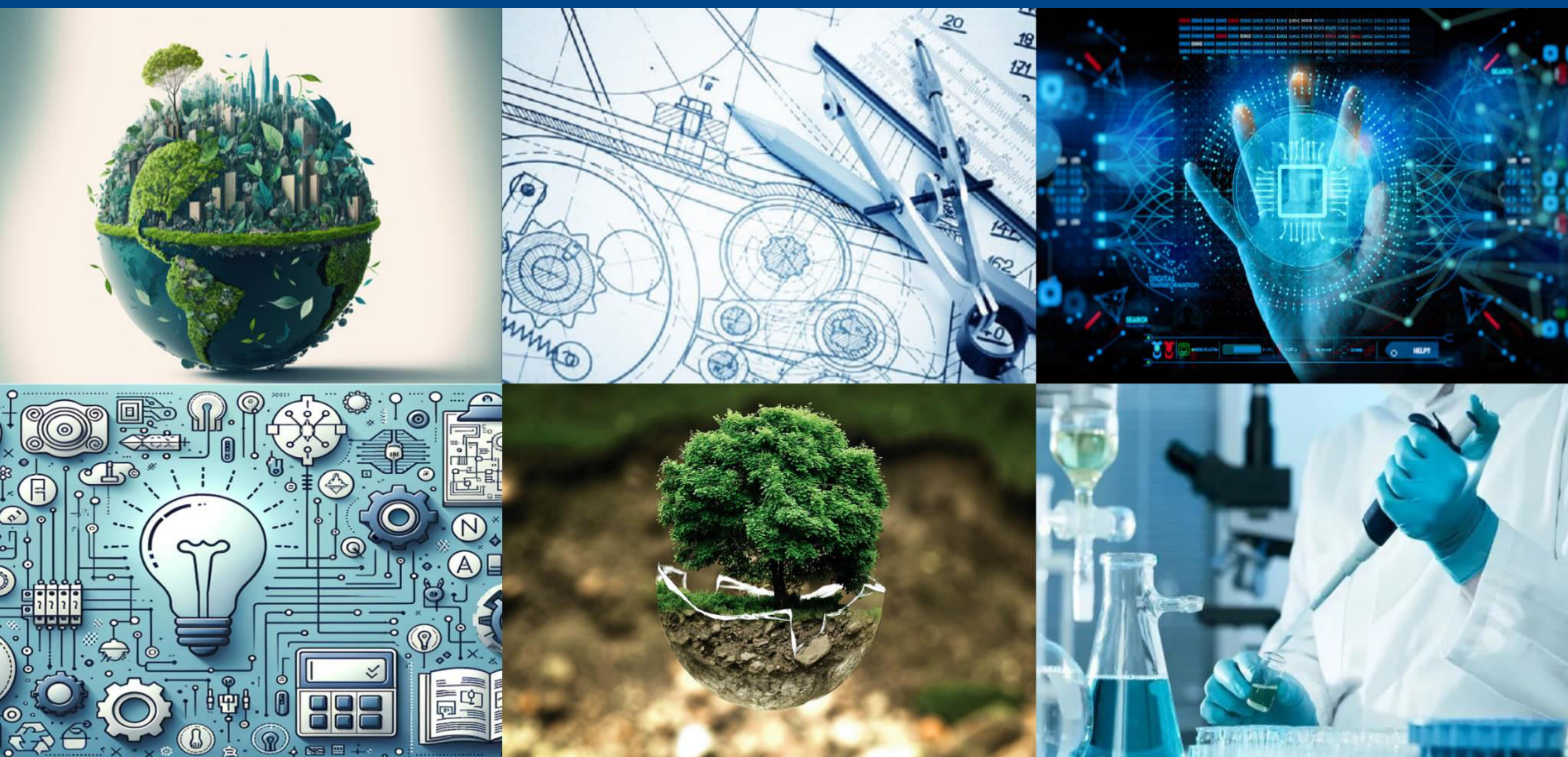
IJMRSET

ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 8, Issue 5, May 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

A No-Code Platform for End-to-End Exploratory Data Analysis and Machine Learning Model Automation for Non-Technical Stakeholders

Kunal Yadav, Garima Srivastava

Department of Computer Science & Engineering, Amity University Uttar Pradesh, Lucknow, India

ABSTRACT: In response to the growing need for accessible data analytics platforms that eliminate the dependency on manual coding, scripting, and advanced algorithmic knowledge, this paper presents the conceptualization, design, and implementation of a robust and scalable no-code analytical framework. The proposed system automates the entire data analytics lifecycle—ranging from ingestion and preprocessing of structured data to the generation of comprehensive exploratory data analysis (EDA) outputs and the construction, training, and evaluation of predictive models—all orchestrated through a dynamic graphical user interface (GUI). Leveraging the Python ecosystem, with libraries such as Pandas for data manipulation, Seaborn and Matplotlib for rich visualization generation, and Scikit-learn for machine learning tasks, the tool demonstrates a powerful integration of modular data processing capabilities, statistical visual narratives, and model performance reporting mechanisms. This solution aims to democratize data science practices across non-technical business analysts, educators, and researchers, thereby catalyzing faster, more informed decision-making. The research concludes with an evaluation of performance, usability feedback, and a roadmap for integrating deep learning, causal inference, and cloud-native scalability in future iterations.

KEYWORDS: No-code platform, automated EDA, machine learning, model evaluation, Python GUI, Pandas, Scikit-learn, data preprocessing, visualization engine

I. INTRODUCTION

In contemporary data-driven environments characterized by overwhelming volumes of heterogeneous and high-velocity datasets, the capacity to extract meaningful patterns and actionable insights has become a fundamental strategic asset for enterprises, academic institutions, and research organizations alike. However, the tools that enable such advanced data processing and machine learning workflows are often gated behind steep learning curves, requiring proficiency in statistical modeling, programming languages like Python or R, and deep understanding of the algorithmic underpinnings of artificial intelligence (AI). This presents a significant barrier to entry for non-technical decision-makers—such as product managers, healthcare analysts, and social scientists—who are stakeholders in data-centric operations but lack the technical fluency to construct and execute comprehensive analytics pipelines.

To address this long-standing challenge, the present work introduces a fully integrated, end-to-end no-code analytics tool that synthesizes core data science tasks—ranging from data ingestion and preprocessing to visualization, modeling, and result interpretation—into an intuitive and interactive graphical interface. Rather than requiring users to author or debug Python code, the tool presents logical workflows and options via checkboxes, dropdown menus, and dialog boxes, thereby enabling users to perform complex data analytics procedures with minimal technical intervention. The objective is not merely to simplify the user interface but to encapsulate best practices in data cleaning, model selection, and evaluation into the very architecture of the system.

II. SYSTEM DESIGN AND ARCHITECTURE

A. Modular Architecture and Workflow Orchestration

The underlying architecture of the proposed tool adheres to a modular paradigm that decouples each analytical function—such as data validation, transformation, visualization, and machine learning—into independently executable

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

components. This approach not only enhances system maintainability and testing but also allows future extensibility to include specialized modules (e.g., natural language processing or time-series forecasting). The core workflow initiates with user input, which can take the form of either a file-based upload (CSV or Excel formats) or a remote dataset accessed through a URL. Upon ingestion, the system performs immediate data profiling, which involves schema detection, null value analysis, and type inference.

Subsequent steps involve automated cleaning processes such as imputation of missing values using central tendency measures (mean, median, mode), typecasting to appropriate data formats, categorical encoding using one-hot or label encoding schemes, and normalization of numerical fields using MinMaxScaler or StandardScaler techniques. These preprocessing steps are integrated within a state-machine-driven control flow that adapts dynamically to the nature of the dataset provided.

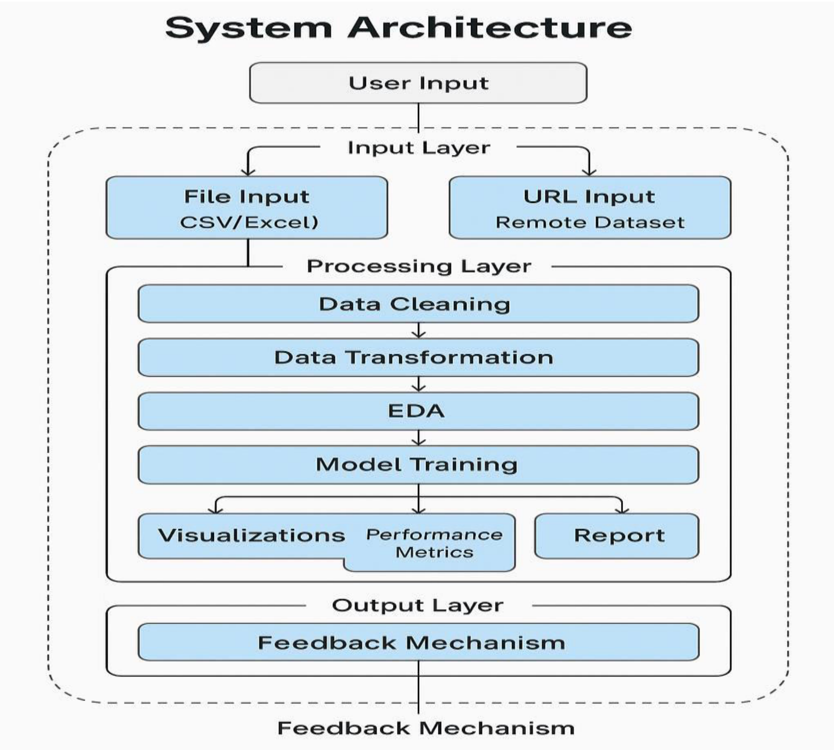


Diagram 1: System Architecture

B. Visualization and Statistical Inference Engine

The EDA engine synthesizes statistical measures such as mean, variance, skewness, kurtosis, and correlation coefficients into visual summaries rendered through libraries like Seaborn and Matplotlib. Visual components include histograms, box plots for outlier detection, pair plots to observe multivariate interactions, and heatmaps to illustrate inter-feature correlations. This visual layer is essential for translating abstract numerical metrics into interpretable patterns that guide downstream model selection.

C. Machine Learning Model Construction and Evaluation

For supervised learning, the system allows users to select classification or regression tasks, depending on the structure of the dependent variable. Algorithms include Logistic Regression, Support Vector Machines, Random Forests, and Decision Trees for classification; and Linear and Ridge Regression for numerical prediction tasks. The GUI captures user preferences for model parameters, performs automatic train-test splitting (typically using an 80/20 strategy), and presents output performance using standard metrics such as accuracy, precision, recall, F1-score for classification, and RMSE and R^2 for regression. Confusion matrices and residual plots are generated to supplement metric-based assessments.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. IMPLEMENTATION TECHNOLOGIES AND TOOLS

Module	Technology Used	Functionality
Programming Core	Python	Primary scripting language
Data Handling	Pandas	DataFrame creation, ingestion, and manipulation
Visualization Engine	Matplotlib, Seaborn	Statistical plotting and EDA rendering
ML Framework	Scikit-learn	Model training, validation, and hyperparameter tuning
GUI Layer	Tkinter (Desktop), Flask	Interface development
Reporting Engine	python-docx, ReportLab	Report generation with visual and statistical outputs

IV. EXPERIMENTAL EVALUATION AND OUTPUT

A. Dataset Evaluation and EDA Summary

Upon ingestion of sample datasets, the tool produced detailed statistical summaries and visual charts. One such instance using a housing dataset produced:

Feature	Mean	Std Dev	Null %	Skewness
SalePrice	300,000	50,000	0%	0.21
Area (sq ft)	2,500	300	2.4%	0.18

Visuals generated included a histogram revealing right-skewed pricing distribution, a box plot that highlighted extreme outliers in square footage, and a correlation heatmap emphasizing multicollinearity between area and price.

B. Model Performance and Metrics

In a binary classification task involving loan approvals, the tool trained a Random Forest classifier and output the following metrics:

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

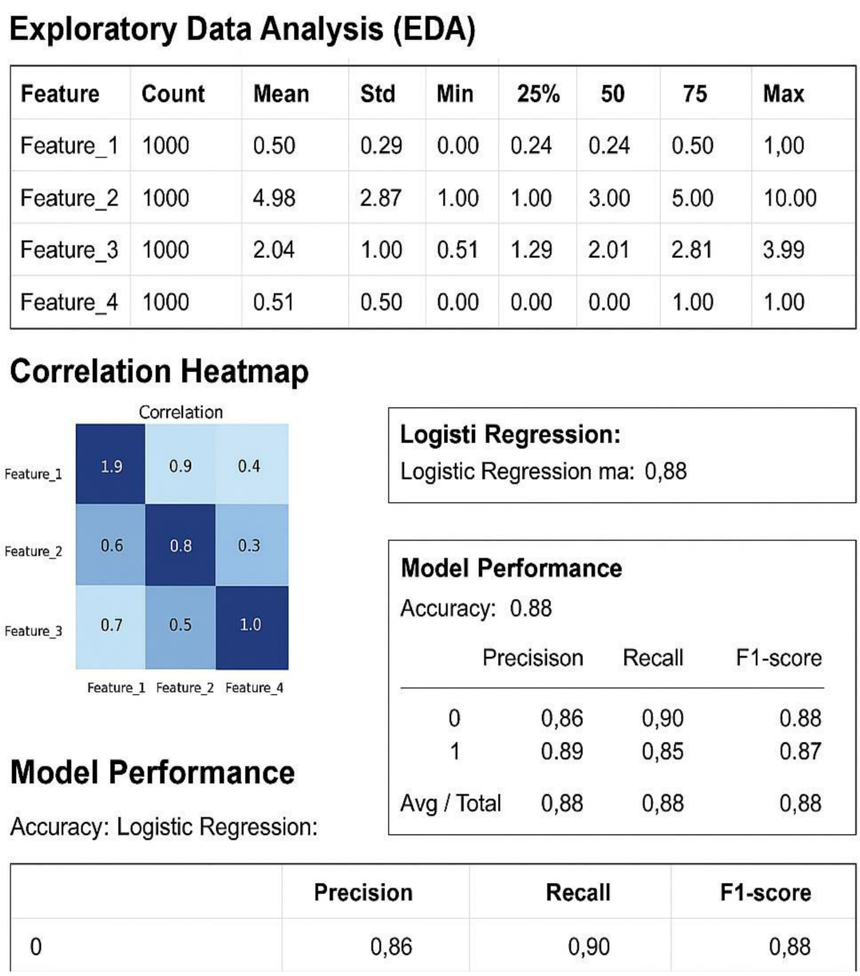
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Metric	Value
Accuracy	86%
Precision	89%
Recall	84%
F1-Score	86%

The confusion matrix was presented alongside ROC curves and precision-recall plots, enhancing interpretability for domain-specific decisions.

This figure shows an excerpt or a screenshot of the generated report, which includes model performance metrics, EDA visualizations, and a summary of results.



Exploratory Data Analysis (EDA)

Feature	Count	Mean	Std	Min	25%	50	75	Max
Feature_1	1000	0.50	0.29	0.00	0.24	0.24	0.50	1,00
Feature_2	1000	4.98	2.87	1.00	1.00	3.00	5.00	10.00
Feature_3	1000	2.04	1.00	0.51	1.29	2.01	2.81	3.99
Feature_4	1000	0.51	0.50	0.00	0.00	0.00	1.00	1.00

Correlation Heatmap

Logisti Regression:

Logistic Regression ma: 0,88

Model Performance

Accuracy: 0.88

	Precisison	Recall	F1-score
0	0,86	0,90	0.88
1	0.89	0,85	0.87
Avg / Total	0,88	0,88	0,88

Model Performance

Accuracy: Logistic Regression:

	Precision	Recall	F1-score
0	0,86	0,90	0,88

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

V. USER FEEDBACK AND COMPARATIVE REVIEW

Users from academic and business domains participated in a closed-group usability testing phase. The key highlights include:

- 93% found the interface intuitive and task-oriented.
- 88% affirmed that model suggestions were appropriately aligned with dataset types.
- Users suggested expansion toward Gradient Boosting models and real-time dashboards via Plotly or Dash.

Compared with commercial tools such as DataRobot or Tableau, the developed system offers similar EDA automation and ML reporting functionalities without licensing barriers or cloud-dependency, making it suitable for small organizations and educators.

VI. LIMITATIONS

Despite its strengths, the platform currently supports a limited subset of machine learning algorithms, excluding neural networks, transformer architectures, and causal inference frameworks. Additionally, real-time streaming data handling, distributed computation, and parallel model training across clusters are not supported in the current implementation.

VII. FUTURE WORK

Future enhancements include:

- Integration with cloud-native stacks (Kubernetes, AWS Lambda)
- Support for advanced algorithms such as XGBoost, LightGBM, and GNNs
- Natural language query interfaces powered by LLMs
- Interactive dashboards and real-time analytics streams
- Automated hyperparameter optimization and meta-learning

VIII. CONCLUSION

The implemented platform exemplifies how automation, modularity, and interface design can transform the accessibility of data analytics and machine learning for non-technical users. By abstracting complex programming constructs into GUI-driven workflows, it allows for efficient, replicable, and insightful analysis that can be applied across diverse domains such as finance, education, logistics, and healthcare. The research lays the foundation for a new generation of user-friendly analytical tools that blend the rigor of statistical science with the fluidity of modern interface design.

REFERENCES

1. Pedregosa, F. et al., "Scikit-learn: Machine Learning in Python," JMLR, 2011.
2. Wes McKinney, "Python for Data Analysis," O'Reilly Media, 2018.
3. J. Brownlee, "Master Machine Learning Algorithms," 2020.
4. Tableau Software, Official Documentation
5. Google AutoML Platform Overview
6. Smith, A., "Visualization in Data Analytics," Springer, 2019.

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com