

PENGELOMPOKAN PRODUKSI TAMBAK GARAM DENGAN METODE CLUSTER K-MEANS DAN OPTIMASI CLUSTER MENGGUNAKAN ELBOW (STUDI KASUS: DINAS KELAUTAN KABUPATEN BANGKALAN)

Ach Muqoddam¹

¹Universitas Trunojoyo Madura

¹ach.adam04@gmail.com

ABSTRACT

Salt ponds in Madura are known by the Indonesian people as salt producers or it can be said that Madura is generally called a salt island, especially salt ponds in Bangkalan Regency in 2014 salt production increased from the previous year. But not long after that in 2020 and 2021 Salt Pond production decreased. The problem that occurs in the field of salt pond business in Kwanyar District, Bangkalan Regency is to analyze the problems that exist in salt ponds regarding the characteristics and potential of the salt pond business and the salt pond production system. And the problems that exist in the Maritime Service of Bangkalan Regency are difficult to classify salt pond production data. Responding to the problem of salt ponds in Bangkalan, a data mining system is needed to classify salt pond production. The grouping of salt ponds is grouped into 3 clusters. Where 3 groups (clusters) are used here, namely the need to increase production (cluster 1), it is sufficient to increase production (cluster 2), there is no need to increase production (cluster 3), which is useful for evaluating and providing solutions. The method used is the K-Means Clustering method, as well as finding an optimal group using Elbow Optimization. The results of the K-Means process of grouping the results obtained are group 1 = 65 data that need to increase production, group 2 = 55 data that simply needs to increase production, and group 3 = 21 data that doesn't need to increase production. Elbow Optimization Test to find the most optimal K, by conducting trials 7 clusters. The result is K = 3 which is optimal from testing comparisons with other Ks, looking at the SSE values and The resulting Elbow Garp.

Keywords: Salt Ponds, Clustering, K-Means Algorithm, Elbow Optimization

I. PENDAHULUAN

Tambak garam yang dikelola rakyat Indonesia berukuran kecil, dan terdapat beberapa contoh yang ada di tambak garam di Kabupaten Bangkalan, Sampang, Pamekasan, Sumenep, di Pulau Madura. dikelola lahan per Kecamatan yang ada di lima Kabupaten berkisar antara 0,3 sampai 1,7 hektar per orang. Pengolahan tambak garam di Pulau Madura mempunyai berbagai ukuran mulai dari 0,4 hingga 1,7 hektar per orang. Data yang di publikasikan Kementerian Kelautan dan Perikanan (KKP) bahwasannya kuantitas dari produksi sentra-sentra garam dari tahun 2017 sampai 2018 menunjukkan peningkatan [1]. Tambak Garam di Madura sudah dikenal oleh masyarakat Indonesia sebagai penghasil Garam atau bisa dikatakan Madura itu secara umum disebut pulau garam, terutama Tambak Garam di Kabupaten Bangkalan ditahun 2014 produksi Garam meningkat dari tahun sebelumnya dengan jumlah produksi 8.000 ton [2]. Tetapi tidak lama kemudian di tahun 2020 dan 2021 produksi Tambak Garam tidak optimal mencapai jumlah 2.000 ribu ton, sedangkan target tahunan seharusnya 4.000 ribu ton per tahun, penyebabnya adalah faktor cuaca yang tidak mendukung sehingga menggagalkan hasil panen tambak garam dan juga tambak garam di Kabupaten Bangkalan tidak adanya perlengkapan rumah produksi garam, akibatnya hasil tambak garam dijual keluar daerah [3]. Menanggapi persoalan tambak garam di Bangkalan, maka seharusnya ada evaluasi dan solusi bagi petani Tambak

Garam di Bangkalan untuk dilakukan pemberdayaan dalam meningkatkan kesejahteraan ekonomi masyarakat, Maka dari itu dibutuhkan suatu clusterisasi yang mempartisi data ke dalam cluster atau kelompok sehingga data yang memiliki karakteristik yang mirip dikelompokkan ke dalam satu cluster yang sama dan data yang memiliki perbedaan dikelompokkan ke dalam kelompok yang lain [4]. Alasan memilih metode K-Means dikarenakan metode tersebut banyak digunakan sebagai teknik pengelompokan dengan berbagai macam jenis data [5].

Berdasarkan beberapa hal yang sudah di jelaskan tentunya memunculkan permasalahan yang ada, maka penulis ingin melakukan pengelempokan Tambak Garam yang ada di Kabupaten Bangkalan dengan 3 kelompok diantaranya tidak perlu meningkatkan produksi, cukup perlu meningkatkan produksi, perlu meningkatkan produksi. Pembuatan sistem data mining menggunakan metode Algoritma K-Means.

II. LANDASAN TEORI

A. DATA MINING

Suatu proses yang berguna sebagai pembelajaran pada sekumpulan data. Sebuah pembelajaran yang didapatkan tersebut berguna untuk berbagai bidang, seperti dilingkup bisnis, pendidikan, kesehatan, dan lainnya. Data mining

jugaidiartikan untuk suatu prosesi yang menggunakan teknik statistiki, matematik, artificial intelegent, dan mechine learning yang berguna untuk mengenali atau mencari informasi [6].

B. CLUSTERING

Clustering digunakan sebagai salah satu proses data mining yang berkarakter tanpaarahan (unsupervised), metode tersebut diimplementasikan tanpai adanyaiilatihan (training). Dalam data mining ada beberapa macam tipe mengklasterkan yang didalam mengelompokan data, yaitu: [7]

- a. Hierarchical Clustering adalah metode yang digunakan sebagai pengelompokan data yang di mulai dengan mengelompokan dua atau lebih objek yang memiliki kemiripan yang terdekat
- b. Partitioning Algorithms adalah dalam kelompok pada algoritma ini partisi akan membentuk ber macam-macam yang akan di evaluasi berdasarkan beberapa kriteria
- c. Density-Based adalah cluster yang membentuk berdasarkan pada akses dan fungsi densitas
- d. Grid-Based adalah sebuah cluster yang
- e. akan membentuk berdasarkan pada struktur multiple level granularity
- f. Model-Based adalah sebuah masing- masing cluster dianggap hipotesa dan hipotesa yang baik akan di pilih

C. METODE K-MEANS

K-Means dijadikan metode sebagai analisa data pada data mining dimana proses pemodelan tanpa supervisi dan merupakan salah satu metode yang mengelompokkan data secara partisi. Pada metode ini data dikelompokkan menjadi beberapa kelompok yang mempunyai karakteristik yang mirip atau sama dengan yang lainnya. Metode K-Means ini meminimalisir perbedaan antara data didalam satu cluster dan memaksimalkan perbedaan dengan cluster yang lain [8]. Pengelompokan K-Means adalah teknik yang bergantung pada pusat cluster. Biasanya dilihat dari rata- rata (Means) dari setiap cluster. Pengelompokan di ukur kesamaan kelompok dengan mengulangi jarak pengukuran antara setiap objek dan pusat setiap cluster menggunakan teori Euclidean. Algoritma K-Means adalah algoritma perulangan yang dapat dijelaskan dengan langkah-langkah sebagai berikut: [9]

- 1. Menentukan jumlah cluster K, untuk jumlah cluster tergantung pada kebutuhan
- 2. Memilih secara random nilai titik tengah (centroid) yang akan dijadikan pusat cluster awal
- 3. Mengalokasikan masing-masing data ke pusat cluster yang terdekat, dengan rumus:

$$D = \sqrt{((O_{ij} - T_{ij})^2 + (O_{ij} - T_{ij})^2)} \quad (1)$$

Dimana:

- D = Jarak data ke pusat cluster;
- O = Data record;
- T = Data centroid;

- 4. Menghitung ulang pusat cluster (centroid) baru dengan rumus:

$$KI = \frac{O_1 + O_2 + O_3 + \dots + O_n}{\sum O} \quad (2)$$

Dimana:

- KI = Centroid baru;
- O1 = Nilai data record ke-1;
- On = Nilai data record ke-n;
- $\sum O$ = Jumlah data record;

- 5. Mengalokasikan kembali data ke pusat cluster yang baru ditemukan. Apabila masih mengalami perubahan, maka proses diulang kembali ke langkah ke 3, tapi apabila pusat cluster tidak mengalami perubahan, maka proses perhitungan dihentikan.

D. OPTIMASI ELBOW

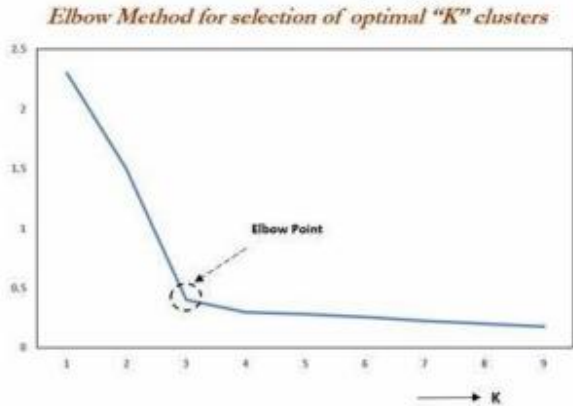
Optmasi Elbow digunakan untuk menghasilkan sebuah informasi dalam menentukan jumlah cluster terbaik dengan cara melihat presentasi hasil perbandingan antara jumlah cluster yang membuat siku pada satu titik. Metode Elbow ini memberikan suatu metode pendukung yang dapat menentukan nilai K yang optimal pada penerapan metode K-Means untuk dijadikan sebagai model data dalam penentuan cluster terbaik [10]. Cara untuk mendapatkan nilai perbandingan adalah dengan menghitung SSE (Sum of Square Error) dari masing-masing nilai cluster, dengan rumus sebagai berikut:

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_K} \|x_i - c_k\|_2^2 \quad (3)$$

Dimana:

- K = Jumlah Cluster;
- Xi = Atribut data;
- $\| \cdot \|$ = Jarak antar data; Ck = Pusat cluster;

Jika ada beberapa nilai k yang mengalami penurunan paling signifikan dan setelah itu hasil dari nilai k akan turun secara signifikan sampai hasil dari nilai k tersebut stabil, maka itu nilai SSE paling optimal.



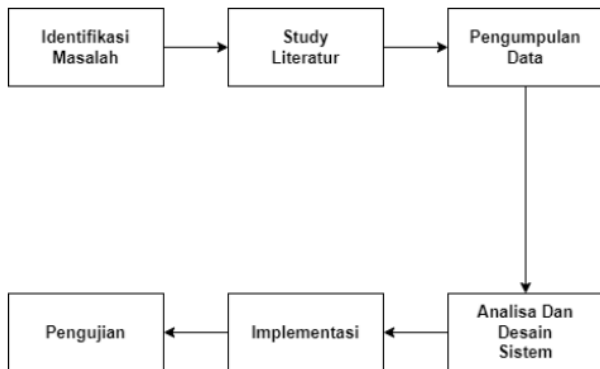
Gambar 1. Grafik Elbow

Algoritma Optimasi Elbow dalam menentukan nilai k pada K-Means pada Gambar 1 :

1. Mulai
2. Inisialisasi awal nilai k
3. Hitung hasil Sum Of Square Error
4. Melihat hasil Sum Of Square Error dari nilai k yang turun secara drastis dan stabil

Mengalokasikan kembali data ke pusat cluster yang baru ditemukan. Apabila masih mengalami perubahan, maka proses diulang kembali ke langkah ke 3, tapi apabila pusat cluster tidak mengalami perubahan, maka proses perhitungan dihentikan.

III. PERANCANGAN SISTEM



Gambar 2. Alur Penelitian

Analisa dan perancangan sistem akan menjelaskan beberapa hal yang berhubungan dengan mengidentifikasi masalah, studi literatur, mengumpulkan data, analisa dan desain sistem, Implementasi sistem yang akan dibangun, serta Pengujian sistem.

3.1 Identifikasi Masalah

Identifikasi masalah yang dilakukan pada Dinas Kelautan Kabupaten Bangkalan yaitu bagaimana cara mengelompokkan data produksi tambak garam dan mencari cluster atau jumlah kelompok yang optimal.

3.2 Studi Literatur

Peneliti melakukan pencarian informasi dengan membaca sebuah artikel atau jurnal atau penelitian terdahulu yang berhubungan dengan clustering, sehingga dapat memperbanyak sumber informasi penelitian yang berkaitan.

3.3 Pengumpulan Data

Pengumpulan data dilakukan dengan cara wawancara dan meneliti ke Dinas Kelautan Kabupaten

Bangkalan.

3.4 Desain Sistem

Flowchart pada Sistem



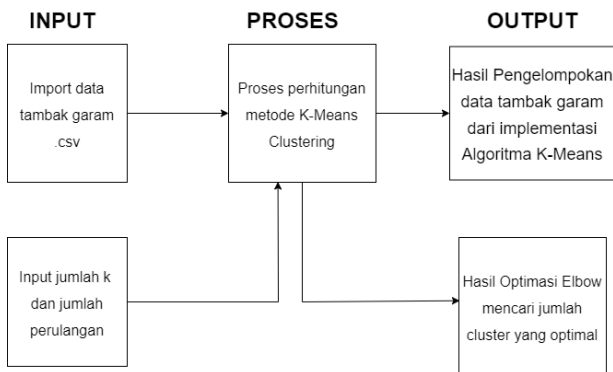
Gambar 3. Flowchart Pada Sistem

Pada Gambar 3. Flowchart algoritma k-means clustering yang terdiri beberapa tahapan, yaitu:

1. Input dataset tambak garam yang akan dimasukkan ke dalam sistem agar terbaca oleh sistem
2. Menentukan berapa jumlah cluster yang diinginkan, sehingga data objek akan dikelompokkan menyesuaikan jumlah cluster yang sudah di tetapkan.
3. Mencari titik awal centroid secara random sebagai pusat cluster perhitungan pertama, dan penentuan titik awal centroid selanjutnya menghitung hasil dari nilai rata-rata pada setiap cluster dibagi jumlah data setiap cluster
4. Selanjutnya menghitung Euclidean Distance
5. Melihat kondisi, apakah hasil dari perhitungan

- Euclidean Distance dari titik pusat cluster apakah ada perubahan data cluster
- 6. Jika ada perubahan data di cluster
- 7. Maka kembali ke tahap 3 sampai data cluster tidak ada perubahan
- 8. Jika tidak ada data cluster yang berubah
- 9. Maka, hasil nilai dari clusterisasi dari perhitungan dari metode k-means clustering sudah ditemukan
- 10. Selanjutnya mencari jumlah kelompok
- 11. (cluster) yang optimal menggunakan Elbow Method
- 12. Selesai.

3.5 Analisa Sistem



Gambar 4. Diagram IPO

Pada Gambar 4 menggambarkan tentang diagram I-P-O Rancangan Sistem yang menjelaskan bagaimana proses berjalannya suatu sistem sebagai berikut :

a. Input

Dimulai dari memasukkan file .csv data tambak garam ke dalam sistem agar bisa terbaca oleh sistem, selanjutnya menentukan jumlah k (cluster) dan jumlah perulangan.

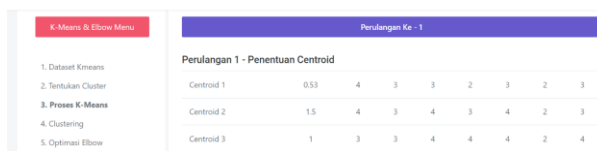
b. Proses

Pada tahapan proses data yang dimasukkan akan dilakukan perhitungan mencari titik awal centroid.

c. Output

Kemudian dari tahap proses menghasilkan objek data yang akan dikelompokkan ke dalam satu kelompok berdasarkan kesamaan data. Maka hasil dari tahapan – tahapan Proses K-Means telah selesai.

3.6 Implementasi Proses K-Means Clustering



Gambar 5. Proses K-Means Clustering

Halaman Proses K-Means merupakan halaman proses yang menampilkan perhitungan metode K- Means

Clustering pada Gambar 5.

3.7 Pengujian

Pada pengujian jumlah 7 cluster ini dilakukan untuk mencari nilai K yang paling optimal yang berguna untuk penentuan Cluster semakin akurat menggunakan Optimasi Elbow. Hasil dari proses metode K-Means Clustering. pengujian ini menggunakan metode elbow. Pengujian menggunakan metode Optimasi Elbow ini menguji 7 nilai K yang sudah ditentukan, berikut tampilan didalam sistem.

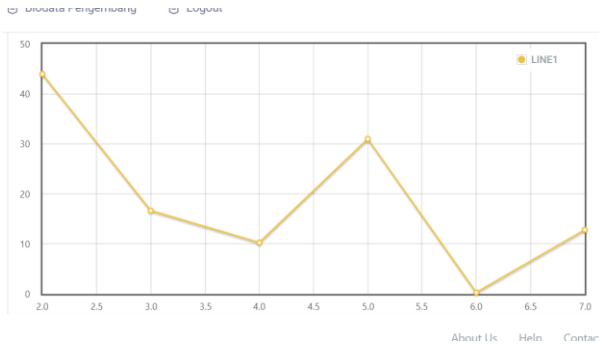
Table 1. Hasil nilai SSE dan Selisih jumlah dari 7 cluster yang didapat dari Optimasi Elbow

Cluster	Hasil Sum Square Error	Selisih
Cluster_2	43.95529	43.95529
Cluster_3	27.29889	16.65641
Cluster_4	37.58069	10.2818
Cluster_5	6.594307	30.98638
Cluster_6	6.294084	0.300223
Cluster_7	19.12541	12.83133

Pada Tabel 1 adalah hasil Optimasi Elbow dari 7 Cluster yang menunjukkan hasil SSE dan beserta selisihnya.

Gambar 6. Optimasi Elbow pada sistem

Pada Gambar 6 menjelaskan sistem yang menghasilkan proses dari Optimasi Elbow dari 7 cluster. Setelah itu hasil dari Optimasi Elbow pada gambar diatas menjelaskan setiap cluster memperoleh nilai SSE (Sum of Square Error) dan mempunyai selisih dari antara cluster ke cluster lainnya.



Gambar 7. Optimasi Elbow dalam bentuk Grafik

Pada Gambar 7 adalah menampilkan hasil dari proses dari Optimasi Elbow dalam bentuk grafik cluster

2 sampai cluster 7 sesuai hasil nilai yang didapat. Untuk hasil garis pada gambar diatas, grafik tersebut menunjukkan garis dari cluster 2 sampai ke cluster 4 menunjukkan penurunan. Dan cluster 5 mengalami kenaikan dan ke cluster 6 mengalami penurun yang signifikan, akan tetapi di cluster 7 mengalami kenaikan kembali. Sehingga cluster yang optimal adalah cluster yang berbentuk siku dan mengalami penurunan secara stabil. Untuk hasil yang menunjukkan cluster yang paling optimal sesuai pada grafik Optimasi Elbow ialah cluster 3. Maka dari pengujian Cluster 7 ini, berguna sebagai acuan untuk menguji coba mencari cluster yang cocok dan sesuai pada data yang akan di olah nantinya menggunakan metode K-Means Clustering.

IV. HASIL PENELITIAN

4.1 Kesimpulan

Hasil dari pengujian mencari nilai K yang paling optimal dengan melakukan Uji Coba dari jumlah nilai K = 7 memperoleh hasil nilai K = 3 yang paling optimal. Jadi dari Optimasi Elbow nilai K = 3 lebih Optimal dibandingkan dengan nilai K lainnya, melihat dari nilai SSE (Sum of Square Error) dan Elbow Garph yang dihasilkan yaitu SSE nilai K = 3 adalah 27.29889 dengan selisih 16.65641 dan Elbow Garph pada nilai K = 3 menunjukkan penurunan yang signifikan dan membentuk siku, maka clusterisasi pada data Tambak Garam Kabupaten Bangkalan dari total jumlah 141 data, dimana 65 data terkelompokkan ke dalam cluster 1 dengan label perlu meningkatkan produksi, 55 data terkelompokkan ke dalam cluster 2 dengan label cukup perlu meningkatkan produksi, dan 21 data terkelompokkan ke dalam cluster 3 dengan label tidak perlu meningkatkan produksi.

4.2 Saran

Hasil penelitian ini diharapkan dapat bermanfaat bagi peneliti selanjutnya yang membahas tentang hal sejenis. Sistem yang dibangun dalam penelitian ini masih memiliki banyak kekurangan yaitu Sistem masih belum sempurna dan perlu banyak perbaikan dalam tampilan sistem, fitur sistem, dan tahapan beserta hasil pemrosesan data kurang sederhana. Penelitian

selanjutnya diharapkan dapat lebih interaktif, serta lebih baik.

V.REFERENSI

- [1] A. Kadir, "Pengenalan Sistem Informasi," Am. Enterp. Inst. Public Policy Res., no. August, pp. 1–19, 2014.
- [2] Bin Ladjamudin, "Analisis dan Desain Sistem Informasi," Anal. dan Desain Sist. Inf., vol. 53, no. 9, pp. 1689–1699, 2013.
- [3] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," Proc. 2008 ACM Symp. Appl. Comput. - SAC '08, p.1556, 2008.
- [4] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," Proc. of SemEval-2007, no. June, pp. 70–74, 2007.
- [5] U. Krcadinac, P. Pasquier, J. Jovanovic, and V. Devedzic, "Synesketch: An Open Source Library for Sentence-Based Emotion Recognition," IEEE Trans. Affect. Comput., vol. 4, no. 3, pp. 312–325, 2013.
- [6] M. A. Suryadi, Kadarsah; Ramdhani, Sistem Pendukung Keputusan. Remaja Rosdakarya, Bandung, 1998

