Comparing Machine Learning Algorithms for Predicting Cancellation of Hotel Bookings

Kevin de Haan
Edmonton, Canada
Email: kdehaan@ualberta.ca

***Abstract*—Hotels around the world have dealt with cancellations since their inception. Reservations require that a space be set aside, and cancellations mean that even the most consistently popular vacation destinations are often floating unused rooms and wasted space. Being able to predict whether a customer is likely to cancel ahead of time allows establishments to operate on significantly tighter occupancy margins, improving their efficiency and freeing up more space for customers who will follow through on their plans. The nature of this binary classification problem is well suited to simple machine learning strategies; however, more advanced techniques still have the opportunity to clearly demonstrate their ability.**

***Index Terms*—Machine Learning; Hotel Booking Cancellation; K-nearest Neighbors; Linear SGD; Multi-layer Perceptron; Scikit-learn**

I. Introduction

II. Problem Formulation

A. Input and Output

B. The Data

III. Approaches and Baselines

For the purposes of this experiment, three machine learning approaches will be evaluated:

- ***k*-nearest Neighbors** (KNN)
- A **Linear Classifier** trained using **Stochastic Gradient Descent** (SGD)
- A **Multi-layer Perceptron** (MLP)

In addition to the machine learning techniques employed, a baseline must be established. As this is a k-category classification problem, each approach will be compared to two naive baseline predictors:

- **Majority Guess** will always predict that a booking will not be cancelled.
- **Uniform Random Guess** will predict each possible classification category (i.e, cancelled and not cancelled) at an equal rate.

IV. Evaluating Performance

Because of the significant class imbalance ratio, evaluating the performance of the different predictors will be performed using the *Area Under the Precision-Recall Curve* (AUPRC)

V. Results

VI. Conclusion

References

[1] N. Antonio, A. de Almeida, and L. Nunes, "Hotel booking demand datasets," *Data in Brief*, vol. 22, pp. 41–49, Feb. 2019. [Online]. Available: https://doi.org/10.1016/j.dib.2018.11.126

[2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.