

Pengelompokan Provinsi di Indonesia Berdasarkan Produktivitas Tanaman Pangan tahun 2015 menggunakan K-means Clustering

Iban Aria Nugraha



Capstone Project / Internship

Data Science Indonesia Bootcamp 2016

Latar Belakang

Pangan merupakan salah satu kebutuhan paling dasar dan penting bagi manusia untuk dapat melanjutkan keberlangsungan hidupnya. Oleh karena itu, pangan juga disebut sebagai kebutuhan primer yang paling prioritas untuk dipenuhi.

Di Indonesia, sumber pangan berasal dari tanaman yang mengandung unsur zat tepung atau karbohidrat yang tinggi, seperti sereal (beras, jagung, sorgum, gandum) dan umbi-umbian (ubi kayu, ubi jalar, kentang, talas). Bahan pangan tersebut berguna bagi tubuh manusia sebagai sumber energi / tenaga untuk dapat beraktivitas.

Berdasarkan penjelasan di atas, ketersediaan pangan menjadi salah satu faktor penting yang perlu diperhatikan oleh suatu negara untuk menjamin kebutuhan pangan penduduknya dapat terpenuhi dengan baik. Terlebih lagi, pertumbuhan penduduk suatu negara yang terus bertambah secara eksponensial mendorong perlu adanya suatu strategi agar ketersediaan pangan untuk penduduk tetap terjaga.

Pengelompokan provinsi berdasarkan tingkat produktivitas tanaman pangan merupakan salah satu solusi yang dapat dilakukan untuk menentukan strategi ke depan bagi pemerintahan suatu negara dalam upaya pemenuhan pangan penduduknya.

Tujuan & Manfaat

Mengetahui hasil pengelompokan provinsi di Indonesia berdasarkan tingkat produktivitas tanaman pangan dari setiap provinsi. Pengelompokan provinsi ini selanjutnya dapat bermanfaat atau digunakan sebagai acuan pendukung bagi pemerintah Indonesia dalam menentukan strategi dan sikap yang tepat dalam upaya pemenuhan kebutuhan pangan penduduk di Indonesia.

Bahan Penelitian

Data

Data sekunder diperoleh peneliti melalui hasil publikasi Badan Pusat Statistik (BPS) Indonesia melalui laman resmi lembaga pemerintahan tersebut.

Laman resmi BPS Indonesia: www.bps.go.id,

Dataset:

www.bps.go.id/Subjek/view/id/53#subjekViewTab3|accordion-daftar-subjek1

(Pembaharuan terakhir: 1 Juli 2016)

Data yang peneliti gunakan dalam penelitian ini adalah data tingkat produktivitas tanaman pangan (jagung, kacang hijau, kacang tanah, kedelai, padi, ubi jalar, ubi kayu) dari setiap provinsi di Indonesia pada tahun 2015.

Data ini sendiri dikumpulkan oleh BPS melalui kegiatan Survei Pertanian (SP) dan Survei Ubinan. Produktivitas tanaman pangan sendiri merupakan formula dari pembagian antara jumlah total hasil produksi dengan luas panen tanaman pangan (diukur dengan satuan kuintal per hektar).

HASIL & PEMBAHASAN

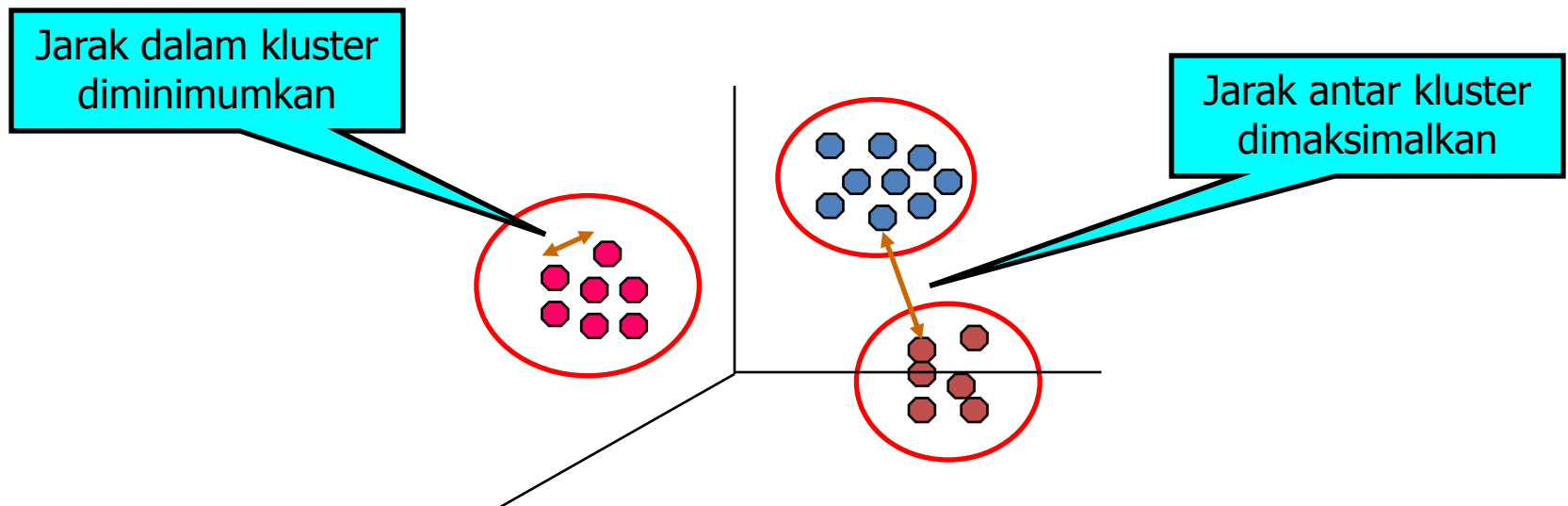
**DATA
SCIENCE
INDONESIA**

Capstone Project / Internship

Data Science Indonesia Bootcamp 2016

Cluster Analysis

Cluster Analysis (atau disebut juga analisis gerombol) merupakan teknik peubah ganda yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya, dimana karakteristik objek-objek dalam suatu gerombol memiliki tingkat kemiripan yang tinggi, sedangkan karakteristik antar objek pada suatu gerombol dengan gerombol lain memiliki tingkat kemiripan yang rendah.



K-means Cluster Analysis

Salah satu metode yang digunakan dalam analisis gerombol tak berhierarki adalah metode k-means. Metode ini cukup relatif sederhana dalam proses pengelompokan sejumlah besar objek dengan atribut tertentu ke dalam kelompok-kelompok (clusters) sebanyak K.

Algoritma K-means Clustering Analysis:

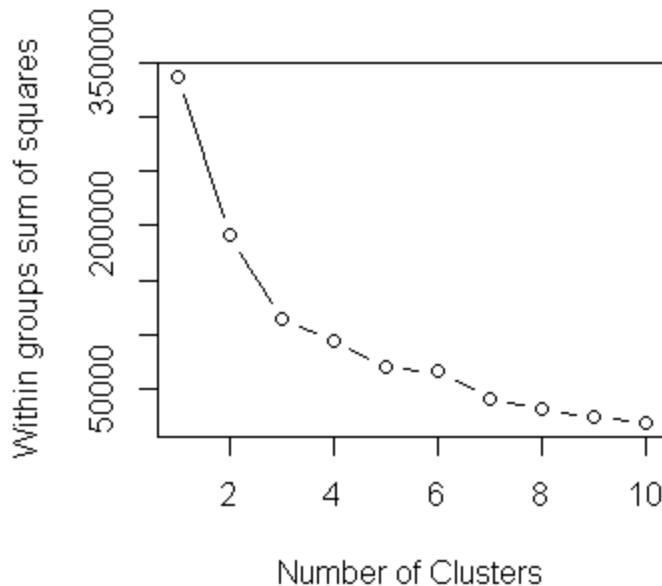
1. Penentuan besarnya K (banyaknya gerombol) dan inisialisasi centroid di setiap gerombolnya
2. Penghitungan jarak antar setiap objek dengan setiap centroid
3. Pengelompokan setiap objek berdasarkan jarak terdekat antara objek dengan centroid-nya
4. Penghitungan nilai centroid (rata-rata) baru berdasarkan hasil pengelompokan terakhir yang terbentuk
5. Lakukan kembali langkah (2) – (4) sampai tidak ada lagi pemindahan objek antar gerombol

Hasil Penelitian: Penentuan “K”

Penentuan nilai “k” pada K-means Clustering Analysis:

> Elbow Method

Melakukan perhitungan *Sum of Squared Error* (SSE) untuk beberapa nilai K. Sum of Squared Error (SSE) sendiri adalah jumlah dari kuadrat jarak antara setiap objek kelompok dengan *centroid* kelompoknya. Konsep dari Elbow Method adalah dengan memilih nilai K dimana terjadi penurunan SSE secara curam.



Berdasarkan pengamatan grafik di samping, **nilai k yang optimum adalah 3.**

Hasil Penelitian: Konvergensi Iterasi

Setelah ditentukan nilai K optimum yang merepresentasikan banyaknya kelompok yang akan terbentuk ($K = 3$), penelitian dilanjutkan dengan melakukan iterasi perhitungan sesuai dengan algoritma k-means clustering analysis menggunakan bantuan SPSS 16.0

Output:

Iteration History^a

| Iteration | Change in Cluster Centers | | |
|-----------|---------------------------|--------|--------|
| | 1 | 2 | 3 |
| 1 | .000 | 45.325 | 62.435 |
| 2 | .000 | 3.313 | 18.505 |
| 3 | .000 | 3.847 | 15.031 |
| 4 | .000 | 5.796 | 19.341 |
| 5 | .000 | .000 | .000 |

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 5. The minimum distance between initial centers is 223.289.

Tabel di atas menunjukkan iterasi dari algoritma *K-means clustering analysis* sampai tidak ada lagi perpindahan objek antar gerombol (konvergen). Perhitungan algoritma dinyatakan konvergen pada iterasi ke-5.

Hasil Penelitian: Pengelompokan Provinsi

Hasil Pengelompokan Provinsi berdasarkan Tingkat Produktivitas Tanaman Pangan:

| Case Number | Provinsi | Cluster | Distance |
|-------------|---------------------------|---------|----------|
| 1 | Nangroe Aceh Darussalam | 2 | 55.48 |
| 2 | Sumatera Utara | 3 | 116.04 |
| 3 | Sumatera Barat | 3 | 144.59 |
| 4 | Riau | 2 | 109.36 |
| 5 | Jambi | 3 | 102.38 |
| 6 | Sumatera Selatan | 2 | 66.43 |
| 7 | Bengkulu | 2 | 45.09 |
| 8 | Lampung | 2 | 81.83 |
| 9 | Kepulauan Bangka Belitung | 2 | 65.65 |
| 10 | Kepulauan Riau | 2 | 68.51 |
| 11 | DKI Jakarta | 1 | - |
| 12 | Jawa Barat | 3 | 57.06 |
| 13 | Jawa Tengah | 3 | 41.19 |
| 14 | DI Yogyakarta | 2 | 49.25 |
| 15 | Jawa Timur | 3 | 72.13 |
| 16 | Banten | 2 | 24.55 |
| 17 | Bali | 2 | 81.28 |

| Case Number | Provinsi | Cluster | Distance |
|-------------|---------------------|---------|----------|
| 18 | Nusa Tenggara Barat | 2 | 69.41 |
| 19 | Nusa Tenggara Timur | 2 | 93.14 |
| 20 | Kalimantan Barat | 2 | 37.15 |
| 21 | Kalimantan Tengah | 2 | 42.62 |
| 22 | Kalimantan Selatan | 2 | 40.92 |
| 23 | Kalimantan Timur | 2 | 40.94 |
| 24 | Kalimantan Utara | 2 | 49.05 |
| 25 | Sulawesi Utara | 2 | 65.43 |
| 26 | Sulawesi Tengah | 2 | 28.14 |
| 27 | Sulawesi Selatan | 2 | 49.02 |
| 28 | Sulawesi Tenggara | 2 | 28.02 |
| 29 | Gorontalo | 2 | 54.66 |
| 30 | Sulawesi Barat | 2 | 41.45 |
| 31 | Maluku | 3 | 58.41 |
| 32 | Maluku Utara | 2 | 45.80 |
| 33 | Papua Barat | 2 | 75.65 |
| 34 | Papua | 2 | 66.80 |

Hasil Penelitian: Pengelompokan Provinsi

Hasil Pengelompokan Provinsi berdasarkan Tingkat Produktivitas Tanaman Pangan

| Group | Banyaknya Provinsi | % | Anggota | | | | | | |
|-------|--------------------|--------|-------------------------|----------------------|---------------------|--------------------|------------------|-------------------|-------------|
| 1 | 1 | 2.94% | DKI Jakarta | | | | | | |
| 2 | 26 | 76.47% | Nangroe Aceh Darussalam | Lampung | Banten | Kalimantan Barat | Kalimantan Utara | Sulawesi Tenggara | Papua Barat |
| | | | Riau | Kep. Bangka Belitung | Bali | Kalimantan Tengah | Sulawesi Utara | Gorontalo | Papua |
| | | | Sumatera Selatan | Kep. Riau | Nusa Tenggara Barat | Kalimantan Selatan | Sulawesi Tengah | Sulawesi Barat | |
| | | | Bengkulu | DI Yogyakarta | Nusa Tenggara Timur | Kalimantan Timur | Sulawesi Selatan | Maluku Utara | |
| 3 | 7 | 20.59% | Sumatera Utara | Sumatera Barat | Jambi | Jawa Barat | Jawa Tengah | Jawa Timur | Maluku |

Berdasarkan tabel di atas, terlihat pengelompokan 34 provinsi ke dalam 3 *clusters* adalah sebagai berikut:

- *Cluster 1* dengan persentase 2.94% hanya beranggotakan DKI Jakarta
- *Cluster 2* dengan persentase 76.47% yang merupakan kelompok dengan objek terbanyak terdiri dari Nangroe Aceh Darussalam, Riau, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta, Banten, Bali, NTT, NTB, semua provinsi di Pulau Kalimantan dan Sulawesi, Maluku Utara, Papua Barat, dan Papua
- *Cluster 3* dengan persentase 20.59% terdiri dari Sumatera Utara, Sumatera Barat, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, dan Maluku

Hasil Penelitian: Signifikansi Variabel

Uji Signifikansi Variabel Tanaman Pangan

ANOVA

| | Cluster | | Error | | F | Sig. |
|----------------------------|-------------|----|-------------|----|--------|------|
| | Mean Square | df | Mean Square | df | | |
| Produktivitas Jagung | 2174.452 | 2 | 170.135 | 31 | 12.781 | .000 |
| Produktivitas Kacang Hijau | 62.924 | 2 | 9.374 | 31 | 6.713 | .004 |
| Produktivitas Kacang Tanah | 78.295 | 2 | 5.178 | 31 | 15.121 | .000 |
| Produktivitas Kedelai | 89.311 | 2 | 6.887 | 31 | 12.967 | .000 |
| Produktivitas Padi | 343.987 | 2 | 89.594 | 31 | 3.839 | .032 |
| Produktivitas Ubi Jalar | 48241.012 | 2 | 1427.146 | 31 | 33.802 | .000 |
| Produktivitas Ubi Kayu | 41174.962 | 2 | 3225.102 | 31 | 12.767 | .000 |

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Berdasarkan tabel ANOVA di atas, terlihat bahwa nilai signifikansi tingkat produktivitas semua tanaman pangan bernilai kurang dari 0.05. Hal ini menunjukkan bahwa ada perbedaan antara *cluster 1*, *cluster 2*, dan *cluster 3* yang berhubungan dengan semua variabel tersebut.

Dari semua variabel tingkat produktivitas tanaman pangan, Ubi Jalar merupakan variabel pembeda paling signifikan antar *cluster* satu dengan *cluster* lainnya dikarenakan memiliki nilai pengujian F yang paling tinggi di antara tanaman pangan lainnya. Sementara, padi merupakan variabel pembeda dengan tingkat signifikansi paling rendah di antara semua tanaman pangan.

Hasil Penelitian: Definisi Cluster

Definisi Cluster

Final Cluster Centers

| | Cluster | | |
|----------------------------|---------|-----|-----|
| | 1 | 2 | 3 |
| Produktivitas Jagung | 0 | 39 | 60 |
| Produktivitas Kacang Hijau | 0 | 9 | 12 |
| Produktivitas Kacang Tanah | 0 | 11 | 13 |
| Produktivitas Kedelai | 0 | 13 | 14 |
| Produktivitas Padi | 56 | 44 | 55 |
| Produktivitas Ubi Jalar | 0 | 113 | 232 |
| Produktivitas Ubi Kayu | 0 | 186 | 273 |

Berdasarkan tabel di atas, kita dapat melakukan pendefinisian setiap *cluster* terbentuk berdasarkan hasil akhir dari iterasi *centroid* pada K-means *clustering analysis*.

Cluster 1 didefinisikan sebagai kelompok yang hanya memproduksi padi saja dengan tingkat produktivitas padi hampir sama dengan *cluster* lainnya. Berbeda dengan *Cluster 1*, *Cluster 2* dan *Cluster 3* memproduksi semua jenis tanaman pangan, dimana tingkat produktivitas semua jenis pangan pada *Cluster 3* lebih tinggi dibandingkan *Cluster 2*.

KESIMPULAN & SARAN

**DATA
SCIENCE
INDONESIA**

Capstone Project / Internship

Data Science Indonesia Bootcamp 2016

Kesimpulan & Saran

Kesimpulan

Berdasarkan hasil analisa K-means Clustering, provinsi di Indonesia terbagi ke dalam 3 kelompok dengan melihat besaran tingkat produktivitas semua tanaman pangan yang diproduksi pada setiap provinsinya. Kelompok 1 hanya beranggotakan Provinsi DKI Jakarta dengan dicirikan sebagai kelompok yang hanya memproduksi padi saja dengan tingkat produktivitas yang hampir setara dengan kelompok lainnya. Kelompok 2 dan Kelompok 3 dicirikan sebagai kelompok provinsi yang memproduksi semua jenis tanaman pangan, namun perbedaan terlihat pada nilai tingkat produktivitas Kelompok 3 yang lebih tinggi pada semua jenis tanaman pangan dibandingkan dengan Kelompok 2.

Saran

Hasil penelitian ini dirasa akan lebih strategik jika dapat digabungkan dengan penelitian lainnya yang berkaitan dengan produksi tanaman pangan, seperti statistik konsumsi tanaman pangan per provinsi, pemetaan jalur distribusi tanaman pangan, dan penelusuran faktor yang mempengaruhi dari tingkat produktivitas tanaman pangan.

Selain itu, tren dari kekonsistenan dan kestabilan pengelompokan provinsi dapat terlihat dengan melakukan analisa serupa dengan menggunakan data pada tahun-tahun sebelumnya.

APPENDIX

**DATA
SCIENCE
INDONESIA**

Capstone Project / Internship

Data Science Indonesia Bootcamp 2016

Definisi (Glosarium)

Luas Panen

adalah luasan tanaman yang dipungut hasilnya setelah tanaman tersebut cukup umur.

Bentuk Produksi

Jenis tanaman yang dilaporkan beserta bentuk produksinya meliputi:

- Padi: kualitas produksi gabah kering giling
- Jagung: kualitas produksi pipilan kering
- Kacang hijau, kacang tanah, dan kedelai: kualitas produksi biji kering
- Ubi jalar dan ubi kayu: kualitas produksi umbi basah

TERIMA KASIH

The logo for Data Science Indonesia is a dark blue circle containing the text "DATA SCIENCE INDONESIA" in white, stacked vertically.

**DATA
SCIENCE
INDONESIA**

Capstone Project / Internship

Data Science Indonesia Bootcamp 2016