

Customer Segmentation Prediction for New Markets

Context

An automobile company plans to enter new markets with existing products (P1-P5). Based on market research, the new market behavior is similar to their existing market. In the existing market, customers are classified into 4 segments (A, B, C, D), and this segmented outreach strategy has been highly successful. The goal is to predict the correct segment for 2627 new potential customers.

Goal

Develop a machine learning model to predict the customer segment (A, B, C, or D).

Phase 1: Setup and Data Loading

```
In [3]: # Import standard Python packages
import numpy as np
import tensorflow as tf
import tensorflow.keras.layers as nn
import os
import math
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler, OneHotEncoder
from sklearn.impute import SimpleImputer
from tensorflow import keras
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.utils import to_categorical

# Model Imports
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier, VotingClassifier
from xgboost import XGBClassifier
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.utils import to_categorical

# Set up environment for better visualization
sns.set_style("whitegrid")
plt.rcParams['figure.figsize'] = (10, 6)

# Define file paths
TRAIN_FILE = "Train.csv"
TEST_FILE = "Test.csv"

In [4]: # Load the datasets
df_train = pd.read_csv(TRAIN_FILE)
df_test = pd.read_csv(TEST_FILE)
print("Data loaded successfully!")

# Rename the target column for clarity
df_train.rename(columns={'Segmentation': 'Segment'}, inplace=True)

# Store the Test IDs for submission
test_ids = df_test['ID']

# Display initial information
print("----- Training Data Info -----")
df_train.info()
print("----- Test Data Info -----")
df_test.info()

# Display the first few rows of the training data
print("----- Training Data Head -----")
print(df_train.head())

# Check the target variable distribution
print("----- Target Variable Distribution (Segment) -----")
print(df_train['Segment'].value_counts(normalize=True) * 100)
```

Data loaded successfully.

```
--- Training Data Info ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2627 entries, 0 to 2626
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
 --- 
 0   ID              2627 non-null   int64  
 1   Gender          2627 non-null   object  
 2   Ever_Married    2627 non-null   object  
 3   Age             2627 non-null   int64  
 4   Graduated      2627 non-null   object  
 5   Profession     2627 non-null   object  
 6   Work_Experience 2627 non-null   float64 
 7   Spending_Score 2627 non-null   float64 
 8   Family_Size     2627 non-null   float64 
 9   Var_1           2627 non-null   object  
dtypes: float64(2), int64(2), object(7)
memory usage: 693.5+ KB
```

```
--- Test Data Info ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2627 entries, 0 to 2626
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
 --- 
 0   ID              2627 non-null   int64  
 1   Gender          2627 non-null   object  
 2   Ever_Married    2627 non-null   object  
 3   Age             2627 non-null   int64  
 4   Graduated      2627 non-null   object  
 5   Profession     2627 non-null   object  
 6   Work_Experience 2627 non-null   float64 
 7   Spending_Score 2627 non-null   float64 
 8   Family_Size     2627 non-null   float64 
 9   Var_1           2627 non-null   object  
dtypes: float64(2), int64(2), object(6)
memory usage: 205.4+ KB
```

```
--- Training Data Head ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2627 entries, 0 to 2626
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
 --- 
 0   ID              2627 non-null   int64  
 1   Gender          2627 non-null   object  
 2   Ever_Married    2627 non-null   object  
 3   Age             2627 non-null   int64  
 4   Graduated      2627 non-null   object  
 5   Profession     2627 non-null   object  
 6   Work_Experience 2627 non-null   float64 
 7   Spending_Score 2627 non-null   float64 
 8   Family_Size     2627 non-null   float64 
 9   Var_1           2627 non-null   object  
dtypes: float64(2), int64(2), object(6)
memory usage: 693.5+ KB
```

```
--- Target Variable Distribution (Segment) ---
Segment          Count
D               829
A               113
C               140
B               50
Name: proportion, dtype: float64
```

```
2.1 Data Cleaning and Missing Values
```

```
In [5]: # Function to check for missing values
def check_missing_values(df, name):
    missing = df.isnull().sum()
    missing = missing[missing > 0]
    missing['Missing %'] = missing / len(df) * 100
    missing_data = pd.DataFrame({'Missing Count': missing, 'Missing Percent': missing_percent})
    print(f"Data {name} Missing Values")
    print(missing_data.sort_values(by='Missing Percent', ascending=False))
    return missing_data
```

```
check_missing_values(df_train, "Train")
check_missing_values(df_test, "Test")
```

```
--- Missing Values in Train Data ---
Missing Count   Missing Percent
Work_Experience  829          10.275161
Family_Size      113          4.301485
Ever_Married     140          1.735250
Profession       50           1.536936
Graduated        78           0.966782
Var_1            76           0.941993
```

```
--- Missing Values in Test Data ---
Missing Count   Missing Percent
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     50           1.903312
Profession       38           1.144577
Var_1            32           1.218120
Graduated        24           0.913590
```

```
Out[5]: Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38           1.144651
Work_Experience  269          10.239817
Family_Size      113          4.301485
Ever_Married     32           1.218120
Graduated        24           0.913590
```

```
--- Missing Count   Missing Percent
Ever_Married     50           0.903312
Graduated        24           0.913590
Profession       38          
```

