

# **KLASTERISASI PROVINSI DI INDONESIA BERDASARKAN PRODUKTIVITAS PANGAN KOMODITAS MENGGUNAKAN ALGORITMA K-MEANS**

**Aditya Novita**

## **ABSTRAK**

Penelitian ini dilakukan bertujuan untuk klasterisasi provinsi berdasarkan luas panen, produksi, dan produktivitas pangan komoditas di Indonesia. Data bersumber dari website Kementerian Pertanian. Data penelitian ini mencakup data luas panen, produksi, dan produktivitas provinsi di Indonesia pada tahun 2017 hingga 2019. Penelitian dilakukan menggunakan K-Means dalam pengelompokan suatu data dan dievaluasi dengan perhitungan (*Sum of Square Error*) SSE agar ditemukan klaster yang optimal. Penelitian ini dieksekusi menggunakan *Google Colaboratory* dan bahasa yang digunakan yaitu pemrograman python. Hasil penelitian klasterisasi provinsi ini dihasilkan klaster optimal pada  $k=3$  dengan selisih nilai SSE sebesar 241.05797006047 . Hasil klasterisasi pada klaster 0 (sedang) berjumlah 29 data dengan karakteristik yaitu provinsi memiliki variabel lebih dominan yang nilainya lebih rendah dari klaster 1 dan lebih tinggi dari klaster 2, pada klaster 1 (tinggi) berjumlah 64 data karakteristik provinsi memiliki variabel lebih dominan nilainya lebih tinggi dari klaster 0 dan 2, pada klaster 2 (rendah) berjumlah 9 data dengan karakteristik provinsi memiliki variabel lebih dominan nilainya lebih rendah dari klaster 0 dan 1.

**Keyword** : Pangan, *Clustering*, *K-Means*, Elbow

# **PROVINCIAL CLUSTERIZATION IN INDONESIA BASED ON FOOD PRODUCTIVITY OF COMMODITIES USING K-MEANS ALGORITHM**

**Aditya Novita**

## **ABSTRACT**

This research was conducted with the aim of clustering provinces based on harvested area, production, and productivity of food commodities in Indonesia. Data sourced from the website of the Ministry of Agriculture. The data of this research include data on harvested area, production, and productivity of provinces in Indonesia from 2017 to 2019. The study was conducted using K-Means in grouping a data and evaluated by calculating (Sum of Square Error) SSE in order to find the optimal cluster. This research was executed using Google Collaboratory and the language used was python programming. The results of this provincial clustering study resulted in the optimal cluster at  $k=3$  with a difference in SSE value of 241.05797006047 . The results of clustering in cluster 0 (medium) amount to 29 data with the characteristics that the province has a more dominant variable whose value is lower than cluster 1 and higher than cluster 2, in cluster 1 (high) there are 64 data characteristic of the province having a more dominant variable whose value is higher. higher than clusters 0 and 2, in cluster 2 (low) there are 9 data with provincial characteristics having more dominant variables whose values are lower than clusters 0 and 1.

*Keyword : Food, Clustering, K-Means, Elbow*