SOLVING BIG DATA ANALYTICS PLATFORM SELECTION BY USING MCDM METHOD

Jabbarova K.I.,
Doctor of Philosophy, associate professor of Azerbaijan State Oil and Industry University
Mammadaliyeva G.N.
master's student of Azerbaijan State Oil and Industry University
https://doi.org/10.5281/zenodo.10899481

Abstract

Choosing the best big data analytics platform is essential for companies looking to gain valuable insights in today's data-driven environment. However, due to the variety of available platforms and different application requirements, this selection procedure is inherently complex. To solve this problem, we use fuzzy TOPSIS method. This method allows you to systematically evaluate and select the most suitable approach for a big data analytics platform based on the company's requirements. The fuzzy number based decision matrix consist of three alternatives and four criteria (Availability and fault tolerance, Data security, Ease of installation and maintenance, User interface and reporting). The getting results show validaty of the considered approach.

Keywords: Big data analytics, Multi-Criteria Decision Making (MCDM), platform selection, fuzzy TOPSIS method

Introduction

The exponential proliferation of data in the modern period has altered the way judgments are made across many different disciplines. The volume, variety, and velocity of data have all dramatically increased as a result of the necessity to develop advanced analytics tools. You may offer insightful data on this platform to help put company plans into action. Solutions for big data analytics are necessary to handle massive data volumes and find patterns, features, and trends [1].

However, selecting a big data analytics platform is difficult for enterprises because of the abundance of possibilities and the intricate interactions between technological, operational, and strategic considerations. Scholars and professionals are increasingly using multi-criteria decision making (MCDM) techniques in addition to a methodical approach to get beyond the difficulties associated with platform selection in order to address this issue.

In order to determine the framework for decision data decision making, this study [2] investigates the numerous applications of MCDM approaches to platform-based big data analysis. Through the integration of quantitative and qualitative criteria, MCDM enables stakeholders to rank their choices in order of preference and assess various platforms according to how well they meet the needs and limitations of the company [3].

By taking into account the dynamic and complex character of decision making, this research also uses computer modeling tools to enhance decision making. A business may evaluate the platform's potential through analysis under various workload, resource, and condition scenarios. They are better equipped to determine if they are eventually appropriate for real deployment as a result [4].

By utilizing data from the literature on big data analytics, decision support systems, and MCDM approaches, this study seeks to expand our understanding of the use of platform selection strategies. This research intends to clarify the concepts, procedures, and real-world application of MCDM in fourth big data analytics, empowering companies to make choices that capitalize on the transformational power of data-driven knowledge [5].

Description of MCDM methods:

When evaluating and ranking options that are dependent on several goals or frequently in conflict, a methodical process known as multi-criteria decision making techniques (MCDM) is employed. Through the simultaneous definition of several criteria, these methods produce organized frameworks for resolving challenging decision-making situations. MCDM techniques offer useful instruments for evaluating and ranking platforms according to different technical, financial, and economic factors when choosing a big data analytics platform [6].

MCDM Approaches in the Selection of Platforms:

1. Process for Analytical Hierarchy (AHP): Saaty (1980) established the Analytical Analytical Process (AMP), which is one of the most well-known applications of the ellipsis method (MCDM). It entails breaking down issues into smaller subsets, making judgments along a hierarchy of acceptable and unacceptable possibilities, and assessing the proper languages between the subsets using numerical scales. The Analytic Hierarchy Process (AHP) with these comparisons entails decision-making, weighing the importance of each criterion, and keeping track of each alternative's overall score.

The Analytic Hierarchy Process (AHP) enables decision makers to give consideration to scalability, performance, cost-effectiveness, and user experience when choosing a big data analytics platform. The MAI offers many platforms based on this data, establishing quantitative options and offering an organized definition strategy for professionals [6].

2. A fast computation is used in the TOPSIS approach to determine selection order by similarity:

The TOPSIS technique proposed by Hwang and Yoon (1981) is applied for alternative ranking. Based on predetermined data, ideal and anti-ideal solutions are found, and as a result, the alternative is calculated to be isolated from these solutions. The following criteria are used to rank the alternatives: the closer they are to the ideal answer, the further away they are from it.

TOPSIS allows decision makers to simultaneously evaluate the relative performance of multiple platforms using various criteria when selecting a big data analytics platform. To help you use mobile solutions, TOPSIS evaluates the advantages and disadvantages of each platform and provides a comprehensive evaluation [7].

3. The PROMETEE approach is a way to rank indicators in order to evaluate enrichment.

A noncompensatory MCDM technique called PROMETHEE was created by Bransom and Vincke (1985). It ranks alternatives based on a comparison of their profile performance under various scenarios. It compares the alternatives using selection functions, which establish which alternative is preferred to what extent. PROMETHEE assigns a total score based on these choices [8].

When selecting a large data analytics platform, PROMETHEE enables decision makers to view the finer points and trade-offs between many factors. The PROMETHEE framework's structure, which guarantees dependability and transparency in decision-making processes, must then be shown [8].

MCDM method(s) specifying criteria:

The choice of MCDM method(s) is influenced by a number of factors, including the decision makers' preferences, the nature of the decision issues, and the data's availability. Factors including computational complexity, simplicity of use, transparency, and dependability are crucial when selecting a big data analytics platform.

Simulation design:

The MCDM problem consist of three alternatives (A1, A2, A3) and four criteria (C1-Availability and fault tolerance, C2- Data security, C3-Ease of installation and maintenance, C4-User interface and reporting). For solving this issue fuzzy TOPSIS method is used.

Step 1. Construct the normalized fuzzy decision matrix (see Table 1).

Table 1.

The normalized fuzzy decision matrix

weights	**0,21**	**0,26**	**0,23**	**0,3**
	Availability and fault tolerance	**Data security**	**Ease of installation and maintenance**	**User interface and reporting**
A1	(0.77,0.94,0.99)	(0.71,0.88,0.99)	(0.77,0.94,0.99)	(0.77,0.94,0.99)
A2	(0.61,0.76,0.92)	(0.66,0.81,0.97)	(0.66,0.81,0.96)	(0.71,0.88,0.99)
A3	(0.77,0.94,0.99)	(0.77,0.94,0.99)	(0.77,0.94,0.99)	(0.77,0.94,0.99)

Step 2. Compute the weighted normalized fuzzy decision matrix (see Table 2)

The weighted normalized fuzzy decision matrix is $\tilde{V} = (\tilde{v}_{ij})$, where $\tilde{v}_{ij} = \tilde{r}_{ij} \times w_j$.

Table 1.

The NFDM

	Availability and fault tolerance	**Data security**	**Ease of installation and maintenance**	**User interface and reporting**
A1	(0.1617,0.1974,0.2079)	(0.1846,0.2288,0.2574)	(0.1771,0.2162,0.2277)	(0.231,0.282,0.297)
A2	(0.1281,0,1596,0.1932)	(0.1716,0.2106,0.2522)	(0.1518,0.1863,0.2208)	(0.213,0.264,0.297)
A3	(0.1617,0.1974,0.2079)	(0.2002,0.2444,0.2574)	(0.1771,0.2162,0.2277)	(0.231,0.282,0.297)

Step 3. Compute the Fuzzy Positive Ideal Solution (FPIS) and Fuzzy Negative Ideal Solution (FNIS) (see Table 3)

$$A^{*} = \tilde{v}_1^{*}, \tilde{v}_2^{*}, \dots, \tilde{v}_n^{*}, \text{ where } \tilde{v}_j^{*} = \max_i \{v_{ij}\}$$

$$A^{-} = \tilde{v}_1^{-}, \tilde{v}_2^{-}, \dots, \tilde{v}_n^{-}, \text{ where } \tilde{v}_j^{*} = \min_i \{v_{ij}\}$$

Table 3.

Positive Ideal and Negative Ideal Solutions

	Availability and fault tolerance	**Data security**	**Ease of installation and maintenance**	**User interface and reporting**
A1	(0.1617,0.1974,0.2079)	(0.1846,0.2288,0.2574)	(0.1771,0.2162,0.2277)	(0.231,0.282,0.297)
A2	(0.1281,0,1596,0.1932)	(0.1716,0.2106,0.2522)	(0.1518,0.1863,0.2208)	(0.213,0.264,0.297)
A3	(0.1617,0.1974,0.2079)	(0.2002,0.2444,0.2574)	(0.1771,0.2162,0.2277)	(0.231,0.282,0.297)
Positive ideal	**(0.1617,0.1974,0.2079)**	**(0.2002,0.2444,0.2574)**	**(0.1771,0.2162,0.2277)**	**(0.231,0.282,0.297)**
Negative ideal	**(0.1281,0.1596,0.1932)**	**(0.1716,0.2106,0.2522)**	**(0.1518,0.1863,0.2208)**	**(0.213,0.264,0.297)**

Step 4. Compute the distance from each alternative to the FPIS and to the FNIS.

$$d_i^* = \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^*)$$

$$d_i^- = \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^-)$$

Be the distance from each alternative A_i to the FPIS and to the FNIS, respectively (see Table 4).

Table 4.

The value of d_i^* and d_i^-

	d_i^*	d_i^-
A1	0.012737	0.081324
A2	0.093805	0
A3	0	0.093805

Step 5. Compute the closeness coefficient CC_i for each alternative (see Table 5)

$$CC_i = \frac{d_i^-}{d_i^- + d_i^*}$$

Table 5.

The results of CC_i

	CC_i
A1	0.864584
A2	0
A3	1

Step 6. Rank the alternatives (see Table 6)

Table 6.

Ranking results

	CC_i	Rank
A1	0.864584	2
A2	0	3
A3	1	1

The third alternative is best one.

Conclusion:

In the quickly developing field of big data analytics, choosing the right platform is essential for companies wishing to use data-driven analytics. The study discovered that platform selection complexity and uncertainty issues may be successfully handled by multi-criteria decision making (MCDM) techniques.

References:

1. Saaty, T. L. 1980. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York.

2. Hwang, C. L., and Yoon, K. 1981. Multiple Attribute Decision Making: Methods and Applications. New York: Springer-Verlag.

3. Brans, J. P., and Vincke, P. H. 1985. A Preference Ranking Organization Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). Management Science, 31(6), 647-656.

4. Kaur, K. and Gupta, A., Pandey, S., 2018. A systematic review of big data analytics: Challenges and insights. Journal of Big Data, 5(1), 1-33.

5. Zhang, Y., Li, C., and Chen, H. 2019. A Review of Big Data Analytics in Industrial Applications. Special Issue on Big Data Analytics, 7(4), 1-13.

6. He, H., Xie, L., and Guo, S. 2021. Multi-criteria decision-making methods for technology selection: A literature review. Journal of Cleaner Production, 297, 126599.

7. Jahanshahloo, G. R., Lotfi, F. H., and Izadikhah, M. 2006. An algorithmic method to extend TOPSIS for decision-making problems with interval data. Applied Mathematics and Computation, 175(2), 1375-1384.

8. Marler, R. T., and Arora, J. S. 2004. Survey of multi-objective optimization methods for engineering. Structural and Multidisciplinary Optimization, 26(6), 369-395.

9. Jabbarova, Aynur I., Jabbarova, K.I. Solving Employee Selection Problem Under Fuzzy-Valued Information. 15th International Conference on Applications of Fuzzy Systems, Soft Computing and Artificial Intelligence Tools – ICAFS-2022. 620-625 (2023).