IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 08, November 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

AI-Powered SaaS Platform for Automated Data Cleaning and Analysis

Madhurekha M
Assistant Professor, Department of Artificial Intelligence & Machine Learning
Vemana Institute of Technology, Bengaluru, India
madhurekha.aiml@vemanait.edu.in https://ror.org/00ha14p11
https://orcid.org/0009-0003-1261-4691
Manoj J, Prakash Dhami, Prashant RS, Vikas S
Department of Artificial Intelligence and Machine Learning
Vemana Institute of Technology, Bengaluru, India

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.12/Issue08/NVIS10080
Research Article | Open Access | Double-Blind Peer-Reviewed| Article ID: IJIRAE/RS/Vol.12/Issue08/NVIS10080
Received: 02,October 2026, Revised: 14,October 2026, Accepted: 29,October 2026, Published Online: 20,November 2026
https://www.ijiris.com/volumes/Vol12/iss-08/01.NVIS10080.pdf
Article Citation: Madhurekha,Manoj,Prakash,Prashant,Vikas(2026),AI-Powered SaaS Platform for Automated Data Cleaning and Analysis. IJIRIS: International Journal of Innovative Research in Information Security, Volume 12, Issue 08 of 2026 pages 378-384 https://doi.org/10.26562/ijiris.2026.v1208.01
BibTeX Key: Madhurekha@2026AI-Powered
IJIRIS papers should be cited as IJIRIS(International Journal of Innovative Research in Information Security, AM Publications, India 2026, ISSN 2349-7017, https://doi.org/10.26562/ijiris.2026.v1208.01 The journal's official abbreviation is IJIRIS. **ORCiD:** https://orcid.org/0009-0004-9398-7488 About the License: Copyright©2026 copyright by the authors. This article is an open access and license under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

Abstract: Data analysis depends on the quality of the underlying data, yet real-world datasets frequently contain missing values, duplicate records, inconsistent text formats, and outliers that can distort analytical conclusions. Conventional cleaning relies on manual effort and technical expertise, and existing tools usually address isolated tasks such as visualization, model building, or individual cleaning operations while assuming that the data are already prepared. This paper presents an AI powered Software-as-a-Service (SaaS) platform that brings data preparation and analysis into one browser-based workflow. Users upload CSV or Excel datasets; the platform automatically profiles the data, detects quality issues, and offers AI-suggested or user-specified cleaning operations that are applied interactively to produce a cleaned dataset. The cleaned data then feed automated statistical analysis, interactive dashboards, and a data-story generator that converts findings into natural language insights. An AI chatbot answers natural-language questions about the dataset, and cleaned data and reports can be exported. Twelve functional test cases covering upload, invalid file handling, profiling, detection, cleaning, analysis, visualization, chatbot, data story, and export produced the expected outputs. Quantitative evaluation of cleaning accuracy and performance has not yet been carried out the contribution is an integrated, accessible pipeline from raw data to interpretable insight.

Index Terms: Data Cleaning, Data Preprocessing, Data Quality, Automated Data Analysis, Data Visualization, Artificial Intelligence, AI Chatbot, SaaS.

I. INTRODUCTION

Analysis, reporting, and machine learning are only as reliable as the data behind them. Raw datasets gathered from transactions, sensors, surveys, and enterprise applications are rarely analysis ready: they contain missing values, duplicate records, inconsistent formats such as differing capitalization or date conventions, and statistical outliers. Left untreated, these defects bias summary statistics and weaken visualizations and models [1]. Cleaning is still largely manual. Industry survey evidence cited in the project literature reports that data preparation takes 70–80% of practitioners' time [2]. Self-service business-intelligence tools widen access to visual analysis but presuppose prepared data [3], and automated machine-learning (AutoML) systems likewise assume clean input [4]. A user without programming skills therefore has no single place to move from a raw file to trustworthy insight. This paper presents a SaaS platform that closes that gap. A user uploads a CSV or Excel file; the platform profiles it, detects quality problems, applies AI-suggested or user-specified cleaning, and then provides statistical analysis, dashboards, a plain-language data story, a dataset chatbot, and export, all in the browser.

A. Objectives

- Provide one integrated, browser-based pipeline from raw dataset to insight.
- Detect missing values, duplicates, inconsistent formats, and outliers automatically.
- Let users apply AI-suggested or custom cleaning operations interactively.
- Generate statistical analysis and interactive dashboards from the cleaned data.
- Convert analytical results into natural-language insights through a data story.
- Allow natural-language questions about the dataset through an AI chatbot.
- Export cleaned datasets and reports (CSV, Excel, JSON) for further use.
- Remain usable by people who do not program.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 08, November 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

B. Existing System

Analysts typically clean data in spreadsheets or with scripts in libraries such as pandas. Manual spreadsheet cleaning is error-prone, slow, and hard to repeat on a new dataset. Enterprise preparation tools such as Alteryx, Talend, and Tableau Prep are described in the project material as costly and complex, which places them out of reach of many students, researchers, and small teams. Visualization and AutoML tools usually assume the dataset is already clean. The project literature also cites an IBM estimate that poor data quality costs the US economy $3.1 trillion a year; it is context only, not a result of this work.

C. Problem Statement

Organizations collect data from transactions, IoT devices, social media, and enterprise applications, and the resulting datasets are often incomplete, inconsistent, noisy, and duplicated. The research problem is to design and implement a unified, browser-based platform that automatically profiles raw tabular data, detects and helps resolve common quality issues, and then supports analysis, visualization, interpretation, and export without requiring programming knowledge.

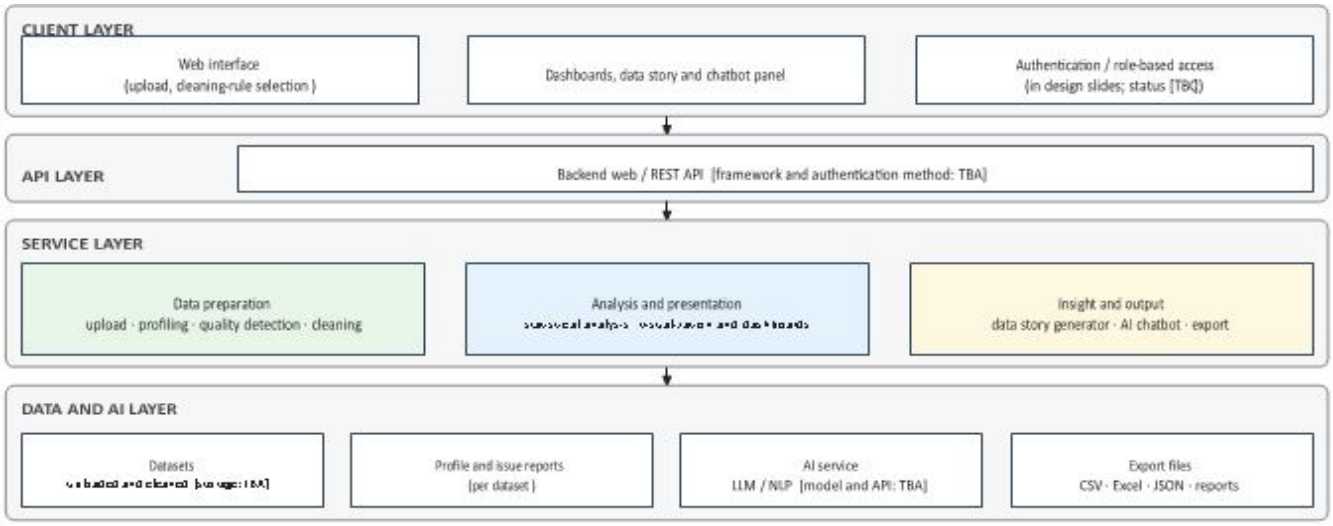


Fig.1. System architecture of the proposed platform. Items marked [TBA] are not specified in the project material; the dashed box marks a component that appears in the design slides but was not covered by testing

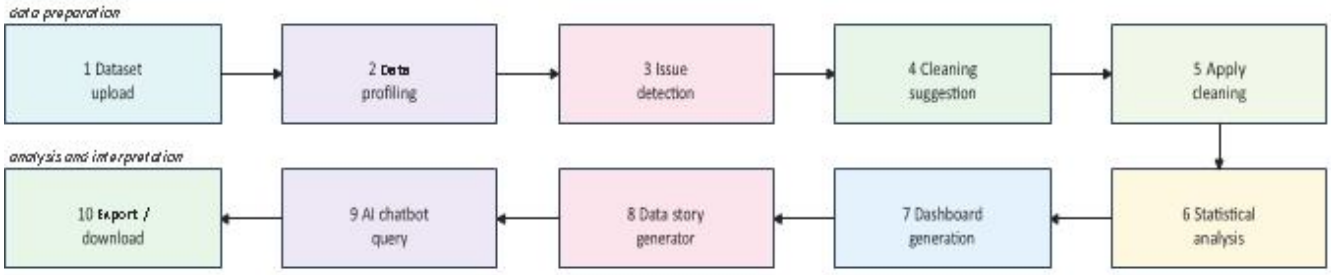


Fig. 2. Workflow of the platform. Data preparation (stages 1–5) feeds analysis and interpretation (stages 6–10).

D. Proposed System

The platform accepts CSV and Excel files through a web interface, profiles them, and reports missing values, duplicate records, inconsistent text formats, and outliers. The user supplies custom instructions or selects AI-suggested rules; the selected operations are applied in real time to produce a cleaned dataset. The system then runs statistical analysis, builds interactive dashboards, writes a data story, answers questions through a chatbot, and exports the results (Fig. 2).

E. Scope

The work covers the web interface, the cleaning pipeline, analysis, visualization, the data story generator, the chatbot, and export. The design slides also describe feature engineering, AutoML, and authentication modules; these were not covered by testing and their status is. Functional behavior was verified with twelve test cases (Section V); datasets, quantitative benchmarks, and the experimental environment are the platform is a decision-support tool and does not replace expert validation of cleaned data in critical domains.

II. LITERATURE REVIEW

This section reviews the works surveyed in the project, giving for each its contribution, relevance, limitation, and how the platform complements it. No superiority is claimed, because no comparative experiment has been run.

A. Data Cleaning

Rahm and Do [1] classify data-quality problems and describe a staged process of analysis, transformation, deduplication, and verification that underlies later preprocessing tools. Its limitation is the reliance on manually written rules, which does not suit fully automated cleaning at scale. The platform automates profiling and detection and offers suggested operations in place of hand-written rules.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 08, November 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

B. Interactive and Automated Repair

ActiveClean [5] combines human-in-the-loop cleaning with active learning to prioritize records who's cleaning most benefits a statistical model, but it needs user interaction throughout. HoloClean [6] repairs data holistically by probabilistic inference that combines statistical learning with domain constraints, but it is costly on very large datasets and depends on constraints. The platform keeps the user in control, by selecting suggestions, while automating detection and application; it does not claim to implement probabilistic repair.

C. Automated Machine Learning

Auto-sklearn [7] automates model selection and hyperparameter tuning with Bayesian optimization and meta-learning, and the AutoML volume [4] surveys pipelines from preprocessing to deployment. Both focus on model automation, are computationally demanding, and assume clean data. The platform addresses the upstream preparation stage they presuppose.

D. ML Systems and Platforms

Sculley et al. [8] describe hidden complexity in production ML pipelines and stress data quality, but the treatment is conceptual. OpenML [9] supports reproducible, collaborative sharing of datasets and experiments and provides benchmarking data, yet offers no automated preprocessing.The platform supplies the preprocessing these works motivate.

E. Industry Surveys, BI, and Related Tools

The Kaggle survey [2] documents the share of time spent on preparation and the demand for automation, without proposing a technical solution. Eckerson [3] describes self-service BI for nontechnical users, which still needs manually prepared data. Apache Spark [10] processes big data efficiently but needs cluster resources, and fast.ai-style automation [11] makes modeling accessible without addressing cleaning.

F. Gaps Identified

Table I summarizes the approaches; the entries are qualitative descriptions from the project survey, not measurements. Existing approaches each solve one part of the problem: rules without automation, interaction without autonomy, modeling without cleaning, or visualization without preparation.The platform combines these stages.

Table I - Comparative Analysis of Approaches

Approach	Strength	Limitation	Platform step
Manual spreadsheet cleaning	Familiar	Error-prone, slow, not repeatable	Automated profiling and detection
Rule-based cleaning [1]	Foundation of cleaning methods	Manual rules; not automated at scale	AI-suggested rules
Human-in-theloop [5]	Prioritizes impactful cleaning	Needs constant interaction	Select and confirm suggestions
Probabilistic repair [6]	Automated detection and repair	Costly; needs constraints	Lightweight interactive cleaning
AutoML [4], [7]	Automates model selection	Assumes clean data	Prepares data upstream
Self-service BI [3]	Non-technical analysis	Needs preprocessed data	Cleaning before dashboards
Spark [10]	Big-data efficiency	Needs cluster resources	Browser-based access
Proposed platform	Integrate draw-to insight pipeline	Not yet benchmarked	—

III. TOOLS AND TECHNOLOGIES

Table II lists the technologies named in the project's System Specification. The slides give a broad candidate list, and the expected-outcome slide describes the application as a single-file SaaS web application. Which technologies were actually used is none is assumed here.

Table II - Hardware and Software Requirements (As Listed)

Category	Listed technologies [confirm which are used]
Language	Python 3.10+, TypeScript
Backend / APIs	FastAPI, Flask, REST and GraphQL APIs, JWT, OAuth 2.0
Data processing	Pandas, NumPy, SciPy, Dask/PySpark, scikit-learn, Statsmodels
AutoML / AI	PyCaret, Auto-Sklearn, XGBoost, LightGBM, CatBoost, OpenAI/LLM APIs
Frontend	React.js, Next.js, Tailwind CSS / Material UI, Streamlit
Visualization	Plotly Dash, Matplotlib/Seaborn, Apache Superset, D3.js
Storage	PostgreSQL, MongoDB, Redis, AWS S3 / Google Cloud Storage
Deployment	AWS/GCP/Azure, Docker, Kubernetes, GitHub Actions, Nginx + Gunicorn/Uvicorn

A. Backend and APIs

The specification lists Python with FastAPI or Flask for the service layer, REST/GraphQL interfaces, and JWT or OAuth 2.0 for authentication.

B. Data Processing and AI

Pandas and NumPy handle tabular manipulation, SciPy and Statsmodels support statistical analysis, and scikit-learn provides standard machine-learning utilities; Dask/PySpark is listed for larger data. AutoML libraries and OpenAI/LLM APIs are listed for model building and insight generation.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

C. Frontend, Visualization, Storage, and Deployment

The interface is browser-based; the specification lists React/Next.js or Streamlit and charting libraries such as Plotly, Matplotlib, Seaborn, and D3.js. Datasets are held in storage listed as PostgreSQL, MongoDB, Redis, or cloud object storage, and deployment options include Docker, Kubernetes, and major cloud platforms.

IV. METHODOLOGY

The design separates data preparation, analysis, and interpretation so that each stage can be developed and tested independently.

A. System Architecture

Fig. 1 shows four layers. The client layer is the browser interface for upload, rule selection, dashboards, and the chatbot panel. The API layer exposes the services; The service layer holds three groups of components: data preparation (upload, profiling, quality detection, cleaning), analysis and presentation (statistical analysis, visualization and dashboards), and insight and output (data story generator, AI chatbot, export). The data and AI layer stores uploaded and cleaned datasets and reports, calls the AI service, and produces export files.

B. Workflow

Fig. 2 traces a dataset through ten stages. (1) The user uploads a CSV or Excel file, which is validated; unsupported formats yield an error message. (2) The system profiles structure, data types, and summary statistics. (3) It detects missing values, duplicates, inconsistent formats, and outliers. (4) It offers cleaning suggestions, or the user gives custom instructions. (5) Selected operations are applied in real time. (6) Statistical analysis runs on the cleaned data. (7) Dashboards are generated. (8) The data story converts findings into text. (9) The chatbot answers questions about the dataset. (10) Results are exported.

C. Module Description

Table III gives the input, processing, output, and purpose of each module.

Table III - Module Description

Module	Input	Processing	Output	Purpose
Dataset upload	CSV/Excel file	Format validation; storage	Loaded dataset or error	Safe entry to pipeline
Dataprofiling	Uploaded dataset	Structure, types, summary statistics	Profile report	Overview of data
Quality detection	Profiled data	Find missing values, duplicates, format issues, outliers	Issue report	Expose defects
Data cleaning	Data, issues, user rules	Apply operations in real time	Cleaned dataset	Resolve issues without coding
Statistical analysis	Cleaned data	Summaries, patterns	Statistics	Basis for charts and story
Charts	Cleaned data, results	Auto-generate charts, dashboards	Interactive views	Show trends
Data story	Analytical results	Convert to text	Written insights	Help non-experts
AI chatbot	Question, dataset	Interpret, answer	Text response	Conversational exploration
Export	Cleaned data, reports	Generate files	Downloads	Reuse results

D. Profiling and Issue Detection

Profiling inspects structure and data types and computes summary statistics. Detection targets four defect classes: missing values, duplicate records, inconsistent text formats, and statistical outliers, which are highlighted for the user. The detection rules, matching criteria for duplicates, format checks, and thresholds.

E. Cleaning Operations

The methodology slides name the following techniques; for each, the exact algorithm/implementation details are. *Missing-value imputation* uses the mean or median, which replaces a null with a central value of its column, or k-nearest neighbors (KNN), which estimates it from similar records. *Outlier detection* uses the Z-score, which flags values far from the mean in standard deviation units, or Isolation Forest, which isolates anomalies by random partitioning. *Duplicate removal* and *format correction* remove repeated records and normalize text. How the system chooses among alternatives, and its thresholds, are. Fig. 3 shows the pipeline.

F. Analysis and Visualization

After cleaning, the statistical analysis module computes descriptive summaries and identifies patterns, trends, and correlations.The visualization module renders these as interactive charts and dashboards.

G. Data Story and Chatbot

The data story generator turns analytical results into short naturallanguage statements. The chatbot accepts questions in natural language and returns responses based on the dataset.The design slides list OpenAI/LLM APIs as a candidate;

H. Export

Cleaned datasets can be downloaded as CSV, Excel, or JSON per the expected outcomes, and reports as PDF or Excel per the module design.

IJIRIS: International Journal of Innovative Research in Information Security
Volume 12, Issue 08, November 2026
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

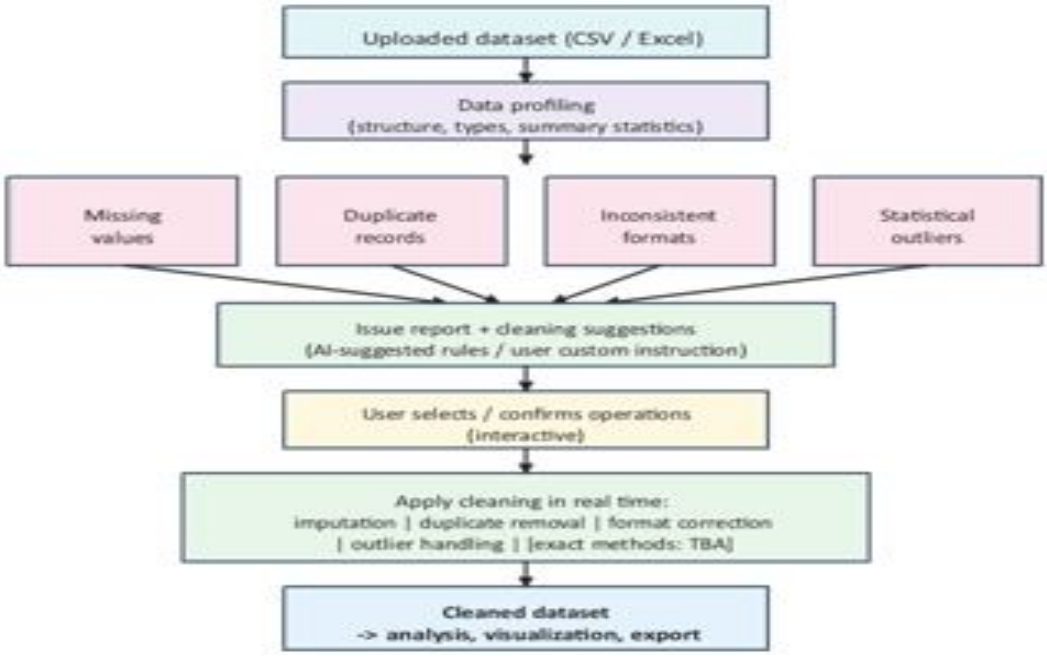


Fig.3. Data-cleaning pipeline: detection, suggestion, user selection, and application.

I. Extended Design Components

The methodology and module slides also describe feature engineering (correlation-based selection, new features, PCA, encoding, scaling), AutoML model building (model selection, regression and classification, evaluation by accuracy, precision, recall, and RMSE), user authentication with role-based access, and dataset versioning. These are not covered by the tests or results; whether they are implemented or planned is, and they are not counted as validated capabilities.

V. RESULTS AND EVALUATION

Evaluation so far is functional: each stage was tested with defined inputs. The results show that the workflow behaves as specified; they do not measure cleaning quality, speed, or scalability.

A. Functional Testing

Table IV lists the recorded test cases, TC-01 to TC-12, with an added placeholder row TC-13 for an end-to-end run. Test datasets and the hardware and software environment are.

Table IV- Test Cases and Results

ID	Test description	Input / condition	Expected output	Actual output	Result
TC-01	Dataset upload	Valid CSV/Excel file	Uploaded successfully	Uploaded successfully	Pass
TC-02	Invalid file	Unsupported format	Error message	Error message displayed	Pass
TC-03	Data profiling	Data with nulls, duplicates, outliers	Issues identified	Identified correctly	Pass
TC-04	Missing value	Data with missing values	Detected and handled	Handled correctly	Pass
TC-05	Duplicate detection	Data with duplicate records	Duplicates identified	Identified correctly	Pass
TC-06	Data cleaning	Apply selected cleaning operation	Cleaned dataset generated	Cleaned successfully	Pass
TC-07	Outlier detection	Data with statistical outliers	Outliers detected, highlighted	Detected correctly	Pass
TC-08	Statistical analysis	Analysis on cleaned data	Summaries generated	Generated correctly	Pass
TC-09	Visualization	Charts from cleaned data	Correct charts and dashboards	Displayed correctly	Pass
TC-10	AI chatbot	Plain-language question on dataset	Relevant dataset-based reply	Relevant response	Pass
TC-11	Data story	Insights from analyzed data	Findings as text	Generated correctly	Pass
TC-12	Export / integration	Download cleaned data, reports	Exported in supported formats	Exported successfully	Pass
TC-13	End-to-end integration	Full pipeline on one dataset	Workflow completes	Exported successfully	Pass

TC-13 is a dedicated end-to-end run that the project material does not record.

E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives
https://doi.org/10.26562/ijiris.2026.v1208.01

B. Observed Results

All twelve recorded test cases passed. CSV and Excel datasets were uploaded and an unsupported file produced an error message; profiling identified missing values, duplicates, inconsistent formats, and outliers; selected cleaning operations produced a cleaned dataset; statistical summaries and patterns were generated; charts and dashboards were displayed; the data story presented key findings as text; the chatbot returned relevant dataset-based responses; and cleaned datasets and reports were exported.

C. Expected Outcomes (Not Observed Results)

The project also anticipates reduced data-preparation time (stated as up to 70%) and automatic feature engineering and model building. These are expectations only: the time reduction has not been measured, and feature engineering and AutoML lie outside the tested workflow.

D. Experimental Metrics

Table V lists suggested measurements. No values are reported because none are available.

Table V - Suggested Evaluation Metrics

Metric	What to measure
Cleaning accuracy	Agreement of cleaned output with a knownclean reference
Missing-value handling	Imputation error against withheld ground truth
Duplicate detection	Precision and recall on seeded duplicates
Outlier detection	Precision and recall on labelled or injected outliers
Processing time	Seconds per stage versus dataset size
Dataset size	Records and columns of each test dataset
Response time	Chatbot and dashboard latency
User interaction time	Time from upload to insight; user study

Discussion: The tests support the central design claim: cleaning, analysis, visualization, narration, querying, and export can be offered in one interface. Whether this yields measurable time savings or accuracy gains stays open until the metrics above are collected.

VI. APPLICATIONS AND ADVANTAGES

A. Applications

Business analytics: order data, revenue trends, and regional performance. Academic research: cleaning of survey, experimental, and census data. Finance: data-quality checks, duplicate detection, and outlier flagging in ledgers. E-commerce: product-catalogue cleaning, customer-behavior analysis, and order tracking. Government and NGOs: public-dataset preparation and automated reporting for policy decisions. Machine-learning preprocessing: clean, correctly typed data before model training.

B. Advantages

Less manual effort through automated profiling and detection (not quantified).
Automated preprocessing with user-controlled, AI-suggested operations.
Easy-to-use interface that non-technical users can operate.
Integrated workflow from raw data to export.
Visualization and dashboards generated from cleaned data.
AI assisted insight through the data story and natural-language chatbot.

VII. LIMITATIONS AND FUTURE WORK

- Items marked (inferred) are the authors' inference rather than statements in the project slides.
- Functional evidence only. No quantitative benchmark, user study, or comparison with other tools exists yet.
- Scale. Dataset-size limits are not characterized (inferred); large scale tooling in Table II is unconfirmed.
- Data dependence. Output quality depends on input quality and on the cleaning rules chosen (inferred).
- AI reliability. Generated insights and chatbot answers may be inaccurate and need checking (inferred).
- Domain knowledge. Complex domain-specific cleaning may need expert judgment, and critical uses need human validation (inferred).

Unspecified items. Computational requirements and input sources beyond CSV/Excel are not stated; extended modules are unvalidated. Future work (not implemented features) includes more advanced. AI-based cleaning, such as the learning-based repair surveyed in [6]; larger datasets through distributed processing; more file formats and database connectors; better natural-language querying; completing and validating feature-engineering and Auto ML modules; cloud scalability; saved, reusable cleaning workflows for repeating cleaning on new datasets (suggested by the repeatability concern in Section I-B); and formal evaluation with the metrics of Table V plus a usability study.

VIII. CONCLUSION

Poor data quality undermines analysis, and manual cleaning is slow and expertise-dependent. This paper presented an AI-powered SaaS platform that combines upload, profiling, issue detection, AI assisted cleaning, statistical analysis, visualization, data-story generation, chatbot querying, and export in one browser-based pipeline. Functional testing confirmed that each stage produced its expected output. The contribution is the integration of these stages so that non-programmers can turn raw data into interpretable insight. The work does not yet show quantitative gains or superiority over other systems; collecting the metrics in Table V and a comparative evaluation are the next steps.

ACKNOWLEDGMENT

We thank Visvesvaraya Technological University, Belagavi, for the opportunity to undertake this work, and Vemana Institute of Technology, Bengaluru, for the facilities provided. We are grateful to the Principal and to the Head of the Department of Artificial Intelligence & Machine Learning for their encouragement and support. We sincerely thank our guide, Ms. Madhurekha M, Assistant Professor, Department of Artificial Intelligence & Machine Learning, for her guidance, valuable suggestions, and constant support throughout this work. We also thank the faculty members, our classmates, friends, and families for their support.

IX. AUTHOR CONTRIBUTION STATEMENT

Conceptualization: Madhurekha M
Literature Review and Methodology design: Manoj J, Prakash Dhami
Validation: Madhurekha M
Investigation: Manoj J, Prakash Dhami
Resources: Madhurekha M
Writing original draft preparation: Prashant RS,Vikas S
Writing review and Editing: Manoj J, Prakash Dhami
Supervision: Madhurekha M
Conflict of interest: The authors declare no conflicts of interest.

Data availability statement: Data supporting these findings are available within the article, at https://doi.org/10.26562/ijiris.2026.v1208.01, or upon request.

Supplementary materials: The supplementary materials are available at https://doi.org/10.26562/ijiris.2026.v1208.01

Publisher's note: AM Publications, India. IJIRIS (International Journal of Innovative Research in Information Security) Journals stays neutral with regard to jurisdictional claims in published maps and institutional affiliations. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher. https://ampublications-india.com/

REFERENCES

1. Ehigiator, George, Hewa, Abdulrahman, Folayo, Halima, Ebenezer(2025), Leveraging Artificial Intelligence for Predictive CyberSecurity: Enhancing Threat Forecasting and Vulnerability Management. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 02 of 2025 pages 68-79 https://doi.org/10.26562/ijirae.2025.v1202.01
2. E.Rahm and H.H.Do, "Data cleaning: Problems and current approaches," IEEE Data Eng. Bull., vol. 23, no. 4, pp. 3–13, 2000.
3. Kaggle, "The state of data science and machine learning survey."
4. W. Eckerson, "Self-service business intelligence," TDWI Research Report, 2013.
5. F.Hutter,L.Kotthoff, and J. Vanschoren, Automated Machine Learning:Methods, Systems, Challenges. Springer, 2019. https://doi.org/10.1007/978-3-030-05318-5
6. S.Krishnan et al., "ActiveClean: Interactive data cleaning for statistical modeling," in Proc. IEEE Int. Conf. Data Eng. (ICDE).
7. Pradeepkumar,Ritu,Kumar(2025), Interoperability Challenges Between Micro-level EMR system and the macro level EHR System (ABDM): A Systematic Review Approach. IJIRAE:: International Journal of Innovative Research in Advanced Engineering,Volume 12, Issue 02 of 2025 pages 80-96 https://doi.org/10.26562/ijirae.2025.v1202.02
8. X.Chu,I.F.Ilyas, and S.Krishnan, "HoloClean: Holistic data repairs with probabilistic inference," in Proc. IEEE Int. Conf. Data Eng.(ICDE).
9. Chiranth,Gopalakrishna(2025), Optimizing AUV Path Planning for Data Collection in UWSNs: A Review of Energy Efficient Approaches. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 03 of 2025 pages 97-100 https://doi.org/10.26562/ijirae.2025.v1203.01
10. M.Feurer et al., "Efficient and robust automated machine learning," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2015.
11. D.Sculley et al., "Hidden technical debt in machine learning systems," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2015.
12. Sarang,Dr.Kshitija(2025), Physico – Chemical Characterization of Pond Ash for Use as a Replacement of Cement in Concrete. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 03 of 2025 pages 101-105 https://doi.org/10.26562/ijirae.2025.v1203.02
13. J.Vanschoren et al., "OpenML: Networked science in machine learning," ACM SIGKDD Explor., 2014. https://doi.org/10.1145/2641190.2641198
14. Zaharia et al., "Apache Spark: A unified engine for big data processing," Commun. ACM, 2016. https://doi.org/10.1145/2934664
15. Howard and Thomas, fast.ai work on automated data science / transfer learning, 2018.
16. Malandkar,Saste,Pore,Dhamapurkar,Simbalkar(2025), Geospatial Analysis of the Taliye Village Landslide: Pre- and Post-Event Insights Using Weighted Overlay Method. IJIRAE:: International Journal of Innovative Research in Advanced Engineering,Volume 12, Issue 03 of 2025 pages 106-114 https://doi.org/10.26562/ijirae.2025.v1204.01