https://github.com/Mallika-Yeturi/Modeling-for-Media-Campaign-Cost-Optimization

Project Phase II Report
Optimizing Media Campaign Costs at Food Mart X

Stakeholder

Who is your stakeholder?

The primary stakeholder is Food Mart X's marketing department, tasked with deploying media campaigns aimed at expanding the customer base efficiently and effectively.

Problem Statement
What is the problem they are trying to solve?

Food Mart X seeks to optimize its media campaign costs to ensure a more strategic allocation of its marketing budget, aiming to enhance customer acquisition rates without proportionally increasing spending.

Data Source
Where is your dataset from?

The dataset was obtained from Kaggle, a comprehensive source for open datasets. Specifically, we utilized the "Media Campaign Cost Prediction" dataset, which includes historical data on campaign spending, outcomes, customer demographics, and product details.

Dataset Link: Media Campaign Cost Prediction Dataset

Models Explored

What models did you try, and why?

Random Forest Regressor: Chosen for its ability to handle high-dimensional data and its robustness against overfitting. It's particularly well-suited for regression tasks involving complex, nonlinear relationships.

Hyperparameters Tuned:
n_estimators: 100, 200, 300
max_depth: 10, 20, None
min_samples_split: 2, 5, 10

XGBoost Regressor: Selected for its efficiency, scalability, and performance in various machine learning competitions. XGBoost offers built-in regularization which helps prevent overfitting.

Hyperparameters Tuned:
n_estimators: 100, 150, 200
learning_rate: 0.01, 0.1, 0.2
max_depth: 3, 6, 9

Feature Selection and Engineering
What features did you select/engineer? How did you choose those?

Based on initial EDA, we focused on features directly influencing campaign costs, such as store_sales, store_costs, and customer_interaction_metrics. Engineered features included cost_per_interaction and sales_to_cost_ratio, designed to capture the efficiency and effectiveness of media campaigns.

The selection process was guided by correlation analysis with the target variable and domain expertise, ensuring relevance to the cost prediction task.

Model Evaluation
How did you evaluate the model? What evaluation metrics did you use? Why?

Models were evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) metrics to provide a holistic view of performance:

MAE offers a straightforward average of absolute errors, making it easy to interpret.
RMSE penalizes larger errors more, which is crucial for cost optimization tasks.
R^2 indicates the proportion of variance explained by the model, offering insights into its explanatory power.
Reflections and Recommendations
Given the results, the XGBoost Regressor, with its optimized hyperparameters, demonstrated superior performance. It struck an excellent balance between prediction accuracy and model complexity, making it highly suitable for this application.

Future Directions
What would you do differently next time?

In future work, we plan to explore more advanced feature engineering techniques and potentially incorporate additional external data sources. Additionally, experimenting with neural networks could uncover more complex patterns within the data.

Conclusion
We recommend the deployment of the optimized XGBoost model for Food Mart X. Its predictive accuracy, coupled with the strategic feature selection and hyperparameter tuning, makes it a

valuable tool for optimizing media campaign costs. The precision and recall metrics, in this case, suggest that the model can reliably predict campaign costs, supporting more informed budget allocation decisions.