CHECKPOINT 1

DAVID SCARIN, FRANCISCO SILVA, TOMAS AZEVEDO
PA2_DS3_GA

DATASET

Airline Passenger Satisfaction Dataset

This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger?

Characteristics:

- 129880 observations
- 24 features

Target Variable:

- Satisfaction: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Problem Type:

- Multi-class classification problem

PRE-MODELLING

Airline Passenger Satisfaction Dataset

Applying methods for interpretable ML, describing the data, setting baseline expectations, trying to find instances that are representative of the data through maximum mean discrepancy (MMD)

Techniques:

- EDA (Exploratory Data Analysis) and Visualization
- Prototypes through data summarization: MMD2-Critic

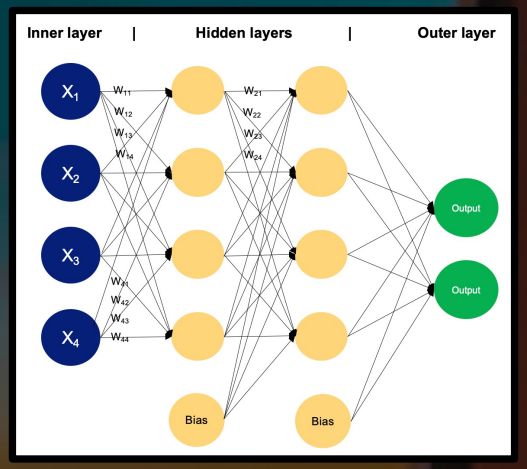
IN-MODELLING

Airline Passenger Satisfaction Dataset

Applying supervised multi-class classification learning models to the dataset; White Box / Black Box models,

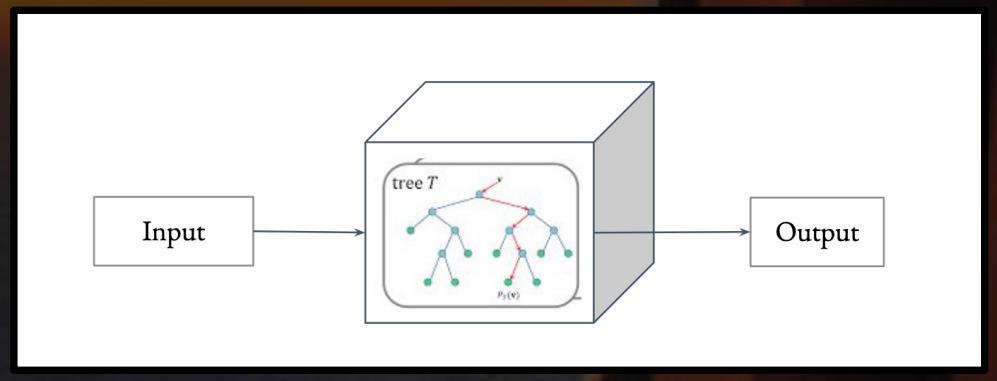
Black Box:

- MLP (Multi-Layer Perceptron)



White Box:

- Decision Tree
- Other rule learners (ILP?)



BIBLIOGRAPHY

Airline Passenger Satisfaction Dataset

- Molnar, C. (2024) Interpretable Machine Learning: A Guide For Making Black Box Models Explainable, Leanpub, 2022. ISBN: 979-8411463330 (https://christophm.github.io/interpretable-ml-book/)

- Belle, V., Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. Frontiers in Big Data, 4, 688969. https://doi.org/10.3389/fdata.2021.688969

- Class Documentation TAAC 2024, Rita Ribeiro, Pedro Ferreira

- Marques-Silva, Joao. "Logic-based explainability in machine learning." *Reasoning Web. Causality, Explanations and Declarative Knowledge: 18th International Summer School 2022, Berlin, Germany, September 27–30, 2022, Tutorial Lectures*. Cham: Springer Nature Switzerland, 2023. 24-104.