# Learning to Rank with Random Forest: a case study in hostel reservations*

Carolina Macedo Moreira[1], Carlos Soares[1], and Isabel Portugal[2]

[1] Faculdade de Engenharia da Universidade do Porto
[2] Hostelworld Group LLC

**Abstract.** Learning to rank is the application of supervised machine learning in the construction of ranking models for information retrieval systems. Hostelworld Services Portugal uses this method to improve the ranking of properties on their listing page, to improve their profits, as well as to boost their costumers satisfaction.
In the online hospitality industry, user search is one of the key factors. Search filters allow the users to easily see which properties they're interested in. Therefore, the improvement of these filters is of maximum importance.
We developed a random forest based approach, with several variations, one focused exclusively on clicks, another on bookings, and another combining the two. We compared this approach to a simple baseline using the static ranking and to the current method, achieving positive results. Among the different variants that were tested, there are no significant differences.

**Keywords:** Machine Learning · Learning to Rank · Information Retrieval.

## 1 Introduction

### 1.1 Hostelworld and Problem

Hostelworld is an online travel agency where you can book hostels from all over the world. They have properties in over 170 countries, with over 16,500 hostels listed on the website. As a company, they pride themselves on their customer service and extensive review database. They would like to optimize their listing page results.

### 1.2 Hypothesis

- a random forest approach is better than the baselines to predict click rankings;
- a random forest approach is better than the baseline to predict bookings;

* Done in partnership with Hostelworld Group LLC.

– the ability of the proposed approach to predict the bookings improves when the ties are broken with the predicted ranking of clicks;
– in case of ties in the rank predicted by the random forest model, the current approach is better for breaking them, since the method is more sophisticated.

## 2 Results

### 2.1 Baseline Results

| | Clicks | | Bookings | |
|---|---|---|---|---|
| | MAP | NDCG | MAP | NDCG |
| Static | 0.1694 | 0.4090 | 0.0308 | 0.3493 |
| Dynamic | 0.3909 | 0.5064 | 0.0991 | 0.2773 |

**Table 1.** Baseline Results

### 2.2 Results

| | MAP | NDCG |
|---|---|---|
| Baseline | 0.3688 | 0.3392 |
| Model | 0.5686 | 0.9839 |
| Improvement | 54% | 273% |

| | MAP | NDCG |
|---|---|---|
| Baseline | 0.1091 | 0.1489 |
| Model | 0.2175 | 0.6186 |
| Model Tie-breaker Clicks | 0.2180 | 0.6204 |
| Improvement Regular Model | 101% | 356% |
| Improvement Clicks Model | 102% | 356% |

**Table 2.** Results for clicks and bookings

## 3 Conclusions

Upon completing the project, the following conclusions were reached; A random forest approach is indeed better than the baselines to predict click rankings and a random forest approach is also better than the baseline to predict bookings. However, the ability of the proposed approach to predict the bookings doesn't particularly improve when the ties are broken with the predicted ranking of clicks, nor, in case of ties in the rank predicted by the random forest model, is the current approach particularly better for breaking them.