Analyzing Digital Marketing Campaigns Using Machine Learning to Predict Sales

Predicting Conversion Rates and Optimizing Campaign Strategies

Christopher Nichols, Isagani Hernandez, Bijay Upadhyaya
DTSC 5502 Online

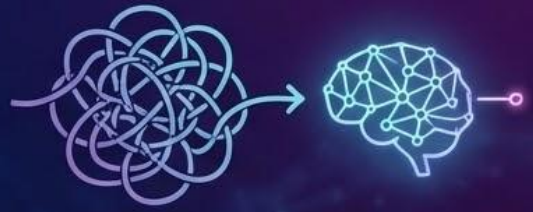
Why is marketing campaign analysis important?

STRATEGY & OPTIMIZATION

STRATEGY & OPTIMIZATION

Marketing campaigns are a crucial part of business strategy. By understanding what works and what doesn't, businesses can optimize their marketing spend and increase conversion rates.

ML-POWERED INSIGHTS



ML-POWERED INSIGHTS

Many businesses struggle to understand the effectiveness of their campaigns, especially with large and complex datasets. This project aims to provide a solution, using machine learning, that can offer insights in what features would be most important in making marketing decisions.

PROJECT GOALS

PRIMARY GOAL

Predict the effectiveness of marketing campaigns by analyzing features such as ad spend, campaign duration, audience reach, and more.

SUB-GOALS

- Predict conversion rates based on various marketing factors.
- Identify the most important features that influence campaign success.
- Create a model that businesses can use to make data-driven decisions.

RESULTS & EVALUATION

MODEL PERFORMANCE

Linear Regression

Moderate Correlation

MSE: **0.111** R^2: **~0.88**

Random Forests

Significant Improvement, Complex Interactions

MSE: **0.00124** R^2: **0.999**

XGBoost

Strong Performance

RMSE: **0.056** R^2: **0.997**

TOP FEATURES

1. ad_spend
2. conversion_ad_spend
3. ad_spend_engagement
4. engagement_metric

(Based on feature importance analysis)

KEY INSIGHTS

Interaction Terms

Pivotal interactions like **ad_spend_engagement** and **conversion_ad_spend** indicate budget's interplay with other factors.

Efficiency Matters

Features emphasizing ROI, e.g., **conversion_ad_spend**, align directly with maximizing conversions for minimal spending.

Audience Behavior

Engagement acts as a critical bridge between ad spend and conversion; ads must engage to convert.

DATA OVERVIEW

DATASET SUMMARY

Dataset: 10,000 samples with 43 features (e.g., ad spend, campaign duration, engagement metrics, audience demographics).

KEY FEATURES

- Ad Spend
- Industry Type
- Marketing Channel
- Audience Type
- Target Variable: Conversion Rate

DATA CLEANING & FEATURE ENGINEERING

- One-hot encoding (e.g., categorical data like industries, etc.)
- Interaction terms (capturing non-linear relationships to capture non-linear

KEY OBSERVATIONS FROM OUR CORRELATION MATRIX:

1. HIGH CORRELATIONS

Engagement Metric vs. Audience Reach (**0.911**)

Strong relationship, indicating potential redundancy. Decision needed on how to combine or select.

2. MODERATE CORRELATIONS

Ad Spend vs. Budget Allocation (**0.445**).

Positive correlation; higher ad spend often means more budget. Makes sense, implies overlap.

LOW/NEAR-ZERO CORRELATIONS

Conversion Rate vs. Ad Spend (**-0.0015**) & Duration vs. Success (**-0.0043**)

Suggests weak linear relationships. Combined 'Conversion Rate' and 'Ad Spend' into new feature 'conversion_ad_spend'.

4. NEGATIVE CORRELATIONS

Budget Allocation vs. Duration (**-0.648**) & Audience Reach vs. Budget Allocation (**-0.0106**)

Longer durations tend to have lower budget. Minor inefficiencies for wider reach.

METHODOLOGY AND MODELS USED

MACHINE LEARNING MODELS

Linear Regression: Used to model the relationship between features and conversion rate.

Random Forests: A tree-based method to capture non-linear relationships and feature importance.

XGBoost: A powerful gradient boosting technique for improving predictive accuracy.

WHY THESE MODELS?

Linear Regression: Gives interpretability and a baseline performance.

Random Forests: Help capture more intricate relationships between features and tell which features are most important.

XGBoost: Powerful for improving predictive accuracy and handling diverse and high-dimensional data. Can also give insight to which features are most important.

FEATURE ENGINEERING

WHY & WHAT OF FEATURE ENGINEERING

Why Feature Engineering?: Some features may not be directly useful, but by combining them, new relationships and patterns can possibly be discovered.

What Feature Engineering Did We Perform?: We determined features that showed colinearity and then created new features from these by multiplying closely related features together. This allows models to make accurate generalizations and see non-linear relationships (learned from Kaggle Flood Challenge).

IMPLEMENTATION EXAMPLES (CODE SNIPPET)

```
# 2. Interaction Feature Creation
digital_market['ad_spend_duration'] = digital_market['ad_spend'] * digital_market['duration']
digital_market['ad_spend_engagement'] = digital_market['ad_spend'] * digital_market['engagement_metric']

# --- Interactions with Conversion Metric ---
digital_market['industry_rat_audience_urban'] = digital_market['conversion_rate'] * digital_market['target_audience_urban']

# Interactions with categoric features
digital_market['industry_e-commerce_social_media'] = digital_market['industry_e-commerce'] * digital_market['marketing_channel_social_media']
digital_market['audience_30?_channel_influencer'] = digital_market['target_audience_25-34'] * digital_market['marketing_channel_influencer']

# Interactions with numeric and categorical features
digital_market['ad_spend_china_finance'] = digital_market['ad_spend'] * digital_market['industry_finance']
digital_market['duration_podcast_suburban'] = digital_market['duration'] * digital_market['target_audience_suburban']

# Interactions based on domain knowledge
digital_market['budget_athlete_audience_reach'] = digital_market['budget_allocation'] * digital_market['audience_reach']
digital_market['conversion_rate_engagement_metric'] = digital_market['conversion_rate'] * digital_market['engagement_metric']

# --- Interactions with Engagement Metrics ---
digital_market['clicks_pitch_25-35'] = digital_market['engagement_metric'] * digital_market['target_audience']
digital_market['engagement_rate_social_media'] = digital_market['engagement_metric'] * digital_market['audience_reach']
digital_market['engagement_engagement_duration'] = digital_market['conversion_rate'] * digital_market['engagement_metric']

# --- Interactions with Conversion Metrics ---
digital_market['clicks_stats_35-44'] = digital_market['engagement_metric'] * digital_market['target_audience_25-34']
digital_market['engagement_time_email_metric'] = digital_market['engagement_metric'] * digital_market['marketing_channel_social_media']
digital_market['engagement_word'] = digital_market['engagement_metric'] / digital_market['duration']

# --- Interactions with Conversion Metrics ---
digital_market['conversion_rate_25-34'] = digital_market['conversion_rate'] * digital_market['target_audience_25-34']
digital_market['conversion_rate_engagement'] = digital_market['conversion_rate'] * digital_market['marketing_channel_search engine']
digital_market['conversion_spend'] = digital_market['conversion_rate'] * digital_market['ad_spend']
```

Key Features: Interaction terms like ad_spend_duration, conversion_rate_audience_urban, and others.

DESIGN DECISIONS

CHOICE OF MODELS

XGBoost: Chosen for its high accuracy in prediction tasks.

Random Forests: Ability to handle complex data without overfitting.

Linear Regression: Baseline model to understand linear Explored powerful models not in class.

DATA PREPOCESSING

One-hot encoding: Used for categorical variables.

Feature scaling: Scaled some features for better performance in gradient-based models.

CHALLENGES IN DESIGN

Feature Selection: Deciding most relevant features, handling multicollinearity and nonlinear relationships.

Hyperparameter Tuning: Handled through grid search and cross-validation to improve model performance (especially Forests and XGBoost).

RESULTS & EVALUATION

MODEL PERFORMANCE

Linear Regression
Moderate Correlation
MSE: **0.111** R^2: **~0.88**

Random Forests
Significant Improvement, Complex Interactions
MSE: **0.00124** R^2: **0.999**

XGBoost
Strong Performance
RMSE: **0.056** R^2: **0.997**

TOP FEATURES

1. ad_spend
2. conversion_ad_spend
3. ad_spend_engagement
4. engagement_metric

(Based on feature importance analysis)

KEY INSIGHTS

Interaction Terms
Pivotal interactions like **ad_spend_engagement** and **conversion_ad_spend** indicate budget's interplay with other factors.

Efficiency Matters
Features emphasizing ROI, e.g., **conversion_ad_spend**, align directly with maximizing conversions for minimal spending.

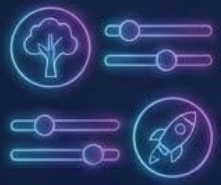
Audience Behavior
Engagement acts as a critical bridge between ad spend and conversion; ads must engage to convert.

WHAT'S NEXT?

IMPROVEMENTS

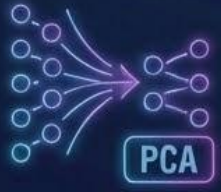
Hyperparameter Tuning

Implement further hyperparameter tuning for **XGBoost** and **Random Forests** to fine-tune the models.



Feature Engineering

Expand feature engineering by including more interaction terms or using advanced techniques like **PCA** for dimensionality reduction.



FUTURE WORK

Deploying the Model

Build a **web application** where businesses can input their campaign details and get predictions on conversion rates.

Additional Metrics

Expand the analysis to include other marketing performance metrics like **customer engagement** or **retention**.