Clustering Potential Area of Fusarium Oxysporum As A Disease of Garlic

V Purwayoga\textsuperscript{1*}, I S Sitanggang\textsuperscript{1}

\textsuperscript{1}Department of Computer Science, Faculty of Natural Science and Mathematics, IPB University (Bogor Agricultural University), Bogor, Indonesia

*Corresponding author: vega_purwayoga@apps.ipb.ac.id

Abstract. The decrease of garlic production can be caused by several factors, one of the main factors is a disease. The main disease of garlic is \textit{Fusarium Oxysporum} (FO) or known as the rotten base. The decrease in garlic production can be reduced by identifying which areas are likely to be attacked by FO. This study aims to determine clusters of the areas in Central Java Province as the largest producer of garlic in Indonesia which is potentially affected by FO so that this disease can be prevented. This study uses the weather data of 2017 including temperature, humidity, and wind velocity for clustering the potential areas attacked by FO. The algorithm used is K-Means clustering. The best clustering results are obtained at number of clusters K= 10, it was obtained from several experiments to determine the number of K. The most potential month for FO breeding is October. The results show that 37 \% of areas in October 2017 fall into the minimum category, 63 \% FO the area belong to the medium attack category and no one is included in the maximum category due to the fact that daily weather conditions do not fit into that category. October became a potential period because that period was the period with the highest number of regions included in the medium category.

1. Introduction
Domestic garlic production in Indonesia has not been able to meet the public demand for garlic. The garlic demands in 2015 reached 479.8 thousand tons, but the domestic production reached only 16.2 tonnes [1]. It is very inadequate to fulfil the community needs on garlic so that the government imports 479.9 thousand tons to handle that problem. From 2016 until 2019 the government undertakes a central development program and import regulation. It is expected to minimize the amount of imports of garlic. According to the Ministry of Agriculture, the onion commodity target is in 2033. The government’s agenda to increase garlic productivity is to utilize suitable land for planting garlic [2].

The preparation to increase productivity are constrained by several factors including the disease which attacked the garlic plant. In the cultivation of garlic, the rotten base disease is one of the factors which makes the very significant decrease of garlic productivity since 1973. This disease grows at the temperature of 20-24. The Onions which are affected by this disease will certainly rot and it causes the productivity of the onion is not optimal [3]. October, November and December must be avoided, because these months are high rainfall months. High rainfall makes FO easy to live. If planting is done in those months, it is necessary to control the garlic growth from the FO [4]. The control can be done by mapping an area based on the time of the growth of the disease so that the disease can be avoided [5].
Mapping an area based on the time of the growth of the disease can be done by data mining techniques namely clustering. One of the clustering algorithms with good performance is K-Means [6], so this study aims to implement the clustering algorithm namely K-Means to group potential area of FO.

2. Materials and Methods

2.1. Study Area and Data

The study areas are all district located in Central Java province. Central Java is the largest producer of garlic in Indonesia on data from https://www.pertanian.go.id/. The data were obtained by application of agricultural commodities solap and data.online.bmkg. The weather data used are average temperature per month and average humidity per month in 2017.

Table 1. Districts in Central Java Province

| No | District ID | District     | No | District ID | District    |
|----|-------------|--------------|----|-------------|-------------|
| 1  | 11          | Banjarnegara | 19 | 127         | Surakarta City |
| 2  | 14          | Banyumas     | 20 | 132         | Tegal City   |
| 3  | 21          | Batang       | 21 | 140         | Kudus        |
| 4  | 30          | Blora        | 22 | 155         | Magelang     |
| 5  | 35          | Brebes       | 23 | 180         | Pati         |
| 6  | 40          | Cilacap      | 24 | 182         | Pekalongan   |
| 7  | 42          | Demak        | 25 | 183         | Pemalang     |
| 8  | 49          | Grobogan     | 26 | 192         | Purbalingga  |
| 9  | 61          | Jepara       | 27 | 194         | Purworejo    |
| 10 | 63          | Boyolali     | 28 | 196         | Rembang      |
| 11 | 65          | Karanganyar  | 29 | 200         | Semarang     |
| 12 | 70          | Kebumen      | 30 | 213         | Sragen       |
| 13 | 73          | Kendal       | 31 | 216         | Sukoharjo    |
| 14 | 77          | Klaten       | 32 | 234         | Tegal        |
| 15 | 107         | Magelang City| 33 | 235         | Temanggung   |
| 16 | 119         | Pekalongan City| 34 | 246         | Wonogiri     |
| 17 | 122         | Salatiga City| 35 | 247         | Wonosobo     |
| 18 | 124         | Semarang City|    |             |              |

2.2. Sensitivity to Disease

Plants have the severity of the disease caused by both infectious agents (protozoa, bacteria, viruses, etc.) and associated very small vector organisms (mosquitoes, ticks, sandflies, etc.). Their temperature and fluid levels are determined directly by the local climate. Therefore, there is a limited range of climatic conditions - the climate cover - in which each species of infective or vector can survive and reproduce. It is important that the incubation time of the vector-borne infective agent in its vector organism is usually very sensitive to temperature changes, usually, it displays exponential relationships. Other climate sensitivities for agents, vectors, and hosts include levels of rainfall, sea level, wind and solar duration [7].

2.3. Fusarium Oxysporum

*Fusarium Oxysporum* (FO) is a disease-causing pathogen in plant species. This fungus is saprophyte and parasite. This fungus can live long in the form of klamidiospora. This fungus infects the roots, especially on the wounded root. The wilting disease can develop between the temperature 20-29. It
will be more easy developed if the nitrogen is mostly contained in soil [8]. FO living requirements are provided in Table 2.

| Attribute       | Minimum | Medium | Maximum |
|-----------------|---------|--------|---------|
| Temperature (°C)| 20      | 24.5   | 29      |
| Humidity (%)    | 70      | 85     | 100     |

2.4. Interpolation
Interpolation is a method for creating new points that do not yet have values. This value is determined based on previously known data points [9]. The variation of interpolation result depends on the location of a point that has been identified [10]. In this case, the interpolation was done on the data of temperature (°C) and humidity (%) at a weather station.

2.4. K-Means Cluster
The K means algorithm is the most popular and widely used clustering algorithm. It was initially determined how many clusters to be formed. The distribution of the object or the first element in the cluster can be selected as the centroid point of the cluster. The K-Means algorithm then will repeat the following steps until there is stability (no movable objects):
1. Determine the coordinates of the centroid point of each cluster,
2. Determine the distance of each object to the coordinates of the centroid point,
3. Grouping the objects according to their minimum distance [11].

2.5. Euclidean Distance
Euclidean Distance is the most commonly used formula to calculate the similarity of two vectors. Euclidean distance formula is the root of the quadratic difference or the difference in the value of 2 vectors.

\[ \text{dist}(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{in} - x_{jn})^2} \]

Where \( x = \text{objects}, i = (x_{i1}, x_{i2}, ..., x_{in}), j = (x_{j1}, x_{j2}, ..., x_{jn}) \) are two-dimensional data objects [12].

2.5. Cluster Evaluation
The cluster evaluation method used in the study is SSE (Error Sum of Squares). Where the purpose of SSE is to obtain clusters that have the smallest internal variance [13].

\[ \text{SSE} = \sum_{K=1}^{K} \sum_{x_i \in S_K} \|x_i - C_K\|^2 \]

With \( k = \text{many cluster formed} C_i = \text{the i-th cluster, } x = \text{the data present in each cluster}[14]. \)

3. Result and Analysis

3.1. Data Collection
Daily weather data including (temperature (°C) and humidity (%)) were collected from the Meteorology, Climatology and Geophysics Agency (data.online.bmkg) in the 2017 period. Interpolation on temperature (°C) and humidity (%) data was implemented by taking temperature (°C)
and humidity (%) data at the nearest weather station from the centroid of each region. The visualization of the centroid of the polygon is presented in Fig. 1.

Figure 1. Cluster of FO living requirements

3.2. Pre-process
Data preprocessing stages include data selection and handling of missing values. Data selection was done by selecting the map of Central Java Province. Handling of missing values was performed to fill the blank values using the mean value in each of the daily weather data (temperature and humidity).

3.3. Clustering using K-Means
The implementation of the K-Means algorithm uses the R programming language. Number of cluster K = 10 results the best clusters. Results of potential area of FO are presented in Table 3. The map plot of clustering results is shown in Fig. 2. The average value of attributes in each cluster is presented in Table 4. Each cluster has the dominant characteristics as follows.
1. The regions in cluster 1 to cluster 3 generally have a temperature ranges from 25.00 – 25.50 °C; humidity ranges from 75 – 80 %.
2. The regions in cluster 4 to cluster 10 generally have a temperature of ranges from 24.00 – 25.50 °C; humidity ranges from 80 – 84 %.

Table 3. Results of potential area of FO

| Cluster ID | Member of cluster |
|------------|------------------|
| 1          | 2 Purbalingga, Purworejo |
| 2          | 4 Banjarnegara, Kebumen, Kota Magelang, Kota Pekalongan |
| 3          | 11 Batang, Boyolali, Karanganyar, Klaten, Kota Salatiga, Kota Semarang, Kota Surakarta, Semarang, Sragen, Sukoharjo, Wonogiri |
| 4          | 1 Demak |
| 5          | 5 Magelang, Pati, Pekalongan, Temanggung, Wonosobo |
| 6          | 1 Kota Tegal |
| 7          | 1 Banyumas |
| 8          | 3 Grobogan, Jepara, Kudus |
| 9          | 2 Blora, Rembang |
| 10         | 5 Brebes, Cilacap, Kendal, Pemalang, Tegal |
3.4. Cluster Verification Based on FO Living Requirements

Each area was grouped based on weather characteristics in each region. After the area is clustered, the centroid of the cluster is measured using Euclidean distance with predetermined criteria. Attack level category is divided into 3, namely minimum, medium and maximum. The maximum category is the most potential category for FO attacking garlic, while the minimum category is the lowest category for FO live. The results of the verification can be seen in Table 5. The attack level of FO in Central of Java Province is visualized in Fig. 3.

Table 5. Result of cluster verification based on FO living requirements

| Cluster ID | Minimum | Medium | Maximum |
|------------|---------|--------|---------|
| 1          | 0       | 2      | 0       |
| 2          | 0       | 4      | 0       |
| 3          | 0       | 11     | 0       |
| 4          | 1       | 0      | 0       |
| 5          | 0       | 5      | 0       |
| 6          | 1       | 0      | 0       |
| 7          | 1       | 0      | 0       |
| 8          | 3       | 0      | 0       |
| 9          | 2       | 0      | 0       |
| 10         | 5       | 0      | 0       |
Weather characteristics classification and verification of the FO living requirements was carried out every month to find out which months the FO was able to breed. Cluster results and verification of FO living requirements for each month can be seen in Fig. 4. The month with the most attacks by FO was October, there were 22 regions affected by FO.

4. Conclusion
Clustering potential area of FO was successfully implemented based on temperature (°C) and humidity (%) data in each area in Central of Java Province. The results of this study showed that the best number of cluster / K = 10 with a cluster evaluation value is 90.8 %. The number of clusters to produce a good cluster evaluation is 10. October is a month with a potential FO to breed. This was evidenced by the 22 regions that were attacked during the month. The categorization results show that 37% of the total regions in Central Java province fall into the minimum category, 63% in the medium category, and none of them fall into the maximum category.

This result indicates that the daily weather conditions in reality not fully in accordance with FO living requirements, as evidenced by the absence of daily weather conditions in an area that falls into
the Maximum category. So that the review related to the criteria of FO living requierements used is needed. Adjacent areas generally have the same characteristics, s the need to use aspects of spatial proximity. Temperature and humidity data at the weather station Central of Java Province is too far away to be used by all regions in Central of Java Province, so it requires weather stations from other provinces.

References

[1] Hariwibowo PA 2014 Permintaan bawang putih di indonesia (in Bahasa) Habitat 25(2)
[2] Direktur Jenderal Holtikultura 2017 Pengembangan Bawang Putih Nasional (in Bahasa) Kementerian Pertanian, 2017, available at: https://riph.pertanian.go.id/asset/media/download/file/547a6106025e209a3517aa07db2f27b7.pdf
[3] Lopez Y 2012 The effect of climate change on plant diseases Queretaro State University
[4] Wiyatiningsih S 2009 Keparahan Penyakit Moler pada Enam Kultivar Bawang Merah Karena Infeksi Fusarium oxysporum di Tiga Daerah Sentra Produksi (in Bahasa) UPN Veteran Jatim
[5] Nathan B 2013 Diseases of garlic: Various pests Plant Disease Diagnostic Clinic
[6] Nuryaman Y, Asistyasari A, Yudha Y 2018 Komparasi Algoritma K-Mean dan AHC untuk Klasifikasi Curah Hujan di Indonesia Ibraith-Informatika. 29 (2)
[7] Ganie SA 2013 Impact of climate change on plant diseases Sher-e-Kashmir University of Agricultural Sciences and Technology
[8] Hidayat I 2014 Evaluasi antagonisme fusarium oxysporum pada kentang (in Bahasa) Indonesian Institute of Sciences
[9] Wang K 2013 A study of cubic spline interpolation Rivier Academic Journal. 9(2): 1-15
[10] Javari M Comparison of interpolation methods for modeling spatial variations of Precipitation in Iran International Journal Of Environmental & Science Education 12(5): 1037-1054
[11] Singh A 2013 K-means with three different distance metrics Amity University Uttar Pradesh
[12] Hasan M 2015 Comparison of euclidean distance function and manhattan distance function using k-mediods Comilla University
[13] Agnestisia I 2013 Clustering Kondisi Laboratorium Beradasarkan Penilaian Mahasiswa dengan Metode Ward Di Jurusan Teknik Industri UPN “Veteran” Jawa Timur (in Bahasa) UPN Veteran Jatim
[14] Syakur MA 2018 Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster IOP Conference Series: Materials Science and Engineering