Tech Mahindra Growth Factories Ltd.

Internship Task

Task Description

Analysis of a Hair Salon establishment

No-shows are a big problem for hair salons the same as airlines and medical facilities. Excessive no-shows increase costs and wait times for businesses and all other customers alike.

This is a dataset that can be used to predict appointment no-shows. The data includes bookings and cancellation information to determine whether a given booking resulted in a "no-show" wherein the client either didn't show up at all or canceled the appointment within 48 hours of the planned booking (i.e., an out-of-policy cancellation).

The description of dataset attributes are as follows:

- Booking index.
- book_tod: The booking time of day.
- book_dow: The booking day-of-week.
- book_category: The booked service category (COLOR or STYLE)
- book_staff: The staff member to provide the service.
- last_category: The client's last booked service category before the current booking or today whichever is greater.
- last_staff: The staff member who provided the client's last service before the current booking or today whichever is greater.
- last_day_services: The number of services provided to the client on their last visit before the current booking or today whichever is greater.
- last_receipt_tot: The amount paid by the client on their last visit before the current booking or today whichever is greater.
- last_dow: The day-of-week of the client's last booking before before the current booking or today whichever is greater.
- last_tod: The time-of-day of the client's last booking before the current booking or today whichever is greater.

- last_noshow: Did the client no-show on their last booking before the current booking or today whichever is greater? (0 - no, 1 - yes)
- last_prod_flag: Did the client buy a retail product on their last booking before the current booking or today whichever is greater? (0 - no, 1 - yes)
- last_cumrev: The client's cumulative service revenue as of their last booking before the current booking or today whichever is greater.
- last_cumbook: The client's cumulative number of bookings as of their last booking before the current booking or today whichever is greater.
- last_cumstyle: The client's cumulative number of STYLE bookings as of their last booking before the current booking or today whichever is greater.
- last_cumcolor: The client's cumulative number of COLOR bookings as of their last booking before the current booking or today whichever is greater.
- last_cumprod: The client's cumulative number of bookings with retail product purchases as of their last booking before the current booking or today whichever is greater.
- last_cumcancel: The client's cumulative number of appointment cancellations as of their last booking before the current booking or today whichever is greater.
- last_cumnoshow: The client's cumulative number of no-shows as of their last booking before the current booking or today whichever is greater.
- noshow: Did the client no-show or execute an out-of-policy cancellation for this booking? (0 - no, 1 - yes)
- recency: The number of days since the client's last booking before the current booking or today whichever is greater.

.

Tasks

Requirements:

1. Perform exploratory Data Analysis and come up with actionable insights from each plot and table. Do data wrangling if required.

2. Predict if the client did a no-show or an out-of-policy cancellation for a booking. Please use Gridsearch Cross validation for any algorithms you use.

Please perform the entire analysis in Python