Supply Chain Analytics Project

Objective
Predicting Future Product Demand

Demand forecasting combines the techniques of art and science to anticipate the future product needs of customers.

Dataset

- **Source:** [Online Retail Dataset from UCI Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/online+retail)
- **Type:** Multivariate, Sequential, Time-Series
- **Context:** Transactions from a UK-based non-store online retail between 2010-2011.
- **Contents:** The dataset includes transaction details from a gift shop and contains fields such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

Correlation Analysis

Feature Engineering from "Description" Column

- **Product Type:** Extracted nouns using a POS tagger to identify product names, added to a new column called 'Product_type'.
- **Colour Type:** Identified product colors and added them to a new column called 'Colour_type'.

New Revenue Column

- **Calculation:** Revenue = UnitPrice * Quantity

Machine Learning Model Approach (Three Steps)

1. **Clustering:** Data is initially clustered.
2. **Classification:** Clustering outputs are used as labeled data for classification.
3. **Prediction:** Sales quantity is predicted using regression, incorporating the cluster number as a feature.

Clustering Details:

Challenge: Handling mixed attributes (numerical + categorical) without losing the significance of categorical data.

Solution: K-Prototypes algorithm, which is suitable for mixed data types.

Clustering with k-prototypes

- Adapts the k-means paradigm to handle both numerical and categorical data.
- Aims to maximize similarity within clusters based on a mix of attribute types.

Clustering Output

The clustering result is utilized as the target variable for training the classification model.

Classification Phase

A Linear SVM algorithm provided effective performance for classifying clusters.

Demand Prediction

- Integrated all features and designated the 'Quantity' column as the target.
- The dataset was divided into training and testing sets to predict item quantities.
- The Random Forest algorithm yielded robust results for forecasting demands.

Summary

- Effective data preprocessing shaped the inputs for modeling.
- Ingenious feature engineering enhanced the interpretation of demand.
- A sequential three-step modeling approach (clustering, classification, and prediction) was employed to refine forecasting techniques and tailor them to specific business needs.