International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

PYTHON DRIVEN DATA STORYTELLER TURNING COMPLEX DATA INTO CLEAR INSIGHTS

[1] J Priyanka, [2] Balla Charrishma, [3] G Sreeja, [4]K Rajkumar, [5] B Santosh
[1]AssistantProfessor, [2345]Students
Department of CSE(Data Science)
Siddhartha Institute of Technology & Sciences, Narapally

Jalagapriyanka.cse@siddhartha.co.in, 24TQ1A6705@siddhartha.co.in, 24TQ1A6715@siddhartha.co.in , 24TQ1A6755@siddhartha.co.in, 24TQ1A6707@siddhartha.co.in

Abstract

The Python Driven Data Storyteller: Turning Complex Data into Clear Insights web application presents an intelligent and user-friendly platform designed to simplify data analysis and enhance decision-making. The system enables users to upload raw datasets into a secure environment, where automated preprocessing techniques clean, validate, and restore data integrity, ensuring reliable analysis. At its core, the application leverages Python-based analytical capabilities to dynamically generate appropriate visualizations such as scatter plots, histograms, and bar charts based on the nature and structure of the data. This automated visualization logic eliminates the need for manual intervention, making the system accessible even to non-technical users. Additionally, the interactive dashboard provides a dedicated Key Metrics section, which computes and displays essential statistical measures, offering meaningful context to the analyzed data.

The integration of a Streamlit-powered interface ensures smooth interaction, real-time updates, and an intuitive user experience. The application also supports automated PDF report generation, enabling users to export insights in a structured and professional format. Furthermore, it incorporates trend analysis and storytelling guidance, helping users interpret patterns and communicate findings effectively. Overall, this system transforms complex datasets into clear, actionable insights by combining automation, visualization, and storytelling techniques. It enhances analytical efficiency, improves clarity, and supports data-driven decision-making across various domains.

I. Introduction

The Python Driven Data Storyteller: Turning Complex Data into Clear Insights web application is an advanced analytical platform designed to transform raw and complex datasets into meaningful, easy-to-understand insights. In today's data-driven world, organizations and individuals often struggle to interpret large volumes of unstructured or inconsistent data. This application addresses that challenge by providing a structured and intelligent workflow that simplifies data analysis while enhancing the quality of insights generated.

The system begins with a comprehensive Data Intelligence Analysis, which evaluates key aspects of the dataset, including distribution skewness, missing values, outliers, correlation strength, and an overall Data Health Score. These analytical components are visually represented through interactive statistical bars, allowing users to quickly

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

assess the condition and reliability of their data. This initial step ensures that users gain a strong foundational understanding before proceeding to deeper analysis.

Based on this evaluation, the application generates a Personalized Analytical Workflow tailored to the specific characteristics and issues within the dataset. The process is divided into two main phases. The Infiltration Phase (Inference) focuses on automated data cleaning, preprocessing, and normalization, ensuring data consistency and readiness for analysis. Following this, the Extraction Phase (Processing) applies statistical techniques and dynamic visualization logic to automatically produce the most appropriate graphical representations, such as scatter plots, histograms, and bar charts, uncovering hidden patterns and relationships.

The application is built with an intuitive interface powered by Streamlit, enabling seamless navigation and real-time interaction. It also features automated PDF reporting and Key Metric Recommendations that are customized based on the dataset's structure and type. By continuously tracking trends and offering optimization suggestions for better visualization and storytelling, the system ensures that users can communicate insights effectively.

II. Literature Survey

The field of data analytics and storytelling has evolved significantly with the integration of machine learning, artificial intelligence, and intelligent visualization systems. Various research works have focused on improving data quality, automating visualization selection, and enhancing reporting systems to make data interpretation more effective and accessible.

The study titled "Automated Data Quality Assessment Using Machine Learning" highlights the critical issue of poor data quality in analytical processes. It emphasizes that datasets often contain noise, missing values, and outliers, which can lead to incorrect conclusions and flawed business decisions. The paper proposes an automated data quality assessment system that uses machine learning techniques for distribution classification and outlier detection. This system not only improves data reliability but also recommends suitable analytical routines. Experimental results indicate high accuracy in automated data type classification, demonstrating the effectiveness of machine learning in improving data preprocessing and analysis.

Another important contribution is presented in "A Study on Dynamic Visualization Selection Using Heuristic Processing." This work addresses the challenge of selecting appropriate visualizations in a rapidly evolving data science environment. With numerous visualization tools and frameworks available, users often face difficulty in choosing the right chart type. The proposed system introduces a heuristic-based dynamic visualization selection mechanism, where users upload datasets and specify basic analytical goals. The system then predicts the most suitable visualization techniques, such as scatter plots, histograms, or bar charts, thereby simplifying the decision-making process and improving analytical clarity.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The paper "Business Intelligence Suggestion Systems Using Deep Learning" focuses on the role of deep learning in enhancing business intelligence and data storytelling. It explains that selecting the right statistical models and visualization techniques is crucial for generating meaningful insights. Traditional approaches often fall short when handling large and complex datasets. The proposed deep learning-based system automates model selection and recommendation, enabling users to choose optimal analytical methods based on data characteristics. This approach improves report quality and ensures more accurate interpretation of data, especially in complex IT and business environments.

In "Artificial Intelligence Based Smart Reporting Systems Based on Data Condition," the authors discuss the growing importance of automated reporting systems in modern analytics. The paper highlights how AI can be used to streamline complex analytical processes and generate insightful reports. The proposed system uses AI algorithms to analyze large volumes of unstructured data and automatically produce structured outputs, including PDF reports. This reduces manual effort and enhances the efficiency of data interpretation. The study also emphasizes the importance of selecting appropriate metrics and processing methods based on data conditions to improve overall analytical performance.

Finally, the study "DataProf: A Web Application for Smart Data and Analytics Users" explores the development of a user-friendly web-based platform for data analysis. It addresses challenges related to the complexity of statistical methods and the steep learning curve associated with modern analytical tools. The proposed application enables users to evaluate data health, detect outliers, and compare different analytical strategies. By simplifying data exploration and visualization, the system encourages wider adoption of data analytics tools and supports more informed decision-making.

In summary, these research works collectively highlight the importance of automation, intelligent visualization, and AI-driven analytics in modern data storytelling. They demonstrate how advanced technologies can address challenges related to data quality, visualization selection, and reporting, thereby improving the effectiveness and accessibility of data-driven insights.

III. System Analysis

The Python Driven Data Storyteller system is designed to transform complex datasets into clear and meaningful insights through automation and intelligent processing. It focuses on simplifying data analysis by integrating data cleaning, statistical evaluation, and visualization into a unified workflow. The system analyzes datasets to detect missing values, outliers, and inconsistencies, ensuring data quality before further processing. It evaluates important statistical properties such as distribution, correlation, and skewness to understand the nature of the data. A Data Health Score is generated to provide an overall assessment of dataset reliability. The system supports automated preprocessing techniques, reducing manual intervention and improving efficiency. It dynamically selects suitable visualization techniques based on data types and analytical needs. Interactive dashboards present key metrics and insights in a user-friendly manner. The system is designed for both technical and non-technical

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

users, making it highly accessible. It also integrates automated reporting features to generate structured outputs. The analytical process is divided into inference and processing phases for better organization. Overall, the system enhances clarity, efficiency, and decision-making in data analysis.

Existing System

The existing systems for data analysis largely depend on manual processes and user expertise. Users are required to clean and preprocess data manually, which can be time-consuming and error-prone. Visualization selection is often based on user knowledge rather than automated suggestions, leading to inappropriate chart choices. Many tools lack integrated data quality assessment features, making it difficult to evaluate dataset reliability. Handling missing values and outliers typically requires additional tools or programming skills. Reporting is often done manually, requiring extra effort to present findings. Existing systems provide limited support for guiding users through the analytical process. Non-technical users face challenges in understanding and using complex tools. There is minimal automation in generating insights or recommendations. The absence of personalized workflows makes analysis less efficient. Users must interpret raw outputs without proper storytelling assistance. As a result, insights generated may be inconsistent, unclear, and less impactful.

Disadvantages of Existing System

- Requires strong technical knowledge and expertise
- Involves time-consuming manual data preprocessing
- Lacks automated data quality assessment mechanisms
- Difficulty in selecting appropriate visualization techniques
- No personalized workflow for different datasets
- High chances of human errors during analysis
- Limited support for non-technical users
- Manual and inefficient reporting process
- Does not provide real-time insights or dashboards
- Poor integration of data storytelling features

Proposed System

The proposed system introduces an intelligent and automated approach to data storytelling using Python technologies. It begins with a Data Intelligence Analysis phase that evaluates dataset quality by detecting missing values, outliers, and correlations. The system generates a Data Health Score to provide an overall understanding of the dataset condition. It offers a Personalized Analytical Workflow tailored to the specific needs of the dataset. The Infiltration Phase focuses on automated data cleaning and normalization to prepare the data for analysis. The Extraction Phase applies statistical techniques and dynamic visualization logic to generate appropriate charts. The system automatically selects suitable visualizations such as scatter plots, histograms, and bar charts. An interactive dashboard displays key metrics and insights in a clear and structured format. The Streamlit-based interface ensures smooth and user-friendly interaction. Automated PDF reporting

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

allows users to export results easily. The system continuously tracks trends and provides optimization tips. Overall, it improves data clarity, efficiency, and decision-making.

Advantages of Proposed System

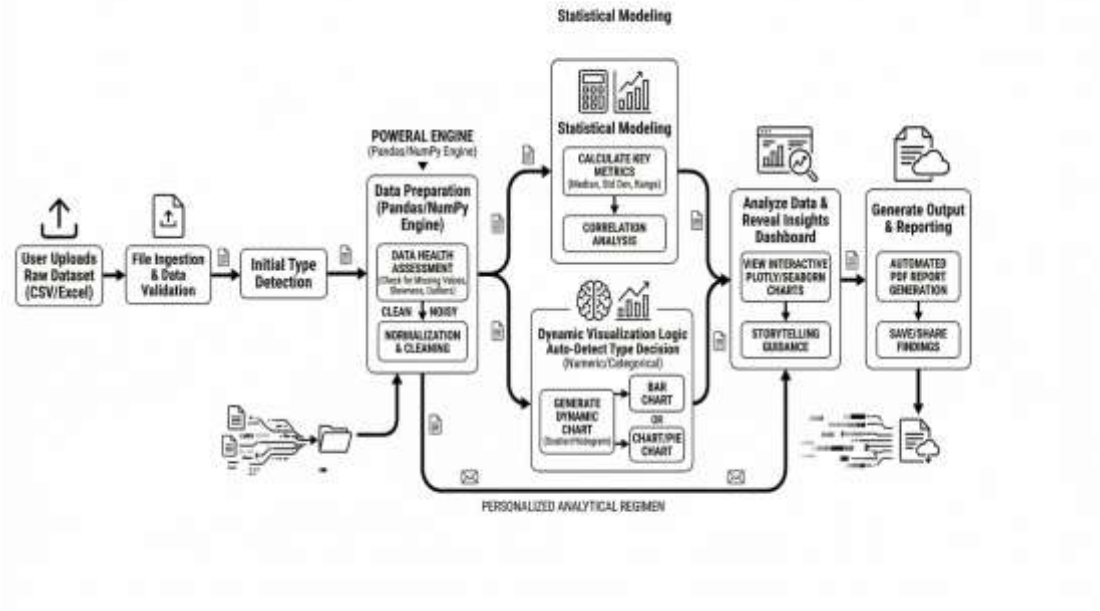
- Automates data cleaning and preprocessing, reducing manual effort
- Provides intelligent data quality assessment (missing values, outliers, correlations)
- Generates a Data Health Score for better understanding of dataset condition
- Dynamically selects the most suitable visualization techniques
- User-friendly interface accessible to both technical and non-technical users
- Offers personalized analytical workflows based on dataset characteristics
- Displays real-time insights through interactive dashboards
- Reduces human errors in data analysis
- Supports automated PDF report generation for easy sharing

IV. Methodology

The system follows a structured methodology to transform raw data into meaningful insights. Initially, users upload their datasets into the application. The preprocessing stage handles missing values, noise, and inconsistencies in the data. Outlier detection techniques are applied to improve data quality. The system then analyzes statistical properties such as distribution, correlation, and skewness. A Data Health Score is calculated to evaluate dataset condition. Based on this analysis, a personalized workflow is generated. In the inference phase, data is cleaned and normalized for consistency. In the processing phase, dynamic visualization logic is applied to generate suitable charts. The system automatically selects appropriate visualizations such as scatter plots and histograms. Key metrics are calculated and displayed on the dashboard. Finally, results are compiled into a structured PDF report for easy sharing and interpretation.

System Architecture

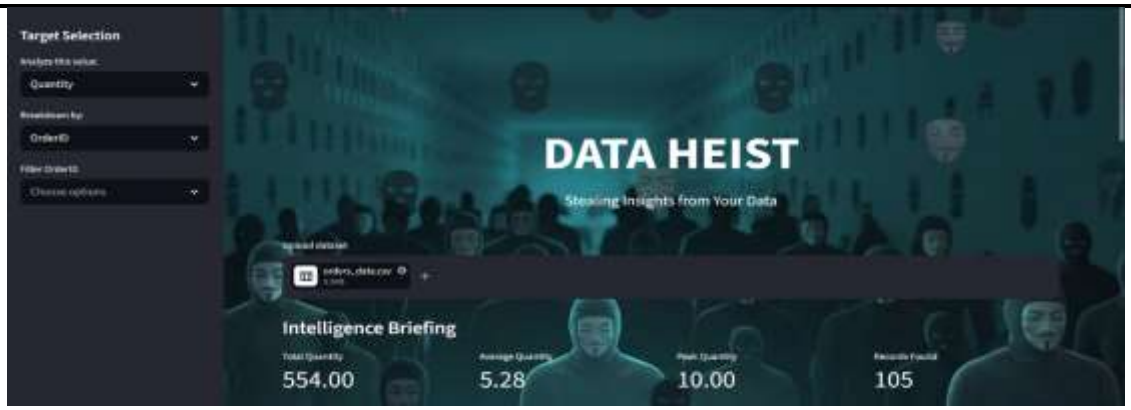
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

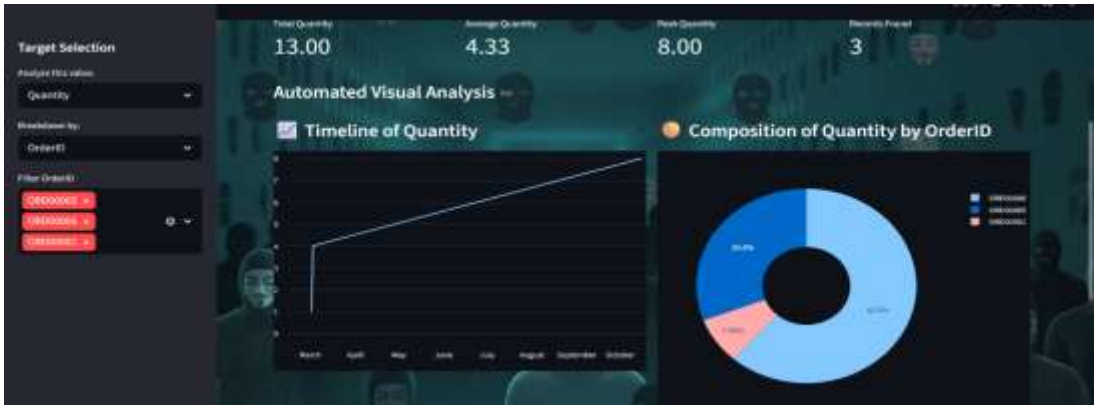


The system architecture consists of multiple interconnected components that work together to deliver efficient data analysis. The User Interface is developed using Streamlit, providing an interactive environment for users. The Data Input Module allows users to upload datasets securely. The Preprocessing Module handles data cleaning, normalization, and transformation. The Data Analysis Engine evaluates statistical properties and identifies patterns in the data. The Visualization Engine dynamically selects and generates appropriate charts. The Recommendation Module provides key insights and analytical suggestions. The Workflow Manager controls the inference and processing phases of analysis. The Reporting Module generates automated PDF reports. A storage layer is used to manage datasets and results securely. The backend is powered by Python libraries for data analysis and machine learning. All components are integrated to ensure smooth, efficient, and intelligent data storytelling.

V. Result and Output

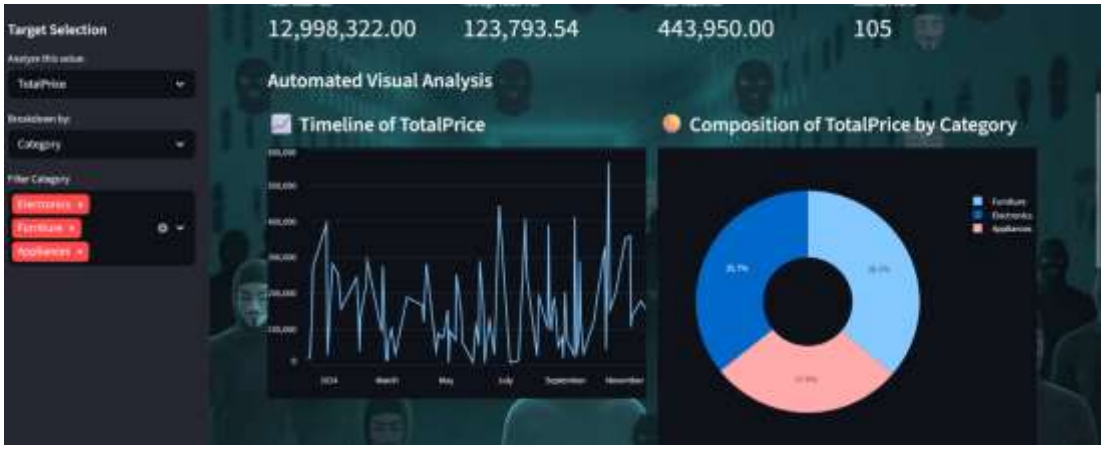
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper





Received: 20-04-2026 | Accepted: 28-05-2026 | Published: 05-06-2026

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

VI. Conclusion

The Data Heist platform marks a revolutionary advancement in personalized data intelligence technology. By utilizing advanced Python-driven logic and statistical analysis, the platform provides precise and tailored analytical advice, addressing various data concerns such as skewness, missing values, outliers, noise, and hidden correlations. This system processes raw datasets to deliver accurate "Data Health" assessments, ensuring users receive highly personalized and effective visualization recommendations.

The platform excels in offering customized analytical workflows, factoring in individual data distributions, variable types, and specific business intelligence issues. Its user-friendly interface features visual aids like interactive statistical bars, key metric dashboards, and health scores, making it easy for users to understand their data conditions. The design ensures accessibility across devices through its web-based deployment, allowing users to access insights and track progress conveniently.

Data privacy is a critical aspect, with robust security measures such as in-memory processing and secure session handling in place, ensuring that sensitive information is managed with integrity. Integration with automated PDF reporting adds professional

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

validation to the logic-driven recommendations, blending high-fidelity technology with actionable strategic insights.

References

1. Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
2. Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
3. Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
4. Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
5. V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
6. B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
7. R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
8. Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
9. Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
10. Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
11. Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

12. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
13. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
14. Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.
15. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
16. Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.
17. Mudusu, S. K. (2026, March 26). A data trust scoring framework for reliable and responsible AI systems. InfoWorld (Foundry Expert Contributor Network).
18. Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.
19. Mudusu, S. K. (2025, April 20). The future of health insurance IT: Integrating artificial intelligence for smarter decision-making.
20. Gajula, S. (2025). Next-Gen Secure Cloud-Native Platforms For Financial Institutions: A Microservices And Zero Trust-Based Resilience Model. Journal of International Crisis & Risk Communication Research (JICRCR), 8.
21. Maturi, S. Y. (2024). Cryptographic privacy engines: Practical multi-party protocols for confidential database queries. Nanotechnology Perceptions, 20(S13), 2770–2785
22. Ranjbareslamloo, S., Dzukey, G. A., Islam Muhit, M. M., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.06.108
23. Manoharan, D. (2026). Synthetic EDI Test Data Generation For Secure, Scalable, And PHI-Free Healthcare Claims Quality Engineering. Journal of International Crisis and Risk Communication Research, 9(1).
24. Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. International Journal of Research in Information Technology and Computing, 8(4).
25. Pavan Kumar Adabala. (2026). IoT-Driven Digital Twins for Manufacturing Optimization: Hybrid Modelling, Reinforcement Learning and Sustainable Operations. International Journal of Computational and Experimental Science and Engineering, 12(1). https://doi.org/10.22399/ijcesen.5050
26. Kavuri, S. (Ed.). (2024). Shift-left and shift-right testing approaches: A practical roadmap for continuous quality in agile and DevOps. Journal of Information Systems Engineering and Management, 9(4). https://doi.org/10.52783/jisem.v9i4.127
27. Srikanth Kavuri. (2023). Machine Learning Approaches for Security Vulnerability Detection in Software Testing. Computer Fraud and Security. https://doi.org/10.52710/cfs.837

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

28. Venkata Pavan Kumar Gummadi. (2025). MuleSoft Architectural Paradigms and Sustainability: A Comprehensive Technical Analysis. Journal of Computer Science and Technology Studies, 7(12), 534–540. https://doi.org/10.32996/jcsts.2025.7.12.59
29. Gummadi, V. P. K. (Ed.). (2025). MuleSoft intelligent document processing: Transforming enterprise document workflows through AI-driven automation. Journal of Computational Analysis and Applications, 34(12). https://doi.org/10.48047/jocaaa.2025.34.12.16
30. Pokala, H. K., & Gummadi, V. P. K. (2026). Autonomous AI-Powered Resource Management for Apache Flink on Amazon EKS. 2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET), 1–4. https://doi.org/10.1109/icaiset66439.2026.11541881