**Developed By: Minha Mahmood**

**Roll No: BSAIM-F24-099**

**Course: Artificial Intelligence**

**Project Documentation: Wholesale Customer Analysis & Prediction System**

## Project Overview

This project is a comprehensive Machine Learning-based Web Application for analyzing wholesale customer data, predicting customer channels (Retail vs. Horeca), and performing customer segmentation using clustering algorithms. The application provides an interactive interface for data exploration, prediction, and visualization.

## 1. Problem Statement

Wholesale distributors need to understand their customers' purchasing patterns to optimise marketing strategies, inventory management, and customer relationship management. The challenge is to:

- Classify customers into appropriate channels (Retail or Horeca)

- Segment customers based on spending behaviour

- Provide actionable insights through an intuitive interface

## 2. Dataset Description

Dataset: Wholesale Customers Data

Source: UCI Machine Learning Repository

Records: 440 customers

Features: 8 attributes

**Features:**

**1. Channel (Target Variable):**

   - 1 = Horeca (Hotel/Restaurant/Cafe)

   - 2 = Retail

**2. Region:**

   - 1 = Lisbon

   - 2 = Oporto

   - 3 = Other

**3. Product Categories (Annual spending in monetary units):**

   - Fresh

   - Milk

   - Grocery

   - Frozen

   - Detergents_Paper

   - Delicassen

## 3. Technical Architecture

### 3.1 Tech Stack

Frontend: HTML5, CSS3, Jinja2 Templates

Backend: Flask (Python Web Framework)

Machine Learning: Scikit-learn, Pandas, NumPy

Data Processing: Pandas, NumPy

Model Persistence: Pickle

Visualisation: HTML Tables

## 3.2 Project Structure

wholesale-project/

- app.py

- model_svc.pkl

- processed_data.csv

- Wholesale customers data.csv

- static/style.css

- templates/*.html

- notebooks/*.ipynb

## 4. Machine Learning Pipeline

## 4.1 Data Preprocessing

Steps:

- Loading & Exploration

- Feature Selection

- Null Value Handling

- Data Type Conversion

- Label Encoding

- Export cleaned data

## 4.2 Model Development

Algorithm: Support Vector Classifier (SVC)

Accuracy: 87.9%

4.3 Clustering Analysis

Algorithm: K-Means (3 clusters)

## 5. Web Application Features

Pages:

- Home

- Prediction

- Results

- Dataset Viewer

- Summary

- Cluster Analysis

## 6. Installation & Setup

Requirements:

Python 3.7+, Flask, Pandas, Scikit-learn, NumPy

## Commands:

pip install Flask pandas scikit-learn numpy

python app.py

## 7. Usage Instructions

How to:

- Make predictions

- Explore dataset

- Interpret outputs


## 8. Business Applications

- Marketing Optimisation

- Inventory Management

- Customer Relationship Management

- Strategic Planning


## 9. Technical Challenges & Solutions

### Challenges:

- Data imbalance

- Feature selection

- Model accuracy

- Web integration

### Solutions:

- Preprocessing

- Feature engineering

- Algorithm selection

- Modular design


## 10. Future Enhancements

- More ML models

- Advanced visualisations

- Database integration

- Authentication

- Export features

- Cloud deployment

## 11. Conclusion

This project demonstrates complete ML workflow, business value, deployment, and documentation.