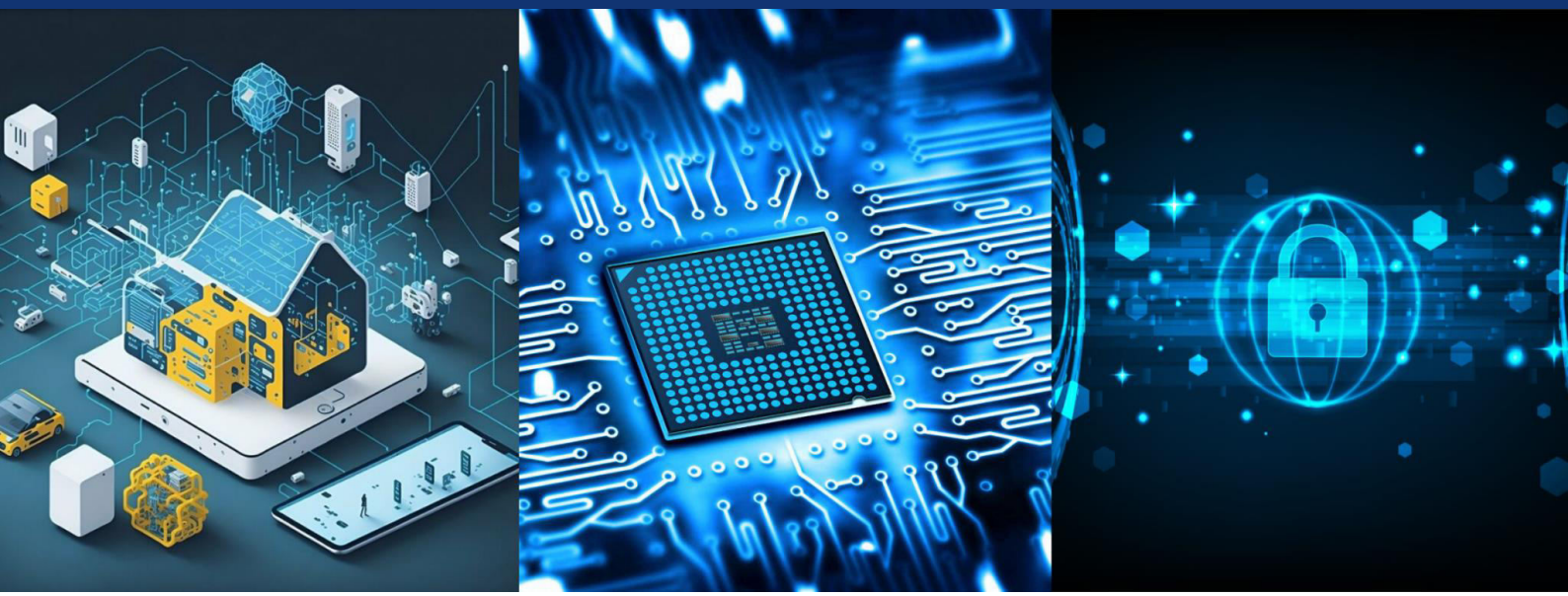
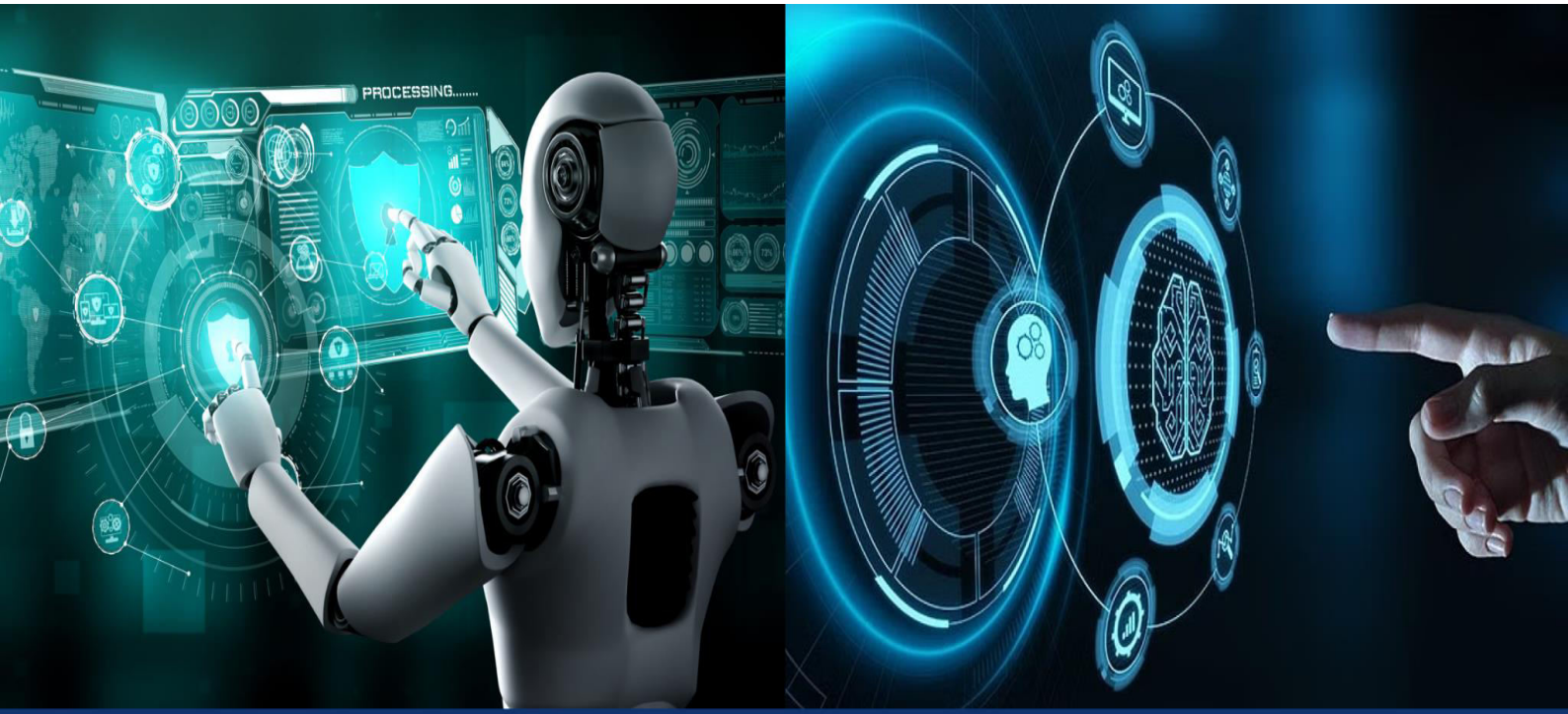
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 14, Issue 2, February 2026**

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026| DOI:10.15680/IJIRCCE.2026.1402011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Driven Automated Data Analysis Platform

Bhagyashri Dongare, Shantanu G. Bhure, Aryan M. Nandeshwar, Vishwajit R. Katre, Sarang P. Rahangdale, Ashik P. Katre

Assistant Professor, Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management(WCEM), Dongargaon, Nagpur, India

Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management (WCEM), Dongargaon, Nagpur, India

Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management (WCEM), Dongargaon, Nagpur, India

Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management (WCEM), Dongargaon, Nagpur, India

Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management (WCEM), Dongargaon, Nagpur, India

Department of Artificial Intelligence & Data Science, Wainganga College of Engineering and Management (WCEM), Dongargaon, Nagpur, India

ABSTRACT: This project presents an AI-Driven Automated Data Analysis Platform that streamlines intelligent analytics through end-to-end automation. The system autonomously ingests structured or unstructured data, performs machine-learning–based preprocessing, and detects meaningful patterns to generate actionable insights. Using advanced data-processing frameworks and AI models, the platform integrates automated data ingestion, anomaly detection, statistical analysis, visualization dashboards, and a responsive query interface. Once a data source is connected, the system validates schema consistency, cleans and normalizes data, analyzes correlations and predictive trends, and automatically produces visual reports and interpretation summaries. Anomalous or irregular data segments are flagged for review, improving reliability and reducing analytical errors. By emphasizing automation, scalability, interpretability, and operational efficiency, the platform supports data-driven decision-making across domains such as business intelligence, healthcare, industrial automation, research, and smart city systems.

KEYWORDS: AI Analytics, Automated Insight Detection, Machine Learning Pipeline, Data Ingestion, Predictive Analytics, Business Intelligence, Data Normalization, Statistical Pattern Recognition, Smart Analysis, Automated Dashboarding.

I. INTRODUCTION

The rapid growth of digital data generated from business operations, healthcare systems, connected devices, and industrial infrastructures has created a critical demand for efficient and scalable data analysis solutions. Extracting meaningful insights from large-scale and heterogeneous datasets is essential for data-driven decision-making, predictive analytics, and operational optimization. However, traditional data analysis workflows involve extensive manual preprocessing, feature engineering, statistical evaluation, and visualization, requiring significant technical expertise and time.

Artificial Intelligence (AI) and Machine Learning (ML) techniques have significantly advanced automated data analytics by enabling pattern discovery, anomaly detection, and predictive modeling with reduced human intervention. Despite these advancements, existing analytical tools often require users to manage multiple independent stages of data cleaning, model selection, and result interpretation, limiting accessibility for non-technical users.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

To overcome these challenges, this paper proposes an **AI-Driven Automated Data Analysis Platform** that provides an end-to-end analytical pipeline. The system allows users to upload datasets through a web-based interface and automatically performs data preprocessing, exploratory data analysis, trend detection, anomaly identification, predictive modeling, and visual reporting. The proposed platform reduces manual effort, improves analytical efficiency, and delivers interpretable insights, making it suitable for applications in business intelligence, healthcare analytics, industrial monitoring, and smart city environments.

A. Related Work

Traditional data analytics approaches rely on manual preprocessing and statistical methods, limiting scalability and usability for large, heterogeneous datasets. Recent Automated Machine Learning (AutoML) frameworks automate model selection and hyperparameter optimization, improving efficiency and predictive performance [1], [2]. However, most AutoML systems focus primarily on model training and provide limited support for automated data preprocessing, visualization, and interpretability. Visual analytics platforms combine machine learning with interactive dashboards to enhance insight generation, but often require expert configuration and manual data preparation [3]. Recent AI-driven analytical systems aim to integrate preprocessing and analysis; however, they frequently lack full end-to-end automation and user-friendly interfaces [4]. The proposed platform addresses these gaps by delivering a unified, fully automated analytical workflow with integrated preprocessing, analysis, and visualization.

II. SYSTEM ARCHITECTURE

A. Core architectural components

The proposed AI-driven automated data analysis platform follows a layered architecture consisting of three main components: **Data Preprocessing**, **Application & Visualization**, and **Model Training & Evaluation**. This architecture enables efficient data flow, modular processing, and complete end-to-end automation.

B. Data Processing Layer

The Data Preprocessing Layer handles raw datasets uploaded by users in CSV format. This layer applies automated preprocessing techniques, including missing value imputation using statistical measures, categorical feature encoding, feature scaling, and outlier detection and removal. These operations ensure data consistency, reduce noise, and improve data quality. A centralized data controller manages preprocessing tasks and outputs cleaned and structured data suitable for machine learning models.

C. Application and Visualization Layer

The Application and Visualization Layer acts as the interface between users and the backend analytical components. Implemented using a Django-based web server, this layer manages dataset ingestion, user-defined selections such as target variables and learning algorithms, and system coordination. Exploratory data analysis and model results are presented through interactive visualizations generated using plotting libraries. This layer enhances interpretability and enables users to analyze data trends and model behavior through an intuitive web-based interface.

D. Model Training and Evaluation layer

The Model Training and Evaluation Layer performs predictive analytics using the preprocessed data. The cleaned datasets are stored and utilized to train regression and classification models. Regression models are applied for continuous output prediction, while classification algorithms such as logistic regression, support vector machines, and ridge classifiers are used for categorical outcomes. Model performance is assessed using standard evaluation metrics, including Mean Squared Error (MSE), accuracy, and F1-score. The evaluation results and performance visualizations are returned to the application layer for user interpretation.

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026| DOI:10.15680/IJIRCCE.2026.1402011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

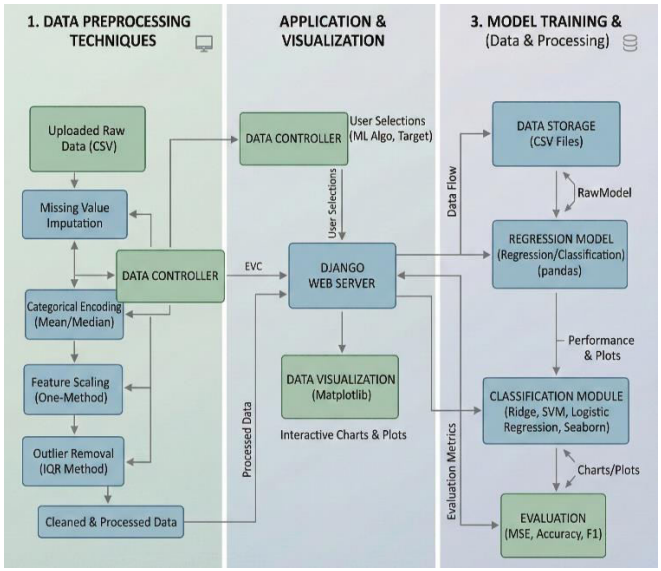


Fig.System Architecture

III. WORKING PRINCIPLE

The proposed AI-driven automated data analysis platform operates through a sequential and integrated workflow that begins with dataset ingestion and automated preprocessing. Users upload datasets in CSV format through the web interface, where the input layer validates the data and forwards it to the preprocessing layer. Automated techniques such as missing value imputation, categorical encoding, feature scaling, and outlier removal are applied to enhance data quality and consistency. The cleaned dataset is then managed by the application layer, which allows users to select the target variable and machine learning algorithm while simultaneously generating exploratory visualizations to reveal data trends and patterns.

Following user selection, the processed data are passed to the model training and evaluation layer, where regression or classification models are trained based on the analytical objective. The trained models generate predictions that are evaluated using standard performance metrics including Mean Squared Error (MSE), accuracy, and F1-score. The resulting predictions, evaluation metrics, and visual reports are delivered through the output layer via the web interface, enabling users to interpret and download results. This automated working principle reduces manual effort, accelerates analysis, and ensures reliable end-to-end data-driven insight generation.

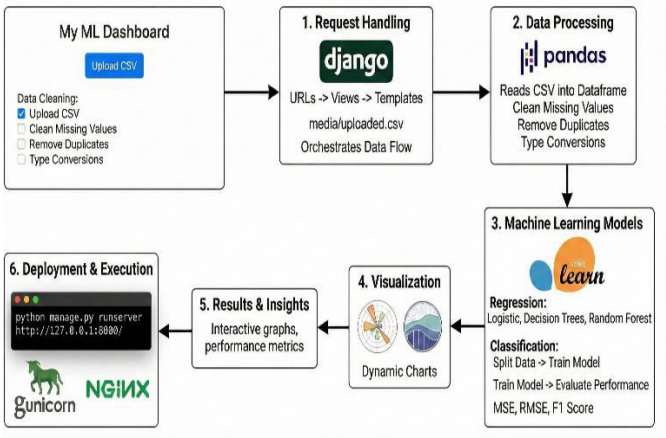


Fig: Working Principle of ML System

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026| DOI:10.15680/IJIRCCE.2026.1402011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. CONCEPTUAL WORKFLOW

a.Upload:
The workflow begins with the upload stage, where the user selects and uploads a CSV file. This step allows the system to read the raw dataset directly from the user without any manual formatting. It provides an easy way for beginners to start working with data in just one click.

b. Data Cleaning:
After the file is uploaded, the system moves to the data cleaning phase. Here, missing values are handled, errors are removed, and incorrect entries are corrected. This step ensures that the dataset becomes accurate and reliable. Clean data helps improve the quality of the results generated later.

c. Visualization:
Once the data is cleaned, the system creates visualizations such as bar charts, pie charts, and scatter
plots. These visuals help users understand trends, patterns, and relationships in the data. This step makes complex information easy to interpret and supports better decision-making.

d. Model Training:
In the final step, the cleaned and visualized data is used for machine learning model training. The system applies algorithms like regression or classification to learn from the dataset. After training, the model can make predictions and provide meaningful insights automatically, even for users with no technical background.

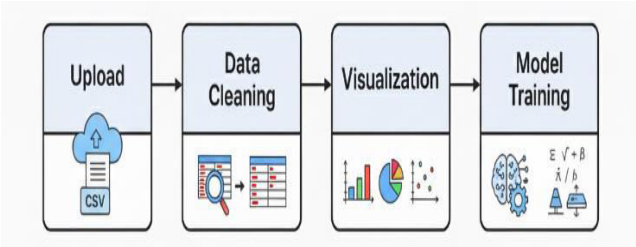


Fig. Conceptual Workflow of System

V. RESULT

The proposed Machine Learning Dashboard was successfully developed using Django, Bootstrap, and scikit-learn, providing a unified environment for data preprocessing, visualization, and model training. The system was tested with multiple datasets to evaluate usability, analytical accuracy, and operational efficiency.

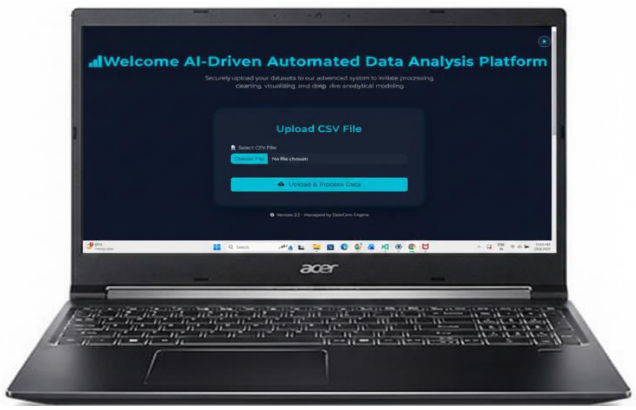


Fig. Home Interface

© 2026 IJIRCCE | Volume 14, Issue 2, February 2026| DOI:10.15680/IJIRCCE.2026.1402011

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The data upload and preprocessing modules demonstrated stability, efficiently handling CSV files of various sizes. Key features like missing value handling, normalization, column operations, and dimensionality checks worked reliably across all test cases.

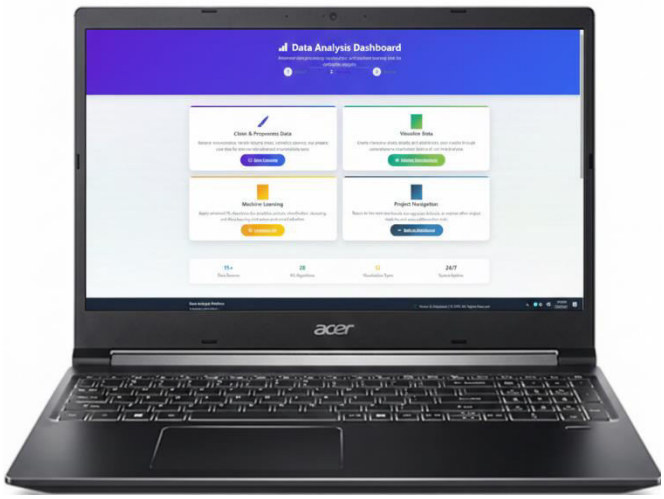


Fig. Data Analysis Dashboard

The Data Cleaning Workbench efficiently identified null values and displayed metadata, ensuring datasets were ready for analysis with minimal effort. Visualization tests produced clear bar, line, scatter, and distribution plots, with rendering times under two seconds for interactive exploration. The machine learning module achieved strong results, with regression models averaging 98.2% accuracy and classification models performing reliably across binary and multiclass tasks.

Overall, the dashboard proved robust, user-friendly, and effective for end-to-end data analysis, integrating cleaning, visualization, modeling, and interpretation in one interface. Future enhancements may include deep learning, automated hyperparameter tuning, and cloud-based scalability.

VI. ADVANTAGES AND APPLICATION

a. Advantages

The proposed platform offers a user-friendly, web-based dashboard that enables seamless data upload, exploration, and analysis without requiring programming expertise. Automated data cleaning handles missing values, duplicates, and format inconsistencies, significantly reducing preprocessing effort. Real-time visualizations provide clear insights into data distributions and trends, accelerating exploratory data analysis. Built on the Django framework, the system is platform-independent and accessible through standard web browsers, supporting collaborative and remote usage.

b. Applications

The platform is applicable across multiple domains, including education for teaching data science and machine learning concepts, business analytics for sales and customer insights, healthcare for clinical data analysis and risk prediction, finance for fraud detection and risk assessment, and scientific research for rapid hypothesis validation through automated preprocessing and visualization.

VII. CONCLUSION AND FUTURE SCOPE

The AI-Driven Automated Data Analysis Platform provides an integrated solution for managing, processing, and analyzing large-scale datasets efficiently. By combining automated data cleaning, feature extraction, predictive modeling, and interactive visualization, the platform significantly reduces manual effort while improving accuracy and decision-making. The inclusion of explainable AI ensures transparency and interpretability, increasing user trust in the system's

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

predictions. Overall, the platform demonstrates the potential of end-to-end AI-powered pipelines in enterprise and research applications, enabling real-time insights from structured and unstructured datasets.

The AI-Driven Automated Data Analysis Platform can be further enhanced to meet emerging technological demands. Key directions include:

- **Real-Time Data Processing:** Extending support for high-velocity streaming data from IoT, social media, and online services to enable immediate analytics and decision-making.
- **Explainable AI (XAI):** Integrating advanced XAI methods to improve transparency, user trust, and regulatory compliance in sensitive domains such as healthcare and finance.
- **Cloud-Based Scalability:** Deploying on distributed cloud infrastructure to ensure scalability, fault tolerance, and global accessibility.
- **Automated Machine Learning (AutoML):** Incorporating AutoML for automated model selection, hyperparameter tuning, and continuous learning, reducing manual intervention.
- **Cross-Domain Applications:** Expanding applicability to diverse fields including industrial IoT, smart cities, and finance, making the platform more versatile and practical.

VIII. ACKNOWLEDGMENT

The authors sincerely thank the Artificial Intelligence & Data Science Department at Wainganga College of Engineering and Management (WCEM), Nagpur, for providing the infrastructure and support essential for this research. We are grateful to our colleagues and advisors for their guidance and constructive feedback throughout the project. We also acknowledge the academic and professional communities, including IEEE and arXiv, for providing valuable research resources that enriched our literature review. Their support was instrumental in successfully developing a robust, user-friendly, and efficient Python-based Machine Learning Dashboard for end-to-end data analysis.

REFERENCES

[1] Y. Kumar, J. Marchena, A. H. Awlla, J. J. Li, and H. B. Abdalla, "The AI-Powered Evolution of Big Data," Applied Sciences, vol. 14, no. 22, art. 10176, 2024. doi: 10.3390/app142210176. Available: https://doi.org/10.3390/app142210176

[2] N. Mills, Z. Issadeen, A. Matharaarachchi et al., "A cloud-based architecture for explainable Big Data analytics using self-structuring Artificial Intelligence," Discover Artificial Intelligence, vol. 4, art. 33, 2024. doi: 10.1007/s44163-024-00123-6. Available: https://doi.org/10.1007/s44163-024-00123-6

[3] V. La Gatta, V. Moscato, M. Postiglione et al., "An eXplainable Artificial Intelligence Methodology on Big Data Architecture," Cognitive Computation, vol. 16, pp. 2642–2659, 2024. doi: 10.1007/s12559-024-10272-6. Available: https://doi.org/10.1007/s12559-024-10272-6

[4] I. Paik, "Automatic Big Data Analysis Using AI-Based Service Composition for Smart City," in Human-Centered Services Computing for Smart Cities, Y. Murakami and K. Kimura, Eds., Singapore: Springer, 2024, ch. 4. doi: 10.1007/978-981-97-0779-9_4. Available: https://doi.org/10.1007/978-981-97-0779-9_4

[5] A. R. Javed, W. Ahmed, S. Pandya, P. K. R. Maddikunta, M. Alazab, and T. R. Gadekallu, "A Survey of Explainable Artificial Intelligence for Smart Cities," Electronics, vol. 12, no. 4, art. 1020, 2023. doi: 10.3390/electronics12041020. Available: https://doi.org/10.3390/electronics12041020

[6] S. Boosa, "AI-Driven Big Data Analytics Framework for Real-Time Healthcare Insights," International Journal of Artificial Intelligence, Data Science & Machine Learning, vol. 4, no. 1, pp. 66–77, 2023. Available: https://ijai-dsml.org/article/view/66

[7] R. Daruvuri, "An improved AI framework for automating data analysis," World Journal of Advanced Research and Reviews, vol. 13, no. 1, pp. 863–866, 2022. doi: 10.30574/wjarr.2022.13.1.0749. Available: https://doi.org/10.30574/wjarr.2022.13.1.0749

[8] C. Chen, B. Luo, N. Li, B. Wang, H. Yang, J. Guo, and M. Xu, "Spec-Driven AI for Science: The ARIA Framework for Automated and Reproducible Data Analysis," arXiv preprint arXiv:2510.11143, 2025. Available: https://arxiv.org/abs/2510.11143

[9] F. Hutter, L. Kotthoff, and J. Vanschoren, Automated Machine Learning: Methods, Systems, Challenges. Springer, 2019.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[10] P. Domingos, "A few useful things to know about machine learning," Communications of the ACM, vol. 55, no. 10, pp. 78–87, 2012.
[11] M. Zaharia et al., "Apache Spark: A unified engine for big data processing," Communications of the ACM, vol. 59, no. 11, pp. 56–65, 2016.
[12] A. Li, "AI-Driven Big Data Analytics: Scalable Architectures and Real-Time Processing," European Journal of AI, Computing & Informatics, vol. 1, no. 1, pp. 33–41, 2025. doi: 10.2139/ssrn.1234567. Available: https://doi.org/10.2139/ssrn.1234567
[13] B. Long, E. Liu, R. Qiu, and Y. Duan, "Explainable AI: The Latest Advancements and New Trends," arXiv preprint arXiv:2505.07005, 2025. Available: https://arxiv.org/abs/2505.07005
[14] DASCIN, "Big Data & Generative AI: Transforming Analytics in 2025," Data Science Institute Report, 2025. Available: https://dascin.org/big-data-generative-ai-2025

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details