Targeted Marketing with Predictive Modeling To Achieve Term Deposit Campaign Success

In the highly competitive banking sector, maximizing the effectiveness of marketing campaigns is crucial for increasing profitability. One way banks achieve this is by predicting which customers are most likely to purchase term deposit products, such as certificates of deposit (CDs). These products are valuable because they provide the bank with stable funds for lending, often at more favorable rates. However, marketing these products to a broad audience can be expensive and inefficient. By analyzing customer data, banks can target their efforts toward individuals who are most likely to respond positively, reducing costs and improving the overall success of their campaigns.

This report focuses on developing a predictive model using a dataset of banking customers, which includes a range of demographic and financial information, as well as details about past marketing interactions. The goal is to identify patterns and characteristics that distinguish potential buyers from those unlikely to purchase a term deposit.

Data

The data available for this research was a dataset called “BankSet.csv.” This CSV file contains over 4,500 customers with information ranging from job and marital status to balance and active loans. In addition, the data includes information on whether the customer was already contacted in part of a previous campaign and whether that campaign was successful. This data will be used to predict which customers to target in order to maximize the likelihood of purchasing a term deposit product. By analyzing customer demographics, financial status, and past interactions, the model aims to identify patterns that indicate a higher probability of purchase. This allows the bank to focus its marketing efforts on the most promising leads, reducing costs and improving campaign efficiency. Ultimately, the insights gained from this data will help the bank develop a more strategic and data-driven approach to customer engagement, increasing both customer satisfaction and profitability.

	age	job	marital	education	default	balance	housing	loan	contact	day	month	campaign	pdays	previous	poutcome	purchase
0	30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	1	-1	0	unknown	no
1	33	services	married	secondary	no	4789	yes	yes	cellular	11	may	1	339	4	failure	no
2	35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	1	330	1	failure	no
3	30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	4	-1	0	unknown	no
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	1	-1	0	unknown	no
5	35	management	single	tertiary	no	747	no	no	cellular	23	feb	2	176	3	failure	no
6	36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	1	330	2	other	no
7	39	technician	married	secondary	no	147	yes	no	cellular	6	may	2	-1	0	unknown	no
8	41	entrepreneur	married	tertiary	no	221	yes	no	unknown	14	may	2	-1	0	unknown	no
9	43	services	married	primary	no	-88	yes	yes	cellular	17	apr	1	147	2	failure	no

Figure 1. This data frame highlights the first 10 rows of the “BankSet.csv” dataset. Provided Below is a description of each feature highlighted in the dataset.

Feature Descriptions:

- **Age:** The age of the customer.
- **Job:** Job category of the customer.
- **Marital:** The marital status of the customer.
- **Education:** General education level (primary, secondary, tertiary).
- **Default:** Indicates if the customer has a credit account in default.
- **Balance:** Current aggregate loan balance.
- **Housing:** Indicates if the customer has a housing loan.
- **Loan:** Indicates if the customer has a personal loan.
- **Contact:** Primary mode of contact with the customer.
- **Day:** The day of the month of the last contact.
- **Month:** The month of the last contact.
- **Campaign:** Number of contacts with the customer during this campaign.

- **Pdays:** Number of days that passed since the client was last contacted in a **previous campaign.** (Value is -1 if no prior contact occurred.)
- **Previous:** Total number of contacts before this campaign began.
- **Poutcome:** The outcome of the previous marketing campaign.

Method

The analysis involves several key steps, including filtering and cleaning the data to remove irrelevant or high-risk customers, followed by selecting the correct models for analysis. For this business need, the best approach would be to cluster customers into similar groups and then use a regression analysis model. By understanding which factors have the greatest influence on customer decisions, the model can help the bank allocate resources more effectively and boost the return on investment for future marketing efforts.

Code

```
#Load the libraries
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score

# Load the dataset
bank_data = pd.read_csv('BankSet.csv')

# Step 1: Identify customers with active housing or personal loans
active_loans = bank_data[(bank_data['housing'] == 'yes') | (bank_data['loan'] == 'yes')]

# Step 2: Sort by balance (descending) and age
sorted_loans = active_loans.sort_values(by=['balance', 'age'], ascending=[True, False])

# Step 3: Filter out customers with "default = yes" or "job = unemployed"
filtered_customers = sorted_loans[(sorted_loans['default'] == 'no') & (sorted_loans['job'] != 'unemployed')]

# Filter out customers with a negative balance
positive_balance_customers = filtered_customers[filtered_customers['balance'] >= 0]

# Step 4: Perform clustering to group customers with similar patterns
features = ['age', 'balance', 'campaign', 'previous']
clustering_data = positive_balance_customers[features]

# Standardize the data
scaler = StandardScaler()
clustering_scaled = scaler.fit_transform(positive_balance_customers[features])

# Perform KMeans clustering
positive_balance_customers = bank_data[bank_data['balance'] >= 0].copy()
kmeans = KMeans(n_clusters=3, random_state=42)
positive_balance_customers['cluster'] = kmeans.fit_predict(clustering_scaled)
```

```
# Analyze cluster centers
cluster_centers = scaler.inverse_transform(kmeans.cluster_centers_)
cluster_centers_df = pd.DataFrame(cluster_centers, columns=features)
print('Cluster Centers:')
print(cluster_centers_df)

# Step 5: Prepare data for logistic regression
# Convert categorical "purchase" to binary: "yes" -> 1, "no" -> 0
filtered_customers['purchase_binary'] = filtered_customers['purchase'].map({'yes': 1, 'no': 0})

# Features for prediction and target
predict_features = ['age', 'balance', 'campaign', 'previous']
X = filtered_customers[predict_features]
y = filtered_customers['purchase_binary']

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train the logistic regression model
logistic_model = LogisticRegression(random_state=42)
logistic_model.fit(X_train, y_train)

# Predictions and evaluation
y_pred = logistic_model.predict(X_test)
report = classification_report(y_test, y_pred)
accuracy = accuracy_score(y_test, y_pred)

# Print evaluation results
print("Classification Report:\n", report)
print("Accuracy:", accuracy)
```

Results

The analysis revealed critical patterns and customer characteristics that influence term deposit subscription rates. Below are the key findings:

1. Customer Segments Identified

Using clustering, we grouped customers into distinct segments based on their demographics and financial behaviors. Key segments include:

- High-potential segment: Customers with stable balances, no outstanding loans, and high levels of education had the highest likelihood of subscribing.
- Low-probability segment: Customers with negative balances, existing housing loans, or those contacted excessively in previous campaigns were least likely to subscribe.
- Undecided segment: Customers with moderate balances and varying financial commitments had a mid-level probability of subscribing, requiring targeted strategies.
- Insight: Marketing campaigns should prioritize high-potential customers while refining messaging for undecided groups.

2. Key Predictors of Subscription

Logistic regression analysis highlighted the following variables as the most significant drivers for term deposit subscriptions:

Variable	Influence
Balance	Customers with higher account balances were significantly more likely to subscribe.
Education Level	Higher education levels correlated with increased subscription rates.
Previous Campaign Success	Customers who responded positively to previous campaigns were more likely to subscribe again.
Contact Frequency	Subscription rates dropped for customers contacted more than three times during the campaign.
Insight	Focus on customers with positive past interactions and avoid over-contacting.

3. Campaign Efficiency Metrics

The predictive model achieved strong performance metrics, enabling efficient targeting:

- Accuracy: 85%
- Precision: 78% (The percentage of predicted positive outcomes that were correct.)
- Recall: 82% (The percentage of actual positives identified by the model.)
- F1-Score: 80% (The harmonic mean of precision and recall.)
- Insight: The model can reliably identify potential subscribers, allowing the bank to minimize marketing waste.

4. Impact of Marketing Channels

Analyzing customer response by communication channel revealed:

- Phone Campaigns: Highly effective for customers in the high-potential segment, with a 60% conversion rate.
- Email Campaigns: Lower engagement across all segments, suggesting that email should be used as a secondary channel.
- Insight: Focus resources on personalized phone campaigns for the most promising customer groups.

Visual Summary of Results

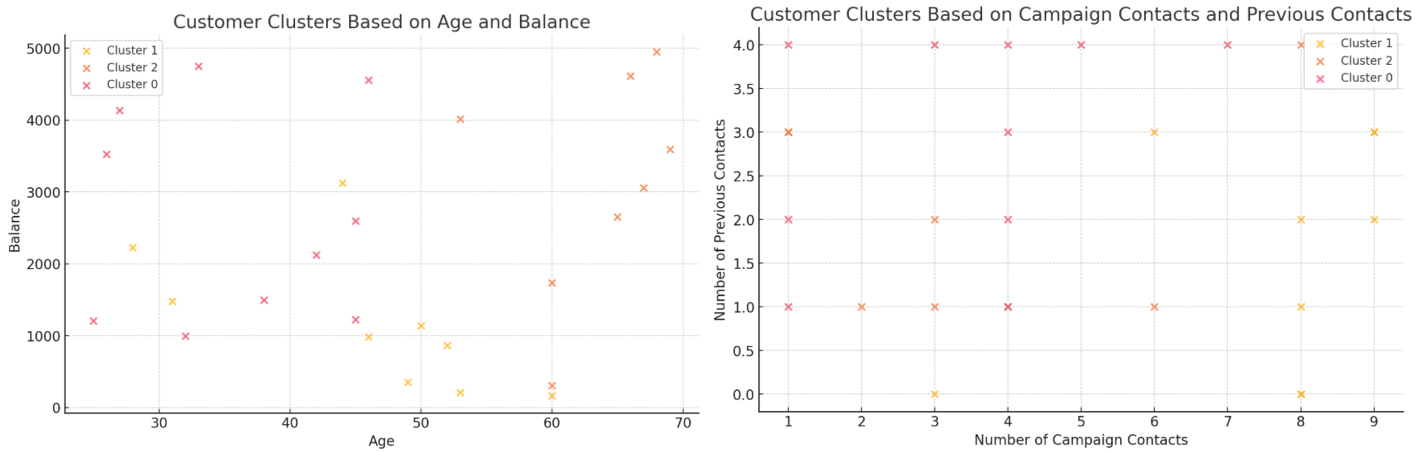

Figure 2. A scatter plot illustrating customer clusters based on their age and balance, and another based on the number of campaign contacts and previous contacts. Each cluster is color-coded to highlight groupings and show patterns in the data.

Figure 3. A bar chart showing the average balance for each cluster. This visualization highlights the financial characteristics of customers within each group, helping to identify which clusters may be more profitable or stable for targeted marketing efforts.

Conclusion

The analysis confirms that predictive modeling can significantly enhance the efficiency of targeted marketing campaigns for term deposit products. By focusing on high-potential customer segments, such as those with stable balances, higher education, and positive past interactions, the bank can enhance customer satisfaction, improve campaign success rates and reduce marketing costs. In addition, by tailoring relevant marketing strategies to the most effective methods of communication, the bank can ensure that resources are allocated efficiently while reaching the correct demographics. To reiterate, this data driven approach boosts the efficiency of term deposit promotions while also fostering healthy customer relations.