Advanced Marketing Segmentation Using Deep Clustering - Research Document

Project Overview

This is a sophisticated customer segmentation tool that combines deep learning (autoencoders) with traditional clustering techniques (K-means) to perform advanced customer segmentation. The application is built using Streamlit for the interface and TensorFlow for the deep learning components.

Key Technologies Used

- Streamlit: For the web interface
- TensorFlow: For deep learning and autoencoder implementation
- Scikit-learn: For PCA and K-means clustering
- Pandas: For data manipulation
- Plotly: For interactive visualizations

Technical Architecture

The system integrates several components seamlessly to form a full pipeline from data ingestion to model training and visualization.

Data Processing Pipeline

1. Data Input:
   - Accepts CSV files via a user-friendly upload interface.
   - Option to download a sample dataset.

- Performs initial data validation and preview.

2. Preprocessing Steps:

- Handles missing values using median imputation.
- Identifies categorical and numerical columns automatically.
- Performs one-hot encoding for categorical variables.
- Standardizes numerical features using StandardScaler.
- Applies PCA for dimensionality reduction, preserving 95% variance.

Deep Learning Architecture

- Input Layer: Matches the dimension of PCA-transformed data.
- Encoder: Three dense layers (128 -> 64 -> 32 neurons).
- Decoder: Three dense layers (32 -> 64 -> 128 -> original dimension).
- Activation: ReLU for hidden layers, Sigmoid for output.
- Optimizer: Adam with mean squared error loss.
- Data Split: 80-20 train-validation.

Clustering Approach

- K-means clustering on encoded representations.
- User-defined cluster count (2-10).
- Combines deep feature learning with traditional clustering.

Advanced Marketing Segmentation Using Deep Clustering - Research Document

Visualization Suite

- Bar Chart: Customer group distribution.
- Scatter Plot: 2D cluster visualization.
- Box Plots: Feature importance.
- Scatter Matrix: Metric relationships.
- 3D Scatter Plot: Cluster visualization.
- Radar Charts: Group profiles.
- Heatmap: Metric comparison.

Model Persistence

- Autoencoder model saved in .keras format.
- K-means model in pickle format.
- Scaler and PCA transformers also saved using pickle.

Key Features

- Interactive web interface.
- Real-time processing and visualization.
- Automated preprocessing pipeline.
- Dimensionality reduction and deep feature learning.
- Model persistence for future use.

Advanced Marketing Segmentation Using Deep Clustering - Research Document

Use Cases

- Customer behavior analysis.
- Market segmentation.
- Target marketing.
- Customer persona development.
- Strategic decision making.

Technical Benefits

- Combines deep learning with traditional clustering.
- Handles both categorical and numerical data.
- Provides interpretable visualizations.
- Reduces manual effort through automation.
- Accessible to non-technical users via interface.

Limitations

- Requires CSV format input.
- Limited to 10 clusters.
- Uses fixed autoencoder architecture.
- Needs sufficient data for effective clustering.