Solution Design Document

Packt Group Project: Customer prioritization for Marketing

1. About Project

1.1 Project Title

Customer prioritization for the sales team.

1.2 Problem Statement

Thousands of potential customers visit our website every day for a free horoscope report, some of which actually result in a conversion.

Due to limited human resources, we are unable to reach out to each one of those thousands of potential customers each day. In addition to being infeasible for us, it is probably not necessary either.

To help filter the long list into something manageable by the sales team, we came up with a baseline model that prioritizes the customers we reach out to each day. The baseline model was built in a hurry without any serious data analysis and it is just a static formula taking as input certain values generated from browsing sessions.

Over the years, we have found that a majority of the potential customers we reach out to do not result in an immediate conversion.

We want to use data and technology to maximize conversions from our contacts each day.

1.3 Expected Solution

The solution should consist of the following:

a) A report containing key summaries and insights from historical data.

b) A data science product using an ML model to prioritize the potential list of customers.

   The predictions report (filename format: top_250_report_ddmmyyyy.csv) should be limited to the top 250 customers with the following fields:
   customer_id, conversion_probability

   The model needs to consider only those customers who visited our website within the last 24 hours.

c) An evaluation report showing at least 50% increase in the conversions when compared with the baseline model (over a period of 7 days).

1.4 Project Participants

Owner : Rutuja Yerunkar

Lead / Instructor : Neeraj Garg

Data Scientists / Developers: Reet Batra, Rishav Sharma

1.5 Project Schedule

Duration: 10 working days (2 calendar weeks)

Start Date: 10th September, 2021

End Date: 23rd September, 2021

1.6 GitHub Repo

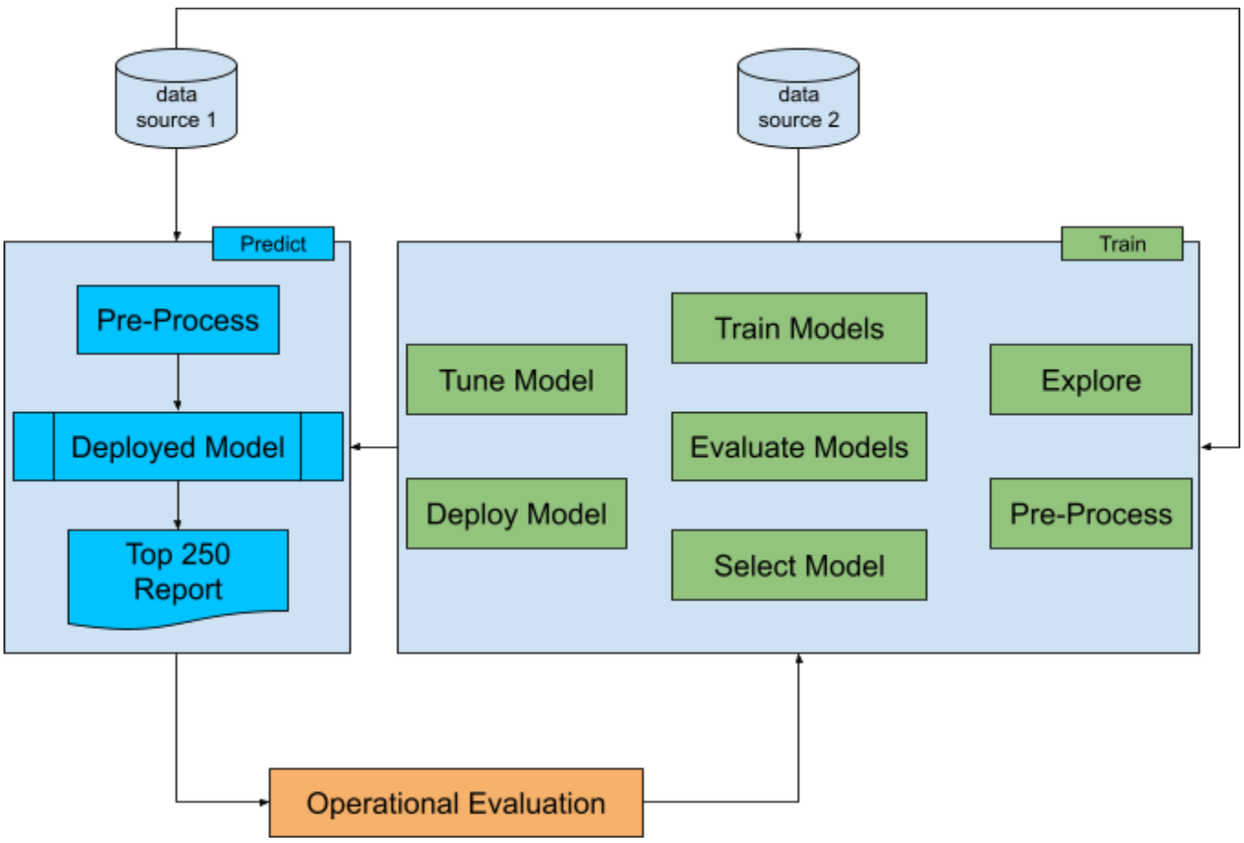
https://github.com/TeamEpicProjects/Customer-Prioritization-for-Marketing

2. Solution Design

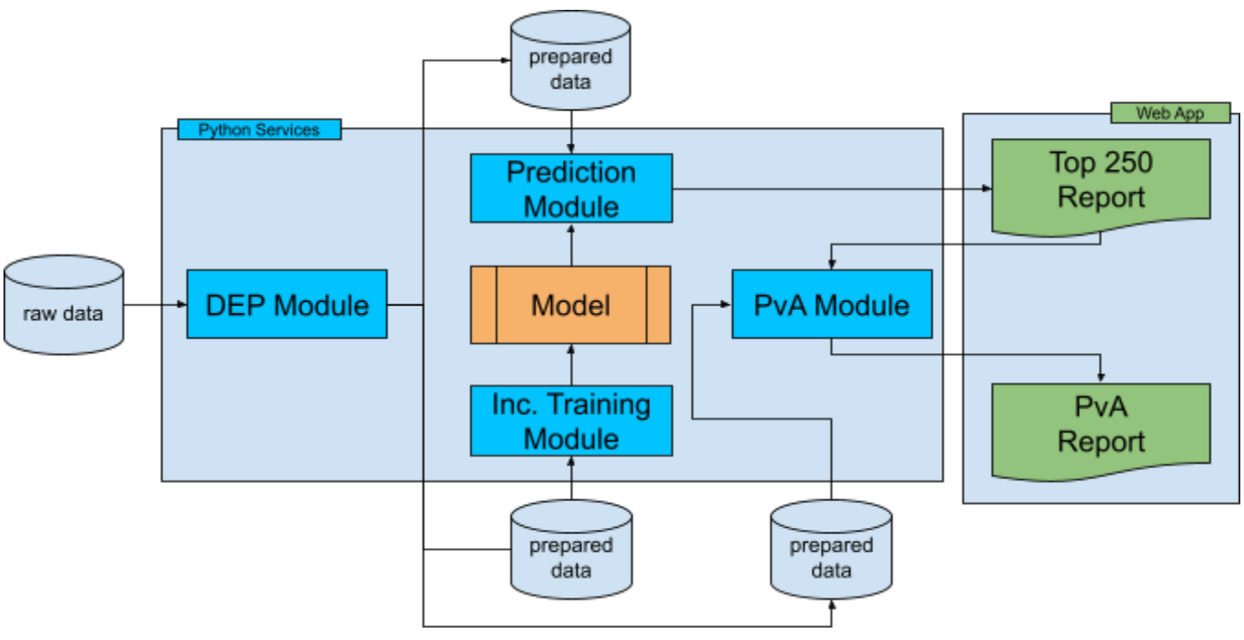
2.1 Technology Stack

Data Science: Python, Pandas, Seaborn, Scikit-learn, Imblearn, Joblib
Web App : Python, Flask, HTML

2.2 Development Process



2.3 High-Level Architecture (solution deployment)



2.4 Module Specifications

2.4.1 Data Extraction & Preparation Module

Purpose:

Extract and provide prepared data to other modules.

Input:

Raw datasets (multiple csv files), requesting module identifier.

Output:

Prediction module: extract and prepare data from previous day's visits.
Incremental Training Module: extract and prepare data from from -3 to -1 days results.
Predicted vs. Actual Module: extract and prepare data from -3 to -1 days results.

2.4.2 Prediction Module

Purpose:

Use the model on previous day's data to predict today's propensities.

Input:

Model, prepared data.

Output:

Top 250 report.

2.4.3 Incremental Training Module

Purpose:

Daily model updation, using results from the previous 3 days to improve the model.

Input:

Prepared data.

Output:

Updated model.

2.4.4 Predicted vs. Actual Module

Purpose:

Daily model evaluation, using results from the previous 3 days to compare predicted vs. actual sales.

Input:

Top 250 Report from 3 days ago, prepared data.

Output:

PvA report.

2.4.5 Web App

Purpose:

Provide a user interface to the solution for the sales team and managers to use.

Input:

Report type: Prediction report or PvA report, report date (for prediction report the range of selectable dates should include current date, but for PvA report it should exclude current and last 3 days).

Output:

Top 250 report or PvA report for the selected date.