

Digitasi Produktivitas Panen Padi Berbasis K-Means *Clustering*

Fitri Marisa¹, Abi Zahma², Adrianus Mui Bau³, Egy Noviansa⁴, Adi Semri Neno⁵, Anastasia Lidya Maukar⁶

^{1,2,3,4,5} Teknik Informatika, Universitas Widyagama Malang Jl. Borobudur No.35, Malang

⁶ Teknik Industri, President University, Jl. Ki Hajar Dewantara, Kota Jababeka, Cikarang Baru, Bekasi

email : fitrimarisa@gmail.com, zahma.abi@gmail.com, muidardy@gmail.com,
ofthedeathdarkness2@gmail.com, nenoady@gmail.com, almaukar@gmail.com

Abstract—Paddy as an ingredient of staple food by people in Indonesia includes in East Java Province. Therefore attention to the production of paddy in East Java is necessary, and this attention will give a piece of knowledge about which region produces paddy optimally or less optimal. This study aim is to do a clustering about paddy production in each region in East Java. K-Means algorithm uses to do clustering. The result is 3 clusters obtained, high, medium, and less productivity cluster. There are six regions in high productivity cluster, 20 regions in medium productivity cluster, and 12 regions in less productivity cluster.

Index Terms—Clustering; K-Means; Paddy; Cluster.

Abstrak—Padi merupakan bahan makanan pokok masyarakat di wilayah negara Indonesia termasuk di wilayah Provinsi Jawa Timur. Oleh karena itu hasil produksi padi perlu diperhatikan untuk mengetahui daerah yang memiliki produksi yang kurang optimal atau yang optimal. Pada penelitian ini bertujuan untuk melakukan *clustering* produksi padi di daerah Jawa Timur. *Clustering* dilakukan dengan algoritma *K-Means* sebanyak 5 iterasi yang menghasilkan 3 klaster yaitu klaster daerah produksi tinggi, daerah produksi sedang, dan daerah produksi rendah. Klaster daerah produksi tinggi terdapat 6 daerah, klaster produksi sedang terdapat 20 daerah, dan klaster produksi rendah terdapat 12 daerah.

Kata Kunci—Klastering; K-Means; Padi; Klaster.

I. PENDAHULUAN

Sandang, pangan, serta papan adalah 3 kebutuhan pokok manusia yang harus dipenuhi agar mempunyai kehidupan yang layak. Pangan sendiri merupakan salah satu makanan pokok yang dibutuhkan setiap manusia untuk bertahan hidup. Hasil olahan padi yang berupa beras merupakan sumber makanan pokok mayoritas warga Indonesia. Berdasarkan data yang dihimpun oleh *Food and Agriculture Organization of the United Nations* (FAO) yang dipresentasikan tahun 2018, Indonesia menjadi wilayah yang memproduksi beras terbanyak ketiga di dunia dengan total produksi sekitar 59,2 juta ton. Hasil panen ini meningkat dibandingkan dengan hasil produksi beras pada tahun 2014 sekitar 59 juta ton.[1]

Provinsi Jawa Timur terdiri dari banyak kota dan kabupaten yang mayoritas masyarakatnya mengkonsumsi nasi dari beras sebagai makanan pokok. Area atau lahan pertanian untuk menanam tanaman padi terdapat di semua daerah kota maupun kabupaten. Hasil keseluruhan panen padi di provinsi ini adalah 10,5 juta ton.[2]. Apabila dibandingkan dengan produksi nasional, maka Provinsi Jawa Timur menghasilkan kurang lebih seper-enam dari produksi padi nasional.

Guna mengatur kebutuhan beras di Provinsi Jawa Timur pemerintah perlu mengoptimalkan produksi padi di tiap kabupaten atau kota di Jawa Timur. Hasil panen atau produksi padi tiap kabupaten dan kota di Jawa Timur dapat digunakan sebagai dasar dari metode untuk melakukan pemetaan produksi padi. Langkah ini dilakukan bertujuan untuk mengetahui daerah kota atau kabupaten yang memiliki hasil produksi padi kurang maksimal. Sehingga nantinya pemerintah Provinsi Jawa Timur dapat melakukan langkah-langkah perbaikan terhadap kota atau kabupaten yang kurang optimal tersebut. Langkah-langkah perbaikan tersebut tentunya harus disertai dan didukung dengan data yang relevan dan juga pengetahuan yang tersedia.

Dalam teknologi informasi terdapat metode *Data Mining* yang diperuntukan guna memperoleh pengetahuan tersembunyi yang belum muncul sebelumnya dalam suatu *dataset* [3]. Dalam *Data Mining*

terdapat metode-metode yang digunakan untuk melakukan pencarian informasi sesuai dataset yang yaitu estimasi, prediksi, klasifikasi, klusterisasi, dan asosiasi. *Clustering* atau proses klusterisasi merupakan Teknik yang digunakan untuk proses pemetaan atau pengelompokkan dalam *Data Mining*. *Clustering* adalah Teknik yang digunakan untuk mengelompokkan data yang mirip satu sama lain ke dalam klaster, data yang terkelompok ini tidak memiliki kemiripan dengan klaster yang lain [4]. Dari beberapa algoritma dalam Teknik *clustering* salah satu yang populer dan sering digunakan adalah algoritma *K-Means*.

Algoritma yang sering digunakan dan populer sebagai teknik pembagi atau pengelompokkan pada teknik *clustering* yaitu Algoritma *K-Means*. Teknik klustering *K-Means*, mengelompokkan data sesuai dengan kedekatan data satu sama lain berdasarkan jarak *Euclidean*. *K-Means* mengambil k sebagai input parameter dan pembagi dari set obyek n dari klaster k . Nilai titik pertengahan dari obyek diambil sebagai parameter kesamaan untuk membentuk klaster. Titik pertengahan atau pusat klaster dibentuk dari pemilihan acak dari obyek k . Dengan membandingkan kemiripan terbanyak obyek akan dimasukkan ke klaster.

Pada penelitian yang dilakukan oleh Reza et al. mengenai penggunaan drone untuk menangkap gambar area tanam padi kemudian mengolah gambar tersebut dan mengelompokkannya menggunakan algoritma *K-Means*. Dari hasil penelitian tersebut didapat peta area pertumbuhan padi sesuai hasil pengelompokkan[5].

Kemudian penelitian selanjutnya Liyantono et al. mengenai estimasi panen padi dengan penggunaan drone untuk menangkap gambar menggunakan. Gambar yang diperoleh diolah kemudian dipetakan menggunakan algoritma *K-Means*. Hasil yang diperoleh adalah periode terbaik observasi sekitar 63 hari setelah tanam dengan nilai regresi 0,27 dan peta estimasi produksi padi.[6].

Berdasarkan hasil penelitian tersebut peneliti bermaksud untuk melakukan penelitian dengan memanfaatkan informasi dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur yang dipresentasikan tahun 2018 terkait produksi padi di wilayah Provinsi Jawa Timur baik di kabupaten maupun kota dengan menggunakan algoritma *K-Means*. Dari penelitian ini diharapkan ditemukan informasi tersembunyi mengenai hasil produksi padi Provinsi Jawa Timur tahun 2018.

II. METODOLOGI

Penelitian yang dijalankan ini memanfaatkan metode *Software Development Life Cycle* (SDLC) dan model yang dipakai adalah air terjun atau *waterfall*, diawali dari melakukan Analisa mengenai kebutuhan yang terdiri dari studi literatur dan pengumpulan data, desain rancangan, implementasi, hasil & pembahasan kemudian diakhiri dengan penarikan kesimpulan. Berikut diagram alur metodologi.



Gambar 1 Diagram alir metodologi

A. Analisis Kebutuhan

Analisis kebutuhan dilakukan saat awal *Software Development Life Cycle (SDLC)* dan menggunakan model air terjun atau *waterfall*. Analisa mengenai kebutuhan yang dilakukan terdiri dari :

1) Studi Literatur

Pencarian informasi terkait penerapan algoritma *K-Means* pada hasil produksi padi Provinsi Jawa Timur tahun 2018. Pencarian informasi dilakukan dari berbagai sumber misal artikel jurnal nasional maupun internasional, buku, portal berita, halaman website terkait dan sebagainya.

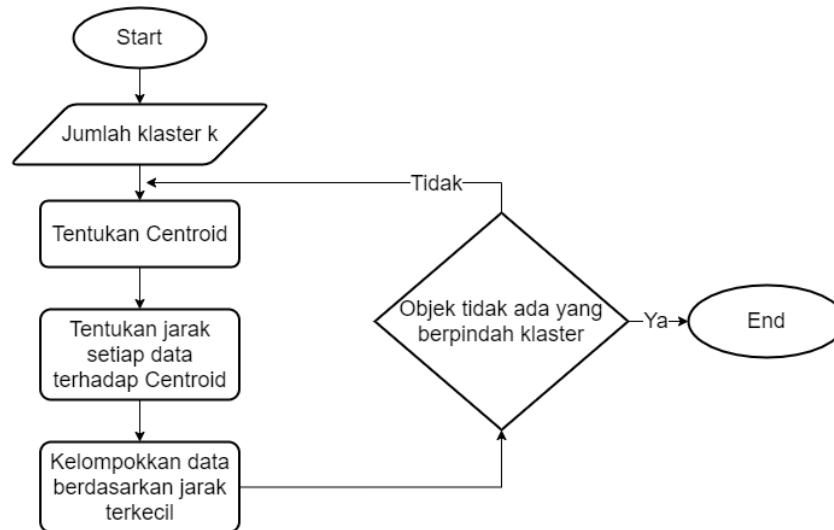
Pada penelitian sebelumnya telah dilakukan menggunakan metode *Clustering Fuzzy C-Means* untuk pemetaan potensi tanaman padi.[7]. Kemudian pemetaan palawija menggunakan algoritma *K-Means*. [8]. *K-Means* juga digunakan dalam pengelompokan prestasi santri di pondok pesantren.[9]. Selanjutnya penelitian mengenai pengelompokan pada variasi anggur untuk memprediksi hasil panen menggunakan *K-Means*. [10]. Pada bidang kesehatan *K-Means* digunakan untuk penelitian pada kluster data pasien [11] dan algoritma pendukung untuk proses prediksi pertumbuhan kanker payudara [12]. Selain itu algoritma *K-Means* juga digunakan untuk pengelompokan Riwayat transaksi penjualan beberapa produk pada HWI [13] dan juga untuk pengelompokan kinerja karyawan kontrak di Universitas Udayana [14].

2) Pengumpulan Data

Proses akumulasi data terkait hasil produksi panen padi di Provinsi Jawa Timur dikerjakan guna memperoleh dataset yang digunakan untuk proses *data mining*.

B. Desain

Data mining adalah satu set tata cara yang dijalankan guna memperoleh informasi anyar terkait suatu dataset yang sebelumnya tidak terlihat. Dalam proses memperoleh informasi anyar yang sebelumnya tidak terlihat, data mining melibatkan peran manusia dan komputer selain itu proses yang dilakukan untuk menemukan informasi tersebut juga bersifat iterative atau mengulang-ulang. Teknik klustering membagi data menjadi kluster-kluster tertentu berdasarkan kemiripan atribut dari semua data dalam dataset. Karakteristik tiap cluster tidak ditentukan sebelumnya melainkan terbagi berdasarkan atribut data-data yang dikelompokkan dalam kluster tersebut. Algoritma *K-Means* akan menentukan setiap data masuk ke dalam kluster tertentu sehingga tidak ada data diluar kluster. Pada setiap iterasi memungkinkan data berpindah dari satu kluster ke kluster lainnya. Berikut diagram alir algoritma *K-Means*.

Gambar 2 Diagram alir algoritma *K-Means*

1) Menentukan Jumlah Kluster

Langkah pertama yaitu menentukan jumlah kluster untuk mengelompokkan dataset. Penelitian ini menggunakan 3 kluster yang terdiri dari C1, C2, dan C3.

2) Menentukan Centroid

Setelah menentukan jumlah dan kategori kluster tahapan berikutnya yang dilakukan adalah menetapkan titik sentral awal yang digunakan untuk setiap kluster. Titik sentral awal ditetapkan secara acak atau bebas. Pada penelitian ini *centroid* yang digunakan adalah C1 adalah (65229,377367), C2 adalah (20863,104712), dan C3 adalah (35230,219252).

3) Menentukan Jarak Data ke Centroid

Setelah menetapkan titik sentral atau pusat bagi setiap kluster, tahapan berikutnya yang dikerjakan pada proses klustering atau pengelompokan data adalah melakukan perhitungan jarak antara titik sentral setiap kluster dengan setiap obyek data dalam dataset. Perhitungan jarak tersebut dikerjakan dengan menggunakan rumus *K-Means* sebagai berikut.

$$d_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

4) Mengelompokkan Data Berdasarkan Jarak Minimum

Kemudian akan diperoleh hasil berupa jarak antara titik sentral dari setiap kluster dengan obyek data sesudah perhitungan dengan rumus *K-Means*. Kemudian objek data akan dimasukkan ke salah satu kluster berdasarkan jarak minimum atau paling kecil.

C. Implementasi

Pada bagian implementasi dilakukan penghitungan jarak tiap titik pusat masing-masing kluster dengan objek data yang terdapat dalam dataset. Sesuai rumus sebelumnya x adalah dataset produksi padi provinsi Jawa Timur pada tahun 2018, dataset yang digunakan adalah dataset bagian kolom luas panen dan produksi. Sedangkan y merupakan titik pusat dari tiap kluster yang telah ditentukan sebelumnya.

Pada iterasi-1 digunakan centroid yang telah ditentukan diawal yaitu C1, C2, dan C3 untuk menghitung jarak dari tiap data. Setelah ketemu jarak dari tiap data maka dikelompokkan berdasarkan jarak minimal terhadap masing-masing centroid. Kemudian iterasi-2 menggunakan centroid nilai rata-rata dari tiap kluster untuk menghitung jarak. Hasil akan dikelompokkan lagi berdasarkan jarak minimal data terhadap centroid.

Langkah tersebut diulang terus pada setiap data dalam dataset produksi pertanian provinsi Jawa Timur tahun 2018 sampai data stabil atau tidak ada perpindahan kluster lagi dari setiap data.

III. HASIL DAN PEMBAHASAN

A. Pembahasan

Pada proses analisis proses mining menggunakan data produksi padi di Provinsi Jawa Timur tahun 2018 terdapat 38 data yang dilakukan proses penghitungan jarak ke tiap titik pusat kluster. Kemudian dilakukan perhitungan nilai jumlah total, rata-rata, jumlah data, min, dan max dari setiap kluster.

B. Hasil

Hasil penelitian klustering menggunakan algoritma *K-Means* pada produksi padi di Jawa Timur tahun 2018 adalah sebagai berikut. Penghitungan ini dilakukan dengan memasukkan rumus algoritma *K-Means* secara manual kedalam *Excel* kemudian menghitungnya sesuai data.

Pada proses penentuan kluster pertama ditentukan titik pusat atau *centroid* dari 3 kluster seperti yang telah ditentukan di bagian desain. Pada iterasi pertama menghasilkan C1 terdapat 15 data, C2 terdapat 12 data, dan C3 terdapat 11 data. Dari data yang dikelompokkan dalam tiap kluster diambil nilai rata-ratanya, nilai rata-rata ini digunakan sebagai titik pusat baru untuk iterasi berikutnya. Berikut data titik pusat dan jumlah data pada kluster untuk tiap iterasi.

Table 1 Tabel hasil iterasi *K-Means*

Iterasi	Kluster 1		Kluster 2		Kluster 3	
	Titik pusat	Jumlah Data	Titik pusat	Jumlah Data	Titik pusat	Jumlah Data
1	65229; 377367	15	20863; 104712	12	35230; 219252	11
2	86573,06; 507624,73	10	6511,33; 33578,75	12	41088; 229146	16
3	101866,7; 595709,4	7	6511,33; 33578,75	12	45743,56; 261117,68	19
4	116103,42; 678588,28	6	6511,33; 33578,75	12	49360; 283413,63	20
5	123682,16; 718519,33	6	6511,33; 33578,75	12	50423,55; 291193,05	20

Hasil iterasi dari tabel diatas menunjukkan bahwa diperlukan 5 iterasi agar data menunjukkan sudah stabil. Iterasi ke 5 menunjukkan hasil seperti pada iterasi ke 4 dimana data tidak ada yang berpindah kluster lagi. Sesuai algoritma *K-Means* apabila tidak ada data yang berpindah maka proses iterasi berhenti dan proses klustering dinyatakan telah selesai sehingga data yang dihasilkan dari iterasi ke 5 merupakan data yang sudah stabil.

Dari hasil klustering menggunakan algoritma *K-Means* terhadap hasil produksi tanaman padi di Provinsi Jawa Timur dengan variabel luas lahan panen (ha) dan produksi (ton) didapatkan hasil sebagai berikut :

Table 2 Data hasil klustering iterasi ke 4

Data nilai	Kluster 1		Kluster 2		Kluster 3	
	Lahan Produksi	Hasil Panen	Lahan Produksi	Hasil Panen	Lahan Produksi	Hasil Panen
Total	742.093,00	4.311.116,00	78.136,00	402.945,00	1.008.471,00	5.823.861,00
Rata-rata	123.682,17	718.519,33	6.511,33	33.578,75	50.423,55	291.193,05
Jumlah	6	6	12	12	20	20
Rasio	1	1	2	2	3,33	3,33
Nilai Maks	151.884	924.212	22.883	118.139	72.604	439.002
Nilai Min	84891	532815	735	4903	33291	179915

Dari data diatas yang diperoleh dari iterasi ke 5 dapat diambil data bahwa pada kluster 1 terdapat 6 daerah dengan produksi tinggi. Selanjutnya adalah kluster 3 terdapat 20 daerah dengan produksi sedang. Kemudian kluster 2 terdapat 12 daerah dengan produksi rendah.

IV. KESIMPULAN

Berdasarkan hasil iterasi *K-Means* dihasilkan 5 tahap iterasi. Pada iterasi ke 5 proses iterasi berhenti dikarenakan rasio data dalam tiap kluster yang dihasilkan sama dengan rasio data pada iterasi ke 4 sehingga kluster yang dihasilkan dari iterasi ke 5 merupakan pembagian kluster yang paling representative dimana hasilnya adalah kluster 1 terdapat 6 daerah dengan produksi tinggi, kluster 2 terdapat 12 daerah dengan produksi rendah, dan kluster 3 terdapat 20 daerah dengan produksi sedang.

Dari permasalahan diatas terkait klustering produksi padi di Provinsi Jawa Timur yang dihimpun pada 2018 dapat diselesaikan dengan menggunakan algoritma *K-Means*. Pada hasil akhir data terbagi dengan

sempurna dimana tidak ada data yang saling tumpang antar kluster. Dari hasil tersebut juga dapat diketahui jumlah total, rata-rata, rasio, nilai maksimal, nilai minimal, dan jumlah data pada setiap kluster.

V. DAFTAR PUSTAKA

- [1] "FAOSTAT," 2019. <http://www.fao.org/faostat/en/#data/QC/visualize> (accessed Jan. 12, 2021).
- [2] "BPS Provinsi Jawa Timur," 2019. <https://jatim.bps.go.id/statictable/2019/10/08/1583/luas-panen-produksi-dan-produktivitas-padi-di-provinsi-jawa-timur-menurut-kabupaten-kota-ha-2018.html> (accessed Jan. 12, 2021).
- [3] J. G. Soumen Chakrabarti, Martin Ester, Usama Fayyad and W. W. Jiawei Han, Shinichi Morishita, Gregory Piatetsky-Shapiro, "Data Mining Curriculum: A Proposal (Version 1.0)," vol. 1, 2006, Accessed: Jan. 12, 2021. [Online]. Available: <https://www.kdd.org/curriculum/index.html>.
- [4] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham: Morgan kaufman Publisher, 2012.
- [5] M. N. Reza, I. S. Na, S. W. Baek, and K. H. Lee, "Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images," *Biosyst. Eng.*, vol. 177, no. 2018, pp. 109–121, 2019, doi: 10.1016/j.biosystemseng.2018.09.014.
- [6] Liyantono, Y. Almadani, Y. Adillah, M. Maulana Yusuf, M. N. Reza Mahbub, and A. Fatikhunnada, "Analysis of Paddy Productivity Using NDVI and K-means Clustering in Cibarusah Jaya, Bekasi Regency," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 557, no. 1, 2019, doi: 10.1088/1757-899X/557/1/012085.
- [7] Winarni, "Penerapan Metode Clustering Fuzzy C-Means Menggunakan Matlab Untuk Memetakan Potensi Tanaman Padi Di Kabupaten Bekasi," *J. Sist. Inf. Manaj. Basis Data*, vol. 01, no. 02, pp. 116–127, 2018.
- [8] B. R. JURISTRA, "PEMETAAN HASIL CLUSTERING PRODUKTIVITAS PADI DAN PALAWIJA DI PULAU JAWA MENGGUNAKAN ALGORITMA K-MEANS," Solo, 2017. [Online]. Available: <https://eprints.uns.ac.id/id/eprint/33524>.
- [9] H. Yuwafi, F. Marisa, and I. D. Wijaya, "Implementasi Data Mining Untuk Menentukan Santri Berprestasi Di Pp . Manaarullhuda Dengan Metode," *J. SPIRIT*, vol. 11, no. 1, pp. 22–29, 2019.
- [10] A. Aquino, B. Millan, M. P. Diago, and J. Tardaguila, "Automated early yield prediction in vineyards from on-the-go image acquisition," *Comput. Electron. Agric.*, vol. 144, no. March 2017, pp. 26–36, 2018, doi: 10.1016/j.compag.2017.11.026.
- [11] N. P. Dharshinni, F. Azmi, I. Fawwaz, A. M. Husein, and S. D. Siregar, "Analysis of Accuracy K-Means and Apriori Algorithms for Patient Data Clusters," *J. Phys. Conf. Ser.*, vol. 1230, no. 1, 2019, doi: 10.1088/1742-6596/1230/1/012020.
- [12] A. Jamal, A. Handayani, A. A. Septiandri, E. Ripmiatin, and Y. Effendi, "Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 9, no. 3, p. 192, 2018, doi: 10.24843/lkjiti.2018.v09.i03.p08.
- [13] R. Risnawati and Rohminatin, "K-MEANS CLUSTERING HWI PRODUCTS (Case Study: HWI Kisaran Distributor)," vol. 4509, pp. 1–7, 2020.
- [14] D. Ardiada, P. A. Ariawan, and M. Sudarma, "Evaluation of Supporting Work Quality Using K-Means Algorithm," *IJEET Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 3–6, 2018.

Fitri Marisa, Meraih gelar sarjana komputer (S.Kom) dari STMIK PPKIA Pradnya Paramita pada tahun 2002. Kemudian meraih gelar Master (M.Pd) dari Universitas Negeri Malang pada tahun 2006. Saat ini Penulis menjadi dosen program studi Teknik Informatika di Universitas Widya Gama.

Anastasia Lidya Maukar, Meraih gelar sarjana Universitas Negeri Surabaya pada tahun 1997. Kemudian meraih gelar Master (M.Sc) dari University Of Hertfordshire pada tahun 2001. Kemudian meraih gelar Master (M.MT) dari Institut Teknologi Sepuluh Nopember pada tahun 2004. Saat ini Penulis menjadi dosen program studi Teknik Industri di Universitas Presiden.