K-means clustering on quality of radial run out tires

B. Biantoro, Hernadewita*
Post Graduate Industrial Engineering. University of Mercubuana, Indonesia
Meruya Selatan, West Jakarta 11650, Indonesia
*Email: hernadewita@mercubuana.ac.id

Abstract. The tire industry entered the era of industrial revolution (IR) 4.0. The process of tire production with electronic transactions produces big data that can be exploited in solving quality problems. Stratification techniques commonly used in big data conditions are less effective in mapping product quality conditions. Data handling and subjectivity in product quality classification are inhibitory factors in exploiting data. K means easy grouping of calculations and implementation can be used to map product quality. Radial runs out tire quality using K-mean clustering and elbow method validation shows that the optimum conditions for clustering the radial quality of exhaust tires are three groups. Each group is related to the construction of tires, i.e. radial outlets located at tread splices, sidewall splices and BEC splices. The priority to improve the quality of radial depletion can be suggested through quality mapping based on the results of K-mean cluster analysis, especially on BEC components. The results of the repairs show that the radial value of running out of tires decreased by 12.5%.

Keywords: IR 4.0, big data, RRO, K-Means, tread, sidewall, BEC splices

1. Introduction
The tire industry is one of the industries that supports the advancement of the automotive industry. Tires as in automotive components were played an important role in driving. The function of tires is to hold back the loading of the vehicle, steering the vehicle speed and movements. The result of the study from Lieh [1] was shown that the tires determined the level of comfort while driving.

The tire industry is currently undergoing a change following the 4.0 industrial revolution. The use of information technology e.g. barcode identification, Radio Frequency Identification (RFID) in the production process changes the process conditions significantly Gao et al. [2]. The production process allows the recording of production process transactions to be carried out electronically. Product process records are stored in big data. The massive and lengthy tire production process can generate tens of thousands of transaction data every day. The availability of big data in the production process in the era of the industrial revolution 4.0 has been exploited by many researchers in various ways, including one is the effort to improve product quality. Big data processing, known as data mining, is integrated with existing problem-solving techniques.

Problem solving techniques are very diverse. The most basic problem-solving technique is Plan-Do- Check-Act (PDCA) and it continues to evolve into comprehensive techniques such as Six Sigma. The basic steps in problem solving are the same, which is to map the problem and see the
extent of the difference between the actual situation and the situation it should be. A common technique for mapping problems is data stratification or grouping of quality problem types so that problems from a large scope can be grouped into a simpler scope. This grouping is fundamental to understanding the source of the problem and determining priority corrective actions. Product quality grouping is also very important to look at product quality globally. Stratification techniques on small amounts of data can be done easily and effectively. However, huge number of data (big data) such as the tire manufacturing process is very difficult to do. Obstacles in stratification are huge number of data as well as subjectivity in determining product grouping. Therefore, we need product quality classification methods for efficient and objective big data environment conditions.

To solve the problem of grouping product quality can use grouping techniques. A product grouping method is the grouping of an entity based on process similarities, performance or other factors. Products with a high degree of similarity indicate similarity of product quality and similarity of behavior experienced in the production process. Product integration is needed to understand the mapping of existing quality problems as exemplified by Shankar and Sundararajan [3] in mapping the quality of car engines based on engine torque performance and fuel consumption.

Product quality grouping techniques in the case of tire factory processes that are already based on big data are needed as part to perfect the existing data stratification techniques. This study focuses on the use of grouping techniques in mapping product quality. The object of study is the quality of radial running tires based on radial running tires is tire performance that contributes to driving comfort Kenny [4]. The ultimate goal is to provide a grouping technique that can be used as a support technique for the tire industry in solving quality problems in the industry revolution 4.0 competitions.

2. Literature review

2.1 Radial run out tires

![Splices position](image-url)
Tires are a composite product consisting of several components. Each component has a splices and this contributes to the uniformity of the tires. Therefore, the position of each joint of the tire component is adjusted to a specific position to avoid accumulation and efforts to level the weight of the material in a single tire loop. The position of each material joint can be illustrated in Figure 1.

Radial run out of tires is a measure of how round the tires are from the center. Tire RRO is measured on the inspection machine and reported on the value in millimeter and the location of the highest RRO point. The higher the RRO value, the lower the tire and vice versa the smaller the RRO value, the more round the tire. Meanwhile, the position of the highest point expressed in degrees with zero degrees is the sidewall splices.

2.2 Cluster analysis
Clustering is the grouping of data based on the original level of similarity by maximizing equations within groups and minimizing equations between groups. This technique is used in many areas, such as determining the segmentation of user behavior carried out Abdi and Abolmakarem [5]. The use of quality manufacturing field to classify defect records is done by Chongwatpol [7] or productivity sector by grouping process machines based on their performance Rajesh et al. [6]. There are various methods of grouping and are generally divided into partition, hierarchical, density and grid methods [8]. In partition-K method is the most commonly used method because the calculation is simple, quick to achieve solution and easy to implement [9].

Weaknesses in K-means clustering is that this method is sensitive to data outlier data, determining the number of clusters and handling data that is not globular or convex-shaped clusters that couldn’t find the optimal number [10]. The general procedures to determine the classification of the cluster of K-means method are:

- Determine the value of K at random as the starting point of the cluster
- Calculate the distance between objects with cluster points and data allocation to the nearest cluster point
- Calculate the new average value of the cluster point
- Repeat steps 2 and 3 until the cluster point value does not change

The starting point of a randomly performed cluster determines the accuracy of the cluster result. Research to determine the starting point of a better cluster is done by Zhang and Xia [11] by turning all negative data into positive, sorting and dividing equally according to the number of K. Similarly, performs sorting and division of data Second- two of them show that the grouping process is faster and more effective than random determination. Other values such as medoid are proposed to determine the initial value known as K-medoid as performed by Oktarina et al. [12].

The challenge for the grouping process is to determine the optimal number of K. As stated by Naeem and Wumaier [13], Yuan and Yang [9], Maheswari [14]. They have made several comparisons of K value assessment techniques such as Elbow, Gap Statistics, Silahout Coefficient and Canopy, and others. The results show that each method produces a different optimum K value for each method. Recent research for big data can be assessed using distance between clusters [15]. The elbow method has a better speed in determining the optimal number of clusters, but in certain cases, such as convex clusters, the elbow point is difficult to determine.

3. Methods
This research is conducting through the following stages:

- Data screening should be done before performing cluster analysis since Abdi and Abolmakarem [5] filter data using 3sigma rules before performing cluster analysis.
• Converts RRO data into a vector in an effort to use a tire shape where a 0-degree location equals a 360-degree location.
• Cluster analysis using K method is done in several stages from K value equal to 2 to specific K value. This level is done repeatedly to see the optimal K value with the elbow method.
• Calculate the total error for each cluster value experiment, then evaluate the cluster value and determine the optimal K value.
• Profile the quality of RRO tire products by calculating the average value and standard deviation in each cluster and linking it to the tire structure. From this stage, different treatments can be done in the process so that RRO tire quality problems can be simplified and it is easier to make repairs.

4. Results and discussion
The first step in the data mining process is to clean the data from incomplete data. The initial amount of data for the RRO ban was 5278 sets of data and 4% of the data were incomplete and removed from the data. The next step is to separate the RRO outlier and extreme tire values from the data as these values can affect the accuracy of the data mining prediction classification method. Data allocation of more than 3 standard deviations is excluded from test data using box plot diagrams. Other techniques can also use control charts, e.g. Chongwatpol [7].

Euclidean calculations were used to measure the distance between a data set and a midpoint. This measurement is commonly used by researchers despite having weaknesses in data outliers. By doing pre-screening outliers, Euclidean still produces consistent clusters.

Data transformation is generally done to avoid data with high levels of variation by standardizing it [14]. A more complex transformation is in cluster analysis for text mining data [12] with the aim of standardizing word forms with similar values. In this study, RRO value data transformation was performed to follow the existing product form. Change the value and angle of radial run out (RRO) tire (Figure 2a and Figure 2b). In Figure 2a, if used in cluster analysis, the RRO group at an angle of 0 degrees to 360 degrees will be separated. As for the data that has been changed in Figure 2b, the location of the highest point of the RRO tire follows the shape of the tire so that the group in the area of 0 degrees allows it to be a cluster with clusters at a 360 degree angle. The values of x and y in Figure 2b have a range between 2 and -2. There is no significant difference in value so no further transformation is required. This condition has similar conditions [16].

![Figure 2. Scatter RRO: a) Normal, b) After conversion](image-url)
Determination of the initial value of the cluster center is done by sharing the data equally after compiling. In general, the value of k in the medium dimensional grouping tested is in the range of 3-4 [17] but in studies where the level of multidimensional data, k value test can be tested more than 10 [18]. In this study of RRO quality, the k value test is 2 to 6, which is involving two variables for the clustering process.

The evaluation of the optimal number of clusters is mostly done by researchers, but some studies show that there are differences in the value of optimum clusters in various methods with similar data conditions [13,14]. This raises doubts in choosing cluster values. The elbow assessment method is the simplest and can produce a consistent number of clusters on various amounts of data, therefore this study uses the elbow method to evaluate the optimal number of clusters [19]. The calculation of the sum of error values in each test is as shown in Table 1. The evaluation of the optimal number of clusters based on the elbow method is the number of cluster values with the most difference values and in Table 1 is shown by the number of Clusters 3.

Number of Clusters 3, as in Figure 3, was shown that the RRO data in this concave study are different from Maheswari [14] and Oktarina et al. [12], where the elbow position is not clearly visible. As for the RRO tire grouping, the ideal number of Clusters is 3, and could be used as the default cluster to determine the group of RRO tire quality.

| Number of Cluster | Sum Square Error | Differences on Sum Square Error |
|-------------------|-----------------|-------------------------------|
| 2                 | 2421            | -                             |
| 3                 | 1259            | 1161                          |
| 4                 | 1036            | 223                           |
| 5                 | 841             | 196                           |
| 6                 | 252             | 88                            |

Figure 3. Value of sum square error

The results of grouping with k = 3 and the position in each cluster are as shown in Figure 4. The profile of each cluster has been done to obtain clearer information about the cluster description so that the differences between the clusters can be understood. The finding of cluster differences was comparing cluster values with the specific performance standards or criteria [17,18]. From this
study, they compared the location of the cluster center with the standard tire structure so that it can be found which point is the RRO center of the tire.

The descriptions for each group of RRO tires are shown in Table 3. The center of each cluster on the X and Y axes are as in Table 3. The conversion to RRO angle measurements is 102 degrees, 228 degrees and 328 degrees. This center point is located at the BEC, Tread and Sidewall splices. These three components contribute in influencing the value of RRO compared to other components. The highest average value of RRO is at BEC position 0.98 and standard deviation of 0.34 so that to improve the quality of RRO tires, improvement activities can be focused to improve the quality of BEC components.

![Figure 4. Cluster K=3](image)

| Cluster | Number of Data | Center of Cluster (x, y) | Value of RRO RRO (deg) | Means | Standard Deviation |
|---------|----------------|--------------------------|------------------------|-------|--------------------|
| Cluster 1 | 1505 | (-0.583, -0.655) | 228 | 0.98 | 0.34 |
| Cluster 2 | 2444 | (-0.187, 0.833) | 102 | 0.93 | 0.28 |
| Cluster 3 | 1086 | (0.775, 0.391) | 328 | 0.95 | 0.33 |

In the activity of improving the quality of BEC splices, it is done by reducing the length variation of BEC components and improving the splices method. The effect of improvement can be seen in Table 4 by performing a cluster analysis back to 771 data with the default K = 3 on products that have been increased. Figure 6 shows the cluster results after improvement was made. The center of the cluster shows, before and after the increasing and shows almost the same value. This indicates that the default cluster K = 3 is still consistent, the same results have come out from the study by Putu et al. [19].
**Table 4. The position of cluster RRO after improvement of BEC Components**

| Cluster | Number of Data | Center of Cluster (x, y) | Value of RRO Means RRO (deg) | The position closest to the center of the RRO on the tire structure |
|---------|----------------|--------------------------|-------------------------------|---------------------------------------------------------------|
| Cluster 1 | 169            | (-0.542, -0.599)          | 169                           | 0.86 0.25 Splices of BEC                                     |
| Cluster 2 | 423            | (-0.176, 0.833)           | 102                           | 0.91 0.25 Splices of Tread                                    |
| Cluster 3 | 179            | (0.685, -0.363)           | 326                           | 0.84 0.24 Splices of Sidewall                                  |

**Figure 6. Cluster K=3 after improvement**

The change in mean value in cluster 1 was occurred, 12.5% compared to the mean value before increase. Other cluster points also have the devaluation in mean and it is pointing that BEC is part of the components that interact in a single product.

**5. Conclusions**

From the data analyse and discussion, it can be concluded:

- In performing cluster analysis, it is necessary to converse the data that can describe the condition of the product being studied, in this case for tires, can be done by changing the value and angle of the RRO into a vector value so that the form of data dissemination resembles the actual situation.
- RRO tire cluster analysis can be divided into 3 clusters and closely related to tire structure, especially the three main components, namely BEC, Tread and Sidewall.
- The formed clusters remain consistent even when tested with different sample numbers or at different product periods if the natural condition of the product is the same.

Further study can be done by looking at whether the existing clusters are still consistent and if the tire structure is different because each tire manufacturer has different standards. Cluster analysis as an initial product description can be combined with other methods such as to obtain other factors that cause the quality of RRO.
6. Acknowledgement
This research is supported by Ministry of Education/Head of the National Research and Innovation Agency Republic of Indonesia through Hibah Penelitian Tesis Magister (PTM) 2020, Contract No SK/26/E1/KPT/2020 and Institution of Research and Community Service University of Mercubuana, Jakarta.

7. References
[1] Lieh, J. (2009). Tire Damping Effect on Ride. Journal of Vibration and Acoustics, 131(June), 1–6. https://doi.org/10.1115/1.3085878
[2] Gao, Y., Yang, D., & Ning, W. (2010). Research on Manufacturing Process Traceability in Tire Enterprise. Applied Mechanic and Material, 47(December), 485–488.
[3] Shankar, R., & Sundararajan, M. (2015). Manufacturing Process Optimization using Data Mining Techniques. International Journal of Data Mining Techniques and Applications, 4(March), 1–6.
[4] Kenny, T. M. (1989). Quantifying Tire, Rim and Vehicle Effect on Ride Quality. In International Congress and Exposition (pp. 1–9). Detroit, Michigan, USA: SAE.
[5] Abdi, F., & Abolmakarem, S. (2018). Customer Behavior Mining Framework (CBMF) using clustering and classification techniques. Journal of Industrial Engineering International, 3. https://doi.org/10.1007/s40092-018-0285-3
[6] Rajesh, K. V. D., Krishna, M. M., Ali, A., & Chalapathi, P. V. (2017). A Modified Hybrid Similarity Coefficient Based Method for Solving the Cell Formation Problem in Cellular Manufacturing System. Materials Today: Proceedings, 4(2), 1469–1477. https://doi.org/10.1016/j.matpr.2017.01.169.
[7] Chongwatpol, J. (2015). Prognostic analysis of defects in manufacturing. Industrial Management and Data Systems, 115(1), 64–87.
[8] Han, J., Kamber, M., & Pei, J. (2012). Data mining: Data mining concepts and techniques. USA: Elsevier.
[9] Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. Multidisciplinary Scientific Journal, 2(16), 226–235.
[10] Shukla, S.S.N. (2014). A Review on K-means Data Clustering Approach. International Journal of Information & Computation Technology, 4(17), 1847–1860. Retrieved from http://www.irphouse.com
[11] Zhang, K., Li, G., Gong, J., & Zhang, M. (2016). Research on Assembly Data Mining Technology of Complex Mechanical System. Procedia CIRP, 44, 97–101.
[12] Oktarina, C., Notodiputro, K. A., & Indahwati, I. (2020). Comparison of K-Means Clustering Method and K-Medoids on Twitter Data. Indonesian Journal of Statistics and Its Applications, 4(1), 189–202. https://doi.org/10.29244/ijlsa.v4i1.599.
[13] Naeem, S., & Wumaier, A. (2018). Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K. International Journal of Computer Applications, 182(31), 7–14. https://doi.org/10.5120/ijca2018918234.
[14] Maheswari, K. (2019). Finding Best Possible Number of Clusters using K-Means Algorithm. International Journal of Engineering and Advanced Technology, 9(1S3), 533–538. https://doi.org/10.35940/ijeat.a1119.1291s419.
[15] Alibuhitto, M. C., & Mahat, N. I. (2020). Distance based k-means clustering algorithm for determining number of clusters for high dimensional data. Decision Science Letters, 9(1), 51–58. https://doi.org/10.5267/j.dsl.2019.8.002
[16] Pascal, C., Ozuomba, S., & Kalu, C. (2015). Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. International Journal of Advanced Research in Artificial Intelligence, 4(10). https://doi.org/10.14569/ijarai.2015.041007.
[17] 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, 7*(1), 292–295. Retrieved from http://arxiv.org/abs/1002.2425

[18] Widiyaningtyas, T., Prabowo, M. I. W., & Pratama, M. A. M. (2017). Implementation of k-means clustering method to distribution of high school teachers. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2017–Decem(September), 19–21.* https://doi.org/10.1109/EECSI.2017.8239083

[19] Putu, N., Merliana, E., & Santoso, J. (2017). Analisa penentuan jumlah cluster terbaik pada metode K means Clustering. In *Prosiding Seminar Nasional Multi Disiplin Ilmu UNISBANK* (pp. 978–979).