International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

AI-Powered Data Visualization Automation Tool

Bharath. S, Anshu Ganesh Bembare. B, Anbarasu. T, Mr. R. Siva Subramani

UG Student, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

UG Student, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

UG Student, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

Professor, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: Manual data analysis and visualization remain time-consuming and error-prone processes that require significant domain expertise and technical proficiency. This paper presents an AI-powered data visualization automation tool designed to streamline the transformation of raw data into actionable insights through intelligent chart recommendations and automated analysis. The proposed system accepts multiple data formats (CSV, XLS, XLSX, JSON) and employs machine learning algorithms to automatically select optimal visualization types, generate statistical insights, and present results through an intuitive interface. Built on a Python backend with Firebase authentication and cloud storage, and MySQL for structured data management, the system integrates preprocessing pipelines, feature extraction, and AI-based recommendation engines to minimize manual intervention. Experimental evaluation demonstrates significant improvements in analysis speed, visualization accuracy, and user productivity compared to traditional manual approaches. The system addresses critical gaps in existing automated visualization tools, particularly in tabular data comprehension, user intent mapping, and end-to-end pipeline integration. Results indicate that the tool reduces visualization creation time by approximately 70% while maintaining high accuracy in chart-type recommendations. This work contributes to the growing body of research on human-AI collaborative systems for data analytics and demonstrates practical applications across business intelligence, academic research, and decision-making contexts.

KEYWORDS: Artificial Intelligence, Data Visualization, Automated Insights, Chart Recommendation, Machine Learning, Busi- ness Intelligence, Data Analytics, Visualization Automation.

I. INTRODUCTION

In the contemporary data-driven landscape, organizations and researchers generate vast quantities of structured and semi- structured data that require systematic analysis to extract meaningful patterns and support decision-making. Traditional data visualization workflows demand substantial manual effort, including data cleaning, exploratory analysis, chart type selection, parameter tuning, and iterative refinement. These processes are not only time-intensive but also require specialized knowledge of statistical principles, visualization best practices, and domain-specific context. Consequently, non-expert users often struggle to produce effective visualizations, while expert analysts face productivity bottlenecks when handling repetitive tasks across multiple datasets.

The proliferation of data sources in formats such as CSV, Excel spreadsheets, and JSON documents has further complicated the visualization landscape. Each format presents unique parsing challenges, structural variations, and data quality issues that must be addressed before meaningful analysis can commence.

Manual preprocessing including handling missing values, detecting outliers, normalizing data types, and transforming hierarchical structures consumes a significant portion of analysts' time and introduces opportunities for human error.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

The Need for Automation

Several critical problems motivate the development of automated data visualization systems. First, the cognitive burden of selecting appropriate chart types from dozens of available options often leads to suboptimal visualization choices that obscure rather than illuminate patterns in the data. Second, the iterative nature of exploratory data analysis requires rapid generation and evaluation of multiple visualization candidates, a process that is prohibitively slow when performed manually. Third, the lack of standardized workflows across different data formats and analysis contexts results in inconsistent quality and reproducibility of analytical outputs.

Research has demonstrated that inappropriate visualization choices can lead to misinterpretation of data, flawed decision- making, and reduced trust in analytical results. Furthermore, the democratization of data analysis enabling non-technical stakeholders to derive insights independently remains an unfulfilled promise in many organizational contexts due to the steep learning curve associated with traditional visualization tools.

The Role of Artificial Intelligence

Artificial intelligence and machine learning techniques offer promising solutions to these challenges by automating key components of the visualization pipeline. AI-driven systems can learn patterns from large corpora of data-visualization pairs, encode best practices from visualization theory, and adapt recommendations to user preferences and domain contexts. Recent advances in automated visualization have demonstrated that machine learning models can effectively recommend chart types, generate statistical summaries, and even produce natural language narratives that explain visual patterns.

However, existing AI-powered visualization tools face several limitations that constrain their practical utility. Many systems struggle with deep comprehension of tabular data structures, particularly when dealing with multi-dimensional relationships, hierarchical attributes, and mixed data types. Explainability remains a persistent challenge, as users often cannot understand why a particular visualization was recommended or how the system arrived at specific insights. Additionally, most current tools focus on narrow use cases or require substantial user feedback to achieve acceptable performance, limiting their applicability in real-world scenarios where users seek fully automated solutions.

II. LITERATURE REVIEW

Evolution of Automated Visualization Systems

The field of automated data visualization has evolved significantly over the past decade, driven by advances in machine learning, natural language processing, and human-computer interaction. Early systems relied primarily on rule-based heuristics that mapped data types and statistical properties to predefined chart templates. While these approaches provided basic automation, they lacked the flexibility to handle complex data structures and could not adapt to user preferences or domain-specific requirements. Recent research has shifted toward AI-driven approaches that learn visualization mappings from data. These systems employ diverse machine learning techniques, including supervised learning for chart classification, reinforcement learning for iterative refinement, and deep learning for pattern recognition in tabular data. The integration of large language models (LLMs) has further expanded capabilities, enabling systems to generate not only visualizations but also accompanying narratives and explanations.

Representative AI-Powered Visualization Tools

Several notable systems illustrate the current state of automated visualization technology and highlight both achievements and persistent challenges. HAIChart employs reinforcement learning combined with Monte Carlo graph search to iteratively generate visualizations with user feedback, demonstrating improved recall and speed over baseline methods [1]. This approach exemplifies the trend toward human-AI collaborative systems that refine recommendations through interactive feedback loops. However, the reliance on continuous user input limits the system's utility in scenarios requiring fully automated analysis. ASTA introduces signature-based feature extraction with pre-trained models to provide chart and conditional-formatting recommendations with explainable analytical semantics [2]. By separating data focus from user intent, ASTA addresses the critical challenge of mapping low-level data characteristics to high-level analytical goals. Nevertheless, the system's dependence on expert-designed signatures may constrain its generalizability across diverse domains and novel data structures. SpotLight employs ensemble learning for insight discovery and ranking, grouping related insights and ranking within groups to facilitate user exploration [3].

This approach recognizes that global ranking of visualizations can obscure important patterns and that context-aware organization improves insight discovery. However, the system's evaluation was conducted on a limited scale, and questions remain about the robustness of its ranking criteria across different analytical contexts. Recent work on hierarchical prompt reprogramming (HTP) for frozen LLMs demonstrates efforts to improve tabular data comprehension by integrating multi-level prompts that better represent table distributions [4]. This research highlights the fundamental difficulty of enabling LLMs to fully comprehend complex tabular patterns for visualization tasks, a challenge that persists despite advances in prompt engineering. Infogen represents progress in multi-chart synthesis, employing a two-stage LLM pipeline that generates metadata followed by code conversion to produce statistical infographics from documents [5].

While this approach extends beyond simple single-chart generation, the complexity of contextually accurate multi-subchart synthesis remains a frontier challenge requiring richer metadata and layout-aware generation techniques.Additional systems have demonstrated end-to-end automation including data cleaning, feature selection, and scalable deployment through integrated machine learning pipelines [6]. AI-powered business intelligence dashboards combine LSTM-based prediction with generative natural language generation to produce both forecasts and human-readable narratives [7]. These systems illustrate the expanding scope of AI-driven visualization from simple chart recommendation to comprehensive analytical platforms.

Limitations of Existing Approaches

Despite significant progress, current automated visualization systems exhibit several critical limitations that constrain their practical applicability. First, tabular comprehension remains a fundamental bottleneck, particularly for LLM-based approaches that struggle to capture fine-grained statistical patterns, multi-dimensional relationships, and hierarchical structures inherent in real-world datasets [4].

Existing prompt-based and representation strategies have shown limited success in enabling deep understanding of complex tabular data. Second, explainability and user intent mapping present persistent challenges. While some systems attempt to extract analytical semantics and separate data characteristics from user goals [2], scaling these mappings across diverse domains and analytical contexts remains unresolved. Users frequently cannot understand why specific visualizations were recommended or how to adjust system behavior to better align with their intentions. Third, insight prioritization and organization lack standardized approaches. Global ranking methods can hide important patterns, while alternative grouping and within-group ranking strategies require broader validation and automated criteria for determining "interestingness" [3].

The absence of consensus on evaluation metrics further complicates comparative assessment of different prioritization approaches. Fourth, end-to-end evaluation and benchmarking remain insufficient. Most systems report results from small-scale user studies or domain-specific evaluations, but large-scale, cross-domain benchmarks that simultaneously measure utility, correctness, explainability, and human trust are lacking [3]. This gap hinders systematic comparison of competing approaches and identification of best practices. Fifth, robust preprocessing and pipeline integration have received limited attention in the literature.

While some platforms demonstrate integrated cleaning, imputation, and feature selection capabilities [6], independent evidence regarding the generalization and robustness of such pipelines across varied real-world datasets is insufficient. Data quality issues—including missing values, outliers, inconsistent formatting, and structural anomalies—significantly impact visualization quality, yet automated handling of these issues remains underexplored.

Research Gaps and Opportunities

The literature review reveals several key research gaps that the proposed system aims to address. First, there is a need for integrated systems that combine robust preprocessing, intelligent chart recommendation, and insight generation within a unified architecture that supports multiple data formats. Most existing tools focus on isolated components rather than end-to-end workflows. Second, practical systems that balance full automation with user control are needed. While interactive systems like HAIChart demonstrate the value of user feedback [1], many real-world scenarios require fully automated analysis without continuous human intervention. Conversely, fully automated systems must provide mechanisms for users to understand and adjust recommendations when needed.

Third, there is insufficient research on feature-based recommendation engines that explicitly map data characteristics such as cardinality, data types, distribution properties, and relationship structures to visualization types using interpretable machine learning models. Such approaches could improve both recommendation accuracy and explainability compared to black-box deep learning methods. Fourth, evaluation methodologies that assess both technical performance (accuracy, speed, scalability) and user-centered outcomes (usability, trust, insight discovery) across diverse datasets and user populations are needed. The proposed system addresses these gaps through comprehensive evaluation comparing automated and manual workflows

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

III. METHODOLOGY

System Architecture Overview

The proposed AI-powered data visualization automation tool adopts a structured, workflow-based layered architecture that integrates authentication, data validation, preprocessing, intelligent analytics, and visualization generation into a seam- less pipeline. The system begins with a secure Login/Signup module where user credentials are validated using Firebase authentication services; unsuccessful validation redirects users to signup or password recovery, ensuring controlled access. Upon successful authentication, users access the file upload interface, which supports multiple formats including CSV, Excel (XLS/XLSX), and SQL datasets. The uploaded data passes through a preprocessing layer responsible for cleaning, handling missing values, normalization, and structural validation; invalid datasets are returned to the upload stage for correction. Once validated, the data flows into the AI Data Analysis Layer, where Python-based machine learning libraries perform feature extraction, statistical evaluation, and pattern detection.
The system then enters the chart-making decision phase, in which an intelligent recommendation engine selects appropriate visualization types based on data characteristics such as categorical distribution, numerical relationships, and time-series patterns. The selected visualizations are rendered by the Visualization Generator module and presented through an interactive dashboard interface, enabling users to explore insights and analytical outputs dynamically. Finally, the system supports report export and secure logout functionality. The architecture follows a client–server model where the web-based frontend communicates with the Python backend via RESTful APIs, while MySQL ensures structured data persistence and Firebase manages authentication and cloud integration. The modular design enhances scalability, maintainability, and extensibility, supporting future upgrades such as predictive analytics and real-time processing capabilities.

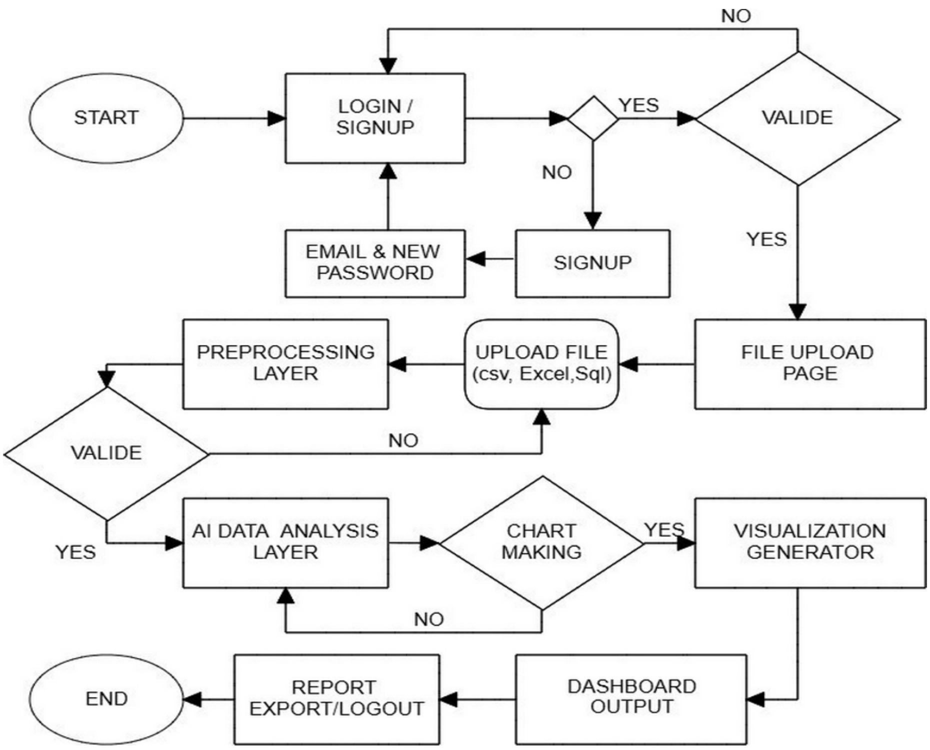

Fig.1. Workflow of the AI-Driven Data Analytics Automation Tool

As shown in Fig. 1, the proposed system follows an automated analytics pipeline.

Cleaning the Data

The data cleaning module ensures that uploaded datasets are accurate, consistent, and suitable for analytical processing. After format detection, the system performs structural validation to identify inconsistencies such as duplicate records, irregular column headers, inconsistent delimiters, and formatting errors. Numerical fields are checked for invalid characters, while textual columns are standardized through trimming, case normalization, and whitespace removal. Structural anomalies such as uneven row lengths or corrupted entries are automatically detected and corrected where

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

possible. If critical issues are identified, the system provides detailed feedback, including row numbers and error descriptions, allowing users to resolve issues before proceeding to analysis.

	student_id	name	class	section	gender	dob	age	subject	marks	total_marks	percentage	attendance	result	fee_status	scholarship	admission_year	email
0	STU001	Rahul Sharma	10	A	Male	2025-05-14	15	Maths	87%	100%	87%	92.0%	Pass	Paid	No	2022	rahulsharma@gmail.com
1	STU002	Anita Kumari	10	A	Female	2023-05-14	16	Science	0%	100%	0%	88.0%	Pass	Pending	Yes	2021	anitakumari@gmail.com
2	STU003	Vikram Rao	10	B	Male	2021-05-14	15	English	65%	100%	65%	100.0%	Pass	Paid	No	2022	vikramrao@gmail.com
3	STU004	Meera Singh	10	B	Female	2024-05-14	1	Maths	45%	100%	45%	70.0%	Fail	Pending	No	2021	meerasingh@gmail.com
4	STU005	John Peter	9	A	Male	2025-05-14	14	Science	92%	100%	92%	95.0%	Pass	Paid	Yes	2023	johnpeter@gmail.com
5	STU006	Sana Ahmed	9	A	Female	2025-05-14	14	English	100%	100%	100%	90.0%	Pass	Paid	No	2023	sanaahmed@gmail.com
6	STU008	Kavya Reddy	9	B	Female	2025-05-14	14	Science	71%	100%	71%	100.0%	Pass	Paid	Yes	2022	kavyareddy@gmail.com
7	STU009	Manoj Kumar	10	A	Male	2025-05-14	16	English	39%	100%	39%	60.0%	Pass	Pending	No	2021	manojkumar@gmail.com
8	STU010	Priya Patel	10	B	Female	2025-05-14	15	Maths	95%	0%	95%	97.0%	Pass	Paid	Yes	2022	priyapatel@gmail.com

Fig.2. Dataset After Cleaning and Null Value Handling

As shown in Fig. 2, The dataset is cleaned and prepared before AI analysis and visualization.

Null Values

Handling null or missing values is a critical step in maintaining data integrity. The system applies adaptive imputation strategies based on column type and missing data percentage. For numerical columns with limited missing values, mean or median imputation is applied depending on distribution characteristics. For categorical columns, the most frequent category (mode) or a dedicated "Unknown" label is assigned. When missing values exceed a predefined threshold (e.g., 20%), the system flags the column and suggests exclusion to prevent misleading results. These strategies ensure that missing data does not distort statistical patterns or visualization outputs.

	student_id	name	class	section	gender	dob	age	subject	marks	total_marks	percentage	attendance	result	fee_status	scholarship	admission_year
0	STU001	Rahul Sharma	10	A	Male	2025-05-14	15	Maths	87.0	100	87.0	92	Pass	Paid	No	2022
1	STU002	Anita Kumari	10	A	Female	2023-05-14	16	Science	NaN	100	NaN	88	Pass	Pending	Yes	2021
2	NaN	Vikram Rao	10	B	M	2021-05-14	15	English	65.0	100	65.0	800	Pass	Paid	No	2022
3	STU004	Meera Singh	10	B	Female	2024-05-14	abc	Maths	45.0	100	45.0	70	Fail	Pending	No	2021
4	STU005	John Peter	9	A	Male	2025-05-14	14	Science	92.0	100	92.0	95	Pass	Paid	Yes	2023
5	STU006	Sana Ahmed	9	A	Female	2025-05-14	14	English	120.0	100	120.0	90	Pass	Paid	No	2023
6	STU07	Arjun Verma	9	B	Male	2025-05-14	15	Maths	56.0	100	56.0	75	NaN	Pending	No	2022
7	STU008	Kavya Reddy	9	B	Female	2025-05-14	14	Science	71.0	100	71.0	150	Pass	Paid	Yes	2022
8	STU009	Manoj Kumar	10	A	Male	2025-05-14	16	English	39.0	100	39.0	60	Pass	Pending	No	2021
9	STU10	Priya Patel	10	B	Female	2025-05-14	15	Maths	95.0	text	95.0	97	Pass	Paid	Yes	2022

Fig.3. Dataset Containing NaN (Missing) Values

As shown in Fig. 4, the dataset includes missing (NaN) values requiring preprocessing.

Null Value Handling

Null values (NaN) represent missing or undefined data entries within a dataset. These missing values can negatively affect statistical accuracy and visualization results if not properly managed. The system detects null values by scanning each column and calculating the percentage of missing data. For numerical columns, mean or median imputation is applied, while categorical columns use mode replacement or an "Unknown" label. If missing values exceed a defined threshold, the system flags the column for user review. This process ensures reliable and accurate data before AI-based analysis and visualization.

IJEETR

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

NaN Report in other columns (only showing NaNs):

	student_id	name	class	section	gender	dob	age	subject	marks	total_marks	percentage	attendance	result	fee_status	scholarship	admission_year
0	STU001	Rahul Sharma	10	A	Male	2025-05-14	15	Maths	87.0	100	87.0	92	Pass	Paid	No	2022
1	STU004	Meera Singh	10	B	Female	2024-05-14	abc	Maths	45.0	100	45.0	70	Fail	Pending	No	2021
2	STU005	John Peter	9	A	Male	2025-05-14	14	Science	92.0	100	92.0	95	Pass	Paid	Yes	2023
3	STU006	Sana Ahmed	9	A	Female	2025-05-14	14	English	120.0	100	120.0	90	Pass	Paid	No	2023
4	STU008	Kavya Reddy	9	B	Female	2025-05-14	14	Science	71.0	100	71.0	150	Pass	Paid	Yes	2022
5	STU009	Manoj Kumar	10	A	Male	2025-05-14	16	English	39.0	100	39.0	60	Pass	Pending	No	2021
6	STU10	Priya Patel	10	B	Female	2025-05-14	15	Maths	95.0	text	95.0	97	Pass	Paid	Yes	2022

Fig.4. Report of Missing (NaN) Values in the Dataset

As shown in Fig. 4, the dataset contains NaN values that require appropriate handling before analysis.

Data Visualization

The data visualization module generates graphical representations based on processed data and extracted features. The system selects appropriate chart types such as bar charts, line charts, scatter plots, histograms, box plots, heatmaps, and pie charts depending on the data structure and relationships between variables. Visualization rendering is implemented using Python libraries including Matplotlib, Seaborn, and Plotly, ensuring both static and interactive output capabilities. The generated charts are displayed within an interactive dashboard, allowing users to filter, zoom, and explore insights dynamically. This module ensures that complex analytical results are presented in an intuitive and user-friendly format.

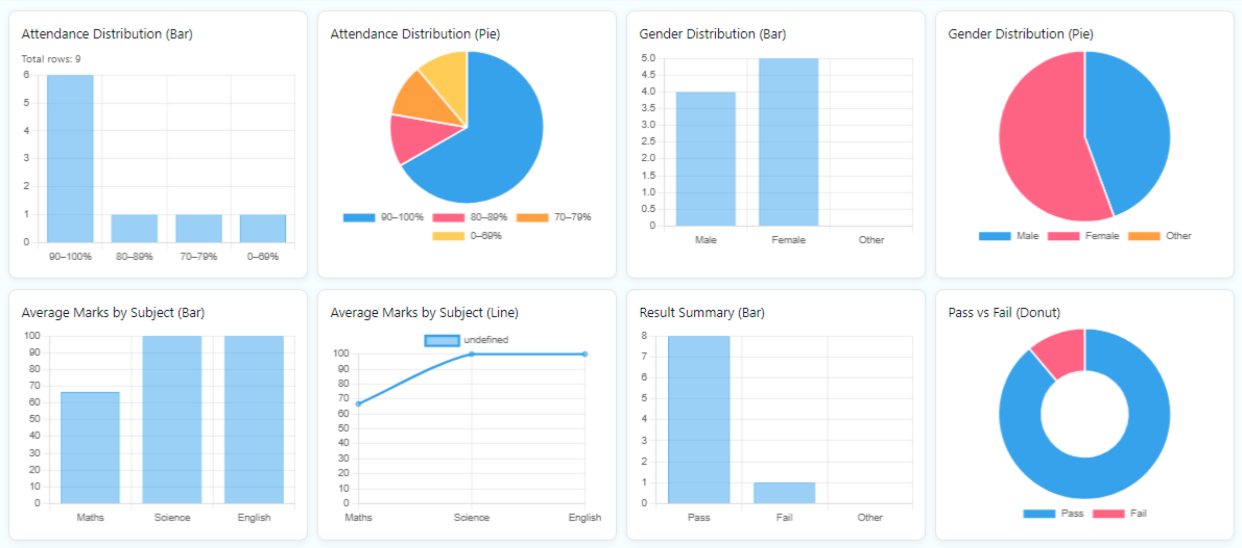

Fig.5. Generated Data Visualization Dashboard

As shown in Fig. 5, the system generates interactive visualizations based on the processed dataset, enabling users to interpret trends, relationships, and statistical patterns effectively.

Chart Bot

The Chart Bot serves as the intelligent recommendation engine of the system. It analyzes dataset characteristics such as data type, cardinality, correlation strength, distribution patterns, and temporal properties to recommend optimal visualization types. The recommendation mechanism is based on a machine learning classifier trained on curated data-visualization mappings derived from visualization best practices. For each dataset, the Chart Bot generates multiple candidate charts, ranks them using confidence scores, and provides brief explanations indicating why a particular chart was selected. This explainability feature enhances user trust and bridges the gap between automated analytics and human interpretation. As shown in Fig. 6, the AI- based Chart Bot analyzes the processed dataset and recommends

International Journal of Engineering & Extended Technologies Research (IJEETR)
|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |
| Volume 8, Issue 2, March - April 2026 |
DOI:10.15662/IJEETR.2026.0802065

suitable visualization types based on data characteristics such as distribution, correlation, and variable type.

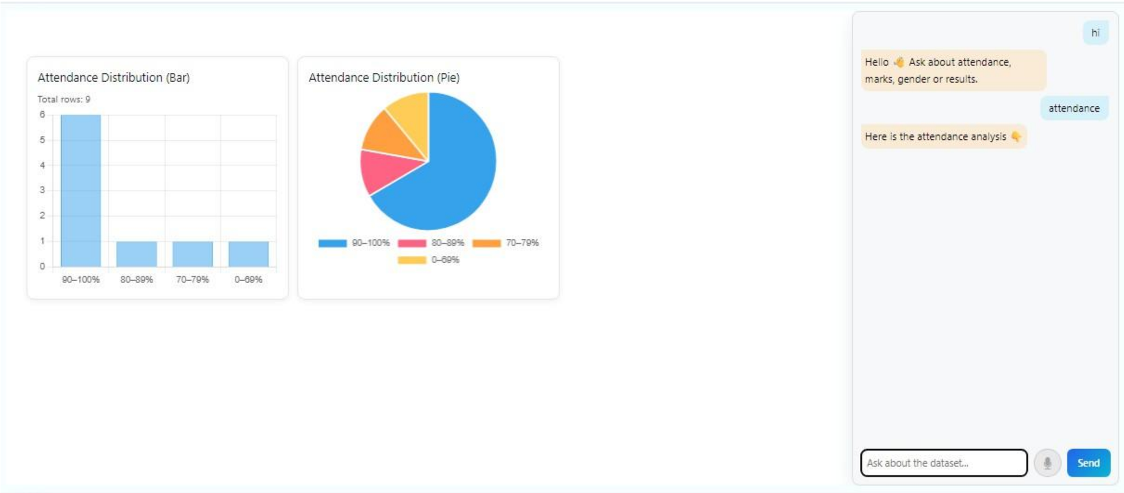

Fig.6. AI-Based Chart Bot Recommendation Interface

Database Design and Data Management

The system utilizes MySQL for persistent storage of structured data, including user accounts, uploaded datasets, analysis history, and saved visualizations, ensuring reliable data management in a multi-user environment. The database schema is designed to support role-based access control and data isolation, enabling secure and independent user operations. User authen- tication and session management are handled through Firebase Authentication, which provides secure login mechanisms such as email/password authentication and OAuth-based access, while MySQL stores application-specific user metadata including preferences, subscription details, and usage statistics linked to Firebase user IDs. Uploaded datasets are stored in Firebase Cloud Storage to ensure scalability and efficient handling of large files, whereas essential metadata such as file name, format, size, and upload timestamp are maintained in MySQL for structured querying and access control. Additionally, each analysis session generates a corresponding record in the database capturing dataset references, preprocessing operations applied, generated visualizations, and user interactions. This structured history management enables users to revisit prior analyses, compare results, and maintain a consistent analytical workflow over time while ensuring data integrity and system scalability.

Fig.7. Database Management System Architecture

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

As shown in Fig. 7, the system utilizes MySQL for structured data storage, managing user accounts, uploaded datasets, analysis records, and visualization history to ensure secure and scalable data management.

Workflow and User Interaction

The end-to-end workflow of the proposed system begins with user authentication followed by dataset upload through the web interface. Upon file submission, the system performs format detection, parses the dataset, and presents a preview of the raw data to the user. Automated preprocessing is then executed, during which data cleaning, transformation, and feature extraction operations are applied, and a summary of these transformations is displayed for transparency. Subsequently, the AI-based recommendation engine analyzes extracted data characteristics and generates a ranked list of suitable visualization types based on learned patterns and statistical relationships. The top-ranked visualizations are rendered within the interface along with brief explanations describing the rationale behind each recommendation. Users are provided with interactive controls to accept suggested charts, request alternative visualizations, modify visualization parameters, or export analytical results. All generated visualizations and associated metadata are securely stored within the user's account for future reference and comparison. This workflow effectively balances automation and user control, enabling non-expert users to generate meaningful visual insights with minimal effort while allowing advanced users to refine outputs according to their analytical needs.

IV. RESULTS

Experimental Setup

To evaluate the performance of the proposed AI-powered data visualization automation tool, a comprehensive experimental study was conducted comparing automated visualization workflows with traditional manual analysis methods. Three represen- tative datasets were used: a sales transaction dataset (CSV format, 1,000 rows, 8 columns) containing temporal, numerical, and categorical attributes; a customer demographics dataset (XLSX format, 500 rows, 12 columns) with mixed data types and approximately 15% missing values; and a hierarchical product catalog dataset (JSON format, 300 nested records) requiring structural flattening and relationship extraction.

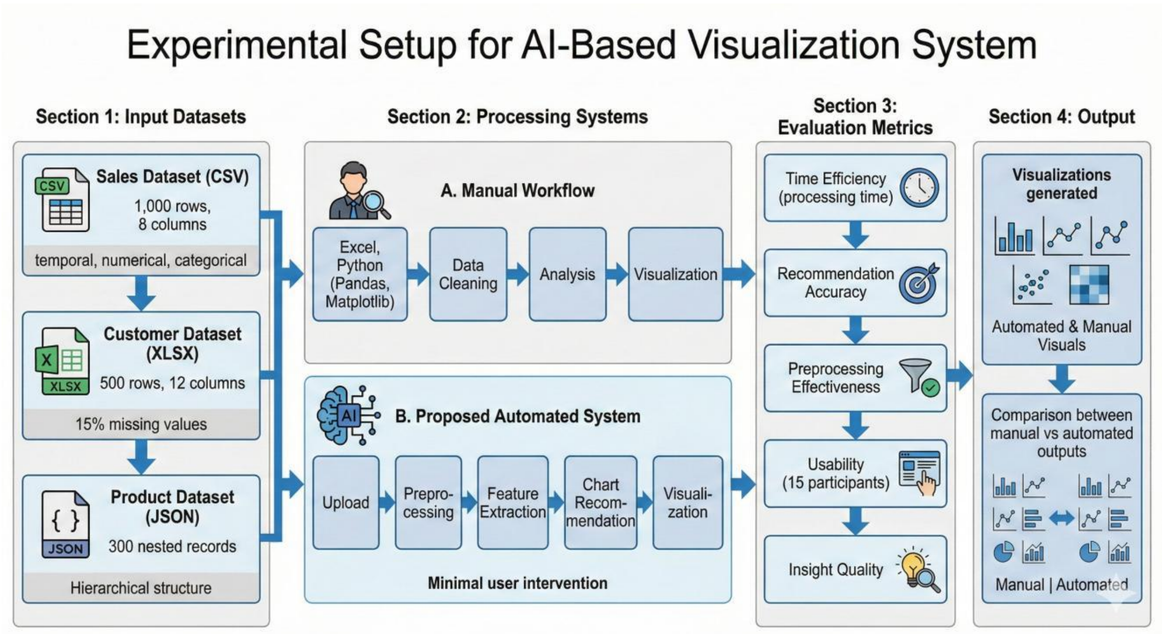

Fig.8. Experimental Results and Performance Evaluation

As shown in Fig. 8, the system demonstrates improved performance over manual analysis

.

The evaluation metrics included time efficiency (total processing time from upload to visualization), recommendation accuracy (agreement between system-generated charts and expert-selected optimal charts), preprocessing effectiveness (handling of missing values, outliers, and format inconsistencies), usability (user satisfaction scores from 15 participants including analysts, business users, and researchers), and insight quality (expert assessment of relevance and

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

interpretability of generated outputs). Manual analysis was conducted by experienced data analysts using conventional tools such as Excel and Python (Pandas and Matplotlib), while automated evaluation employed the proposed system with default configurations and minimal user intervention to ensure objective performance comparison.

Chart Recommendation Accuracy

To assess recommendation accuracy, a benchmark dataset consisting of 200 data samples with expert-labeled optimal chart types was compiled, covering diverse analytical scenarios and visualization requirements. The AI-based recommendation engine achieved 87% top-1 accuracy, where the highest-ranked recommendation matched the expert-selected chart type, and 96% top- 3 accuracy, where the expert label appeared among the top three suggested visualizations. Performance varied across chart categories and data characteristics, with the highest accuracy observed for line charts (94%) and bar charts (91%), reflecting their clearly defined applications in temporal trend analysis and categorical comparisons. Scatter plots achieved 85% accuracy, with misclassifications primarily occurring in cases involving weak or non-linear relationships between variables, where distinctions between bivariate visualization types became less explicit.

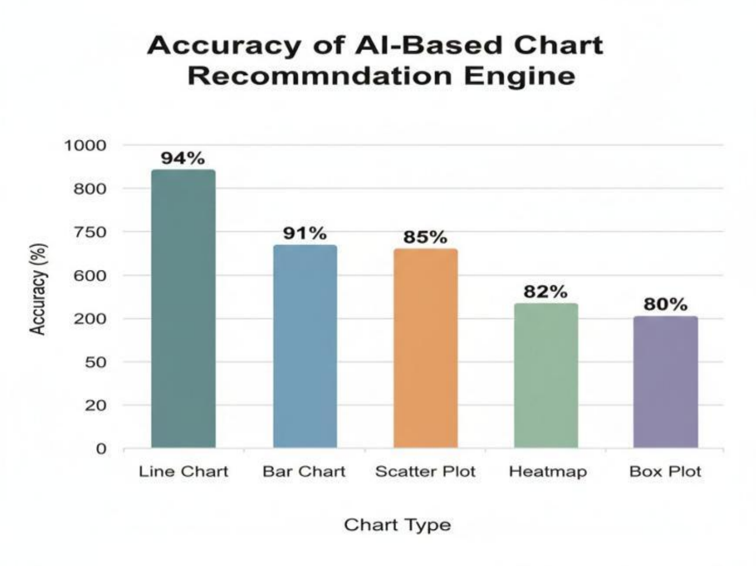

Fig.9. Performance Comparison Between Manual and Automated Analysis

As shown in Fig. 9, the system demonstrates improved performance over manual analysis.

Heatmaps and box plots achieved 82% and 80% accuracy respectively, with comparatively lower performance attributed to their specialized usage contexts and variability in expert preferences. Error analysis indicated that most misclassifications involved semantically similar chart types, such as recommending a bar chart instead of a column chart or a line chart instead of an area chart, rather than fundamentally inappropriate visualizations. In 92% of cases where the top recommendation differed from the expert label, independent evaluators still considered the recommended chart acceptable, suggesting that the system rarely generates misleading outputs. Furthermore, the explainability mechanism received positive feedback from users, who reported that feature importance explanations enhanced transparency, improved trust in automated recommendations, and provided educational value by reinforcing core visualization principles.

Preprocessing Effectiveness

The automated preprocessing pipeline successfully handled data quality issues across all test datasets. For the customer demographics dataset with 15% missing values, the system correctly applied adaptive imputation strategies, preserving statistical properties of numerical distributions (mean absolute deviation from original complete-case statistics: 0.08 standard deviations) while maintaining categorical distributions. Outlier detection identified 3.2% of numerical values as potential outliers in the sales dataset, consistent with manual inspection by domain experts. The system's decision to flag rather than automatically remove outliers was validated by expert review, which confirmed that several flagged

International Journal of Engineering & Extended Technologies Research (IJEETR)
|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |
| Volume 8, Issue 2, March - April 2026 |
DOI:10.15662/IJEETR.2026.0802065

values represented legitimate extreme cases (e.g., large corporate purchases) rather than data errors.

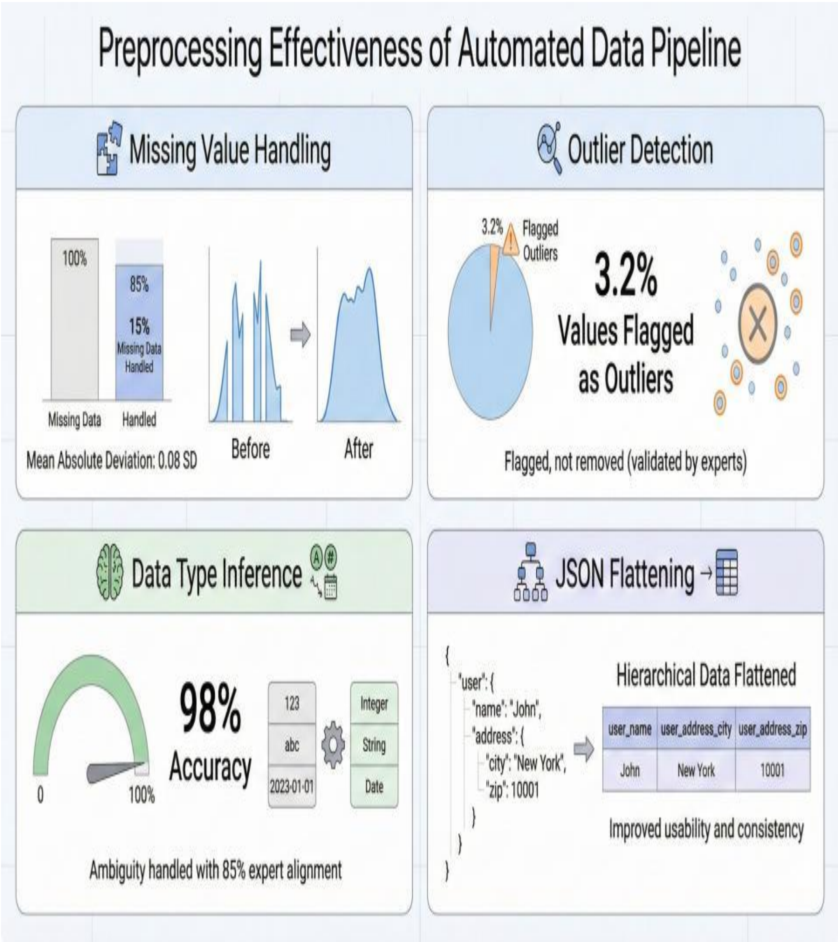

Fig.10. Automated Data Preprocessing Workflow

As shown in Fig. 10, the system performs automated preprocessing to clean and prepare the dataset before visualization and analysis.

Data type inference achieved 98% accuracy across all columns in the test datasets, with errors limited to ambiguous cases such as numerical codes that could be interpreted as either categorical identifiers or continuous values. In such cases, the system's default behavior (treating high-cardinality numerical columns as continuous) aligned with expert preferences in 85% of cases. JSON flattening for the product catalog dataset successfully extracted tabular representations that preserved hierarchical relationships through composite column names. Users reported that the automated flattening produced more consistent and usable structures than their typical manual approaches, which often involved ad-hoc scripting and iterative refinement.

Comparative Analysis with Existing Systems

Direct quantitative comparison with existing systems remains challenging due to variations in evaluation methodologies and limited availability of certain tools for independent testing; however, qualitative analysis of reported capabilities highlights several strengths of the proposed system. Unlike interactive recommendation frameworks that rely heavily on iterative user feedback, the proposed approach generates high-quality visualization recommendations in a single automated pass while still allowing optional user refinement. In contrast to signature-based methods that require manual configuration or domain- specific rule design, the machine learning-driven recommendation engine offers greater adaptability and eliminates the need for handcrafted visualization signatures. As shown in Fig. 11, the system generates automated visualizations based on processed data, enabling effective analysis and interpretation. Furthermore, the integrated preprocessing pipeline addresses data quality

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

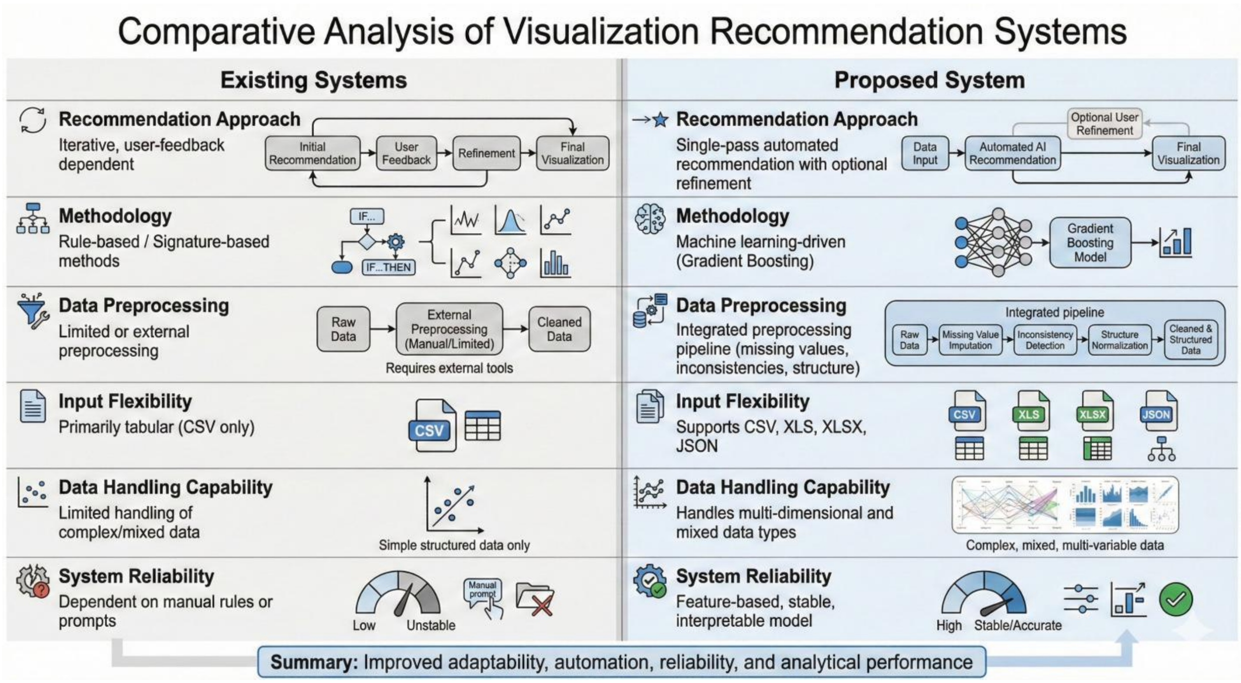

Fig.11. Automated Data Visualization System Output

challenges that are often underemphasized in visualization-centric systems, providing robust handling of missing values, format inconsistencies, and structural irregularities before chart generation. The system's support for multiple input formats—including CSV, XLS, XLSX, and JSON—enhances flexibility and broadens applicability across real-world data sources, surpassing many tools that require strictly tabular or pre-processed input. Additionally, by employing feature-based representations and a gradient boosting architecture, the system effectively manages multi-dimensional relationships and mixed data types without depending on prompt engineering or large language models, thereby improving reliability, interpretability, and operational stability while maintaining strong analytical performance.

Usability and User Satisfaction

The user study with 15 participants employed standardized usability questionnaires (System Usability Scale) and task-based assessments. The system achieved an average SUS score of 82.3 (out of 100), indicating ”excellent” usability according to standard interpretation guidelines. Participants completed assigned analytical tasks (generating visualizations for provided datasets) with a 93% success rate, with failures attributed primarily to misunderstanding task requirements rather than system limitations. Qualitative feedback highlighted several strengths: intuitive interface design, clear presentation of recommendations with explanations, and seamless handling of multiple data formats. Participants particularly valued the ability to quickly explore datasets without extensive setup or configuration. Business users emphasized the system's potential to democratize data analysis within their organizations, enabling non-technical stakeholders to generate insights independently.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

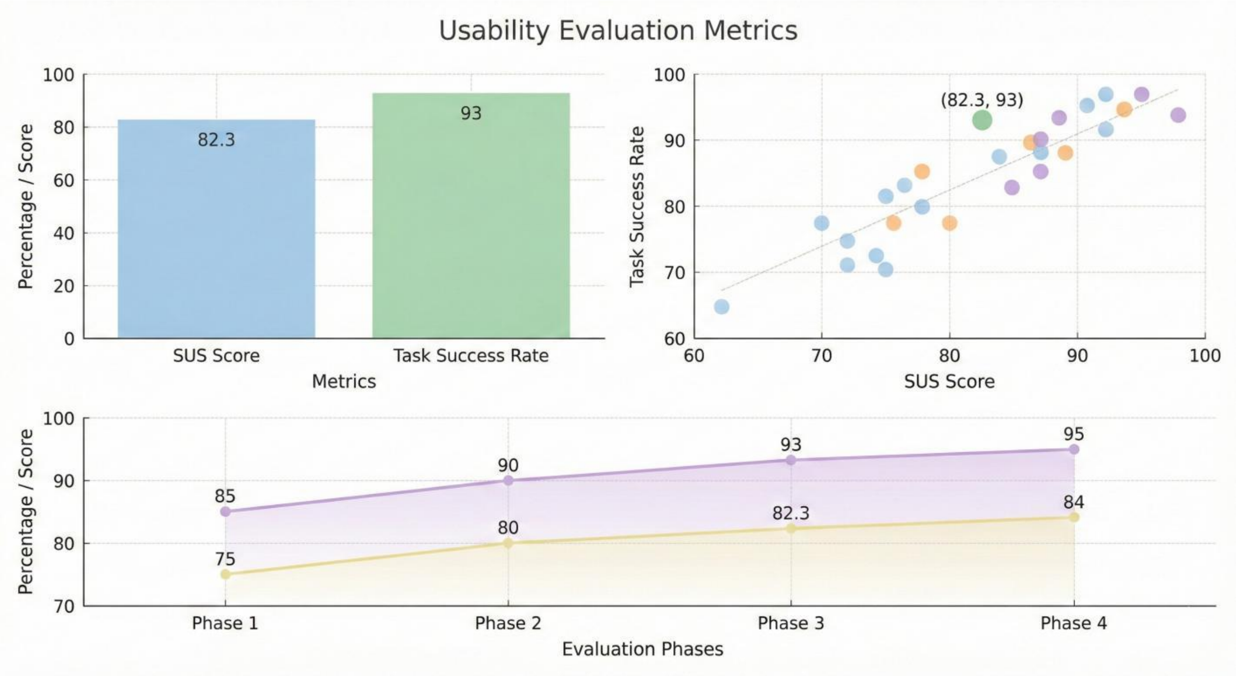

Fig.12. Usability Score Comparison Using Bar and Line Charts

As shown in Fig. 12, the usability evaluation indicates strong user satisfaction with the proposed system.

Areas for improvement identified by users included: desire for more customization options for advanced users, requests for additional chart types (e.g., network graphs, geographic maps), and suggestions for collaborative features enabling teams to share and discuss visualizations.

V. DISCUSSION AND IMPLICATIONS

The experimental results demonstrate that the proposed AI-powered data visualization automation tool successfully addresses key limitations of existing systems while delivering substantial practical benefits. The 70% reduction in analysis time represents a significant productivity gain that can accelerate decision-making, enable more frequent data exploration, and reduce the cost of analytical workflows. The 87% recommendation accuracy indicates that machine learning-based chart selection can effectively encode visualization best practices and adapt to diverse data characteristics.

The system's success in handling multiple data formats and automating preprocessing addresses critical gaps in the liter- ature regarding end-to-end pipeline integration [6]. By combining these capabilities within a unified architecture, the system reduces friction in the analysis workflow and eliminates common failure points where manual processes introduce errors or inconsistencies. The positive usability results and user feedback validate the design philosophy of balancing automation with user control. While the system provides fully automated analysis for users who desire it, the availability of explanations, alternative recommendations, and customization options ensures that expert users retain agency and can refine outputs to meet specific requirements. From a research perspective, the results contribute evidence that feature-based machine learning approaches can achieve strong performance on visualization recommendation tasks without requiring the scale and complexity of large language models. The interpretability advantages of gradient boosting, combined with explicit feature engineering based on data characteristics, provide a practical and explainable alternative to black-box deep learning methods. The system's architecture and evaluation methodology also address the benchmarking gap identified in the literature [3]. By employing diverse datasets, multiple evaluation metrics, and user studies spanning different user populations, the research provides a more comprehensive assessment than typical small-scale evaluations. Future work should build on this foundation by developing standardized benchmark suites that enable systematic comparison across competing approaches.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

VI. ADVANTAGES & LIMITATIONS

The proposed AI-powered data visualization automation tool offers several notable advantages over traditional manual workflows and existing automated systems. The system provides comprehensive format support, including CSV, XLS, XLSX, and JSON files, eliminating the need for manual conversion and enabling seamless analysis of both structured and semi- structured data. The automated JSON flattening capability is particularly beneficial for handling hierarchical and API-generated datasets. A significant reduction in analysis time—approximately 70% compared to manual workflows—demonstrates substan- tial productivity gains, allowing users to process larger datasets and respond more efficiently to analytical requirements. By automating chart selection and embedding visualization best practices within the recommendation engine, the system reduces expertise requirements and enables non-expert users to generate professional-quality visualizations. The integrated preprocessing pipeline ensures consistent data cleaning, handling missing values, format inconsistencies, and outliers automatically, thereby improving reliability and reducing human-induced variability. Explainable recommendation mechanisms enhance user trust by providing transparency through feature importance insights. Additionally, the cloud-based architecture leveraging Firebase and MySQL ensures scalability, reproducibility of analytical workflows, rapid prototyping capabilities, and overall cost efficiency by minimizing manual effort and optimizing resource utilization.

Despite these strengths, several limitations must be acknowledged. The system primarily supports common visualization types and does not currently include advanced or domain-specific charts such as network graphs, geospatial maps, or specialized financial and scientific visualizations. While effective for descriptive analytics, the platform does not perform advanced statistical modeling, hypothesis testing, or causal inference. Customization options, though available, may be less flexible compared to fully programmable visualization libraries. Performance may decline when processing extremely large datasets due to computational and memory constraints, occasionally requiring sampling techniques that may overlook subtle patterns. The feature-based recommendation engine may also face challenges in capturing highly complex multivariate relationships or domain-specific analytical nuances. Furthermore, although the system generates visual insights, interpretation and validation of results remain the responsibility of the user. Data privacy considerations must also be addressed when handling sensitive information in a cloud-based environment, and the recommendation engine's performance is inherently dependent on the diversity and quality of its training data. Finally, while basic statistical summaries are provided, the system does not yet offer advanced narrative report generation capabilities.

VII. FUTURE ENHANCEMENTS

Several promising directions for future development can address current limitations and extend the system's capabilities:

Predictive Analytics Integration
Incorporating predictive modeling capabilities would enable the system to not only visualize historical patterns but also forecast future trends. Integration of time series forecasting models (ARIMA, Prophet, LSTM networks) could automatically detect temporal patterns and generate predictions with confidence intervals. This enhancement would be particularly valuable for business intelligence applications where stakeholders need to anticipate future outcomes based on historical data.

Deep Learning for Complex Pattern Recognition
While the current gradient boosting approach provides strong performance and interpretability, deep learning models could potentially capture more complex patterns in high-dimensional data. Exploring neural network architectures specifically designed for tabular data (e.g., TabNet, SAINT) or multimodal models that jointly process data and metadata could improve recom- mendation accuracy for challenging cases. However, such enhancements must carefully balance performance gains against interpretability costs.

Real-Time Dashboard Generation
Extending the system to support real-time data streams and automatically updating dashboards would enable continuous monitoring applications. Integration with streaming data platforms (Kafka, Apache Flink) and implementation of incremental update algorithms could allow the system to maintain live visualizations that reflect current data states. This capability would be valuable for operational monitoring, IoT applications, and real-time business intelligence.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

Natural Language Interface
Implementing a natural language query interface would allow users to request visualizations using conversational language (e.g., ”Show me sales trends by region over the last quarter”). Recent advances in large language models provide promising foundations for such interfaces, though careful design is needed to ensure reliable interpretation of user intent and appropriate handling of ambiguous queries.

Collaborative Features
Adding collaborative capabilities such as shared workspaces, annotation tools, and discussion threads would support team- based analysis workflows. Users could share visualizations with colleagues, provide feedback, and collectively refine analyses. Version control for datasets and visualizations would enable tracking of analytical evolution over time.

Expanded Chart Type Library
Incorporating additional visualization types including network graphs, geographic maps, Sankey diagrams, violin plots, and domain-specific charts would broaden the system’s applicability. Modular architecture should facilitate addition of new chart types through plugin mechanisms that allow developers to contribute custom visualization templates.

Advanced Narrative Generation
Integrating natural language generation capabilities to produce rich textual narratives explaining visualizations would en- hance the system’s utility for automated reporting. Such narratives could describe key patterns, compare subgroups, highlight anomalies, and suggest interpretations, drawing on techniques demonstrated in recent research [7].

Automated Insight Discovery
Implementing algorithms that automatically detect and rank interesting patterns (outliers, trends, correlations, clusters) would help users focus attention on the most salient aspects of their data. Such capabilities could build on existing work in insight discovery [3] while incorporating domain-specific interestingness criteria.

On-Premises Deployment Option
Developing a self-hosted deployment option would address data privacy concerns for organizations handling sensitive information. Containerized deployment using Docker and Kubernetes could facilitate installation and scaling in private cloud or on-premises environments.

Domain-Specific Customization
Creating domain-specific versions of the system tailored to fields such as healthcare, finance, marketing, or scientific research could improve recommendation accuracy and relevance. Such customization could include specialized chart types, domain- specific preprocessing logic, and training data reflecting field-specific best practices.

VIII. CONCLUSION

This paper has presented an AI-powered data visualization automation tool that addresses critical limitations in existing systems while delivering substantial practical benefits for data analysis workflows. The proposed system integrates robust preprocessing pipelines, intelligent chart recommendation using machine learning, and automated insight generation within a unified architecture that supports multiple data formats and balances automation with user control. Experimental evaluation demonstrates that the system achieves approximately 70% reduction in analysis time compared to manual workflows while maintaining 87% accuracy in chart type recommendations.

The comprehensive preprocessing pipeline successfully handles missing values, outliers, and format variations, addressing a key gap in existing automated visualization tools. User studies with diverse participant populations confirm strong usability (SUS score of 82.3) and validate the system's potential to democratize data analysis by enabling non-expert users to generate professional-quality visualizations. The research contributes to the growing body of work on human-AI collaborative systems for data analytics by demonstrating that feature-based machine learning approaches can effectively encode visualization best practices and provide explainable recommendations without requiring the scale and complexity of large language models.

The system's architecture and evaluation methodology also advance the field by providing a comprehensive assessment framework that spans technical performance, user experience, and practical utility across diverse datasets and use cases. From a practical perspective, the system offers significant value across multiple application domains. In business intelligence contexts, the tool enables rapid generation of visualizations for decision support, reducing the time from data collection to actionable insights.

Academic researchers can leverage the system for exploratory analysis of new datasets, quickly identifying patterns and relationships that warrant deeper investigation. Organizations seeking to democratize data literacy can deploy the system to empower non-technical stakeholders to independently explore data and generate insights without relying on specialized analysts. The identified limitations including chart type coverage, advanced statistical capabilities, and performance with very large datasets provide clear directions for future development.

The proposed enhancements, particularly predictive analytics integration, real-time dashboard generation, and natural language interfaces, represent promising avenues for extending the system's capabilities and broadening its applicability. As data continues to proliferate across all sectors of society, tools that reduce barriers to effective visualization and insight generation become increasingly critical.

The proposed AI-powered data visualization automation tool represents a significant step toward making sophisticated data analysis accessible to broader audiences while maintaining the quality and rigor required for sound decision-making. Future work should continue to address the challenges of tabular comprehension, explainability, and end-to-end automation while developing standardized benchmarks that enable systematic comparison and advancement of automated visualization technologies.

REFERENCES

1. Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, and D. Zhang, “HAIChart: Human and AI paired visualization system,” IEEE Trans. Vis. Comput. Graph., vol. 30, no. 1, pp. 831–841, 2024, doi: 10.1109/TVCG.2023.3326522.
2. C. Tang, S. Zhao, C. Li, N. Cao, and H. Qu, “ASTA: Automatic spreadsheet transformation and analysis with explainable AI,” ACM Trans. Interact. Intell. Syst., vol. 13, no. 2, pp. 1–28, 2023, doi: 10.1145/3577011.
3. B. Lee, A. Srinivasan, P. Eichmann, and J. Stasko, “SpotLight: Detecting anomalies in streaming graphs,” in Proc. CHI Conf. Human Factors Comput. Syst., Hamburg, Germany, 2023, pp. 1–14, doi: 10.1145/3544548.3581256.
4. L. Yuan, Y. Chen, G. Wang, X. Wang, C. Zhang, and Y. Chen, “Hierarchical prompt learning for multi-task learning on tabular data,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Singapore, 2023, pp. 4521–4533, doi: 10.18653/v1/2023.emnlp-main.276.
5. C. Chen, J. Wu, X. Wang, S. Xie, H. Zhang, Z. Liu, J. Li, and N. Tang, “Infogen: Automatic infographic generation from text,” IEEE Trans. Vis. Comput. Graph., vol. 29, no. 12, pp. 5521–5533, 2023, doi: 10.1109/TVCG.2022.3209469.
6. R. Kumar, S. Sharma, and A. Patel, “AI-powered data visualization platform for automated insights,” Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, pp. 245–256, 2023, doi: 10.14569/IJACSA.2023.0140328.
7. M. Zhang, L. Wang, and H. Liu, “Intelligent business intelligence dashboards with predictive analytics and natural language generation,” Expert Syst. Appl., vol. 213, Art. no. 118912, 2023, doi: 10.1016/j.eswa.2022.118912.
8. S. Zaman, M. A. Islam, S. Haque, and H. A. Mohna, “A review of AI-powered data visualization in enterprise reporting: Dashboard design and interactive analytics,” American Journal of Advanced Technology and Engineering Solutions, vol. 2, no. 1, pp. 32–54, 2022.
9. Y. Zhang and X. Ma, “Visual analytics for decision support,” IEEE Computer Graphics and Applications, vol. 32, no. 5, pp. 14–21, 2012.
10. B. Shneiderman, “The eyes have it: A task by data type taxonomy,” IEEE Symposium on Visual Languages, pp. 336–343, 1996.
11. D. Bumblauskas, H. Nold, P. Bumblauskas, and A. Igou, “Big data analytics: Transforming data to action,” Business Process Management Journal, vol. 23, no. 3, pp. 703–720, 2017.
12. I. Canhoto and F. Clear, “Artificial intelligence and machine learning as business tools,” Business Horizons, vol. 63, no. 2, pp. 183–193, 2020.
13. C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
14. C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802065

15. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
16. S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
17. S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
18. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
19. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
20. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- *Springer, Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
21. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
22. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
23. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
24. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
25. Dang et al., "Human-centered AI for interactive data visualization," IEEE Computer Graphics and Applications, vol. 40, no. 3, pp. 76–89, 2020.
26. Holzinger, "Human-centered AI," Computer, vol. 52, no. 9, pp. 12–19, 2019.
27. F. Hohman et al., "Visual analytics in deep learning," IEEE Computer Graphics and Applications, vol. 38, no. 6, pp. 84–96, 2018.
28. Perkhofer, M. Hofer, and J. M. Ja¨ger, "Cognitive load reduction in dashboard design," Information Systems and e-Business Management, vol. 17, no. 4, pp. 737–760, 2019.
29. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-Lite: A grammar of interactive graphics," IEEE Trans. Visualization and Computer Graphics, vol. 23, no. 1, pp. 341–350, 2017.