# Yelp Business Analytics, Recommendations and Rating Prediction

# Project Synopsis

# CDAC DBDA

**Guided By:**

**Ishan sir**

**Submitted By:**

**Group 2**

****

**Centre for Development of Advanced Computing (CDAC)** Post Graduate Diploma in Big Data Analytics (PG-DBDA) Mumbai Centre – USM VITA

**Academic Year: Feb – Aug 2025**

## Team Members:

The project team consists of seven dedicated and skilled members:

* Karan Anpat– Team Leader
* Mrunmai Jawalekar– Team Member
* Tushar Nandurkar – Team Member
* Mayuri Kakde – Team Member
* Yogeshwar Patil – Team Member
* Sanjana Dalvi – Team Member
* Santosh – Team Member

**Title:**

**Yelp Business Data Analytics, Recommendation and Rating Prediction**

**Problem Statement:**

Yelp hosts millions of reviews, ratings, and user interactions across various businesses like restaurants, salons, and shops. With such a vast amount of data, users often face difficulty finding the most relevant and trustworthy businesses based on their preferences. Therefore, there is a need to analyze Yelp’s data to:

* Understand user and business trends,
* Build an intelligent recommendation system,
* Predict user ratings for businesses using machine learning models.

**Objectives:**

1. Analyze Yelp business, user, and review data to extract useful insights.
2. Build a recommendation system to suggest relevant businesses to users.
3. Predict user ratings for businesses using machine learning models.
4. Evaluate model performance using metrics like RMSE and MAE.

**Dataset Description:**

The Yelp dataset contains a wealth of information, including:

* Business data: name, location, category, average rating
* Review data: review text, user ID, business ID, stars (rating)
* User data: number of reviews, average rating, and more

**Scope of Work:**

1. **Data Cleaning and Preprocessing:**
   * Load and merge business, review, and user data
   * Handle missing values, extract relevant features
2. **Exploratory Data Analysis (EDA):**
   * Analyze user behavior, business trends, and rating distributions
   * Visualize insights using charts and plots
3. **Recommendation System:**
   * **Content-Based Filtering:** Suggest similar businesses based on category and location
   * **Collaborative Filtering:** Use matrix factorization (SVD) to recommend businesses based on user-user similarity
4. **Rating Prediction:**
   * Train ML models like Random Forest, XGBoost to predict user ratings
   * Optionally include review text sentiment as a feature (using NLP)
5. **Model Evaluation:**
   * Evaluate performance using metrics such as RMSE, MAE, R² Score
   * Use train-test split and cross-validation for reliability

**Tools & Technologies:**

* **Programming Language:** Python
* **Libraries:** pandas, NumPy, NLTK, Scikit-learn, Matplotlib, Seaborn, TextBlob, SpaCy, Transformers
* **Visualization:** Plotly, Streamlit, PowerBI
* **IDE:** Jupyter Notebook
* **Dataset Source:** Yelp Open Dataset

**Expected Outcomes:**

* Visual and statistical insights from Yelp data
* A recommendation engine suggesting personalized businesses
* A predictive model forecasting user ratings

**Conclusion:**

This project bridges data analytics with practical machine learning by using real-world data from Yelp. It demonstrates how intelligent systems can assist users in decision-making by offering personalized recommendations and predictions, which are widely used in platforms like Amazon, Netflix, and Yelp itself.