**INDIAN INSTITUTE OF TECHNOLOGY JODHPUR**

**NH 62 Nagaur Road, Karwar, Jodhpur, Rajasthan**



**MTech – Data Engineering**

**Machine Learning with Big Data**

**Project Proposal – Book Recommendation System**

**Submitted By:**

**Aneerban Chowdhury Roll No: G23AI2059**

**Pavani Racham Roll No: G23AI2032**

**Naresh Vemuri Roll No: G23AI2025**

**Project Proposal: Book Recommendation System**

**Explanation**:

The Book Recommendation System aims to enhance user experience by providing personalized book recommendations based on user preferences and reading history. This system will leverage Google Cloud services such as **BigQuery**, Cloud Storage (GCS), and Machine Learning models to process and analyze large datasets efficiently. The recommendation engine will employ **collaborative** **filtering** and **content**-**based** **filtering** **techniques** to generate accurate suggestions. The solution will also feature a user-friendly frontend for interaction and a backend for data processing and model inference.

**List of Key Technology Challenges:**

* **Data Collection & Storage:** Handling large volumes of book metadata, user interactions, and preferences efficiently in BigQuery and GCS.
* **Data Processing & Pipelines:** Creating scalable ETL pipelines for cleaning, transforming, and processing data.
* **Recommendation Model Accuracy:** Balancing collaborative and content-based filtering approaches for high accuracy.
* **Real-Time Processing:** Ensuring low-latency recommendations through optimized query execution and model inference.
* **Scalability & Performance:** Managing high user traffic and large datasets without performance bottlenecks.
* **Frontend Integration:** Delivering seamless interaction between users and the recommendation engine.

**Technology Stack:**

* **UI**: Flask
* **Database**: Google BigQuery
* **Storage**: Google Cloud Storage
* **Machine** **Learning**: Scikit-Learn, Google AI Platform
* **Visualization**: Google Looker Studio, Plotly

**List of Deliverables:**

* **Data Ingestion & Processing Pipelines**: ETL pipelines for ingesting book data and user interactions.
* **Recommendation** **Engine**: Collaborative & content-based filtering for book recommendations.
* **User** **Interface**: Interactive frontend for users to browse and receive recommendations.
* **Performance** **Optimization**: Efficient query execution and inference for real-time suggestions.
* **Deployment** **&** **Scaling**: Fully deployed system on Google Cloud ensuring scalability and reliability.
* **Visualization** **&** **Reporting**: Dashboards providing insights into user preferences
* **Technical** **Documentation**: Comprehensive documentation for system architecture, API endpoints, and ML model workflows.

**Data Sources:**

* Kaggle Dataset - Collaborative Filtering Books Recommendation
* Size of Dataset & Attributes
  + Books 271360 rows X 8 columns – ISBN, Title, Author, Publication, Year of Publishing
  + Users 278858 rows X 3 columns – User ID, Location, Age
  + Ratings 1149780 rows X 3 columns – User ID, ISBN, Rating