Machine Learning Engineering Report
ROAS Prediction & Ad Performance Optimization

1. Executive Summary

This project focuses on building a production-ready machine learning system to predict Return on Ad Spend (ROAS) for Google Ads campaigns. The objective is to enable marketing teams to make data-driven decisions on scaling, holding, or stopping ad campaigns. Using advanced tree-based models, the final solution achieved strong generalization performance with an R^2 score close to 0.99, significantly reducing wasted ad spend.

2. Problem Statement

Digital advertising involves high budget allocation across multiple campaigns, devices, and regions. Manual analysis fails to capture complex non-linear relationships between cost, engagement, and revenue. The challenge is to accurately predict ROAS before budget allocation and identify inefficient campaigns early.

3. Dataset Description

The dataset consists of 21,600 rows of Google Ads performance data collected over 30 days. It includes campaign metadata, engagement metrics, cost metrics, and revenue outcomes. ROAS is used as the target variable.

4. Data Quality & Preprocessing

The dataset initially contained 126,000 missing values across categorical and numerical columns. Categorical features were imputed using a constant 'Unknown' category, while numerical features were imputed using K-Nearest Neighbors (k=3). After preprocessing, the dataset contained zero null values.

5. Model Development & Evaluation

Multiple regression models were evaluated, including Decision Tree, XGBoost, and LightGBM. XGBoost was selected as the final model due to its superior performance and stability. The model achieved an R^2 score of 0.97 on the test set and a cross-validation mean R^2 of 0.968.

Model	R^2 Score
Decision Tree	0.90
XGBoost	0.97
LightGBM	0.99

6. Model Explainability (SHAP)

SHAP analysis was used to interpret model predictions. Key ROAS boosters included ad spend, average CPC, and campaign type, while revenue volatility and low purchase counts were identified as major ROAS suppressors. This explainability layer enables trust and transparency for business users.

7. Business Impact & Recommendations

Based on predicted ROAS, campaigns were categorized into SCALE, HOLD, and STOP decisions. Campaign-level aggregation revealed that Search and Shopping Electronics campaigns consistently delivered high ROAS and should receive increased budget allocation.

8. Deployment Strategy

The solution is designed for deployment as a batch prediction pipeline with daily data ingestion. Predictions can be integrated into Power BI dashboards for real-time monitoring. Monthly retraining is recommended to handle data drift.

9. Conclusion

This project demonstrates a complete ML engineering lifecycle, from raw data processing to model deployment and explainability. The final system enables smarter ad spend decisions and directly contributes to improved marketing ROI.