Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 10, 2025

BIG DATA FOR PRODUCTIVITY: A POLICY-TO-PRACTICE FRAMEWORK FOR FIRMS

Saadia Tahir [1], Komal Bashir [*2], Faria Kanwal [3]

*[1,*2,3]Lahore College for Women University, Lahore*

[*2]komal.bashir@gmail.com

DOI: https://doi.org/10.5281/zenodo.17411858

Keywords
Big Data, Big Data Analysis, Big Data tools, Business Intelligence, data linkage, OLAP

Article History
Received: 30 August 2025
Accepted: 09 October 2025
Published: 22 October 2025

Corresponding Author: *
Komal Bashir

Abstract
Technology generates optimal outcomes when employed wisely and comprehensively. Big data as a technology tool converts raw, unstructured information into a strategic asset that can generate competitive advantage and measurable value for the organization. By addressing varied information and implementing advanced analytics, organizations can improve predictive models for consumer behavior and progress in medical, educational, commercial, scientific, and many other social sectors. Nonetheless, these benefits are unevenly distributed: productivity improvements are focused in technologically advanced nations, while many developing countries receive only a fraction of the potential. This paper asserts that realizing the potential of big data requires interdisciplinary expertise—merging domain knowledge with statistics, computing, and data governance—alongside careful attention to confounders and outliers, rather than depending solely on single-domain intuition. We propose a conceptual framework for converting data assets into sustainable advantages, founded on three pillars: (1) interoperable infrastructure and standards to enable dataset integration; (2) human capital and literacy for the design, validation, and implementation of models; and (3) governance frameworks that ensure quality, privacy, equity, and responsible use. The document highlights essential research goals, including methods for secure data integration, metrics for productivity spillovers, and policy interventions that promote adoption in resource-constrained settings. These components illustrate how businesses and nations can transform data capabilities into substantial economic growth.

INTRODUCTION

Big data represents large data collection having size beyond the capability of frequently used software tools for capturing, managing, and processing the data within the elapsed time [1], with the growing time companies investing more and more in this subject to generate revenue and get a competitive advantage [2]. Big data focuses on understanding and targeting customers, optimizing business processes, personal quantification, as well as performance optimization, improving public healthcare systems, improving and predicting sports performance, promoting science and research, upgrading device and machine performance, reforming security and law enforcement, refining and updating cities and countries, and financial trading. With their increasing size, difficulties like capturing, memory management, visualization, analytics, and sharing, while extracting some meaningful knowledge, are also arising. After collection, storage, and analysis of relevant data, its utilization by quantifying and communicating the value is another troublesome task, since the process is costly; thus, the revenue generation must be worth the investment. In short, moving from data to insight is a major challenge. Data scientists and engineers are employed to observe the quality of data, as poor and erroneous data might lead to poor business decisions. Semantics' lack in repositories is due to schemaless properties, which results in hindrance for data analysts to extract some meaningful structure and parse it. Semantic webs are helpful in such scenarios since they define concepts and their relationships in the form of semantic graphs [3][4].

The idea is to devise and introduce a framework that makes the collected data completely useful for the organization, and despite consuming it fractionally, which might miss some critical turning points for decisions. Since millions of data points out of billions of data sets are used thus it's a wastage of resources and energies. Thus, the concept of this research is to propose a solution that assists in devising effective rules while considering multiple attributes at a time to get insight into business trends and the effects of attributes. Considering the social and economic business issues, related data sets were gathered, and their relationships were identified. After removing noisy and polluted data sets, useful data sets were extracted that were further utilized in designing dimensions and cubes where data could be analyzed in a better manner in the presence of multiple dependent attributes.

TRADITIONAL DATA VS BIG DATA

Traditional data systems, such as relational databases, were the major source for business organizations and companies to store and analyze data just a few years ago. The systems were primarily designed to handle structured data, and one of the most defining characteristics of the system was its well-organized data. Other data storage techniques existed, but these were the most commonly utilized [5][6]. In a single computer, traditional data handles vast and complicated issues. It relied on centralized design, which itself is expensive and inefficient for huge data volumes; on the other hand, big data is based on a distributed database system. Big volumes of datasets are split into tiny groups and handled under this architecture; the answer to a particular issue is computed by various components existing in a specific computer network. Those computers communicate with each other and come up with a solution to the problem [7][8].

Distributed databases are more affordable, work better, and offer superior computations. The distributed architecture uses microprocessors. This is cheaper than centralized architecture, which relies on mainframes. However, distributed architecture has a greater computing power than traditional architecture [9][10].

Traditional database systems work with structured data, but big data works with both semi-structured and unstructured data. Traditional databases hold modest amounts of data ranging from a few gigabytes to a few terabytes, whereas big data can store and analyze data in the hundreds of terabytes or petabytes range and beyond. Storing a significant volume of data lowers costs, which aids Business Intelligence (BI) [11][12].

WHY TO ANALYZE BIG DATA?

The word analysis refers to the detailed examination of anything complex to generate the required results [13]. The process of analysis involves the in-depth study of a particular subject with the help of various tools and methods [14]. Big data refers to a huge amount of data that is being analyzed through different techniques, tools, and provides the results that directly facilitate the decision-making process. These results can be used by business organizations for the improvement of their strategies [15].

Decision-making process for business organizations and cooperate sectors makes use of data analysis techniques to generate the required information that gives the details of their customers and helps them to streamline customer preferences to revise or design their business strategies [16]. In this way, big data analysis gives them better knowledge and understanding of their customers and their preferences [17]. Similarly, the data gathered through various social media resources can also be analyzed by the retailers and corporate sector, which aids in the progress of their business [18].

In previous research to improve business decisions, a data set was gathered through surveys and interviews, but the results obtained by this methodology are not satisfactory because they vary from person to person. A big force is required to carry out this research. This process is time-consuming, and data integrity is compromised. Organizations must focus on adopting big data for reliable and accurate analysis [19].

Now, people not only desire to get information, but they would like to know the significance and need for their data, and also put it to use to support them in decision-making [19]. Big data analytics would be the Procedure for implementing calculations to analyze datasets and extract anonymous and useful patterns, connections, along with information [20].

Big data analysis involves the use of applications and tools to analyze data. Different organizations use tools and applications for analyzing huge data sets to gather useful information from them. This information is used to draw conclusions and aids in the development of successful business strategies [21].

Big data analysis gives opportunities for business strategy makers to discover the hidden patterns of information in huge data sets. It also enables the use of information to find the correlations that are useful for business strategies [22].

BIG DATA TOOLS AND TECHNIQUES

1. OLTP

Online Transaction Processing (OLTP) uses traditional DBMS methods to capture, store, and process data. They aren't well adapted to analyzing and displaying massive volumes of data in real-time [23][24].

2. DATA MINING

Data mining is the core process for knowledge discovery to extract useful patterns from data and identify outliers. It has limitations when it comes to accuracy. Information gathered can be inaccurate, making. Data mining, if combined with big data frameworks, might assist the policy makers in making more accurate decisions [25].

no chance of missing significant facets. Firstly, for

3. CLUSTERING

Clustering is an unsupervised machine learning technique for finding and classifying related data points in big datasets without regard for the result. Clustering (also known as cluster analysis) is a technique for organizing data into structures that are easier to comprehend and manage. The outcomes of several clustering techniques are generally extremely varied. Except for simple linkage, the sequence in which the parameters are sorted has an impact on the outcomes, which leads to less accuracy [25].

4. TALEND OPEN STUDIO

Talend Open Studio is used to compute large volumes of datasets. It gives user a graphical environment to perform their analysis visually, but this tool has insufficient predictive capacities and management issues. The system becomes slow after its installation [26].

5. QLIK VIEW

The Qlik View tool is used for the purpose of converting raw data into knowledge. It provides data integration and data analytics solutions. The limitations of this tool are that it's not enterprise-ready, not able to execute complex queries, and risky to current customers [27].

PROPOSED FRAMEWORK

Consider data as a highly valuable commodity, but singular data points are useless, no matter how much data one has, until and unless they identify the linkages and relations between them

to identify the current interests and trends while observing the affecting dominant factors [28]. The critical part is to identify the links between the data sets because the commonalities, relations, and outliers are the driving forces.

To produce useful outcomes, one must generate insight into data promptly. Since this is an expensive process thus the results must be equivalent or greater in terms of revenue and profits. Data gathered must be utilized completely while extracting the sample sets, so that there is

designing rules assisting decision-making, data was gathered from multiple sources. Since the target was market analysis thus the data concerning yearly business sales, employee sales, and customer preferences or customer satisfaction was compiled and observed. For the sake of data analysis and cleaning, Excel files were imported into the data warehouse, and before identifying relationships among entities, data was cleaned against duplicate entries, special characters, and null values. Without any queries, relations among data entities were established for the better analysis of data sets, since the rules' accuracy and reliability are directly related to the quality of data. Data views and dimensions were formed while considering the related entities that were further used in designing cubes. Figure 1 shows the proposed framework.

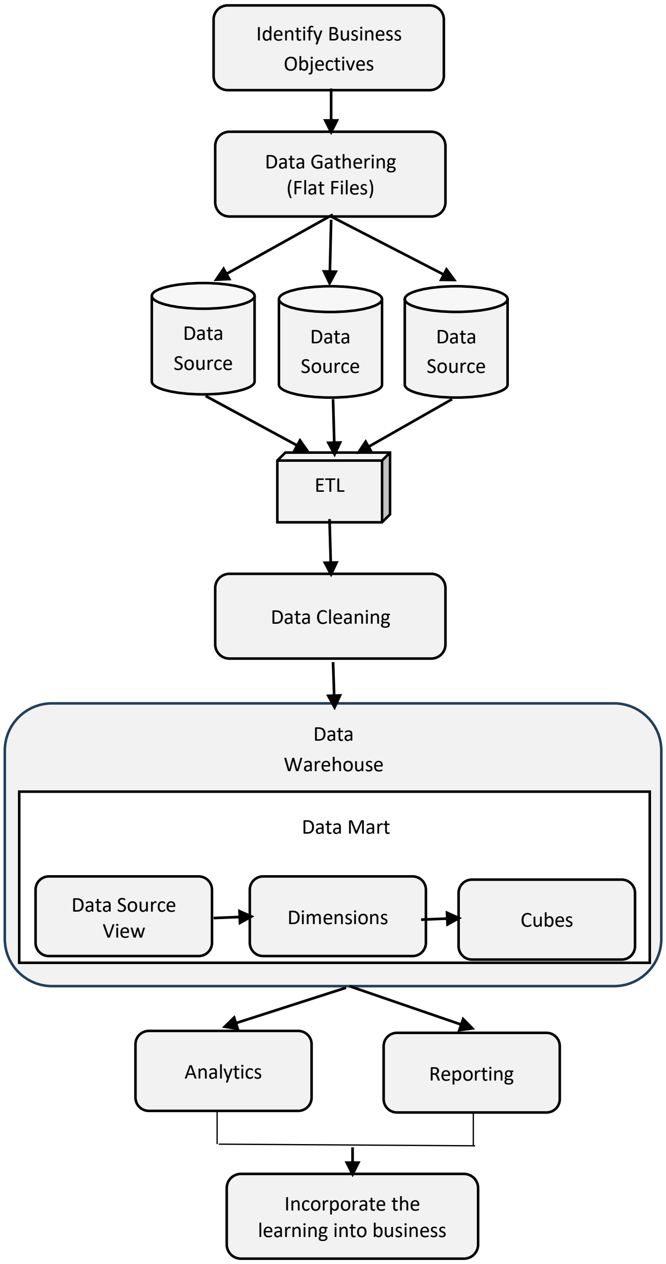

Figure 1: Proposed Architecture

ONLINE ANALYTICAL PROCESSING (OLAP)

Online analytical processing is designed for the multidimensional analysis of significant business data. OLAP ensures an opportunity to perform calculations on typically small amounts of information. It provides its users with essential insights based on the data provided by various sources and databases. The main feature of OLAP is its ability to represent data and item relations in a multidimensional structure to be displayed in the form of cubes. This makes it possible to view data from various perspectives. Users can carry out holistic analysis by receiving responses to ad-hoc queries. OLAP Queries run very fast, even given large volumes of data, owing to precomputed aggregations.

Both OLAP and data mining allow handling and analyzing large amounts of data from numerous angles, such as sales by segment, by market, by country, by employee, by product, etc. In general, OLAP enables organizations to successfully perform reporting, analysis, budgeting, etc. The information obtained by this BI system commonly provides privileged information contributing to decision-making.

RESULTS AND WORKFLOW

The targeted data set is for market-based analysis that highlights the facts that could be used to improve customer satisfaction, business sales, and employee performance by assisting in designing such rules that could facilitate decision makers in their decisions.

The flow of working and results of the implementation process are shown as follows:

1. DATASET GATHERING

DATASET 1:

The initial step is to gather a targeted dataset, so a dataset is gathered that is about a global superstore that mainly contains information about worldwide customers, employees, products, and business sales.

DATASET 2:

Then we created another Excel file by the name of Returns, which mainly contains information about the products returned in the market and the Order ID of the Orders table as a foreign key for linking data.

2. ETL (extract, transform, and load)

Importing datasets from different Excel source files into a single platform, i.e., data warehouse, is obtained by ETL. The SQL Server integration services (SSIS) package is created in Visual Studio, through which ETL (extract, transform, and load) is performed.

Data mapping facilitates data integration and migration from Excel source files to destination files. The ETL process is depicted in Figure 2.

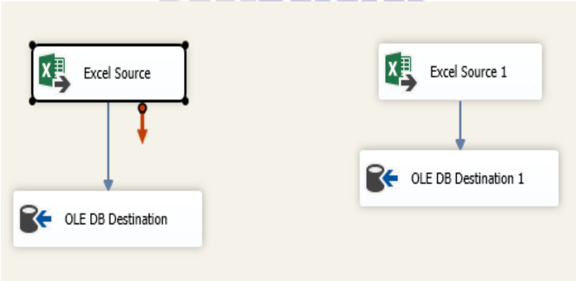

Figure 2: Data Warehouse

3. DATA CLEANING

Data cleaning, as shown in Figure 3, is performed by which Null values and special characters were identified and removed from the Excel source file. So that whenever we apply any query, it can easily extract data from the dataset file without causing any error. This was done by using trim queries.

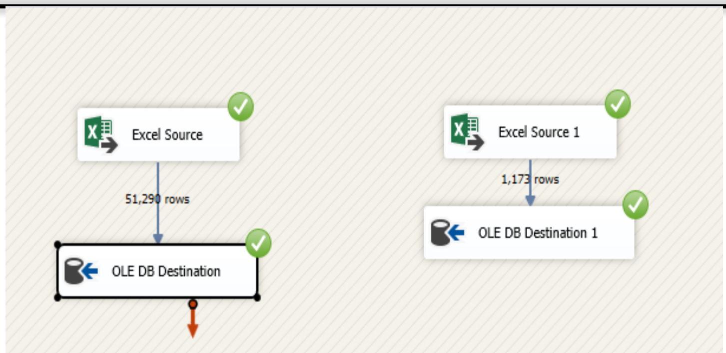

Figure 3: Data Cleaning

4. IDENTIFYING RELATIONSHIPS

After importing data, the logical data model is created in SSMS, where relationships are established between the fact table and dimension tables. The logical data model from the perspective of the Global Market is given in Figure 4.

5. OLAP IMPLEMENTATION

SQL Server Analysis Services (SSAS), which is an MSBI tool, was utilized in business intelligence for creating OLAP data source views, dimensions, and a cube of data.

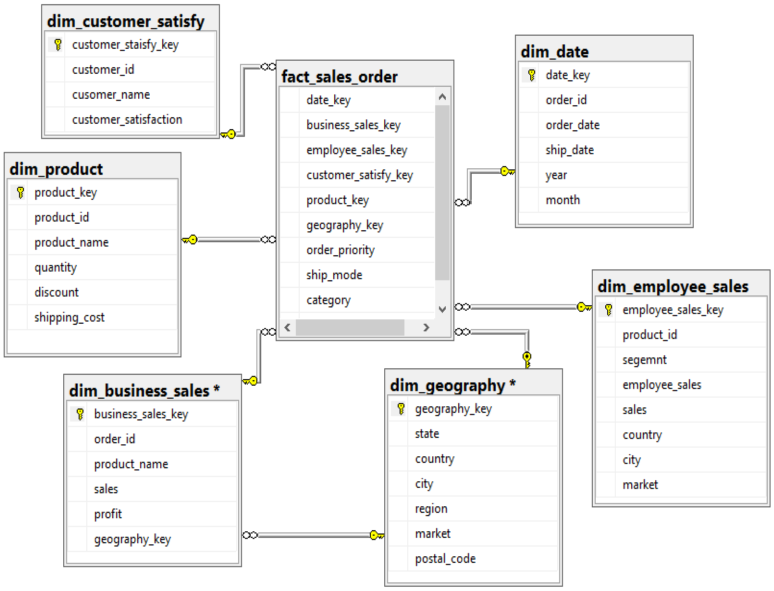

Figure 4: Star schema data mart design

DATA SOURCES AND DATA SOURCE VIEW

Data source view in SQL Services Analysis Services (SSAS) is a collection of tables or views from a database that is needed for designing a cube (multidimensional dataset), e.g. if the data warehouse comprises 10 tables but we need only 5 tables for the designing of the cube then there isn't any point in adding 10 tables, rather than the data source view give space for adding the required 5 tables. So, therefore, after specifying the relationships, data sources and their views were formed as Figure 5.

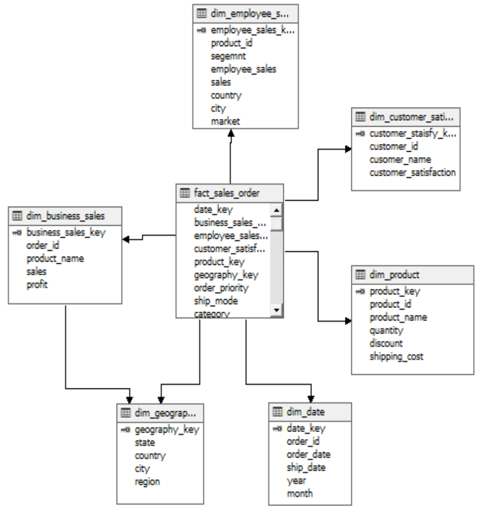

Figure 5: OLAP Data Source View

DIMENSIONS

Dimensions are a collection of relevant terms utilized for organizing data, e.g., a region dimension is made up of state, country, city, and region. Dimensions are created for forming data structures for a cube because cubes are made up of dimensions.

To analyze the yearly business sales, dimensions were designed using attributes date, products, countries, and their sales rate, shown in Figure 6.

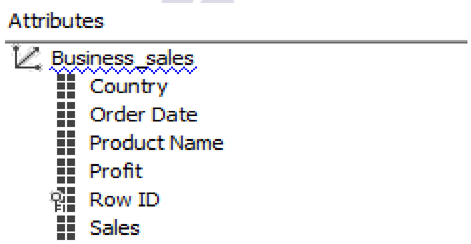

Figure 6: Business Sales Dimension

To analyze the sales made by employees, dimensions were designed using attributes of products, segments, countries, and their employee sales shown in Figure 7.

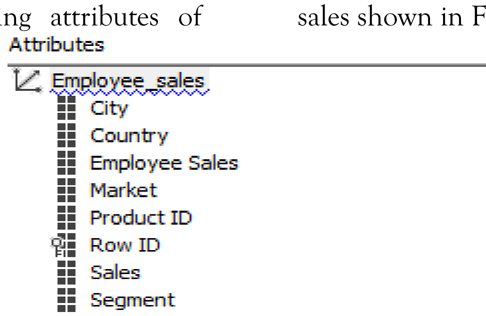

Figure 7: Employee Sales Dimension

To analyze customer satisfaction, dimensions were designed using attributes of customers, products, countries, market, and their satisfaction levels shown in Figure 8.

Other dimensions could also be designed to explore further dependencies among entities for refining results.

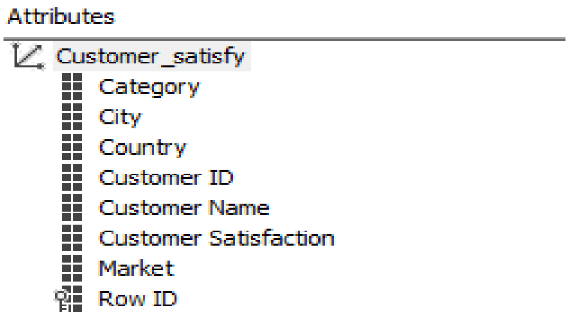

Figure 8: Customer Satisfaction Dimension

CUBE

A multi-dimensional cube as in Figure 9, for reporting market information globally is made up of seven dimensions. Each dimension represents some attributes in the database.

Figure 9: OLAP Cube

DATA ANALYSIS AND REPORTING

For reporting's SQL Server Reporting Services (SSRS) package was installed in Visual Studio, MDX queries were applied to the data cube in SSMS, and then reports were generated, including Yearly Business Sales Figure 10, Employee Sales Figure 11, and Customer Satisfaction Figure 12.

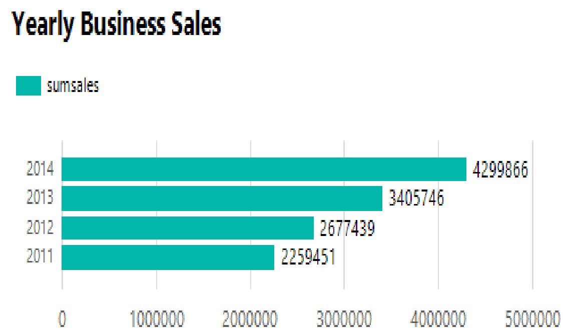

Figure 10: Market sales by year

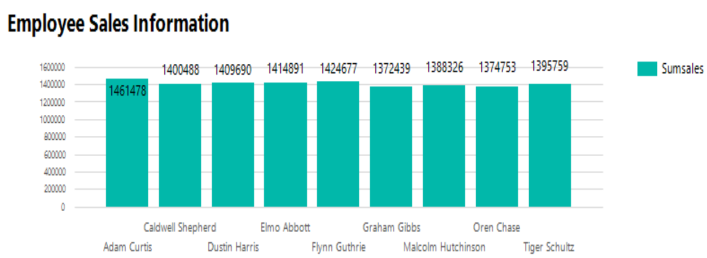

Figure 11: Employee Performance by Sales

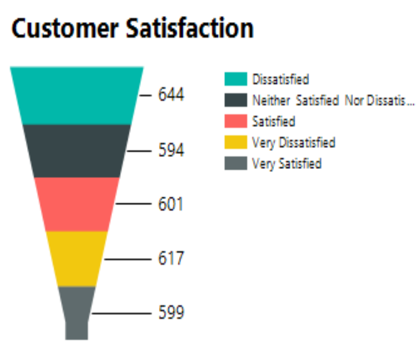

Figure 12: Customer Loyalty

According to the reports and the above figure, the customer loyalty in percentage is given in Table 1.

Table 1: A Percentage measure of Customer satisfaction

Count	Customer Loyalty Measures	Percentage
644	Dissatisfied	21%

594	Neither satisfied nor dissatisfied	19%
601	Satisfied	20%
617	Very Dissatisfied	20%
599	Very satisfied	20%
Total responses = 3055		

DISCUSSION

Therefore, the browsing of all these contents could promote decision making, which means customers are very likely to promote which behavior in the past, employee performance in business, and business revenue in past years. This confirms that related recommendations provided by big data analytics could capture the intention and behavior of consumers, skilled and less dedicated employees, strengths and weaknesses of businesses. Decision-makers can offer recommendations fitting the demand of customers, employees, thus improving market purchasing power, and hence increasing the country's economy, and the unemployment rate would be reduced.

The more the attributes and dimensions are involved in designing cubes, the more reliable and effective results can be generated, and their quality can be improved by introducing improved data cleaning techniques.

CONCLUSION

Through the integration of heterogeneous datasets using ETL, data cleansing, warehousing, and OLAP (cubes and dimensions), organizations can uncover high-value signals. These include factors like loyal customer segments, high-performing and skilled employees, profitable regions, and reliance on import-dependent inputs. Moreover, it was found that such parameters directly inform strategic decisions. Utilizing these data facilitates focused retention and acquisition, workforce enhancement, optimal site selection, and more strategic sourcing. At scale, data-driven decisions can enhance corporate growth, create job opportunities (particularly for the young), and mitigate macro-level vulnerabilities, thereby bolstering national productivity.

Recognizing this value depends on the quality of execution across the data lifecycle. Processing substantial volumes with suitable latency; assuring availability "when required"; verifying data as it

transitions between systems; and maintaining integrity and privacy consistently. These criteria support this research's three foundational elements, such as interoperable infrastructure, human capital, and governance. Furthermore, these are considered essential prerequisites for converting analytics into a competitive advantage. Incorporating these measures, firms may proactively address market fluctuations, adapt to swift product and service cycles, and methodically transform data-identified opportunities into quantifiable results. Future research must emphasize resilient, automated data-quality controls; privacy-preserving integration techniques; and near-real-time analytical pipelines. Decision frameworks that amalgamate predictive models with causal evaluation to guarantee that proposed actions effectively yield the desired outcomes may also be suggested.

REFERENCES

[1] A. Kaur, "Big Data: A Review of Challenges, Tools and Techniques," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 1090–1093, 2017.
[2] S. MacFeely, "The Big (data) Bang: Opportunities and Challenges for Compiling SDG Indicators," *J. Big Data*, vol. 10, no. 5, pp. 121–133, Jan. 2019.
[3] J. N. Rao and M. Ramesh, "A Review on Data Mining & Big Data, Machine Learning Techniques," *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 2277–3878, 2018.
[4] N. Elgendy and A. Elragal, "Big data analytics: A literature review paper," *J. Big Data*, vol. 4, no. 3, pp. 214–227, 2019.
[5] S. Borodo, S. M. Shamsuddin, S. Hasan, and S. M. Borodo, "Big Data Platforms and Techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 1, no. 1, pp. 191–200, 2017.

[6] S. Bejjam, M. Seshashayee, B. Suvarnamukhi, and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.
[7] A. B. Sriramoju, N. Vijay, and G. Ramesh, "An Overview of Big Data Challenges, Tools and Techniques," *Int. J. Res. Appl.*, vol. 4, no. 16, pp. 596–601, 2017.
[8] S. Verma, "Big Data and advance analytics: Architecture, techniques, applications, and challenges," *Int. J. Bus. Anal.*, vol. 4, no. 4, pp. 21–47, 2018.
[9] P. Prasdika and B. Sugiantoro, "A Review Paper on Big Data and Data Mining Concepts and Techniques," *Int. J. Informatics Dev. (IJID)*, vol. 7, no. 1, p. 35, 2018.
[10] J. Samosir, M. Indrawan-Santiago, and P. D. Haghighi, "An evaluation of data stream processing systems for data driven applications," *Procedia Comput. Sci.*, vol. 8, no. 5, pp. 439–449, Jan. 2016.
[11] E. Battisti and S. Shams, "Big data and risk management in business processes: implications for corporate real estate," *Bus. Process Manag. J.*, vol. 26, no. 5, pp. 1141–1155, Oct. 2020.
[12] P. Mannava, "A Study on the Challenges and Types of Big Data," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 2, no. 8, pp. 4145–4149, 2017.
[13] B. Punjabi and H. Sonal, "A Research on Big Data Analysis & Processing With Data Mining Techniques," *Int. J. Sci. Eng. Appl. Sci.*, vol. 1, no. 9, pp. 115–122, 2016.
[14] S. Singh, T. Firdaus, D. A. K. Sharma, and M. Tech Scholar, "Survey on Big Data Using Data Mining," *Int. J. Eng. Dev. Res.*, vol. 3, no. 4, pp. 2321–9939, 2018.
[15] N. Garg, S. Singla, and S. Jangra, "Challenges and Techniques for Testing of Big Data," *Procedia Comput. Sci.*, vol. 85, pp. 940–948, Jan. 2016.
[16] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 989–1037, Feb. 2020.
[17] R. Kitchin and G. Mcardle, "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets," *Big Data Soc.*, vol. 3, no. 1, pp. 1–10, 2016.
[18] P. Géczy, "Big data characteristics," *J. Big Data*, vol. 3, no. 6, pp. 94–104, 2019.
[19] J. Bughin, "Big data, Big bang?," *J. Big Data*, vol. 3, no. 1, pp. 1–14, Dec. 2016.
[20] M. Abdullah, N. M. Alotaibi, and M. A. Abdullah, "Big data mining: A classification perspective," *Commun. Manag. Inf. Technol.*, vol. 7, no. 4, pp. 687–695, 2018.
[21] M. Uma and D. V. B. Deepa, "Big Data Analytics in Data Mining—A Review," *Int. J. Appl. Eng. Res.*, vol. 13, no. 7, pp. 15386–15396, 2018.
[22] K. Siddardha and C. Suresh, "Big Data Analytics: Challenges, Tools and Limitations," *Int. J. Eng. Tech. Res.*, vol. 6, no. 3, pp. 40–44, 2018.
[23] M. Vesterinen, J. Mero, and M. Skippari, "Big data analytics capability, marketing agility, and firm performance: A conceptual framework," *J. Mark. Theory Pract.*, vol. 33, no. 2, pp. 310–330, 2024, doi: 10.1080/10696679.2024.2322600.
[24] I. A. Ajah and H. Friday Nweke, "Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications," *Big Data Cogn. Comput.*, vol. 3, no. 2, pp. 1–30, 2019.
[25] R. Jony, R. Rony, A. Rahat, and R. M., "Big data characteristics, Value chain and challenges," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 3, no. 2, pp. 1–6, 2018.
[26] A. Nath, R. Toshniwal, and K. Ghosh Dastidar Asoke Nath, "Big Data Security Issues and Challenges," *Int. J. Innov. Res. Adv. Eng.*, vol. 2, no. 2, pp. 15–20, 2019.
[27] J. Zakir, T. Seymour, and K. Berg, "Big data analytics," *Issues Inf. Syst.*, vol. 16, no. 2, pp. 81–90, 2017.
[28] S. Zeng and K. J. Fox, "Productivity Measurement With Big Data: A Data-Driven Approach Capturing Firm Heterogeneity," early view, Jun. 2025, doi: 10.1111/1467-8462.70014.