

ANALYSIS AND IMPLEMENTATION DATE MINING OF K MEAN CLUSTERING MODEL FOR CONSTRUCTION SERVICE PROVIDERS IN ASSOCIATION OF BANTEN PROVINCE CONTRACTORS

Tubagus Dian Hardiansya¹, Dana Santoso², and Erry Rimawan³

¹²³*Industrial Engineering, Faculty of Engineering, MercuBuana University, Jakarta, Indonesia*

Abstract

The executor of construction services in this case is the contractor in which the contractor consists of small, medium and large classifications. The contractor is an institutional institution that acts as the executor of infrastructure project activities; the contractor consists of construction consultant services and construction implementing services. The unitary state of the Indonesian republic in the last five years has been very intensively developing infrastructure by the government. In this study trying to make a clustering model with k means clustering from construction providers in the contractor association using the WEKA application so that experts can be obtained for cluster 2, as many as 33 Provision of construction services from 98 Provision of construction services (33%), grade for cluster 4, 31 Provision of construction services from 98 Provision of construction services (30%).

Keywords: *Date Mining, Clustering, K means, WEKA, Clustering, Construction service, Contractor.*

1. Introduction

Development of infrastructure in the last 10 years has been very rapid and the importance of the construction industry for the national economy can be seen from the following impacts (Hillebrandt, 2001). With the condition of infrastructure building being a contractor, and construction service provider companies growing very rapidly (Soeparto&Trigunansyah, 2005) (A Malik 2010), this is certainly also inseparable from the role of construction service companies / contractors in terms of implementing on projects the project The construction industry has a complex and dynamic environment (Wilson, 1975). It consists of various professional and interrelated industries (Tanesia, Syriac, Yudha, &Ramba, 2017) for this requires contractor productivity (Utomo, 2014). The construction industry is very competitive in the construction market (Safa et al., 2015). Market share of the construction business (Tanesia et al., 2017).

As it has been regulated in Indonesian regulations through Law No.18 of 1999 and elaborated by the derivation of Government Regulation No. 28/2000 concerning Business and Role of Construction Service Communities, and Government Regulation of the Republic of Indonesia No. 29/2000 concerning Implementation of Construction Services, and explanation of Government Regulation of the Republic of Indonesia Number 30 Year 2000 Concerning Implementation of Construction Services Development, the first amendment to Government Regulation Number 04 of 2010 concerning Amendment to Government Regulation Number 28 of 2000 concerning Business and Role of Construction Service Communities, Government Regulation of the Republic of Indonesia Number 59 of 2010 concerning Amendment to Government Regulation Number 29 of 2000 concerning the Implementation of Construction Services, Government Regulation Number 92 of 2010 concerning the Second Amendment to Government Regulation Number 28 of 2000 concerning the Business and Role of Construction Services

Communities, as well as the latest amendment to Law No. 2 of 2017. Regulations no.03 construction services development agency 2017 regarding certification and registration of implementing construction services.\

II. LITERATURE REVIEW

2.1. Competitiveness

Companies that have no competitiveness will be abandoned by the market (Wethyavivorn, Charoenngam, & Teerajetgul, 2009). Competitiveness is related to the efficiency and effectiveness of an organization or company in the competitive market, compared to other organizations that offer the same or similar products or services (Kaming, Raharjo, & Swantari, 2017).

company (business unit) costs and strategies covering the expenditure of raw materials, labor, storage, capital and human resources including remuneration, training, career development (Luu, Kim, Cao, & Park, 2008) (Wethyavivorn et al., 2009) (Lu, Ye, Flanagan, & Jewell, 2013).

2.2. Data mining

Data mining is an analysis of data to find a clear relationship and concludes what has not been known before (Schuh et al., 2019) (Degree, Science, & Lecture, 2012) (Ian H, nd) (Jiménez, Anupol, Cajal, & Gervilla, 2018) in the latest way understood and useful for owners Data mining Data mining is the method used for extracting. predictive information is hidden in the data base (Degree et al., 2012). In data mining there are two types of grouping methods used in data grouping are hierarchical clustering and non-hierarchical clustering (Singh & Singh, 2012). Data mining is a process of finding patterns in a database with certain methods (Degree et al., 2012). Clustering or classification is a method used to divide data sets into groups based on similarities (Erwinsyah, Chin, Palaloi, Nguyen, & Shankar, 2019) that have been determined before (Bala, Sikka, & Singh, 2014). Clustering is analyzing a group of objects and grouping these objects into clusters based on similarity values. Clustering is a data mining method with an unsupervised approach (unsupervised learning) that does not require output targets. Data mining is the process of extracting information that is used to search for unknown patterns from large data sets (Jiménez et al., 2018). Objects in a cluster have similar data and will be very different from data in other cluster objects (Ng & Han, 1994).

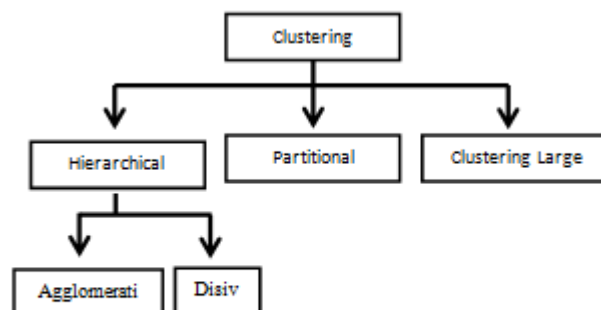


Figure 2.1. Clustering Algorithm Category

III. K-Means algorithm

K-Means algorithm is an unsupervised classification method that divides data items into one or more clusters (Novaliendry, Hendriyani, Yang, & Hamimi, 2015). K-means clustering is one of the non-hierarchical data grouping methods that classify data in the form of one or more groups / groups. (Verma, Srivastava, Chack, Diswar, & Gupta, 2012). Data that has the same characteristics are grouped in one group / group and data have different characteristics, data are grouped with other groups so that data in one cluster has a small level of variation (Borhade & Mulay, 2015).

According to (Teknik & Jambi, 2018) the steps of grouping by the K-Means method are as follows: a) Select the number of clusters k. b) Initialization of this cluster center can be done in various ways. But what is most often done is randomly. The central group is given a initial value with a random number. c) Allocate all data objects to the nearest cluster. The closeness of two objects is determined by the distance of both objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data center cluster and the top 10 algorithms (Wu et al., 2008).

K-Means algorithm is considered quite efficient (Novaliendry et al., 2015). K-Means in this study is used to classify (Ramadhani, 2014), to improve document representation by applying the text mining algorithm (Valls et al., 2018) (Prilianti & Key, 2014). This approach is also by applying the k-Means clustering approach, the implementation of the K-Means algorithm is also used to apply the design of stock prices and commodity prices, K-Means clustering can also be used to group the serum (Köttgen et al., 2013).

IV. Analysis

3.1. Results and Analysis

Testing analysis, it is important to determine whether the analysis results are in accordance with the expected decisions. To test the correctness of data processing done manually, it can use one of the Weka 3.8.2 application software.

3.2. Data Testing

The test data used is a table that has the following components:

1. Has the attributes of the company name, company location, finance, experts, fields, experience, grade and experts.
2. Number of interactions 98

3.3. Implementation step

All variables consist of attributes that are used to determine the strategy of competitiveness of construction service providers in the contractor association kept at Microsoft under the name Data Construction Company M.xls (which contains cases or criteria in generating rules). Next, it is the process of data transformation where the M.xls construction company data is then stored with CSV extension, then the file is opened with notepad or other text editors and the data has changed to a comma-restricted format. Then the data is adjusted by adding preliminary information and the data can already be used as input in Weka 3.8.2. Next in the Weka application, click Explorer, Open File (select CSV file), click Select, choose Cluster and click Start. At this stage the data mining process is carried out by selecting an algorithm that will be used to produce the centroid cluster.

3.4. Implementation and Results

3.4.1. Centroid cluster value and cluster instance

Figure 1 shows the results of the centroid cluster and cluster instance with Weka.

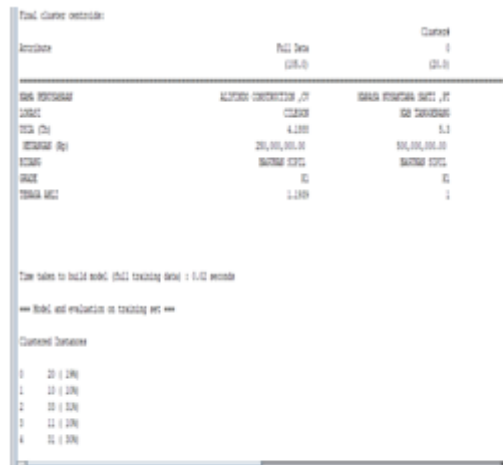


Fig. 1: Centroid Cluster Results and Clustered Machines with Weka.

Figure 2 shows the results of the centroid cluster and cluster instance with Weka.

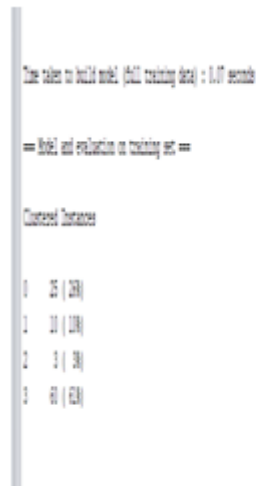


Fig. 2: Centroid Cluster Results and Clustered Machines with Weka.

Time taken to build model (full training data) : 0.02 seconds

==== Model and evaluation on training set ====

Clustered Instances

0 20 (19%)

1 10 (10%)

2 33 (31%)

3 11 (10%)

4 31 (30%)

Testing the data with Weka Software based on the reasons Providing construction services produces data forms:

1) Provision of construction services that have competitiveness from company locations for cluster 0, such as 20 Provision of construction services from 98 Provision of construction services (19%).

2) Provision of construction services having grade competitiveness for cluster 1, 10 Provision of construction services from 98 Provision of construction services (10%).

3) Provision of construction services that have competitiveness from experts for cluster 2, as many as 33 Provision of construction services from 98 Provision of construction services (33%).

4) Provision of construction services that have competitiveness from the age of the company to cluster 3, 11 Provision of construction services out of 100 Provision of construction services (10%).

5) Provision of construction services that have competitiveness from finance to cluster 4, 31 Provision of construction services from 98 Provision of construction services (30%).

Then cluster 3 with those with financial competitiveness and cluster 5 with those with competitiveness from experts and grade can be used as a competitiveness strategy. Graph grouping and plot graph grouping.

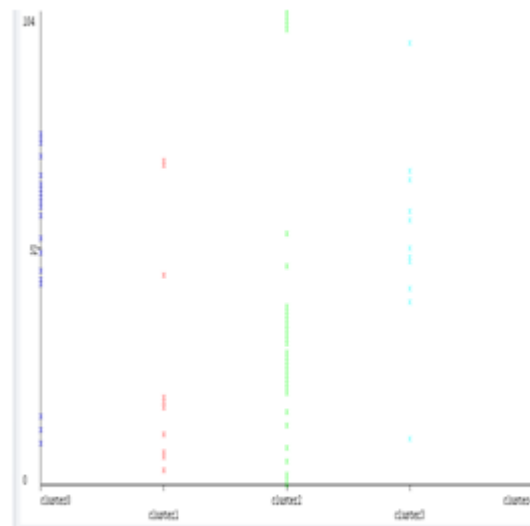


Fig. 3: Graph of Clustering Results with Weka.

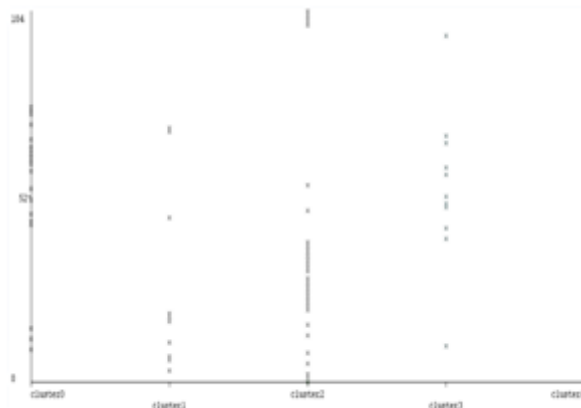


Fig. 4: Plot Graph Clustering with Weka

The clustering graphs and clustered plot plots showing the strategic exposures of competitiveness as shown in Figure 3 and Figure 4. Strategic exposure to competitiveness (Utomo, 2014) (Tanesia et al., 2017) (Kazaz&Ulubeyli, 2009) must be considered in particular as an attempt to advance that enhanced human and financial resources (Wethyavivorn et al., 2009) (Luu et al., 2008) (Sahin, Ulubeyli, &Kazaza, 2015) (Ling, Pham, & Hoang, 2009) by adopting innovation technology (Smyth, 2004) to improve performance personalization (Deng, Liu, & Jin, 2013) (Lu et al., 2013) (Luu et al., 2008).

V. Conclusion

The results of this study found that there are two competitiveness strategies that can be applied by construction service providers (Contractors). They are:

1. Conduct an Expert strategy by increasing the number of skilled personnel and giving bonuses to experts to have good competitiveness for the company.
2. Company grade determines the competitiveness of construction service providers where contracting companies must apply their companies by making continuous innovations.
3. Conducting strategies in finance by cooperating with banks so that capital is sufficient.

References

- [1] A Malik 2010 Pengantar Construction Services Business Services. Yogyakarta Publisher Andi.
https://books.google.co.id/books?hl=en&lr=&id=pTA_WAh3aXEC&oi=fnd&pg=PR3&dq=related:VXfRj5QOf0oJ:scholar.google.com/&ots=-oolahyv4g&sig=KJLcjmCNhbuty_cq_&q=
- [2] Bala, R., Sikka, S., & Singh, J. (2014). A Comparative Analysis of Clustering Algorithms. *International Journal of Computer Applications*, 100(15), 35–39. <https://doi.org/10.5120/17603-8293>
- [3] Borhade, M., & Mulay, P. (2015). Online interactive data mining tool. *Procedia Computer Science*, 50, 335–340. <https://doi.org/10.1016/j.procs.2015.04.039>
- [4] Construction Services Development Agency Regulation No. 03 of 2017 concerning certification and registration of implementing construction services.
- [5] Degree, M. M., Science, C., & Lecture, A. C. (2012). *Data Mining : Concepts and*. 05, 703. <https://doi.org/10.1093/nar/gku1019>

- [6] Deng, F., Liu, G., & Jin, Z. (2013). Factors formulating the competitiveness of the chinese construction industry: Empirical investigation. *Journal of Management in Engineering*, 29(4), 435–445. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000161](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000161)
- [7] Erwinsyah, A., Chin, J., Palaloi, I. A., Nguyen, P. T., & Shankar, K. (2019). Multigroup classification using privacy preserving data mining. *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue), 922–926. <https://doi.org/10.35940/ijeat.F1174.0886S19>
- [8] Hughes, Will, Hillebrandt, Patricia, Lingard, Helen and Greenwood, David (2001) The impact of market and supply configurations on the costs of tendering in the construction industry. In: CIB World Building Congress, 2-6 April 2000, Wellington. This version was downloaded from Northumbria Research Link: <http://nrl.northumbria.ac.uk/32130/>
- [9] Ian H, W. & E. F. (n.d.). Data mining-1.pdf.
- [10] Jiménez, R., Anupol, J., Cajal, B., & Gervilla, E. (2018). Data mining techniques for drug use research. *Addictive Behaviors Reports*, 8(May), 128–135. <https://doi.org/10.1016/j.abrep.2018.09.005>
- [11] Kaming, P., Raharjo, F., & Swantari, P. I. (2017). Kajian daya saing kontraktor besar indonesia. 26–27.
- [12] Kazaz, A., & Ulubeyli, S. (2009). Strategic management practices in turkish construction firms. *Journal of Management in Engineering*, 25(4), 185–194. [https://doi.org/10.1061/\(ASCE\)0742-597X\(2009\)25:4\(185\)](https://doi.org/10.1061/(ASCE)0742-597X(2009)25:4(185))
- [13] Köttgen, A., Albrecht, E., Teumer, A., Vitart, V., Hundertmark, C., Pistis, G., ... Conall, M. (2013). NIH Public Access associated with serum urate concentrations. 45(2), 145–154. <https://doi.org/10.1038/ng.2500>. Genome-wide
- [14] Ling, F. Y. Y., Pham, V. M. C., & Hoang, T. P. (2009). Strengths, weaknesses, opportunities, and threats for architectural, engineering, and construction firms: Case study of vietnam. *Journal of Construction Engineering and Management*, 135(10), 1105–1113. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000069](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000069)
- [15] Lu, W., Ye, K., Flanagan, R., & Jewell, C. (2013). Developing construction professional services in the international market: SWOT analysis of China. *Journal of Management in Engineering*, 29(3), 302–313. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000144](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000144)
- [16] Luu, T. Van, Kim, S. Y., Cao, H. L., & Park, Y. M. (2008). Performance measurement of construction firms in developing countries. *Construction Management and Economics*, 26(4), 373–386. <https://doi.org/10.1080/01446190801918706>
- [17] Ng, R. T., & Han, J. (1994). Efficient and Effective Clustering Methods for Spatial Data Mining. *Proceedings of the 20th International Conference on Very Large Data Bases*, 144–155. Retrieved from ftp://ftp.cs.ubc.ca/.snapshot/sv_weekly.2/local/techreports/1994/TR-94-13.pdf
- [18] Novaliendry, D., Hendriyani, Y., Yang, C. H., & Hamimi, H. (2015). The optimized K-means clustering algorithms to analyzed the budget revenue expenditure in Padang. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2(August), 61–66.
- [19] Patil, Swati S., and P. P. Bhangale. "Overhead Cost in Construction Industry." *International Journal of Industrial Engineering & Technology (IJJET)* 4 (2014): 1-6.

- [20] Prilianti, K. R., & Kunci, K. (2014). Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering. 2(1), 1–6.
- [21] Republic of Indonesia Law Number 18 Year 1999 concerning construction services.
- [22] Republic of Indonesia Government Regulation Number 28 of 2000 Concerning Business and Role of Construction Services Communities.
- [23] Government Regulation of the Republic of Indonesia Number 29 Year 2000 Regarding the Implementation of Construction Services.
- [24] Republic of Indonesia Government Regulation Number 30 Year 2000 Concerning Implementation of Construction Services Development.
- [25] Government Regulation of the Republic of Indonesia Number 04 of 2010 concerning Amendments to Government Regulation of the Republic of Indonesia Number 28 of 2000 concerning Business and Role of Construction Services Communities.
- [26] Government Regulation of the Republic of Indonesia Number 59 of 2010 concerning Amendment to Government Regulation of the Republic of Indonesia Number 29 of 2000 concerning the Implementation of Construction Services.
- [27] Government Regulation of the Republic of Indonesia Number 92 of 2010 concerning the Second Amendment to Government Regulation of the Republic of Indonesia Number 28 of 2000 concerning Business and Role of the Construction Services Community.
- [28] Hasyim, Muhammad. "Myth And Ideology Construction In Indonesia Television Advertising: A Semiotic Based Approach." *Jurnal: International Journal of Communication and Media Studies (IJCMS) ISSN (P) 5.1* (2015).
- [29] Prabhu, P. Ganesh, and D. Ambika. "Study on behaviour of workers in construction industry to improve production efficiency." *International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development* 3.1 (2013): 59-66.
- [30] Republic of Indonesia Law No.2 Year 2017 Construction services.
- [31] Ramadhani, R. D. (2014). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. *Industrial Marketing Management*, 1(1), 1–9. <https://doi.org/10.1016/j.indmarman.2016.05.016>
- [32] Safa, M., Shahi, A., Haas, C. T., Fiander-McCann, D., Safa, M., Hipel, K., & MacGillivray, S. (2015). Competitive intelligence (CI) for evaluation of construction contractors. *Automation in Construction*, 59, 149–157. <https://doi.org/10.1016/j.autcon.2015.02.009>
- [33] Sahin, S., Ulubeyli, S., & Kazaza, A. (2015). Innovative Crisis Management in Construction: Approaches and the Process. *Procedia - Social and Behavioral Sciences*, 195, 2298–2305. <https://doi.org/10.1016/j.sbspro.2015.06.181>
- [34] Schuh, G., Reinhart, G., Prote, J. P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data mining definitions and applications for the management of production complexity. *Procedia CIRP*, 81, 874–879. <https://doi.org/10.1016/j.procir.2019.03.217>
- [35] Singh, N., & Singh, D. (2012). Performance Evaluation of K-Means and Heirarichal Clustering in Terms of Accuracy and Running Time. *International Journal of Computer Science and Information Technologies*, 3(3), 4119–4121.

- [36] SINGH, JAGMEET, JASPAL SINGH, and MANPREET KAUR."APPLICATIONS OF NANOTECHNOLOGY IN CONSTRUCTION INDUSTRY."International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development (IJCSEIERD) 5. 6, Dec 2015, 31-34
- [37] Singh, D. E. V. I. N. D. E. R., and J. A. S. P. A. L. Singh."Use of agrowaste in concrete construction." Int J Environ Ecol Family Urban Stud 6.1 (2016): 119-130.
- [38] Smyth, H. (2004). Competencies for Improving Construction Performance: Theories and Practice for Developing Capacity. International Journal of Construction Management, 4(1), 41–56. <https://doi.org/10.1080/15623599.2004.10773050>
- [39] Soeparto, H. G., & Trigunansyah, B. (2005). Industri konstruksi Indonesia: Masa depan dan tantangannya. (January 2005), 1–9. <https://doi.org/10.13140/2.1.1801.8568>
- [40] Tanesia, R. K., Suryani, D., Yudha, F. M., & Ramba, J. (2017). KAJIAN SEGMENTASI PASAR DAN STRATEGI PEMASARAN USAHA JASA KONSTRUKSI DI NEGARA-NEGARA ASEAN (Studi Kasus : Indonesia vs Filipina). Jurnal Teknik Sipil, 13(3), 216. <https://doi.org/10.24002/jts.v13i3.878>
- [41] Teknik, F., & Jambi, U. A. (2018). Imti Tsalil Amri , Santoso. 1(2), 25–30.
- [42] Udhayakumar, R., and P. Karthikeyan. "Integrating Project Management Tools For Augmenting The Profession Of Civil Engineers In M/S Urc Construction (P) Ltd., India." Trans Stellar, International Journal of Civil, Structural, Environmental, Infrastructure Engineering Research and Development (2014).
- [43] Utomo, P. (2014). Kesiapan Sumber Daya Manusia (Tenaga Kerja) Bidang Konstruksi di Indonesia Menghadapi Masyarakat Ekonomi Asean. Urnal Ilmiah Ilmu Hukum Qistie, 7(November), 85–97.
- [44] Valls, F., Redondo, E., Fonseca, D., Torres-Kompen, R., Villagrasa, S., & Martí, N. (2018). Urban data and urban design: A data mining approach to architecture education. Telematics and Informatics, 35(4), 1039–1052. <https://doi.org/10.1016/j.tele.2017.09.015>
- [45] Verma, M., Srivastava, M., Chack, N., Diswar, A. K., & Gupta, N. (2012). A Comparative Study of Various Clustering Algorithms in Data Mining. International Journal of Engineering Research and Applications Www.Ijera.Com, 2(3), 1379–1384.
- [46] Wethyavivorn, P., Charoenngam, C., & Teerajetgul, W. (2009). Strategic assets driving organizational capabilities of thai construction firms. Journal of Construction Engineering and Management, 135(11), 1222–1231. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000091](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000091)
- [47] Wilson, A. (1975). Economic Theory and the Construction Industry. Journal of the Operational Research Society, 26(3), 574–575. <https://doi.org/10.1057/jors.1975.125>
- [48] Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H, Steinberg, D. (2008). Top 10 algorithms in data mining. In Knowledge and Information Systems (Vol. 14). <https://doi.org/10.1007/s10115-007-0114-2>.