**Clustering Algorithms in Relation**

**to Word Analysis to Understand Context –**

link: https://www.iasj.net/iasj/download/9f2c8448d0838c39

Jephone Torre

**"Analysis of H-index and Papers Citation in Computer Science Field using K-Means Clustering Algorithm" by Omar Ibrahim Obaid focuses on analyzing the H-index and paper citations in the field of computer science. “**

ABSTRACT: This paper provides an analysis of H-index and paper citations in the computer science field using K-Means Clustering Algorithm. By leveraging cutting edge visual analytics through the use of Power BI and Orange Data Mining tool with K-Means clustering algorithm, we are able to present a comprehensive review of how H-index and citations impact the scholarly evaluation of authors in computer science field. The analysis obtained will assist academics seeking to expand their research influence while providing additional context to the general audience seeking to understand how machine learning algorithms can contribute to data exploration. The analysis conducted has revealed that area of research which has the highest value of citation and H-index include the following topics: Artificial-Intelligence, Computational-Intelligence, Data-mining, Evolutionary-Algorithms, and Big Data Analytics. Finally, the analysis of this research clearly demonstrated that paper citations remain an important factor for determining scientific impact in disciplines such as computer science given its close relationship with established metrics such as h-index scores. Keywords: H-index, Papers Citation, Computer Science, K-means Clustering, Power BI, Analysis

**\*\*Introduction:\*\***

- The H-index is a metric used to measure the productivity and impact of academic papers.

- Paper citations are important in assessing the relevance and importance of academic publications.

- High numbers of citations indicate influential papers, and journals with high rejection rates tend to publish more highly cited works.

- Paper citations are used to evaluate the quality and significance of research papers in the academic field.

**\*\*Abstract: \*\***

- The paper uses the K-Means Clustering Algorithm, along with Power BI and Orange Data Mining tools, to analyze H-index and paper citations in computer science.

- The analysis reveals that research areas with the highest citation and H-index values include Artificial Intelligence, Computational Intelligence, Data Mining, Evolutionary Algorithms, and Big Data Analytics.

- The paper emphasizes the importance of paper citations in determining scientific impact in computer science.

**\*\*Literature Review:\*\***

- Studies have shown that higher H-index values indicate greater impact in academic literature.

- Strong correlations exist between the total number of citations and the H-index.

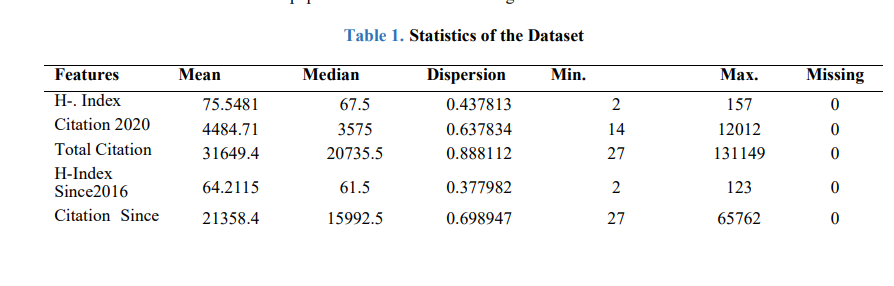
- Relative measures such as correlations between H-index and paper citations serve as indicators for scientific fields.

- H-index and paper citations offer valuable tools for assessing scholarly influence and scientific prestige.

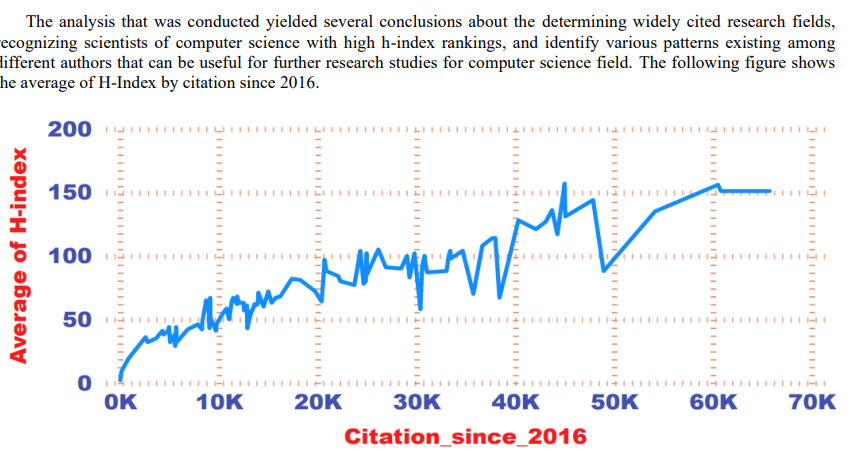
**\*\*Materials and Methods:\*\***

- The paper uses a dataset of 103 researchers in the Computer Science field, including information such as their names, affiliations, H-index, and citation counts.

- Microsoft Power BI is used for data analysis, while Orange Data Mining is used to implement the K-Means clustering algorithm.

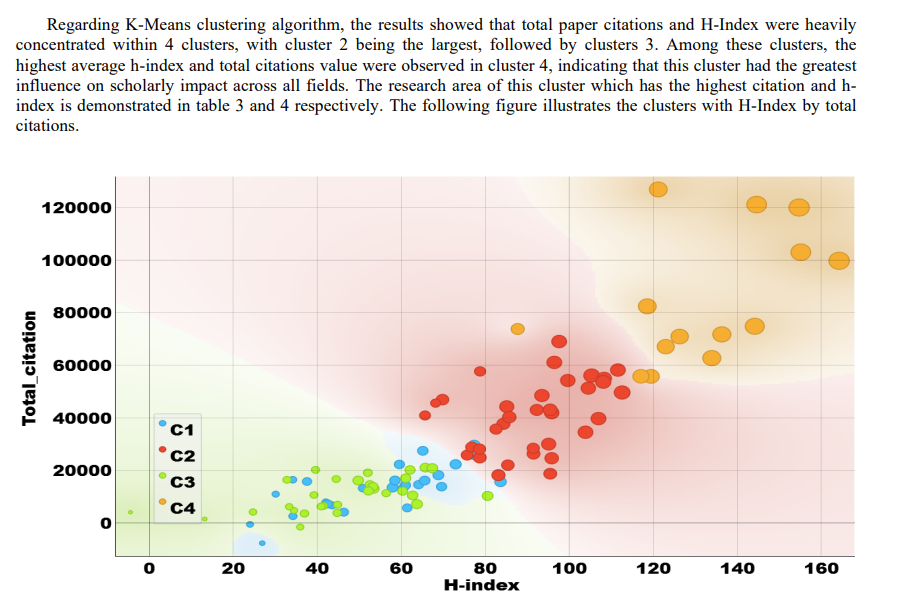
- Silhouette analysis is used to assess the separation of clusters within the dataset. 

**\*\*Results and Discussion: \*\***

- The analysis identified clusters with high H-index and total citation values, showing that certain research areas had a greater impact. 

- The paper presents tables of scientists with higher H-index, research areas with the highest H-Index, and research areas with the highest total citation.

- K-Means clustering showed that paper citations in software engineering and information technology had relatively lower overall influence compared to clusters in artificial intelligence and mathematical computing methods.



**\*\*Conclusion: \*\***

- The paper concludes that K-Means clustering and data visualization tools like Power BI provide effective methods for measuring and identifying research performance in the computer science field.

- It highlights the importance of assessing research output and provides insights into scholarly trends and influence within the field.

The paper offers a comprehensive analysis of the H-index and paper citations in the computer science field, using K-Means clustering to identify influential research areas and patterns among researchers. It emphasizes the significance of these metrics in assessing scholarly impact and research performance.

Funding: None

ACKNOLEDGEMENT:

The author we would like to thank the reviewers for their valuable contribution in the publication of this paper. CONFLICTS OF INTEREST The author declares no conflict of interest. REFERENCES   
[1] H-index (2023) Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/H-index (Accessed: February 4, 2023).

[2] Hirsch, J. E. Does the H index have predictive power? Proceedings of the National Academy of Sciences of the United States of America, 104(49), 19193– 19198, 2007.

[3] X. Bai, F. Zhang, and I. Lee, “Predicting the citations of scholarly paper,” Journal of Informatics, vol. 13, no. 1, pp. 407–418, 2019.

[4] T. Yu, G. Yu, P. Li, and L. Wang, “Citation impact prediction for scientific papers using stepwise regression analysis,” Scient metrics, vol. 101, no. 2, pp. 1233–1252, 2014.

[5] C. Castillo, D. Donato, and A. Gionis, “Estimating number of citations using author reputation,” in International Symposium on String Processing and Information Retrieval, N. Ziviani and R. Baeza-Yates, Eds., vol. 4726 of Lecture Notes in Computer Science, pp. 107–117, Springer, Berlin Heidelberg, 2007.

[6] Bai, X., Xia, F., Lee, I., Zhang, J., & Ning, Z. Identifying Anomalous Citations for Objective Evaluation of Scholarly Article Impact. PloS One, 11(9), e0162364, 2016.

[7] V. Jai Kumar , Y.V.Bhaskar Reddy , V.S.Nishok , Vivek K., Customer Data Processing in Internet of Things Environments Using RFM Analysis and K-means Clustering, Journal of Intelligent Systems and Internet of Things, Vol. 7 , No. 1 , (2022) : 08-19 (Doi : <https://doi.org/10.54216/JISIoT.070101>)

[8] Robbert van Haselen, the h-index: A new way of assessing the scientific impact of individual CAM authors, Complementary Therapies in Medicine, Volume 15, Issue 4, Pages 225-227, ISSN 0965-2299, https://doi.org/10.1016/j.ctim, 2007.

[9] Acuna, D. E., Allesina, S., & Kording, K. P. Future impact: Predicting scientific success. Nature, 489(7415), 201– 202. ISSN 0028-0836, 2012.

[10] Devastator, T. Papers citations vs h-index, Kaggle. Available at: https://www.kaggle.com/datasets/thedevastator/cs-researchers-h-index-and-citation-analysis (Accessed: February 4, 2023).

[11] Raheela zaib, & OURLIS Ourabah. (2023). Large Scale Data Using K-Means. Mesopotamian Journal of Big Data, 2023, 38–47. <https://doi.org/10.58496/MJBD/2023/006>

[12] Bioinformatics Laboratory, Data Mining, Orange Data Mining - Data Mining. Available at: https://orangedatamining.com/ (Accessed: February 4, 2023).

[13] Youguo Li, Haiyan Wu, A Clustering Method Based on K-Means Algorithm, Physics Procedia, Volume 25, Pages 1104-1109, ISSN 1875-3892, https://doi.org/10.1016/j.phpro.2012.03.206, 2012. W. Medhat et al.

[Sentiment analysis algorithms and applications: A survey](https://www.sciencedirect.com/science/article/pii/S2090447914000550)

Ain Shams Engineering Journal (2014)

S. Poria et al.

[A review of affective computing: From unimodal analysis to multimodal fusion](https://www.sciencedirect.com/science/article/pii/S1566253517300738)

Information Fusion (2017)

S. Poria et al.

[Fusing audio, visual and textual clues for sentiment analysis from multimodal content](https://www.sciencedirect.com/science/article/pii/S0925231215011297)

Neurocomputing (2016)

S. Poria et al.

[Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis](https://www.sciencedirect.com/science/article/pii/S0925231217302023)

Neurocomputing (2017)

K. Ravi et al.

[A survey on opinion mining and sentiment analysis: Tasks, approaches and applications](https://www.sciencedirect.com/science/article/pii/S0950705115002336)

Knowledge-Based Systems (2015)

M. Soleymani et al.

[A survey of multimodal sentiment analysis](https://www.sciencedirect.com/science/article/pii/S0262885617301191)

Image and Vision Computing (2017)

Alper Kursat Uysal et al.

[The impact of preprocessing on text classification](https://www.sciencedirect.com/science/article/pii/S0306457313000964)

Information Processing & Management (2014)

Rui Zhao

[Deep learning and its applications to machine health monitoring](https://www.sciencedirect.com/science/article/pii/S0888327018303108)

Mechanical Systems and Signal Processing (2019)