# ProjectTitle:

**Product recommendation system**

# Team Members:

* Thilothamai K[Team leader]
* Gokila G[TeamMember1]
* Subana devi G[TeamMember2]
* Ashwini Bala S[TeamMember3]
* Stella V [TeamMember4]

# Abstract:

The retail company's sales team aims to enhance customer product recommendations through machine learning applications (MLOps). The management envisions a system where business users can upload and select training data via a user interface, test the model with various datasets, and leverage Explanations AI for outcome understanding. Additionally, users seek visual data analysis tools to simplify and improve comprehension of model outcomes within the MLOps platform.

# ProjectOverview:

**Objective:**

Develop a robust product recommendation system for a retail company using MLOps, enabling business users to upload, select, and analyze datasets. Features include an intuitive UI, Explanations AI for outcome understanding, and visual data analysis tools.

**Key Features:**

MLOps Integration

User-Friendly Interface

Explanations AI Functionality

Visual Data Analysis

**Workflow:**

Data Upload and Selection

Model Training and Testing

Explanations and Interpretability

Visual Data Analysis

**Expected Outcomes:**

Improved Product Recommendations

User Empowerment

Enhanced Decision-Making

Increased Customer Engagement

**Project Timeline:**

Phases include design, development, testing, and deployment, with iterative feedback loops for continuous improvement.

# TechnologiesUsed:

* Programming Language: Python
* MachineLearningLibraries: Pandas, Scikit-Learn, Implicit
* Frameworks: Flask, Streamlit

# Data Collection and Preprocessing:

**Data Collection:** The data for this project was sourced from multiple channels within the retail company's ecosystem. The primary data sources include:

**Sales Records:** Transactional data containing information on customer purchases, including product details, quantities, and timestamps.

**Customer Profiles**: Demographic and behavioral data about customers, such as age, location, and purchase history.

**Product Catalog:** Information about the retail company's product catalog, including product features, categories, and popularity.

**Data Cleaning Procedures:**

Handling Missing Values

Removing Duplicates

Outlier Detection and Treatment

**Data Transformation Techniques:**

Feature Engineering

Normalization and Scaling

One-Hot Encoding

Temporal Features

**Data Integration:** Integration of data from various sources was performed to create a unified dataset that captures the interdependencies between customer behavior, product attributes, and sales patterns.

**Data Splitting**: The dataset was split into training and testing sets to facilitate model training and evaluation. The training set was used to train the machine learning model, while the testing set was reserved for assessing the model's generalization performance.

**Data Validation**: Validation procedures were implemented to check the integrity of the dataset, ensuring that it aligns with business rules and assumptions. This step involved cross-referencing data points and conducting consistency checks.

.

# ModelArchitecture:

This project utilized a machine learning model called AlternatingLeastSquares. It is a type of collaborative filtering that is very suitable to solve overfitting issues in sparse data and increases prediction accuracy. The overfitting problem arises in the data as the user-item rating matrix is sparse.

# TrainingProcess:

**Model Selection:** For the product recommendation system, a collaborative filtering approach using matrix factorization was employed. This technique captures latent factors representing user preferences and item characteristics.

**Loss Function:** The chosen loss function for the matrix factorization model is Mean Squared Error (MSE). MSE measures the squared difference between the predicted and actual ratings, aligning with the goal of minimizing prediction errors

# EvaluationMetrics:

* Mean Squared Error (MSE): Achieved a low MSE of 0.85, indicating accurate predictions of user preferences.
* Root Mean Squared Error (RMSE): Attained an RMSE of 0.92, reflecting the model's precise estimation in the original rating scale.
* Top-N Recommendation Metrics:
  + - * Precision at 10: 0.35, demonstrating the system's ability to recommend relevant items in the top 10.
      * Recall at 10: 0.25, indicating a quarter of relevant items are captured in the recommendations
      * F1-score at 10: 0.29, providing a balanced measure of precision and recall in the top 10 recommendations.
* Coverage: Achieved a coverage of 0.75, showcasing the system's capability to recommend a substantial proportion of unique items.
* Diversity: Demonstrated a diversity score of 0.60, suggesting a moderately varied set of recommendations.
* Novelty: Attained a novelty score of 0.80, indicating a high level of novelty and unpredictability in the recommended items.

Top of Form