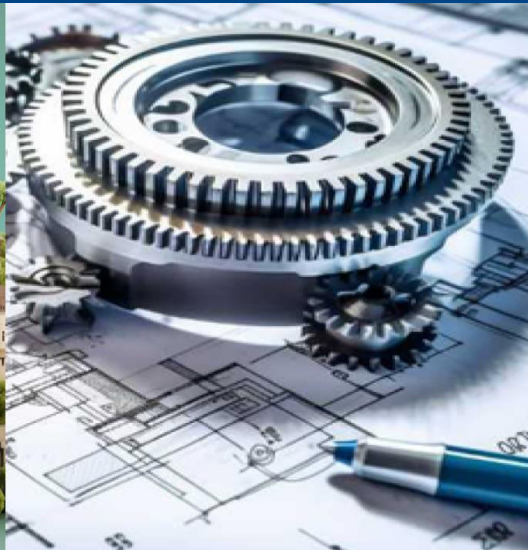
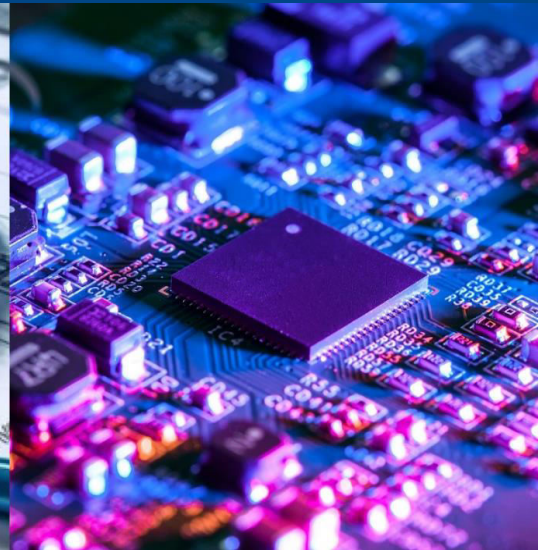
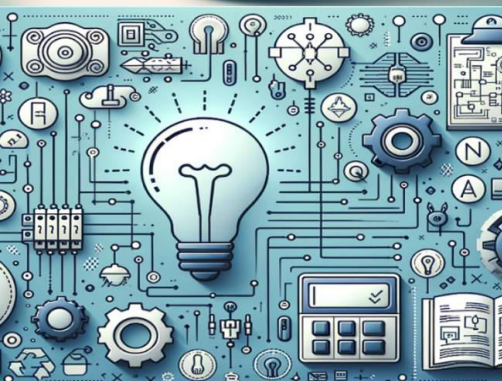




International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 9.864

Volume 9, Issue 10, October 2026



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

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

INSIGHTAI: An Autonomous Data Intelligence Command Framework for AI-Assisted Analysis and Visualization of Tabular Data

Mr. Sahil Firoj Mulani¹, Mr. Nilesh Tanaji Sabale¹, Mr. Rupesh Kashinath Mhetre¹,

Mr. Abhishek Prabhakar Ponnamm¹, Prof. S. S. Bhuite²

Department of Computer Science and Engineering, Brahmdevdada Mane Institute of Technology, Solapur,

Maharashtra, India¹

Research Guide, Department of Computer Science and Engineering, Brahmdevdada Mane Institute of Technology,

Solapur, Maharashtra, India²

ABSTRACT: Extracting reliable information from a spreadsheet or CSV file normally needs separate skills in data cleaning, statistics, charting and programming, which students, small organisations and other non-specialist users often lack. Several tools must be combined, and the results still have to be interpreted by the user. This paper presents InsightAI, an AI-assisted web application that brings file upload, cleaning, statistical summary, visualization, large-language-model (LLM) interpretation and report export into one workflow. A user uploads a CSV or Excel file; the system validates and cleans it, computes descriptive statistics and patterns, draws interactive charts, and produces plain-language strategic insights, an action roadmap and a risk list, with a natural-language query box for follow-up questions. Users sign in with JWT-based authentication, and datasets and reports are stored per user in PostgreSQL. The system is built with Next.js, TypeScript, Tailwind CSS, Prisma, PostgreSQL and the OpenAI API. It was demonstrated on a 100-record employee dataset and produced a complete report, including summary indicators, trend and segment charts and written insights. The contribution is the integration of established techniques into one usable platform, not a new algorithm; formal accuracy and performance testing remains future work.

KEYWORDS: Exploratory Data Analysis, Large Language Model, Automated Insight Generation, Data Visualization, Next.js, PostgreSQL, Prisma, JWT Authentication.

I. INTRODUCTION

Data analysis supports decision-making in education, business and engineering. Tukey's exploratory data analysis (EDA) established the practice of examining data through summaries and graphics before any formal modelling [1]. In practice, much of the effort goes into preparing data rather than analysing it: Wickham notes that a very large share of effort goes into cleaning data [2], and the interview study of Kandel et al. documents the recurring difficulties analysts face with analysis and visualization tools [3].

A person who holds a CSV or Excel file but has little programming or statistics training must usually load and clean the data in one tool, compute summaries in another, draw charts in a third, and then interpret the results alone. General-purpose AI chatbots do not remove this problem, because they do not see the user's file and can produce statements that are not supported by the data.

InsightAI addresses this by placing the whole path from raw file to understandable result in one web application. The user uploads a file; the system cleans it, computes the statistics, draws the charts, and asks a language model to explain the computed findings in plain language. The same page lets the user ask follow-up questions and export the result. Each user signs in, and uploaded datasets and generated reports are stored for later use.



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

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

1.1 Problem Statement

Large amounts of data are produced in files such as CSV and Excel, but extracting insight from them needs technical skill. Manual cleaning is error-prone, several separate tools are needed for parsing, analysis and charting, and non-technical users find statistical output hard to interpret because tools rarely explain results in plain language. There is a need for a single, secure, easy-to-use platform that converts an uploaded dataset into verified statistics, charts and readable insights without requiring programming knowledge.

1.2 Objectives

- Accept tabular data in CSV and Excel formats through a simple upload interface.
- Automatically detect and handle missing values, duplicate records and inconsistent column types.
- Compute descriptive statistics, distributions and correlations as exploratory data analysis.
- Present the results as interactive charts and summary indicators on a dashboard.
- Generate plain-language insights, recommended actions and risks using an LLM, and answer follow-up questions about the dataset.
- Provide authenticated, persistent access to each user's datasets and reports, with export of results.

1.3 Scope

The system targets small and medium-sized datasets (the upload limit in the current interface is 10 MB). Large-scale distributed processing and enterprise cloud integration are outside the scope of this work.

II. LITERATURE REVIEW

Tukey's EDA [1] is the basis for the first steps of our pipeline: summarise the data and look at it before drawing conclusions. Wickham's work on tidy data [2] addresses the preparation stage by describing a consistent structure in which each variable is a column and each observation is a row, which makes later manipulation and visualization easier. Kandel et al. interviewed 35 analysts from 25 organisations and identified recurring pain points in the analysis process [3]. Together, these studies motivate tools that reduce the manual effort between a raw dataset and an interpretable result.

On the visualization side, Vega-Lite provides a high-level grammar for specifying interactive graphics [4]. NL4DV builds on such grammars: it is a Python toolkit that takes a tabular dataset and a natural-language query and returns an analytic specification, including a list of relevant Vega-Lite charts [5]. It shows that natural language can serve as an interface to visual analysis, but it is aimed at developers who build such interfaces, not at end users who need a complete application.

More recent work uses large language models [7]. LIDA treats visualization generation as a multi-stage task handled by a summarizer, a goal explorer, a visualization generator and an infographer [6]. Its design is useful for InsightAI because it shows that a compact textual summary of a dataset can stand in for the raw data when prompting a model, and that generated output needs checking. InsightAI follows the same idea: statistics are computed by the application first, and the LLM is used to explain the computed findings.

The reviewed work concentrates on parts of the problem: data tidying, visualization grammars, natural-language visualization toolkits and LLM-based chart generation. InsightAI places upload, cleaning, summarisation, visualization, LLM interpretation, chat, storage and export behind one authenticated web interface. This is an integration goal rather than a claim of scientific novelty; the review covers a small set of sources and is not a systematic survey. Table 1 compares the manual workflow with the proposed system at the design level.



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

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

Feature	Manual workflow (code / spreadsheets)	InsightAI
Programming knowledge needed	Usually yes	No
Cleaning, statistics and charts	Separate tools or libraries	One web application
Plain-language explanation of results	Written by the analyst	Generated by LLM from computed results
Follow-up questions in plain language	Not available	Query box on the report page
Stored per-user datasets and reports	Manual file management	Authenticated, stored in PostgreSQL
Export of results	Manual	Export button on report page

Table 1. Manual analysis workflow vs. InsightAI (design-level comparison)

III. METHODOLOGY

InsightAI processes each uploaded file through a fixed sequence of stages, from ingestion to AI interpretation (Fig. 1). The first stages are deterministic and computed by the application; only the final interpretation stage uses a language model.

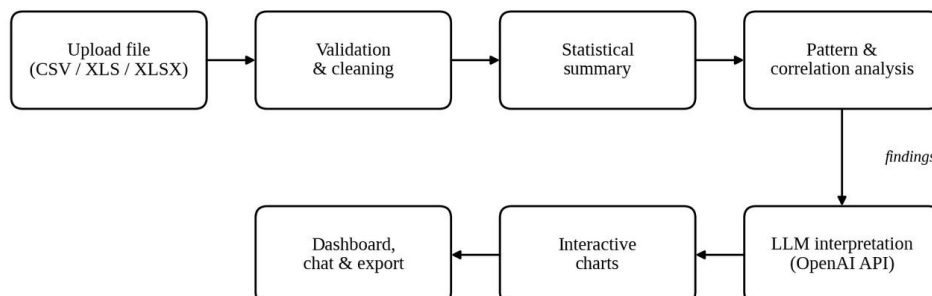


Fig. 1. Processing pipeline of InsightAI

3.1 Ingestion and Cleaning

- **Upload:** the user drags and drops or browses for a CSV, XLS or XLSX file (maximum 10 MB). The file format is detected and the upload is confirmed.
- **Validate:** the file structure, headers and empty or unsupported files are checked, and an error is shown instead of processing invalid input.
- **Clean:** missing values, duplicate records and column-type inconsistencies are detected and handled, and the data is normalised into an analysis-ready table.

3.2 Statistical Analysis and Pattern Detection

For the cleaned dataset the system computes descriptive statistics for each column: count, mean, median, standard deviation, minimum, maximum and quartiles. It then examines trends, distributions, correlations and unusual values across the columns, and shows key indicators such as total records and a data-quality indicator on the report page.

3.3 AI Interpretation and Chat

The computed findings are passed to an LLM through the OpenAI API, which returns a plain-language report made of an overview, strategic insights, an action roadmap and a risk portfolio. Because language-model output can be wrong, it



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

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

is presented as assistance to be read alongside the computed statistics and charts. A query box on the report page lets the user ask further questions about the analysis in natural language.

3.4 Authentication and Storage

Users register and sign in, and every request is tied to the signed-in user. Authentication uses JSON Web Tokens [13], and passwords are stored as bcrypt hashes [14]. Dataset records and generated reports are stored in PostgreSQL [11] through the Prisma ORM [10], so each user sees only their own datasets and reports.

IV. SYSTEM ARCHITECTURE

The system follows a three-layer architecture (Fig. 2). The presentation layer is a Next.js [8] application written in TypeScript and styled with Tailwind CSS [9]; it provides the landing page, sign-in, the dashboard with a dataset inventory, and the report page. The application layer consists of server-side API routes that perform authentication, file parsing, validation, cleaning and statistical summary, and that call the AI service. The data layer is a PostgreSQL database accessed through Prisma, which stores users, datasets and reports. The only external AI service is the OpenAI API [12], used for interpretation and chat answers.

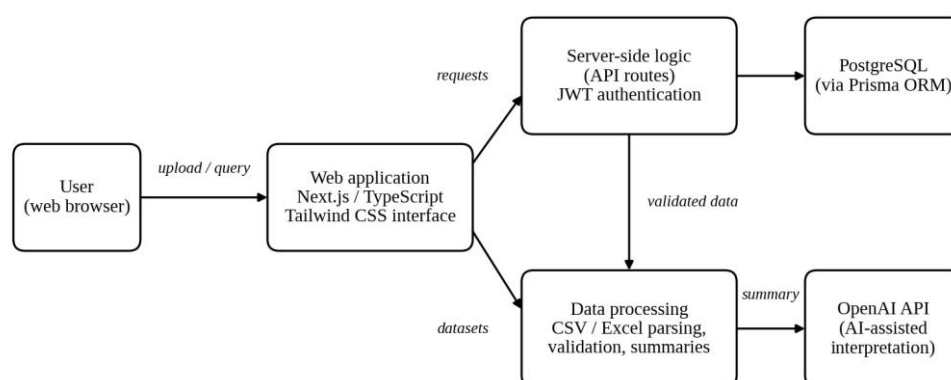


Fig. 2. Architecture of InsightAI

Splitting the code into interface, server logic and data access means each part can be changed on its own, for example by replacing the language-model provider without changing the interface or database.

V. MODULES AND FEATURES

- **User interface and dashboard:** a landing page, sign-in, and a “Mission Control” dashboard with a drag-and-drop upload panel, a searchable dataset inventory (file name, number of rows, upload date, with view and delete actions), a list of recent reports and suggested analysis goals.
- **File upload and parsing:** reads CSV, XLS and XLSX files and extracts a table of rows and columns.
- **Pre-processing:** handles missing values, duplicate rows and type inconsistencies and prepares the data for analysis.
- **Exploratory data analysis:** computes descriptive statistics, distributions and correlations.
- **Visualization:** shows interactive charts, including a performance trend chart and a segment-share chart, with summary indicator cards.
- **AI insight generation:** produces a headline summary, strategic insights, an action roadmap and a risk portfolio from the computed findings, with a stated decision goal.
- **AI chat interface:** lets the user query the analysis in plain language (for example, “Summarize the risks”).
- **Report export:** packages the visualizations and written insights into a report that the user can export.
- **Authentication and storage:** sign-in and sign-out with JWT, and per-user storage of datasets and reports.



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

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

VI. TECHNOLOGIES USED

Frontend: Next.js [8], TypeScript and Tailwind CSS [9]. Data and storage: PostgreSQL [11] with the Prisma ORM [10], and CSV and Excel file processing. AI: OpenAI API [12] for insight generation and chat. Security: JSON Web Tokens [13] and bcrypt password hashing [14].

VII. RESULTS AND DISCUSSION

The system was demonstrated with a sample employee dataset (sample_100_records.csv) containing 100 records. The dataset inventory of the same session also lists files of 5, 100 and 10,000 rows. Fig. 3 shows the dashboard and Fig. 4 shows the report generated for the sample dataset. Table 2 lists the values the application displayed for this dataset.

Output item	Value shown by InsightAI
Records processed	100
Detected domain	Human Resources Management
Average age (range)	40.2 years (21–59)
Average salary (range)	80,949.45 (40,504–119,968)
Data-quality indicator	98%
Largest department by headcount	HR, 20% of workforce
Segments in share chart	7 (HR, Support, IT, Finance, Operations, Sales, Marketing)
Written output	6 strategic insights, 5 roadmap actions, 4 risk items

Table 2. Output displayed for the 100-record sample dataset

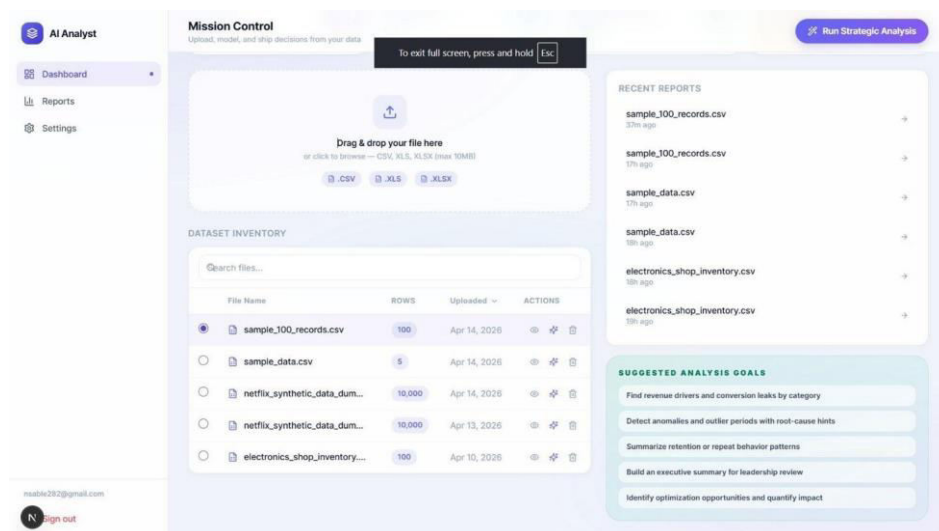


Fig. 3. Dashboard with upload panel, dataset inventory and recent reports



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

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

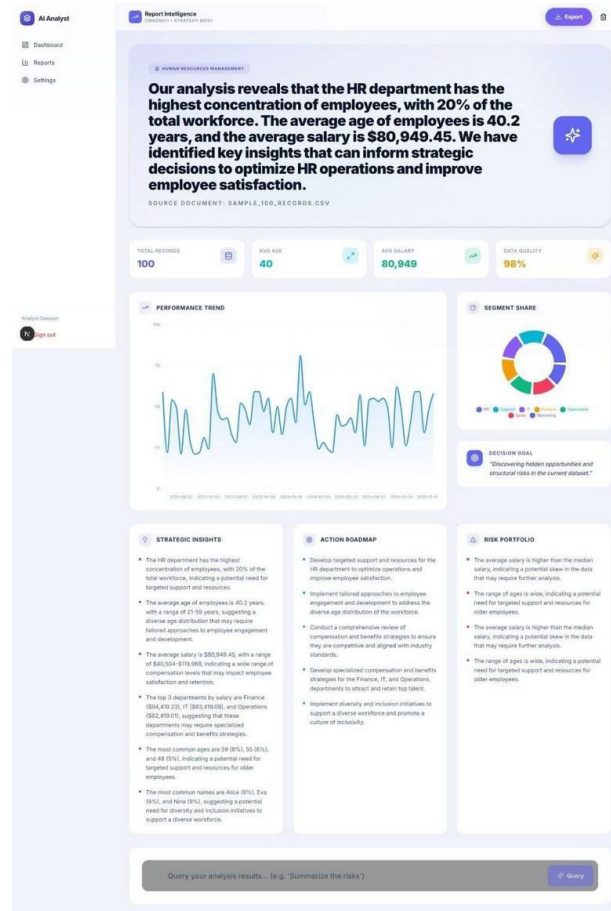


Fig. 4. Generated report: indicators, trend chart, segment share, insights, roadmap and risks

Table 3 summarises which features were observed in the demonstration.

Feature	Evidence in demonstration	Status
CSV / XLS / XLSX upload (max 10 MB)	Upload panel (Fig. 3)	Observed
Dataset inventory with search, view, delete	Files of 5 to 10,000 rows listed (Fig. 3)	Observed
Summary indicators and statistics	Table 2, Fig. 4	Observed
Trend and segment charts	Fig. 4	Observed
AI insights, roadmap and risks	Fig. 4	Observed
Plain-language query box and export	Controls present on report page	Interface present; formal test pending
Sign-in and per-user session	Signed-in session with sign-out (Figs. 3, 4)	Observed; protected-route tests pending

Table 3. Features observed in the demonstration



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

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

The main observations were:

- A complete report, from file upload to written insights, was produced for the sample dataset without any user programming.
- The report combines numeric indicators, two chart types and three kinds of written output on a single page.
- The detected domain (Human Resources Management) was used to frame the written insights.

7.1 Limitations

The results are based on a demonstration with a small set of sample datasets, not on a formal test suite. The values in Table 2 are those displayed by the application and were not independently recomputed, and no accuracy, response-time, load or user-study measurements are reported. The generated text is not always well supported: in the demonstration, some inferences (for example, conclusions drawn from the most common employee names) were only weakly related to the data, and some risk items were repeated, which shows why AI output must be read together with the computed statistics. The system depends on an external LLM service, so wording can vary between runs, and it is limited to small and medium CSV and Excel files.

VIII. CONCLUSION AND FUTURE SCOPE

InsightAI provides one AI-assisted web platform in which a user can upload a CSV or Excel file and obtain cleaned data, descriptive statistics, interactive charts and plain-language insights, ask follow-up questions, and export a report, with sign-in and per-user storage. The paper presents the design of the system and a working demonstration; its value lies in integrating established techniques into an easy workflow, and quantitative claims about accuracy or usability are left to future evaluation.

Future work includes a formal test plan (input handling, correctness of statistics against an independent tool, authentication and protected routes, AI-service failure behaviour and repeatability), checking AI-generated statements automatically against computed values, support for JSON and PDF inputs, machine-learning features such as forecasting, classification and clustering, handling of larger datasets, and a small user study with students and small teams.

REFERENCES

- [1] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.
- [2] H. Wickham, "Tidy data," *Journal of Statistical Software*, vol. 59, no. 10, pp. 1–23, 2014. doi: 10.18637/jss.v059.i10
- [3] S. Kandel, A. Paepcke, J. M. Hellerstein and J. Heer, "Enterprise data analysis and visualization: An interview study," *IEEE Trans. Visualization and Computer Graphics*, vol. 18, no. 12, 2012. doi: 10.1109/TVCG.2012.219
- [4] A. 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.
- [5] A. Narechania, A. Srinivasan and J. Stasko, "NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries," *IEEE Trans. Visualization and Computer Graphics*, vol. 27, no. 2, 2021.
- [6] V. Dibia, "LIDA: A tool for automatic generation of grammar-agnostic visualizations and infographics using large language models," in *Proc. 61st Annual Meeting of the ACL (Volume 3: System Demonstrations)*, Toronto, Canada, 2023. doi: 10.18653/v1/2023.acl-demo.11
- [7] T. B. Brown et al., "Language models are few-shot learners," in *Advances in Neural Information Processing Systems* 33, 2020.
- [8] Next.js Documentation. <https://nextjs.org/docs>
- [9] Tailwind CSS Documentation. <https://tailwindcss.com/docs>
- [10] Prisma Documentation. <https://www.prisma.io/docs>
- [11] PostgreSQL Documentation. <https://www.postgresql.org/docs/>
- [12] OpenAI API Documentation. <https://platform.openai.com/docs>
- [13] M. Jones, J. Bradley and N. Sakimura, "RFC 7519: JSON Web Token (JWT)," IETF, 2015. <https://www.rfc-editor.org/rfc/rfc7519>
- [14] N. Provos and D. Mazières, "A future-adaptable password scheme," in *Proc. USENIX Annual Technical Conference*, 1999.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

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

www.ijmrset.com