ZEMASAi (MLFlow: Bridging Data to Deployment)

1. Project Description

Our project, **ZEMASAi**, introduces a unified Machine Learning Lifecycle platform — **MLFlow** — designed to streamline the process from raw data ingestion to model deployment.
It automates data cleaning, exploratory data analysis (EDA), model training, evaluation, and deployment through a seamless interface.
The main goal is to reduce complexity, improve data quality, and speed up the journey from data to actionable insights.

---

2. Group Members & Roles

Member Name	Role / Responsibility
Mohamed Adel	Data Loading Function (responsible for reading and importing datasets in CSV, Excel, and JSON formats).
Eyad Sherif	Data Cleaning Function (handles missing values, duplicates, outliers, and feature preprocessing).
Ahmed Mohamed	Exploratory Data Analysis (EDA) Function (creates data visualizations and statistical insights).
Ziad Hawana	Model Training Function (builds and tunes ML models).
Salim Mohamed	Evaluation Function (assesses model performance using metrics like accuracy, F1-score, and error rate).
Ahmed Akram	Deployment & UI Function (deploys the model as an API and builds a user-friendly interface).

---

3. Team Leader

Ahmed Akram — Project Manager and Evaluation Function Lead.
Responsible for overall coordination, task integration, and progress tracking.

---

4. Objectives

- Automate the **Machine Learning Lifecycle** from data ingestion to deployment.
- Ensure **high data quality** and consistency across all stages.
- Improve **model accuracy and explainability** through clean EDA insights.
- Provide a **user interface** for seamless deployment and monitoring.
- Enhance **team collaboration** and code reusability across modules.

---

5. Tools & Technologies

- **Programming Language:** Python
- **Libraries & Frameworks:** Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn, TensorFlow/PyTorch
- **Deployment Tools:** FastAPI / Flask
- **Version Control:** Git & GitHub
- **UI Development:** Streamlit

- **Data Storage:** CSV / Excel / JSON files
- **Project Management:** Trello

---

6. Milestones & Deadlines

Milestone	Description	Deadline
M1: Planning & Research	Architecture & UI design	Week 1
M2: Data Module Completion	Develop and test data loading & cleaning modules.	Week 2
M3: EDA Visualization	Implement EDA and visualization functions.	Week 3
M4: Model Training	Create and test training pipeline.	Week 4
M5: Evaluation Metrics	Implement model evaluation metrics and testing.	Week 5
M6: Deployment & UI	Build API and user interface for deployment.	Week 6
M7: Integration & Testing	Combine all modules and finalize report.	Week 7

---

7. KPIs (Key Performance Indicators)

1. Data Quality

- Percentage of missing values handled: **≥ 98%**
- Data accuracy after preprocessing: **≥ 95%**
- Dataset diversity (representation of categories): **≥ 90%**

2. Model Performance

- Model accuracy (Accuracy/F1-Score): **≥ 90%**
- Model prediction speed (Latency): **< 50 milliseconds**
- Error rate (False Positive/Negative): **≤ 10%**

3. Deployment & Scalability

- API uptime: **≥ 99%**
- Response time per request: **< 200 milliseconds**
- Real-time processing (if applicable): N/A (Tabular Model)

4. Business Impact & Practical Use

- Reduction in manual effort: **≥ 70%**
- Expected cost savings: **≥ 40%**
- User satisfaction: **≥ 85%**

8. Summary

This proposal presents **ZEMASAi (MLFlow)** — an integrated ML lifecycle platform combining **data engineering, analysis, model training, evaluation, and deployment** in one unified workflow. Our modular design enables teamwork, efficiency, and scalability, ensuring a smoother path from **data to production**.