

Building an explainable  
model for ordinal classification  
on Eucalyptus dataset.  
Meeting black box  
model performance levels.

Karol Saputa  
Małgorzata Wachulec  
Aleksandra Wichrowska



# Presentation scope

1. Research purpose
2. Ordinal classification
3. Dataset
4. Methodology
5. Results
6. Model explanation



# Research purpose

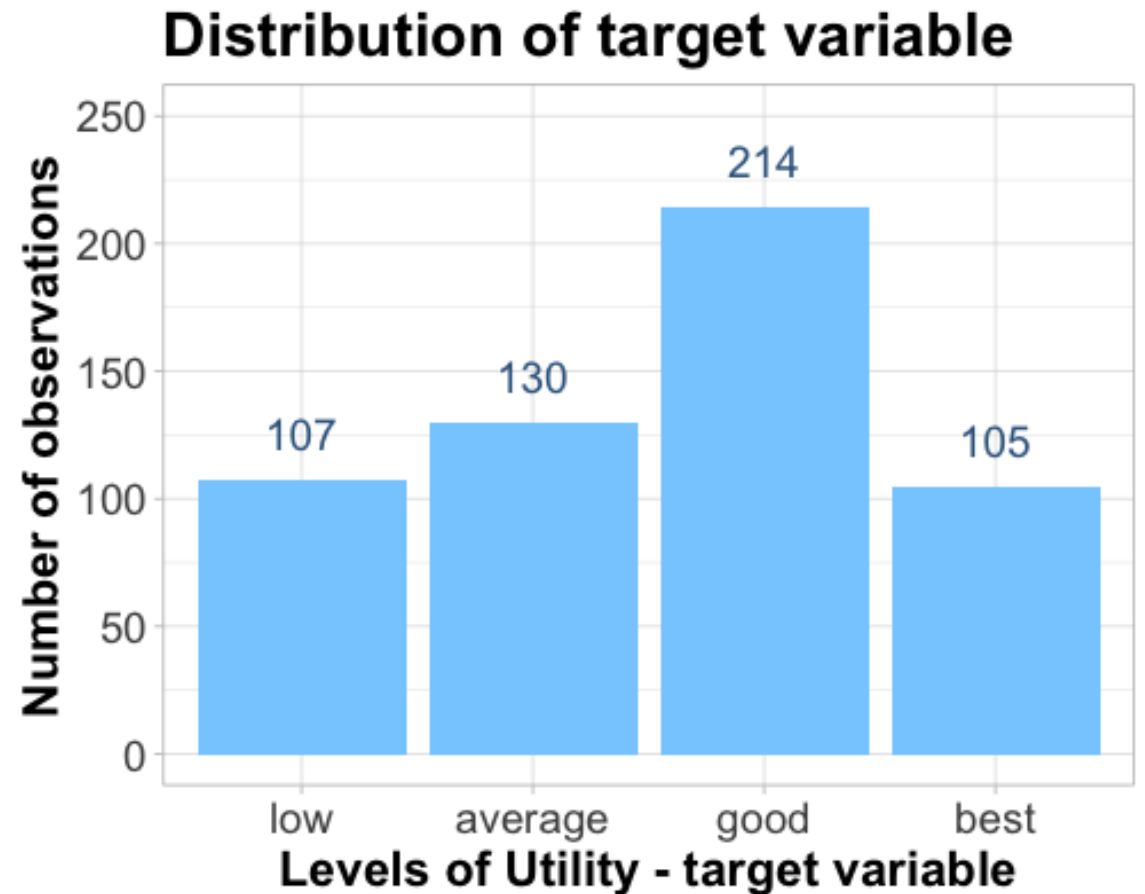
- Make an explainable machine learning model for ordinal classification problem
- Test some existing approaches to ordinal classification
- Achieve better results than 'black box' model

# Ordinal classification

- Problem between regression and classification
- Target variable is ranked:  
e.g. has values: ('low', 'medium', 'high')
- Applied to problems like predicting user rating

# Dataset - Eucalyptus

- OpenML
- 556 observations of 87 variables
- Ordinal target: Low, Average, Good, Best
- Target meaning: suitability for soil conditions



# Methodology

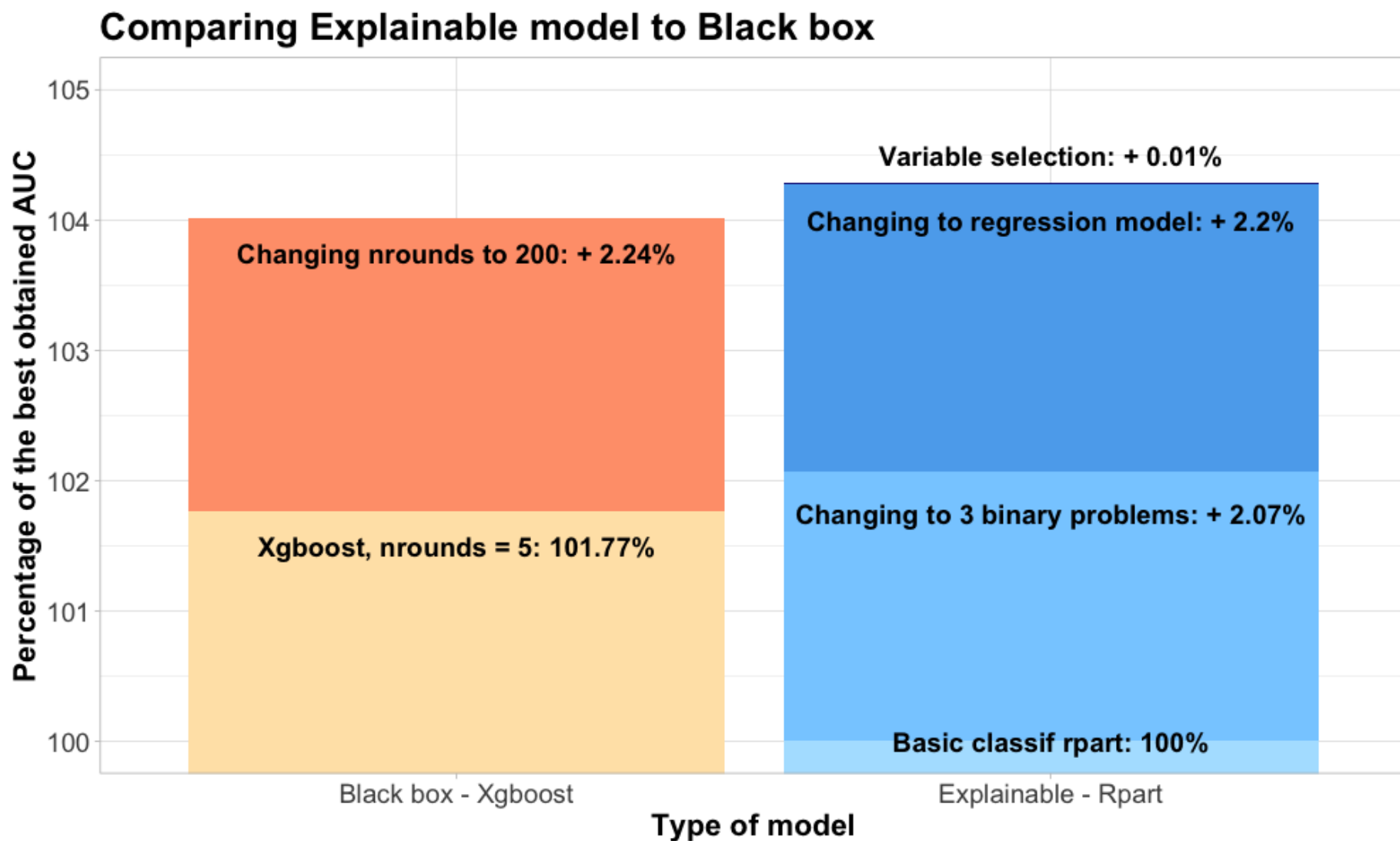
- Trying all mentioned ordinal classification approaches
- In addition:
  - Missing data imputation
  - Changing variable types
  - Variable selection
- Keeping changes that increased the AUC score
- Deleting the site names and specific location tags

# Results

Best model: Rpart for regression problem, with variable selection

Model	AUC	MSE	ACC	ACC1	% Best AUC
Basic rpart	0.8259	0.5284	0.5835	0.9797	95.89%
Regr rpart var-sel.	0.8613	0.4996	0.5815	0.9323	100.00%
Xgboost	0.8590	0.4467	0.6248	0.9873	99.73%
Xgboost small	0.8405	0.4998	0.6044	0.9830	97.59%

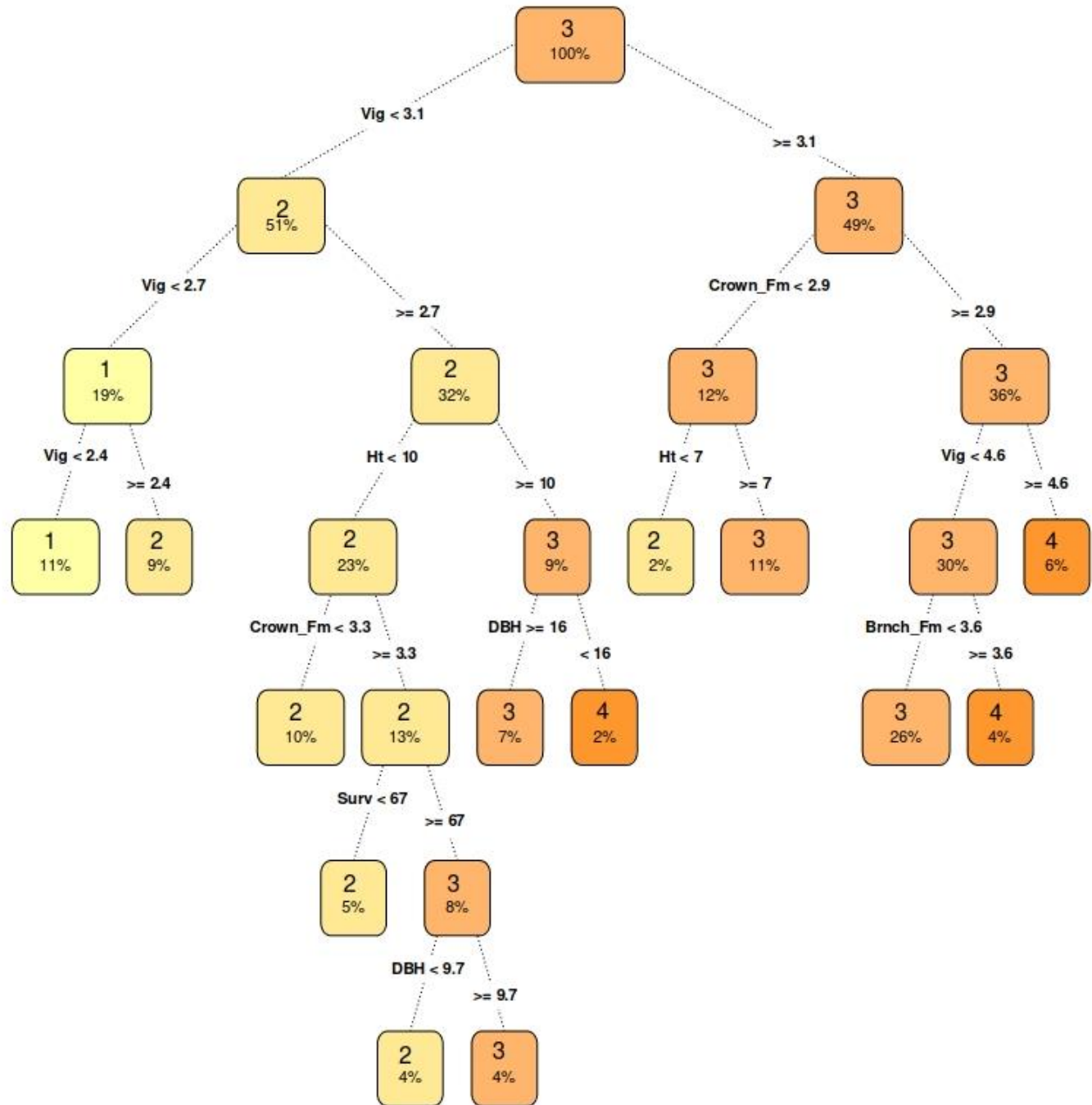
# Results





# Model explanation

- Tree model
- Depth = 6
- Easily explainable:  
e.g. small vigour value => low utility
- Independent from location



# Summary

Our model:

- lower computational cost compared to black box
- comparable/better results after presented methods applied
- ability to explain decisions

Building an explainable model for ordinal classification on Eucalyptus dataset. Meeting black box model performance levels, Saputa, Wachulec, Wichrowska, 2020

