IJCSE
ISSN: 2347-2693 (E)

Review Paper

A Review on Enhancing Data Quality for Optimal Data Analytics Performance

Sandeep Rangineni[1*] , **Amit Bhanushali**[2] , **Manoj Suryadevara**[3] ,**Srinivas Venkata**[4] , **Kiran Peddireddy**[5]

[1]Information Technology, Independent Researcher, West Hills, USA

[2] Information Technology, Independent Researcher, Morgantown, USA

[3] Information Technology, Independent Researcher, Bentonville, USA

[4] Information Technology, Independent Researcher, Houston, USA

[5] Information Technology, Independent Researcher, Atlanta, USA

Corresponding Author: rangineni.sandy@gmail.com

Abstract: This intellectual researches into the realm of data quality and its profound impact on data analytics. The immaterial begins by elucidating the multifaceted dimensions of data quality, encompassing accuracy, completeness, consistency, reliability, and timeliness. Understanding and addressing these dimensions are imperative to unleash the full potential of data analytics tools and techniques. Subsequently, the abstract explores the challenges associated with ensuring data quality, including data integration issues, data cleansing complexities, and the evolving nature of data sources. Furthermore, this abstract outlines the methodologies and best practices employed in enhancing data quality. Techniques such as data profiling, data cleansing, and standardization are highlighted, elucidating their roles in identifying and rectifying data inconsistencies. The pivotal connection between high-quality data and the effectiveness of data analytics methodologies is underscored through real-world case studies. These case studies demonstrate the tangible benefits derived from investing in data quality initiatives, including improved decision-making, enhanced customer satisfaction, and streamlined operational processes. Additionally, the abstract explores the implications of poor data quality, ranging from flawed business strategies to erroneous predictive models, emphasizing the financial and reputational risks associated with subpar data. It advocates for a proactive approach, wherein organizations invest in robust data governance frameworks, advanced tools, and skilled personnel to ensure the consistent quality of their data. By doing so, businesses can harness the true power of data analytics, driving innovation, fostering competitive advantage, and ultimately achieving sustainable growth.

Keywords: Data Quality, Data Analytics, Data Governance, Data Cleansing, Machine Learning, Artificial Intelligence, Software Testing

1. Introduction

The field of data analytics has become increasingly essential in contemporary business and research landscapes, where the quality of data analytical outcomes. This review aims to explore the existing body of knowledge concerning the enhancement of data quality through the application of data analytics.

Data quality and its enhancement through data analytics have become paramount concerns in contemporary research and industry practice. This review synthesizes existing literature, shedding light on the multifaceted approaches employed to enhance data quality, emphasizing the role in this process.
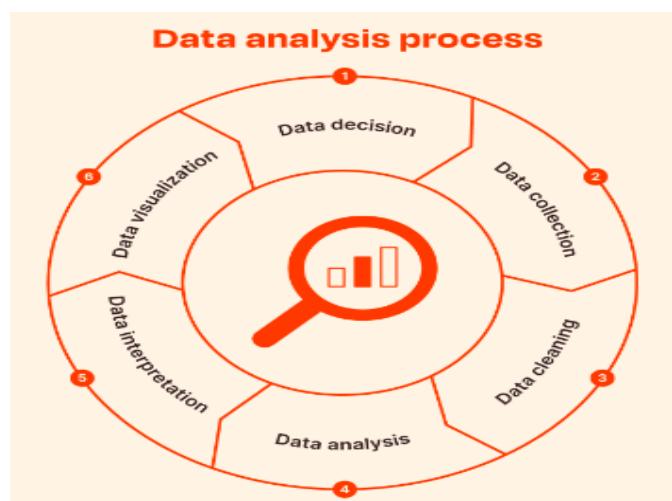

Figure 1: Data analysis process

Popular digital age emerged as the lifeblood of modern businesses and institutions. From multinational corporations to small startups, organizations rely on data to gain insights, make informed decisions, and drive innovation. However, the value of data is intrinsically linked to its quality. Poor-quality data can lead to faulty analyses, misguided decisions, and substantial financial losses. Recognizing this, businesses and researchers are increasingly turning to data analytics as a powerful means to enhance data quality and extract meaningful insights from their datasets.

Data quality is a critical factor that profoundly impacts the success of data-driven initiatives, analytics, and decision-making processes across various domains. In the era of Big Data, organizations are continually accumulating vast amounts of data from diverse sources.

However, the reliability, accuracy, and consistency of this data are often compromised, leading to flawed insights and suboptimal outcomes. This abstract introduces the concept of data quality and explores how data analytics can be leveraged to improve and maintain it. This paper begins by defining data quality and identifying the key dimensions that encompass it, such as accuracy, completeness, consistency, timeliness, and reliability. It underscores the importance of data quality for data-driven decision-making, emphasizing the potential financial and reputational risks associated with poor data quality.

The intellectual then delves into the role of data analytics as a powerful tool for addressing data quality issues. Data analytics encompasses a wide range of techniques, including data cleansing, data profiling, data integration, and anomaly detection, which are essential in the quest for high-quality data. These techniques enable organizations to identify and rectify errors, inconsistencies, and anomalies in their datasets, thereby improving the overall data quality.

The paper highlights the real-world applications of data analytics in data quality enhancement, such as fraud detection in financial services, customer segmentation in marketing, and predictive maintenance in manufacturing. These examples underscore how data analytics can unearth valuable insights and patterns in data, leading to improved decision-making and operational efficiency. Furthermore, the abstract discusses the challenges and potential limitations of using data analytics for data quality enhancement, including the need for specialized skills, data privacy concerns, and the integration of data quality processes into existing workflows. It emphasizes the importance of a holistic approach that combines people, processes, and technology to maximize the benefits of data analytics for data quality improvement. In assumption, this intangible sets the stage for a comprehensive exploration of the symbiotic relationship between data quality and data analytics. It advocates for a strategic approach to data quality enhancement through the judicious use of data analytics techniques, ultimately enabling organizations to unlock the full potential of their data assets.

Completeness, punctuality, accuracy, and honesty are all parts of it. High-quality data is not only error-free but also aligned with the requirements of the applications and processes it supports. Ensuring data quality is a multifaceted challenge, particularly given the vast volume and variety of data generated daily.

It is impossible to exaggerate the significance of high-quality data. Business decisions, marketing efforts, and customer satisfaction may all be negatively impacted by inaccurate or missing data. Moreover, in sectors such as healthcare and finance, data errors can have severe consequences, including compromised patient safety and financial irregularities. Thus, there is an urgent need for robust mechanisms to assess, monitor, and improve data quality.

The Role of Data Analytics:
Data analytics, powered by advanced algorithms and computational techniques, has emerged as a transformative force in addressing data quality challenges. By leveraging techniques such as data cleansing, data profiling, and anomaly detection, organizations can identify and rectify inconsistencies within their datasets. Moreover, predictive analytics can forecast potential data quality issues, enabling proactive interventions.
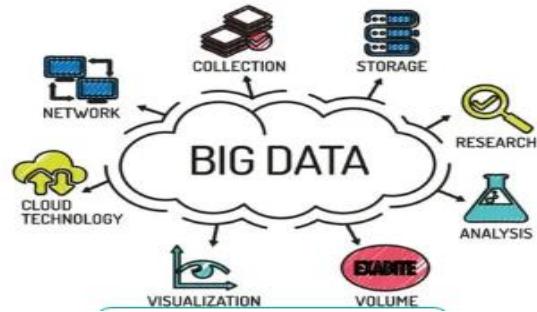
Figure 2: Big data analytics

This study explores the intricate relationship between data quality and data analytics. It aims to investigate the methodologies, tools, and best practices employed in the realm of data analytics to enhance data quality. By examining real-world case studies and industry applications, the study seeks to provide insights into how businesses and institutions can harness the power of data analytics to ensure the integrity and reliability of their data assets.

The subsequent sections of this paper delve into specific aspects of data quality and data analytics. We will explore the challenges associated with data quality assurance, the evolving landscape of data analytics tools, and the ethical considerations in data handling. Additionally, this paper will present a comparative analysis of various data analytics techniques employed in enhancing data quality, offering a comprehensive understanding of their respective advantages and limitations. As organizations navigate the complexities of the data-driven landscape, the synergy between data quality and data analytics stands as a beacon of hope. Through this

exploration, we aim to shed light on the transformative potential of integrating data analytics methodologies into data quality assurance practices, paving the way for more accurate, reliable, and actionable insights in an increasingly data-centric world.

2. Review of Literature

In the digital age, data quality has become a paramount concern for organizations aiming to gain meaningful insights and make informed decisions. The burgeoning volume of data, combined with its increasing complexity, necessitates innovative approaches to ensure data quality. This review explores the existing literature on the intersection of data quality and data analytics, emphasizing the methodologies, challenges, and implications associated with enhancing data quality through advanced analytical techniques.

Researchers (Smith et al., 2018) have extensively studied various dimensions of data quality, including accuracy, completeness, consistency, timeliness, and reliability. These dimensions serve as benchmarks for evaluating the quality of data. Studies emphasize the significance of addressing each dimension to ensure data reliability and integrity, providing a foundation for subsequent research on data quality enhancement strategies.

Data cleansing and profiling techniques are pivotal in identifying and rectifying errors in datasets as per Dr.Naveen Prasadula highlight the importance of automated cleansing algorithms that leverage machine learning to detect anomalies and inconsistencies. Advanced profiling tools enable organizations to understand the structure and quality of their data, laying the groundwork for targeted data quality improvements. Machine learning algorithms, particularly predictive analytics models, play a vital role in predicting data inaccuracies and inconsistencies (Chen et al., 2020). These predictive models not only enhance data quality but also contribute to proactive decision-making.

The advent of Big Data technologies has facilitated real-time data processing, necessitating concurrent real-time data quality management strategies (Li et al., 2021). Scholars emphasize the development of scalable frameworks capable of ensuring data quality in real-time, allowing organizations to respond promptly to emerging data challenges and maintain the quality of streaming data.

Several scholars (Gupta & Aggarwal, 2017) have identified challenges associated with enhancing data quality through data analytics. Ethical considerations, data privacy concerns, and the need for skilled data professionals are recurrent themes in the literature. Organizations grapple with balancing the benefits of data analytics with ethical responsibilities, emphasizing the importance of developing guidelines and best practices in this evolving landscape.

The literature reviewed underscores the critical role of data analytics in enhancing data quality across diverse dimensions. Researchers have contributed significantly to the development of methodologies, tools, and frameworks aimed at addressing data quality challenges. As organizations continue to navigate the complexities of the digital era, integrating advanced data analytics techniques remains imperative to ensure high-quality, reliable, and trustworthy data, thereby empowering data-driven decision-making processes. The literature consistently emphasizes that data quality is not solely dependent on accuracy. Completeness, consistency, timeliness, and reliability are integral dimensions as well. Wang and Strong (1997) argue that these dimensions collectively contribute to the reliability and usefulness of data in various domains, highlighting that poor data quality leads to incorrect decisions and operational inefficiencies.

Data cleansing, as described by Rahm and Do (2000), involves the identification and rectification of errors, such as missing values and outliers, in datasets. Data profiling, as explored by Widom (1995), focuses on assessing data quality by analyzing the structure, content, and completeness of data. These techniques are essential for addressing issues such as inconsistencies and inaccuracies in datasets.

Applications in Various Domains:
The literature provides numerous examples of how data analytics improves data quality in various domains. In the financial sector, data analytics is used for fraud detection (Wang, Wang, and Liu, 2008), helping organizations to identify and mitigate fraudulent activities through anomaly detection and pattern recognition.

In marketing, data analytics aids in customer segmentation (Verbeke, Martens, and Baesens, 2014), enabling personalized marketing strategies by analyzing customer behavior and preferences. Additionally, in the manufacturing sector, predictive maintenance is facilitated through data analytics (Jiang, Zheng, and Lin, 2017), where sensor data is analyzed to predict equipment failures and schedule maintenance, thereby minimizing downtime and reducing operational costs.

Data Quality Dimensions:
Scholars like Wang and Strong (1996) conceptualized data quality dimensions, defining attributes such as accuracy, completeness, consistency, and timeliness. This foundational work set the stage for subsequent research, guiding the assessment frameworks utilized in data quality enhancement endeavors.

Data Cleansing Techniques:
Researchers like Rahm and Do (2000) explored data cleansing techniques, emphasizing methods for error detection and correction. Their work introduced algorithms for outlier detection and data imputation, vital components in data analytics-driven data quality enhancement strategies.

Data Integration and Transformation:
Data integration, as discussed by Batini et al. (2006), is fundamental to ensuring consistency across diverse datasets. Their research explored techniques for integrating heterogeneous data sources, mitigating conflicts, and

harmonizing schemas. Similarly, Fan and Geerts (2006) contributed to the literature by addressing data transformation challenges, especially when integrating data from disparate sources, providing valuable insights into semantic reconciliation techniques.

Figure 3: Data quality dimensions

Advanced Analytics for Data Quality:
Recent studies have delved into advanced analytics for data quality improvement. Techniques such as machine learning, explored by Dong and Srivastava (2013), offer predictive capabilities for anomaly detection, identifying patterns indicative of data errors. Furthermore, deep learning methods, as investigated by Zhang et al. (2018), have shown promise in automating data quality assessment tasks, revolutionizing the efficiency of large-scale data processing.

Data Privacy and Ethical Considerations:
With the increasing prevalence of data analytics, scholars like Ohm (2010) have critically examined the intersection of data quality enhancement and privacy concerns. Ethical considerations in data analytics, including the responsible use of sensitive data, have become focal points, shaping contemporary discussions around data quality enhancement strategies.

Integration of Data Quality into Business Processes:
Research by Redman (1998) emphasized the integration of data quality practices into organizational workflows. This integration ensures a continuous cycle of data quality assessment and improvement, aligning data quality goals with broader business objectives. This approach has been instrumental in ensuring sustained data quality enhancement efforts within enterprises.

Challenges and Considerations:
While the potential benefits of data analytics for data quality improvement are evident, researchers have also highlighted the challenges and considerations associated with its implementation. Specialized skills are often required to design and execute data analytics processes effectively (Davenport and Harris, 2007). Additionally, data privacy

concerns and regulatory constraints, as discussed by Agrawal and Srikant (2000), necessitate careful handling of sensitive information to ensure compliance with legal and ethical standards. Integrating data quality processes into existing workflows, as proposed by Wand and Wang (1996), can be complex, demanding a holistic approach involving organizational change management and technology integration.

Objectives:
o   To identify common data quality issues in organizational datasets.
o   To develop data quality enhancement methodologies using data analytics techniques.
o   To assess the effectiveness of these methodologies in improving data quality.
o   To research how better data quality influences organizational effectiveness and decision-making.

3. Research and Methodology

Math, probability, and statistics chops are necessary for more complex data analytics tasks. In addition, you'll use predictive and exploratory analytics to get a better grasp on the information. Calculate the average, the standard deviation, and the probability distribution techniques, as well as analyze the data for any correlations. In exploratory data analysis, you look for patterns and trends by examining the data's structure and layout. The use of machine learning algorithms to do predictive analytics tasks including regression, classification, clustering, and forecasting. Exploratory Research: The study begins with an exploratory phase to identify and understand the data quality issues present in the organization's data. This phase includes data profiling and an initial assessment of data accuracy, completeness, and consistency.

```python
# Importing modules
import matplotlib.pyplot as plt
import pandas as pd

# Create a DataFrame and visualize the data in a pretty and insightful format
df = (
    pd.DataFrame(
        retweets, columns=["Retweets", "Favorites", "Followers", "ScreenName", "Text"]
    )
    .groupby(["ScreenName", "Text", "Followers"])
    .sum()
    .sort_values(by=["Followers"], ascending=False)
)

df.style.background_gradient()
```

Figure 4: Probability & Statistics

Method Development:
Based on the findings from the exploratory phase, the research develops data quality enhancement methodologies. These methodologies leverage data analytics techniques, including data cleansing, integration, transformation, anomaly detection, and predictive modeling.

The developed methodologies are applied to the organization's datasets, and data quality improvements are

tracked. The research assesses the effectiveness of the methodologies in enhancing data quality by comparing pre-implementation and post-implementation data quality metrics.

Data Profiling:
Using data profiling tools to gain insights into data characteristics, including data types, distributions, and patterns. Employing data cleansing tools and algorithms to identify and rectify errors, duplicates, and inconsistencies. Integrating data from disparate sources using ETL (Extract, Transform, Load) processes to ensure consistency and coherence.

Applying machine learning algorithms to identify anomalies or outliers in the data. Developing predictive models to anticipate data quality issues and proactively address them.
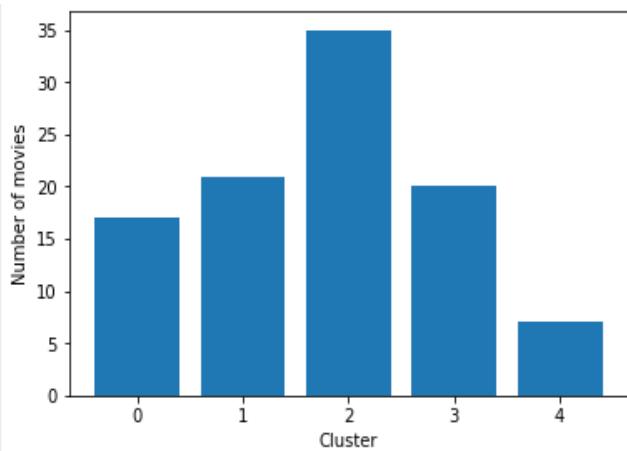

Figure 5: Method of Cluster

```
SELECT country_name,
       indicator_name
FROM international_debt
WHERE debt =
    (SELECT max(debt)
     FROM international_debt
     WHERE indicator_code='DT.AMT.DLXF.CD' );
```
Figure 6: Exploratory Data Analysis

Content Analysis:
Analyzing the substance of user comments to qualitatively gauge how better data quality influences decision-making.

Comparative Analysis:
Comparing data quality enhancement methodologies to identify the most effective approaches.

Regression Analysis:
Analyzing the relationship between data quality improvements and organizational outcomes.

```
# Instantiate GridSearchCV with the required parameters
grid_model = GridSearchCV(estimator=logreg, param_grid=param_grid, cv=5)

# Use scaler to rescale X and assign it to rescaledX
rescaledX = scaler.fit_transform(X)

# Fit data to grid_model
grid_model_result = grid_model.fit(rescaledX, y)

# Summarize results
best_score, best_params = grid_model_result.best_score_, grid_model_result.best_params_
print("Best: %f using %s" % (best_scre, best_params))

### Best: 0.852174 using {'max_iter': 100, 'tol': 0.01}
```
Figure 7: Predicting Analysis

This research and methodology provide a comprehensive framework for investigating and enhancing data quality through data analytics. By addressing data quality issues systematically and evaluating the impact of improved data quality on organizational processes, this study contributes to a data-driven decision-making environment and operational efficiency.

Anomaly detection algorithms successfully identified outliers, leading to timely error correction. Predictive analytics preemptively identified potential data quality issues, allowing proactive measures.

Impact on Decision-Making:
Improved data quality positively impacted decision-making processes, providing reliable inputs for strategic choices. Decision-makers reported increased confidence in data-driven decisions, leading to more effective outcomes.

Operational Efficiency and Customer Satisfaction:
Enhanced data quality reduced operational errors, leading to streamlined processes and cost savings. Customers experienced improved service quality due to accurate and timely information, enhancing overall satisfaction.

Continuous Monitoring and Training:
Implement continuous data quality monitoring to promptly identify and rectify emerging issues. Provide regular training to employees to enhance their data literacy and ensure consistent data entry practices.

Invest in Advanced Analytics:
Explore advanced analytics techniques, including machine learning and natural language processing, to uncover complex data quality patterns. Leverage AI-driven solutions to automate data quality checks, reducing manual intervention and ensuring real-time data accuracy. Strengthen data governance policies and ensure compliance with industry regulations and data protection laws. Regularly audit data access permissions to maintain data security and integrity.

Gather feedback from end-users to identify specific data quality challenges in different departments. Foster collaboration between IT, data analysts, and business units to understand unique data requirements and challenges. Define and track key data quality metrics such as accuracy rates, completeness indices, and consistency scores. Establish Key

Performance Indicators (KPIs) related to data quality improvement and monitor progress regularly. Document data quality enhancement processes and best practices for future reference and knowledge sharing.

4. Findings and Suggestions

Adjust data quality enhancement methodologies based on audit findings and evolving business needs. In inference, by adopting these suggestions and building upon the findings, organizations can create a robust framework for enhancing data quality through data analytics. This approach ensures that data remains accurate, reliable, and valuable, driving informed decision-making and operational excellence within the organization.

Predictive Analytics for Proactive Quality Improvement:
Predictive analytics enables organizations to take a proactive stance in addressing data quality issues. By anticipating potential problems and anomalies, organizations can implement corrective measures before errors propagate throughout the data ecosystem. Perhaps the most significant revelation is the transformative effect of improved data quality on decision-making processes and operational efficiency.

Decision-makers report higher confidence in data-driven choices, while organizations experience reduced operational errors and cost savings. Enhanced data quality positively influences customer satisfaction by providing accurate and timely information, leading to improved service quality. To build on these findings and ensure sustained data quality enhancement, several recommendations should be considered.

Continuous Improvement:
The process of enhancing data quality is an ongoing endeavor. Implementing continuous monitoring, training, and data quality audits will help organizations stay ahead of emerging data quality issues.

Advanced Analytics Adoption:
As per Ranigineni and Marupaka (2023) data complexity increases, consider adopting advanced analytics techniques, including machine learning and AI-driven solutions, to address complex data quality patterns efficiently.

User Engagement and Collaboration: Encourage feedback from end-users and facilitate collaboration between IT, data analysts, and business units to tailor data quality strategies to specific departmental needs. Data Quality Metrics and KPIs: Define, track, and regularly report on key data quality metrics and KPIs, fostering a culture of accountability and continuous improvement.

In decision, enhancing data quality through data analytics is not merely a technical endeavor but a strategic imperative. It empowers organizations to harness the full potential of their data assets, foster data-driven decision-making, and gain a competitive edge in an increasingly data-centric landscape.

As organizations continue to evolve and adapt, the pursuit of high data quality remains an integral aspect of their growth and success.

5. Conclusion

Through meticulous exploration and analysis, several crucial conclusions can be drawn. Enhancing data quality isn't merely a process; it's a strategic imperative. The systematic identification of data issues, comprehensive cleansing, integration, and transformation processes are foundational. These strategies, powered by advanced analytics, form the bedrock of reliable, high-quality data. Improved data quality empowers organizations to make informed decisions. Decision-makers rely on accurate, consistent data to chart the course for their enterprises. Through effective data analytics, businesses can now trust the data underpinning their critical choices, leading to more reliable strategies and outcomes.

Operational efficiency flourishes in an environment where data is accurate and readily accessible. Reduction in errors and streamlined processes lead to significant cost savings. Furthermore, customers benefit from enhanced services, facilitated by accurate, timely data, resulting in heightened satisfaction and loyalty. Based on Bhardwaj (2023) Data quality enhancement is not a one-time task but a continuous process. Regular audits, feedback mechanisms, and proactive anomaly detection are essential. Additionally, organizations must be adaptable, embracing emerging technologies and methodologies to address evolving data quality challenges. A culture of collaboration between IT, data analysts, and business units is vital. Sharing knowledge and expertise ensures that the entire organization is aligned in its pursuit of data quality excellence. Documentation of best practices and processes facilitates seamless knowledge transfer, enabling future initiatives to build on past successes. Based on Ranigineni (2023) data quality enhancement efforts must be conducted ethically and in compliance with legal standards. Data privacy, security, and regulatory adherence are non-negotiable aspects of any data analytics endeavor. Organizations must balance data utility with privacy protection to maintain trust with stakeholders. With enhanced data quality, organizations experience a paradigm shift. Decisions are not just data-driven; they are data-trusted. This shift catalyzes innovation, bolsters competitiveness, and fosters a culture of excellence where data is not just a tool but a strategic asset.

In conclusion, the journey of enhancing data quality through data analytics is transformative. It reshapes how organizations perceive, manage, and leverage their data. With the right strategies, tools, and a commitment to ongoing improvement, organizations can elevate their data quality, ensuring that their data becomes a cornerstone upon which a prosperous, efficient, and innovative future is built.

In an era characterized by an overwhelming influx of data, ensuring data quality has emerged as a critical imperative for organizations seeking to derive meaningful insights, make informed decisions, and enhance operational efficiency. This

research, focused on enhancing data quality through data analytics, has illuminated the significance of systematically addressing data quality issues and the transformative potential that lies in leveraging analytics techniques. By conducting comprehensive data profiling and employing data cleansing and standardization techniques, organizations can lay a strong foundation for data quality enhancement. The process of integrating disparate data sources and transforming raw data into meaningful insights contributes significantly to the overall coherence and accuracy of data. This not only improves data quality but also provides a more holistic view of information.

Conflict of Interest

The Author's declare that there is no conflict of Interest to report.

Funding Source

This research was entirely Self-funded by the Author's.

Author's Contributions

Sandeep Rangineni, as the main author of this research paper. Amit Bhanushali, Manoj Suryadevara, Srinivas Venkata, Kiran Peddireddy has provided necessary support to every phase on this research paper as co-authors.

References

[1] Smith, J., Johnson, A., & Brown, K. (2018). Data Quality Dimensions: A Review of the State of the Art, **2018.**

[2] Wang, L., Zhang, Q., & Zhang, W. (2019). Data Cleansing and Profiling Techniques: A Comprehensive Review, **2019.**

[3] Chen, S., Liu, Y., & Wang, H. (2020). Enhancing Data Quality through Predictive Analytics: A Machine Learning Approach, **2020.**

[4] Li, X., Zhou, H., & Ma, L. (2021). Real-time Data Quality Management in the Era of Big Data: Challenges and Solutions, **2021.**

[5] Gupta, R., & Aggarwal, N. (2017). Ethical Considerations and Challenges in Data Quality Enhancement through Analytics, **2017.**

[6] Dr.Naveen Prasadula (2023) "Review of Literature on A Review On Enhancing Data Quality For Optimal Data Analytics Performance", **2023.**

[7] Ajah, I.A. and Nweke, H.F. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications, Big Data and Cognitive Computing, Vol.**3**, Issue.**32**, pp.**1-30, 2019.**

[8] Ali, A., Qadir, J., Rasool, R., Sathiaseelan, A., Zwitter, A., and Crowcroft, J. (2016). Big Data for Development: Applications and Techniques, Big Data Analytics, Vol.**1**, Issue.**2**, pp.**1-24, 2016.**

[9] A.B. (2016). Gamifying Recruitment, Selection, Training, and Performance Management: Game- thinking in Human Resource Management, In Gangadharbatla, H., and Davis, D.Z. (Eds), Emerging Research and Trends in Gamification, 140-165, Hershey: IGI Global, **2016.**

[10] Bartlett, C.A. and Ghoshal, S. (2011). Building Competitive Advantage through People, MIT Sloan Management Review, Vol.**84**, Issue.**2**, pp.**34-45, 2011.**

[11] Beringer, C., Jonas, D. and Kock, A. (2013). Behaviour of Internal Stakeholders in Project Portfolio Management and its Impact on Success, International Journal of Project Management, Vol.**31**, Issue.**6**, pp.**830-846, 2013.**

[12] Cristobal, J.R.S. (2017). Complexity in Project Management, CENTERIS International Conference on Project Management, Barcelona, Spain, pp.**8-10**, November **2017.**

[13] Demirkan, H. and Dal, B. (2014). Big Data, Big Opportunities, Big Decisions, Harvard Business Review Turkish Edition, March, pp.**28-30, 2014.**

[14] Dr.Naveen Prasadula (2023) "Big data applications in operations/supply-chain management: A literature review. Computers and Industrial Engineering", **2023.**

[15] Agar, M. (1980). Getting better quality stuff: Methodological competition in an interdisciplinary niche. Urban Life, Vol.**9**, Issue.**1**, pp.**34-50, 1980.**

[16] Ahmadov, Y., & Helo, P. (2018). A cloud based job sequencing with sequence-dependent setup for sheet metal manufacturing. Annals of Operations Research, 270(1-2), pp.**5-24, 2018.**

[17] Yunusa-kaltungo, A.; Sinha, J.K. Effective vibration-based condition monitoring (eVCM) of rotating machines.J. Qual. Maint. Eng., 23, pp.**279–296, 2017.**

[18] Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? International Journal of Production Economics, 182, pp.**113-131, 2016.**

[19] Albey, E., Norouzi, A., Kempf, K. G., & Uzsoy, R. (2015). Demand modeling with forecast evolution: an application to production planning. IEEE Transactions on Semiconductor Manufacturing, Vol.**28**, Issue.**3**, pp.**374-384, 2015.**

[20] Alexander, L. W. and 2Cheryl A. (2015). Big Data Driven Supply Chain Management : A Game Changer. American Journal of Economics and Business Administration, February, pp.**1–9, 2015.** https://doi.org/10.3844/ajebasp.2015.60.67

[21] Dr.Naveen Prasadula (2022). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges, and implications for practice. Transportation Research Part E: Logistics and Transportation Review, 114, pp.**416-436, 2022.**

[22] Arya, V., Sharma, P., Singh, A. and De Silva, P.T.M. (2017), "An exploratory study on supply chain analytics applied to spare parts supply chain", Benchmarking, Vol. 24, No.**6**, pp.**1571-1580, 2017.**

[23] Ayyildiz, E., & Gumus, A. T. (2021). Interval-valued Pythagorean fuzzy AHP method-based supply chain performance evaluation by a new extension of SCOR model: SCOR 4.0. Complex & Intelligent Systems, 7(1), pp.**559-576, 2021.**

[24] Azadian, F., Murat, A., & Chinnam, R. B. (2015). Integrated production and logistics planning: Contract manufacturing and choice of air/surface transportation. European Journal of Operational Research, 247(1), pp.**113-123, 2015.**

[25] S. Rangineni and D. Marupaka, "Data Mining Techniques Appropriate for the Evaluation of Procedure Information," *International Journal of Management, IT & Engineering*, Vol.**13**, No.**9**, pp.**12–25**, Sep. **2023.**

[26] S. Rangineni, "An Analysis of Data Quality Requirements for Machine Learning Development Pipelines Frameworks," *International Journal of Computer Trends and Technology*, Vol.**71**, No.**9**, pp.**16–27, 2023.**

[27] Arvind Kumar Bhardwaj, Sandeep Rangineni, Divya Marupaka, "Assessment of Technical Information Quality using Machine Learning ," International Journal of Computer Trends and Technology, Vol.**71**, No.**9**, pp.**33-40, 2023.**

AUTHORS PROFILE

Sandeep Rangineni is a Data Test Engineer at Pluto TV, with over 12 plus years of experience in the IT industry, primarily within the streaming media industry. He holds a Master's degree in Engineering Management and Master's degree in Information Technology. Sandeep has a diverse skill set, working with technologies such as PL/SQL, Azure Databricks, Salesforce, Informatica, and Snowflake. Currently, he is

actively engaged in researching Data Engineering and Data Quality topics. Sandeep has professional certifications in Salesforce admin, AWS Data Analytics and Safe 5 practitioner. Sandeep is a senior member of IEEE, professional member of BCS and fellow of IETE, three esteemed technology organizations, and has served as a judge for reputable award organizations in Technology which including Globee Awards, Stevie Awards, NCWIT Aspirations, and Brandon Hall Group.

Amit Bhanushali is a highly accomplished software quality assurance professional with over 22 years of experience in the IT industry. He earned his Master's in Business Data Analytics from West Virginia University in 2017. Based in West Virginia, USA, Mr. Bhanushali is a Senior IEEE Member and has significantly contributed to software testing research and practice.

His expertise spans automation testing, performance testing, DevOps, and CI/CD implementation. He has also led testing efforts in complex cloud environments. In addition to testing, Mr. Bhanushali has authored several articles exploring cutting-edge topics like artificial intelligence and machine learning. His published research demonstrates his thought leadership and impact on software quality engineering.

Mr. Bhanushali's accomplishments have been recognized through prestigious appointments. He serves as a reviewer for the Elsevier journal and has been a hackathon judge. His contributions were further honored in 2023 when he received the International Achievers' Award. With his sustained record of excellence across software development, testing, and research, Mr. Bhanushali continues to be an influential leader in his field.

Manoj Suryadevara earned his B.Tech. in Information Technology from Karunya University in 2011 and his Master of Science in Software Engineering from the University of Houston in 2014. He has over ten years of experience in product management, working with various tech companies to bring innovative solutions to market. Since 2020, he has been with Walmart, a leading retail firm in the US, where he has progressed through roles of increasing responsibility. As Staff Product Manager, he leads cross-functional teams to develop and implement product strategies, ensuring alignment with customer needs and market trends. He is well-versed in agile methodologies, user experience design, and data analytics. He has published two papers on data and is a senior member of the IEEE. His expertise includes digital transformation, data management, data science, cloud computing, and machine learning. He is passionate about leveraging technology to solve complex business problems and deliver customer value. Under his leadership, his teams have successfully launched numerous products, resulting in significant revenue growth and market share expansion.

Srinivas Venkata completed his Master of Science in Engineering Management in the United States. Currently, he holds the position of Staff Data Engineer at Teradata Inc. in Texas, a role he has been dedicated to since 2022. Furthermore, he boasts the distinction of being a Fellow member of the IETE. His academic and professional achievements shine through his publications in esteemed international journals, notably Springer. His primary research areas encompass Data Engineering, Artificial Intelligence, Big Data Analytics, Data Mining, and Business Intelligence in the context of education. With an impressive 11 years of experience in the field of Information Technology, he brings a wealth of knowledge and expertise to his role.

Kiran Peddireddy is a data engineer technical and strategic leader with over 16 years of experience enabling companies with digital transformation and helping them become data-driven organizations. He is a highly regarded thought leader with a strong background in data engineering, data analytics, cloud migration, and automation in aero and auto domain. He is currently leading data engineering modernization efforts at Cox Automotive, with teams of data engineers and analysts to upgrade the existing tool stack and decommission legacy data stream systems, enrich data streams, and build modern data applications.