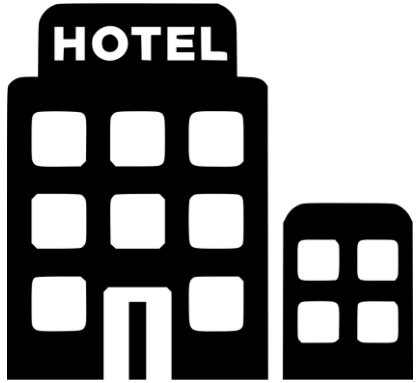

Hotel Booking Cancellation Predictions using a Machine Learning

Problem

- Reservations for accommodations are often made well ahead of time at hotels.
- Nonetheless, this practice exposes hotels to potential risks. They must commit to providing rooms to customers who fulfill their bookings, but they also face the financial burden of unoccupied rooms in cases of cancellations or no-shows by customers.
- The act of canceling bookings has an adverse impact on both the revenue of hotels and their ability to make precise predictions about future occupancy levels.

Model

Model	**Model Type**	**Hyper Parameters**	**Selection Method**
Logistic Regression	Regresion	Maximum iterations = 10000	10 fold cross validation
Random Forest	Tree-Based	n_estimators = 100, criterion = “entropy”	minimize cross validation loss metrics by random search
XG Boost	Ensemble	Learning rate = 0.300 Max Depth = 6	Minimize cross validation loss metrics by random search
kNN	Lazy-Learning	K = 2	Minimize error rate of ks from 1 to 40
Ensemble classifier	Ensemble	Voting method = majority	--

Conclusions

- 1. It's possible to anticipate hotel booking cancellations with an accuracy exceeding 80%.
- 2. When it comes to modeling hotel cancellations, ensemble methods prove to be superior compared to alternative approaches.
- 3. Among the critical factors in these models, deposit type, lead time to booking, average daily rate, and arrival date take precedence.
- 4. The most optimal performance is achieved by a composite model that merges kNN, logistic regression, random forest, and XG Boost. This amalgamated model is further refined by segmenting it based on whether the hotel is situated in a resort area or a city setting.