

Master research project/thesis topic

Interpretable algorithms for flight delay prediction

Context

Flight delays are a common hassle for travelers every day. Just in 2017 around 20% of the flights in the US were delayed. Delays are not only inconvenient to the passengers but also increase airplane fuel consumption. Machine learning algorithms have been used in different forms to produce predictions of flight delays, but these can vary for each region of the globe and do not give a thorough insight into the mechanism of the delays.

Black box models have dominated the machine learning landscape in the recent years. A major issue associated with these algorithms is their lack of transparency and untrustworthiness when dealing with high-stake decisions, as in the case of the aviation industry, where each decision can impact thousands of lives in unexpected ways.

Goal

The objective of this project is to apply recently developed interpretable models that not only can perform to the level of state-of-the-art models, but can also provide extra insight into the problem. The project should start with an overview of the prediction of flight delay literature and based on this a comparison setting for the algorithms should be selected. As a final step, the best models should be analyzed and extra knowledge about flight delays should be extracted.

Research questions

Example of a main research question:

- What are the best interpretable models for flight delays?

Secondary questions:

- How to define/measure interpretability of the models?
- How can we extract knowledge from the models?
- What extra insights do these models give?

Realization

- Literature study
- Summarize the landscape of flight delay predictions
- Define a comparison setting based on existing literature and available data
- Choose models to apply
- Apply the selected models
- Evaluate the results, derive recommendations
- Write report /thesis

Student profile

Good understanding of data mining and machine learning problems and algorithms; experience/interest in interpretability of models; programming experience (preferably in Python and/or R).

Supervisors

Hugo Manuel Proença (h.manuel.proenca@liacs.leidenuniv.nl)

Matthijs van Leeuwen (m.van.leeuwen@liacs.leidenuniv.nl)

Relevant literature

Example of flight delay prediction literature:

- **Deep learning for flight delay prediction**
Kim, Young Jin, et al. "A deep learning approach to flight delay prediction." *Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th*. IEEE, 2016.
- **A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks**
Gopalakrishnan, Karthik, and Hamsa Balakrishnan. "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks." *USA/Europe Air Traffic Management Seminar*. 2017.

Example of interpretable models for classification:

- **Interpretable Classification Models for Recidivism Prediction**
Zeng, Jiaming, Berk Ustun, and Cynthia Rudin. "Interpretable classification models for recidivism prediction." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 180.3 (2017): 689-722.
- **Scalable Bayesian rule lists**
Yang, Hongyu, Cynthia Rudin, and Margo Seltzer. "Scalable Bayesian rule lists." *arXiv preprint arXiv:1602.08610* (2016).
- **Interpretable decision sets: A joint framework for description and prediction**
Lakkaraju, Himabindu, Stephen H. Bach, and Jure Leskovec. "Interpretable decision sets: A joint framework for description and prediction." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016.