📓 Implementation Document

School Marksheet → AI Insights Generator

---

1. Introduction

The **School Marksheet → AI Insights Generator** is a data analytics and AI-powered application designed to automate the analysis of academic marksheets used by schools. The system converts raw marksheet files (Excel, CSV, or PDF) into structured analytics and generates meaningful insights for teachers and school administrators.

The primary goal of this project is to **reduce manual effort**, **avoid calculation errors**, and **provide data-driven academic insights** in a user-friendly manner.

---

2. System Architecture

The application follows a **modular architecture**, where each component has a single responsibility. This improves maintainability, scalability, and clarity.

Core Modules

1. **app.py**
    - Acts as the main entry point.
    - Handles the Streamlit UI and user workflow.
    - Coordinates data flow between modules.
2. **parser.py**
    - Responsible for reading marksheet files.
    - Supports Excel, CSV, and PDF formats.
    - Automatically detects header rows.
    - Cleans empty rows and columns.
    - Normalizes inconsistent column names.
    - Performs auto column detection (roll no, name, attendance, subjects).
3. **analysis.py**
    - Performs statistical computations.
    - Calculates student-wise metrics such as total marks, average marks, and weak student flags.
    - Computes subject-wise averages.
    - Generates class-level summaries.

4. **insights_llm.py**
    - Builds structured context from numeric analysis.
    - Integrates AI models to generate human-readable insights.
    - Supports Gemini API, Hugging Face models, and fallback logic.

---

3. Data Flow

The system follows a clear end-to-end data pipeline:

1. User uploads a marksheet file.
2. File is parsed and cleaned using the parser module.
3. Columns are auto-detected and shown to the user for verification.
4. Confirmed columns are passed to the analysis module.
5. Statistical metrics are calculated.
6. A structured summary is built.
7. AI generates a narrative insights report.
8. Results are displayed in the Streamlit UI.

This ensures transparency, accuracy, and reliability at every stage.

---

4. Column Detection Strategy

Since school marksheets vary widely in format, the system uses a **hybrid detection strategy**:

- **Keyword matching** for roll number, name, and attendance columns.
- **Numeric validation** to detect subject columns.
- Threshold checks to exclude ID-like numeric columns.
- Manual override via UI for full flexibility.

This approach balances automation with human verification.

---

5. AI Usage

Purpose of AI

AI is used to transform numeric statistics into **teacher-friendly, actionable insights**, rather than raw numbers.

AI Inputs

The AI model receives a structured context containing:

- Class average
- Number of students
- Weak student count
- Subject-wise averages
- Top and weak performers

AI Outputs

The generated report includes:

- Overall performance summary
- Identification of weak subjects
- Strong subject recognition
- Actionable teaching recommendations

Supported AI Options

- Google Gemini API
- Hugging Face Inference API
- Fallback static insights (if no API is configured)

The fallback mechanism ensures that the application **never fails to generate insights**, even in offline or API-limited environments.

---

6. User Interface Design

The UI is built using **Streamlit** with a step-by-step workflow:

1. File upload and preview
2. Automatic column detection display
3. Manual column confirmation
4. Analytics dashboard
5. AI insights generation

This design prioritizes clarity, usability, and error prevention.

---

7. Limitations

Despite its robustness, the system has some limitations:

- PDF table extraction accuracy depends on the PDF structure.
- AI insight quality depends on the quality of input data.
- Currently supports analysis of a single class at a time.

- No built-in authentication or role management.

These limitations are acceptable for a prototype and internship-level project.

---

8. Future Improvements

The project can be extended in several ways:

- Export reports as PDF or Excel.
- Student-wise personalized feedback.
- Multi-class and multi-year performance comparison.
- Predictive analytics for student performance.
- Integration with school management systems.
- Cloud deployment with authentication and user roles.

---

9. Conclusion

The School Marksheet → AI Insights Generator successfully demonstrates the integration of **data analytics, intelligent automation, and AI-based insights** into a real-world education problem.

The project highlights skills in Python, data processing, UI development, and AI integration, making it a strong practical solution for academic analysis and a solid internship submission.

---

10. Git Repo

sohamkhule/Marksheet: Marksheet Insights is a tool that processes student marksheets and generates useful insights such as total marks, percentage, grade, and performance trends. The project helps in quickly analyzing academic results in a structured and automated way.
