[IN104]
Explainability Project
Titanic Survivor Prediction and Explainable Artificial Intelligence Techniques

Silvia Tulli

April 2021

1 Introduction

Machine learning (ML) has touched almost every scientific discipline (and industry) and is now steadily used throughout academia. However, machine learning models appear to be opaque to humans. This motivates an increasing number of studies on developing methodologies to explain the reasoning and outputs of machine learning models. Predict survival on the Titanic is a legendary problem to familiarize with ML basics. First, we will use the Titanic database to build a model that predict what sorts of people were more likely to survive. Second, we will investigate methods to explain the functioning of our models.

2 Interesting Articles

- Lipton, Z.C. (2018). The Mythos of Model Interpretability. Queue, 16, 31 - 57. pdf
- Gunning, D., Aha, D., (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44-58. pdf
- Doshi-Velez, F., Kim, B. (2017). A roadmap for a rigorous science of interpretability.pdf
- Miller, T., (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. In: Artificial Intelligence Journal.pdf

3 Online Resources

- SHAP - Github repository
- Lime - Github repository
- Partial Dependence Plot (PDP) - Github repository
- Accumulated Local Effects (ALE) - Github repository
- Predicting Titanic Survivors with Machine Learning - Ju Liu Tutorial
- Intepretable Machine Learning, Partial Dependence Plot (PDP) - Christoph Molnar
- Intepretable Machine Learning, Accumulated Local Effects (ALE) Plot - Christoph Molnar
- Machine Learning Explainability - Kaggle short course
- Explaining AI Models Using Shap - Xyonix post