SEGMENTASI WILAYAH PERTANIAN BERBASIS MACHINE LEARNING DENGAN METODE K-MEANS

SKRIPSI

Untuk Memenuhi Persyaratan Memperoleh Gelar Sarjana Pada Program Studi
Sistem Informasi Fakultas Sains dan Teknologi
Universitas Labuhanbatu

OLEH :

NADYA SURATIKA

2209100086

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS LABUHANBATU
RANTAUPRAPAT

2026

LEMBAR PENGESAHAN/PERSETUJUAN SKRIPSI

JUDUL	:	SEGMENTASI WILAYAH PERTANIAN BERBASIS MACHINE LEARNING DENGAN METODE K-MEANS
NAMA	:	NADYA SURATIKA
NPM	:	2209100086
PROGRAM STUDI	:	SISTEM INFORMASI

Disetujui pada tanggal: 13 Juli 2026

Pembimbing I	Pembimbing II
Dr. Ibnu Rasyid Munthe, S.Kom , M.Kom	Asriani Hasibuan, S Pd., M.Hum
NIDN. 113028702	NIDN. 0116078802

LEMBAR PENGESAHAN NASKAH SKRIPSI

JUDUL	:	SEGMENTASI WILAYAH PERTANIAN BERBASIS MACHINE LEARNING DENGAN METODE K-MEANS
NAMA	:	NADYA SURATIKA
NPM	:	2209100086
PROGRAM STUDI	:	SISTEM INFORMASI

Telah Diuji Dan Dinyatakan Lulus Dalam Ujian Sarjana
Pada Tanggal 13 Juli 2026

Penguji I (Ketua) **Tanda Tangan**

Nama : Dr. Ibnu Rasyid Munthe, S.Kom., M.Kom
NIDN : 0113028702

Penguji II (Anggota)
Nama : Asriani Hasibuan, S.Pd., M.Hum
NIDN : 0116078802

Penguji III (anggota)
Nama : Sudi Suryadi, S.Kom., M.Kom
NIDN : 0128017903

Rantauprapat, 13 Juli 2026

Dekan Fakultas Sains Dan Teknologi

(Assoc. Prof. Dr. Iwan Purnama, S.Kom., M.Kom)
NIDN. 0112029202

Ka. Program Studi Sistem Informasi

(Fadillah Hamdani, S.Kom., M.Kom)
NIDN. 0[illegible]7003

PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama	:	NADYA SURATIKA
NPM	:	2209100086
Judul Skripsi	:	Segmentasi Wilayah Pertanian Berbasis Machine Learning dengan Metode K-Means

Dengan ini penulis menyatakan bahwa skripsi ini disusun sebagai syarat untuk memperoleh gelar sarjana pada program studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Labuhanhatu adalah hasil karya tulis penulis sendiri. Semua kutipan maupun rujukan dalam penulisan skripsi ini telah penulis cantumkan sumbernya dengan benar sesuai dengan ketentuan yang berlaku.

Jika di kemudian hari ternyata ditemukan seluruh atau sebagian skripsi ini bukan hasil karya penulis atau plagiat, penulis bersedia menerima sanksi pencabatan gelar akademik yang disandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Rantauprapat, 13 Juli 2026
Yang Membuat Pernyataan,



NADYA SURATIKA

NPM. 2209100086

ABSTRAK

Segmentasi wilayah pertanian diperlukan untuk mengidentifikasi kelompok wilayah yang memiliki karakteristik produksi padi yang relatif serupa. Penelitian ini bertujuan mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan data luas panen, produktivitas, dan produksi padi tahun 2023. Data yang digunakan bersumber dari Badan Pusat Statistik dan terdiri atas 32 kabupaten/kota yang memiliki data lengkap. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan algoritma K-Means clustering. Jumlah cluster ditetapkan sebanyak tiga untuk membentuk kategori kapasitas produksi padi relatif rendah, menengah, dan tinggi. Perhitungan dilakukan menggunakan Euclidean Distance dengan centroid awal yang berasal dari Nias, Mandailing Natal, dan Tapanuli Utara, serta diverifikasi menggunakan Python pada Google Colab. Proses K-Means mencapai konvergensi pada iterasi keempat karena tidak terdapat lagi perpindahan anggota cluster. Hasil akhir menunjukkan bahwa Cluster 1 terdiri atas 18 wilayah dengan kapasitas produksi padi relatif rendah, Cluster 2 terdiri atas 12 wilayah dengan kapasitas produksi padi relatif menengah, dan Cluster 3 terdiri atas 2 wilayah dengan kapasitas produksi padi relatif tinggi, yaitu Deli Serdang dan Serdang Bedagai. Hasil segmentasi ini dapat digunakan sebagai informasi awal untuk mendukung penentuan prioritas pengembangan pertanian berdasarkan karakteristik kapasitas produksi setiap wilayah.

Kata kunci: segmentasi wilayah, pertanian, machine learning, K-Means clustering, produksi padi

ABSTRACT

Agricultural area segmentation is needed to identify groups of regions with relatively similar rice production characteristics. This study aims to group regencies and cities in North Sumatra Province based on harvested area, productivity, and rice production data for 2023. The data were obtained from Statistics Indonesia and consisted of 32 regencies and cities with complete records. This study employed a descriptive quantitative approach using the K-Means clustering algorithm. The number of clusters was set to three to represent relatively low, medium, and high rice-production capacity. The calculation used Euclidean Distance with initial centroids selected from Nias, Mandailing Natal, and North Tapanuli, and the results were verified using Python in Google Colab. The K-Means process converged in the fourth iteration because no region changed cluster membership. The final results show that Cluster 1 consists of 18 regions with relatively low rice-production capacity, Cluster 2 consists of 12 regions with medium rice-production capacity, and Cluster 3 consists of 2 regions with relatively high rice-production capacity, namely Deli Serdang and Serdang Bedagai. The segmentation results can serve as preliminary information to support agricultural development priorities according to the production capacity characteristics of each region.

Keywords: regional segmentation, agriculture, machine learning, K-Means clustering, rice production.

KATA PENGANTAR

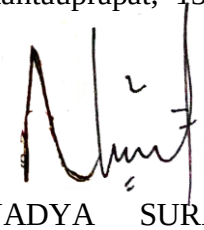
Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Segmentasi Wilayah Pertanian Berbasis Machine Learning dengan Metode K-Means". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu. Penulis menyampaikan terima kasih kepada Bapak Ibnu Rasyid Munthe, S.T., M.Kom. selaku dosen Pembimbing I dan Ibu Asriani Hasibuan, S.Pd., M.Hum. selaku dosen Pembimbing II yang telah memberikan arahan, bimbingan, dan motivasi selama proses penyusunan skripsi. Penulis juga menyampaikan terima kasih dan penghargaan kepada:

1. Dr. H. Amarullah Nasution, S.E., M.B.A., selaku Pendiri Yayasan Universitas Labuhanbatu.
2. Bapak Halomoan Nasution, S.H., M.H., selaku Ketua Yayasan Universitas Labuhanbatu.
3. Assoc. Prof. Ade Parlaungan Nasution, S.E., M.Si., Ph.D., selaku Rektor Universitas Labuhanbatu yang telah memberikan dukungan kepada mahasiswa.
4. Dr. Iwan Purnama, S.Kom., M.Kom., selaku Dekan Fakultas Sains dan Teknologi Universitas Labuhanbatu.
5. Budianto Bangun, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi yang telah memberikan pengarahan kepada mahasiswa.
6. Ibnu Rasyid Munthe, S.T., M.Kom., selaku dosen Pembimbing I yang telah memberikan bimbingan, motivasi, dan pengarahan selama penyusunan skripsi.
7. Asriani Hasibuan, S.Pd., M.Hum., selaku dosen Pembimbing II yang telah memberikan bimbingan, motivasi, dan pengarahan selama penyusunan skripsi.
8. Orang tua penulis yang selalu memberikan doa, semangat, dan dukungan.

9. Semua pihak yang telah membantu dalam penyusunan skripsi ini dan tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan penelitian ini. Semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya penerapan machine learning dalam analisis data pertanian, serta menjadi bahan pertimbangan bagi pihak yang berkepentingan.

Rantauprapat, 13 Juli 2026



NADYA SURATIKA

NPM: 2209100086

DAFTAR ISI

LEMBAR PENGESAHAN/PERSETUJUAN SKRIPSI i
LEMBAR PENGESAHAN NASKAH SKRIPSI ii
PERNYATAAN iii
ABSTRAK iv
ABSTRACT v
KATA PENGANTAR vi
DAFTAR ISI viii
BAB I PENDAHULUAN 1
1.1. Latar Belakang Masalah 1
1.2. Rumusan Masalah 3
1.3. Tujuan Penelitian 4
1.4 Hipotesis Penelitian 4
1.5. Manfaat Penelitian 5
BAB II LANDASAN TEORI 7
2.1 Teori–Teori dan Konsep K-Means pada Segmentasi Wilayah Pertanian 7
2.1.1 Teori Segmentasi Data dan Wilayah Pertanian 8
2.1.2 Peran *Unsupervised Learning* dalam Analisis Data Pertanian 8
2.1.3 Penerapan Algoritma *K-Means Clustering* dalam Penelitian Pertanian 9
2.1.4 Hubungan Segmentasi Wilayah dengan Pengelolaan Sumber Daya Pertanian 10
2.2 Algoritma yang Digunakan 11
2.2.2 Keunggulan dan Keterbatasan Algoritma *K-Means* 12
2.2.3 Penerapan *K-Means* dalam Analisis Data Pertanian 13
2.2.4 Faktor yang Mempengaruhi Kinerja Algoritma *K-Means* 14
2.2.6 Evaluasi Kualitas Klaster dengan *Elbow Method* dan *Silhouette Score* 15
2.3 Teknik Evaluasi Algoritma di *Machine Learning* 19
2.4 Langkah-Langkah *Machine Learning* dalam Segmentasi Wilayah Pertanian 22
2.4.1 Tahap Persiapan dan Praproses Data 22
2.4.2 Tahap Pemilihan dan Pelatihan Model 23
2.4.3 Tahap Evaluasi dan Validasi Model 23

2.4.4 Tahap Interpretasi dan Visualisasi Hasil 24
2.5 Alat Bantu Pemrograman dan *Tools* Pendukung 25
2.5.2 Keunggulan Penggunaan *Open-Source Tools* dalam Penelitian Pertanian27
2.5.3 Otomatisasi dan Skalabilitas Pemrosesan Data 27
2.6 Penelitian Terdahulu dan Kelebihan Penelitian Terdahulu 28
2.6.1 Kelebihan dan Inovasi Penelitian Ini 29
2.6.2 Ringkasan Penelitian Terdahulu 29
2.7 Flowchart Kerangka Penelitian 31
2.7.1 Struktur dan Komponen Flowchart Penelitian 33
BAB III METODOLOGI PENELITIAN 35
3.1 Metode Penelitian 35
3.2 Variabel Penelitian 36
3.2.1 Dataset yang Digunakan 37
3.3 Teknik Pengumpulan Data 38
3.3.1 Sumber Data 39
3.3.2 Prosedur Pengumpulan Data 40
3.3.3 Implikasi Metodologis Penggunaan Data Sekunder 43
3.4 Jenis Penelitian 45
3.4.1 Rancangan Penelitian Berdasarkan Kerangka Segmentasi Wilayah 45
3.4.2 Tujuan Operasional Penelitian 47
3.5 Perhitungan Rumus Algoritma dan Teknik Evaluasi 49
3.5.1 Persiapan Data dan Preprocessing 49
3.5.2 Langkah-Langkah Penerapan Algoritma K-Means 51
BAB IV HASIL DAN PEMBAHASAN 70
4.1 Selection 70
4.2 Preprocessing 72
4.3 Transformation 72
4.4 Data Mining 73
4.5 Interpretation/Evaluation 80
4.5.1 Interpretasi Karakteristik Cluster 80
4.5.2 Scatter Plot Hasil Clustering 82

4.5.3 Box Plot Karakteristik Cluster 83
4.5.4 Evaluasi Konvergensi Algoritma K-Means 85
4.5.5 Implikasi Hasil Segmentasi 86
4.5.6 Keterbatasan Evaluasi 86
4.6 Perbandingan Jumlah Cluster 87
4.6.1 Hasil Pengelompokan Menggunakan Empat Cluster 87
4.6.2 Hasil Pengelompokan Menggunakan Lima Cluster 89
4.6.3 Perbandingan Kualitas Tiga, Empat, dan Lima Cluster 91
4.6.4 Pertimbangan Pemilihan Tiga Cluster sebagai Model Utama 94
4.7 Evaluasi Jumlah Cluster Menggunakan Elbow Method dan Silhouette Score 95
4.7.1 Elbow Method 95
4.7.2 Silhouette Score 97
4.7.3 Perbandingan Metrik dan Dasar Pemilihan Model Utama 98
4.7.4 Kesimpulan Evaluasi 99
BAB IV KESIMPULAN DAN SARAN 101
5.1 Kesimpulan 101
5.2 Saran 103
DAFTAR PUSTAKA 105