Determine the clustering of cities in Indonesia for disaster management using K-Means by excel and RapidMiner

Rienna Oktarina*, Junita

Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480.

Email: rienna.oktarina@binus.ac.id

Abstract. The impact of disasters can disrupt people's lives, both natural and non-natural, resulting in human casualties, environmental damage, property loss, and psychological impact. Besides that, disasters that occur can also cause damage to health facilities, worship, education, and damage to homes, both severely, moderately, and lightly. The impact of disasters is so large, so a logistics warehouse is needed to handle the disaster. One of the countries prone to disasters, Indonesia which has the fourth largest population in the world with 34 provinces and 502 regions or cities. The purpose of this research is to determine the clustering of areas in Indonesia with a very high-risk, high-risk, moderate risk, low risk, and very low risk of disaster based on disaster data in Indonesian National Agency for Disaster Management 2010-2019 using K-Means calculations by Excel and the RapidMiner application. The results of both clustering methods are 6 cities that have a very high-risk index, 79 cities that have a high-risk index, 29 cities that have a medium risk index, 19 cities that have a low-risk index, and 369 cities have a very low-risk index. This result can be considered for the construction of logistics warehouses for disaster management and K-Means method also can be used to know the clustering risk.

Keywords: Disaster, Risk, Clustering, K-Means, RapidMiner

1. Introduction

Disaster according to the Indonesian National Agency for Disaster Management (www.bnpb.go.id) is an event or series of events that disrupt people's lives, both natural and non-natural, resulting in human casualties, environmental damage, property loss, and psychological impact. One of the countries prone to disasters is Indonesia with the potential for natural disasters such as floods, landslides, tornadoes, drought, forest and land fires, earthquakes and tsunamis, and non-natural disasters such as fires, transportation accidents, conflicts, or social unrest.

Based on data from the Indonesian National Agency for Disaster Management (www.bnpb.go.id), natural disasters in Indonesia increased from 3,622 natural disaster events in 2018 to 3,768 natural disaster events in 2019. The most frequent disasters include tornadoes 1,370 events, 764 floods, forest, and land fires up to 746 incidents. In general, the trend of disaster events increased by 12% in 2019. The provinces that were most affected by disasters based on 2019 data were Central Java with 914 incidents, West Java with 691 events, East Java with 612 events, Aceh with 180 incidents, South Sulawesi with 164 incidents. Disasters that occur in Indonesia can result in victims being injured, dead, missing, affected, and displaced. Besides that, disasters that occur can also cause damage to health facilities, worship, education, and damage to homes, both severely, moderately, and lightly.
The impact of a disaster is very large, so disaster management is needed. Disaster management can be overcome by building logistics warehouses in areas with the highest potential disaster risk. The location of the warehouse can be determined by using Data Mining. Data mining is an unsupervised classification method that aims at creating groups of an object, or cluster, in such a way that objects in the same cluster are very similar and objects in different clusters are quite distinct [1]. This research utilizes data mining by using a K-Means method for the main purposes of (a) finding the clustering of the risk index of cities in Indonesia by using calculation K-Means and RapidMiner (b) compare both results to determine the areas in Indonesia that have high disaster risk index.

2. Literature Study

2.1. Data Mining
Data mining is a branch of science that combines databases, statistics, artificial intelligence, and machine learning [2]. Data mining is one of the processes undertaken to discover patterns and knowledge from large data [3]. The ultimate goal of data mining is to obtain important information from raw data [1].

2.2. Clustering
One of the techniques usually used in Data Mining is Clustering. Clustering is a technique of grouping data into different groups by their similar measurement. It means data in the same group that is called cluster are more similar to each other than to those in other groups. The Clustering technique mainly used two algorithms such as hierarchical algorithm and partition algorithm. In the hierarchical algorithm, the dataset is divided into smaller subset in a hierarchical manner whereas in partition algorithm dataset is partitioned into the desired number of sets in a single step [4]. The Partitioning method aims to find clusters contained in the data by optimizing the function of specific objectives to improve the quality of the partition [5].

2.3. K-Means
K-Means is an unsupervised learning algorithm that is used to classify the given dataset that is unlabeled. The goal of this algorithm is to find similar groups represented by variable k (number of clusters). The steps of K-Means Clustering are [6]:
1. Randomly select ‘c’ cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is a minimum of all the cluster centers.
4. Recalculate the new cluster center using:
   \[ Z_t = \frac{1}{\sum_{j=1}^{c_i} x_t} \sum_{j=1}^{c_i} x_t \]
   Where ‘c_i’ represents the number of data points in the cluster.
5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3.

2.4. RapidMiner
RapidMiner is a data mining software, which can be used as a standalone application for data analysis or be integrated as a data mining engine into other products. This tool has unique features such as (IJCSIS) [7]:
- Data integration, analytical Extract Transform Load (ETL), data analysis, and reporting into a single suite;
- Powerful intuitive Graphical User Interface (GUI) for the design of analytical process;
- Repository for a process, data, and metadata management;
- Metadata transformation which results in inspection available during design;
- Support on-the-fly error detection and quick fixes; and
Compete and flexible with hundreds of methods available for data integration, data transformation, modeling, and visualization.

3. Method
The data used in this study are the natural, non-natural and social disasters of each city in Indonesia from 2010 – 2019. The flowchart is as in figure 1.

Figure 1. Flowchart of K-Means using Microsoft Excel and RapidMiner

The steps of this research are:
1. Collect the disaster data in Indonesia from 2010 – 2019. The data include the number of incidents, damage to health facilities, damage to educational facilities, damage to religious facilities, missing victims, injured victims, displaced victims, dead victims, affected victims,
and severely damaged houses, minor and moderate damage. The data include the natural disasters, non-natural disaster and social disasters. Natural disasters such as earthquakes, tsunamis, volcanic eruptions, floods, drought, tornadoes, landslides, abrasion, forest and land fires. Non-natural disasters such as fire and transportation accidents. Social disasters include conflict and acts of terror.

2. Find the average number of disaster data in Indonesia from 2010 – 2019.

3. Determine the number of clusters. The number of the cluster are divide into 5 clusters, such as very high, high, medium, low, and very low.

4. Choose the method using RapidMiner Software or MicrosoftExcel.

5. If using K-Means by Excel:
   5.1 Set the centroid, such as Cilacap, Central Java as the centroid of a very-high risk cluster, KotaKupang, East Nusa Tenggara as the centroid of the high-risk cluster, Sampang, East Java as the centroid of medium risk cluster, Barito Selatan, Central Kalimantan as the centroid of the low-risk cluster, and Yapen, Papua as the centroid of very-low risk cluster.
   5.2 Calculate the distance between centroid with other object using Euclidean distance.

\[ D_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \]

Explanation:
- \( D_{ij} \) = Distance of data in \( i \) to centroid \( j \)
- \( x_{ki} \) = \( i \) data to attribute \( k \) data
- \( x_{kj} \) = Centroid \( j \) to attribute \( k \)

5.3 Find the minimum value of each cluster to know to risk clustering.

5.4 After know the risk category then find the average of each data and repeat the 5.2 methods to find the new centroid.

5.5 Analyze the result of the centroid. If the result is changing back to 5.4 but if the result is not changed, compare both results of RapidMiner and K-Means.

6. If using RapidMiner Software:
   6.1 Open the RapidMiner and drag the Read Excel operators.
   6.2 Input the average disaster data in Indonesia from 2010 – 2019.
   6.3 Find K-Means in clustering operators.
   6.4 Input the number of maximum optimization steps two times and click play to compare both of the results.
   6.5 Analyze the result of the centroid. If the result is changed, must increase the number of maximum optimization steps and back to 6.4 but if the result already the same compare both results of RapidMiner and K-Means.

7. Compare both results of RapidMiner and K-Means.

8. Analyze the result, if the result is the same it means the process finish, and if the result is not the same repeat step 4.

4. Result and Discussion

The result of configuration in RapidMiner as Figure 2.

![Figure 2.Configuration RapidMiner](image_url)
The number of the cluster is 5, with the result of clustering is same until the number of maximum optimization is 8 with the result as Figure 3.

**Cluster Model**

- Cluster 0: 10 items
- Cluster 1: 20 items
- Cluster 2: 369 items
- Cluster 3: 6 items
- Cluster 4: 79 items

**Total number of items: 502**

**Figure 3:** Clustering Result in RapidMiner

The result of iteration 1-10 using RapidMiner is shown in Table 1 below.

**Table 1. Summary of Iteration 1 - 10 using RapidMiner**

| Iteration | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Total |
|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| 1         | 33        | 12        | 279       | 60        | 118       | 502   |
| 2         | 96        | 10        | 146       | 42        | 208       | 502   |
| 3         | 20        | 42        | 353       | 7         | 80        | 502   |
| 4         | 18        | 36        | 355       | 6         | 87        | 502   |
| 5         | 18        | 32        | 358       | 6         | 88        | 502   |
| 6         | 18        | 29        | 362       | 6         | 87        | 502   |
| 7         | 18        | 29        | 366       | 6         | 83        | 502   |
| 8         | 19        | 29        | 369       | 6         | 79        | 502   |
| 9         | 19        | 29        | 369       | 6         | 79        | 502   |
| 10        | 19        | 29        | 369       | 6         | 79        | 502   |

The result of iteration by using Excel is optimized when 11 iterations and show in table 2 below.

**Table 2. Summary of Iteration 1 - 12 using K-Means Excel**

| Iteration | Very High | High | Medium | Low | Very Low | Total |
|-----------|-----------|------|--------|-----|----------|-------|
| 1         | 41        | 71   | 17     | 232 | 141      | 502   |
| 2         | 19        | 60   | 51     | 141 | 231      | 502   |
| 3         | 12        | 71   | 52     | 99  | 264      | 502   |
| 4         | 7         | 81   | 47     | 74  | 293      | 502   |
| 5         | 6         | 87   | 40     | 58  | 311      | 502   |
| 6         | 6         | 89   | 34     | 36  | 337      | 502   |
| 7         | 6         | 89   | 30     | 27  | 350      | 502   |
| 8         | 6         | 84   | 29     | 24  | 359      | 502   |
| 9         | 6         | 84   | 29     | 24  | 359      | 502   |
| 10        | 6         | 84   | 29     | 24  | 359      | 502   |
| 11        | 6         | 79   | 29     | 19  | 369      | 502   |
| 12        | 6         | 79   | 29     | 19  | 369      | 502   |

From Tables 1 and 2, the result of iteration was different by using RapidMiner and K-Means Excel because the centroid used in Excel is random, so more iteration is needed to optimize the clustering. It different from RapidMiner, RapidMiner just needs input the data, find the K-Means operator, and input the number of maximum optimization steps. Below are the percentage result of clustering using a pie chart as shown in Figure 4.
Based on the result in the pie chart, the highest percentage is 73% of cities in Indonesia has very-low risk, 16% of cities in Indonesia has high risk, 6% of cities in Indonesia has medium risk, 4% of cities in Indonesia has low risk and 1% of cities in Indonesia has very high risk.

From the K-Means Excel calculation and RapidMiner, the cluster of each city can be found with the result as shown in table 3 up to table 7.

**Table 3. Very High-Risk Cluster**

| No | City         | Clustering Excel | Clustering Rapid Miner |
|----|--------------|------------------|------------------------|
| 1  | BANDUNG      | Very high        | Very high              |
| 2  | BANDUNG BARAT| Very high        | Very high              |
| 3  | BIMA         | Very high        | Very high              |
| 4  | GORONTALO UTARA| Very high      | Very high              |
| 5  | LOMBOK TENGAH| Very high        | Very high              |
| 6  | RUMBUGANG    | Very high        | Very high              |

**Table 4. High-Risk Cluster**

| No | City         | Clustering Excel | Clustering Rapid Miner |
|----|--------------|------------------|------------------------|
| 7  | ACEH BARAT   | High             | High                   |
| 8  | ACEH JAYA    | High             | High                   |
| 9  | ACEH SELATAN | High             | High                   |
| 10 | ACEH SINGKIL | High             | High                   |
| 11 | ACEH UTARA   | High             | High                   |
| 12 | ALOR         | High             | High                   |
| 13 | BANDAR LAMPUNG| High            | High                   |
| 14 | BANTAR       | High             | High                   |
| 15 | BELOU        | High             | High                   |
| 16 | BONTANDIRING| High             | High                   |
| 17 | BANGIRI      | High             | High                   |
| 18 | BEKASARI     | High             | High                   |
| 19 | BENGKAL      | High             | High                   |
| 20 | BENGKAPAN    | High             | High                   |
| 21 | KAIMA        | High             | High                   |
| 22 | KALENG       | High             | High                   |
| 23 | KAMAS        | High             | High                   |
| 24 | KARANGAN     | High             | High                   |
| 25 | KEPATUKAN    | High             | High                   |
| 26 | KEPING      | High             | High                   |
| 27 | KERIBUN      | High             | High                   |
| 28 | KETAPANG     | High             | High                   |
| 29 | KIRING       | High             | High                   |
| 30 | KOTA BARU    | High             | High                   |
| 31 | KOTA BARU    | High             | High                   |
| 32 | KOTA BENGKAL | High             | High                   |
| 33 | KOTA BULU    | High             | High                   |
| 34 | KOTA KOBAN   | High             | High                   |
| 35 | KOTA KUNING  | High             | High                   |
| 36 | KOTA PASIRUAN| High             | High                   |
| 37 | KUANTAN SENGING | High         | High                   |
| 38 | KUPANG       | High             | High                   |
| 39 | KUTAI BARAT  | High             | High                   |
| 40 | KUTAI BANGAN | High             | High                   |
| 41 | LAMPUNG SELATAN| High          | High                   |
| 42 | LANDAK       | High             | High                   |
| 43 | LANGKAT      | High             | High                   |
| 44 | LEBARAN      | High             | High                   |
| 45 | KAWU UTARA   | High             | High                   |
| 46 | KEDANG       | High             | High                   |

**Table 5. High-Risk Cluster (Continue)**

| No | City         | Clustering Excel | Clustering Rapid Miner |
|----|--------------|------------------|------------------------|
| 47 | MANGGIS      | High             | High                   |
| 48 | MANGGIS TIMUR| High             | High                   |
| 49 | MAMAROS      | High             | High                   |
| 50 | MURI BAYA    | High             | High                   |
| 51 | MURIBAYA     | High             | High                   |
| 52 | NAGA KOTO    | High             | High                   |
| 53 | NIAS UTARA   | High             | High                   |
| 54 | OGAN RAYA    | High             | High                   |
| 55 | PANGANDAAN   | High             | High                   |
| 56 | PATTI        | High             | High                   |
| 57 | PEKALONGAN   | High             | High                   |
| 58 | PEMAUTAI     | High             | High                   |
| 59 | PONTIANAK    | High             | High                   |
| 60 | ROKOKAN      | High             | High                   |
| 61 | SRIWORO      | High             | High                   |
| 62 | SRIWORO      | High             | High                   |
| 63 | SRIWORO      | High             | High                   |
| 64 | TAMAN SARI   | High             | High                   |
| 65 | TAMAN SARI   | High             | High                   |
| 66 | TAMAN SARI   | High             | High                   |
| 67 | TAMAN SARI   | High             | High                   |
| 68 | TAMAN SARI   | High             | High                   |
| 69 | TAMAN SARI   | High             | High                   |
| 70 | TAMAN SARI   | High             | High                   |
| 71 | TAMAN SARI   | High             | High                   |
| 72 | TAMAN SARI   | High             | High                   |
| 73 | TAMAN SARI   | High             | High                   |
| 74 | TAMAN SARI   | High             | High                   |
| 75 | TAMAN SARI   | High             | High                   |
| 76 | TAMAN SARI   | High             | High                   |
| 77 | TAMAN SARI   | High             | High                   |
| 78 | TAMAN SARI   | High             | High                   |
| 79 | TAMAN SARI   | High             | High                   |
| 80 | TAMAN SARI   | High             | High                   |
| 81 | TAMAN SARI   | High             | High                   |
| 82 | TAMAN SARI   | High             | High                   |
| 83 | TAMAN SARI   | High             | High                   |
| 84 | TAMAN SARI   | High             | High                   |
| 85 | TAMAN SARI   | High             | High                   |
| 86 | TAMAN SARI   | High             | High                   |
| 87 | TAMAN SARI   | High             | High                   |
| 88 | TAMAN SARI   | High             | High                   |
| 89 | TAMAN SARI   | High             | High                   |
| 90 | TAMAN SARI   | High             | High                   |
| 91 | TAMAN SARI   | High             | High                   |
| 92 | TAMAN SARI   | High             | High                   |
| 93 | TAMAN SARI   | High             | High                   |
| 94 | TAMAN SARI   | High             | High                   |
| 95 | TAMAN SARI   | High             | High                   |
| 96 | TAMAN SARI   | High             | High                   |
| 97 | TAMAN SARI   | High             | High                   |
| 98 | TAMAN SARI   | High             | High                   |
| 99 | TAMAN SARI   | High             | High                   |
| 100| TAMAN SARI   | High             | High                   |

**Figure 4. Percentage Clustering Result in Pie Chart**

![Pie Chart](image-url)
Table 6. Medium Risk Cluster

| No | City       | Clustering Excel | Clustering Rapid Miner |
|----|------------|------------------|------------------------|
| 1  | BENGKALIS  | Medium           | Medium                 |
| 2  | BIDARA     | Medium           | Medium                 |
| 3  | BOGOR      | Medium           | Medium                 |
| 4  | BOHONEGORO | Medium           | Medium                 |
| 5  | CANDI      | Medium           | Medium                 |
| 6  | CIREBON    | Medium           | Medium                 |
| 7  | GUNUNG KEUL | Medium           | Medium                 |
| 8  | JAKARTA BARAT | Medium       | Medium                 |
| 9  | JAKARTA TIMUR | Medium       | Medium                 |
| 10 | KAMPAR     | Medium           | Medium                 |
| 11 | KARAWANG   | Medium           | Medium                 |
| 12 | KEBUMEN    | Medium           | Medium                 |
| 13 | KOTA CIMAH | Medium           | Medium                 |
| 14 | LOMBOK TIMUR | Medium       | Medium                 |
| 15 | LOMBOK UTARA | Medium       | Medium                 |
| 16 | MANGGARAI BARAT | Medium       | Medium                 |
| 17 | MEDAN      | Medium           | Medium                 |
| 18 | MUSI RAWAS UTARA | Medium     | Medium                 |
| 19 | NGADA      | Medium           | Medium                 |
| 20 | PAMEKASAN  | Medium           | Medium                 |
| 21 | PANDEGLANG | Medium           | Medium                 |
| 22 | PASURUAN   | Medium           | Medium                 |
| 23 | SABU RAJUA | Medium           | Medium                 |
| 24 | SEMARANG   | Medium           | Medium                 |
| 25 | SUMBA BARAT | Medium           | Medium                 |
| 26 | SUMBA TIMUR | Medium           | Medium                 |
| 27 | SUMBAWA    | Medium           | Medium                 |
| 28 | TOLI TOLI  | Medium           | Medium                 |
| 29 | TUBAN      | Medium           | Medium                 |

Table 7. Low-Risk Cluster

| No | City        | Clustering Excel | Clustering Rapid Miner |
|----|-------------|------------------|------------------------|
| 1  | BIREUEN     | Low              | Low                    |
| 2  | BOYOLALI    | Low              | Low                    |
| 3  | BONGGALA    | Low              | Low                    |
| 4  | HALMAHERA SELATAN | Low   | Low                    |
| 5  | JEPARA      | Low              | Low                    |
| 6  | KARO        | Low              | Low                    |
| 7  | KLATEN      | Low              | Low                    |
| 8  | MAGELANG    | Low              | Low                    |
| 9  | MALUKU TENGAH | Low           | Low                    |
| 10 | NAGAN RAYA | Low              | Low                    |
| 11 | PALU        | Low              | Low                    |
| 12 | PENSIR SELATAN | Low           | Low                    |
| 13 | PODE JAYA   | Low              | Low                    |
| 14 | SERAM BAGIAN TIMUR | Low    | Low                    |
| 15 | SLEMBEN     | Low              | Low                    |
| 16 | SINTANG     | Low              | Low                    |
| 17 | SLEMAN      | Low              | Low                    |
| 18 | TARAKAN     | Low              | Low                    |

List of Cities with the result of a very low-risk index based on K-Means Excel and RapidMiner as shown in table 8.

Table 8. Very Low-Risk Cluster

| No | City          | Excel | Rapid Miner |
|----|---------------|-------|-------------|
| 1  | BENGKALIS     | Medium| Medium      |
| 2  | BIDARA        | Medium| Medium      |
| 3  | BOGOR         | Medium| Medium      |
| 4  | BOHONEGORO    | Medium| Medium      |
| 5  | CANDI         | Medium| Medium      |
| 6  | CIREBON       | Medium| Medium      |
| 7  | GUNUNG KEUL   | Medium| Medium      |
| 8  | JAKARTA BARAT | Medium| Medium      |
| 9  | JAKARTA TIMUR | Medium| Medium      |
| 10 | KAMPAR        | Medium| Medium      |
| 11 | KARAWANG      | Medium| Medium      |
| 12 | KEBUMEN       | Medium| Medium      |
| 13 | KOTA CIMAH    | Medium| Medium      |
| 14 | LOMBOK TIMUR  | Medium| Medium      |
| 15 | LOMBOK UTARA  | Medium| Medium      |
| 16 | MANGGARAI BARAT | Medium| Medium    |
| 17 | MEDAN         | Medium| Medium      |
| 18 | MUSI RAWAS UTARA | Medium| Medium |
| 19 | NGADA         | Medium| Medium      |
| 20 | PAMEKASAN     | Medium| Medium      |
| 21 | PANDEGLANG    | Medium| Medium      |
| 22 | PASURUAN      | Medium| Medium      |
| 23 | SABU RAJUA    | Medium| Medium      |
| 24 | SEMARANG      | Medium| Medium      |
| 25 | SUMBA BARAT   | Medium| Medium      |
| 26 | SUMBA TIMUR   | Medium| Medium      |
| 27 | SUMBAWA       | Medium| Medium      |
| 28 | TOLI TOLI     | Medium| Medium      |
| 29 | TUBAN         | Medium| Medium      |
Table 8. Very Low-Risk Cluster (Continue)

| No | City                | Cluster | Risk | Year       | Type of Disaster | No | City                | Cluster | Risk | Year       | Type of Disaster |
|----|---------------------|---------|------|------------|------------------|----|---------------------|---------|------|------------|------------------|
| 61 | BOGOR               | 413     | 2.2  | 2010       | 1.6              | 62 | KOTA DAMAK         | 428     | 2.6  | 2011       | 1.8              |
| 63 | BENGKULUEN         | 435     | 2.6  | 2011       | 2.0              | 64 | SUMSEL             | 448     | 2.8  | 2012       | 2.2              |
| 65 | BANTUL             | 457     | 2.4  | 2012       | 2.4              | 66 | MADIUN             | 468     | 2.4  | 2013       | 2.4              |
| 67 | ENDOG              | 479     | 2.2  | 2013       | 2.2              | 68 | MESTER             | 490     | 2.1  | 2014       | 2.1              |

K-Means method can be used for analyzing the risk of each city besides the calculation of Indonesia’s Disaster Risk Index. The result between Excel and RapidMiner compared with Indonesia’s Disaster Risk Index is different because RapidMiner and Excel use the data from 2010-2019 and the Indonesia’s Disaster Risk Index uses the data from 2015-2018, and the data in Indonesia’s Disaster Risk Index consists of hazards per type of disaster, casualties per type of disaster, and rupiah losses per type of disaster.
disaster, environmental damage (ha) per type of disaster and local government capacity per city. It
different with the K-Means that use the data of the number of incidents, damage to health facilities,
damage to educational facilities, damage to religious facilities, missing victims, injured victims,
displaced victims, dead victims, affected victims, and severely damaged houses, minor and moderate
damage. Besides that, the data of disaster that use are different in Indonesia’s Disaster Risk Index
include the data of floods, earthquakes, tsunamis, volcanic eruptions, forest and land fires, landslides,
abrasion, drought, extreme weather. In K-Means method use the natural disasters, non-natural disaster
and social disasters. Natural disasters such as earthquakes, tsunamis, volcanic eruptions, floods,
drought, tornadoes, landslides, abrasion, forest and land fires. Non-natural disasters such as fire and
transportation accidents. Natural disasters include conflict and acts of terror.

The K-Means and the calculation of Indonesia’s Disaster Risk Index has own strength and
weakness. The K-Means method is more efficient than the calculation of Indonesia’s Disaster Risk
Index because it only takes historical data from disaster events and it doesn’t take long time to find the
risk if the data is needed urgent. But in this case, the K-Means didn’t analyse from the rupiah losses
per type of disaster and local government capacity per city. The strength of Indonesia’s Disaster Risk
Index analyse from the rupiah losses per type of disaster and local government capacity per city but
the data Indonesia’s Disaster Risk Index is only update every five years. The K-Means method can be
used as another alternative to find the risk clustering from other perspective.

5. Conclusion
The clusters are divided into 5, which are very high, high, medium, low, and very low. In this case,
clustering with RapidMiner is recommended because using RapidMiner is faster than Excel to find the
result. In the RapidMiner just need to input the data and K-Means operators until the iteration is
optimized. The number of iteration in RapidMiner is 8 and K-Means Excel is 11. The difference of the
number iteration is because the centroid used in Excel is random, so more iteration is needed to
optimize the clustering.

The result of the cluster is the same both K-Means Excel and RapidMiner with 6 cities that have a
very high-risk index, 79 cities with high-risk, 29 cities with medium risk, 19 cities with low-risk, and
369 cities with very low-risk. The percentage of each risk is the highest is 73% of cities in Indonesia
has very-low risk, 16% of cities in Indonesia has high risk, 6% of cities in Indonesia has medium risk,
4% of cities in Indonesia has low risk and 1% of cities in Indonesia has very-high risk. So, this result
can become another alternative for clustering the risk of cities in Indonesia and the government or
Indonesian National Agency for Disaster Management can consider this result for disaster
management.

References
[1] Bansal A, Sharma M and Goel S 2017 Improved K-Mean Clustering Algorithm for Prediction
Analysis using Classification Technique in Data Mining *Int. J. Comput. Appl.* 157 35–40
[2] Ramadhani M and Fitriah D 2019 Implementation of Data Mining Analysis to Determine the
Tuna Fishing Zone Using DBSCAN Algorithm *9* 706
[3] Virk M and Chauhan V Big Data and Shipping-managing Vessel Performance *2* 73–5
[4] Yadav A and Dhingra S 2016 An Enhanced K-Means Clustering Algorithm to Remove Empty
Clusters *4* 901
[5] Ad zipa K, Bustaman A, Aldila D 2019 The implementation of K-Means partitioning algorithm
in HOPACH clustering method *1* 243
[6] Kumar S and Kaur S 2017 Modified K-Means Clustering Algorithm for Disease Prediction
Abstract : *3* 197
[7] Zainal K, Sulaiman N, Jali M 2015 An Analysis of Various Algorithms For Text Spam
Classification and Clustering Using RapidMiner and Weka *13* 4