Clicks & Bookings Forecast

Problem

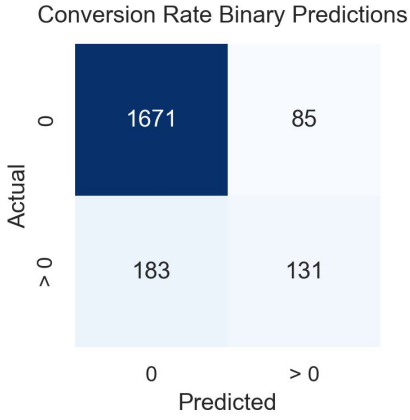
Forecast the conversion rate (ratio between bookings and clickouts) in 2023.08.11 for each hotel-advertiser couple present in dataset.

Solution

After gaining some insights through EDA, the forecast problem is tackled in two steps. In the first one, an XGBoost binary classification model will predict if the conversion rate is going to be zero (0) or higher (1). In the second step, an XGBoost regression model will improve the strictly positive conversion rate predictions.

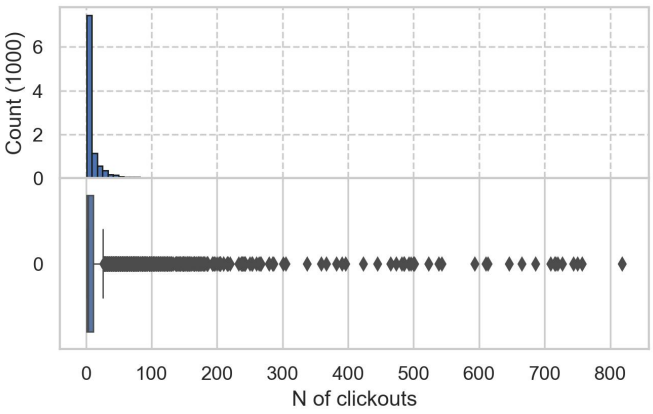
Key Insights

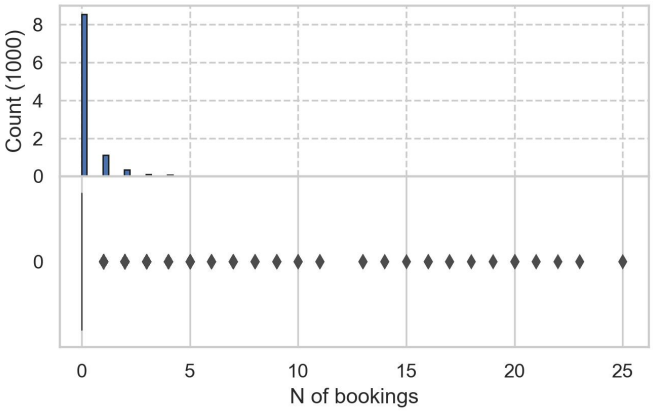
1. Top 5 hotels' IDs by bookings: 73, 36, 26, 33, 35.
2. Top 5 cities' IDs by bookings: 34, 3, 51, 67, 46.
3. Top 5 advertisers' IDs by bookings: 5, 39, 37, 24, 1.
4. N of bookings $\gtrsim 5$ when N of clickouts $\gtrsim 200$.
5. The best XGBoost classification model created, has an accuracy and a recall of 87.1%, and a precision of 85.7% on the test dataset. The model requires only three predictors: hotel ID, advertiser ID, and city ID.
6. The best XGBoost regression model created, has a R^2 score of 46.9% on the test dataset and requires only two predictors: hotel ID, and advertiser ID.
7. Both models are therefore time-independent.
8. The highest feature importance lays on the hotel ID, followed by the advertiser ID, and finally the city ID.

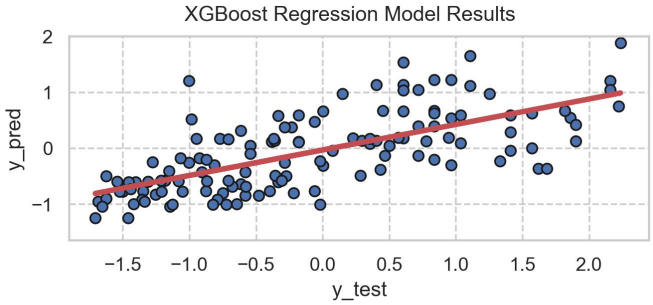


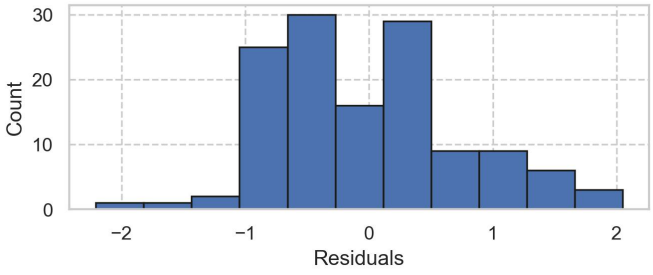
Next Steps

1. Apply oversampling techniques to reduce the class imbalance (null, non-null conversion rate).
2. Apply hyperparameter tuning to the models via GridSearchCV to improve the models' metrics.
3. Test the performance of a Neural Network model.









03.03.2024, V. Ganci, GitHub Portfolio