**Email Marketing Campaign – Final Report**

**Name**: Satyam Kumar
**Role**: Machine Learning Intern
**Project Title**: Email Campaign Optimization Using Data Science

---

### 1. Objective

The goal of this project is to analyze an e-commerce email campaign and propose a data-driven strategy to optimize future email sends using machine learning. The objectives include:

- Measuring open and click-through rates (CTR)

- Building a prediction model for link clicks

- Estimating the improvement using predictive targeting

- Analyzing user behavior across different segments

---

### 2. Dataset Description

The project uses three datasets:

◆ **email_table.csv**

Contains metadata about emails:

- email_text: short or long

- email_version: personalized or generic

- hour, weekday: when the email was sent

- user_country: user's country

- user_past_purchases: user's previous purchases

◆ **email_opened_table.csv**

- Email IDs that were **opened** by users.

◆ **link_clicked_table.csv**

- Email IDs where users **clicked** on a link.

---

## 3. Campaign Performance Summary

After processing and merging data:

- **Total Emails Sent**: 20,000

- **Open Rate**: **10.35%**

- **Click-Through Rate (CTR)**: **2.12%**

These results show relatively low engagement, especially in CTR, making it a good case for optimization.

---

## 4. Predictive Modeling

A **Random Forest Classifier** was trained to predict if a user would click on the link.

◆ **Features Used:**

- email_text, email_version

- hour, weekday

- user_country, user_past_purchases

**Model Performance**

|  | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.98 | 1.00 | 0.99 | 19576 |
| 1 | 0.07 | 0.01 | 0.02 | 424 |
| accuracy | | | 0.98 | 20000 |
| macro avg | 0.53 | 0.50 | 0.50 | 20000 |
| weighted avg | 0.96 | 0.98 | 0.97 | 20000 |

- **Accuracy**: 98%

- **Note**: The model suffers from **class imbalance** (only ~2% clicked), resulting in poor recall for positive class.

---

**5. CTR Improvement Using Model**

By targeting only the **top 20% most likely to click**, based on predicted probability:

- **CTR in Top 20% Group**: **9.25%**

- **Original CTR**: **2.12%**

- **Estimated Improvement**: ~**4.36×**

This proves machine learning can significantly boost the effectiveness of future campaigns with selective targeting.

---

**6. Segment Analysis Insights**

➤ **Click Rate by Email Text:**

- **Short Emails**: **2.39%**

- **Long Emails**: **1.85%**

✅ **Insight**: Shorter emails drive better user engagement.

---

➤ **Click Rate by Email Version:**

- **Personalized**: **2.73%**

- **Generic**: **1.51%**

**Insight**: Personalization improves CTR significantly.

---

➤ **Click Rate by Weekday:**

| Day | CTR (%) |
| --- | --- |
| Monday | 2.29% |
| Tuesday | 2.49% |
| Wednesday | 2.76% |
| Thursday | 2.44% |

| Day | CTR (%) |
|---|---|
| Friday | 1.40% |
| Saturday | 1.78% |
| Sunday | 1.68% |

**Insight**: Best days to send emails are **Wednesday and Tuesday**.

---

➤ **Click Rate by Purchase History:**

| Purchases | CTR (%) |
|---|---|
| 0 | 0.05% |
| 1 | 1.12% |
| 2–5 | 1.85% |
| 6–10 | 3.65% |
| 10+ | 6.90% |

**Insight**: Users with **higher past purchases** are far more likely to click.

---

## 7. Conclusion

The email campaign had a modest performance with an overall CTR of 2.12%. However, by leveraging machine learning, we can:

- Increase CTR to over **9%** by targeting only high-probability users
- Focus on **personalized**, **short emails**, sent mid-week
- Prioritize users with **higher purchase history** and engagement potential

This analysis demonstrates a powerful use case of machine learning for **marketing optimization** and **ROI improvement**.

---