Predicting Advertisement Clicks Using Logistic Regression

A Data-Driven Approach to Optimize Digital Marketing

Team Members: Aymar Donald Shemeza
Chris Lin
Dylan Liu
Laurie Ye

Executive Summary

Brief overview of the project

Key outcomes:

- Data preprocessing and feature engineering
- Model performance with a log loss of 0.426
- Business impact: Improved resource allocation and user engagement

Introduction

- Importance of predicting advertisement click-through rates
- Dataset:
  - 13 million entries
  - 24 categorical variables
  - Target variable: `click`
- Objective: Create a logistic regression model to minimize log loss and optimize predictions.

Data Processing

- **Steps Taken:**
  - Transformation of the `hour` column into `day_of_week` and `hour_of_day`
  - Encoding techniques:
    - One-hot encoding for low-cardinality variables
    - Frequency encoding for high-cardinality variables
  - Removal of ID-related and highly correlated variables

Model Building

- **Model Used:** Logistic Regression
- **Why Logistic Regression?**
  - Handles binary classification effectively
  - Scalable for large datasets
  - Works well with categorical variable encoding
- **Feature Selection:**
  - Backward elimination
  - P-value significance testing

Model Evaluation

- **Metric:** Log loss
  - Penalizes overconfident wrong predictions
  - Encourages calibrated probability estimates
- **Results:**
  - Log loss: 0.426 on validation data
  - 90%-10% split for training and validation

Business Impact

- Enables targeted advertising by predicting user interaction
- Helps allocate resources effectively
- Improves ad placement strategies and revenue generation

Future Improvements

- Explore advanced machine learning techniques:
    - Random Forest
    - Gradient Boosting
- Integration with real-time advertising platforms for dynamic updates

Conclusion

- Logistic regression model optimized for CTR prediction
- Data preprocessing and feature engineering significantly impact outcomes
- Model aligns with business goals to enhance digital marketing efforts

Thank you!!

Q&A