

Project 7

Propensity (Propensity Model to identify how likely certain target groups customers respond to the marketing campaign)

Problem Statement:

Are you aware of what, when, and why your customers will make a purchase? Many businesses undertake an intense pursuit to discover these answers, dedicating valuable resources to data-driven campaigns and high-cost strategies - yet the actual outcomes often remain elusive and disappointing.

Customer information is considered to be a valuable asset, however its true worth can only be established when it is used. Many companies have large collections of data that appear to be impressive, but upon further examination, they may contain outdated or unimportant information.

Propensity modeling is a method that aims to forecast the chance that individuals, leads, and customers will engage in specific actions. This method uses statistical analysis which takes into account all the independent and confounding factors that impact customer behavior.

Suppose you are working for a company as a Data Scientist. Your company is commissioned by an insurance company to develop a tool to optimize their marketing efforts.

This project is aimed at building a **propensity model to identify potential customers**.

Data:

The insurance company has provided you with a historical data set (train.csv). The company has also provided you with a list of potential customers to whom to market (test.csv). From this list of potential customers, you need to determine yes/no whether you wish to market to them.

(Note: Ignore any additional columns available other than the listed below in the table)

Type	Name	Description
Input Variables	custAge	The age of the customer (in years)
Input Variables	profession	Type of job
Input Variables	marital	Marital status
Input Variables	schooling	Education level
Input Variables	default	Has a previous defaulted account?
Input Variables	housing	Has a housing loan?
Input Variables	loan	Has a personal loan?
Input Variables	contact	Preferred contact type
Input Variables	month	Last contact month
Input Variables	day_of_week	Last contact day of the week
Input Variables	campaign	Number of times the customer was contacted
Input Variables	pdays	Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
Input Variables	previous	Number of contacts performed before this campaign and for this client
Input Variables	poutcome	Outcome of the previous marketing campaign
Input Variables	emp.var.rate	Employment variation rate - quarterly indicator
Input Variables	cons.price.idx	Consumer price index - monthly indicator
Input Variables	cons.conf.idx	Consumer confidence index - monthly indicator
Input Variables	euribor3m	Euribor 3 month rate - daily indicator
Input Variables	nr.employed	Number of employees - quarterly indicator
Input Variables	pmonths	Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
Input Variables	pastEmail	Number of previous emails sent to this client
Target Variables	responded	Did the customer respond to the marketing campaign and purchase a policy?

Your focus in this project should be on the following:

The following is recommendation of the steps that should be employed towards attempting to solve this problem statement:

- 📖 **Exploratory Data Analysis:** Analyze and understand the data to identify patterns, relationships, and trends in the data by using Descriptive Statistics and Visualizations.
- 📖 **Data Cleaning:** This might include standardization, handling the missing values and outliers in the data.
- 📖 **Dealing with Imbalanced data:** This data set is highly imbalanced. The data should be balanced using the appropriate methods before moving onto model building.
- 📖 **Feature Engineering:** Create new features or transform the existing features for better performance of the ML Models.
- 📖 **Model Selection:** Choose the most appropriate model that can be used for this project.
- 📖 **Model Training:** Split the data into train & test sets and use the train set to estimate the best model parameters.
- 📖 **Model Validation:** Evaluate the performance of the model on data that was not used during the training process. The goal is to estimate the model's ability to generalize to new, unseen data and to identify any issues with the model, such as overfitting.
- 📖 **Model Deployment:** Model deployment is the process of making a trained machine learning model available for use in a production environment.

Timeline

We expect you to do your best and submit a solution within 2 weeks.

Deliverables

Please share the following deliverables in a zip file.

- ◆ A report (PDF) detailing:
- ◆ Description of design choices and Performance evaluation of the model
- ◆ Discussion of future work
- ◆ The source code used to create the pipeline

Tasks/Activities List

Your code should contain the following activities/Analysis:

- ◆ Collect the time series data from the CSV file linked here.
- ◆ Exploratory Data Analysis (EDA) - Show the Data quality check, treat the missing values, outliers etc if any.
- ◆ Get the correct datatype for date.
- ◆ Balancing the data.
- ◆ Feature Engineering and feature selection.
- ◆ Train/Test Split - Apply a sampling distribution to find the best split.
- ◆ Choose the metrics for the model evaluation
- ◆ Try multiple classification models and choose the best one.
- ◆ Model Selection, Training, Predicting and Assessment
- ◆ Hyperparameter Tuning/Model Improvement
- ◆ Please add a column to the testingCandidate.csv file. In this column, for each observation indicate a 1 (yes) or a 0 (no) whether you wish to market to that candidate.
- ◆ Model deployment plan.

Success Metrics

Below are the metrics for the successful submission of this case study.

- ✦ The accuracy of the model on the test data set should be $> 85\%$ (Subjective in nature)
- ✦ Add methods for Hyperparameter tuning.
- ✦ Perform model validation.

Bonus Points

- ✦ You can package your solution in a zip file included with a README that explains the installation and execution of the end-to-end pipeline.
- ✦ You can demonstrate your documentation skills by describing how it benefits our company.

Data

The dataset for this project can be accessed by clicking the link provided below.

[Propensify.zip](#)