Ensemble Methods

Elínborg Ásbergsdóttir
İpek Korkmaz
Luca Modica
Patrícia Marques

Group 30
Room HC3

02.05.2024

Part 1: Summary

Ensemble Methods

- Bagging (Random Forest)
- Boosting (Gradient Boosting)

Model Parameters

- **Number of estimators**: 15, 50, 100, 150, 250 (**bagging** and **boosting**)
- **Maximum depth**: None, 3, 5, 10 (**bagging** and **boosting**)
- **Learning rate**: 0.05, 0.1, 0.5 (**boosting**)

Pipeline

- Standard scaling
- Noise (0%, 30%, 60%, 90%)
- Train (80%) and test (20%) splitting
- Stratified fold splitting (3 folds)

Analysis

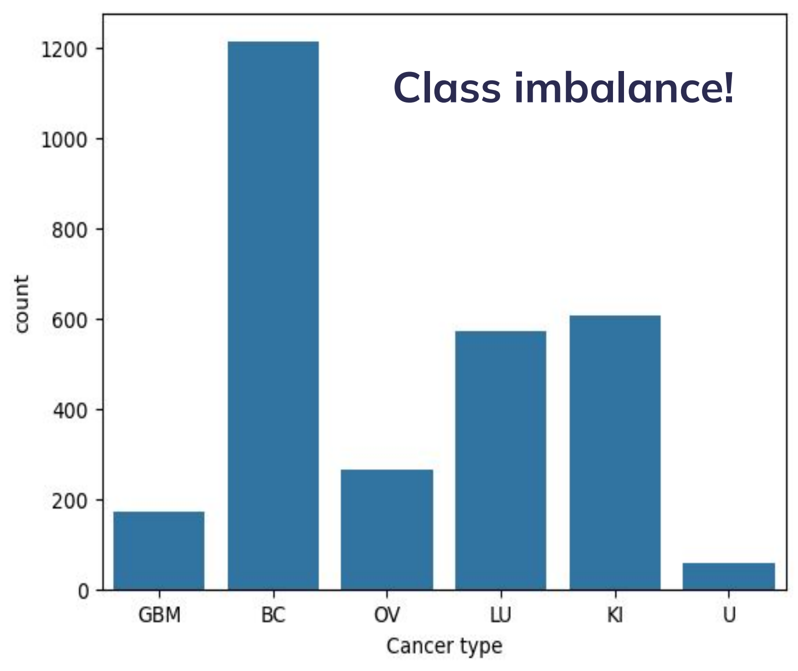
- **F1 Score** (cross-validation, train and test)

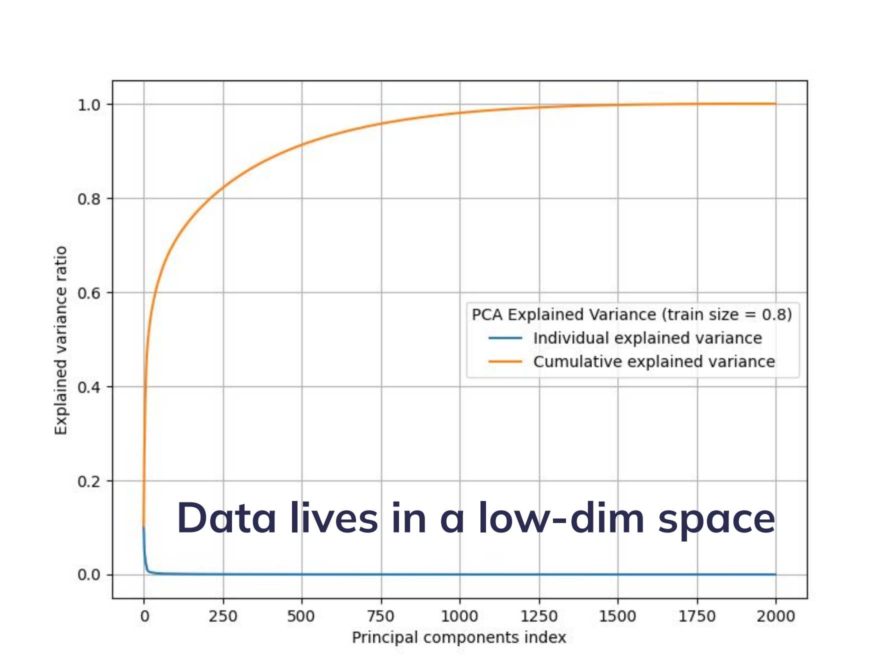
Cancer

Dataset

Cancer dataset exploration

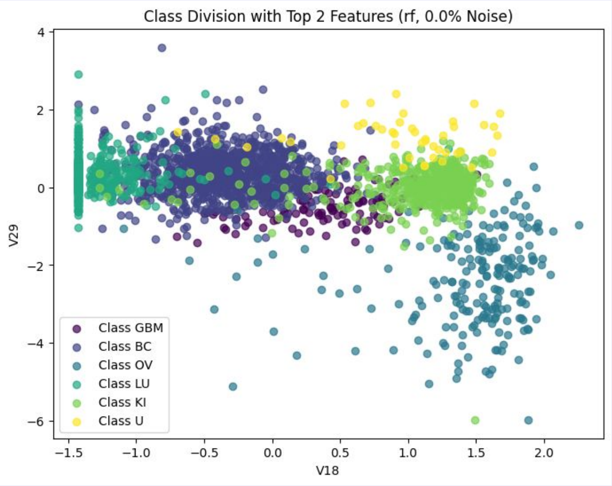
Meaningless features



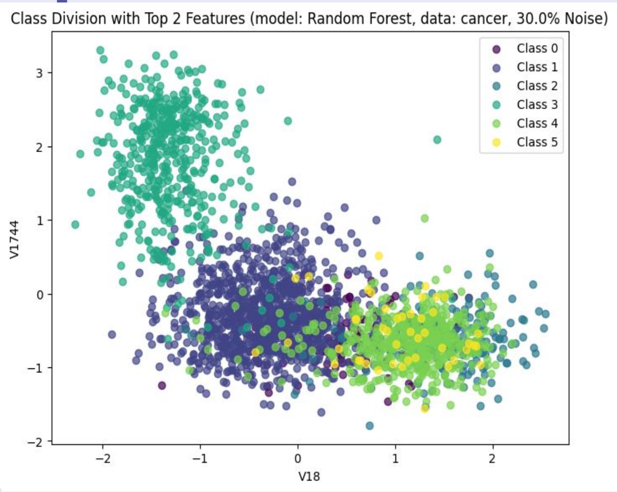


Random Forest: Top 2 Features

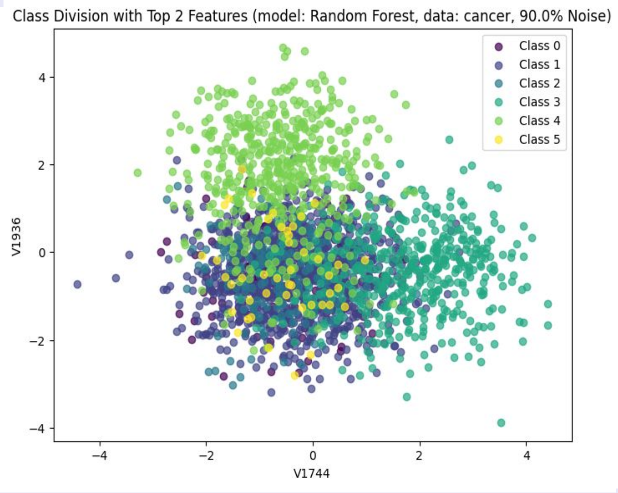
0% Noise



30% Noise



90% Noise



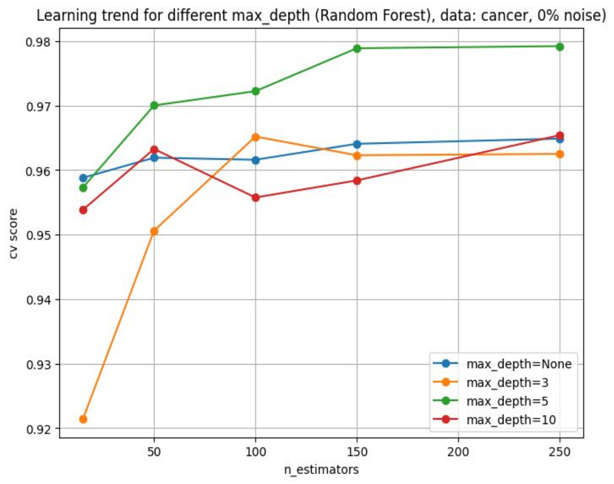
2 features allow to see class **separation**

Three classes are **still separable**

Classes **intertwined** with only 2 features

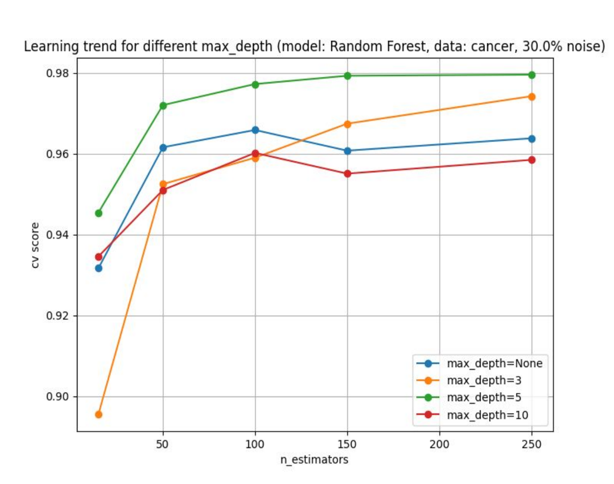
RF: Nr. of Trees and Maximum Depth

0% Noise



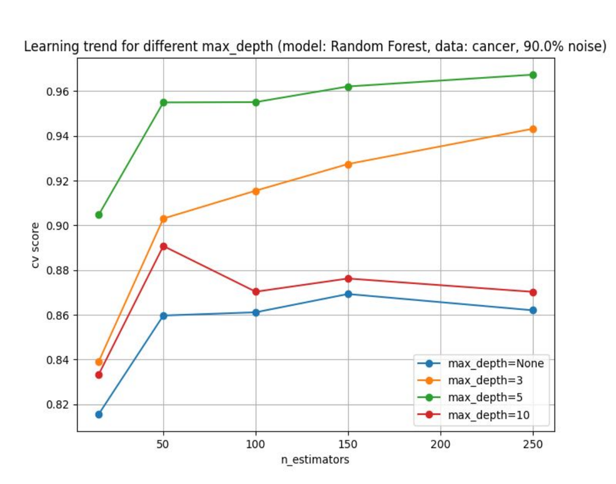
A small number of **deep trees** is sufficient

30% Noise



Good results especially for **enough estimators**

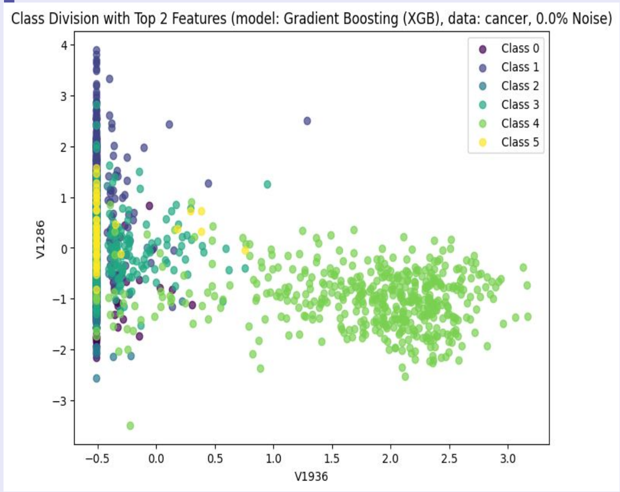
90% Noise



No benefits from a **depth** larger than 5 levels

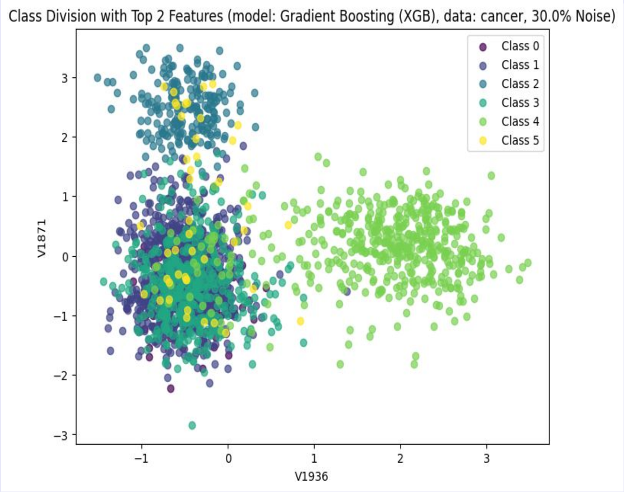
Gradient Boosting: Top 2 Features

0% Noise



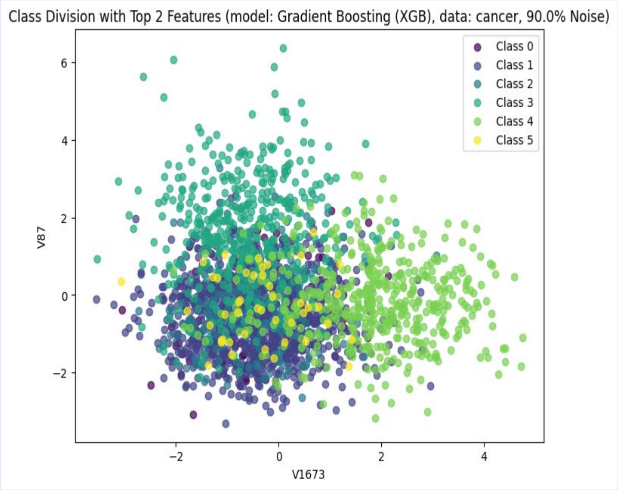
Gene V1936 allows to separate **class 4**

30% Noise



With only 2 genes some classes are **separable**

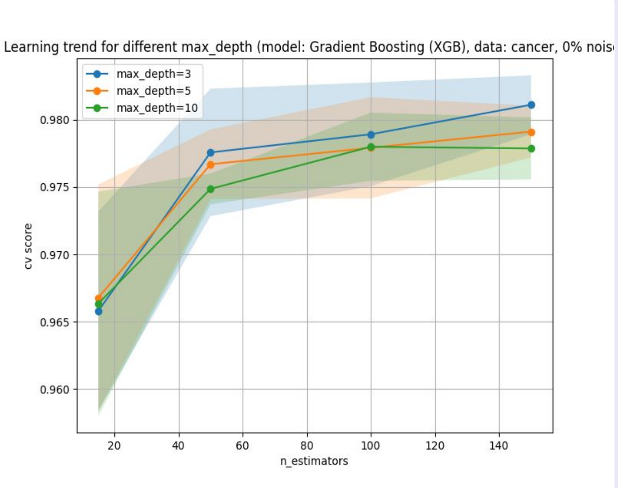
90% Noise



No clear **separation** with only two features

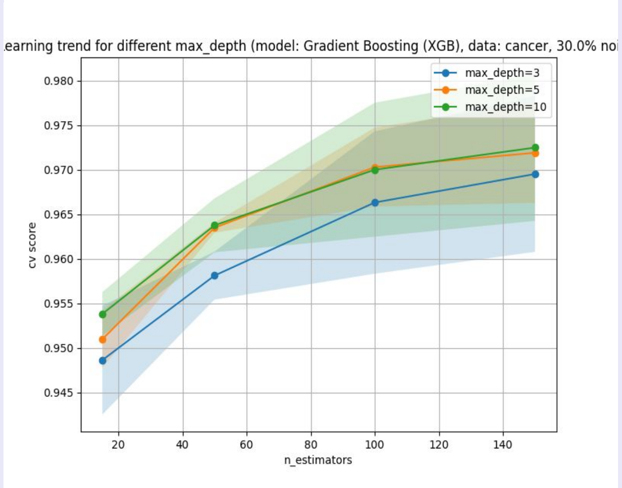
GB: Nr. of Trees and Maximum Depth

0% Noise



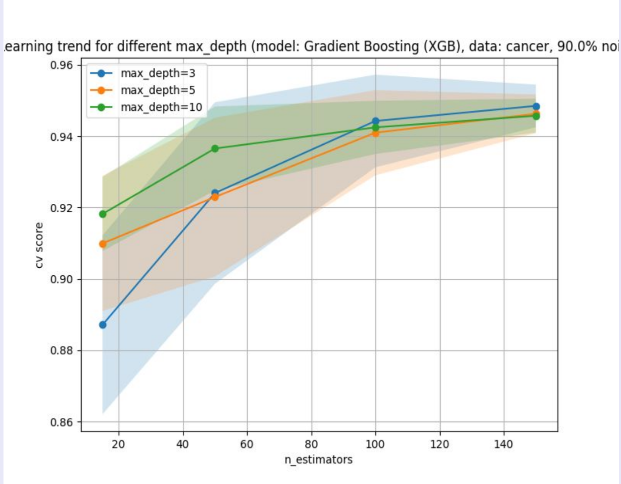
Maximum depth **does not change** F1 score

30% Noise



Results **worsen** but still **good**

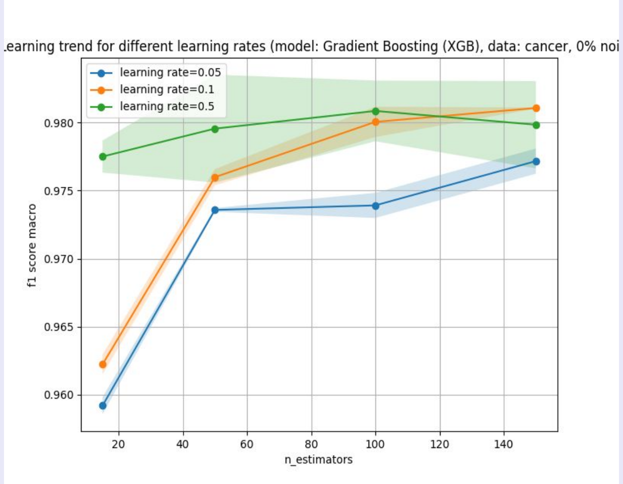
90% Noise



Similar scores with **enough** estimators

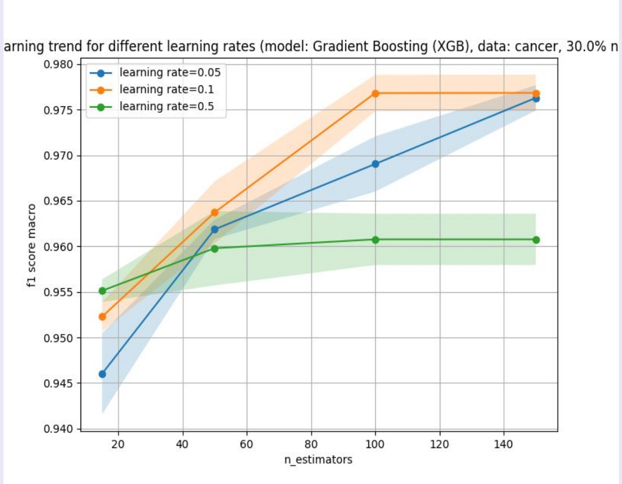
GB: Nr. of Trees and Learning Rate

0% Noise



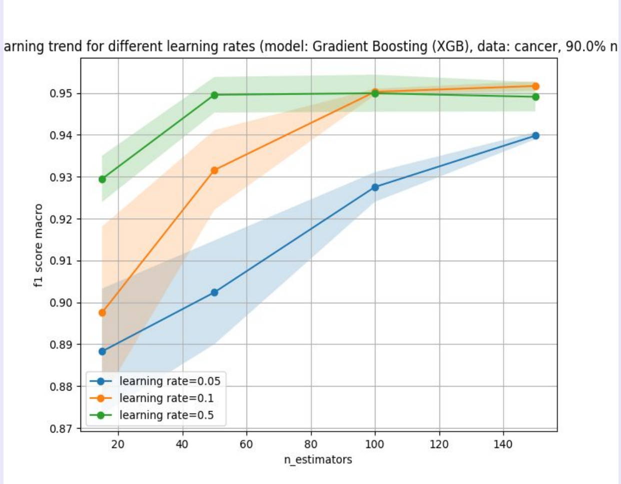
50 trees are enough
for **large learning rates**

30% Noise



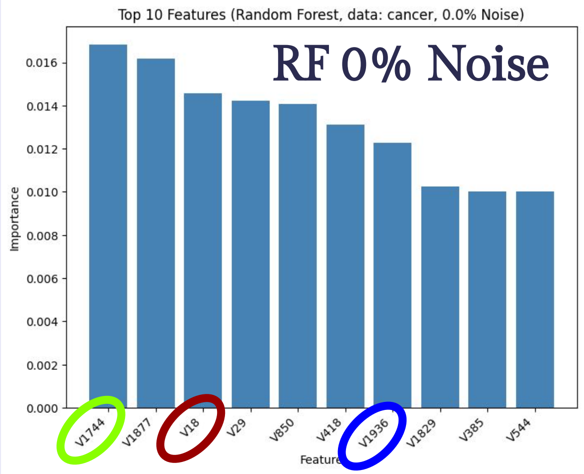
Best results for a **small learning rate**

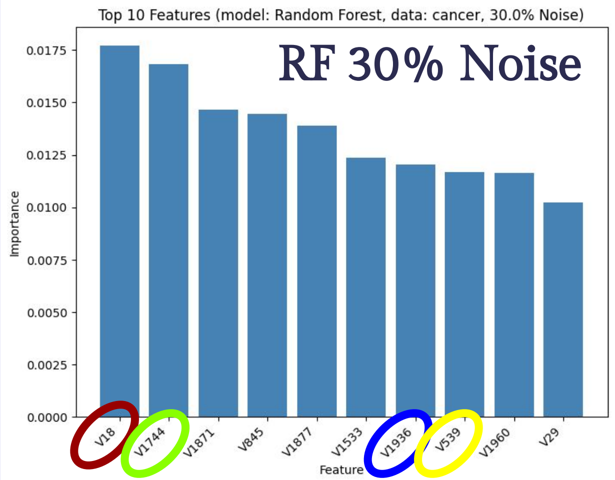
90% Noise

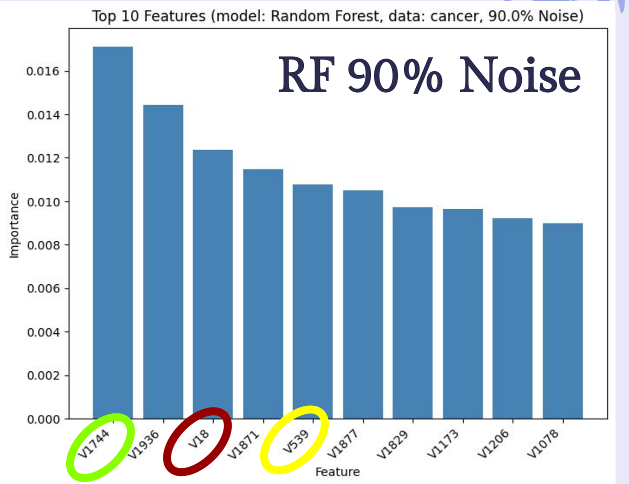


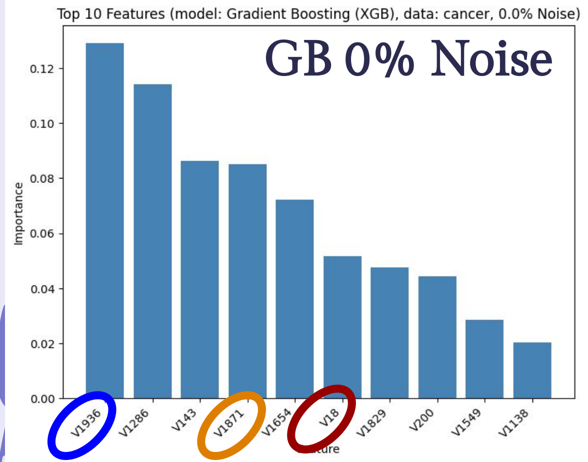
Early stopping effect
for small learning rates

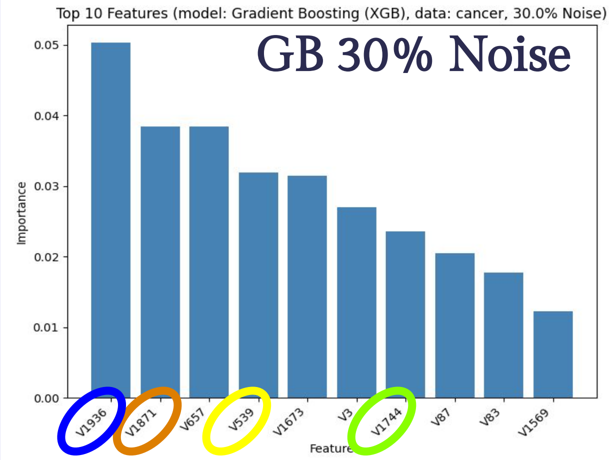
RF vs GB: Top 10 Features

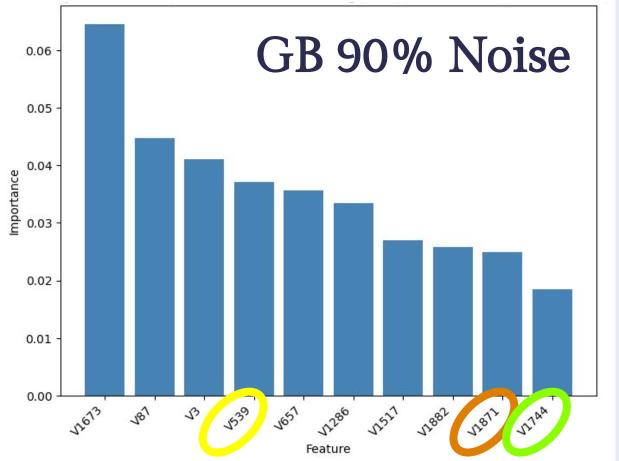




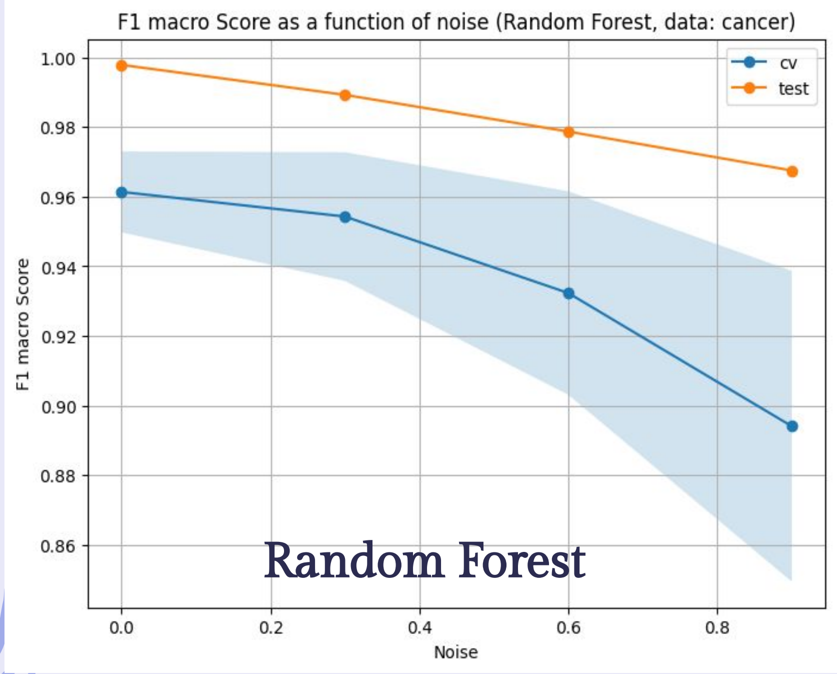


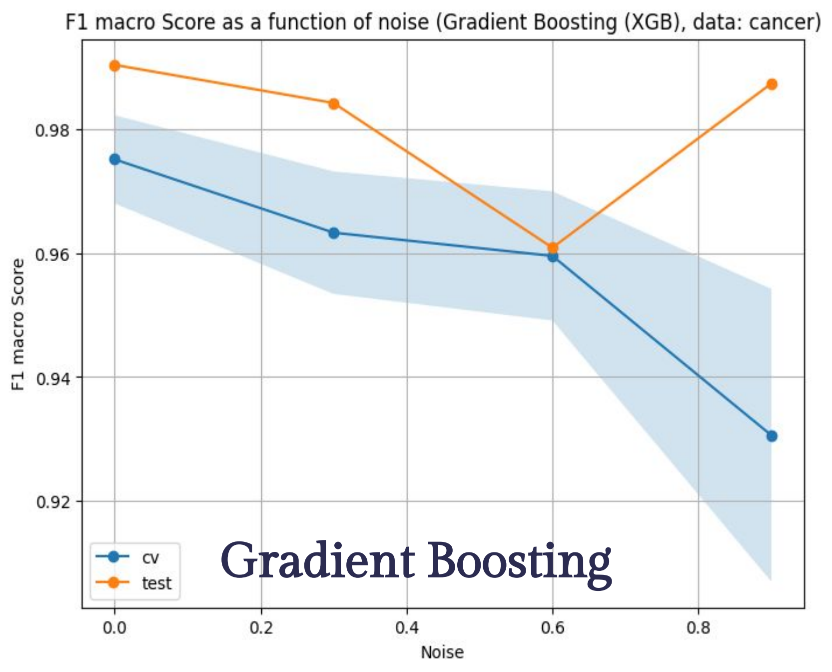






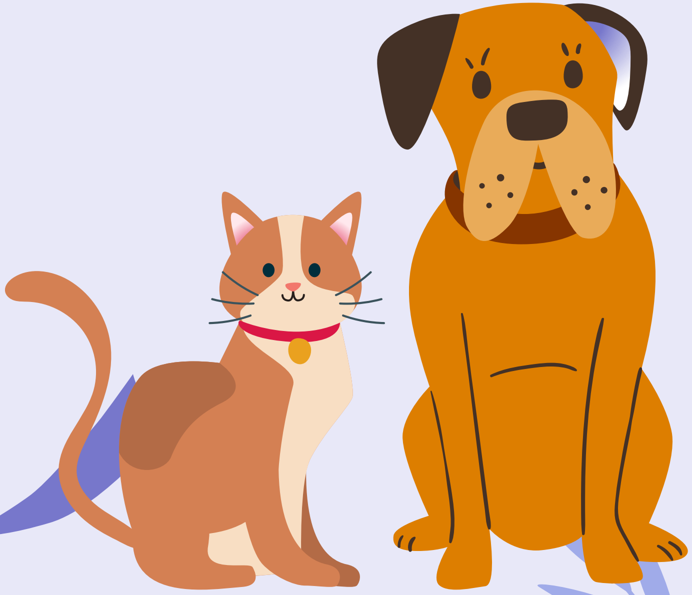
RF vs GB: Performance Comparison





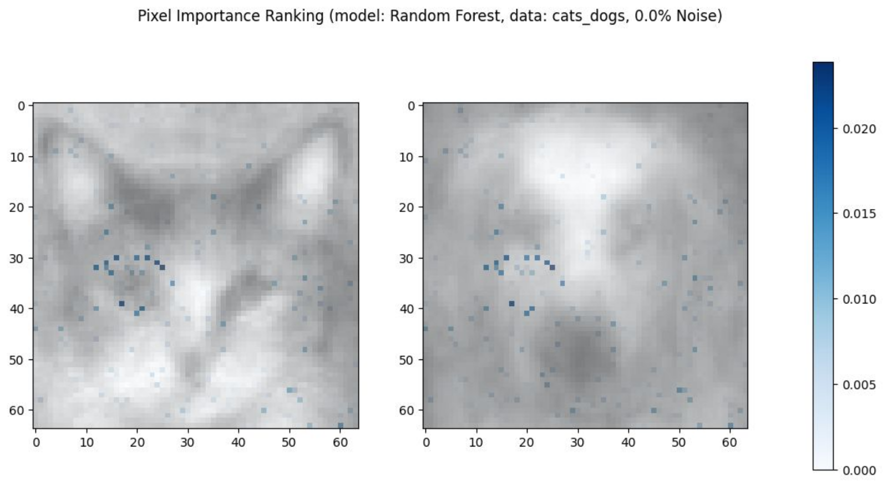
Cats and Dogs

Dataset



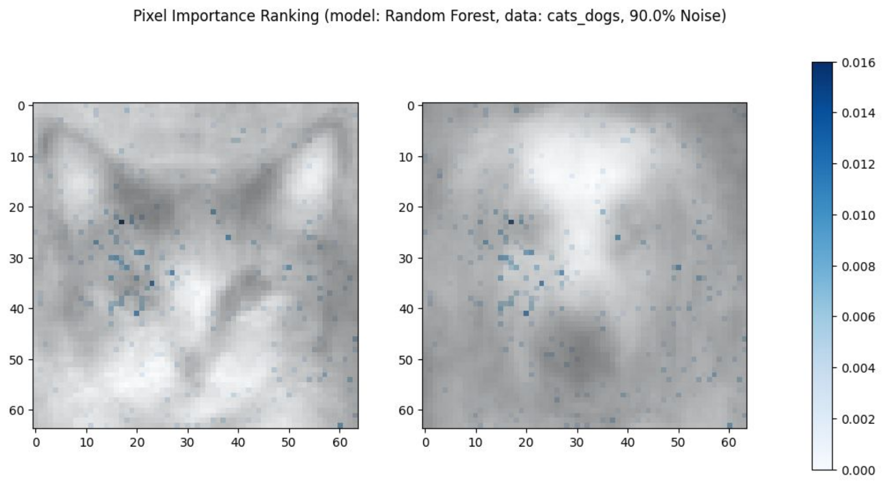
Random Forest: Top Pixels

0% Noise



Area of focus around the **cat's eye** and the **dog's cheek**

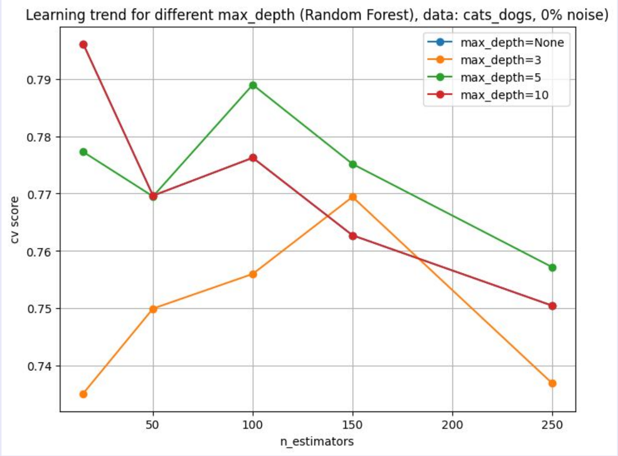
90% Noise



With noise, **more pixels** gain **importance**

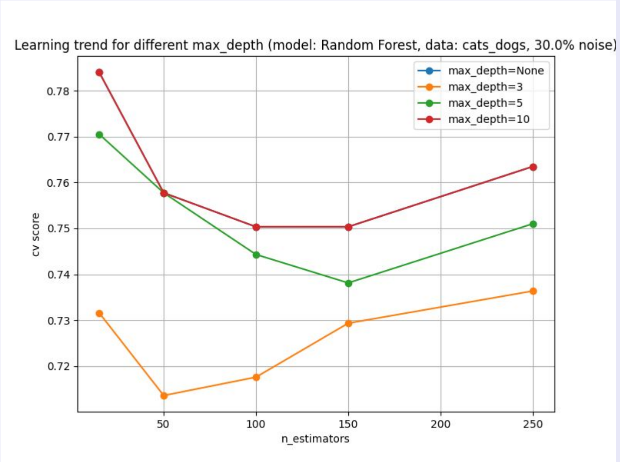
RF: Nr. of Trees and Maximum Depth

0% Noise



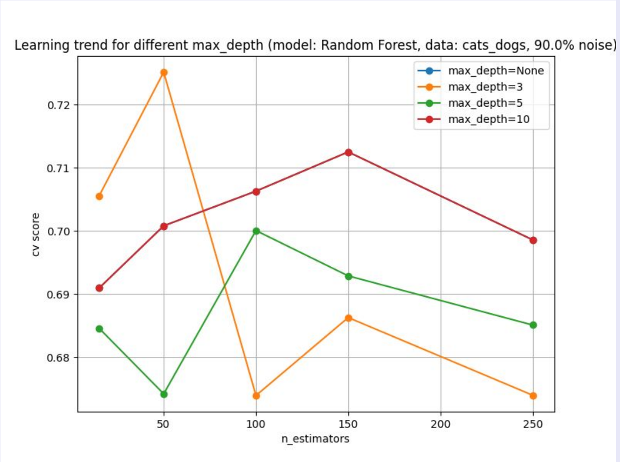
Depths **larger than 3** allow good training

30% Noise



Maximum depth of 10 **outperforms**

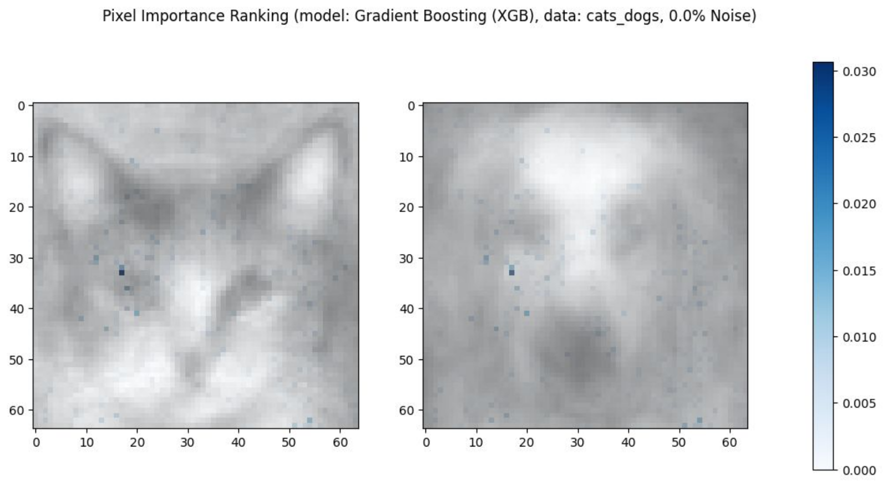
90% Noise



Performance **worsens**, even with more trees

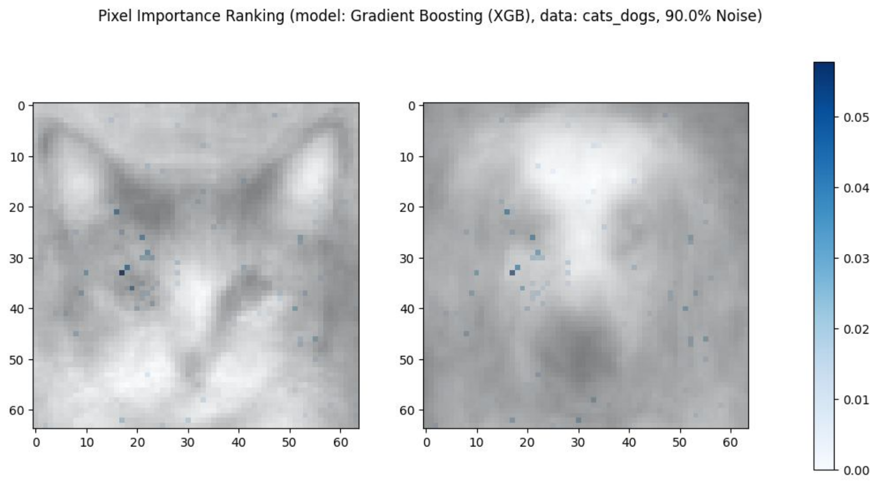
Gradient Boosting: Top Pixels

0% Noise



Same area of focus as with RF: around the **cat's eye** and the **dog's cheek**

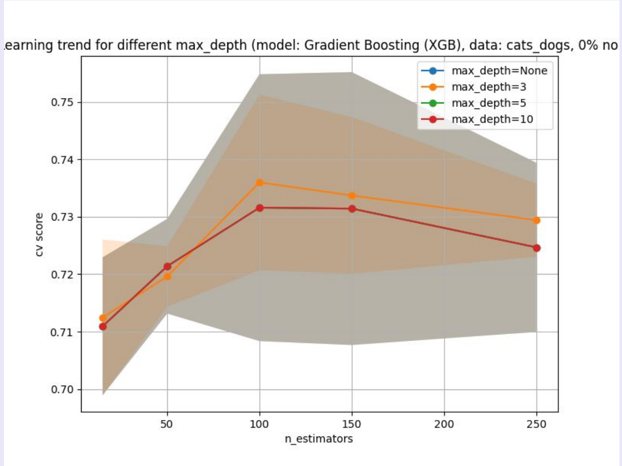
90% Noise



With noise, **maximum** value of importance **increased**

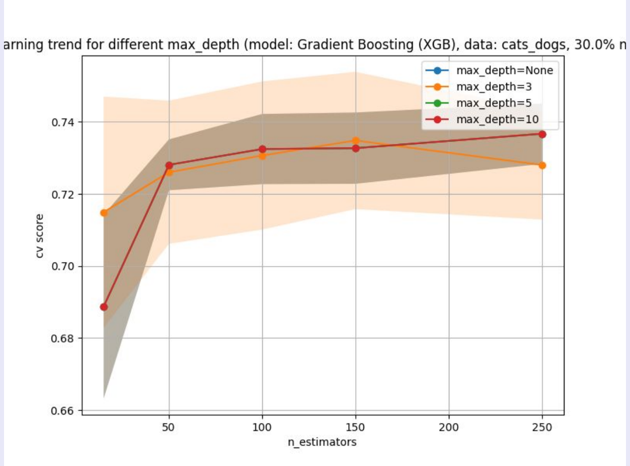
GB: Nr. of Trees and Maximum Depth

0% Noise



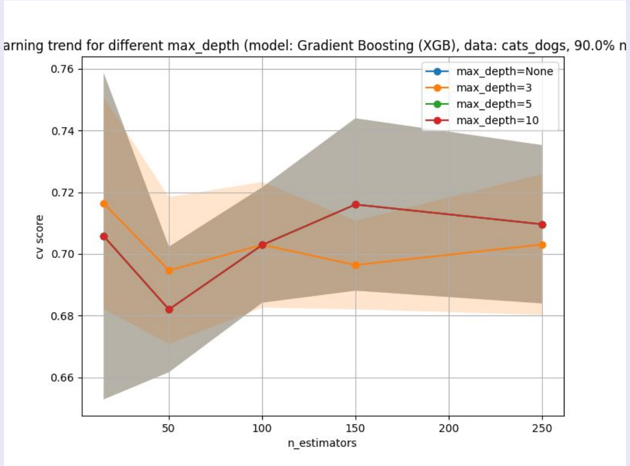
F1 score always **below 0.74**

30% Noise



Minimal changes for different learning rates

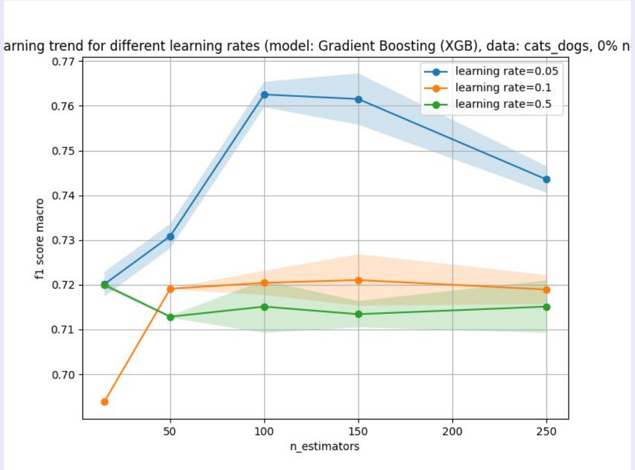
90% Noise



Unstable F1 scores for all estimators

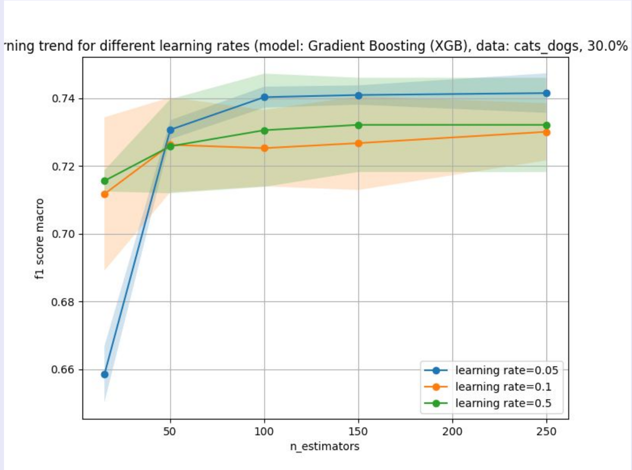
GB: Nr. of Trees and Learning Rate

0% Noise



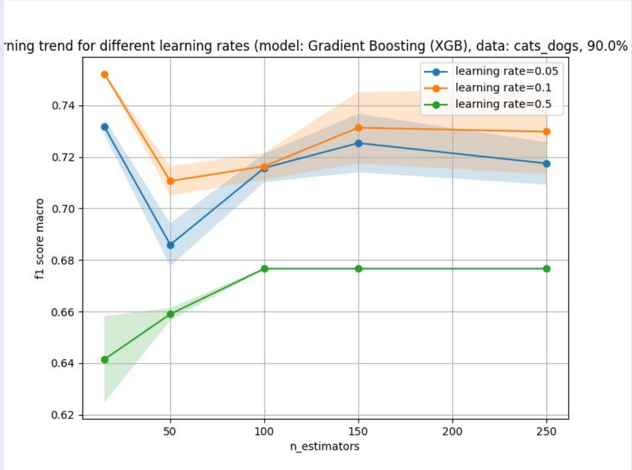
A **low learning rate** outperforms

30% Noise



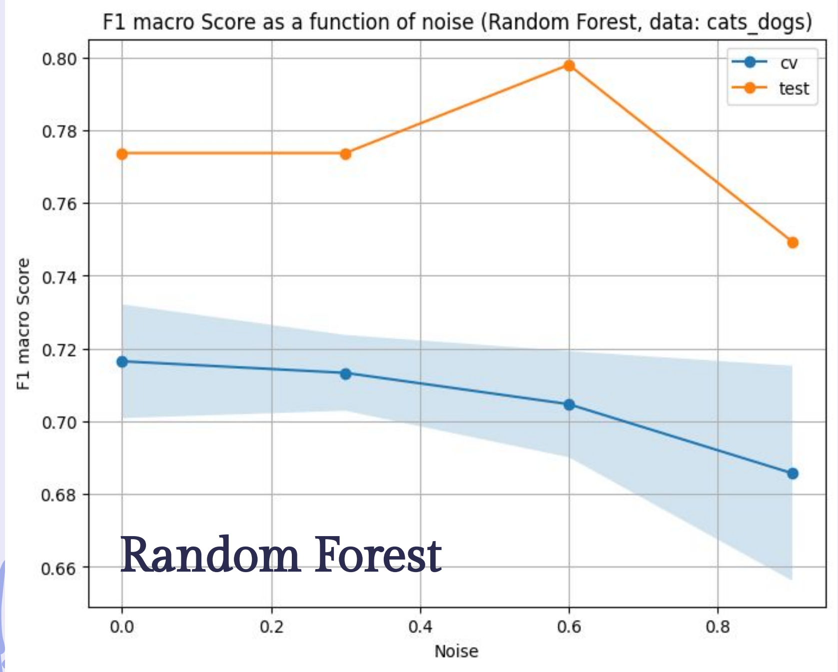
More than 50 trees **don’t improve** F1 score clearly

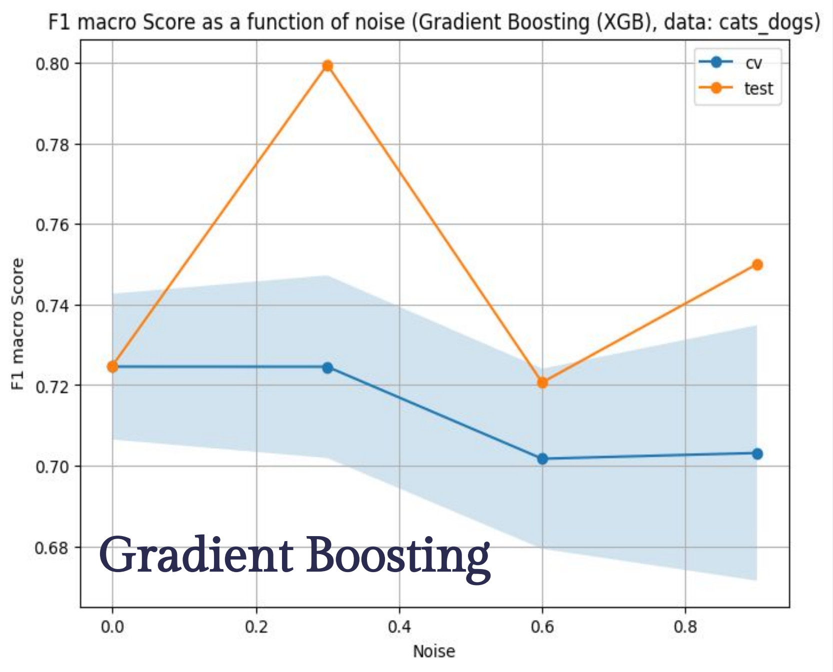
90% Noise



No improvements for forests with more trees

RF vs GB: Performance Comparison





Simulated data

Gradient Boosting

Part 2: Summary

Simulated dataset

- 400 samples for 400 features:
- 100 informative features
- 3 classes with weights 20%, 30%, 50%
- 70% of class separation

Model Parameters

- **Number of estimators**: 100 (default)
- **Maximum depth**: 1, None (default)
- **Learning rate**: 0.1 (default), 10

Ensemble Method

- Boosting (Gradient Boosting)

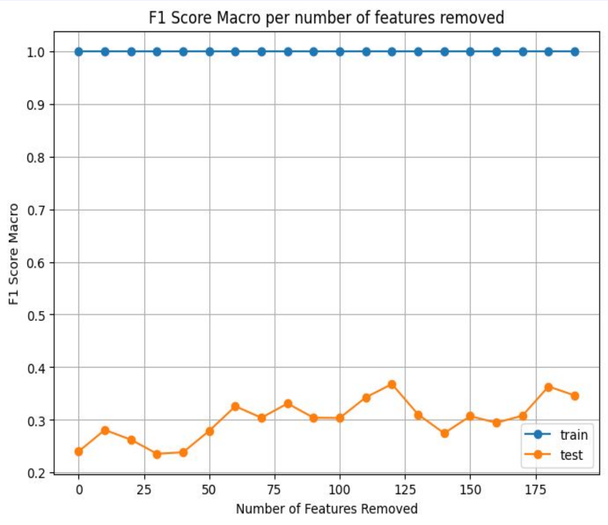
Procedure

- Recursively eliminate features

Analysis

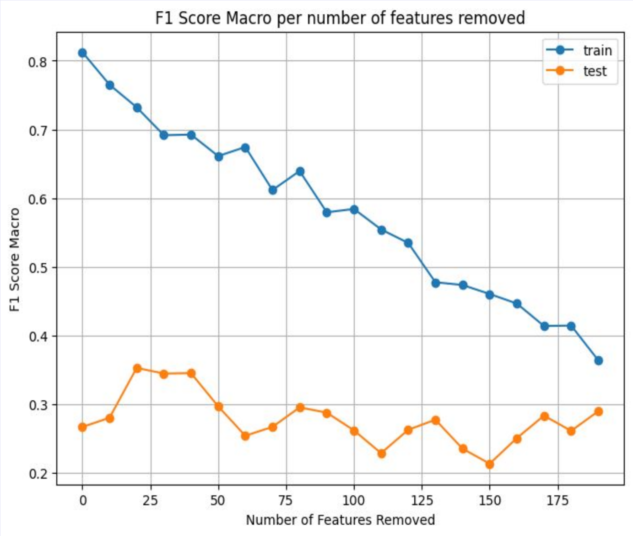
- **F1 Score**

GB: Maximum Depth and Nr. of Trees



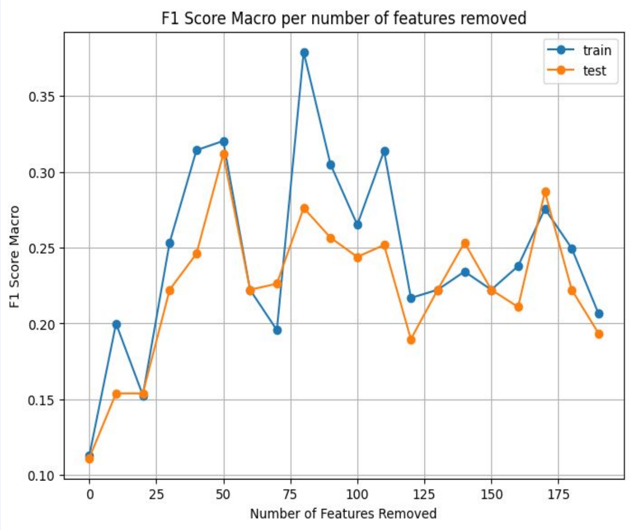
Immediate overfitting

Default parameters
Maximum Depth = None
Learning rate = 0.1



Train scores worsen

Shallow decision trees
Maximum Depth = 1
Learning rate = 0.1



Low F1 scores

High learning rate
Maximum depth = None
Learning rate = 10

Thanks

Does anyone have any questions?

Elínborg Ásbergsdóttir
İpek Korkmaz
Luca Modica
Patrícia Marques