

CLUSTERING PROVINSI DI INDONESIA BERDASARKAN PRODUKTIVITAS PANGAN DENGAN ALGORITMA FUZZY C-MEANS

Ayu Merlinda¹, Ginanjar Abdurrahman², Hardian Oktavianto³

Progam Studi Teknik Informatika, fakultas Teknik,

Universitas Muhammadiyah Jember e-mail:

ayumerlinda1998@gmail.com¹

ABSTRAK

Maraknya pembangunan perumahan dan gedung sehingga berpengaruh pada penyempitan luas areal sawah yang berakibat pada ketersediaan produksi pangan yang semakin rendah dan sebagai bahan evaluasi untuk peningkatan produksi dan produktivitas pangan di Indonesia. Sejalan dengan peningkatan pembangunan nasional khususnya dalam memenuhi kebutuhan pangan maka kebutuhan pangan juga akan semakin meningkat, mengingat besarnya sumber daya alam di sektor pertanian, ke depan sektor ini masih akan menjadi sektor yang penting dalam memberikan kontribusi bagi pertumbuhan ekonomi nasional. Dalam hal ini diperlukan suatu metode pengelompokan data terhadap hasil panen berdasarkan Produktivitas Pangan, salah satu algoritma *clustering* yang dapat digunakan adalah *Fuzzy CMeans* (FCM). Data yang digunakan yaitu data Produktivitas Pangan tahun 2016 di 34 Provinsi di Indonesia. Dari kumpulan pengujian dimulai dari 2 *cluster* hingga 10 *cluster*, menghasilkan *cluster* yang optimum terdapat pada 9 *cluster* berdasarkan nilai terendah yang dihitung dengan metode *Davies Bouldin Index*. *Cluster* 1 terdiri dari 3 provinsi, *cluster* 2 terdiri dari 4 provinsi, *cluster* 3 yang terdiri 3 provinsi, *cluster* 4 yang terdiri 1 provinsi, *cluster* 5 yang terdiri 1 provinsi, *cluster* 6 yang terdiri 1 provinsi, *cluster* 7 yang terdiri 1 provinsi, *cluster* 8 yang terdiri 15 provinsi, dan *cluster* 9 yang terdiri 4 provinsi. Jumlah anggota pada tiap cluster dipengaruhi oleh jarak antara pusat *cluster* dan kemiripan data terhadap pusat *cluster* pada saat *clustering*.

Kata Kunci: Produktivitas Pangan , *Clustering*, *Fuzzy C-Means*, *Davies Bouldin Index*

**CLUSTERING PROVINCES IN INDONESIA BASED ON PRODUCTIVITY FOOD
WITH ALGORITHMS FUZZY C-MEANS**

Ayu Merlinda¹, Ginanjar Abdurrahman², Hardian Oktavianto³

*Informatics Engineering Study Program , Faculty of Engineering,
University of Muhammadiyah Jember e-mail:
ayumerlinda1998@gmail.com*

ABSTRACT

The area under rice fields in Indonesia has been increasingly narrow with the flourishing of housing and building. This has led to the availability of the kind of food production that is inferior and of as an evaluation for the increased production of food and productivity in Indonesia. In line with the increase in national development especially in meet their food needs and demand for food will increase, given the size of the natural resources in the agricultural sector, forward sector into sector will still be essential in contributing to economic growth. In this is a method of grouping based on data for the harvest food productivity, one clustering algorithms can be used is Fuzzy CMeans (FCM). The data used the food productivity data of 2016 34 provinces in Indonesia. Of a series of testing starts on 2 to 10 clusters, steady produced clusters in the 9 clusters based on the lowest value being calculated by Davies method Bouldin index. Clusters of 1 consisting of 3, province clusters of 2 consisting of 4, province clusters of 3 consisting 3, province clusters 4 consisting 1, province cluster of 5 consisting 1, province clusters of 6 consisting 1, province clusters of 7 consisting 1, province clusters 8 consisting 15, province and clusters of 9 consisting 4. Province the number of numbers in each cluster influenced by the distance between the data center clusters and toward a center at the clustering clusters.

Keywords : Food productivity, Clustering Fuzzy C-Means, Davies Bouldin Index

1. PENDAHULUAN

Luas lahan sawah di Indonesia semakin sempit dengan banyaknya pembangunan perumahan dan gedung-gedung yang mengakibatkan ketersediaan produksi pangan yang semakin rendah. Sebagai evaluasi untuk meningkatkan produksi dan produktivitas pangan di Indonesia, kegiatan impor beras dapat terminimalisasi (Nur Afifah, 2016).

Indonesia adalah negara agraris yang sebagian besar penduduknya bermata pencaharian di pertanian. Sektor pertanian mempunyai peranan yang sangat penting dalam perekonomian nasional, hal ini dilihat dari banyaknya jumlah penduduk Indonesia yang hidup dan bekerja di sektor tersebut. Tujuan pembangunan pertanian adalah untuk meningkatkan produksi pertanian, memperluas lahan pekerjaan dan mendorong pemerataan. Menurut Adiwilaga (1992), seiring dengan meningkatnya pembangunan nasional utamanya dalam memenuhi kebutuhan pangan maka permintaan bahan pangan pun meningkat, mengingat sumber daya alam yang besar pada sektor pertanian maka di masa mendatang sektor ini masih merupakan sektor penting dalam memberikan kontribusi pada pertumbuhan ekonomi nasional.

2. PENELITIAN TERKAIT

Produktifitas Pangan

Padi merupakan salah satu komoditas tanaman pangan yang dipandang utama di Indonesia. Beras masih di pandang sebagai produk kunci bagi kestabilan perekonomian dan politik. Tanaman ini berasal dari dua benua yaitu Asia dan Afrika Barat tropis dan subtropis. Sejarah membuktikan bahwa tanaman padi sudah sejak 3000 tahun SM di Zhejiang (Cina). Tanaman padi termasuk golongan graminæ, yaitu sejenis rumput yang berumpun. Dalam 1 bibit tumbuh anakan hingga 20 lebih anakan. Sebagai besar masyarakat Indonesian menjadikan tanaman padi menjadi sumber tanaman pokok (Purnamaningsih, 2006).

Jagung merupakan salah satu tanaman pangan terpenting di dunia selain gandum dan beras. Masyarakat di beberapa daerah di Indonesia juga memanfaatkan jagung sebagai makanan pokok selain sebagai sumber karbohidrat, jagung hasil rekayasa genetika juga kini ditanam sebagai bahan obat. Selain itu jagung saat ini sedang mengalami peningkatan, hal ini terlihat dari sisi produksi dimana permintaan pasar domestik dan internasional sangat besar untuk kebutuhan pangan atau pakan. Sehingga hal ini mendorong peneliti untuk menghasilkan varietas jagung unggul guna lebih meningkatkan produktivitas dan kualitas sehingga persaingan di pasar dapat meningkat, sehingga menghasilkan 2 kelompok areal produksi jagung tinggi dan areal produksi jagung rendah (Nanda Erlangga 2019).

A. Data Mining

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies,2004).

B. Clustering

Menurut Deka (2014) *Clustering* merupakan teknik *data mining* yang digunakan untuk mendapatkan sekumpulan objek yang memiliki karakteristik yang sama dengan data yang cukup besar. Tujuan utama dari metode *clustering* adalah mengelompokkan sejumlah data atau objek ke dalam cluster atau grup sehingga setiap *cluster* akan memuat data semaksimal mungkin. *Clustering* melakukan pengelompokan data berdasarkan kemiripan antar objek, oleh karena itu pengelompokan tersebut diklasifikasikan sebagai metode *unsupervised learning*.

C. Fuzzy C-Mean

Fuzzy C-Means pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. *Fuzzy C-Means* merupakan metode *clustering* dengan pendekatan *fuzzy*, artinya setiap data yang di *cluster* memungkinkan menjadi anggota lebih dari satu *cluster*. Konsep dasar *Fuzzy C-Means* adalah menentukan pusat *cluster*, pada kondisi awal pusat *cluster* ini masih belum akurat dan setiap data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster*.

Ada beberapa algoritma *clustering* data, salah satu diantaranya adalah *Fuzzy CMeans*. *Fuzzy C-Means* adalah salah satu teknik klasterisasi data di mana keberadaan tiap-tiap titik data diperkenalkan oleh Jim Bezdek pada tahun 1981. *Output* dari *Fuzzy CMeans* bukan merupakan *fuzzy inference system*, namun merupakan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system* (Kusumadewi, 2010).

Algoritma *Fuzzy C-Means* adalah sebagai berikut:

1. *Input* data yang akan di *cluster* X, berupa matriks berukuran n x m (n=jumlah sampel data, m=atribut setiap data). =data sampel ke-i(i=1,2,..,n),atribut ke- j(1,2,..,n).
2. Menentukan beberapa *input* yang dibutuhkan dalam perhitungan *Fuzzy C- Means*, yaitu:
 - Jumlah *cluster* (c)
 - Pangkat (w)
 - Maksimum iterasi (MaxIter)
 - *Error* terkecil (ξ)
 - Fungsi objektif awal
 - Iterasi awal (t = 1)
3. Membangkitkan bilangan *random* , i=1,2,.....,n; k=1,2, ,c; sebagai elemen-elemen matriks partisi awal U. Hitung jumlah setiap kolom:

$$Q_i = \sum_{k=1}^c \mu_{ik} \dots \dots \dots (2.1)$$

Q_i ialah jumlah setiap kolom dari nilai *random* sebuah matrik, jumlah Q tergantung dari beberapa jumlah kriteria penilaian.

4. Menghitung pusat *cluster* ke-k: , dengan k=1,2,.....,c; dan j=1,2,.....,m

$$V_{ki} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \dots \dots \dots (2.2)$$

- i = iterasi
- V_{kj} = perubahan matriks
- x_{ij} = atribut

V_{kj} ialah titik pusat setiap *cluster*, jumlah V_{kj} tergantung dari berapa *cluster* yang akan dibentuk dan n ialah jumlah proposal.

5. Menghitung fungsi objektif pada iterasi ke- t, P_t

$$P_t = \sum_{i=1}^{n_i} \sum_{k=1}^{n_k} ([\sum_{j=1}^{m_j} (x_{ij} - v_{ij})^2] (\mu_{ik})^w) \dots \dots \dots (2.3)$$

P_t = fungsi objektif

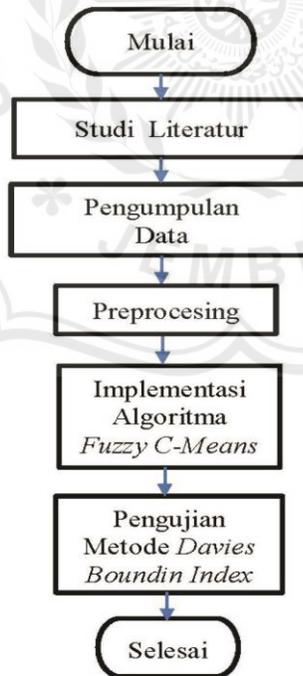
$\sum_{i=1}^{n_i}$ = jumlah data di *cluster*

$\sum_{i=1}^{n_i}$ = jumlah perhitungan *cluster* awal

D. Metode Analisis Data

Pada penelitian ini, metode analisa data yang digunakan adalah *Fuzzy Cmeans*. Dalam melakukan sebuah pengelompokan data yang diperoleh dari *Fuzzy Cmeans* agar menjadi kelompok dari sebuah cluster yang tergabung dalam tingkat keanggotaannya tidak sama seperti nol sampai satu. Kelebihan dari *Fuzzy C-means* yaitu melakukan clustering lebih dari banyaknya huruf dengan cara bersamaan. Penelitian dengan metode *Fuzzy C-Means* yang telah banyak diterapkan untuk bermacam-macam keperluan dalam mengatasi masalah yang dihadapi.

E. Diagram Alur Penelitian



Gambar 3.1 Diagram Alur Penelitian

F. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil resmi Badan Pusat Statistik yang berjudul Produktivitas Pangan Menurut Provinsi di Indonesia 2016 <https://bps.go.id> untuk melengkapi. Data yang digunakan dalam penelitian ini adalah data Produktivitas Pangan Tahun 2016 yang terdiri dari 34 Provinsi Indonesia.

3. HASIL DAN PEMBAHASAN

Data pengujian membahas mengenai hasil yang akan diperoleh dari pengujian data, data tersebut akan diolah menggunakan algoritma Fuzzy C-Means. Berikut hasil dan pembahasan dari Algoritma Fuzzy C-means untuk mengelompokkan Provinsi di Indonesia berdasarkan Produktivitas Pangan. Data yang digunakan adalah data Produktivitas Pangan yaitu Padi, Jagung, Kedelai, Kacang Tanah, Kacang Hijau, Ubi Kayu, Ubi Jalar di Indonesia pada tahun 2016.

Tabel 4.1 Data Produktivitas Pangan di Indonesia

Provinsi	Produktivitas (Kuintal/Hektar)						
	Padi	Jagung	Kedelai	Kacang tanah	Kacang Hijau	Ubi Kayu	Ubi Jalar
ACEH	50,56	42,76	14,61	12,52	10,51	130,87	112,67
SUMATERA UTARA	51,74	62,33	12,35	11,60	11,28	338,54	136,69
SUMATERA BARAT	50,25	68,61	11,93	14,59	12,47	391,85	313,87
RIAU	36,63	24,85	14,15	9,58	10,38	289,54	82,75
.....
PAPUA	43,95	24,36	12,76	11,01	10,91	121,37	123,83

a. Fuzzy C-Means(FCM) pada Jupyter

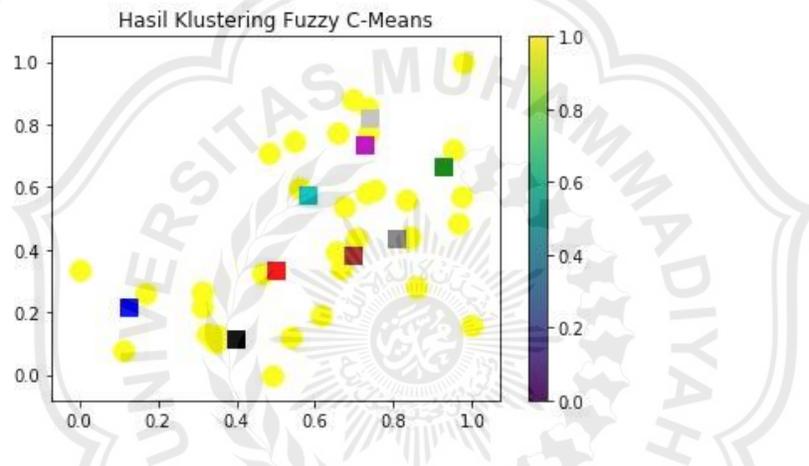
Pengolahan data menggunakan tools Jupyter yang di cluster menggunakan algoritma Fuzzy C-Means dari 2 cluster sampai dengan 10 cluster. Pengolahan tersebut menghasilkan output berupa bilangan random, data yang di cluster, pusat cluster, fungsi objektif, derajat keanggotaan, nilai DBI dan total keseluruhan items.

b. Penentuan Jumlah Cluster Optimum

Setelah melalui proses cluster dengan menggunakan algoritma Fuzzy C- Means, kemudian dilakukan proses dengan menggunakan metode Davies Bouldin-Index untuk menentukan cluster optimum. Berikut ini adalah hasil metode Davies Bouldin-Index :

Tabel 4.2 Hasil Nilai Metode *Davies Bouldin-Index*

Cluster	Nilai <i>Davies Bouldin-Index</i>
2	1,518
3	1,498
4	1,449
5	1,454
6	1,490
7	1,675
8	1,485



```
In [32]: print("Nilai DBI 9 cluster: ",davies_bouldin_score(x_scaled, fcm9_labels))
```

Nilai DBI 9 cluster: 1.2494137787542627

9	1,247
10	1,292

4. KESIMPULAN

Berdasarkan uraian permasalahan dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan hasil metode Davies Bouldin Index dari proses clustering data Produktivitas Pangan di Indonesia untuk mengelompokkan provinsi di Indonesia menggunakan algoritma Fuzzy C-M nilai cluster yang paling optimum adalah 9 cluster yang memiliki nilai terendah pada Davies Bouldin- Index yaitu 1,247.
2. Cluster 1 yang memiliki 3 anggota provinsi, Cluster 2 yang memiliki 4 anggota provinsi, Cluster 3 yang memiliki 3 anggota provinsi, Cluster 4 yang memiliki 1 anggota provinsi , Cluster 5 yang memiliki 1 anggota provinsi, Cluster 6 yang

memiliki 1 anggota provinsi, Cluster 7 yang memiliki 1 anggota provinsi, Cluster 8 yang memiliki 15 anggota provinsi dan Cluster 9 yang memiliki 4 anggota provinsi. Jumlah anggota pada tiap Cluster dipengaruhi oleh jarak antara pusat Cluster dan kemiripan data terhadap pusat Cluster pada saat proses Clustering. Data lengkap terdapat pada lampiran halaman 56 tabel cluster 9

5. SARAN

Saran dari penelitian ini adalah:

1. Validitas cluster digunakan untuk mencari cluster yang terbaik, dapat menggunakan validitas yang lain seperti Elbow, Silhouette, Gap Statistic, dll.
2. Perhitungan manual cluster Fuzzy C-Means pada Microsoft Excel dapat digunakan sebagai alternatif pembandingan hasil cluster pada Jupyter.



DAFTAR PUSTAKA

- Nur Afifah (2016). Pengklasteran Lahan Sawah Di Indonesia Sebagai Evaluasi Ketersediaan Produksi Pangan Menggunakan Fuzzy C-Means. *Jurnal Matematika "MANTIK"* Edisi: Oktober 2016. Vol. 02 No. 01
- Ragapadmi Purnamaningsih (2006). Induksi Kalus dan Optimasi Regenerasi Empat Varietas Padi melalui Kultur In Vitro. *Jurnal ArgoBiogen* 2(2):74-80
- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceeding – 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Bezdek, J. C., Ehrlich, R., & Full W. (1984). FCM: The Fuzzy c-means Clustering algorithm. *Computers & Geosciences*, 10(2), 191-203
- Kusumadewi, S., Hartatik, S., Harjoko, A., Wardoyo, R., 2006, Fuzzy Multi-Attribute Decision Making, Graha Ilmu, Yogyakarta.