

Tourism Purchase Prediction Project

Project Session

Objective

- Explore and visualize the dataset.
- Preprocess the data i.e. treating data entry error, missing values, or outliers
- Build ensemble models - bagging, boosting and stacking to predict whether a person will purchase the newly introduced travel package or not.
- Identify most important features of the model.
- Generate a set of insights and recommendations that will help the business.

Problem Statement

You are a Data Scientist for a tourism company named "Visit with us". The Policy Maker of the company wants to enable and establish a viable business model to expand the customer base.

A viable business model is a central concept that helps you to understand the existing ways of doing the business and how to change the ways for the benefit of the tourism sector.

One of the ways to expand the customer base is to introduce a new offering of packages.

Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking at the data of the last year, we observed that 18% of the customers purchased the packages.

However, the marketing cost was quite high because customers were contacted at random without looking at the available information.

The company is now planning to launch a new product i.e. Wellness Tourism Package. Wellness Tourism is defined as Travel that allows the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.

However, this time company wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

You as a Data Scientist at "Visit with us" travel company has to analyze the customers' data and information to provide recommendations to the Policy Maker and Marketing Team and also build a model to predict the potential customer who is going to purchase the newly introduced travel package.

Data Dictionary

- CustomerID: Unique customer ID
- ProdTaken: Product taken (Yes/No)
- Age: Age of a customer
- TypeofContact: How customer was contacted (Company Invited or Self Inquiry)
- CityTier: City tier of a customer
- DurationOfPitch: Duration of pitch by a salesperson to a customer
- Occupation: Occupation of a customer
- Gender: Gender of a customer
- NumberOfPersonVisited: Total number of person with a customer
- NumberOfFollowups: Total number of follow ups has been done by salesperson after the pitch
- ProductPitched: Product pitched by salesperson
- PreferredPropertyStar: Preferred hotel property rating by customer
- MaritalStatus: Marital status of customer
- NumberOfTrips: Average number of trip in a year by customer
- Passport: Customer has passport or not
- PitchSatisfactionScore: Sales pitch satisfactory score
- OwnCar: Customers owns a car flag
- NumberOfChildrenVisited: Total number of children visiting with customer
- Designation: Designation of customer in current organization
- MonthlyIncome: Gross monthly income of customer

Step to Follow

- Import the libraries
- Load and overview the data
- Find the data types and summary of different columns
- Calculate the percentage of missing values in each column
- Univariate analysis - Plot different graphs for all variables and provide insights
- Bivariate analysis - Compare all attributes visually to check for relationships that can be exploited by making plots and provide insights
- Check for correlated variables
- Impute missing values
- Prepare the data for modelling
- Build bagging, boosting, and stacking classifiers choose the best metric to evaluate it
- Hyperparameter tuning for models using GridSearchCV
- Compare the performance of all the models and give business recommendations.

Business Recommendations

- Our analysis shows that very few customers have passports and they are more likely to purchase the travel package. The company should customize more international packages to attract more such customers.
- We have customers from tier 1 and tier 3 cities but very few from tier 2 cities. Company should expand its marketing strategies to increase the number of customers from tier 2 cities.
- We saw in our analysis that people with higher income or at high positions like AVP or VP are less likely to buy the product. The company can offer short term travel packages and customize the package for higher income customers with added luxuries to target such customers.
- When implementing a marketing strategy, external factors, such as the number of follow ups, time of calling, should also be carefully considered as our analysis shows that the customers who have been followed up more are the ones buying the package.
- After we identify a potential customer, the company should pitch packages as per the customer's monthly income, for example, do not pitch king packages to a customer with low income and such packages can be pitched more to the higher income customers.
- We saw in our analysis that young and single people are more likely to buy the offered packages. The company can offer discounts or customize the package to attract more couples, families, and customers above 30 years of age.

greatlearning
Power Ahead

Happy Learning !

