SmartPred: Unsupervised Hard Disk Failure Detection

MLHPCS 2020

PHILIPP ROMBACH & JANIS KEUPER
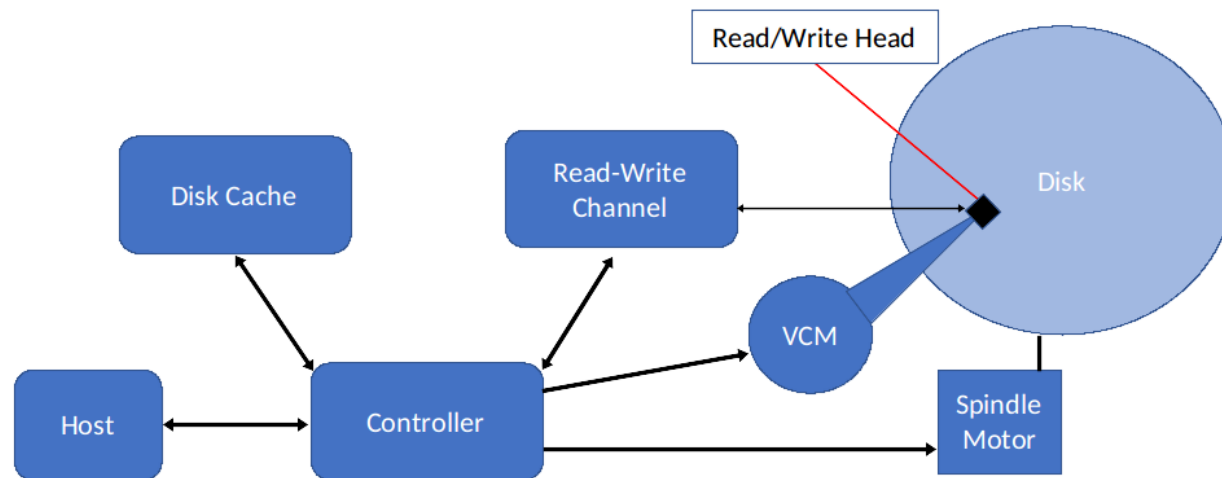
INSTITUTE FOR MACHINE LEARNING AND ANALYTICS

Hochschule Offenburg
offenburg.university

Agenda

- ▶ Hard Disk Drive
- ▶ Isolation Forest
- ▶ SmartPred
  - ▶ Dataset
  - ▶ Feature Selection
  - ▶ Setup
  - ▶ Evaluation
 - ▶ Replacement Process

Hard Disk Drive

Structure

Read/Write Head

Disk

Disk Cache

Read-Write Channel

VCM

Host

Controller

Spindle Motor

Failures

▶ Disk

▶ Mechanical

▶ Electrical

▶ Controller Firmware

Hard Disk Drive – SMART

Attribute	Definition
SMART 1	Read Error Rate
SMART 3	Spin Up Time
SMART 4	Start/Stop Count
SMART 5	Reallocated Sectors Count
SMART 7	Seek Error Rate
SMART 9	Power-On Hours
SMART 10	Spin Retry Count
SMART 12	Power Cycle Count
SMART 184	End-to-End error / IOEDC
SMART 187	Reported Uncorrectable Errors
SMART 188	Command Timeout

Attribute	Definition
SMART 190	Airflow Temperature
SMART 191	G-sense Error Rate
SMART 192	Power-off Retract Count
SMART 193	Load Cycle Count
SMART 194	Temperature
SMART 197	Current Pending Sector Count
SMART 198	Uncorrectable Sector Count
SMART 199	UltraDMA CRC Error Count
SMART 240	Head Flying Hours
SMART 241	Total LBAs Written
SMART 242	Total LBAs Read

# Isolation Forest
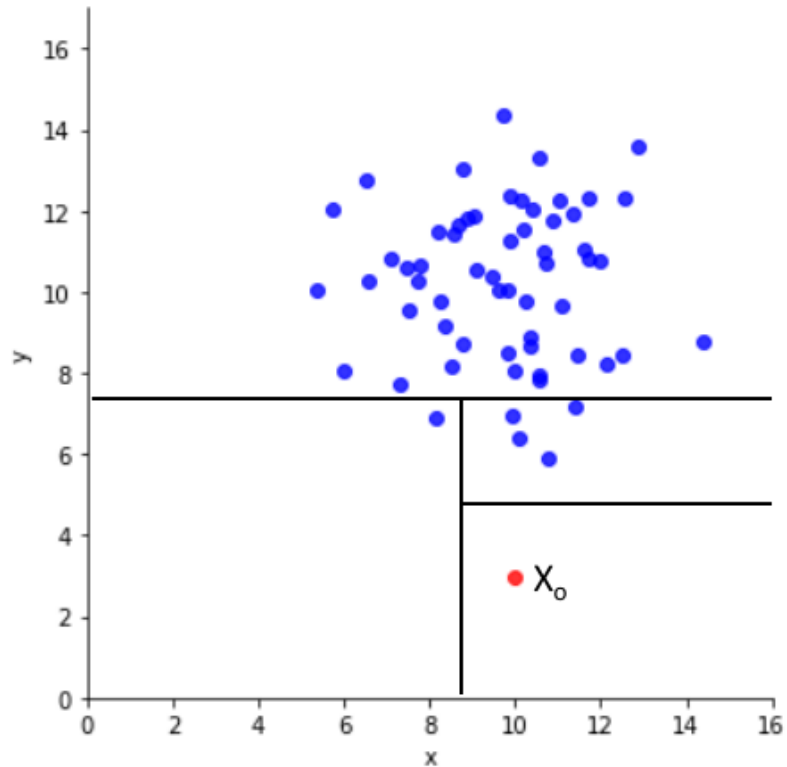
Anomaly X_0
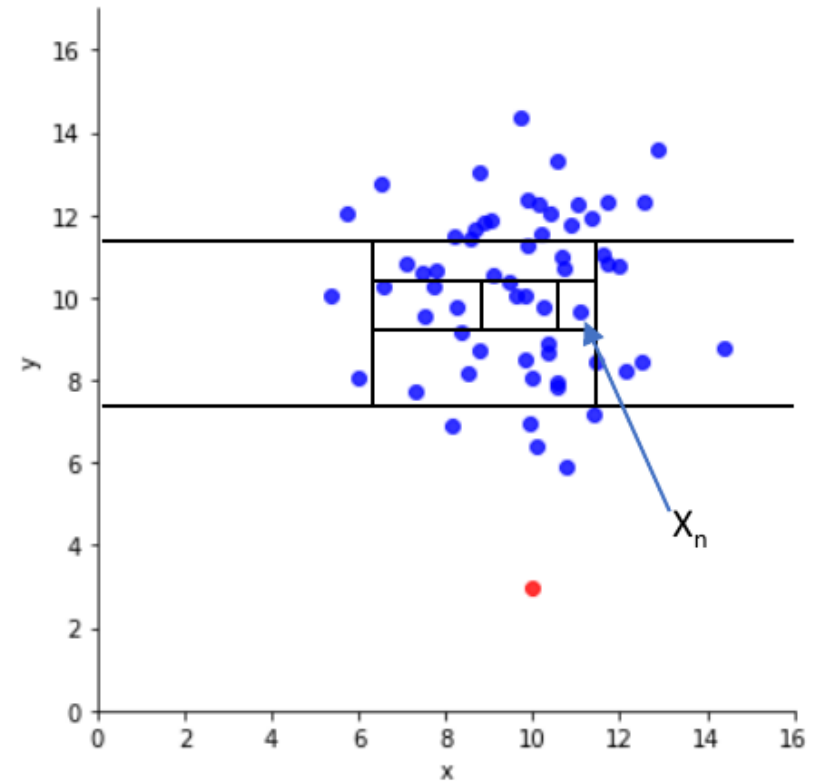
Normal Instance X_n

SmartPred - Dataset

date	serial_number	model	capacity_bytes	failure	smart_5_normalized	smart_5_raw
01.01.2019	Z305B2QN	ST4000DM000	4,00079E+12	0	119.0	0
01.01.2019	ZJV0XJQ4	ST12000NM0007	1,20001E+13	0	82.0	85
01.01.2019	ZJV0XJQ3	ST12000NM0007	1,20001E+13	0	80.0	255
01.01.2019	ZJV0XJQ0	ST12000NM0007	1,20001E+13	1	78.0	1034
01.01.2019	PL1331LAHG1S4H	HGSTHMS5C4040ALE64	4,00079E+12	0	100.0	0

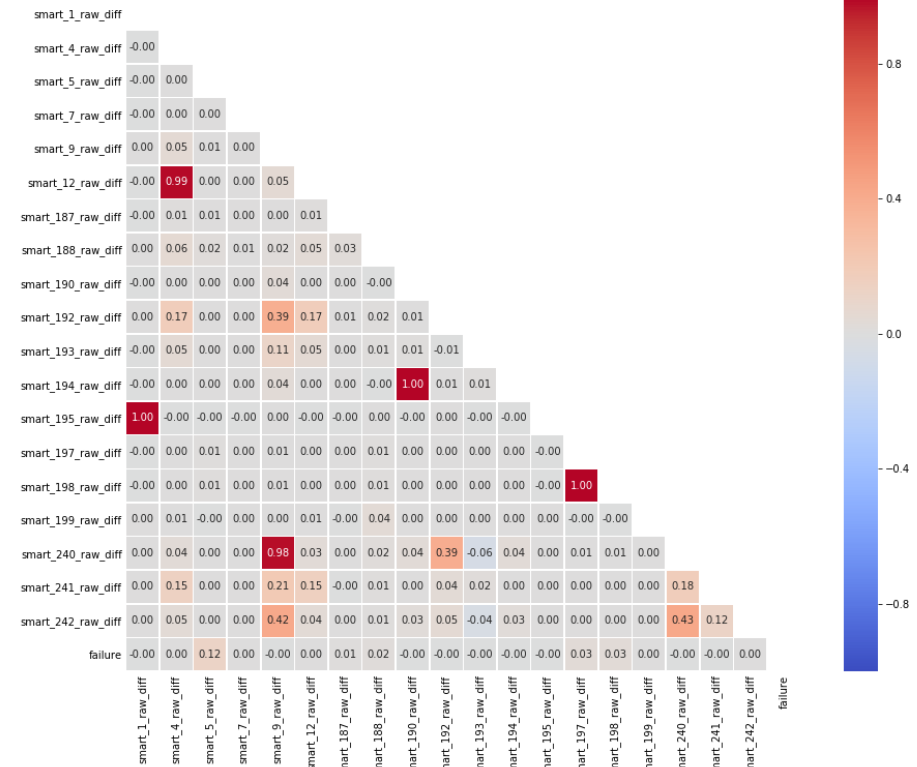▶ Samples: 12614746   ▶ Disks: 37768   ▶ Failed Disks: 868
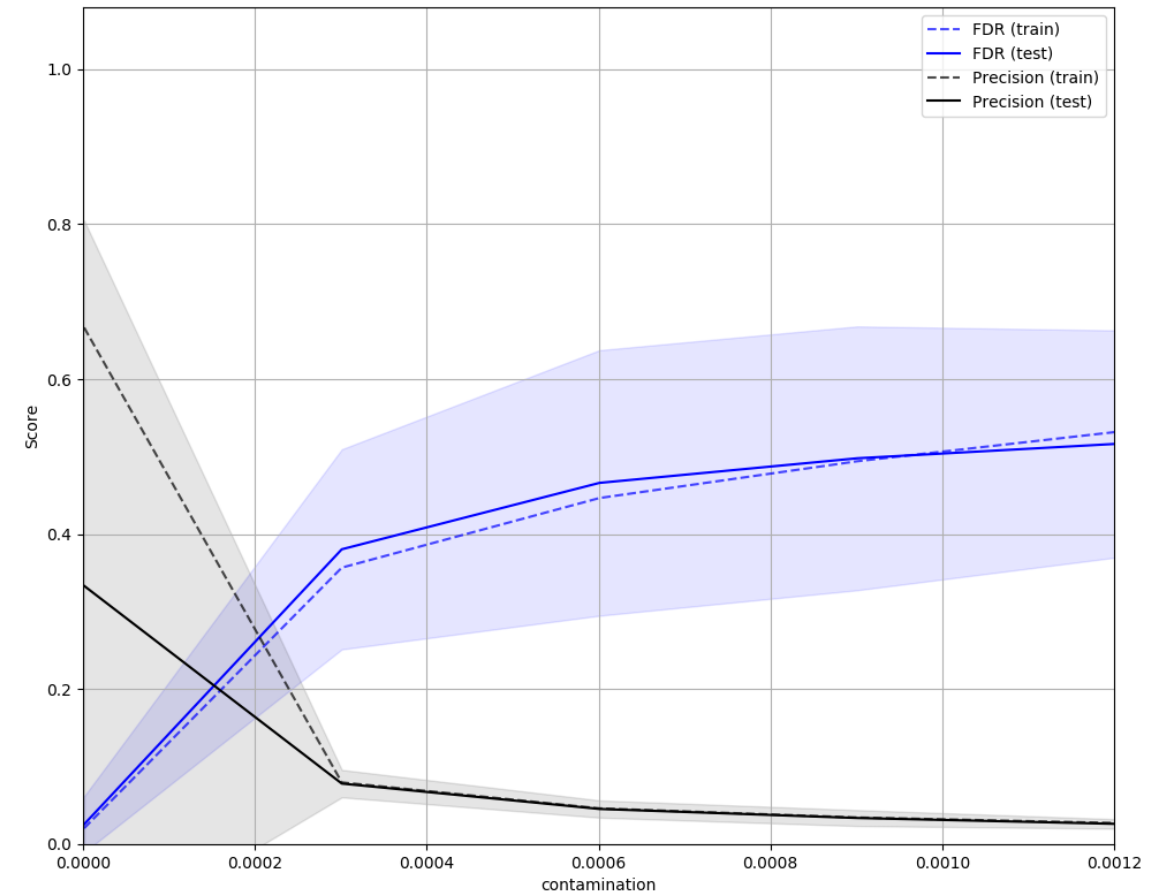
SmartPred - HDD Operating Time
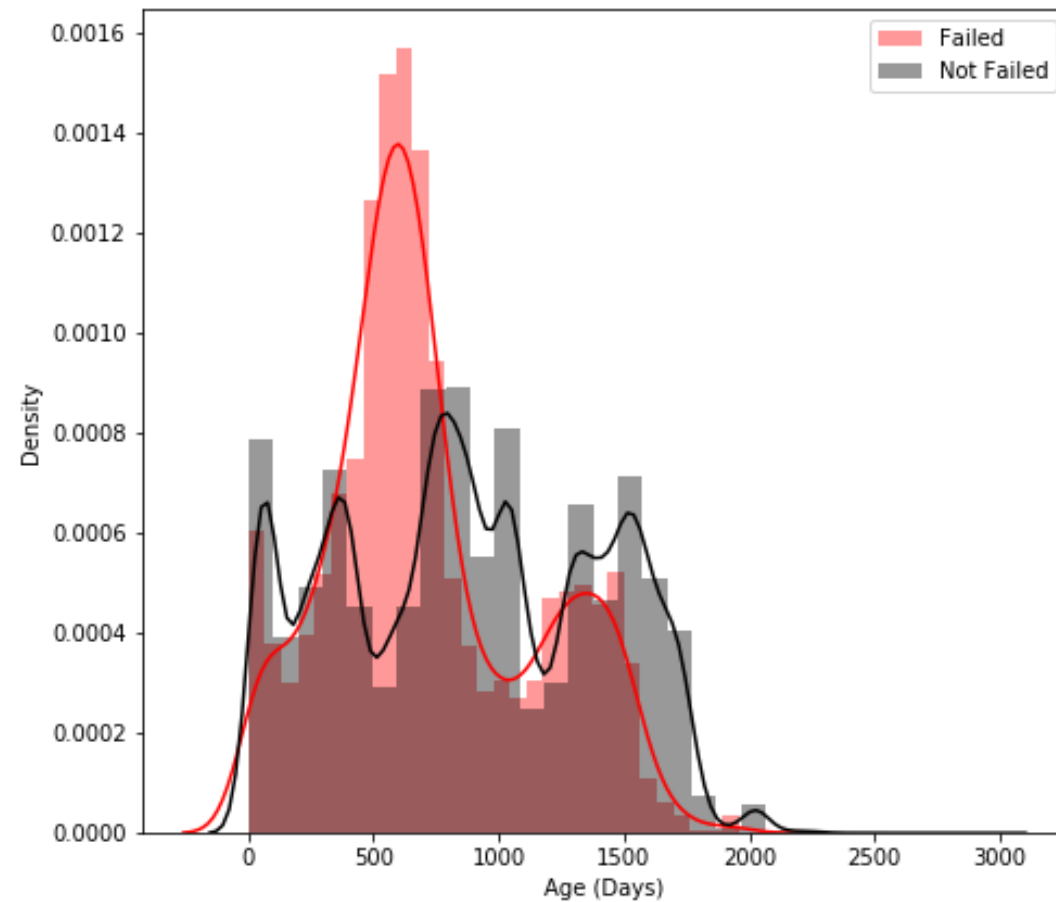
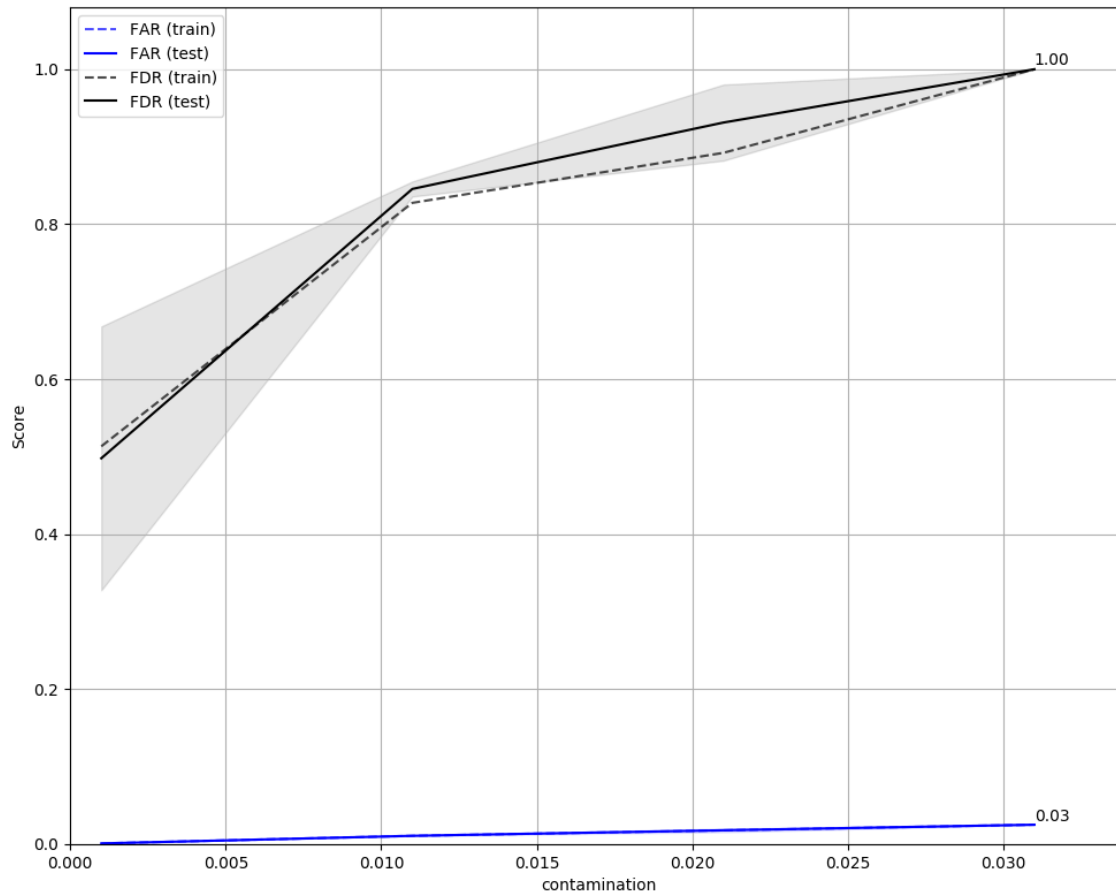# SmartPred - Feature Selection

Correlation coefficient to feature *Failure*
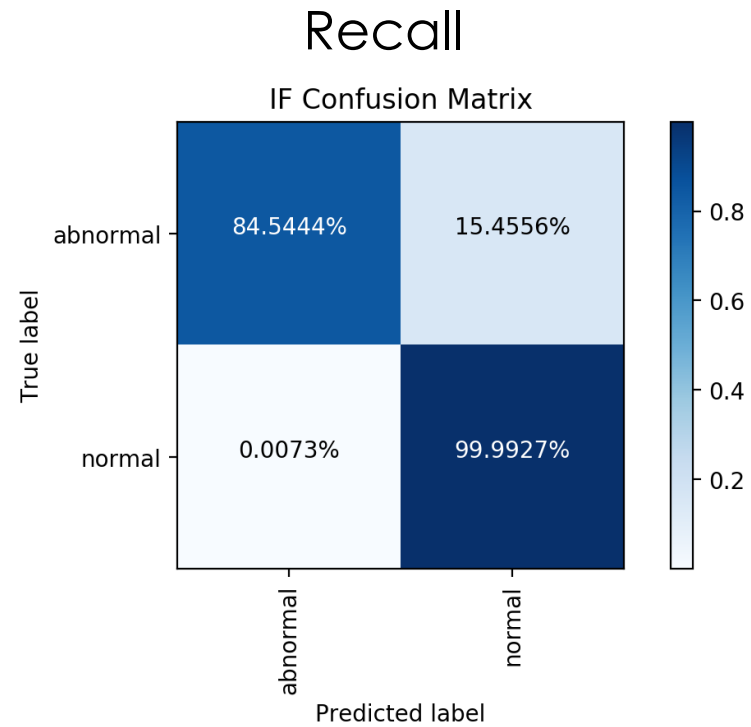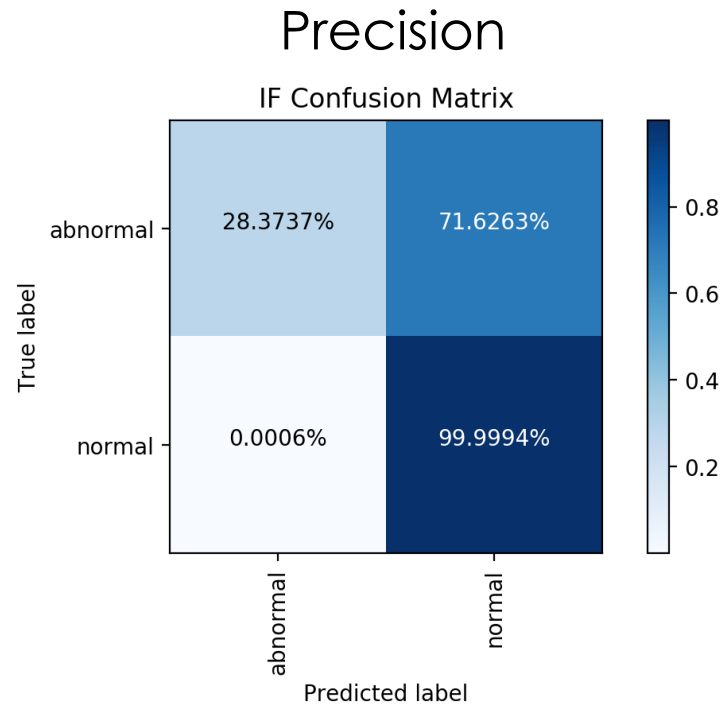
▶ SMART_5_raw: 0.07

▶ SMART_187_raw: 0.02

▶ SMART_197_raw: 0.03

▶ SMART_5_diff: 0.12

▶ SMART_187_diff: 0.01

▶ SMART_197_diff: 0.03

SmartPred - Hyperparameter
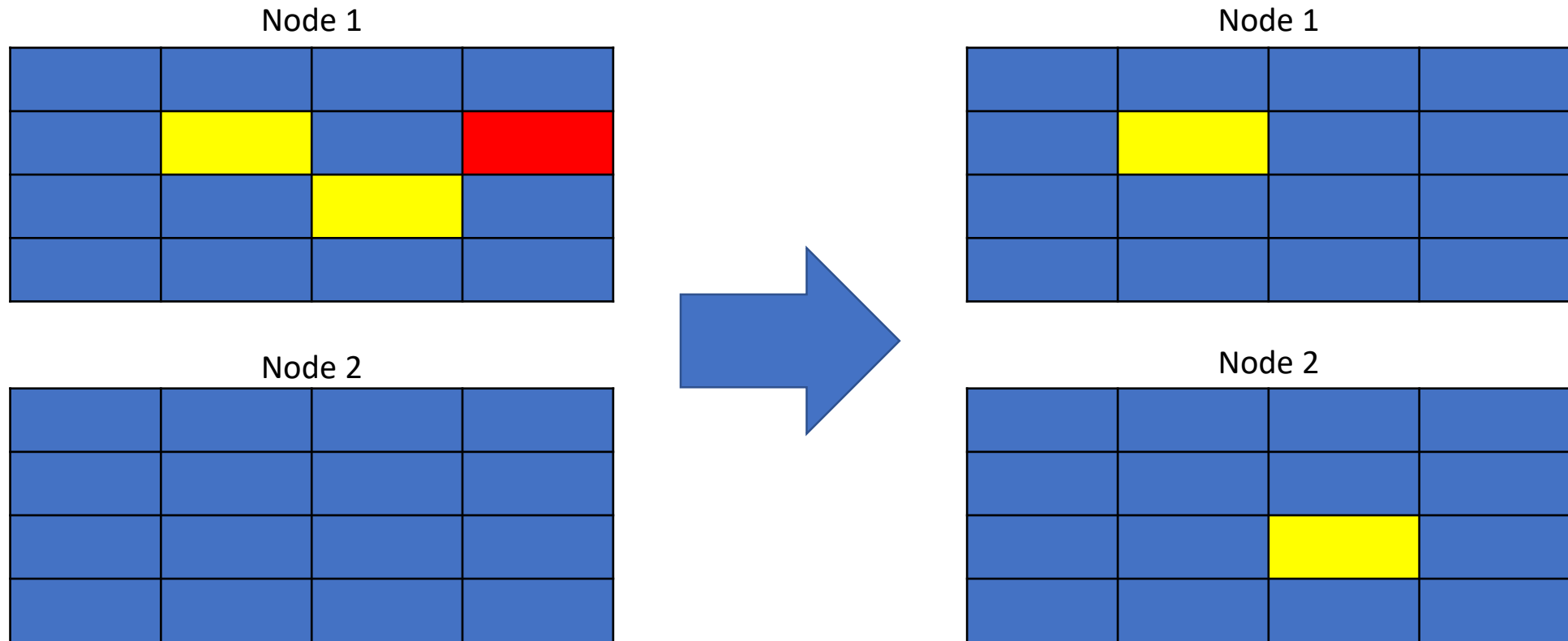
SmartPred - Evaluation

Precision — IF Confusion Matrix

Recall — IF Confusion Matrix

SmartPred - Evaluation

	FDR	FAR	Precision	Computation Time
IF_FDR	84.54%	0.0073%	44.21%	1.10 s ± 10 ms
IF_FAR	28.37%	0.0006%	77.85%	1.10 s ± 10 ms
OneClassSVM	95.39%	0.2925%	2.19%	15.9 s ± 780 ms
LocalOutlierFactor	96.19%	1.1916%	0.55%	16.0 s ± 671 ms

SmartPred - Replacement Process

Thank you for watching!