MARKETING CAMPAIGN ANALYSIS AND OPTIMIZATION

Project Report

By

Athira. C.S

Under the guidance and supervision of

Mrs. Amritha. M. R

TECHOLAS TECHNOLOGIES

Palakkad, Kerala, 678001

Executive Summary

This project focuses on analysing and optimizing marketing campaigns using **Machine Learning**, **Business Intelligence** tools and advanced data analysis techniques. The aim was to analyse customer data and understand customer purchasing behaviours, identify influential features, uncover insights, predict customer responses to marketing campaigns and visualize actionable metrics for decision-makers.

The project was divided into two parts:

1. **Machine Learning Pipeline** using Python in Jupyter Notebook.
2. **Interactive Dashboard Development** using Power BI.

Using the **Marketing Campaign Dataset** from Kaggle, I performed extensive data cleaning, preprocessing, exploratory data analysis (EDA), feature engineering, and visualization. Various classification algorithms were applied, and their performances were compared to select the best model for predicting customer responses.

After parameter tuning using K-Fold Cross Validation, **Random Forest** emerged as the most effective model, achieving an accuracy of **91.8 %**. This project demonstrates my ability to work with end-to-end machine learning pipelines to solve real-world business problems.

By merging predictive analytics with interactive data visualization using Power BI, this project provides a well-rounded solution for marketing campaign optimization, supporting both decision-making and operational strategies. Key metrics such as Total customers, Average Income,Total Campaign Response Rate, Average Age and Total Complaints are calculated using DAX and presented visually. The dashboard also includes slicers for real-time filtering for Education, Age group, Marital status and Products amount.

Objectives

1. Analyze Customer Purchasing Patterns and Campaign Responses

- Conduct in-depth exploratory data analysis (EDA) to understand customer demographics, purchasing behaviors, and interactions with past marketing campaigns.
- Identify key trends and correlations among features such as education level, marital status, income, age, product spending, and campaign acceptance.

2. Predict Customer Responses to Marketing Campaigns Using Machine Learning Models

- Apply various classification algorithms to build predictive models that can forecast customer responses to marketing campaigns.
- Train, validate, and compare models such as Logistic Regression, Random Forest, Decision Tree, Support Vector Machine, K-Nearest Neighbors, and Gaussian Naive Bayes.
- Optimize the best-performing model using K-Fold Cross Validation to ensure robust and accurate predictions.
- Use predictive insights to recommend targeted marketing strategies for improving campaign success rates.

3. Visualize Key Marketing Metrics through an Interactive Power BI Dashboard

- Develop a dynamic and interactive Power BI dashboard to present essential marketing KPIs in a visually intuitive manner.
- Provide a comprehensive view of customer demographics, product spending habits, campaign response rates, and overall campaign effectiveness.
- Enable business stakeholders to explore data through interactive slicers, charts, and filters for deeper analysis.
- Support marketing decision-makers in monitoring campaign performance and identifying high-value customer segments quickly.

4. Identify Target Customer Groups for Optimized Campaign Planning

- Combine findings from machine learning models and dashboard analysis to pinpoint high-response customer groups.
- Identify demographic segments (such as education, marital status, and income level) that are more likely to respond positively to marketing efforts.
- Recommend actionable strategies for campaign personalization and resource allocation based on these customer segments.
- Provide a data-driven foundation for future marketing initiatives aimed at maximizing return on investment (ROI) and improving customer engagement.

Methodology

Part 1 : Machine learning Analysis

1. Data Collection & Loading

- Sourced the **Marketing Campaign Dataset** from Kaggle.
- Imported the dataset into Jupyter Notebook using Pandas.

2. Initial Data Exploration

- Reviewed dataset structure, data types, and statistical summaries using: info, describe , shape and datatypes.
- Identified missing values and duplicate records.

3. Data Cleaning

- Removed duplicates.
- Imputed missing values with **median** to prevent data skewness.

4. Feature Engineering & Preprocessing

- Created new features:
    - Total_Children = Kid_home + Teen_home
    - Total_Spent = Sum of product-related spending.
    - Customer_Tenure_Years based on tenure information.
    - Age = Current year – Year_birth
- Removed irrelevant columns.
- Applied **Label Encoding** and **One-Hot Encoding** for categorical variables.
- Detected and removed outliers from numerical columns.

5. Exploratory Data Analysis (EDA)

Generated several visualizations:

- Histograms for Age distribution.
- Boxplots for Income distribution and outliers.
- Count plots for Education Levels and Response distributions.
- Bar charts showing campaign response by Education and Marital Status.
- Scatter plot showing Income vs. Total Spending.
- Bar charts showing campaign acceptance rates.

6. Model Development & Evaluation

Tested multiple classification algorithms:

- **Logistic Regression**
- **Random Forest Classifier**

- **Decision Tree Classifier**
- **Support Vector Classifier (SVC)**
- **K-Nearest Neighbors (KNN)**
- **Gaussian Naive Bayes**

Applied **K-Fold Cross Validation** to ensure model robustness.

Best Model: Logistic Regression (CV=50) with **92% accuracy**.

Part 2: Power BI Dashboard Development

1. Data Preparation & Loading

- Loaded the cleaned dataset into Power BI.
- Unpivoted product-related columns for easier analysis.
- Prepared age group categories for segmentation.

2. Dashboard Features

Designed an interactive dashboard with the following visuals and components:

- **KPI Cards:**
  - ✓ Total Customers
  - ✓ Average Age
  - ✓ Average Income
  - ✓ Total Complaints
  - ✓ Total Campaign Response Rate
- **Stacked Column Chart:** Customer Count by Age Group and Response**.**
- **Stacked Bar Chart:** Response Rate by Education Level**.**
- **Pie Chart:** Customer distribution by Education**.**
- **Donut Chart:** Spending distribution across Product Categories**.**
- **Slicers (Filters):**
  - ✓ Education
  - ✓ Age Group
  - ✓ Marital Status
  - ✓ Amount Spent on each Product Category

3. Insights from Dashboard:

- Identified customer segments with higher response rates by age, education, and marital status.
- Highlighted product categories with higher spending.
- Provided easy-to-use filtering to explore data by different customer characteristics.

Tools Used

Tools	Purpose
Python	Data Cleaning, Feature Engineering, ML Modeling
Jupyter Notebook	Development Environment
Pandas, NumPy	Data Manipulation
Matplotlib, Seaborn	Data Visualization
Scikit-learn	ML Models & Evaluation
Power BI	Dashboard Development & Interactive Reporting
Kaggle	Dataset Source

Visuals

Python (Jupyter Notebook) Visuals:

- Histogram: Age Distribution.
- Boxplot: Income.
- Count plot: Education and Campaign Responses.
- Bar Charts: Campaign Response by Education, Spending by Education, Response Rate by Marital Status.
- Scatter Plot: Income vs. Total Spending.
- Bar Plots: Campaign Acceptance Rates.

Power BI Dashboard Visuals:

- KPI Cards: Total Customers, Average Age, Average Income, Total Complaints, Campaign Response Rate.
- Stacked Column Chart: Customers by Age Group and Response.
- Stacked Bar Chart: Response Rate by Education.
- Pie Chart: Customer Distribution by Education.
- Donut Chart: Spending by Product Category.
- Slicers: Education, Age Group, Marital Status, Product Spending.

Conclusion & Future Work

Conclusion:

The **Marketing Campaign Analysis and Optimization Project** successfully combined machine learning techniques with interactive data visualization to derive meaningful insights and actionable strategies for campaign improvement.

The analysis of customer data revealed strong relationships between attributes such as education**,** marital status**,** income, and spending behavior, and their influence on marketing campaign responses. Extensive data cleaning, feature engineering, and transformation were performed to prepare the dataset for modeling. Multiple classification algorithms were implemented and evaluated using cross-validation. While Random Forest slightly outperformed Logistic Regression during cross-validation tuning (91.8% vs. 91.7% accuracy), the **Logistic Regression model** demonstrated better generalization on the test set, achieving a **Precision of 1.0**, **Recall of 0.31**, and an **F1 Score of 0.48**, compared to much lower values from Random Forest. This suggests that Logistic Regression is more reliable for predicting positive campaign responses for supporting targeted marketing strategies especially given the class imbalance in the dataset.

Complementing the machine learning work, an **interactive Power BI dashboard** was developed to visualize and communicate key marketing metrics and trends. The dashboard highlights critical KPIs including:

- **Total Customers**: 2,216
- **Average Income**: 52,247
- **Average Age**: 56
- **Total Campaign Response Rate**: 15%
- **Total Complaints**: 21

Key insights derived from the dashboard include:

- Higher campaign response rates were observed among customers with PhD and Master’s degrees.
- Marital status played a significant role, where categories like "Alone" and "YOLO" showed higher response rates, although these groups had fewer individuals. This demonstrates that response percentage can reveal patterns not visible from raw counts alone.
- Product spending analysis showed that "MntWines" accounted for the largest share of spending (50.26%), followed by "MntMeatProducts" (27.51%).
- The dashboard provided effective segmentation based on education level**,** age group**,** marital status, and spending behaviour, aiding in identifying high-value customer segments.

Overall, this project showcased a complete data science workflow—from data preprocessing and predictive modeling to data-driven storytelling via visualization. It equips marketing teams with both a predictive model and a visual tool to improve decision-making, optimize campaign targeting, and enhance customer engagement in future campaigns.

Future Work:

- Apply advanced models like **XGBoost** or **LightGBM** for potential accuracy improvements.
- Automate Power BI data refreshes for live reporting.
- Integrate external data sources (social media, customer reviews) for enriched insights.
- Deploy predictive model as an API or web app for real-time campaign recommendations.
- Conduct A/B testing for campaign strategies based on model predictions.