Selection of the Best Data Analytics Tool for Data Analysts in Nepal Using the Analytic Hierarchy Process

Author 1: *Prince Manandhar*[1]
Author 2: *Prabal Sapkota*[2]

Highlights

- AHP is utilized to identify the best data analytics software/tool for Nepalese data analysts.
- Five important factors and their sub-factors were determined.
- Alternatives include Tableau, Power BI, Python, and RStudio.
- The study aims to assist Nepalese organizations to select the most appropriate data analytics tool.

ABSTRACT

Data is one of the most important assets an organization possesses. When properly analyzed, data can optimize business processes and provide a competitive advantage. With the growth in adoption of computerized technology in business, there was a massive rise in data generation and storage. This gave birth to data analytics. Data analytics uses tools/software that can streamline business processes and enhance decision-making to drive growth and efficiency. However, this only holds true with the application of appropriate tools. Selecting an appropriate tool is not an easy task as multiple data analytic software/tools have emerged in the last two decades. In order to select an appropriate tool, one has to consider several factors/subfactors associated with it. This research aims to identify the most important factors/subfactors and the alternatives associated with the selection of data analytic tools/software in the Nepalese context. This is a case of Multi Criteria Decision Making (MCDM) and Analytic Hierarchy Process (AHP) has been adopted in this work.

Keywords: data, data analytics, MCDM, AHP

1. Introduction

[1] Prince Manandhar, Student, Department of Management Informatics Communication, Kathmandu University, Dhulikhel, Kavre, Nepal, e-mail: 2011019_prince@kusom.edu.np.

[2] Prabal Sapkota, Assistant Professor, Department of Management Informatics Communication, Kathmandu University, Dhulikhel, Kavre, Nepal, e-mail: prabal@ku.edu.np (ORCID: 0000-0002-0628-9373).

The dictionary definition of data, "facts or information, especially when examined and used to find out things or make decisions," clearly indicates that it is significantly crucial in decision-making. In addition, it is essential for optimizing business processes, as proper data analysis can identify bottlenecks, inadequacies, and areas for improvement, enabling organizations to take alternative actions to resolve the issue. However, data is useless if not analyzed properly, which led to the emergence of the discipline - data analytics. Data analytics involves examining, cleaning, transforming, and modeling data to uncover insights and facilitate decision-making. At the beginning, data analytics was a simple manual process used for recording transactions and monitoring inventory. With the advancement of computers, it became faster and more efficient. However, in the 1970s and '80s, the development of relational databases and Structured Query Language (SQL) transformed the way of storing and retrieving data. By 1990 the exponential rise in data generation led to the coining of the term "big data". This was mainly driven by the growth and expansion of internet usage. The challenge of handling huge volumes of data led to the development of Hadoop and MapReduce which allowed for distributed processing of data across multiple computers. The progression in cloud computing and open-source technology by the 2010s made data analytics more accessible and in 2013, Tableau software went public after ten years of its establishment.

Including Tableau, there are numerous paid and open-source data analytics software/tools available for example, Microsoft Power BI, RStudio, Python, WEKA, Orange, KNIME, Excel, Google Sheets, Rapid Miner, and Google Analytics. There are no specific guidelines regarding the selection of these tools.

The choice of data analytics tools can have profound consequences for a business. The appropriate tools can streamline business processes and enhance decision-making to drive growth and efficiency. However, the wrong choice can lead to missed opportunities. Therefore, it is important to sensibly assess various factors to confirm that the chosen tools address the needs and goals of the organization.

Identifying the best tools is not a straightforward decision. It comes with various factors and these factors could be equally important. This increases the complexity in the decision making which leads to a true case of multi criteria decision making (MCDM).

Rationality of the Study

Data analytics could play a vital role in the success of businesses irrespective of size. It can help make better decisions regarding resource allocation, pricing their products and services, and understanding their customer base and trends they are following. This is possible only with selecting the right tools. Understanding data analytics is new in Nepal, both in academia and business, but the trend of learning and adopting it is growing. This work aims to identify the key factors that drive the data analysts to use their preferred data analytics tools. In addition, the work also aims to rank the data analytics software/tools preferred by Nepalese analysts, considering multiple factors. This work can provide a guideline for academics, business houses, government agencies and students regarding data analytics tool selection.

Research Question

- What is the most important factor the Nepalese data analysts consider while selecting data analytics tools/software?
- Which is the most appropriate analytics tool/software in the Nepalese context?

2. Literature Review

As the volume, variety, and velocity of data continue to expand, many organizations have invested in data analytics tools to generate and share valuable insights that improve decision-making. The ability to effectively utilize these tools has become a critical factor for business success, enabling companies to better detect and respond to potential threats and opportunities, and ultimately shape and capitalize on them for a competitive edge (Ghasemaghaei, 2019). Data analytics tools also facilitate knowledge sharing, enabling companies to distribute insights gained from analyzing data gathered from both internal and external sources (Côrte-Real, 2016)

Embracing big data analytics is reshaping today's industrial landscape, promoting precise decision-making and enhancing overall performance (Maroufkhani, 2020). According to Forbes, 53% of companies were adopting big data analytics in 2017. This trend is increasing globally, and organizations are determined not to miss this opportunity or fall behind.

Selecting appropriate data analytics tools is challenging as several criteria need to be considered and multiple alternatives are available. For example, visualization tools such as Tableau, Power BI, and D3.js were evaluated considering factors such as ease of use, flexibility, and output quality (Dogadina & Voronin, 2024). Similarly, other tools such as RapidMiner, Orange, and KNIME were analyzed focusing on features, specifications, and limitations (Kawade, & Deoskar, 2019). Other factors studied were speed, visualization capabilities, memory usage, pricing, learning curves, and community support and the tools such as Python, and SPSS were explored (Christa, 2012; Bansal, 2018).

Identifying the most important criteria for organizations when selecting an appropriate data analytics tool requires consideration of several factors and alternatives. Saaty (2008) emphasized the use of a model that is not overly complex, legitimately aggregates across scales, and maintains consistency in judgments from multiple participants. The Analytic Hierarchy Process (AHP) can break down complex problems into smaller, more manageable parts (Saaty, 2008).

3. Objectives

The objectives of this research are:

- ➢ To determine the best data analytics software/tool in the Nepalese context
- ➢ To determine the most important factor to consider while selecting data analytics tools/software.

4. Research Design

The research was conducted in two stages. In the first stage, literature review was conducted to identify the alternative software/tools and the key factors influencing their selection. The factors and alternatives observed in the literature were further validated by a group of experts professionally working in Nepal as data analysts.

In the second stage, the Analytic Hierarchy Process was employed. The following steps were taken:

- An AHP model was developed based on the factors, subfactors and alternatives, collected from the literature and expert opinions.
- A questionnaire based on pairwise comparisons of different hierarchical levels was created.
- Respondent selection: The respondents for the pairwise questionnaire were the data analysts actively working in big data analytics.
- Data collection was conducted individually, with proper demonstration of the pairwise comparison process and explanation of potential inconsistencies. Data with an inconsistency ratio of less than 0.1 was accepted.
- A total of 15 valid samples were used in the survey, all with an inconsistency ratio of less than 0.1.
- The geometric mean of the valid data collected from individual respondents was calculated for aggregation. This mean value was used for further calculations and prioritization of alternatives.

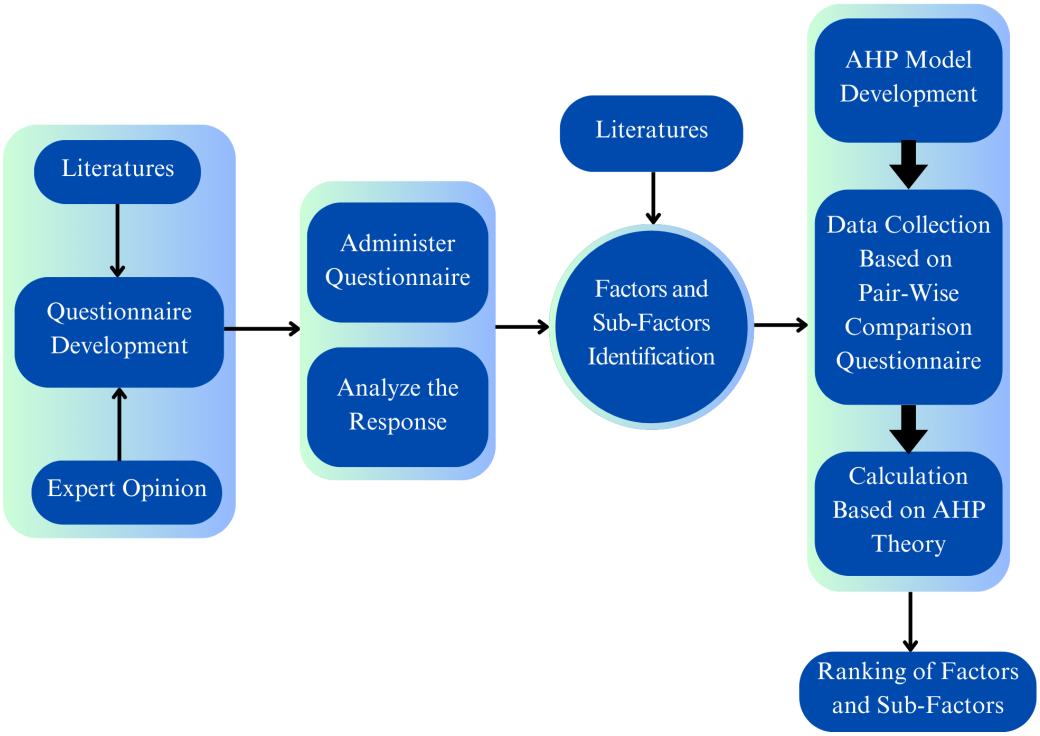

Figure 1 Research Design

5. Model Analysis

An AHP model was developed based on the responses collected from the first-stage questionnaire and expert opinions. The goal was set as “Best tool for Data Analytics” which

is placed at level 1 of the hierarchy. Five factors were incorporated at level 2, with their respective subfactors at level 3, and four alternatives were placed at level 4.

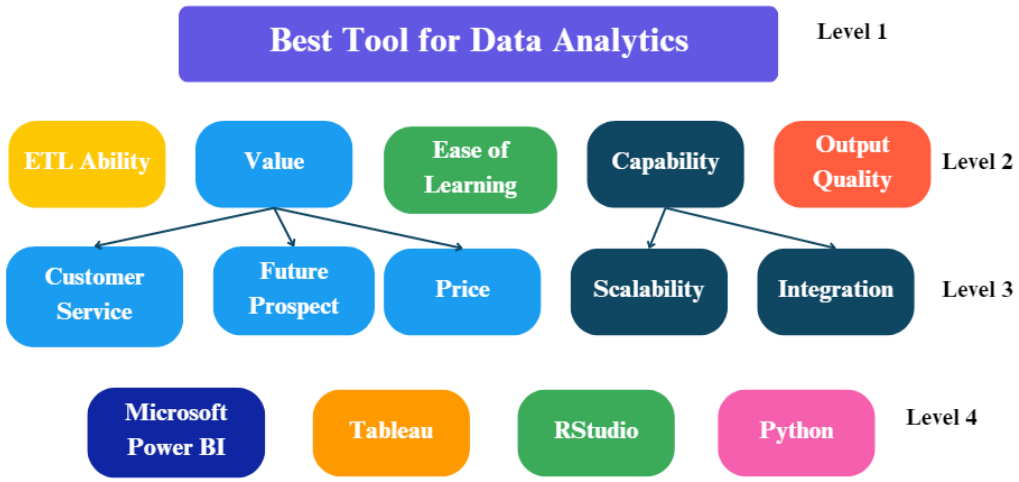

Figure 2 Hierarchy for prioritizing factors and subfactors and alternatives

The responses obtained from the pairwise comparisons were aggregated by calculating the geometric mean using Microsoft Excel Office 365. The calculated geometric mean values were organized into a matrix. A sample matrix as shown in Table 1 was developed. Further calculations were performed to determine the priority vector, row averages, and finally, the consistency ratio. The consistency ratio obtained was 0.01, which is below the threshold value of 0.1. In all cases, the consistency ratio remained below 0.1.

Table 1 Pair wise comparison of factors with respect to the goal

	ETL Ability	Value	Capability	Ease of Learning	Output Quality
ETL Ability	1	1.46	1.27	1.29	1.06
Value	0.70	1	0.9	0.91	0.758
Capability	0.8	1.16	1	1.01	0.832
Ease of Learning	0.77	1.09	0.987	1	0.826
Output Quality	0.93	1.38	1.19	1.213	1

6. Conclusions

The study aims to explore data analytics perspectives on selection of best data analytics software/tool using MCDM and AHP approaches. This work could provide valuable insights for data science students and business owners to select the right tools. In addition, this can also guide policymakers and educational institutions in developing targeted training programs, ensuring future analysts are skilled in the most relevant tools. Although the study has been conducted in a Nepalese context, it can be easily adapted to other locations. Additionally, while the research findings are subjective, as they are based on the

judgment of a select group of individuals, the developed model is applicable and useful in most cases.

7. Limitations

This paper has some limitations. The sample size is small, and the method relies on subjective judgments from the respondents. Since only a select few people were respondents, the weights assigned to the decision- makers may reflect the opinions of a limited group. Inconsistent input from a single respondent can affect the overall rankings of factors, subfactors and alternatives. Additionally, the same model may not be applicable to other locations where selection criteria could differ.

8. Key References

1. Bansal, A., Srivastava, S., (2018). Tools Used in Data Analysis: A Comparative Study, International Journal of Recent Research Aspects, 5, 15- 18

2. Christa, S., Madhuri, K. L., & Suma, V. (2012). A comparative analysis of data mining tools in agent based systems. arXiv preprint arXiv:1210.1040.

3. Côrte-Real, N., Oliveira,T., Ruivo, P. (2016). Assessing business value of Big Data Analytics in European firms, *Journal of Business Research*, 70, 379- 390

4. Dogadina, V. V., & Voronin, A. V. (2024). Comparative analysis of data visualization tools. *International Journal of Computing, Programming and Database Management*, 5(1), 49- 51.

5. Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency, *Decision Support Systems*, 120, 14- 24

6. Kawade, B., & Deoskar, A. (2019). Comparative study of data analytics open source tools for educational data analytics. *Journal of Emerging Technologies and Innovative Research*, 6(2).

7. Maroufkhani, P., Tseng, M-L, Iranmanesh, M. et al, Big data analytics adoption: Determinants and performances among small to medium-sized enterprises, *International Journal of Information Management*, 54, 109021

8. Saaty, T.L. (2008). Decision making with the analytic hierarchy process. *Int. J. Services Sciences*, 1(1), 83- 98.