150 Successful Machine Learning Models: 6 Lessons Learned

Gülsen Zeynep Erdoğan

Emre Balaban

Outline

1. Introduction
2. Inception: Machine Learning as a Swiss knife for Product Development
3. Modeling: Offline model performance is just a health check
4. Modeling : Before solving a Problem, design it
5. Deployment: Time is Money
6. Monitoring: Unsupervised Red Flags
7. Evaluation: Experiment design sophistication pays
8. Conclusion

1. INTRODUCTION

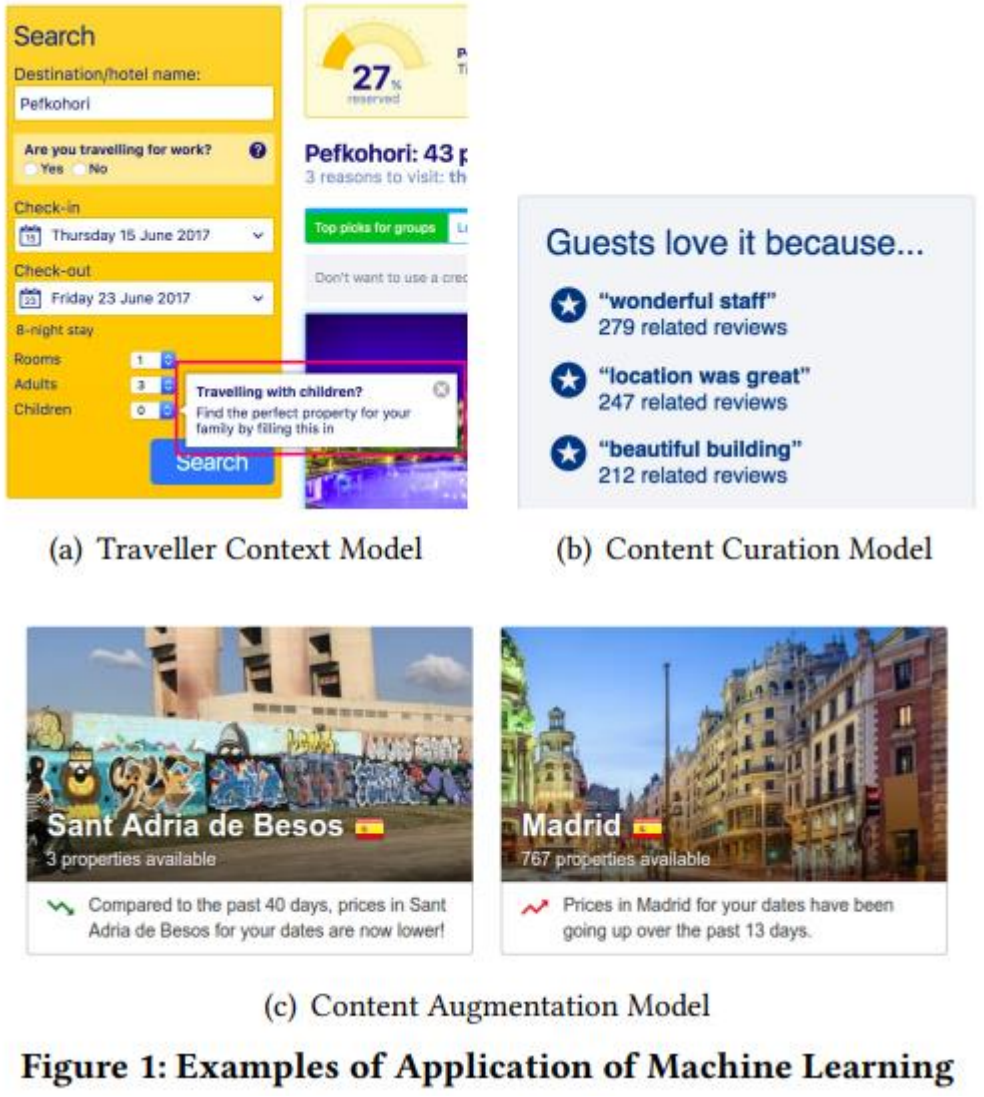
- Study to analyze the impact of machine learning models from the business perspective using machine learning models in use at the website Booking.com

2. INCEPTION: MACHINE LEARNING AS A SWISS KNIFE FOR PRODUCT DEVELOPMENT

- Machine learning can be used for many and very different products in widely different contexts.
- Created 2 types of models:
  - Models which are very specific for a use case.
  - Models which act as a meaningful semantic layer.
- On average each semantic model generated twice as many use cases as the specialized ones.

2.1 Model Families

- Traveller Preference Models
- Traveller Context Models
- Item Space Navigation Models
- User Interface Optimization Models
- Content Curation
- Content Augmentation
  - Great Value
  - Price Trends



(a) Traveller Context Model (b) Content Curation Model

(c) Content Augmentation Model

Figure 1: Examples of Application of Machine Learning

2.2 All model families can provide value

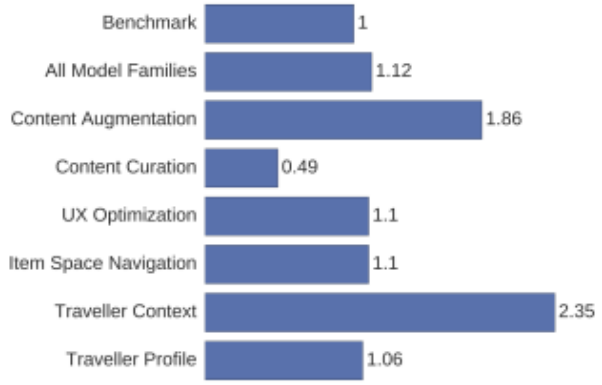


Figure 2: Model Families Business Impact relative to median impact.

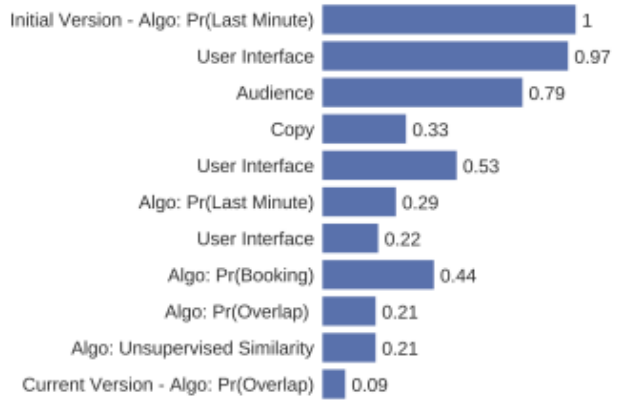
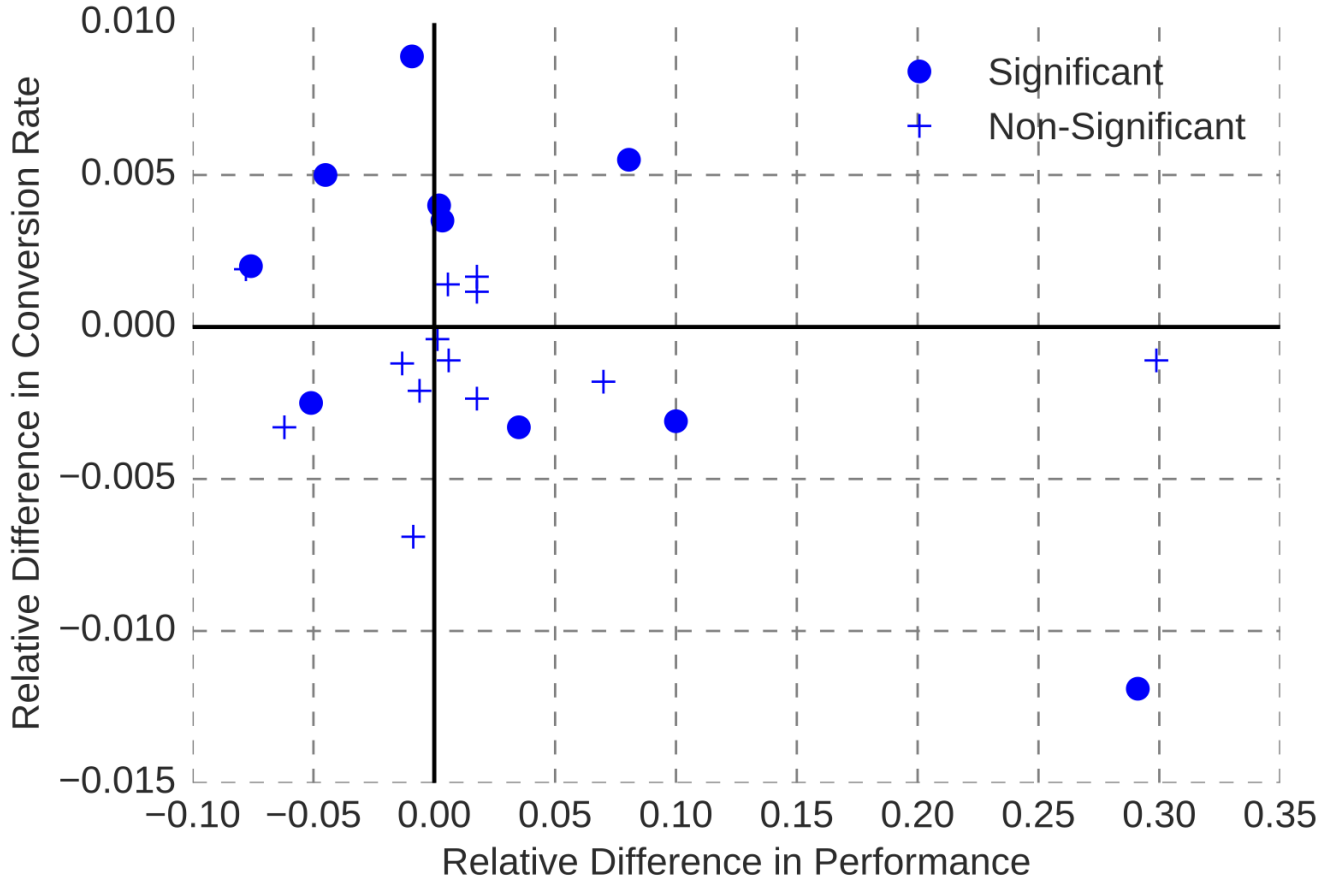


Figure 3: A sequence of experiments on a Recommendations Product. Each experiment tests a new version focusing on the indicated discipline or ML Problem Setup. The length of the bar is the observed impact relative to the first version (all statistically significant)

3. MODELING: OFFLINE MODEL PERFORMANCE IS JUST A HEALTH CHECK

- Model gains aren't always the business gains.



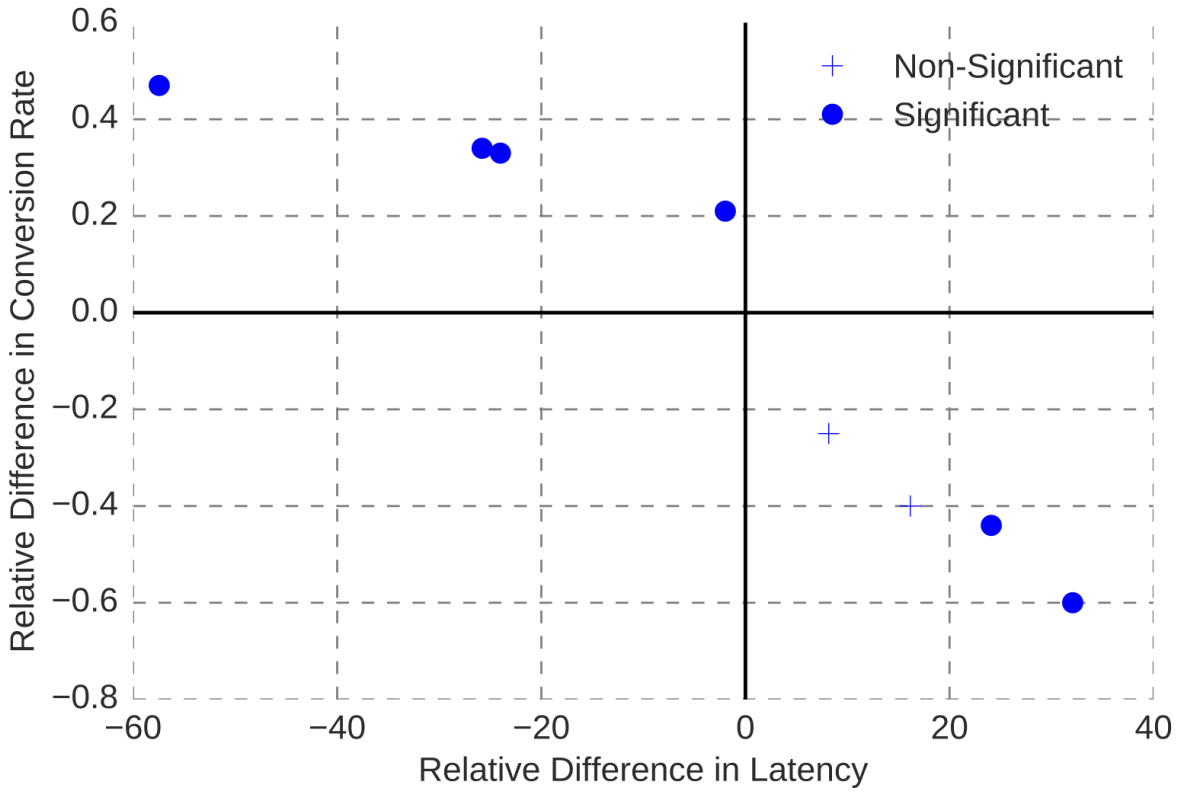
Relative difference in a business metric vs relative performance
difference between a baseline model and a new one.

4 MODELING: BEFORE SOLVING A PROBLEM, DESIGN IT

- Learning Difficulty
- Data to Concept Match
- Selection Bias

5. DEPLOYMENT: TIME IS MONEY

- Analyzing the impact of models on time



Impact of latency on conversion rate

6 MONITORING: UNSUPERVISED RED FLAGS

- Incomplete feedback
- Delayed feedback
- Response Distribution Analysis

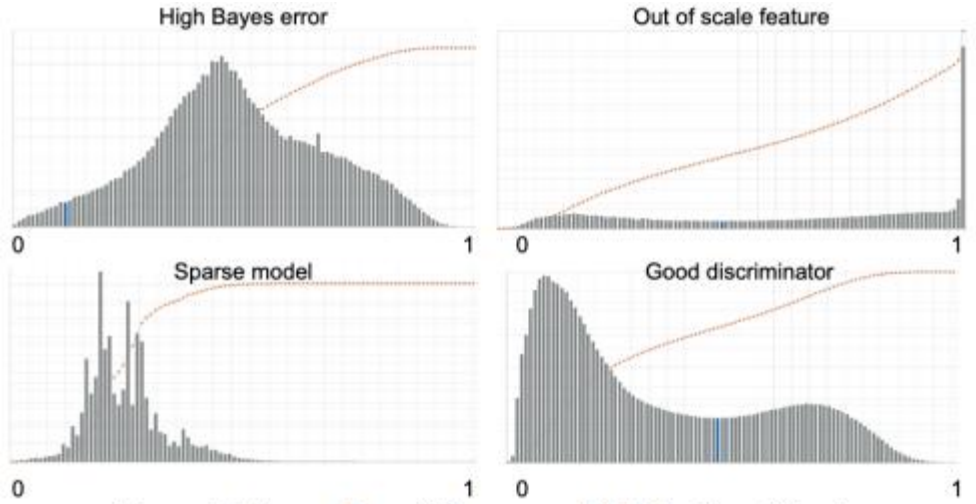
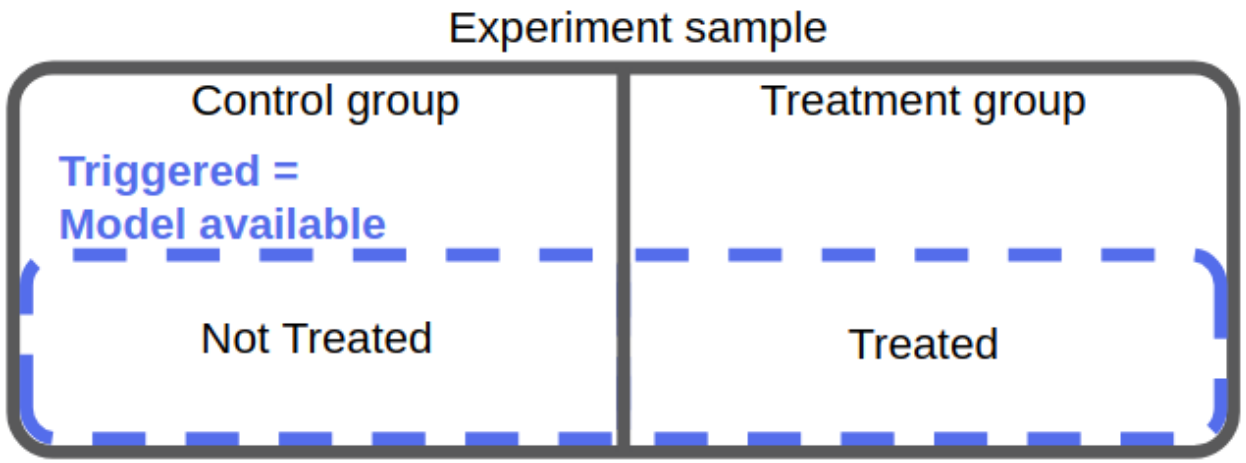


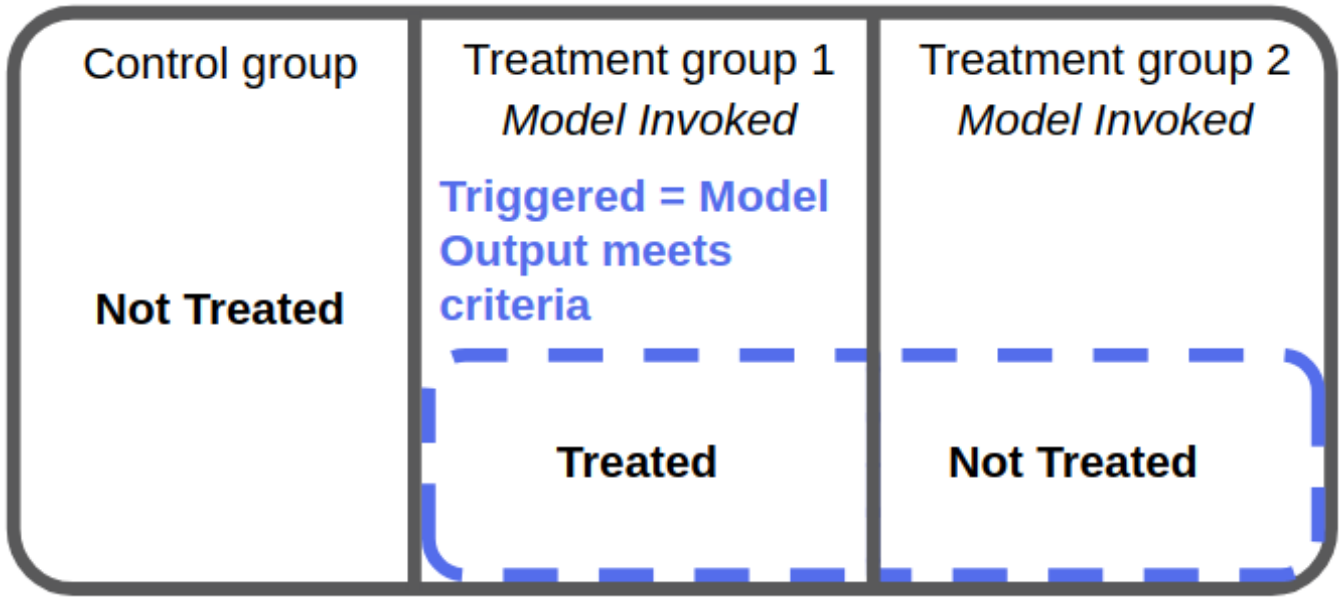
Figure 7: Examples of Response Distribution Charts

7. EVALUATION: EXPERIMENT DESIGN SOPHISTICATION PAYS OFF

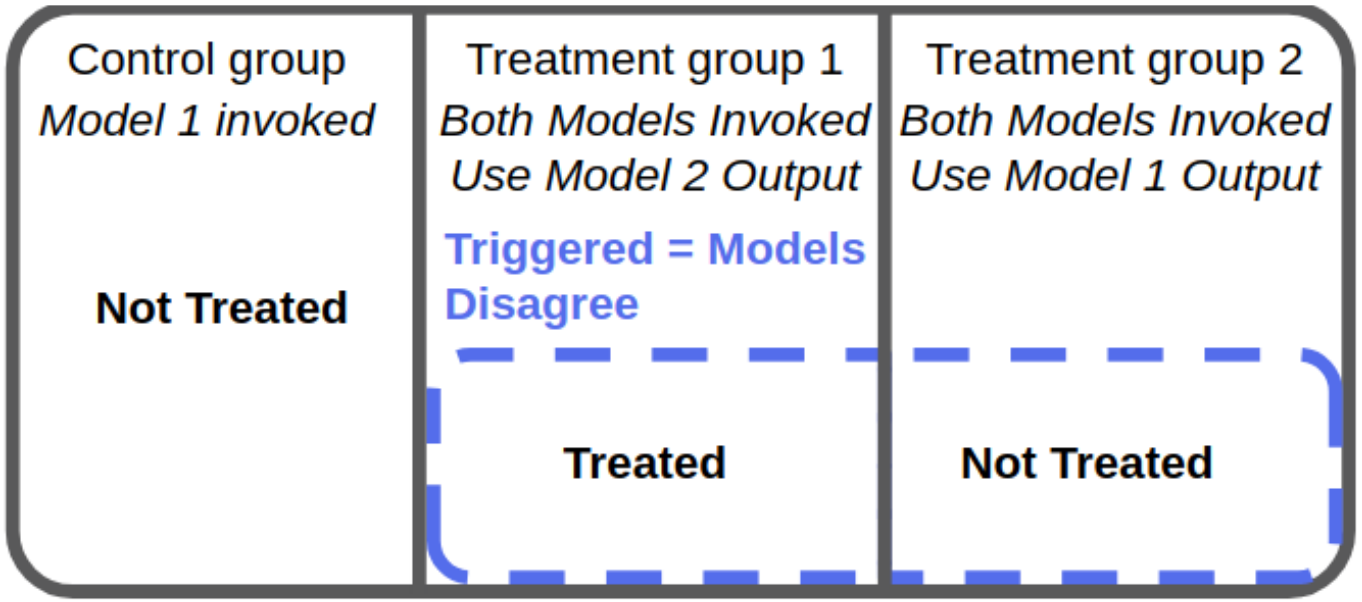
- How to set up experiments



Experiment design for selective triggering.



Experiment design for model-output dependent triggering and control for performance impact.



Experiment design for comparing models.

Summary

- 6 lessons learned:
  - projects introducing machine learned models deliver strong business value
  - model performance is not the same as business performance
  - be clear about the problem you're trying to solve
  - prediction serving latency is important
  - get early feedback on model quality
  - test the business impact of your models through randomised controlled trials