| **Project Title** | **Swiggy’s Restaurant Recommendation System using Streamlit** |
| --- | --- |
| **Skills take away From This Project** | * **Data Preprocessing** * **One-Hot Encoding** * **Clustering (K-Means, Cosine Similarity, or Similar Methods)** * **Streamlit Application Development** * **Python** |
| **Domain** | **Recommendation Systems and Data Analytics** |

**Problem Statement:**

The objective is to build a recommendation system based on restaurant data provided in a CSV file. The system should recommend restaurants to users based on input features such as city, rating, cost, and cuisine preferences. The application will utilize clustering or similarity measures to generate recommendations and display results in an easy-to-use Streamlit interface.

By the end of this project, you will:

1. Perform data cleaning and preprocessing.
2. Encode categorical features using One-Hot Encoding.
3. Apply clustering or similarity-based methods for recommendations.
4. Build a Streamlit application to showcase recommendations.

**Business Use Cases:**

1. **Personalized Recommendations**: Help users discover restaurants based on their preferences.
2. **Improved Customer Experience**: Provide tailored suggestions to enhance decision-making.
3. **Market Insights**: Understand customer preferences and behaviors for targeted marketing.
4. **Operational Efficiency**: Enable businesses to optimize their offerings based on popular preferences.

**Approach:**

* The dataset is provided as a CSV file with the following columns:
* **['id', 'name', 'city', 'rating', 'rating\_count', 'cost', 'cuisine',**
* **'lic\_no', 'link', 'address', 'menu']**
* **Categorical**: name, city, cuisine
* **Numerical**: rating, rating\_count, cost

### **1. Data Understanding and Cleaning**

* **Duplicate Removal**: Identify and drop duplicate rows.
* **Handling Missing Values**: Impute or drop rows with missing values.
* Save the cleaned data to a new CSV file (cleaned\_data.csv).

### **2. Data Preprocessing**

* **Encoding**: Apply One-Hot Encoding to categorical features (name, city, cuisine).
  + Save the encoder as a Pickle file (encoder.pkl).
* Ensure all features are numerical after encoding.
* Create a preprocessed dataset (encoded\_data.csv).
* Ensure the indices of cleaned\_data.csv and encoded\_data.csv match.

### **3. Recommendation Methodology**

* **Clustering or Similarity Measures**:
  + Use **K-Means Clustering** or **Cosine Similarity** to identify similar restaurants based on input features.
  + Use the encoded dataset for computations.
* **Result Mapping**:
  + Map the recommendation results (indices) back to the non-encoded dataset (cleaned\_data.csv).

### **4. Streamlit Application**

* Build an interactive application with the following components:
  + **User Input**: Accept user preferences (e.g., city, cuisine, rating,price,etc).
  + **Recommendation Engine**: Process the input, query the encoded data, and generate recommendations.
  + **Output**: Display recommended restaurants using cleaned\_data.csv.

**Results:**

* **Data Preprocessing**
  + Cleaned Dataset (cleaned\_data.csv):
    - Categorical and numerical features with missing values and duplicates removed.
  + Encoded Dataset (encoded\_data.csv):
    - Preprocessed numerical dataset with categorical features One-Hot Encoded.
  + Encoder File (encoder.pkl):
    - Serialized One-Hot Encoder for Streamlit use.
* **Recommendation System**
  + Clustering or Similarity-based recommendation engine.
  + Mapping results from encoded\_data.csv to cleaned\_data.csv for interpretation.
* **Streamlit Application**
  + User-friendly interface for input and output.
  + Clear display of recommendations from the cleaned dataset.
* **Report**
  + Summarize:
    - Data Cleaning and Preprocessing Steps
    - Methodology for Recommendations
    - Key Results and Insights

**Project Evaluation metrics:**

1. **Recommendation Quality**:
   1. Evaluate recommendations for relevance and diversity.
2. **Application Usability**:
   1. Ensure the Streamlit app is user-friendly.
3. **Data Alignment**:
   1. Verify indices between cleaned\_data.csv and encoded\_data.csv.

## **Timeline**

**Complete the project within 7 days of initiation:**

1. **Day 1-2: Data Cleaning and Preprocessing**
2. **Day 3-4: Build Recommendation Engine**
3. **Day 5-6: Develop Streamlit Application**
4. **Day 7: Testing and Documentation**

**Data Set:**

Dataset : [Data](https://drive.google.com/file/d/1oFn9nfrL1sx2XvRBwysffu66waHX7Sm4/view?usp=sharing)

**Project Deliverables:**

1. **Source Code:** Python scripts or notebooks with clear documentation.
2. **Report:** A comprehensive report summarizing:
   * Approach
   * Data analysis
   * Insights and recommendations

**Project Guidelines:**

1. **Coding Standards:**
   * Use consistent naming conventions and comments for readability.
   * Follow best practices for Python programming.
2. **Version Control:**
   * Use Git for version tracking.
   * Maintain a clean and organized repository.
3. **Testing and Validation:**
   * Validate models using cross-validation techniques.
   * Ensure reproducibility by setting random seeds.

**Timeline:**

The project must be completed and submitted **within 7 days from the assigned date**

**References:**

| **Project Live Evaluation** | [**Project Live Evaluation**](https://docs.google.com/document/u/0/d/1QisLD2kqDWFZJG2oDknKn2eMGi-Xq8oFPgA7UWSbcIQ/edit) |
| --- | --- |
| **EDA Guide** | [**Exploratory Data Analysis (EDA) Guide**](https://docs.google.com/document/d/1tHiTU1X9UwXSLySpJ-FVCohlf_8xpXwa75vlK9S6wl8/edit?usp=sharing) |
| **Capstone Explanation Guideline** | [**Capstone Explanation Guideline**](https://docs.google.com/document/d/1gbhLvJYY7J73lu1g9c6C9LRJvYemiDOdRDAEMe632w8/edit) |
| **GitHub Reference** | [**How to Use GitHub.pptx**](https://docs.google.com/presentation/d/1XHCbgUOqbcXNUyQ87vTlKdKRgAbBxtkA/edit?usp=sharing&ouid=109735616107417446342&rtpof=true&sd=true) |
| **Project Orientation (English)** | [**English Orientation**](https://drive.google.com/file/d/1Gw9BwgU3hoJzounrNPTiTEa-9Zh6dTyu/view?usp=sharing) |
| **Project Orientation (Tamil)** | [**Orientation Rec tamil**](https://drive.google.com/file/d/1vCP9LrIh8kQvpfBQw2sH_G8FASKwVMHL/view?usp=sharing) |

**PROJECT DOUBT CLARIFICATION SESSION ( PROJECT AND CLASS DOUBTS)**

**About Session:** The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

**Note: Book the slot at least before 12:00 Pm on the same day**

**Timing: Monday-Saturday (4:00PM to 5:00PM)**

**Booking link :**[**https://forms.gle/XC553oSbMJ2Gcfug9**](https://forms.gle/XC553oSbMJ2Gcfug9)

**LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)**

**About Session:** The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

**Note: This form will Open only on Saturday (after 2 PM ) and Sunday on Every Week**

**Timing: Monday-Saturday (05:30PM to 07:00PM)**

**Booking link :** [**https://forms.gle/1m2Gsro41fLtZurRA**](https://forms.gle/1m2Gsro41fLtZurRA)

| **Created By:** | **Verified By:** | **Approved By:** |
| --- | --- | --- |
| [**Asvin Selvarajan**](mailto:asvinselvarajan@gmail.com) | **Shadiya** | [**Nehlath Harmain**](mailto:Nehlath@guvi.in) |