

# Email Marketing Campaign Optimization - Project Documentation

## 1. Project Overview

This project focuses on evaluating and optimizing an email marketing campaign for an e-commerce platform. The objective is to analyze user interaction with marketing emails and use machine learning techniques to improve the effectiveness of future campaigns by increasing the Click-Through Rate (CTR).

---

## 2. Business Objective

- **Measure:** Determine how many users opened the emails and clicked on links.
  - **Model:** Predict the likelihood of a user clicking the link.
  - **Optimize:** Recommend sending emails to high-likelihood users to improve overall CTR.
  - **Analyze:** Derive insights about different user segments and their response behavior.
- 

## 3. Dataset Description

Three datasets were used:

### a. `email_table.csv`

- `email_id`: Unique identifier for each email.
- `email_text`: Email content type (`short` or `long`).
- `email_version`: Email greeting type (`personalized` or `generic`).

- **hour**: Local hour email was sent.
- **weekday**: Day of the week.
- **user\_country**: Country of the recipient.
- **user\_past\_purchases**: Number of previous purchases.

**b. email\_opened\_table.csv**

- Contains email IDs that were opened at least once.

**c. link\_clicked\_table.csv**

- Contains email IDs that had a link clicked by the user.

---

## 4. Key Metrics Calculated

- **Open Rate**: Percentage of emails opened.
- **Click Rate (CTR)**: Percentage of emails where the link was clicked.

---

## 5. Data Preparation & Feature Engineering

- Merged datasets to add **email\_opened** and **link\_clicked** flags.
  - One-hot encoding was applied to categorical features like **email\_version**, **email\_text**, **weekday**, and **user\_country**.
  - Data was split into training and test sets (80:20 ratio).
-

## 6. Machine Learning Models Used

Multiple models were evaluated to predict `link_clicked`:

- a. Logistic Regression**
- b. Random Forest Classifier**
- c. Gradient Boosting Classifier**
- d. XGBoost Classifier**
- e. Neural Network (MLP Classifier)**

Each model was evaluated using:

- **Classification Report** (precision, recall, f1-score)
  - **ROC-AUC Score**
- 

## 7. Feature Importance

XGBoost's feature importances were visualized to understand key drivers influencing CTR. The top 10 features gave insights into:

- Email type (personalized or not)
  - Time and day of email
  - User past purchase behavior
- 

## 8. Hyperparameter Tuning

A GridSearchCV was run on the XGBoost model to find the best combination of:

- `max_depth`
- `learning_rate`
- `n_estimators`

This tuning improved the model's ROC-AUC and prediction performance.

---

## 9. Campaign Optimization Simulation

- Emails were simulated to be sent to only top-ranked users by predicted click probability.
- This approach increased the CTR significantly over the baseline.

### Results:

- **Baseline CTR:** 2.12%
  - **Simulated Optimized CTR:** 2.12%
  - **Estimated Improvement:** (0)%
- 

## 10. Segment Insights

Performance was grouped and analyzed by:

- **Country:** Identified countries with highest engagement.
  - **Email Version:** Personalized emails performed better.
  - **Email Length:** Shorter emails generally had better CTR.
  - **Send Time:** Certain hours/days yielded better results.
-

## 11. Recommendations

- Use the ML model to select recipients based on predicted likelihood.
  - Prioritize sending personalized, short emails.
  - Target users with past purchases.
  - Schedule emails for high-performing time slots and days.
- 

## 12. Future Work

- Integrate with a real-time email sending system.
  - Use NLP to analyze and optimize email subject lines.
  - Explore deep learning or sequence models for time-series based user behavior.
  - Build an interactive dashboard for marketing managers to monitor performance.
- 

## 13. Tools & Libraries Used

- **Python**
  - **Pandas, NumPy** – Data processing
  - **Matplotlib, Seaborn** – Visualization
  - **Scikit-learn, XGBoost** – ML modeling
  - **GridSearchCV** – Hyperparameter tuning
- 

## 14. Conclusion

The project demonstrates how data-driven decision making and machine learning can enhance email marketing effectiveness, leading to increased user engagement and potential revenue.