Food Self-sufficiency Decision Support Model Based on Provinces in Indonesia Using the Clustering Method

M. Safii¹, Mochamad Wahyudi¹, Solikhun¹, Muhammad Zarlis², Syahril Effendi²

¹ Doctoral Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Indonesia
² Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Indonesia

* m.safii@amiktunasbangsa.ac.id

Abstract. In Indonesia rice is a staple food as a major source of carbohydrates. Rice is also a staple food of the people of most countries in Asia and even most of the population in the world. Food security for Indonesia is closely linked to the adequacy of rice provision. The indicators of national food security are (1) the availability of food every time; (2) the ability or purchasing power of the people towards food; (3) guarantee of distribution and supply of food; and (4) the ability to import in urgent conditions. Rice as a rice-producing crop is a very important commodity for Indonesia, besides being the main food producer, rice commodity is also the main source of income from millions of farmers. The availability of rice at affordable prices to communities is an important factor for national security, security and stability. Indonesia has a large population, the fourth largest in the world after China, America and India. Dependence on rice imports has a high risk not only of food dependency on other countries, but also limited availability of rice in the international market. In 1977 Indonesia imported 2 million tons of rice and accounted for one third of the rice available on the international market, affecting the price of rice. Increased rice production can be achieved through increased productivity with the application of technology varieties and better cultivation and expansion of planting areas with increased intensity of planting and printing or opening of new rice cultivation area. Based on the above explanation, the researcher uses Data Mining Clustering with K-Means to get the assessment result of all rice production provinces. With this method the results of the research can be used as a model of decision makers for the government to give priority attention to areas that have the potential to increase rice production.

1. Introducing
In Indonesia the main source of food is rice as a source of carbohydrates. Rice is also a staple food for people from most countries in Asia and even most of the population in the world. Food security for Indonesia is closely related to the adequacy of rice supply. The indicator of national food security is: 1. food availability at all times; 2. the ability or purchasing power of people towards food; 3. guaranteed distribution and supply of foodstuffs; and 4. ability to import under urgent conditions [1]. Rice as a rice-producing crop is a very important commodity for Indonesia. as a producer of staple food, rice is also the main source of income for millions of farmers. The availability of rice at affordable prices for the community is an important factor for national security, security and stability of the government.
Thus rice is not only important as a food commodity, but also important as an economic commodity, cultural commodity, strategic commodity, and political commodity. Indonesia has a large population, the fourth largest in the world after China, America and India. Dependence on rice imports contains high risk not only dependence on food for other countries, but also the limited volume of rice availability in international markets.

In 1977 Indonesia imported 2 million tons of rice and reached one third of rice available in international markets, which affected the price of rice. To realize good food security, efforts to increase rice production need to be continued and improve. In addition, it is also necessary to reduce demand through reducing population growth and food diversification. Increased rice production can be achieved through increased productivity with better application of varieties and cultivation technologies and expansion of planting areas by increasing cropping intensity and printing or opening new areas of rice cultivation [2]. Based on the forecast figures of the Central Statistics Agency (BPS) and the Ministry of Agriculture, 2016 rice production reached 79.14 million tons of DUP, an increase of 3.74 million tons compared to 2015. In 2016 corn production of 23.16 million tons of dry shelled or an increase of 3.55 million tons compared to 2015. Based on the explanation above, this study was conducted using Data Mining Clustering to get the results of grouping of provinces as productive or unproductive rice producers [3][4]. There have been many clustering techniques proposed but K-means is one of the oldest and most popular clustering techniques. In this method the number of cluster (k) is predefined prior to analysis and then the selection of the initial centroids will be made randomly and it followed by iterative process of assigning each data point to its nearest centroid [5][6]. It is hoped that the government can make the right decision to prioritize areas that must be increased in production so that the self-sufficiency goals can be implemented [7]. With this grouping, it is hoped that the government can make the right decision to prioritize areas that must be increased in production so that the self-sufficiency goals can be implemented.

2. Methodology
Cluster analysis; Clustering and segmentation actually partition the database, so each partition or group is the same according to certain criteria or metrics. If similarity measurements are available, then there are a number of techniques for forming clusters. In the Clustering method the main concept emphasized is the iterative search for cluster centers, where the cluster center is determined based on the minimum distance of each data at the center of the cluster [8][9]. The data used in this study is based on documents produced by the National Statistics Agency through the https://www.bps.go.id site. In this case the researcher raised the topic of the amount of rice production by province in 2011-2015. The results of the cluster can be used as input for the government so that provinces that enter the low cluster receive special attention as a policy to achieve self-sufficiency goals. The design of the use case diagram is shown in Figure 2.

![Figure 1. Use Case Diagram of K-Means Data Mining](image-url)
3. Results and Discussion
In Clustering, the data obtained will be calculated in advance based on the amount of rice production by province in 2011-2015. The sum result is based on 1 assessment criteria, namely the amount of rice production as shown in table 1.

### Table 1. Data on Rice Production by Province

| Province                  | Land area | Yields |
|---------------------------|-----------|--------|
|                           | 2011      | 2012   | 2013   | 2014   | 2015   | 2012   | 2013   | 2014   | 2015   |
| Aceh                      | 307556    | 308973 | 300808 | 294129 | 290337 | 380869 | 387803 | 419183 | 376137 |
| Sumut                     | 467138    | 448722 | 438346 | 430443 | 423465 | 757547 | 765099 | 742968 | 717318 |
| Sumbar                    | 231463    | 230775 | 224182 | 225890 | 226377 | 461709 | 476422 | 478820 | 503198 |
| Riau                      | 115897    | 109585 | 93338  | 87594  | 71910  | 145242 | 144015 | 115818 | 106037 |
| Jambi                     | 113577    | 113579 | 113546 | 101195 | 94735  | 157415 | 149369 | 153243 | 145990 |
| Sumsel                    | 629355    | 610314 | 612424 | 616753 | 620632 | 784820 | 769725 | 800036 | 810900 |
| Bengkulu                  | 90217     | 88877  | 93382  | 88756  | 85131  | 127934 | 144448 | 147680 | 147572 |
| Lampung                   | 350949    | 364111 | 362037 | 363055 | 377463 | 606973 | 614876 | 638090 | 648731 |
| Bangka Belitung           | 5932      | 6133   | 5358   | 7490   | 10654  | 5299   | 7995   | 10252  | 9943   |
| Kep. Riau                 | 393       | 559    | 487    | 405    | 246    | 387    | 382    | 379    | 385    |
| Jakarta                   | 1098      | 1001   | 895    | 778    | 650    | 1723   | 1897   | 1744   | 1400   |
| Jateng                    | 930507    | 923575 | 925042 | 924307 | 912794 | 1964466| 1918799| 2029891| 1979799|
| Yogyakarta                | 55291     | 55023  | 55126  | 54417  | 53553  | 150827 | 152912 | 159266 | 158903 |
| Jawa Timur                | 1106449   | 1105550| 1102921| 1101765| 1091752| 1926796| 1975719| 2037020| 2072630|
| Banten                    | 197165    | 195951 | 194716 | 190840 | 199492 | 396263 | 362663 | 393704 | 386398 |
| Bali                      | 80164     | 79399  | 78425  | 76655  | 75922  | 152558 | 149000 | 150380 | 142697 |
| NTB                       | 240180    | 246569 | 253208 | 254298 | 264666 | 418062 | 425448 | 438035 | 433712 |
| NTT                       | 144574    | 148610 | 169063 | 172954 | 177238 | 195201 | 200994 | 222469 | 246750 |
| Kalbar                    | 318581    | 322541 | 323959 | 330724 | 444353 | 427798 | 464988 | 452242 | 433944 |
| Kalteng                   | 202237    | 226903 | 225836 | 215545 | 196553 | 214161 | 251787 | 247473 | 242488 |
| Kalsel                    | 457155    | 451869 | 440429 | 431437 | 450152 | 489134 | 496082 | 479721 | 498133 |
| Kaltim                    | 90518     | 90887  | 63323  | 55485  | 57000  | 140215 | 142573 | 109212 | 100026 |
| Kalut                     | -         | 21762  | 21775  | 21448  | -      | -      | -      | -      | -      |
| Sulut                     | 56181     | 56173  | 56157  | 60475  | 55820  | 122108 | 126931 | 127413 | 130438 |
| Sulteng                   | 137786    | 143475 | 146721 | 141448 | 128323 | 221846 | 229080 | 224326 | 219613 |
| Sulsel                    | 576559    | 592376 | 602728 | 623139 | 628148 | 889232 | 981394 | 983107 | 1040024|
| Sultara                   | 85855     | 92280  | 95378  | 96826  | 103812 | 118916 | 124511 | 132945 | 1404048|
| Gorontalo                 | 28707     | 30728  | 32239  | 32116  | 32058  | 32811  | 51193  | 56894  | 62690  |
| Sulbar                    | 55016     | 59020  | 61070  | 62312  | 61292  | 76347  | 83796  | 91195  | 94351  |
| Maluku                    | 14085     | 15972  | 15042  | 13519  | 13394  | 21227  | 20489  | 24399  | 21623  |
| Maluku Utara              | 9093      | 9359   | 10510  | 10516  | 11802  | 16783  | 17794  | 19281  | 21192  |
| Papua Barat               | 7648      | 8330   | 9587   | 9587   | 10126  | 8283   | 7750   | 7523   | 6880   |
| Indonesia                 | 27756     | 27756  | 42350  | 42843  | 44462  | 29262  | 37149  | 41111  | 45493  |

Source: Central Bureau of Statistics

3.1. Centroid Data
Determination of the cluster point is done by taking the largest value (maximum) for the highest level rice production cluster (C1), the average value for the medium level rice production cluster (C2) and the highest value (minimum) for the low level rice production cluster (C3). The point value can be seen in the following Table 3:

### Table 2. Centroid Initial Data

| Cluster          | High cluster | Medium cluster | Low cluster |
|------------------|--------------|----------------|-------------|
|                  | 1101687,4    | 101493,6       | 418         |
|                  | 2032847,2    | 150979,3       | 359,2       |
3.2. Clustering Data

Cluster process by taking the closest distance from each data that is processed. From the data on the amount of rice production by province in 2011-2015, it was found that the grouping in iterations 1 for the 3 clusters. High-level rice production cluster (C1), namely West Java, Central Java, East Java), 10 medium-level rice production cluster provinces (C2) namely (Aceh, North Sumatra, West Sumatra, South Sumatra, Lampung, Banten, West Nusa Tenggara, West Kalimantan, South Kalimantan, South Sulawesi) and 21 low-level rice production cluster provinces (C3) namely (Riau, Jambi, Bengkulu, Kep. Bangka Belitung, Kep.Riau, DKI Jakarta, DI Yogyakarta, Bali, East Nusa Tenggara, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Papua). The process of finding the shortest distance, grouping data on iteration 1 and Clustering data used by Euclidian Distance can be seen in the following table:

Table 3. Calculation of Center Cluster Distance

| Province | Land area | Yields | C1     | C2     | C3     | Distance Shortest | Group |
|----------|-----------|--------|--------|--------|--------|-------------------|-------|
| Aceh     | 300360,6  | 404973,8 | 1.814.414 | 322.585 | 503665,1048 | 322.585 | C2     |
| Sumut    | 442142,8  | 752940,2 | 1.439.848 | 691.664 | 872639,078  | 691.664 | C2     |
| Sumbar   | 227737,4  | 487338,8 | 1.775.496 | 359.270 | 537422,7763 | 359.270 | C2     |
| Riau     | 95664,8   | 124271,6 | 2.157.485 | 27.336  | 156288,9497 | 27.336  | C3     |
| Jambi    | 107322,4  | 145651,4 | 2.133.136 | 7.897   | 180383,963  | 7.897   | C3     |
| Sumsel   | 617895,6  | 807643,6 | 1.317.262 | 835.392 | 1016359,527 | 835.392 | C2     |
| Bengkulu | 89272,6   | 139293,4 | 2.147.214 | 16.909  | 164917,7124 | 16.909  | C3     |
| Lampung  | 363163    | 648887,2 | 1.568.947 | 562.214 | 742821,2941 | 562.214 | C2     |
| Bangka Belitung | 7113,4 | 9063,4 | 2.300.824 | 170.434 | 10981,4155 | 170.434 | C3     |
| Kep. Riau | 418      | 359,2   | 2.311.666 | 181.391 | 0           | 181.391 | C3     |
| Jakarta  | 884,4     | 1580,2   | 2.310.371 | 180.117 | 1307,04623  | 180.117 | C3     |
| Jabar    | 923245    | 1950113,4 | 196.689  | 1.977.918 | 2157116,388 | 1.977.918 | C1     |
| Jateng   | 961629,6  | 1803990,4 | 268.313  | 1.863.405 | 2043774,265 | 1.863.405 | C1     |
| Yogyakarta | 54682    | 155549,2 | 2.149.527 | 47.034  | 164403,5212 | 47.034  | C3     |
| Jawa Timur | 1101687,4 | 2032847,2 | -       | 2.131.153 | 2311666,447 | 2.131.153 | C1     |
| Banten   | 197560,8  | 385287   | 1.879.335 | 253.237 | 432475,08   | 253.237 | C2     |
| Bali     | 78113     | 146409,4 | 2.146.241 | 23.823  | 165430,2691 | 23.823  | C3     |
| NTB      | 251784,2  | 436556,4 | 1.808.447 | 322.710 | 503441,1225 | 322.710 | C2     |
| NTT      | 162527,8  | 226151,2 | 2.036.215 | 96.830  | 277595,7354 | 96.830  | C3     |
| Kalbar   | 325337,6  | 444647   | 1.767.795 | 369.252 | 550422,016  | 369.252 | C2     |
| Kalteng  | 213414,8  | 242118,8 | 1.998.937 | 144.334 | 322201,6301 | 144.334 | C3     |
| Kalsel   | 446208,4  | 494856,6 | 1.671.846 | 486.909 | 667575,3069 | 486.909 | C3     |
| Kalim    | 71442,6   | 17034,2  | 2.175.257 | 45.336  | 136592,6405 | 45.336  | C3     |
| Kalut    | 216616,6667 | 36371  | 2.269.884 | 139.672 | 41810,80139 | 139.672 | C2     |
| Sulut    | 56961,2   | 128863,6 | 2.171.775 | 49.722  | 140394,1391 | 49.722  | C2     |
| Sulteng  | 139550,6  | 220784,4 | 2.051.653 | 79.505  | 260662,9033 | 79.505  | C3     |
| Sulsel   | 604590    | 987557,4 | 1.157.470 | 976.201 | 1157404,031 | 976.201 | C2     |
| Sultara  | 94776,2   | 131432   | 2.151.569 | 20.669  | 161504,021  | 20.669  | C3     |
| Gorontalo | 31169,6   | 56651,2  | 2.247.523 | 117.657 | 64143,97997 | 117.657 | C3     |
| Subar    | 59742     | 87831,8  | 2.206.521 | 75.702  | 105691,971  | 75.702  | C2     |
| Maluku   | 14402,4   | 21775,8  | 2.286.175 | 155.815 | 25578,00225 | 155.815 | C3     |
| Maluku Utara | 10256   | 19297,6  | 2.290.328 | 160.201 | 21341,25672 | 160.201 | C3     |
| Papua Barat | 9055,6   | 7522,8   | 2.301.258 | 170.660 | 11221,13353 | 170.660 | C3     |
| Papua    | 37033,4   | 38873,2  | 2.260.402 | 129.317 | 53141,90371 | 129.317 | C3     |

The K-Means process will continue to operate until the data grouping is the same as the previous grouping of iteration data. The iteration process stops at the second iteration, in the iteration 2 the process of finding the midpoint or centroid value can be found in Table 6 below:

Table 4. Centroid Iteration Data 1

| Group   | High cluster | Medium cluster | Low cluster |
|---------|--------------|----------------|------------|
|         | 1031659      | 1918418,8      | 651,2      |
|         | 201806,2     | 328514,81      | 969,7      |
After getting the midpoint or centroid value, the same process is done by looking for the closest distance. The process of finding the shortest distance, grouping data in iteration 2 and Clustering data can be described in the following tables and figures:

Table 5. Calculation of Center Cluster Iteration Distance 2

| Province   | Land area | Yields | C1   | C2   | C3   | Distance Shortest | Group |
|------------|-----------|--------|------|------|------|-------------------|-------|
| Aceh       | 300.361   | 404.974| 1.680.867 | 124.735 | 503.035.821 | 124.735 | C2       |
| Sumut      | 442.143   | 752.940| 1.306.089 | 487.748 | 871.994.533 | 487.748 | C2       |
| Sumbar     | 227.737   | 487.339| 1.641.426 | 160.927 | 536.770.9415 | 160.927 | C2       |
| Riau       | 95.665    | 124.272| 2.023.623 | 230.177 | 155.662.9138 | 155.662 | C3       |
| Jambi      | 107.322   | 145.651| 1.999.275 | 205.831 | 179.754.1077 | 179.754 | C3       |
| Sumsel     | 617.896   | 807.644| 1.185.336 | 634.582 | 101.573.952  | 634.582 | C3       |
| Bengkulu   | 89.273    | 139.293| 2.013.300 | 220.156 | 164.277.809  | 164.278 | C3       |
| Lampung    | 363.163   | 648.587| 1.435.047 | 358.444 | 742.174.6367 | 358.444 | C2       |
| Bangka Belitung | 7.113 | 9.063 | 2.166.871 | 374.105 | 103.257.0217 | 103.257 | C3       |
| Kep. Riau  | 418       | 359    | 2.177.707 | 385.024 | 653.523.1365 | 653.523 | C3       |
| Jakarta    | 884       | 1.580  | 2.176.411 | 383.739 | 653.523.1365 | 653.523 | C3       |
| Jabar      | 923.245   | 1.950.113| 112.951 | 1.774.840 | 215.646.811 | 112.951 | C3       |
| Jateng     | 961.630   | 1.803.990| 134.156 | 1.659.626 | 204.315.823 | 134.156 | C3       |
| Yogyakarta | 54.682    | 155.549| 2.015.488 | 227.074 | 163.750.2646 | 163.750 | C3       |
| Jawa Timur | 1.101.687| 2.032.847| 134.156 | 1.927.314 | 231.018.583 | 134.156 | C3       |
| Banten     | 197.561   | 385.287| 1.745.340 | 56.931 | 431.825.4018 | 56.931 | C3       |
| Bali       | 78.113    | 146.409| 2.012.279 | 220.142 | 164.781.7854 | 164.782 | C3       |
| NTB        | 251.784   | 436.556| 1.674.551 | 119.041 | 502.795.7407 | 119.041 | C3       |
| NTB        | 162.528   | 226.151| 1.902.408 | 109.641 | 277.327.8593 | 109.641 | C2       |
| Kalbar     | 325.338   | 444.647| 1.634.287 | 169.548 | 549.791.6013 | 169.548 | C2       |
| Kalteng    | 213.415   | 242.116| 1.865.346 | 87.175 | 321.589.4759 | 87.175 | C3       |
| Kalsel     | 440.208   | 494.857| 1.539.347 | 295.63S | 665.165.7601 | 295.63S | C3       |
| Kaltim     | 71.443    | 117.034| 2.041.323 | 248.433 | 135.949.9558 | 135.950 | C2       |
| Kalut      | 21.662    | 36.371| 2.135.930 | 343.220 | 411.663.3936 | 41.167 | C3       |
| Sulut      | 56.961    | 128.864| 2.037.779 | 246.659 | 139.741.4246 | 139.741 | C3       |
| Sulteng    | 139.551   | 220.784| 1.917.764 | 124.425 | 260.022.2022 | 124.425 | C3       |
| Sulsel     | 604.590   | 987.557| 1.024.154 | 772.381 | 115.676.1585 | 772.381 | C3       |
| Sultara    | 94.776    | 131.432| 2.017.689 | 224.270 | 160.872.3946 | 160.872 | C3       |
| Gorontalo  | 31.170    | 56.651| 2.113.565 | 320.978 | 634.964.7377 | 63.496 | C3       |
| Sulbar     | 39.742    | 87.832| 2.072.600 | 279.483 | 105.053.9235 | 105.053 | C3       |
| Maluku     | 14.402    | 21.776| 2.152.223 | 359.457 | 249.397.1228 | 249.397 | C3       |
| Maluku Utara | 10.256 | 19.298| 2.156.368 | 363.740 | 206.921.2656 | 206.921 | C3       |
| Papua Barat | 9.056  | 7.522 | 2.167.312 | 374.418 | 106.566.7192 | 10.657 | C3       |
| Papua      | 37.033    | 38.874| 2.126.492 | 333.230 | 525.398.688 | 52.539 | C3       |

Figure 2. Clustering of Iteration Data 2

3.3. Data analysis
In the iteration 2, data grouping performed on 3 clusters with 2 iterations obtained the same results. Of the 34 rice production data by province in 2011-2015 based on the province it can be seen, 3 provinces of high level rice production cluster (C1) namely (West Java, Central Java, East Java), 10 medium-
level rice production cluster provinces (C2) namely (Aceh, North Sumatra, West Sumatra, South Sumatra, Lampung, Banten, West Nusa Tenggara, West Kalimantan, South Kalimantan, South Sulawesi) and 21 low-level rice production cluster provinces (C3) namely (Riau, Jambi, Bengkulu, Kep. Bangka Belitung, Kep.Riau, DKI.Jakarta, DI.Yogyakarta, Bali, East Nusa Tenggara, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Papua).

4. Conclusions
From the results of these calculations, rice production by province in 2011-2015 results in the lowest 21 rice production provinces, namely (Riau, Jambi, Bengkulu, Kep. Bangka Belitung, Kep.Riau, DKI.Jakarta, DI.Yogyakarta, Bali, East Nusa Tenggara, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua, Papua). Based on this, the government must mobilize existing potentials by maximizing tidal land and swampy swamps and the use of agricultural machinery for the provinces that have the lowest rice production.

References
[1] Hidayat, A., S. Ritunga, and A. M. Fagi.2009. Classification of soil type of rice crop. p 185-214. In Suyamto, I.N. Widiarta, and Satoto (ed). Rice. Technological innovation and Food Security. Book 1. Rice Research Center, Agricultural Research and Development Agency. LIPI Press. Jakarta.
[2] Kathryn T. Morrison, Trisalyn A. Nelson, Farouk S. Nathoo & Aleck S. Ostry. 2012. Application of Bayesian spatial smoothing models to assess agricultural self-sufficiency, International Journal of Geographical Information Science, Volume 26
[3] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C, 2010, Application Of K-Means Clustering Algorithm For Prediction Of Students’ Academic Performance, International Journal Of Computer Science And Information Security, Vol. 7
[4] R.Refianti, A.B. Mutiara, A. Juarna, S.N. Ikhsan (2012) Analysis And Implementation Of Algorithm Clustering Affinity Propagation And K-Means At Data Student Based On Gpa And Duration Of Bachelor-Thesis Completion, Journal of Theoretical and Applied Information Technology, Vol. 35 No.1
[5] Unnati R. Raval, Chaita Jani,2016,Implementing & Improvisation of K-means Clustering Algorithm, International Journal of Computer Science and Mobile Computing, Vol.5, pg. 191-203
[6] Sudesh Kumar, Nancy, Efficient K-Mean Clustering Algorithm for Large Datasets using Data Mining Standard Score Normalization, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 10
[7] Fang, Cheng and Beghin, John C., "Food Self-Sufficiency, Comparative Advantage, and Agricultural Trade: A Policy Analysis Matrix for Chinese Agriculture" (2000). CARD Working Papers. 270.
[8] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” IOP Conference Series: Materials Science and Engineering, vol. 420, no. 12089, pp. 1–8, 2018.
[9] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. P. Windarto, “Application of Classification Method C4 . 5 on Selection of Exemplary Teachers,” in IOP Conference Series, 2018, pp. 1–6.