

SURPASSING BLACK BOX MODEL'S PERFORMANCE ON UNBALANCED DATA WITH AN INTERPRETABLE ONE USING ADVANCED FEATURE ENGINEERING

Lighting Talk

Witold Merkel, Adam Rydelek, Michał Stawikowski

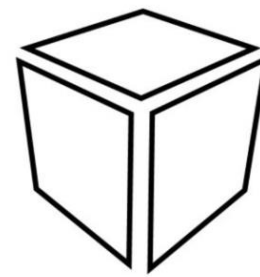
INTRODUCTION TO THE PROBLEM

Interpretability

White Boxes vs Black Boxes

Feature Engineering

Unbalanced Data



or





DATA

Census data

Earnings prediction

Missing values

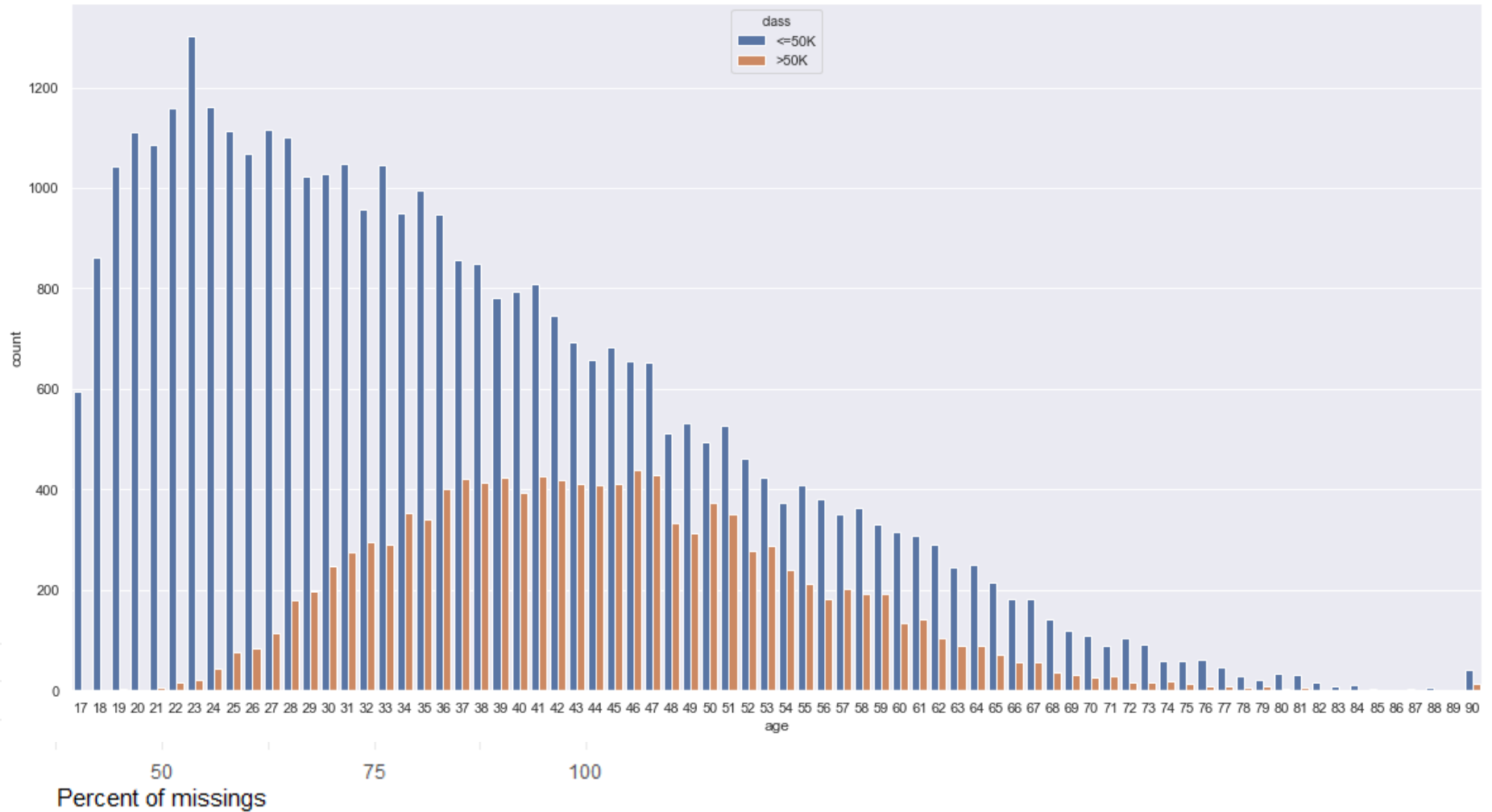
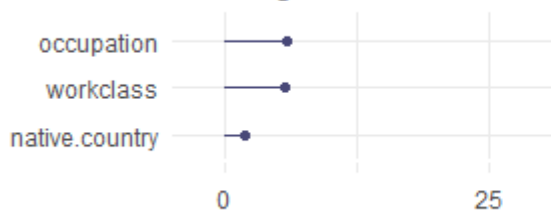
Skewness

EDA AND INITIAL DATA PREPARATION



Features

Missing dataset



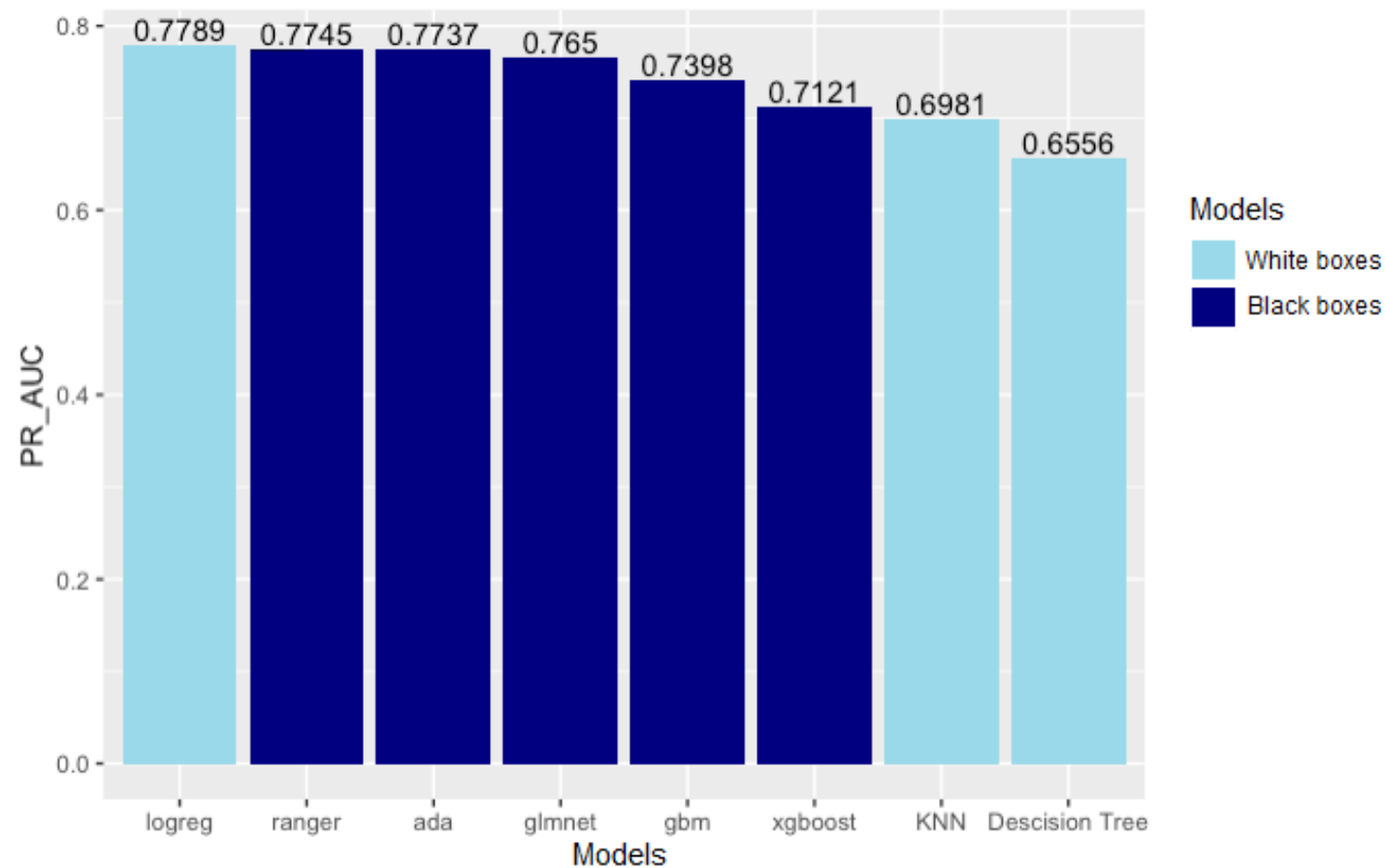
FEATURE ENGINEERING AND TUNING

rSAFE

Changing task to
regression

Variable Selection

Tuning





RESULTS AND SUMMARY

