
Customer Segmentation using Self-Organizing Map (SOM)

Project Description

This project applies a Self-Organizing Map (SOM) to perform customer segmentation on the UCI Online Retail Dataset. The goal is to group customers based on their purchasing behavior, which can help businesses optimize marketing strategies and enhance customer engagement.

Dataset

- **Source**: UCI Online Retail Dataset
- **Description**: This dataset contains transactional data of a UK-based online retail business between 2010 and 2011. The dataset includes features such as:
    - `InvoiceNo`: Invoice number
    - `StockCode`: Unique product code
    - `Description`: Product description
    - `Quantity`: Number of items purchased
    - `InvoiceDate`: Purchase date
    - `UnitPrice`: Price of each unit
    - `CustomerID`: Unique identifier for each customer
    - `Country`: Customer's country

The dataset is located at `/Users/neharavindran/Downloads/Online Retail.xlsx`.

Software Requirements

To run the project, ensure that you have the following software installed on your machine:

- **Python 3.x** (Python 3.10 or higher recommended)
- **Required Python Libraries**:
    - `numpy`
    - `pandas`
    - `matplotlib`
    - `minisom`

Install the required libraries by running the following command:

```
pip install -r requirements.txt
```

Hardware Requirements

Recommended hardware for running the project:

- **System**: MacBook Air M1 or equivalent
- **Memory**: 8 GB RAM or more
- **Processor**: M1 Chip or Intel i5 and above
- **Storage**: At least 2 GB of free space

Setup and Execution

Step 1: Clone the Repository

```
git clone https://github.com/<your-reg-no>/<repo-name>.git
cd <repo-name>
```

Step 2: Set up Virtual Environment (Optional)

```
python3 -m venv myenv
source myenv/bin/activate  # For Mac/Linux
```

Step 3: Install Dependencies

```
pip install -r requirements.txt
```

Step 4: Update Dataset Path

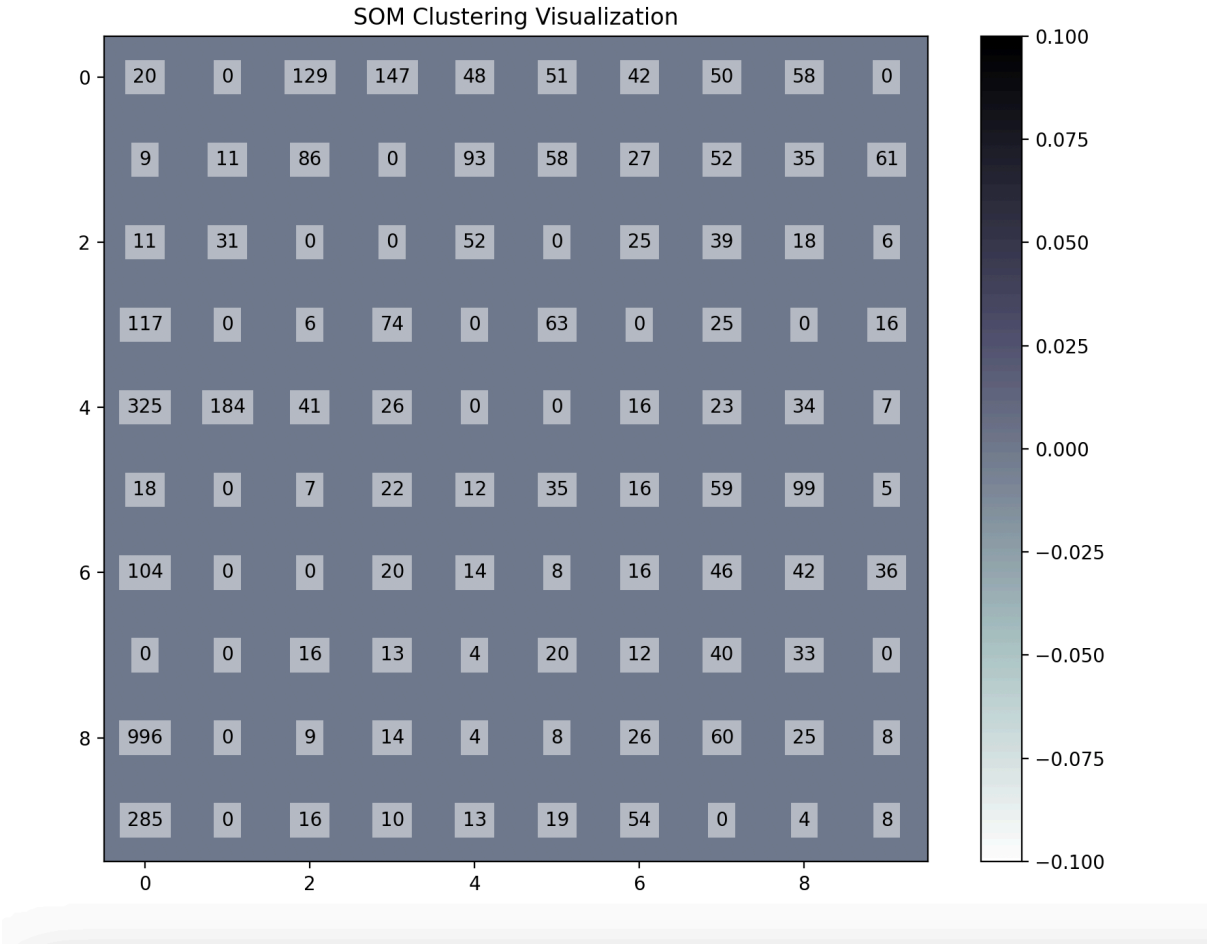
Ensure that the dataset is placed in `/Users/neharavindran/Downloads/Online Retail.xlsx`. If the path is different, modify the path in the code accordingly.

Step 5: Run the Clustering Script

```
python som_clustering.py
```

Outputs

- The script will generate a **visualization** showing customer clusters using the Self-Organizing Map.
- A CSV file `clustered_customers.csv` will be saved containing customer data with assigned cluster labels.



Project Files

- **`som_clustering.py`**: The main Python script that reads the dataset, trains the SOM, and visualizes the clustering results.
- **`requirements.txt`**: Lists all necessary dependencies to run the project.
- **`README.md`**: This file, containing the instructions to execute the project.

How to Execute the Project

1. Clone the repository.
2. Install the required packages using `pip`.
3. Update the dataset path if needed.
4. Run the SOM clustering script to visualize customer segmentation.

Contact Information

- **Name**: NEHA.R- 22MIS1088
- **Email**:neha.r2022@vitstudent.ac.in