Customer Segmentation and Campaign Success Analysis using Bank Marketing Dataset

Project Overview

This project aims to analyse the **Bank Marketing Dataset (UCI)** to identify customer segments that are more likely to subscribe to a term deposit. By leveraging clustering techniques, the goal is to provide actionable insights that can guide marketing strategies for improved customer engagement and subscription rates.

The main focus is to predict whether a customer will subscribe to a term deposit (variable `y` with values `yes` or `no`) based on various demographic and behavioural features. The analysis involves data cleaning, feature engineering, clustering, and performance evaluation, offering insights into which customer segments yield higher subscription rates.

Data Preprocessing and Feature Engineering

Handling Outliers

In the preprocessing phase, outliers in numerical columns such as `balance` and `campaign` were identified and replaced with the **95th percentile values** to minimize their impact on the clustering results. This approach helped reduce the skewness in the data and ensured more reliable and interpretable clusters.

Feature Transformation

- **Binarisation of `previous` and `pdays`**: These features exhibited skewed distributions, with many zero values. To address this, the values were converted into binary features: any non-zero value was coded as `1`, and zero values remained as `0`.
- **Encoding of Ordinal Feature (`education`)**: The `education` feature was transformed based on its importance level. The encoding order was set from least to most important: `unknown` < `primary` < `secondary` < `tertiary`.

- **One-Hot Encoding for Categorical Features**: Features like `job`, `marital`, and `housing` were one-hot encoded to transform them into numerical data that could be used effectively in clustering algorithms.
- **Standardisation of Numerical Features**: Numerical features such as `age`, `balance`, and `campaign` were standardised using the Standard Scaler to ensure they were on the same scale, which is crucial for distance-based algorithms like KMeans.

Clustering Analysis

Methodology

KMeans clustering was used to segment customers based on their attributes. The optimal number of clusters was determined using a combination of the **Silhouette Score** and the **Elbow Method**. The **Silhouette Score** indicated that five clusters were the most optimal for this dataset, providing well-separated and meaningful customer segments.

Cluster Profiling

Each cluster was analysed to identify distinct customer profiles. These profiles were created by examining key features such as age, balance, education, housing status, and previous campaign contact frequency. The profiling process helped reveal patterns that are critical for targeted marketing strategies.

Results and Insights

Subscription Analysis

The main goal was to identify which clusters had the highest proportion of customers who subscribed to the term deposit (i.e., `y = yes`). The following key insights were derived from the cluster profiles:

1. **Cluster 3**:
   - **Highest Proportion of yes Subscriptions**
   - Characteristics:
     - Higher average balances.
     - Predominantly customers with tertiary education.
     - Lower frequency of previous contacts.
     - Positive history with housing and personal loans.

- ○ **Marketing Insights**:
    - ■ This cluster represents a **premium customer segment** that is more likely to subscribe to term deposits. Marketing campaigns targeting this segment should focus on offering exclusive benefits, premium services, and long-term savings options.
2. **Cluster 2**:
    - ○ **Moderate Proportion of yes Subscriptions**
    - ○ Characteristics:
        - ■ Middle-aged customers.
        - ■ Moderate balances and frequent past contact attempts.
    - ○ **Marketing Insights**:
        - ■ This segment shows potential for conversion. Campaigns should emphasise financial stability and investment growth, considering their higher balance and prior interactions with the bank.
3. **Cluster 4**:
    - ○ **Lowest Proportion of yes Subscriptions**
    - ○ Characteristics:
        - ■ Low balance and high frequency of contact during the campaign.
    - ○ **Marketing Insights**:
        - ■ This segment appears less responsive, possibly due to over-contacting. A more tailored and less frequent approach may improve engagement and conversion rates.

Visualisation of Clusters

Cluster distribution and subscription rates were visualised to illustrate the varying success rates across segments. The visualisation demonstrated the clear contrast between clusters, particularly in terms of their likelihood to subscribe to term deposits.

Evaluation Metrics

To validate the effectiveness of the clustering model, several evaluation metrics were used:

- **Silhouette Score**: A score of 0.307 indicated that the clusters were reasonably well-separated.
- **Davies-Bouldin Index**: A value of 1.277 suggested that the clusters were relatively distinct.
- **Calinski-Harabasz Index**: A high value of 14,934.82 further supported the clustering quality.

These metrics suggest that the clustering process was successful in identifying meaningful customer segments that can inform targeted marketing strategies.

Cluster targets:

1. **Target Cluster 3 for Premium Offers**:
   a. Focus marketing efforts on customers in this cluster, offering them premium services, higher interest rates, and long-term investment opportunities. Their high average balance and tertiary education level make them ideal candidates for term deposit subscriptions.
2. **Engage Cluster 2 with Financial Growth Campaigns**:
   a. For middle-aged customers in this segment, highlight the benefits of financial growth, savings, and security through term deposits. Tailor the messaging around their moderate balance and prior contact history to improve conversion rates.
3. **Optimise Contact Strategies for Cluster 4**:
   a. Reduce the frequency of contact with customers in Cluster 4, as they exhibit the lowest subscription rates. Implement targeted campaigns with personalised offers and less frequent touchpoints to avoid customer fatigue.

Conclusion

By applying clustering techniques, this analysis provides valuable insights into customer behaviour, particularly in relation to the likelihood of subscribing to a term deposit. The findings emphasise the importance of segmenting customers based on their financial and behavioural attributes to optimise marketing strategies. Future campaigns should focus on targeting **Cluster 3**, the premium segment, while fine-tuning approaches for other segments based on their unique characteristics.

With these insights, the bank can refine its marketing strategies, increase engagement, and ultimately improve subscription rates for term deposits.