IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data-Driven Decisions: A Deep Dive Into Big Data Applications And Decision Trees

1. **Basavaraju S**
Associate Professor
Department of Computer Applications
Global Institute of Management Sciences
Rajarajeshwari Nagar, Bangalore-560098

2. **Ranjana Kumari**
Professor
Department of Computer Applications
Global Institute of Management Sciences
Rajarajeshwari Nagar, Bangalore-560098

Abstract

The rise of digital technologies has led to the exponential growth of data, necessitating advanced analytical frameworks to manage, process, and derive meaningful insights. Big Data, characterized by its high volume, velocity, and variety, serves as a foundation for informed decision-making across sectors such as business, healthcare, finance, and e-commerce. Concurrently, decision trees have emerged as intuitive tools for classification and prediction, offering a transparent and structured approach to decision-making. This paper explores the fundamental concepts of Big Data, its classifications, and key applications, and demonstrates the construction and use of decision trees through practical case scenarios. Additionally, the study incorporates the use of EdrawMind, an open-source visualization tool, to develop clear and interpretable decision tree diagrams that enhance understanding and presentation. The study highlights how these tools collectively contribute to data-driven decisions and outlines their potential for academic and industrial applications.

Keywords
Big Data, Decision Trees, Data Analytics, Data-Driven Decision-Making.

1. Introduction

The digital era has led to a rapid increase in the amount of data produced every second. This growth, known as Big Data, creates both opportunities and challenges. It allows organizations to gain useful insights for decision-making but also brings difficulties in storing, processing, and analyzing information with traditional methods. As a result, decision-making frameworks need to adapt to handle this scale and complexity.

Among various analytical tools, decision trees are widely used because of their simple and visual structure. They break down complex decisions into clear, step-by-step models that can be understood by both technical and non-technical users. This paper focuses on combining Big Data analytics with decision tree techniques and highlights their practical applications through examples and case studies.

2. Objectives

The objectives of this study are to:

1. Explain the concept and significance of Big Data in modern analytics.
2. Classify Big Data into structured, unstructured, and semi-structured forms.
3. Highlight real-world applications of Big Data across multiple industries.
4. Introduce decision trees as a predictive and prescriptive analytical tool.
5. Demonstrate the construction and interpretation of decision trees using illustrative scenarios.

3. Literature Survey

[1] Big Data Analytics Across Emerging Domains

Taha (2025) provides a comprehensive review of Big Data analytics in emerging domains such as IoT, social media, NLP, and information security. The study evaluates multiple machine learning models, including decision trees and XGBoost, across large-scale datasets. A major contribution of this work is the benchmarking of models in terms of interpretability, scalability, and computational performance. These insights align with this paper's focus on combining Big Data frameworks with interpretable analytical models for practical deployment.

[2] Modern Surveys on Decision Tree Algorithms

Mienye and Jere (2024) offer a detailed survey covering decision tree algorithms from classic methods (ID3, CART, C4.5) to modern ensembles, highlighting applications in healthcare and fraud detection. Their work emphasizes interpretability and scalability, echoing the objectives of this study.

[3] Parallel Decision Tree Approaches for Big Data Streams

Shiralizadeh (2025) reviews stream-based and parallel decision tree algorithms optimized for large-scale data environments. Approaches like pdsCART are discussed in the context of MapReduce frameworks, which address scalability challenges and enable real-time decision-making — directly relevant to Big Data analytics workflows.

[4] Data Mining Techniques and Decision Trees

Han, Kamber, and Pei (2023) in *Data Mining: Concepts and Techniques* extensively discuss data preprocessing, classification, and prediction methods, emphasizing decision trees as a core technique for interpretable analytics. Their work highlights the scalability challenges of traditional decision tree algorithms when applied to Big Data, thereby motivating research on distributed frameworks like MapReduce for large datasets. This study provides foundational understanding for combining Big Data characteristics (Volume, Velocity, Variety) with structured decision-making approaches.

[5] Adaptive Decision Trees for Dynamic Data

Bressan & Sozio (2024) propose dynamic decision trees that can be updated in response to adversarial data changes while maintaining update-time guarantees and decision quality—particularly relevant for Big Data streaming and real-time analytics.

[6] Learning Accurate and Interpretable Decision Trees

Balcan and Sharma (2024) present advanced decision tree learning techniques that balance accuracy and interpretability. Their approach introduces tunable splitting criteria and pruning mechanisms optimized for specific datasets, enabling better control over model complexity in Big Data applications.

[7] Stability-Focused Decision Tree Models

Bertsimas and Digalakis Jr. (2023) propose methods for improving the stability of decision trees in critical applications such as healthcare analytics. Their findings highlight that enhancing model consistency can be achieved with minimal trade-off in predictive performance, which is crucial for high-stakes decision-making environments.

[8] Dynamic Decision Trees for Streaming Big Data

Vidal et al. (2024) introduce dynamic decision tree algorithms capable of adapting to evolving data streams in real time. This work addresses the challenges of scalability and continuous learning in Big Data environments, aligning closely with this paper's focus on practical and interpretable analytics frameworks.

[9] Open-Source Visualization Tools for Analytics

Wondershare EdrawMind (2024) offers an open-source platform for creating mind maps and decision trees. Research on such tools emphasizes the need for simplified visual representations of complex analytical models, enabling better collaboration between technical and non-technical stakeholders. This supports the paper's approach of integrating visualization to improve decision transparency in Big Data projects.

4. Types of Big Data

Big Data can be broadly classified into three main categories based on the nature and organization of the information:

1. Structured Data: Data that is organized in rows and columns, like in databases or spreadsheets (e.g., customer names, phone numbers, transaction records).
2. Unstructured Data: Data that does not have a fixed format and is often text-heavy or media files (e.g., emails, videos, social media posts).
3. Semi-Structured Data: Data that is not fully organized but contains tags or markers to separate elements (e.g., XML files, JSON data, log files).

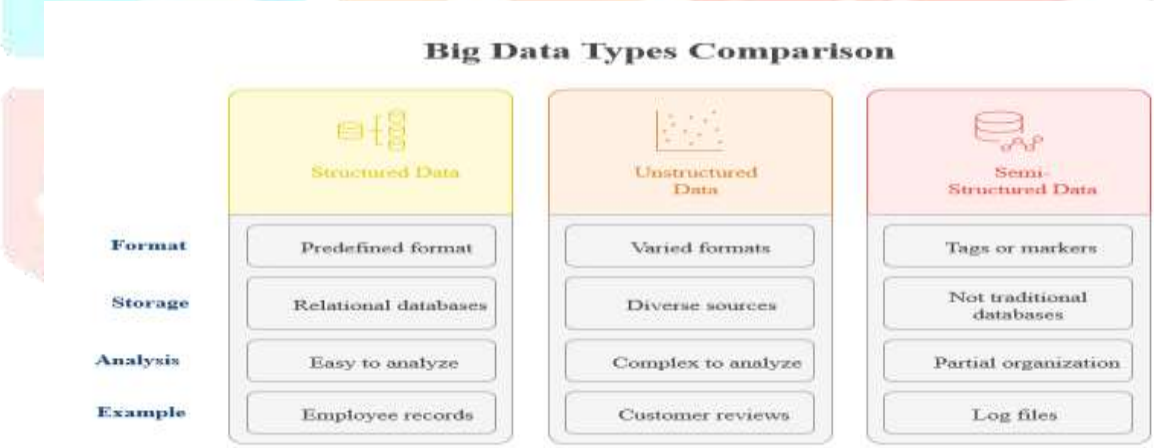

Figure 4.1 Big Data Types Comparison (Image generated using Napkin.ai)

5. Applications of Big Data

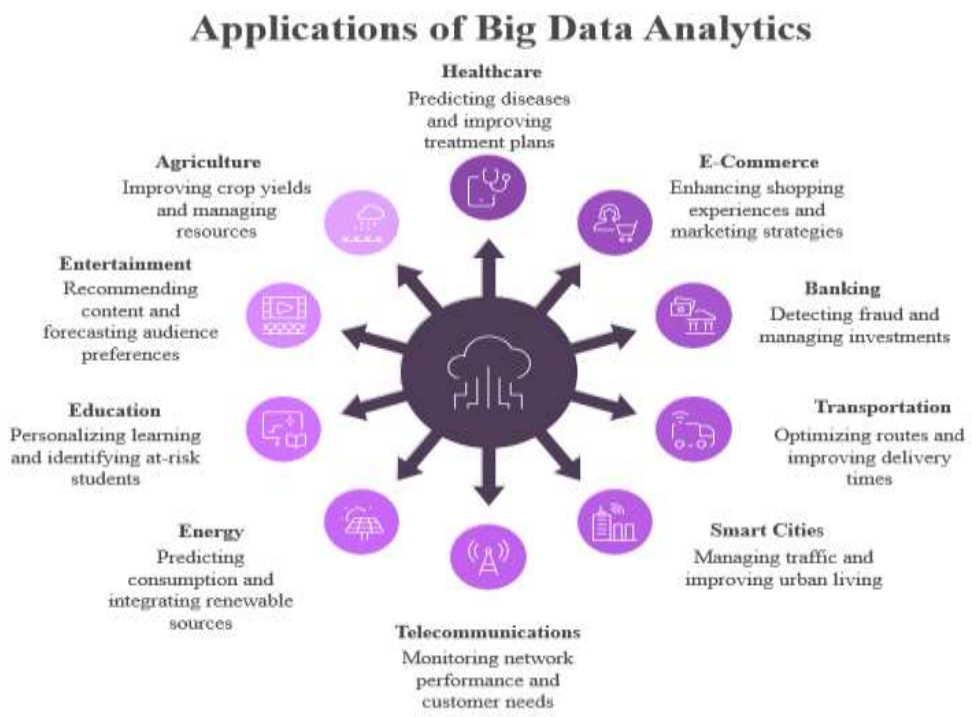

Figure 5.1 Applications of Big Data Analytics (Image generated using Napkin.ai)

6. Decision Trees: Concept and Structure

A decision tree is a visual representation of decision-making. It consists of:

- Root Node (initial decision criterion)
- Internal Nodes (intermediate decisions)
- Leaf Nodes (final outcomes)

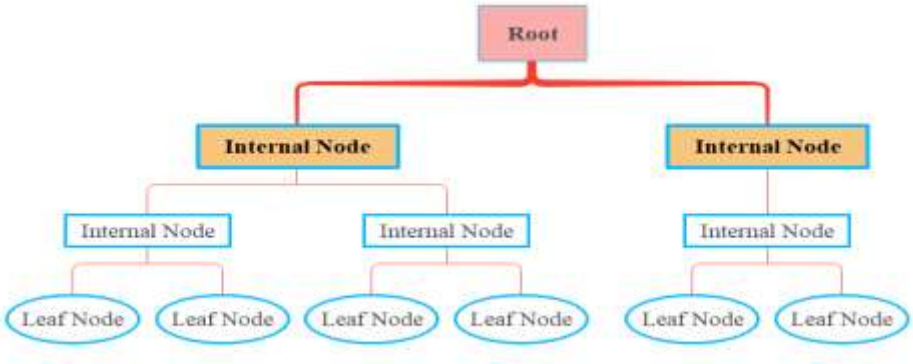

Figure 6.1 Decision Tree Structures

7. Building Decision Trees: Illustrative Examples

Example 1:

Build a decision tree for following scenario

Background: An e-commerce company wants to enhance its user experience by implementing a decision tree-based product recommendation system. The goal is to personalize the shopping experience for both logged-in and guest users.

Scenario: The company has two types of users:

Logged-in users who may have a history of previous purchases and recommend them based on past history

Guest users who are new to the site and may not have any prior interaction. Recommend them trending category product

Figure 7.1 Decision Tree for E-Commerce Product Recommendations

Figure 7.2 Outline of Decision Tree for E-Commerce Product Recommendations

8. Case Studies

Case Study 1: **Build a decision tree for the following scenario**

A company evaluates candidates for a teaching position using a structured decision-making process. The decision tree is based on three sequential criteria: required qualifications, interview performance, and salary acceptance.

- Candidates without required qualifications are rejected immediately.
- Qualified candidates proceed to interviews; weak performance leads to consideration of other candidates.
- Strong performers are assessed for salary acceptance; those accepting are hired, while those rejecting are declined.

Insight: Qualifications act as the primary filter, followed by interview strength and salary considerations, ensuring a structured and fair recruitment process.

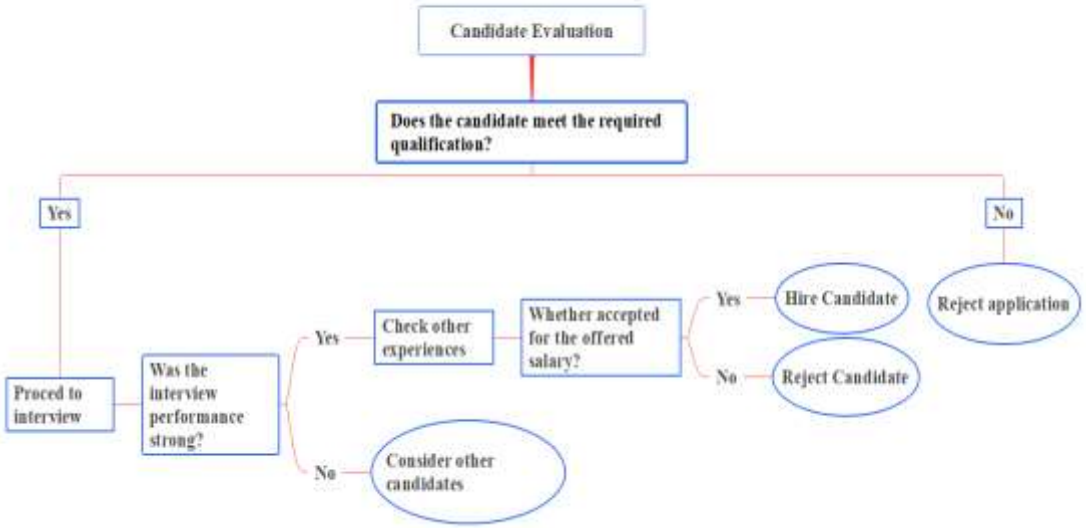

Figure 8.1 Decision Tree for Hiring Process Evaluation

- **Candidate Evaluation**
  - **Does the candidate meet the required qualification?**
    - **Yes**
      - **Proceed to interview**
        - **Was the interview performance strong?**
          - **Yes**
            - **Check other experiences**
              - **Whether accepted for the offered salary?**
                - **Yes**
                  - **Hire Candidate**
                - **No**
                  - **Reject Candidate**
          - **No**
            - **Consider other candidates**
    - **No**
      - **Reject application**

Figure 8.2 Outline of Decision Tree for Hiring Process Evaluation

Case Study 2: Job Offer Acceptance Based on Salary and Benefits

Background: A company seeks to model candidate decisions regarding job offers based on salary and benefits. The decision tree uses four attributes: offered salary (₹20,000–30,000), provident fund (PF) facility, annual hike, and health insurance.

- If the salary is below ₹20,000, the offer is declined.
- For salaries within range, candidates evaluate PF facility, annual hike, and health insurance sequentially.
- Lack of any benefit leads to declining the offer; presence of all benefits results in acceptance.

Insight: Salary is the primary determinant, followed by benefits, with health insurance acting as the final influencing factor for acceptance.

Figure 8.3 Decision Tree for Job Offer Acceptance Based on Salary and Benefits

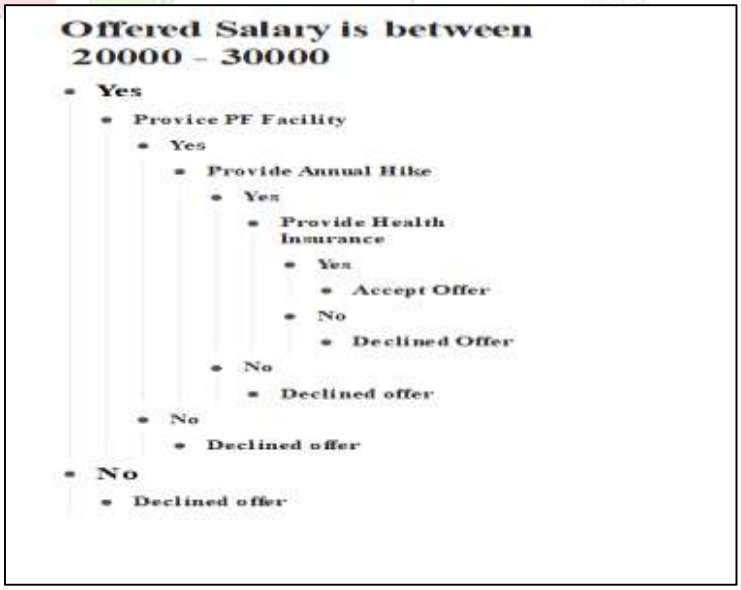

Figure 8.4 Decision Tree for Job Offer Acceptance Based on Salary and Benefits

9. Significance and Discussion

Big Data and decision trees together enable transparent, scalable, and actionable decision-making. They are particularly useful in scenarios where interpretability is as important as accuracy (e.g., HR decisions, healthcare diagnostics).

10. Conclusion

Big Data and decision trees are critical components of modern analytics frameworks. Their integration provides structured decision-making capabilities and is applicable across multiple industries. Future work includes exploring ensemble methods (e.g., Random Forests) and integrating deep learning for advanced applications.

References

[1] K. Taha, "Big Data Analytics in IoT, Social Media, NLP, and Information Security: Trends, Challenges, and Applications," *Journal of Big Data*, vol. 12, art. 150, 2025.

[2] D. I. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," IEEE Access, vol. 99, pp. 1–1, Jan. 2024.

[3] Z. Shiralizadeh, "A Review and Analysis of a Parallel Approach for Decision Tree Learning from Large Data Streams," *arXiv preprint*, May 2025.

[4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. San Francisco, CA: Morgan Kaufmann, 2023.

[5] M. Bressan and M. Sozio, "Fully-Dynamic Approximate Decision Trees With Worst-Case Update Time Guarantees," in Proc. 41st Int. Conf. Mach. Learn., vol. 235, pp. 4517-4541, Jul. 2024.

[6] M.-F. Balcan and D. Sharma, "Learning Accurate and Interpretable Decision Trees," Proc. 40th Conf. Uncertainty in Artificial Intelligence, vol. 244, pp. 288–307, Jul. 2024.

[7] D. Bertsimas and V. Digalakis Jr., "Improving Stability in Decision Tree Models," arXiv preprint, May 2023.

[8] J. Vidal, S. Jha, Z. Liang, E. Delgado, B. S. Deneke, and D. Shasha, "Dynamic Decision Trees," Knowledge, vol. 4, no. 4, pp. 506–542, Jul. 2024.

[9] Wondershare, "EdrawMind: Open-Source Mind Mapping and Diagramming Tool," 2024. [Online]. Available: https://www.edrawmind.com

[10] Napkin.ai. (2025). *AI-powered diagramming and visualization tool*. Retrieved from https://www.napkin.ai