Build a Marketing Budget Optimization Model for B2C sales Project Overview

Overview

In today's competitive market, a company specializes in B2C sales, offering top-quality Data Science and Data Engineering courses. To attract customers, we invest a significant amount of money in various marketing campaigns across different channels, including social media, email, and search advertising. However, our marketing team faces a challenge when it comes to optimizing the allocation of our marketing budget across these channels, aiming to maximize revenue and return on investment (ROI).

To tackle this challenge, our project aims to develop a data-driven approach that leverages the power of machine learning algorithms. By analyzing the data and utilizing predictive models, we can empower our marketing team with valuable insights to make informed decisions about budget allocation. Our objective is to maximize revenue and ROI by strategically distributing our marketing budget across the most promising channels.

With the implementation of this data-driven approach, our company can optimize marketing campaign performance and enhance resource allocation. By accurately predicting revenue and ROI for each marketing channel, we can ensure that our marketing budget is efficiently utilized, resulting in higher returns on our investments. This project will not only empower our marketing team but also strengthen our overall business performance in the highly competitive market landscape.

Business Impact of Marketing Budget Optimization:

- Increased Product Conversions: By optimizing the marketing budget and targeting the right users through the appropriate channels and assets, we can significantly improve our product conversions. This means that more users who are genuinely interested in our offerings will engage with our marketing campaigns and take the desired actions, resulting in a higher conversion rate.

- Increased Revenue: The improved conversions mentioned above directly translate into increased revenue. When we effectively target users who are more likely to engage with our ads and make purchases, the probability of generating more sales and revenue significantly rises. For instance, if we identify that a particular user is highly active on Instagram, directing our marketing efforts

towards that channel can lead to higher click-through rates and more added products to their cart, ultimately boosting revenue.

- Improved Budget Allocation: Optimization of the marketing budget helps avoid wasting valuable resources on non-efficient channels. By analyzing data and leveraging machine learning algorithms, we can identify which channels and strategies deliver the highest returns on investment. This enables us to allocate our budget more wisely, focusing on the channels that yield the best results and avoiding unnecessary expenses on less productive avenues.

- Improved Customer Acquisition Cost (CAC): By targeting customers through the most effective channels, we can improve our Customer Acquisition Cost (CAC). When we reach the right customers with the right message, we increase the likelihood of acquiring them as customers at a lower cost. Moreover, targeting customers through their preferred channels often leads to better customer retention and repeat rates, further optimizing the CAC and enhancing the overall profitability of our marketing efforts.

Aim

The aim of this project is to leverage data-driven approaches and machine learning algorithms to optimize the allocation of the marketing budget across various channels and to empower the marketing team with informed decisions for efficient budget allocation and strategic targeting.

Data Description

The dataset contains the following columns:

- Lead Id: A unique identifier assigned to each lead in the dataset.
- Lead Owner: The internal salesperson associated with the lead.
- Interest Level: Indicates the level of interest expressed by the lead. This information is entered manually.
- The lead created: The date when the lead was created.
- Lead Location(Auto): The location of the lead, is automatically detected.
- Creation Source: The source from which the lead was generated.
- Next activity: The date scheduled for the next activity with the lead.
- What do you do currently?: Describes the current profile or occupation of the lead.

- What are you looking for in a Product?: Specifies the specific requirements or expectations of the lead regarding the product.
- Website Source: The source from which the lead visited the website.
- Lead Last Update time: The timestamp of the last update made to the lead's information.
- Marketing Source: The marketing source through which the lead was acquired.
- Lead Location(Manual): The manually entered location of the lead.
- Demo Date: The date scheduled for a product demonstration with the lead.
- Demo Status: Indicates the status of the demo booked with the lead.
- Closure date: The date when the lead was successfully closed or converted into a customer.
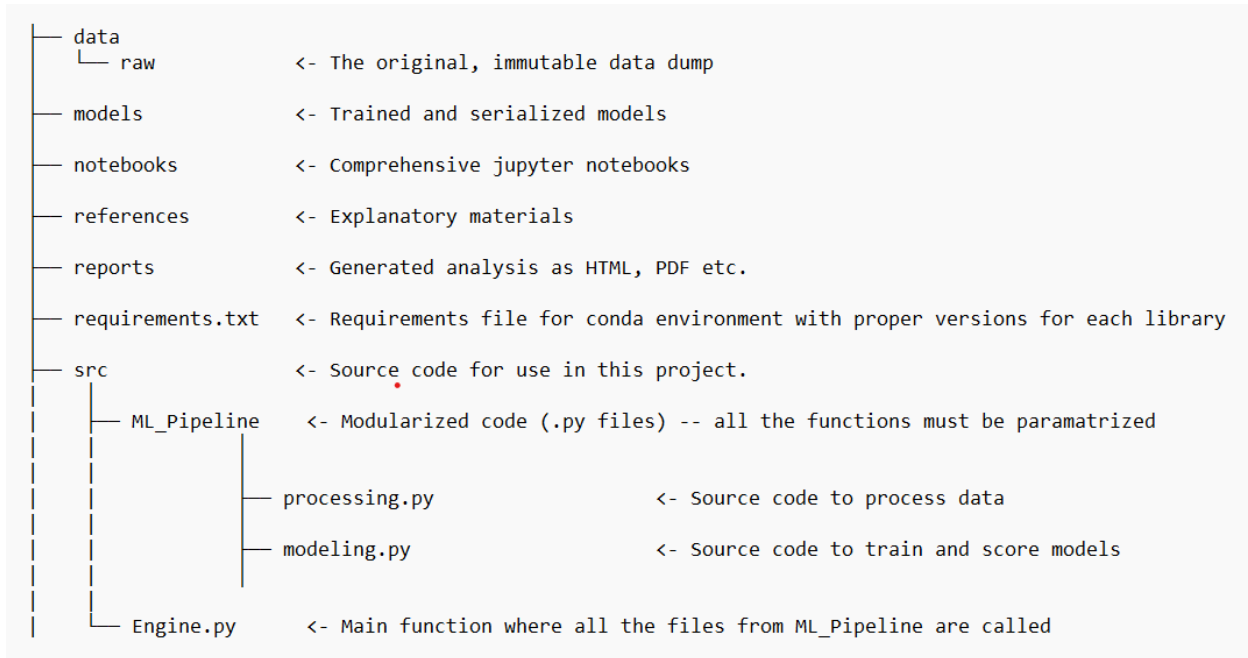
Tech Stack
➔ Language: Python
➔ Libraries: pandas, numpy, matplotlib, scikit-learn, xgboost, lightgbm

Approach

- Exploratory Data Analysis (EDA):
    - Understand the features and their relationships with target variables
    - Check for missing or invalid values and their imputation

- Data Preprocessing:
    - Encode the variables using label encoding
    - Split the dataset into training and testing sets

- Model Building and Testing:
    - Random Forest
    - Light Gradient Boosting
    - Extreme Gradient Boosting

Modular code overview:

```
├── data
│   └── raw            <- The original, immutable data dump
├── models             <- Trained and serialized models
├── notebooks          <- Comprehensive jupyter notebooks
├── references         <- Explanatory materials
├── reports            <- Generated analysis as HTML, PDF etc.
├── requirements.txt   <- Requirements file for conda environment with proper versions for each library
├── src                <- Source code for use in this project.
│   │
│   ├── ML_Pipeline    <- Modularized code (.py files) -- all the functions must be paramatrized
│   │
│   │           ├── processing.py          <- Source code to process data
│   │           ├── modeling.py            <- Source code to train and score models
│   │
│   └── Engine.py      <- Main function where all the files from ML_Pipeline are called
```

Once you unzip the modular_code.zip file, you can find the following folders.

1. data
2. models
3. src
4. reports
5. references
6. notebooks
7. requirements.txt
8. readme.md

1. The requirements.txt file has all the required libraries with respective versions. Kindly install the file using the command **pip install -r requirements.txt**

2. **All the instructions for running the code are present in readme.md file**

Project Takeaways

1. Understanding the importance of data-driven decision-making in marketing budget allocation.
2. Leveraging machine learning algorithms to optimize marketing budget allocation and maximize revenue and ROI.
3. Identifying the significance of targeting the right users through the appropriate channels and assets for improved conversions.
4. Data preprocessing, encoding, and splitting into training and testing sets
5. Recognizing the impact of targeted marketing on customer acquisition costs and improving customer retention.
6. Understanding and Implementing Tree-based models like Random Forest, Light Gradient Boosting, and Extreme Gradient Boosting to effectively predict interest.
7. Improving overall business performance by aligning marketing efforts with customer preferences and behavior.