Sangat diminati", "Cukup diminati", dan "Kurang diminati", dan divalidasi menggunakan *Davies-Bouldin Index* (DBI). Nilai DBI terbaik ditemukan pada $k = 3$ untuk setiap kategori, seluruhnya menunjukkan nilai di bawah 1.0, yang menandakan struktur klasterisasi yang baik dan valid. Selanjutnya, algoritma KNN diaplikasikan pada produk dalam hasil klasterisasi untuk memprediksi tingkat penjualan ke dalam dua kategori: "Penjualan Tinggi" dan "Penjualan Rendah", dengan validasi menggunakan *Confusion Matrix*.

Hasil penelitian menunjukkan bahwa metode ini efektif dalam menghasilkan wawasan yang mendalam mengenai performa produk di *Shopee*. Model *K-Means* mampu membentuk kluster yang representatif terhadap pola minat konsumen, sementara KNN menunjukkan performa klasifikasi yang tinggi dengan akurasi hingga 96%, *precision* 96%, dan *recall* 100% pada kategori tertentu. Pendekatan ini tidak hanya memberikan dasar pengambilan keputusan berbasis data bagi penjual, tetapi juga memberikan kontribusi terhadap pengembangan literatur akademik di bidang analisis *marketplace* menggunakan teknik *data mining*.

Kata Kunci: *Shopee*, *Data Mining*, *K-Means Clustering*, *K-Nearest Neighbor*, Segmentasi Produk, Prediksi Penjualan

ABSTRACT

Name : Berta Firani
NIM : 41821110020
Study Program : Sistem Informasi
Title Thesis : Analisis Segmentasi dan Klasifikasi Penjualan Produk
Marketplace Menggunakan Metode K-Means Clustering dan K-Nearest Neighbor (Studi Kasus: Shopee)
Counsellor : Suhendra, S.Kom, M.Kom

Advances in information technology have changed the pattern of buying and selling transactions from traditional methods to online businesses through marketplace platforms such as Shopee. However, many novice sellers face challenges in determining potential products to sell due to the lack of data-based insights. To overcome this problem, this study proposes a approach that combines the K-Means Clustering and K-Nearest Neighbor (KNN) algorithms for segmentation and prediction of product sales levels in a more targeted and comprehensive manner. Product data on Shopee is collected through a web scraping process, with attributes including product name, price, category, city, number of sales, and rating. The K-Means method is used to group products into several clusters, then three main clusters are taken labeled as "Highly in demand", "Quite in demand", and "Less in demand", and validated using the Davies-Bouldin Index (DBI). The best DBI value is found at $k = 3$ for each category, all showing values below 1.0, indicating a good and valid clustering structure. Next, the KNN algorithm is applied to products in the result of cluster to predict sales levels into two categories: "High Sales" and "Low Sales", with validation using the Confusion Matrix. The results show that this approach is effective in generating in-depth insights into product performance on Shopee. The K-Means model is able to form clusters that are representative of consumer interest patterns, while KNN shows high classification performance with an accuracy of up to 96%, a precision of 96%, and a recall of 100% in certain categories. This approach not only provides a basis for data-based decision making for sellers, but also contributes to the development of academic literature in the field of marketplace analysis using data mining techniques.

Keywords: *Shopee, Data Mining, K-Means Clustering, K-Nearest Neighbor, Product Segmentation, Sales Prediction*

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS.....	v
AKHIR UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Permasalahan.....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Teori/Konsep Terkait.....	7
2.1.1 <i>K-Means Clustering</i>	7
2.1.2 <i>K-Nearest Neighbor</i>	7
2.1.3 <i>Web Scraping</i>	7
2.1.4 <i>Python</i>	8

2.1.5	<i>Preprocessing</i>	8
2.1.6	<i>Shopee</i>	8
2.1.7	<i>Sum Square Error (SSE)</i>	9
2.1.8	<i>Davies-Bouldin Index (DBI)</i>	9
2.1.9	<i>Confusion Matrix</i>	9
2.2	Penelitian Terdahulu.....	10
2.3	<i>Analisis Literatur Review</i>	32
BAB III METODE PENELITIAN.....		34
3.1	Deskripsi Sumber Data	34
3.2	Teknik Pengumpulan Data	34
3.3	Diagram Alir Penelitian.....	35
BAB IV HASIL PENELITIAN		39
4.1	<i>Data Collection</i>	39
4.2	<i>Preprocessing</i>	39
4.2.1	<i>Data Information</i>	39
4.2.2	<i>Data Cleaning</i>	40
4.2.3	<i>Data Transformation</i>	41
4.2.4	<i>Data Setelah Preprocessing</i>	43
4.3	<i>K-Means Clustering</i>	45
4.3.1	<i>K-Means Clustering</i> Grafik <i>Elbow</i> untuk Menentukan Nilai K.....	45
4.3.2	Pengelompokan Berdasarkan Jarak <i>Euclidean</i>	49
4.3.3	Pelabelan pada hasil Klastering	52
4.3.4	Hasil Segmentasi <i>K-Means Clustering</i>	53
4.4	Validasi <i>K-Means Clustering</i> dengan <i>Davies-Bouldin Index</i>	54
4.5	<i>Data Balance</i> dengan Teknik <i>Undersampling</i>	56
4.5.1	Proses <i>Undersampling</i> berdasarkan klaster	57

4.5.2	Definisi fitur dan target, serta <i>split data</i>	57
4.6	<i>K-Nearest Neighbor</i>	57
4.6.1	<i>Labeling Target</i>	58
4.6.2	Pemilihan Fitur dan Target.....	59
4.6.3	Pembagian Data (<i>Train Test Split</i>)	59
4.6.4	Pelatihan Model <i>K-Nearest Neighbor</i>	60
4.6.5	Hasil Klasifikasi <i>K Nearest Neighbor</i>	61
4.7	Pengujian Akurasi dengan <i>Counfusion Matrix</i>	65
4.7.1	Hasil <i>Confusion Matrix</i> pada klaster “Sangat Diminati”	67
4.7.2	Hasil <i>Confusion Matrix</i> pada klaster “Cukup Diminati”	67
4.7.3	Hasil <i>Confusion Matrix</i> pada klaster “Kurang Diminati”	68
4.8	Hasil <i>K-Means Clustering</i> dan <i>K-Nearest Neighbor</i>	68
4.9	<i>Tools Pemrograman</i>	74
4.9.1	<i>Python</i>	75
4.9.2	<i>Google Colab</i>	75
4.9.3	<i>Pandas</i>	75
4.9.4	<i>NumPy</i>	75
4.9.5	<i>Scikit-learn (sklearn)</i>	75
4.9.6	<i>Matplotlib dan Seaborn</i>	76
4.9.7	<i>Instant Data Scraper</i>	76
BAB V KESIMPULAN DAN SARAN.....		77
5.1	Kesimpulan	77
5.2	Saran.....	78
DAFTAR PUSTAKA.....		79
LAMPIRAN.....		82

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	10
<i>Tabel 3. 1 Data Jumlah Produk</i>	<i>34</i>
Tabel 4. 1 Atribut Data	39
Tabel 4. 3 Tabel Hasil Uji Nilai K Optimal.....	46
Tabel 4. 4 Hasil klasterisasi Perlengkapan Rumah dengan K-Means	51
Tabel 4. 5 Hasil klasterisasi Makanan dan Minuman dengan K-Means	51
Tabel 4. 6 Hasil klasterisasi Perawatan dan Kecantikan dengan K-Means.....	52
Tabel 4. 7 Atribut Data untuk proses KNN	58
Tabel 4. 8 Rata-Rata Performansi KNN Berdasarkan Nilai k.....	60
Tabel 4. 9 Produk dengan Prediksi Penjualan Tinggi Kategori Perlengkapan Rumah	62
Tabel 4. 10 Produk dengan Prediksi Penjualan Tinggi Kategori Makanan dan Minuman	63
Tabel 4. 11 Produk dengan Prediksi Penjualan Tinggi Kategori Perawatan dan Kecantikan.....	64
Tabel 4. 12 Validasi KNN pada klaster “Sangat Diminati”	67
Tabel 4. 13 Validasi KNN pada klaster “Cukup Diminati”.....	67
Tabel 4. 14 Validasi KNN pada klaster “Kurang Diminati”	68
Tabel 4. 15 Hasil K-Means dan KNN kategori Perlengkapan Rumah.....	69
Tabel 4. 16 Tabel Hasil Confusion Matrix kategori Perlengkapan Rumah.....	70
Tabel 4. 17 Hasil K-Means dan KNN kategori Makanan dan Minuman.....	71
Tabel 4. 18 Tabel Hasil <i>Confusion Matrix</i> kategori Makanan dan Minuman	72
Tabel 4. 19 Hasil <i>K-Means</i> dan KNN kategori Perawatan dan Kecantikan.....	73
Tabel 4. 20 Tabel Hasil <i>Confusion Matrix</i> kategori Perawatan dan Kecantikan...	73

DAFTAR GAMBAR

Gambar 1. 1 Marketplace dengan Pengunjung Terbanyak di Indonesia (Januari-Desember 2023)	1
Gambar 3. 1 Scraping data menggunakan Instant Data Scraper	35
Gambar 3. 2 Diagram Alir.....	36
Gambar 4. 1 Data Perlengkapan Rumah sebelum preprocessing	40
Gambar 4. 2 Data Makanan dan Minuman sebelum preprocessing.....	40
Gambar 4. 3 Data Perawatan dan Kecantikan sebelum preprocessing	40
Gambar 4. 4 Menghapus data duplikat	41
Gambar 4. 5 Mengisi nilai kosong dengan rata-rata	41
Gambar 4. 6 Mengisi nilai kosong dengan default "Tidak Diketahui"	42
Gambar 4. 7 Transformasi data pada kolom "Terjual"	42
Gambar 4. 8 Mengubah kolom "Harga" ke tipe data numerik.....	43
Gambar 4. 9 Data Perlengkapan Rumah setelah preprocessing.....	43
Gambar 4. 10 Data Makanan dan Minuman setelah preprocessing.....	44
Gambar 4. 11 Data Perawatan dan Kecantikan setelah preprocessing	44
Gambar 4. 12 Menentukan jumlah kluster optimal (nilai k)	45
Gambar 4. 13 Normalisasi menggunakan Standard Scaler.....	46
Gambar 4. 14 Menentukan jumlah kluster optimal dengan metode Grafik Elbow.....	47
Gambar 4. 15 Menghitung turunan kedua dari perubahan SSE.....	47
Gambar 4. 16 Grafik Elbow jumlah kluster optimal Perlengkapan Rumah.....	47
Gambar 4. 17 Grafik Elbow jumlah kluster optimal Makanan dan Minuman.....	48
Gambar 4. 18 Grafik Elbow jumlah kluster optimal Perawatan dan Kecantikan .	49
Gambar 4. 19 Penerapan model k-means.....	49
Gambar 4. 20 Menghitung centroid	50
Gambar 4. 21 Klasterisasi K-Means Berdasarkan Rating dan Jumlah Terjual	50
Gambar 4. 22 Scatter plot hasil klasterisasi dengan K-Means Perlengkapan Rumah	50
Gambar 4. 23 Scatter plot hasil klasterisasi dengan K-Means Makanan dan Minuman	51

Gambar 4. 24 Scatter plot hasil klasterisasi dengan K-Means Perawatan dan Kecantikan.....	52
Gambar 4. 25 Pemberian label pada klaster pada 3 klaster	53
Gambar 4. 26 Menyimpan hasil klasterisasi	53
Gambar 4. 27 Dataset hasil K-Means Clustering.....	54
Gambar 4. 28 Validasi model dengan DBI.....	54
Gambar 4. 29 Grafik validasi dengan DBI kategori Perlengkapan Rumah	55
Gambar 4. 30 Grafik validasi dengan DBI kategori Makanan dan Minuman	55
Gambar 4. 31 Grafik validasi dengan DBI kategori Perawatan dan Kecantikan..	56
Gambar 4. 32 Metode Undersampling	57
Gambar 4. 33 Definisi fitur dan target, serta split data	57
Gambar 4. 34 Penentuan label penjualan berdasarkan median.....	59
Gambar 4. 35 Pemilihan fitur dan variable dependen.....	59
Gambar 4. 36 Pembagian Data Latih dan Uji	60
Gambar 4. 37 Pelatihan Model K-Nearest Neighbor	61
Gambar 4. 38 Menambahkan hasil prediksi ke dataframe	61
Gambar 4. 39 Mengambil 3 produk dengan label “Penjualan Tinggi.....	62
Gambar 4. 40 Validasi KNN dengan Confusion matrix.....	65
Gambar 4. 41 Output evaluasi KNN	65
Gambar 4. 42 Grafik validasi KNN dengan Confusion matrix.....	66
Gambar 4. 43 Hasil K-Means dan KNN kategori Perlengkapan Rumah.....	69
Gambar 4. 44 Hasil K-Means dan KNN kategori Makanan dan Minuman.....	71
Gambar 4. 45 Hasil K-Means dan KNN kategori Perawatan dan Kecantikan	73

DAFTAR LAMPIRAN

Lampir 1 Bimbingan	82
Lampiran 2 Curriculum Vitae.....	83
Lampiran 3 Surat Pernyataan HKI.....	84
Lampiran 4 Sertifikat BNSP	86
Lampiran 5 Form Revisi 1	88
Lampiran 6 Form Revisi 2	89
Lampiran 7 Pernyataan Similarity Check	90

