AutoML and Explainable AI

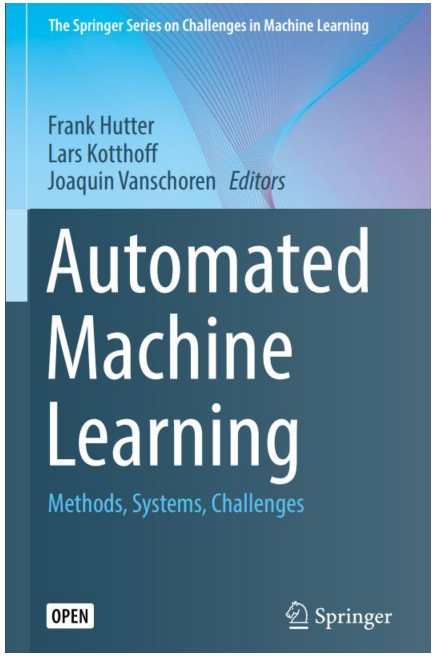
Assoc. Prof. Karl Ezra Pilario, Ph.D.
Process Systems Engineering Laboratory
Department of Chemical Engineering
University of the Philippines Diliman

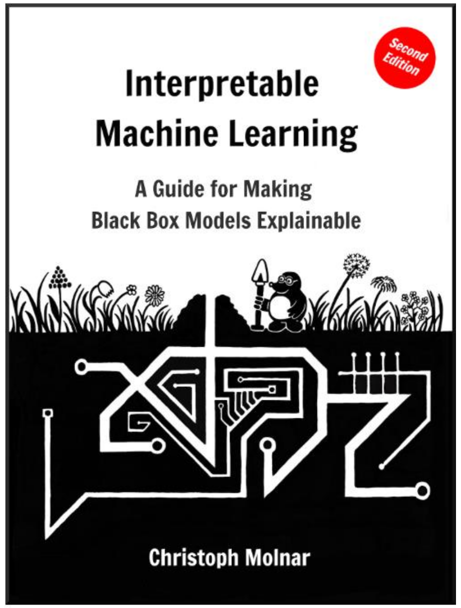
Outline

- AutoML Packages
  - Lazy Predict
  - Auto-sklearn
  - Optuna
  - TPOT
  - PyCaret

- Explainable AI (XAI)
  - Definitions and Concepts
  - Permutation Feature Importance
  - Drop-column Feature Importance
  - Mean-Decrease-in-Impurity
  - Shapley Additive Explanations

Hutter, Kotthoff, Vanschoren (2019)





Molnar (2022)

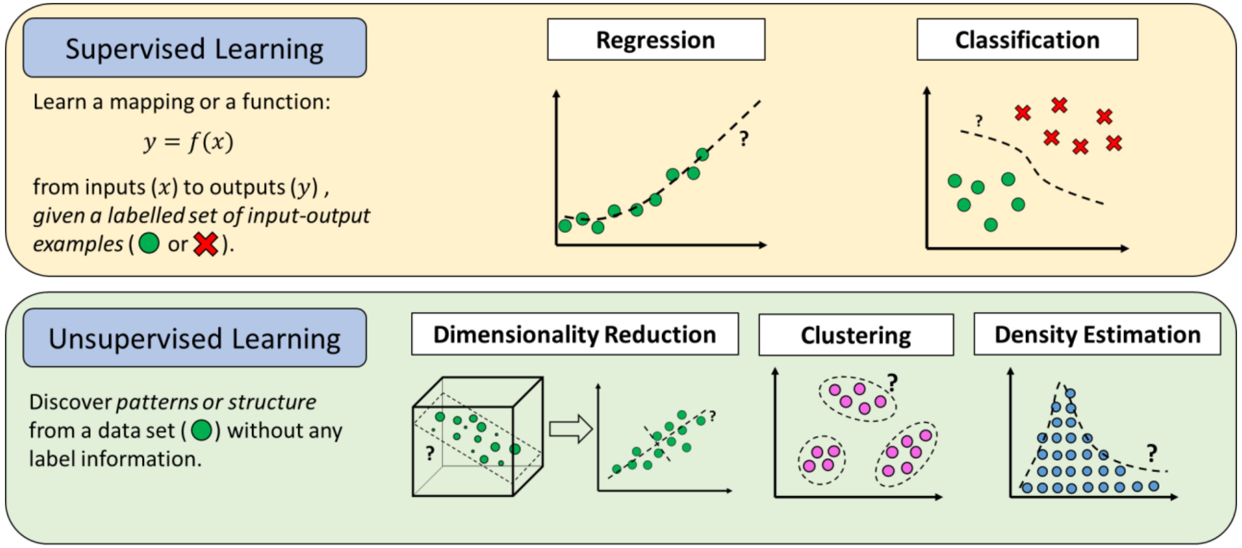
AutoML

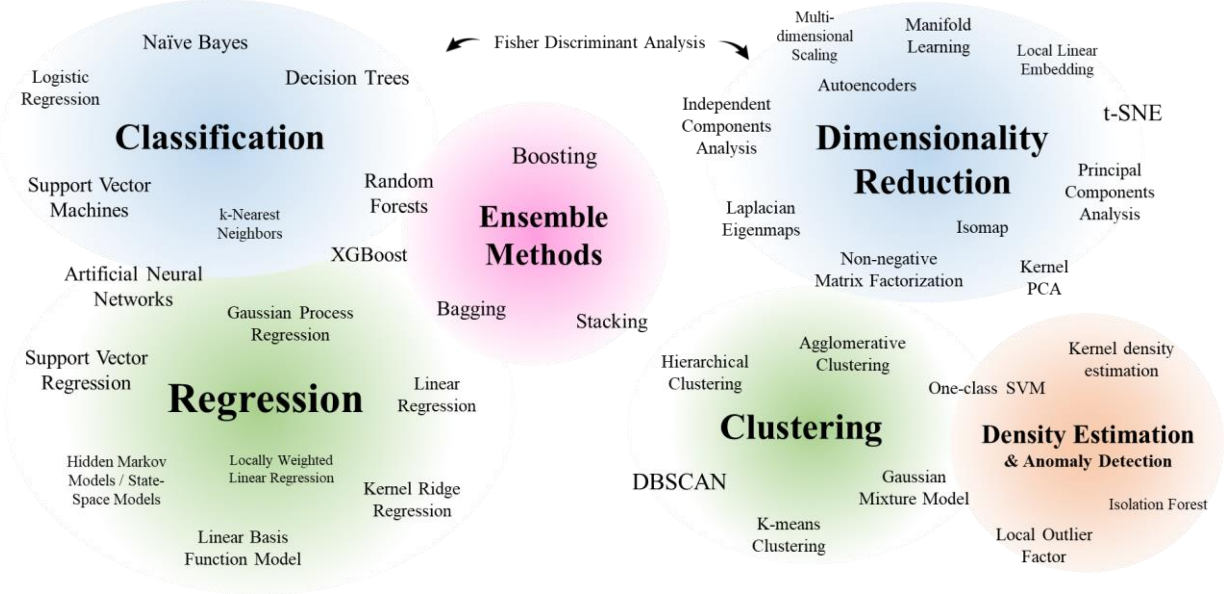
- Automatically discover best-performing models with little user involvement.

- For model comparison, AutoML offers a single hyper-parameter optimization toolkit for all models.

- **Meta-learning:** *Learning to Learn*
  - The science of systematically observing how different ML approaches perform on a wide range of tasks, then learning from this experience to improve ML itself.

- **CASH:** ***C**ombined **A**lgorithm **S**election and **H**yper-parameter Optimization (Kotthoff et al., 2019)*
  - Automatically and simultaneously choosing a learning algorithm and setting its hyperparameters to optimize empirical performance.

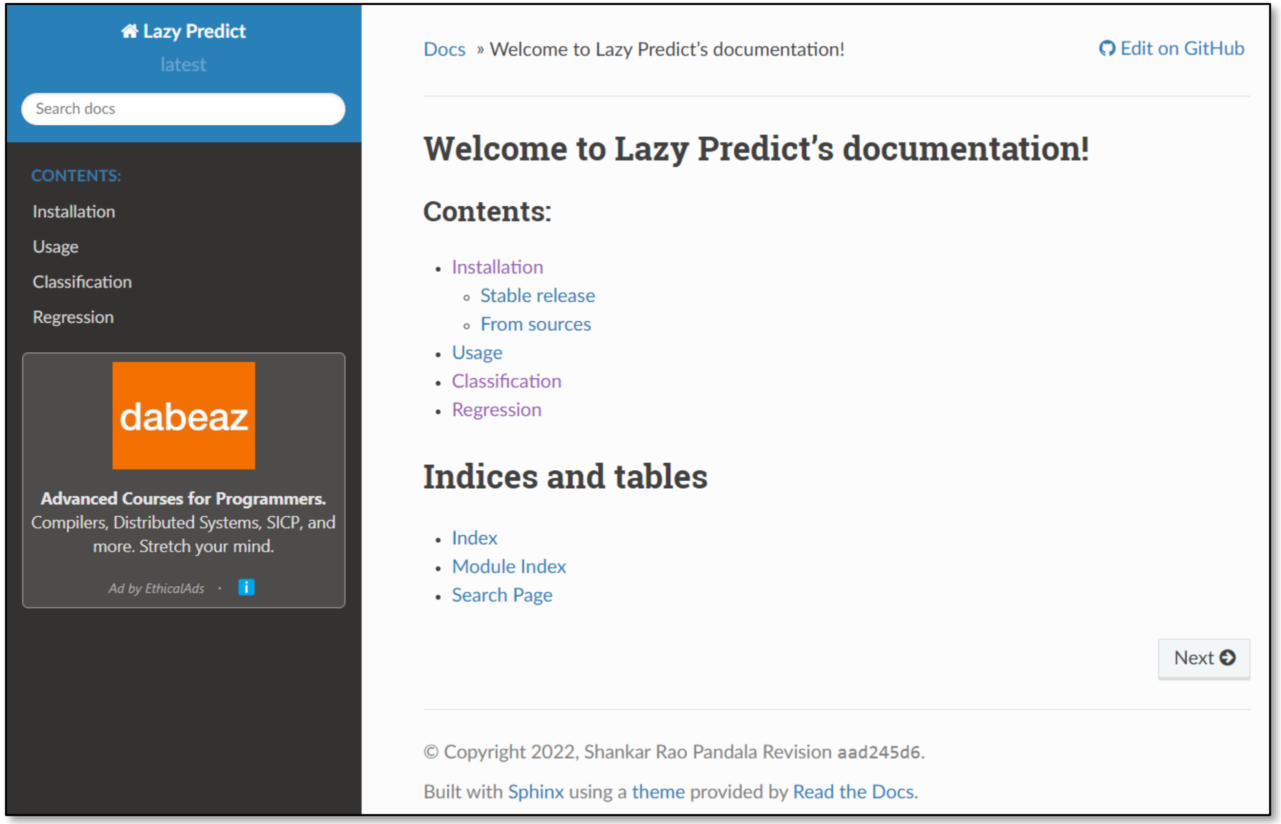




Lazy Predict

- Shankar Rao Pandala (Last Update: 2022)
- https://github.com/shankarpandala/lazypredict/tree/master
- https://lazypredict.readthedocs.io/en/latest/

- Fits a number of **scikit-learn** models on the data with default settings for all.
- Results: Accuracy, R2, F1-score, etc.
- No automatic model selection nor hyper-parameter tuning.
- For classification or regression only.



Lazy Predict

Example:

Use LazyClassifier on the **Breast Cancer Data Set** (this is the example from the website)

Ranked by Accuracy

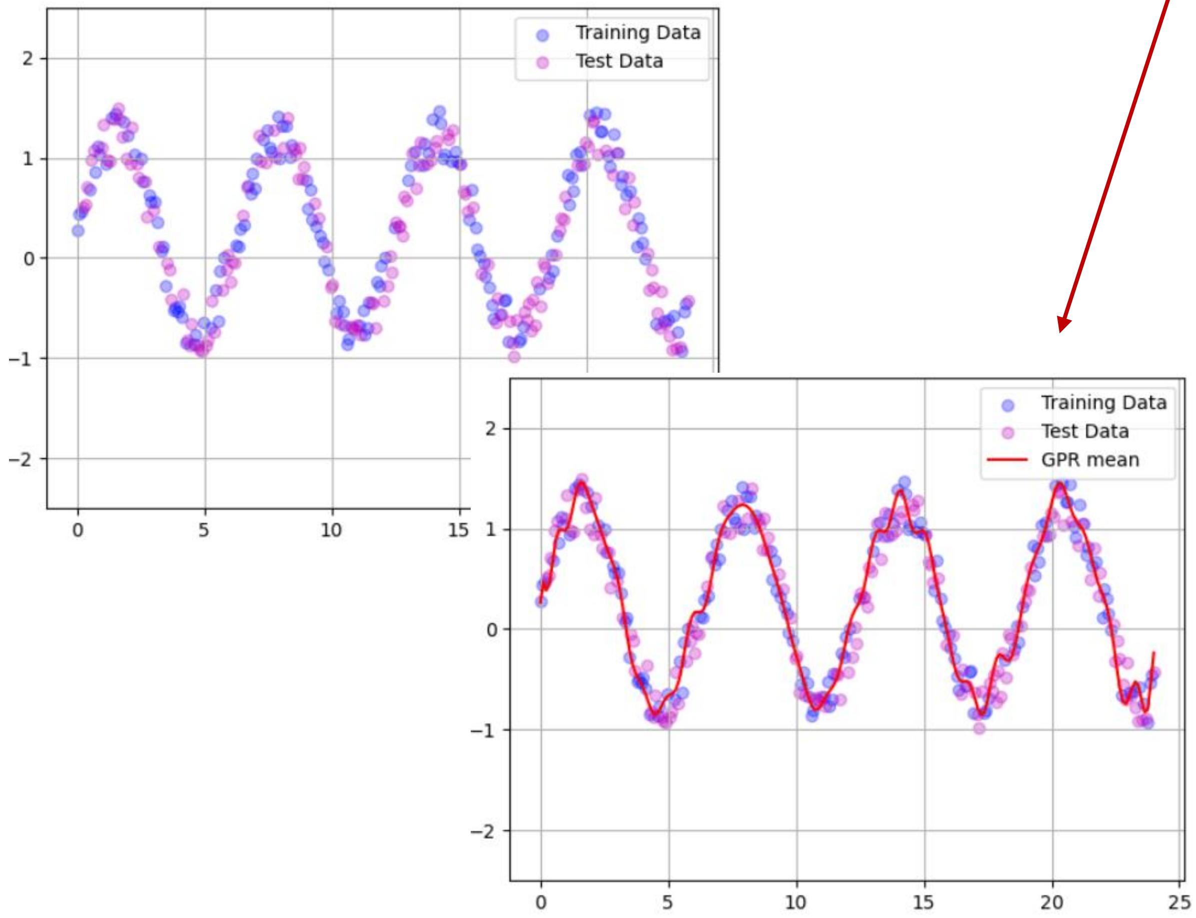
```
100%|██████████| 29/29 [00:01<00:00, 16.79it/s]
```

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
LinearSVC	0.99	0.99	0.99	0.99	0.02
Perceptron	0.99	0.98	0.98	0.99	0.02
LogisticRegression	0.99	0.98	0.98	0.99	0.03
SVC	0.98	0.98	0.98	0.98	0.02
XGBClassifier	0.98	0.98	0.98	0.98	0.13
LabelPropagation	0.98	0.97	0.97	0.98	0.03
LabelSpreading	0.98	0.97	0.97	0.98	0.03
BaggingClassifier	0.97	0.97	0.97	0.97	0.07
PassiveAggressiveClassifier	0.98	0.97	0.97	0.98	0.02
SGDClassifier	0.98	0.97	0.97	0.98	0.02
RandomForestClassifier	0.97	0.97	0.97	0.97	0.29
CalibratedClassifierCV	0.98	0.97	0.97	0.98	0.07
LGBMClassifier	0.97	0.97	0.97	0.97	0.17
QuadraticDiscriminantAnalysis	0.96	0.97	0.97	0.97	0.03
ExtraTreesClassifier	0.97	0.96	0.96	0.97	0.21
RidgeClassifierCV	0.97	0.96	0.96	0.97	0.02
RidgeClassifier	0.97	0.96	0.96	0.97	0.02
AdaBoostClassifier	0.96	0.96	0.96	0.96	0.29
KNeighborsClassifier	0.96	0.96	0.96	0.96	0.04
BernoulliNB	0.95	0.95	0.95	0.95	0.02
LinearDiscriminantAnalysis	0.96	0.95	0.95	0.96	0.03
GaussianNB	0.95	0.95	0.95	0.95	0.02
NuSVC	0.95	0.94	0.94	0.95	0.03
ExtraTreeClassifier	0.94	0.93	0.93	0.94	0.02
NearestCentroid	0.95	0.93	0.93	0.95	0.02
DecisionTreeClassifier	0.93	0.93	0.93	0.93	0.02
DummyClassifier	0.64	0.50	0.50	0.50	0.02

Lazy Predict

Example:

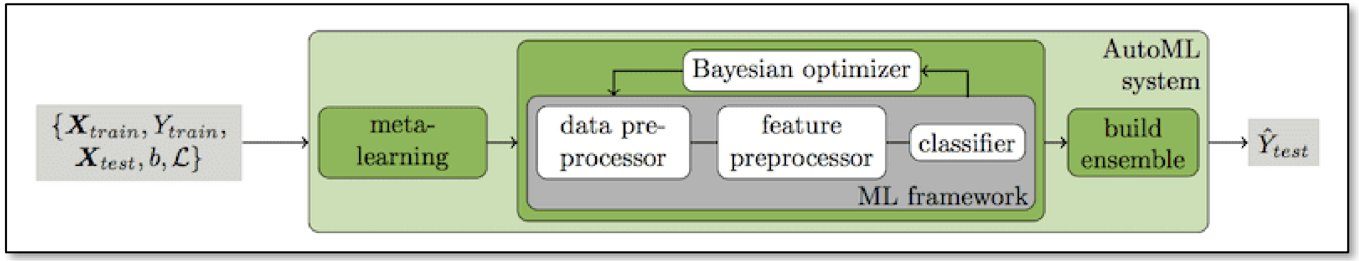
Use LazyRegressor on the **Sine Data Set.**

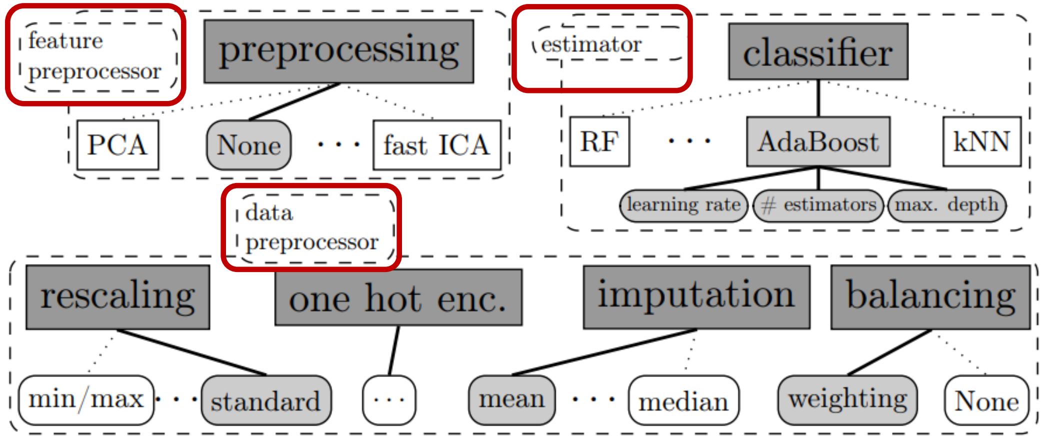



```
100%|██████████| 37/37 [00:01<00:00, 28.27it/s]
```

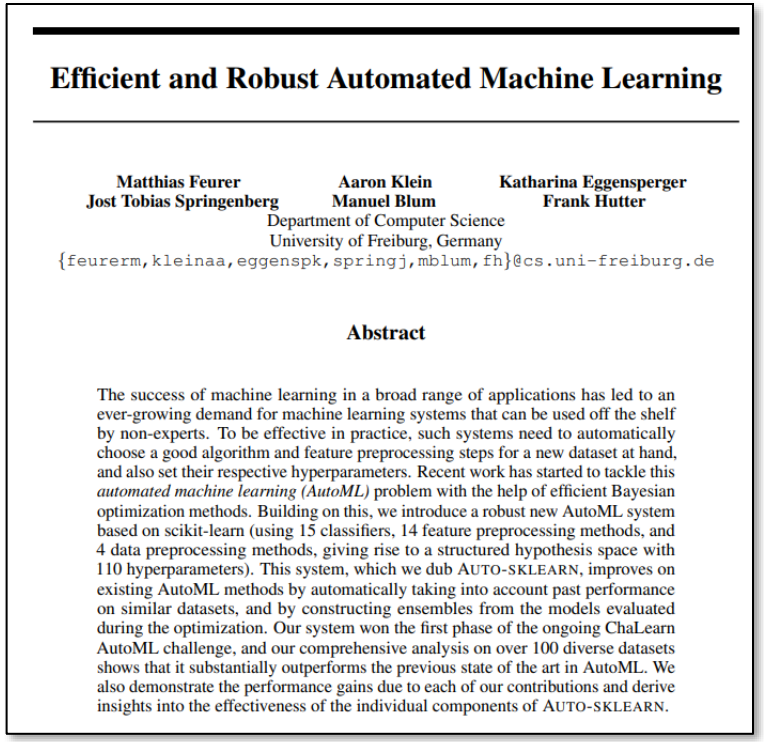
Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
GaussianProcessRegressor	0.95	0.95	0.16	0.03
KNeighborsRegressor	0.94	0.94	0.17	0.02
ExtraTreesRegressor	0.94	0.94	0.19	0.18
BaggingRegressor	0.93	0.93	0.19	0.05
GradientBoostingRegressor	0.93	0.93	0.20	0.08
ExtraTreeRegressor	0.91	0.91	0.22	0.01
DecisionTreeRegressor	0.91	0.91	0.22	0.01
XGBRegressor	0.90	0.90	0.23	0.07
HistGradientBoostingRegressor	0.78	0.78	0.34	0.10
LGBMRegressor	0.76	0.76	0.36	0.07
AdaBoostRegressor	0.55	0.56	0.49	0.08
NuSVR	0.12	0.12	0.69	0.02
SVR	0.11	0.11	0.70	0.02
MLPRegressor	0.04	0.05	0.72	0.10
LinearSVR	0.02	0.03	0.73	0.02
HuberRegressor	0.01	0.02	0.73	0.02
SGDRegressor	0.01	0.02	0.73	0.01
Lars	0.01	0.01	0.73	0.01
TransformedTargetRegressor	0.01	0.01	0.73	0.02
OrthogonalMatchingPursuit	0.01	0.01	0.73	0.01
LinearRegression	0.01	0.01	0.73	0.02
RidgeCV	0.01	0.01	0.73	0.02
TweedieRegressor	-0.00	0.00	0.74	0.01
BayesianRidge	-0.02	-0.01	0.74	0.02
ElasticNetCV	-0.02	-0.01	0.74	0.09
LassoLarsIC	-0.02	-0.01	0.74	0.01
LassoLarsCV	-0.02	-0.01	0.74	0.02
LassoLars	-0.02	-0.01	0.74	0.01
LassoCV	-0.02	-0.01	0.74	0.08
DummyRegressor	-0.02	-0.01	0.74	0.01
LarsCV	-0.02	-0.01	0.74	0.02
ElasticNet	-0.02	-0.01	0.74	0.01
Lasso	-0.02	-0.01	0.74	0.01
KernelRidge	-0.08	-0.07	0.76	0.01
PassiveAggressiveRegressor	-0.90	-0.89	1.02	0.02

Auto-Sklearn





- Feurer et al. (2015) and Feurer et al. (2022)
- https://automl.github.io/auto-sklearn/master/
- For regression and classification with pre-processing.
- A total of 110 tunable hyper-parameters across all models (2015).
- Can discover ensembles.
- Uses Bayesian Optimization and meta-learning.



Efficient and Robust Automated Machine Learning

Matthias Feurer **Aaron Klein** **Katharina Eggensperger**
Jost Tobias Springenberg **Manuel Blum** **Frank Hutter**
Department of Computer Science
University of Freiburg, Germany
{feurerm,kleinaa,eggenspk,springj,mblum,fh}@cs.uni-freiburg.de

Abstract

The success of machine learning in a broad range of applications has led to an ever-growing demand for machine learning systems that can be used off the shelf by non-experts. To be effective in practice, such systems need to automatically choose a good algorithm and feature preprocessing steps for a new dataset at hand, and also set their respective hyperparameters. Recent work has started to tackle this *automated machine learning (AutoML)* problem with the help of efficient Bayesian optimization methods. Building on this, we introduce a robust new AutoML system based on scikit-learn (using 15 classifiers, 14 feature preprocessing methods, and 4 data preprocessing methods, giving rise to a structured hypothesis space with 110 hyperparameters). This system, which we dub AUTO-SKLEARN, improves on existing AutoML methods by automatically taking into account past performance on similar datasets, and by constructing ensembles from the models evaluated during the optimization. Our system won the first phase of the ongoing ChaLearn AutoML challenge, and our comprehensive analysis on over 100 diverse datasets shows that it substantially outperforms the previous state of the art in AutoML. We also demonstrate the performance gains due to each of our contributions and derive insights into the effectiveness of the individual components of AUTO-SKLEARN.

Optuna

- Akiba et al. (2019)
- Suitable for CASH (algorithm selection + hyper-parameter tuning)
- Models and hyper-parameters are user-defined.
- Uses Bayesian Optimization.

OPTUNA

Optuna: A hyperparameter optimization framework

Akiba et al., (2019) Optuna: A Next-generation Hyperparameter Optimization Framework. https://arxiv.org/pdf/1907.10902.pdf

Optuna has modern functionalities as follows:

Lightweight, versatile, and platform agnostic architecture
- Handle a wide variety of tasks with a simple installation that has few requirements.

Pythonic search spaces
- Define search spaces using familiar Python syntax including conditionals and loops.

Efficient optimization algorithms
- Adopt state-of-the-art algorithms for sampling hyperparameters and efficiently pruning unpromising trials.

Easy parallelization
- Scale studies to tens or hundreds of workers with little or no changes to the code.

Quick visualization
- Inspect optimization histories from a variety of plotting functions.

- Grid Search implemented in `GridSampler`
- Random Search implemented in `RandomSampler`
- Tree-structured Parzen Estimator algorithm implemented in `TPESampler`
- CMA-ES based algorithm implemented in `CmaEsSampler`
- Algorithm to enable partial fixed parameters implemented in `PartialFixedSampler`
- Nondominated Sorting Genetic Algorithm II implemented in `NSGAIISampler`
- A Quasi Monte Carlo sampling algorithm implemented in `QMCSampler`

The default sampler is `TPESampler`.

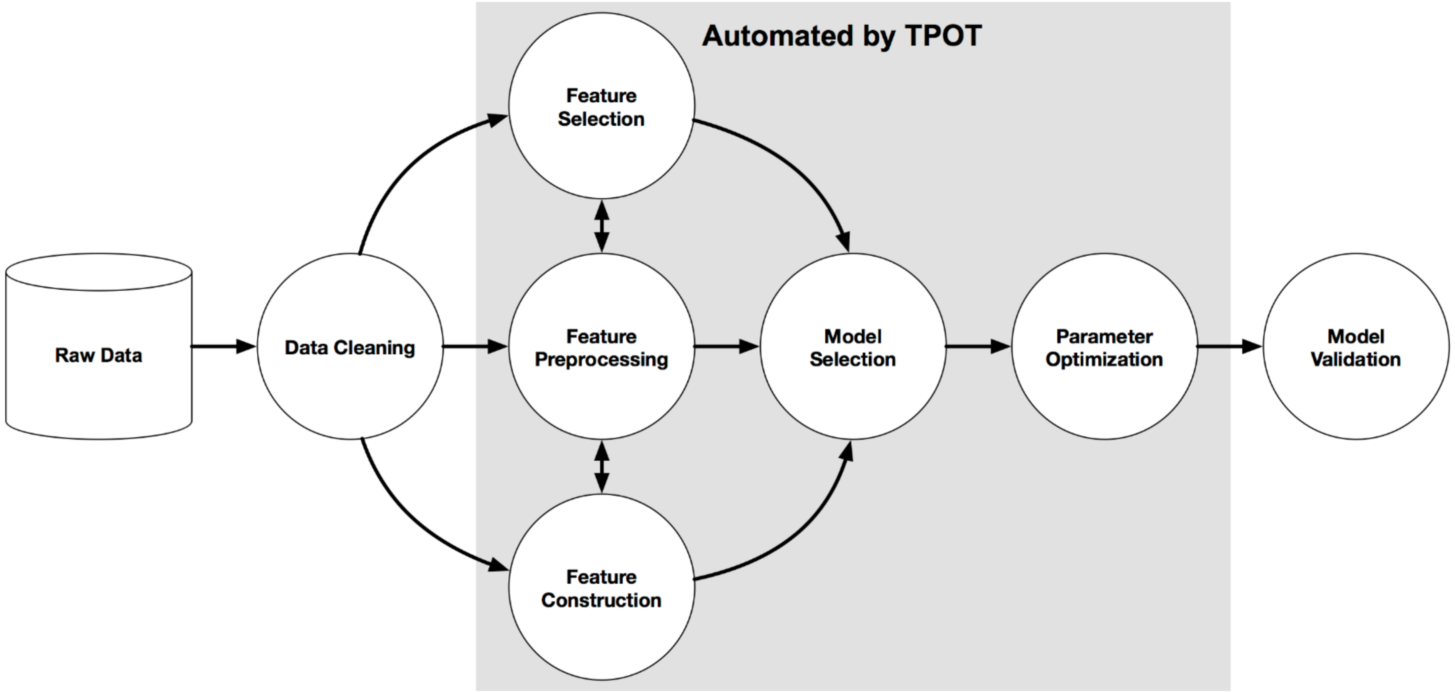
Main algorithm:

TPE (Tree-structured Parzen Estimator)

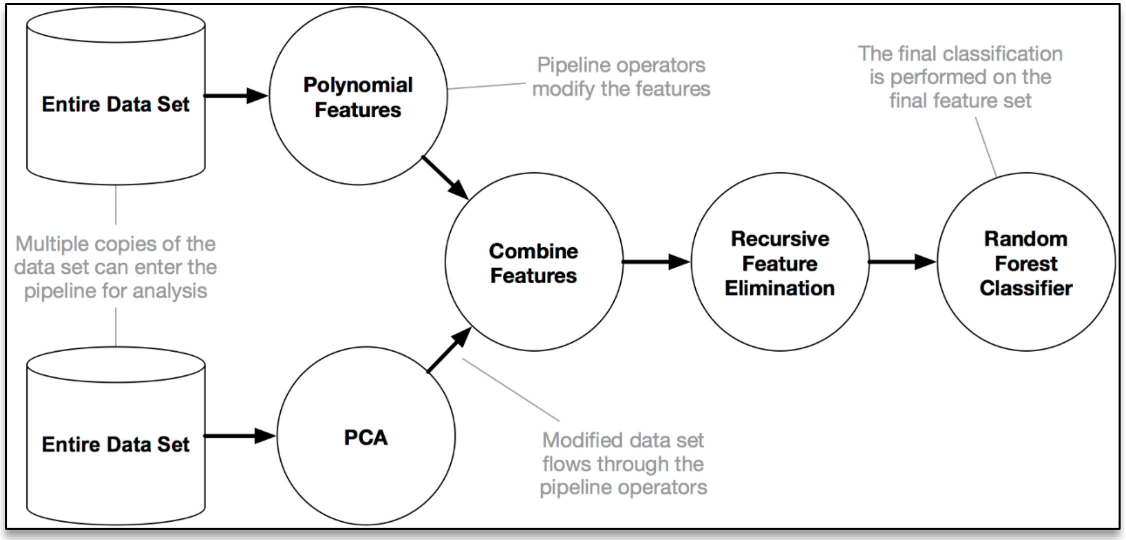
- A variant of Bayesian Optimization

TPOT

- Olson and Moore (2016)
- http://epistasislab.github.io/tpot/
- TPOT = **T**ree-based **P**ipeline **O**ptimization **T**ool
- TPOT optimizes machine learning pipelines using genetic programming.

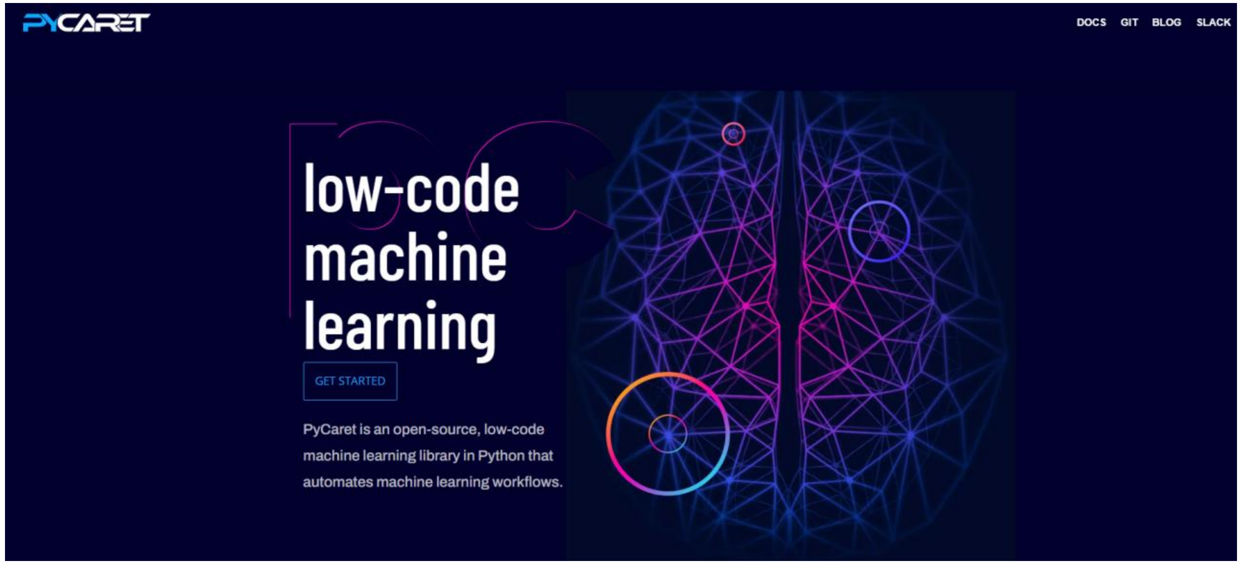


Sample Result:



PyCaret

- Moez Ali (2020)
- https://pycaret.gitbook.io/docs/
- low-code library: replace hundreds of lines of code with a few lines only.



Example: **Regression**

```python
# Regression Functional API Example

# loading sample dataset
from pycaret.datasets import get_data
data = get_data('insurance')

# init setup
from pycaret.regression import *
s = setup(data, target = 'charges', session_id = 123)

# model training and selection
best = compare_models()

# evaluate trained model
evaluate_model(best)

# predict on hold-out/test set
pred_holdout = predict_model(best)

# predict on new data
new_data = data.copy().drop('charges', axis = 1)
predictions = predict_model(best, data = new_data)

# save model
save_model(best, 'best_pipeline')
```

Example: **Classification**

```python
# Classification Functional API Example

# loading sample dataset
from pycaret.datasets import get_data
data = get_data('juice')

# init setup
from pycaret.classification import *
s = setup(data, target = 'Purchase', session_id = 123)

# model training and selection
best = compare_models()

# evaluate trained model
evaluate_model(best)

# predict on hold-out/test set
pred_holdout = predict_model(best)

# predict on new data
new_data = data.copy().drop('Purchase', axis = 1)
predictions = predict_model(best, data = new_data)

# save model
save_model(best, 'best_pipeline')
```

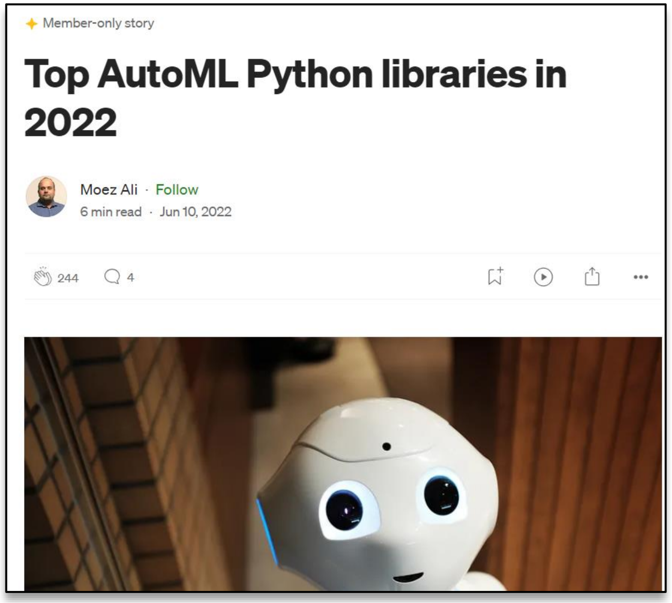
Example: **Anomaly Detection**

```python
# Anomaly Detection Functional API Example

# loading sample dataset
from pycaret.datasets import get_data
data = get_data('anomaly')

# init setup
from pycaret.anomaly import *
s = setup(data, session_id = 123)

# model training
iforest = create_model('iforest')

# assign labels from trained model
results = assign_model(iforest)

# evaluate trained model
evaluate_model(iforest)

# predict on new_data
new_data = data.copy()
predictions = predict_model(iforest, data = new_data)

# save model
save_model(iforest, 'iforest_pipeline')
```

Other AutoML Libraries

According to Moez Ali (from PyCaret), here are the **top AutoML libraries** in 2022.

1. PyCaret
2. H_2O AutoML
3. TPOT
4. Auto-sklearn
5. FLAML
6. EvalML
7. AutoKeras
8. Auto-ViML
9. AutoGluon
10. MLBox



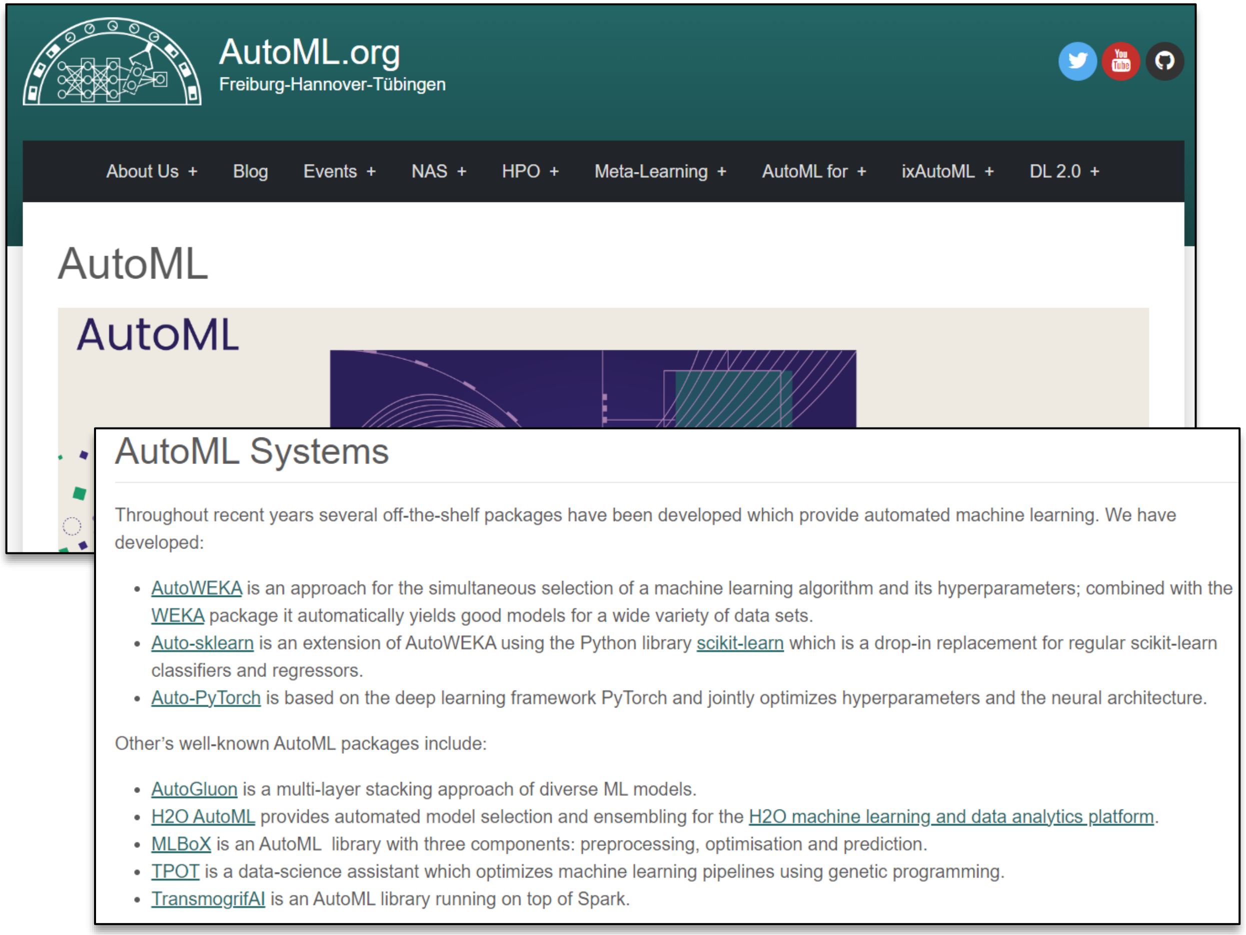
Member-only story

Top AutoML Python libraries in 2022

Moez Ali · Follow
6 min read · Jun 10, 2022

244 4

https://www.automl.org/automl/



AutoML.org
Freiburg-Hannover-Tübingen

About Us + Blog Events + NAS + HPO + Meta-Learning + AutoML for + ixAutoML + DL 2.0 +

AutoML

AutoML

AutoML Systems

Throughout recent years several off-the-shelf packages have been developed which provide automated machine learning. We have developed:

- AutoWEKA is an approach for the simultaneous selection of a machine learning algorithm and its hyperparameters; combined with the WEKA package it automatically yields good models for a wide variety of data sets.
- Auto-sklearn is an extension of AutoWEKA using the Python library scikit-learn which is a drop-in replacement for regular scikit-learn classifiers and regressors.
- Auto-PyTorch is based on the deep learning framework PyTorch and jointly optimizes hyperparameters and the neural architecture.

Other's well-known AutoML packages include:

- AutoGluon is a multi-layer stacking approach of diverse ML models.
- H2O AutoML provides automated model selection and ensembling for the H2O machine learning and data analytics platform.
- MLBoX is an AutoML library with three components: preprocessing, optimisation and prediction.
- TPOT is a data-science assistant which optimizes machine learning pipelines using genetic programming.
- TransmogrifAI is an AutoML library running on top of Spark.

Comparison of AutoML Libraries

A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost

Luís Ferreira, EPMQ - IT, CCG ZGDV Institute, ALGORITMI Center, University of Minho, Guimarães, Portugal, luis.ferreira@ccg.pt
André Pilastri, EPMQ - IT, CCG ZGDV Institute, Guimarães, Portugal, andre.pilastri@ccg.pt
Carlos Manuel Martins, WeDo Technologies, Braga, Portugal, carlos.mmartins@mobileum.com
Pedro Miguel Pires, WeDo Technologies, Braga, Portugal, pedro.mpires@mobileum.com
Paulo Cortez, ALGORITMI Center, Dep. Information Systems, University of Minho, Guimarães, Portugal, pcortez@dsi.uminho.pt

***Abstract*—This paper presents a benchmark of supervised Automated Machine Learning (AutoML) tools. Firstly, we analyze the characteristics of eight recent open-source AutoML tools (Auto-Keras, Auto-PyTorch, Auto-Sklearn, AutoGluon, H2O AutoML, rminer, TPOT and TransmogrifAI) and describe twelve popular OpenML datasets that were used in the benchmark (divided into regression, binary and multi-class classification tasks). Then, we perform a comparison study with hundreds of computational experiments based on three scenarios: General Machine Learning (GML), Deep Learning (DL) and XGBoost (XGB). To select the best tool, we used a lexicographic approach, considering first the average prediction score for each task and**

algorithm selection; Deep Learning (DL) selection and XGBoost (XGB) hyperparameter tuning. Each tool is measured in terms of its predictive performance (using an external 10-fold cross-validation) and computational cost (measured in terms of time elapsed). Moreover, the best AutoML tools are further compared with the best public OpenML predictive results (which are assumed as the "gold standard").

The paper is organized as follows. Section 2 presents the related work. Next, Section 3 describes the AutoML tools and datasets. Section 4 details the benchmark design. Then,

Reference: Ferreira et al. (2021). A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost. *Proceedings of the International Joint Conference on Neural Networks.*

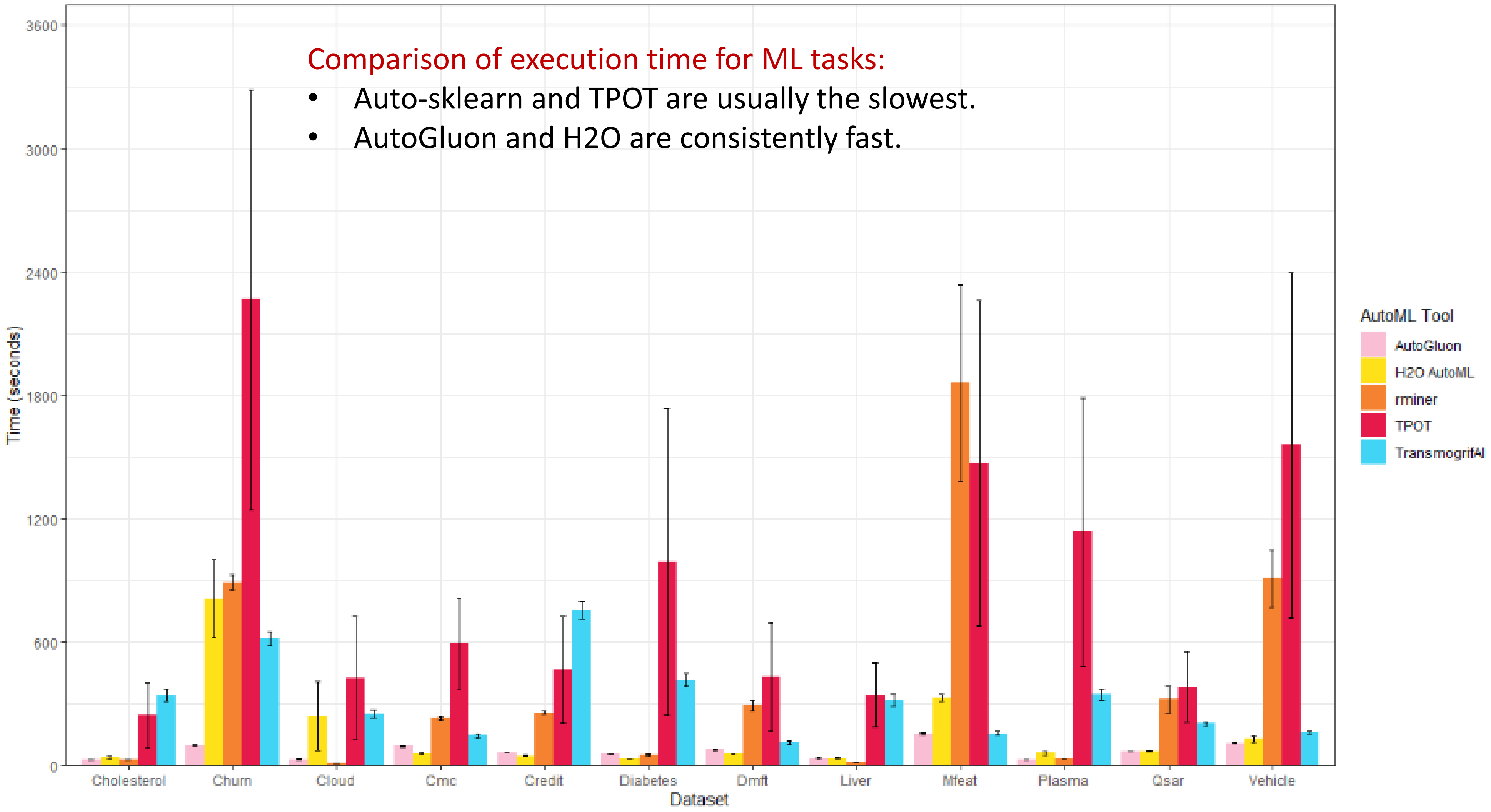
AutoML Tool	Framework	API Lang.	Operating Systems	DL	Scenario GML	Scenario DL	Scenario XGB
Auto-Keras	Keras	Python	MacOs Linux Windows	Yes (only)		✓	
Auto-PyTorch	PyTorch	Python	MacOs Linux Windows	Yes (only)		✓	
Auto-Sklearn	Scikit-Learn	Python	Linux	No	✓		
AutoGluon	PyTorch	Python	MacOS (P.) Linux	Yes	✓	✓	
H2O AutoML	H2O	Java Python R	MacOs Linux Windows (P.)	Yes	✓	✓	✓
rminer AutoML	rminer	R	MacOs Linux Windows	No	✓		✓
TPOT	Scikit-Learn	Python	MacOs Linux Windows	No	✓		
TransmogrifAI	Spark (MLlib)	Scala	MacOs Linux Windows	No	✓		

- In this work, the authors compared 8 different AutoML tools (see Table).
- Twelve different OpenML data sets were used to benchmark the AutoML tools.
    - Binary classification
    - Multi-class classification
    - Regression

Comparison of AutoML Libraries

Comparison of execution time for ML tasks:

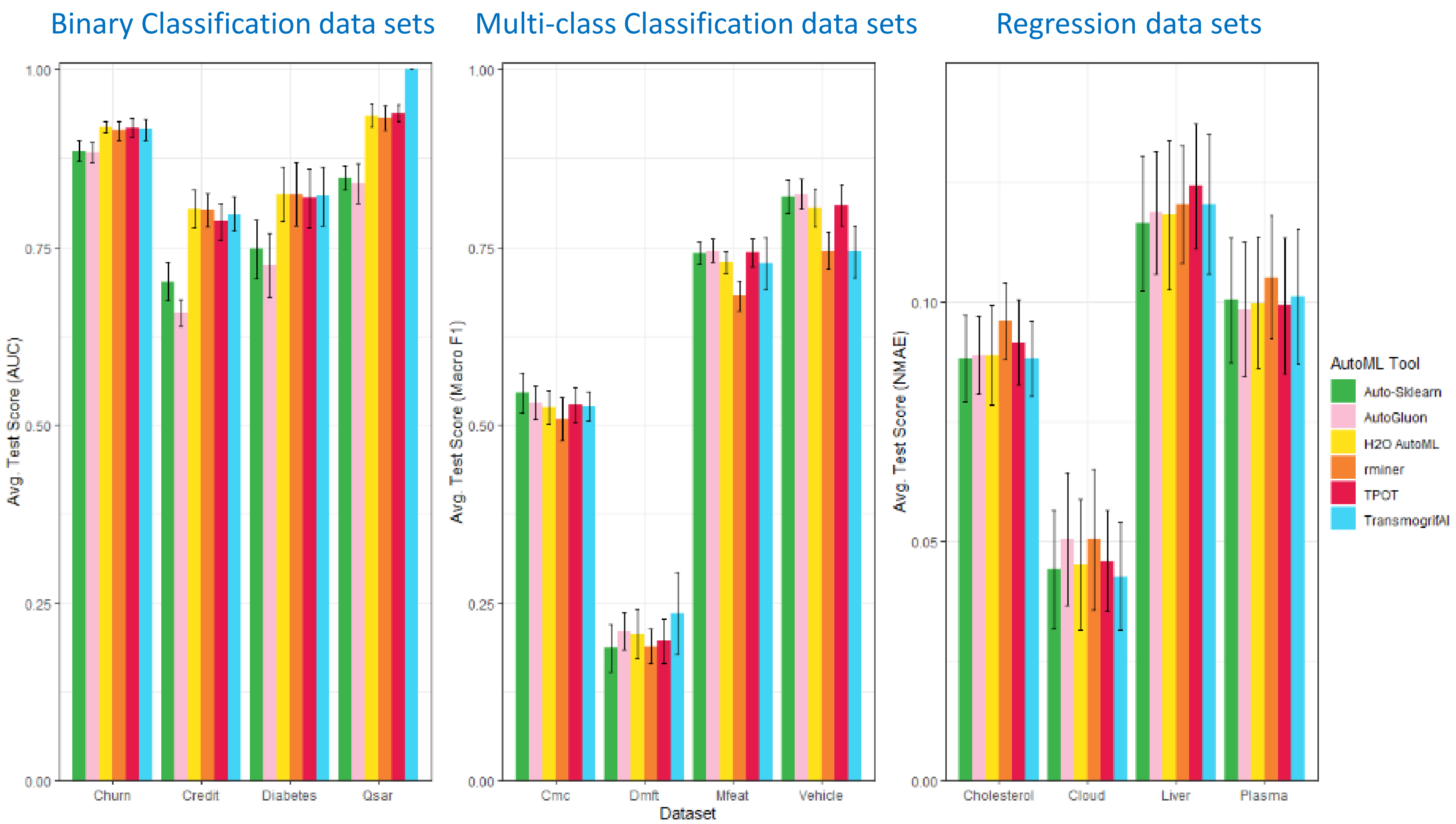
- Auto-sklearn and TPOT are usually the slowest.
- AutoGluon and H2O are consistently fast.



Comparison of AutoML Libraries

Comparison of performance for ML tasks:

- For binary classification, **TransmogrifAI** is best for 3 out of 4 data sets. **AutoGluon** and **Auto-sklearn** produced the worst overall results.
- For multi-class classification, **AutoGluon** and **Auto-sklearn** are best.
- For regression, differences between tools are not that significant. But the best overall is **rminer**.



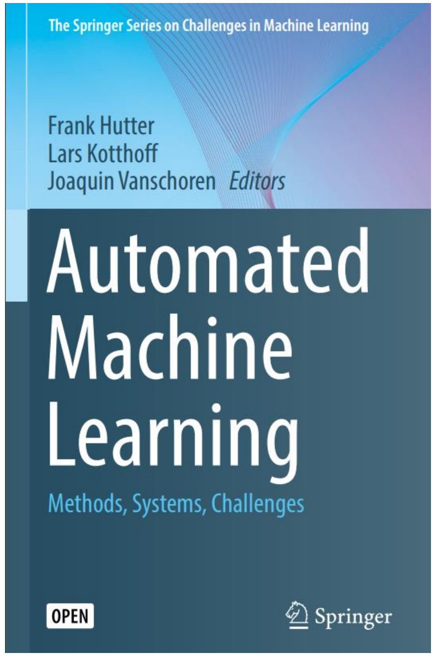
Reference: Ferreira et al. (2021). A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost. *Proceedings of the International Joint Conference on Neural Networks.*

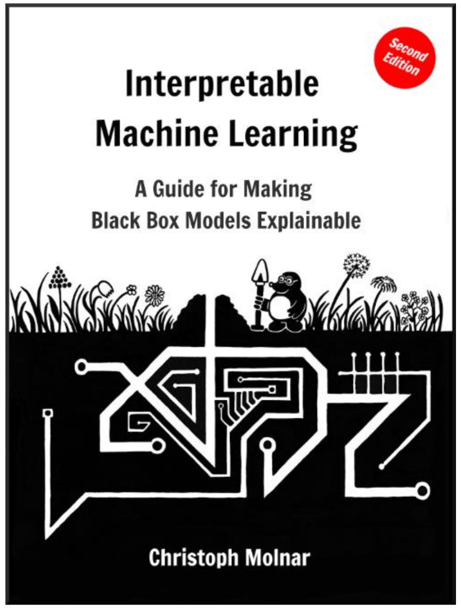
Outline

- AutoML Packages
  - Lazy Predict
  - Auto-sklearn
  - Optuna
  - TPOT
  - PyCaret

- Explainable AI (XAI)
  - Definitions and Concepts
  - Permutation Feature Importance
  - Drop-column Feature Importance
  - Mean-Decrease-in-Impurity
  - Shapley Additive Explanations

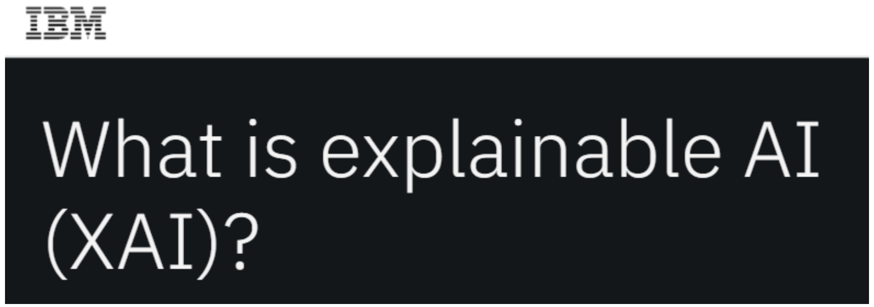
Hutter, Kotthoff, Vanschoren (2019)





Molnar (2022)

Explainable AI (XAI)

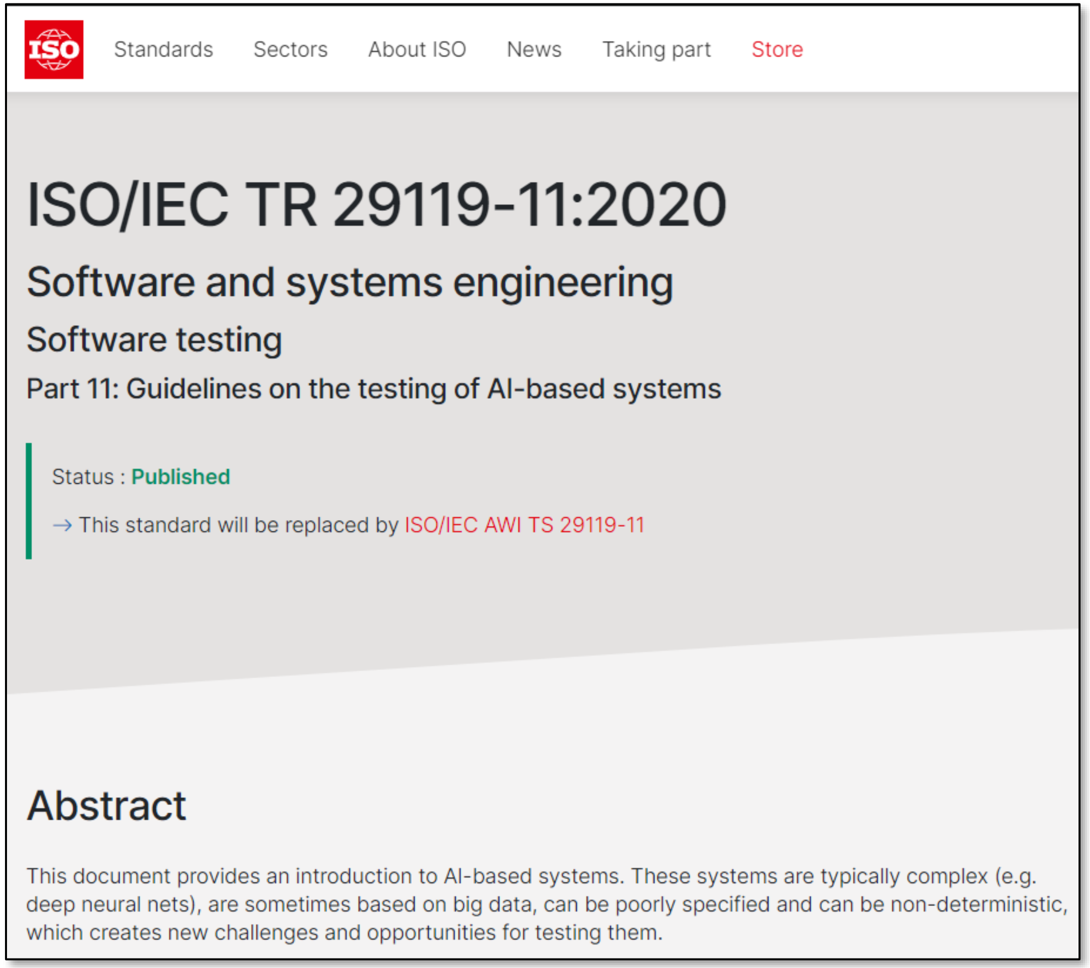


Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to **comprehend and trust** the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected **impact** and **potential biases**. It helps characterize model **accuracy**, **fairness**, **transparency** and outcomes in AI-powered decision making.

Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explainability also helps an organization adopt a **responsible** approach to AI development.

- IBM (https://www.ibm.com/watson/explainable-ai)

Explainable AI (XAI)


Standards Sectors About ISO News Taking part Store

ISO/IEC TR 29119-11:2020

Software and systems engineering

Software testing

Part 11: Guidelines on the testing of AI-based systems

Status : Published

→ This standard will be replaced by ISO/IEC AWI TS 29119-11

Abstract

This document provides an introduction to AI-based systems. These systems are typically complex (e.g. deep neural nets), are sometimes based on big data, can be poorly specified and can be non-deterministic, which creates new challenges and opportunities for testing them.

Source: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:tr:29119:-11:ed-1:v1:en

According to an ISO Standard, the following terms are defined:

Artificial Intelligence
The capability of an engineered system to acquire, process, and apply knowledge and skills.

Machine Learning
A process using computational techniques to enable systems to *learn from data or experience*.

AI-based System
A system including one or more components implementing AI.

Explainability
Level of understanding of how the AI-based system *came up with a given result.*

Interpretability
Level of understanding of how the underlying (AI) technology works.

Transparency
Level of accessibility to the algorithm and data used by the AI-based system.

Explainable AI (XAI)

Understandability

Ability of a model to make a human understand its *internal structure* and how it works *algorithmically*.

Comprehensibility

Ability of a learning algorithm to represent its learned knowledge in a human understandable fashion.

Interpretability

Refers to how accurate a machine learning model can associate a cause to an effect.

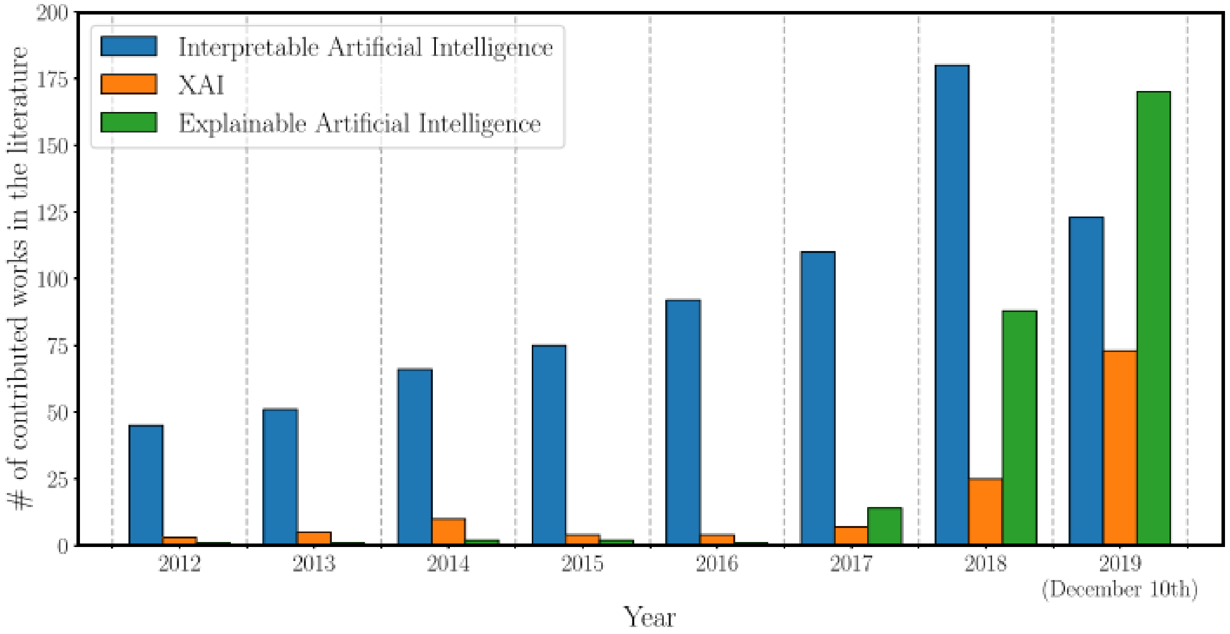
Transparency

A model is transparent if, by itself, it is already understandable.

Explainability

Ability of a model to explain its results to humans:

- How did it arrive at its decisions?
- Which inputs in the data prompted the decision to change?
- Which features have a significant effect on the prediction?

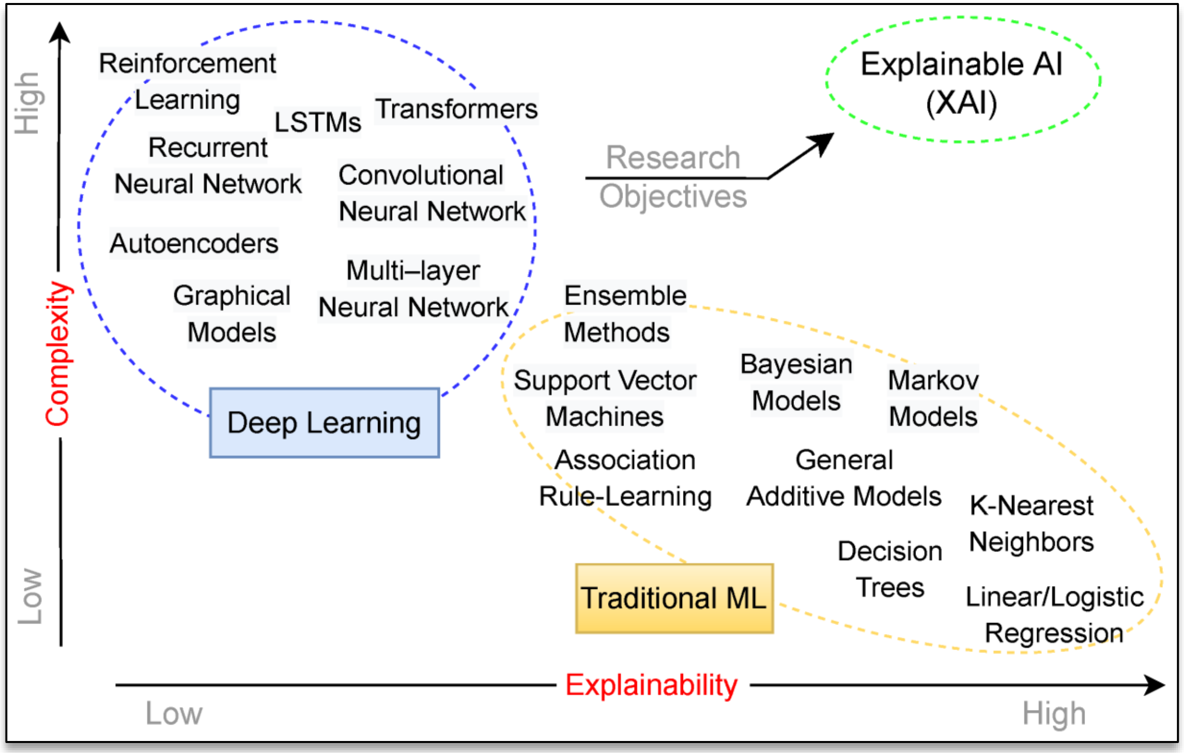


Number of Papers in Literature that mentioned XAI

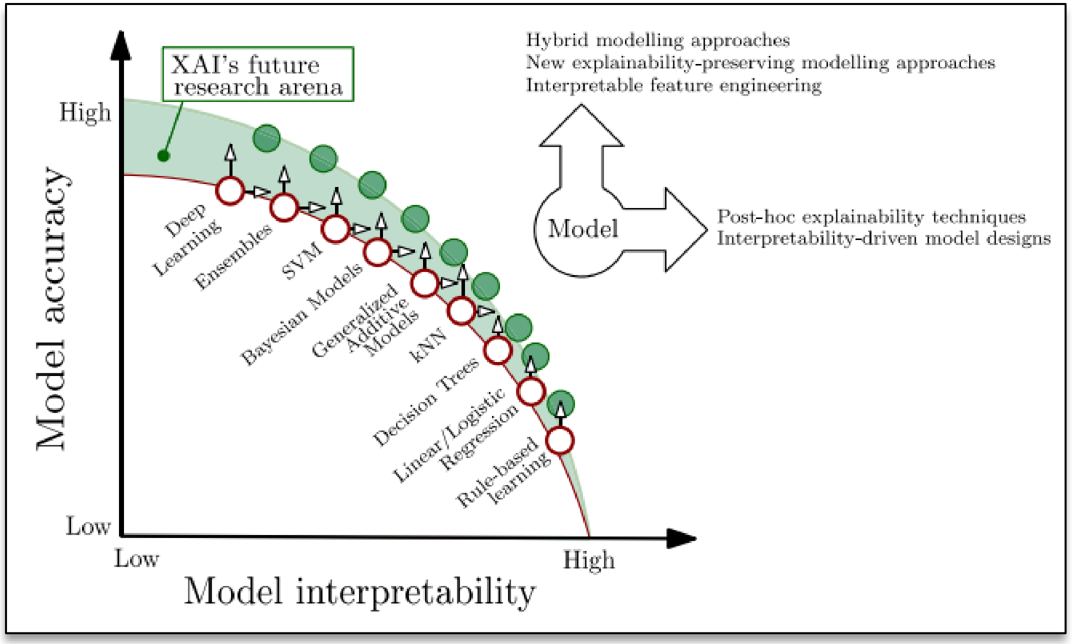
Reference: Arrieta et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, Vol. 58, June 2020, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Explainable AI (XAI)

- It is said that traditional ML models are explainable, but are low-performing.
- On the other hand, deep learning models are not explainable but high-performing.
- Explainable AI aims to provide models that are *explainable yet high-performing*.



Reference: Clement, T.; Kemmerzell, N.; Abdelaal, M.; Amberg, M. XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process. Mach. Learn. Knowl. Extr. 2023, 5, 78-108. https://doi.org/10.3390/make5010006



Reference: Arrieta et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, Vol. 58, June 2020, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Explainable AI (XAI)

The role of explainable AI in the context of the AI Act

Cecilia Panigutti*
cecilia.panigutti@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Ronan Hamon*
ronan.hamon@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Isabelle Hupont*
isabelle.hupont-torres@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

David Fernandez Llorca
david.fernandez-llorca@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Delia Fano Yela
delia.fano-yela@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Henrik Junklewitz
henrik.junklewitz@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Salvatore Scalzo
salvatore.scalzo@ec.europa.eu
European Commission
Brussels, Belgium

Gabriele Mazzini
gabriele.mazzini@ec.europa.eu
European Commission
Brussels, Belgium

Ignacio Sanchez
ignacio.sanchez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Josep Soler Garrido
josep.soler-garrido@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

Emilia Gomez
emilia.gomez-gutierrez@ec.europa.eu
European Commission, Joint Research Centre (JRC)
Sevilla, Spain

ABSTRACT

The proposed EU regulation for Artificial Intelligence (AI), the AI Act, has sparked some debate about the role of explainable AI (XAI) in high-risk AI systems. Some argue that black-box AI models will have to be replaced with transparent ones, others argue that using XAI techniques might help in achieving compliance. This work aims to bring some clarity as regards XAI in the context of the AI Act and focuses in particular on the AI Act requirements for transparency and human oversight. After outlining key points of

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; • **Applied computing → Law**.

KEYWORDS

explainable artificial intelligence, XAI, AI Act, EU regulation, trustworthy AI, transparency, human oversight

ACM Reference Format:

ARTICLE Part of Section 2: Requirements for High-Risk AI Systems

Article 13: Transparency and Provision of Information to Deployers

Feedback – We are working to improve this tool. Please send feedback to Risto Uuk at risto@futureoflife.org

1. High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable deployers to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured with a view to achieving compliance with the relevant obligations of the provider and deployer set out in Section 3 of this Title.

2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information

ARTICLE Part of Section 2: Requirements for High-Risk AI Systems

Article 14: Human Oversight

Feedback – We are working to improve this tool. Please send feedback to Risto Uuk at risto@futureoflife.org

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.

2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Section.

3. The oversight measures shall be commensurate to the risks, level of autonomy and context of use of the AI system and shall be ensured through either one or all of the

Source: https://dl.acm.org/doi/10.1145/3593013.3594069
https://artificialintelligenceact.eu/article/13/

Are AI companies compliant?

Grading Foundation Model Providers' Compliance with the Draft EU AI Act

Source: Stanford Research on Foundation Models (CRFM), Institute for Human-Centered Artificial Intelligence (HAI)

	OpenAI	cohere	stability.ai	ANTHROPIC	Google	BigScience	Meta	AI21 labs	ALEPH ALPHA	EleutherAI	
Draft AI Act Requirements	**GPT-4**	**Cohere Command**	**Stable Diffusion v2**	**Claude**	**PaLM 2**	**BLOOM**	**LLaMA**	**Jurassic-2**	**Luminous**	**GPT-NeoX**	**Totals**
Data sources	●○○○	●●●○	●●●●	○○○○	●●○○	●●●●	●●●●	○○○○	○○○○	●●●●	22
Data governance	●●○○	●●●○	●●○○	○○○○	●●●○	●●●●	●●○○	○○○○	○○○○	●●●○	19
Copyrighted data	○○○○	○○○○	○○○○	○○○○	○○○○	●●●○	○○○○	○○○○	○○○○	●●●●	7
Compute	○○○○	○○○○	●●●●	○○○○	○○○○	●●●●	●●●●	○○○○	●○○○	●●●●	17
Energy	○○○○	●○○○	●●●○	○○○○	○○○○	●●●●	●●●●	○○○○	○○○○	●●●●	16
Capabilities & limitations	●●●●	●●●○	●●●●	●○○○	●●●●	●●●○	●●○○	●●○○	●○○○	●●●○	27
Risks & mitigations	●●●○	●●○○	●○○○	●○○○	●●●○	●●○○	●○○○	●●○○	○○○○	●○○○	16
Evaluations	●●●●	●●○○	○○○○	○○○○	●●○○	●●●○	●●○○	○○○○	●○○○	●○○○	15
Testing	●●●○	●●○○	○○○○	○○○○	●●○○	●●○○	○○○○	●○○○	○○○○	○○○○	10
Machine-generated content	●●●○	●●●○	○○○○	●●●○	●●●○	●●●○	○○○○	●●●○	●○○○	●●○○	21
Member states	●●○○	○○○○	○○○○	●●○○	●●●●	○○○○	○○○○	○○○○	●○○○	○○○○	9
Downstream documentation	●●●○	●●●●	●●●●	○○○○	●●●●	●●●●	●●○○	○○○○	○○○○	●●●○	24
Totals	25 / 48	23 / 48	22 / 48	7 / 48	27 / 48	36 / 48	21 / 48	8 / 48	5 / 48	29 / 48	

Source: https://crfm.stanford.edu/2023/06/15/eu-ai-act.html

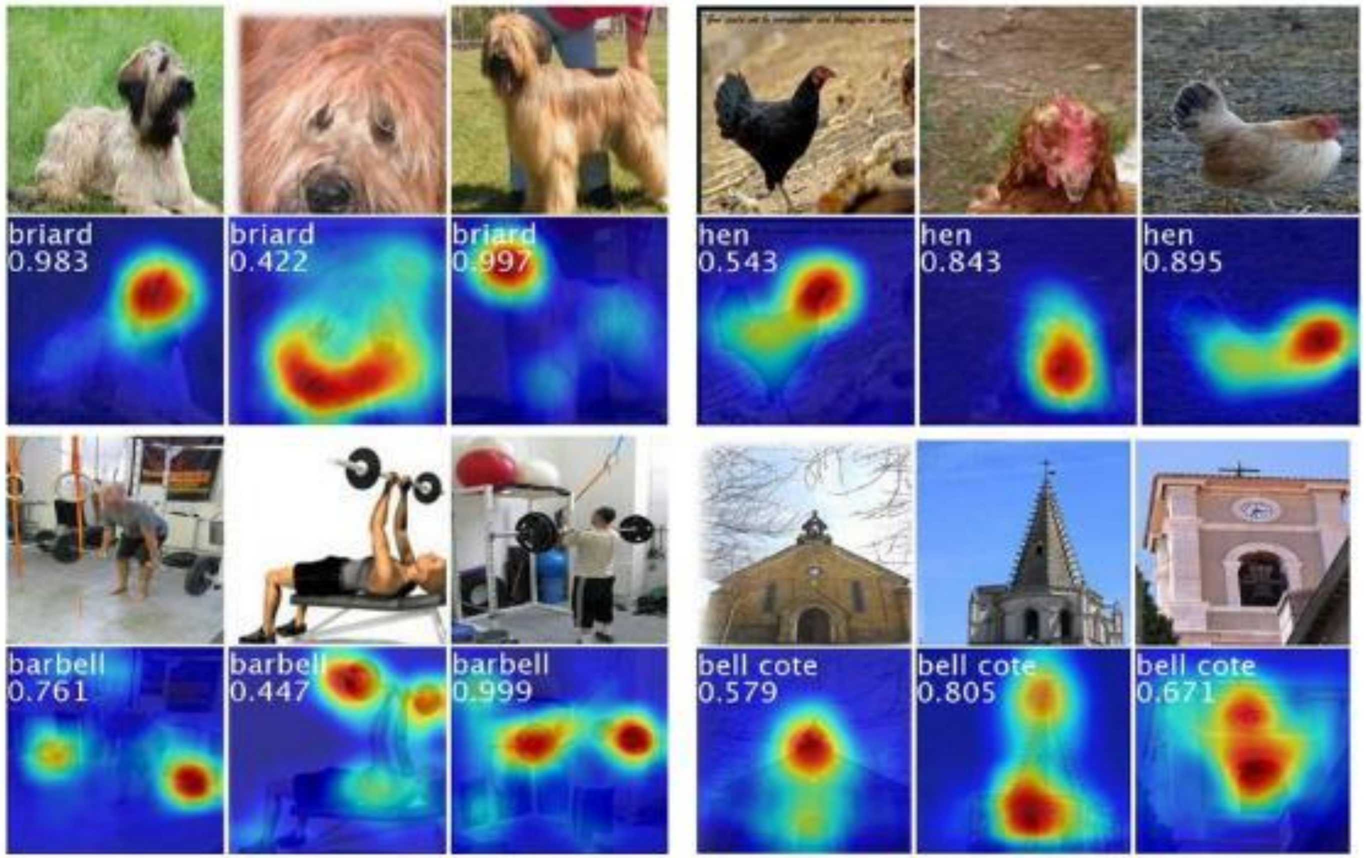
How to Explain ML models?

Intrinsically Explainable
Some models are explainable / interpretable on their own.

Post-hoc Explainability
If an ML model is not transparent, additional analysis must be done *after training the model* in order to provide an explanation.

Some examples of **post-hoc explainability** methods:

- Visual explanation
- Saliency maps (images)
- Model simplification
- Uncertainty Quantification
- **Look at the Features!**
  - Feature Importance
  - Feature Relevance
  - Feature Attribution
  - Feature Significance



https://debuggercafe.com/saliency-maps-in-convolutional-neural-networks/

How to Explain ML models?

ML explainers can be categorized into:

Model-specific Explainers vs. **Model-agnostic Explainers**

The explainability method is only applicable to a certain ML model only.

The explainability method is applicable to any ML model.

Local Explainers vs. **Global Explainers**

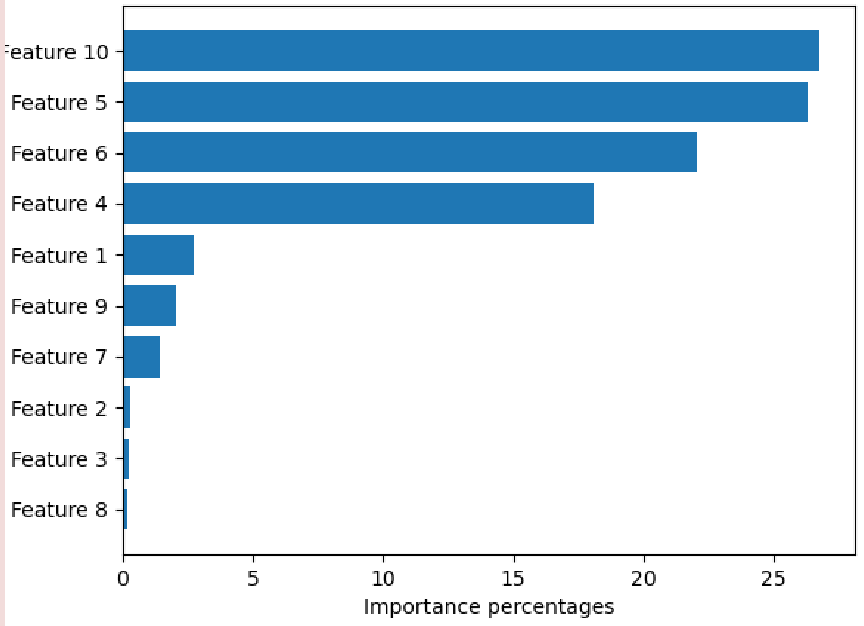
An explanation is provided for a specific data sample only.

An explanation is provided for the model behavior across the entire data space.

Feature Importance

- A mechanism to identify features that have the most *relevant impact* to the model predictions.
- Typically model-agnostic; can be local or global
- Packages: LIME, SHAP, DeepLIFT, etc.

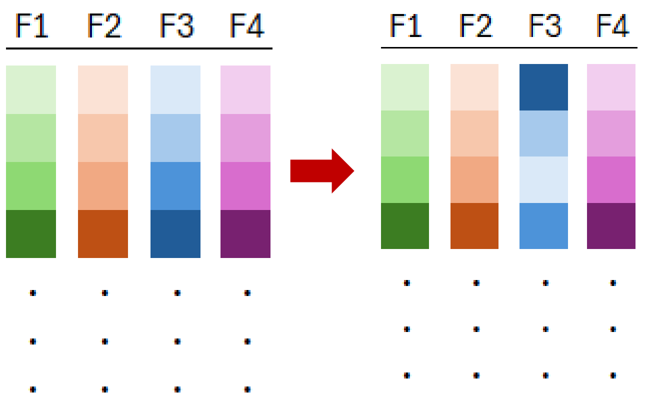
Sample Result



How to Explain ML models?

Permutation Feature Importance (PFI)

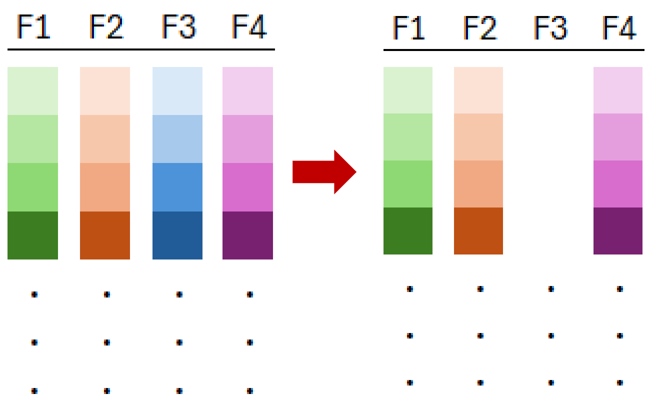
- PFI is defined as the decrease in the model score when a single feature value is randomly shuffled.
- If 2 or more features are correlated, PFI is biased to give them lower importance.



F3 is most important if it gave the largest drop in accuracy.

Drop-Column Feature Importance

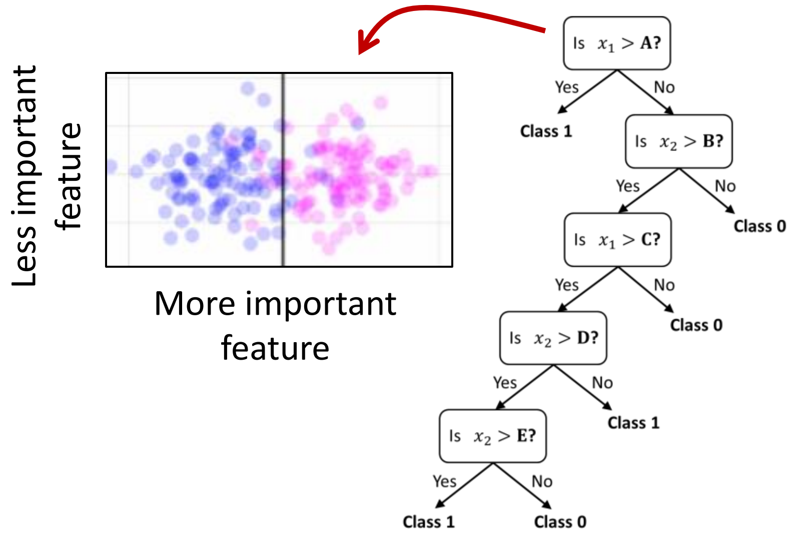
- Defined as the decrease in the model score when a single feature is removed from the data set.
- Requires model to be re-trained. Hence, it is not purely a post-hoc explainer.



F3 is most important if it gave the largest drop in accuracy.

Mean-Decrease-in-Impurity Feature Importance

- Applicable to tree-based models only (e.g. Random Forest, Decision Tree)
- "Decrease in impurity" quantifies the fraction of the samples a feature contributes to.



How to Explain ML models?

Shapley Additive Explanations (SHAP)

- Uses "Shapley values" from coalitional game theory.

- Feature importance, I_j, is computed as:

$$I_j = \sum_{i=1}^{n} \left|\phi_j^{(i)}\right|$$

where ϕ is the Shapley value (contribution of the jth feature at the ith sample).

- Calculation of the Shapley value involves iterating over all possible subsets of the feature set, then checking the changes in the model score.

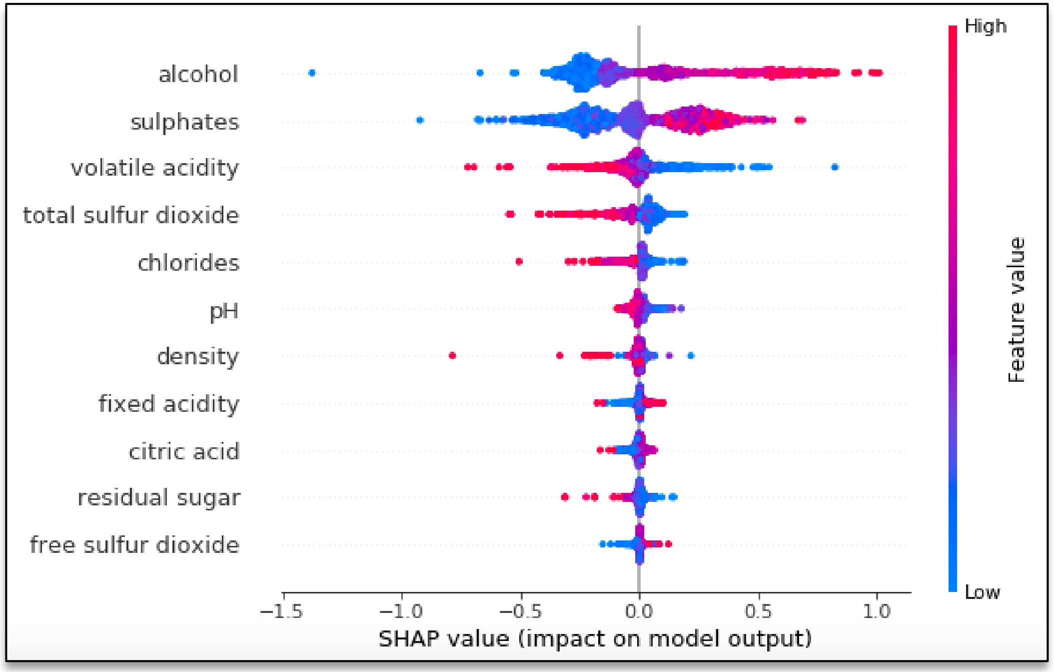
- Computation time grows exponentially with the number of features. Hence, we can approximate the Shapley value using only local samples →Kernel SHAP.

Shapley values are calculated as:

$$\phi_j = \sum_{S \in j} \frac{|S|!\ (|F| - |S| - 1)!}{|F|!} \left[f_{S\cup j}(\boldsymbol{x}_{S\cup j}) - f_S(\boldsymbol{x}_S)\right]$$

F = set of all input features
S = coalition which is a subset of F
$|\cdot|$ = cardinality of the set

$f_{S\cup j}(\boldsymbol{x}_{S\cup j}) - f_S(\boldsymbol{x}_S)$ = marginal contribution of feature j in coalition S.

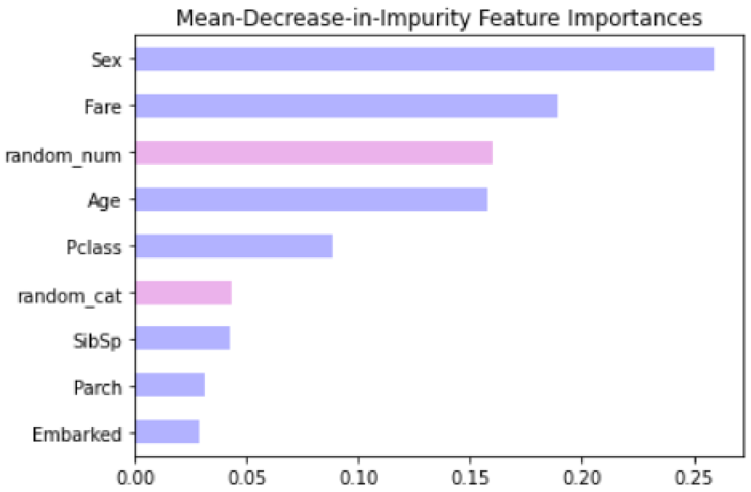


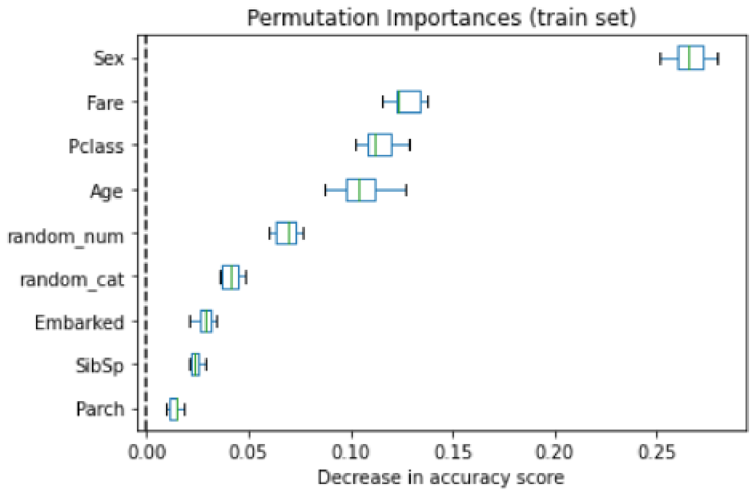
How to Explain ML models?

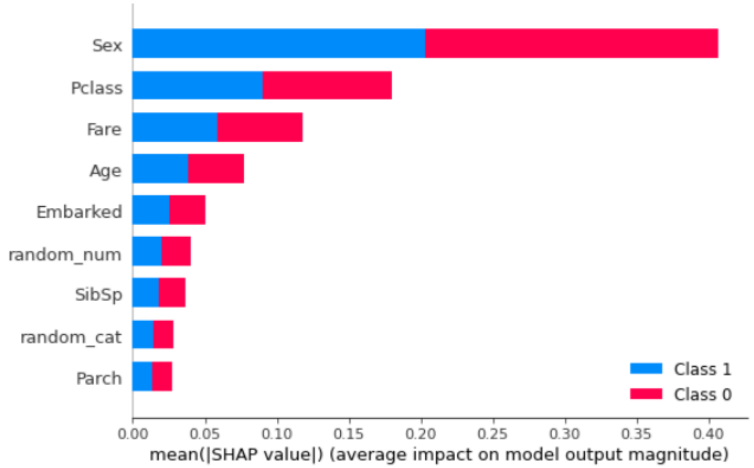
Example:

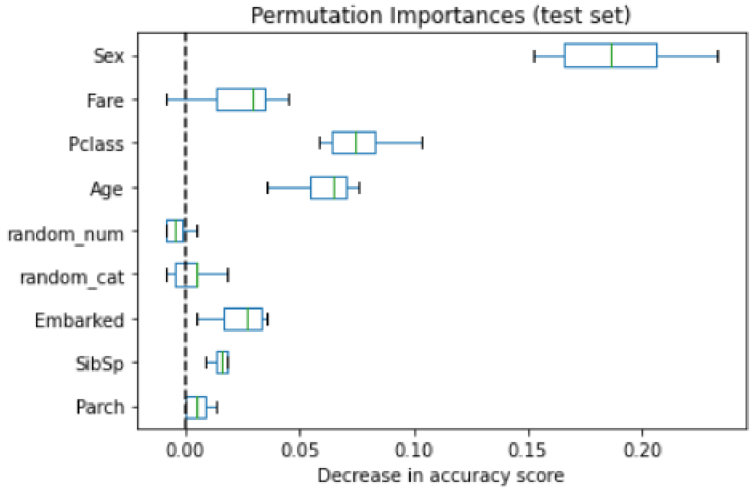
Apply feature importance techniques on a Random Forest classifier trained on the Titanic data set.

Type of feature	Feature	Feature values
	Survived	If survived or no (0 = No, 1 = Yes) (Target variable)
Numeric variables	PassengerId	Unique ID of each passsenger (in integers)
	Age	Age in years
	SibSp	Number of siblings / spouses aboard the Titanic
	Parch	Number of parents / children aboard the Titanic
	Fare	Passenger fare
Strings:	Name	Name of passenger
	Cabin	Cabin number
	Ticket	Ticket number
Categorical variables:	Pclass	Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
	Sex	Sex (string : 'male' 'female')
	Embarked	Port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)





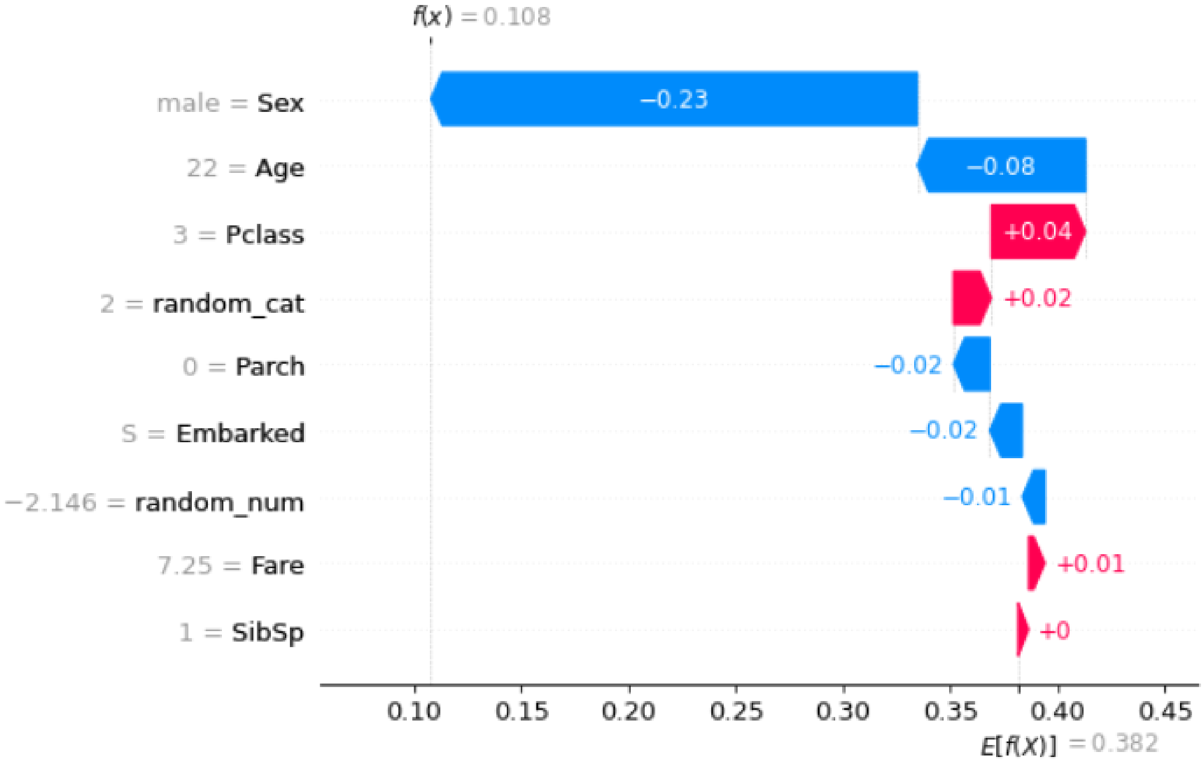




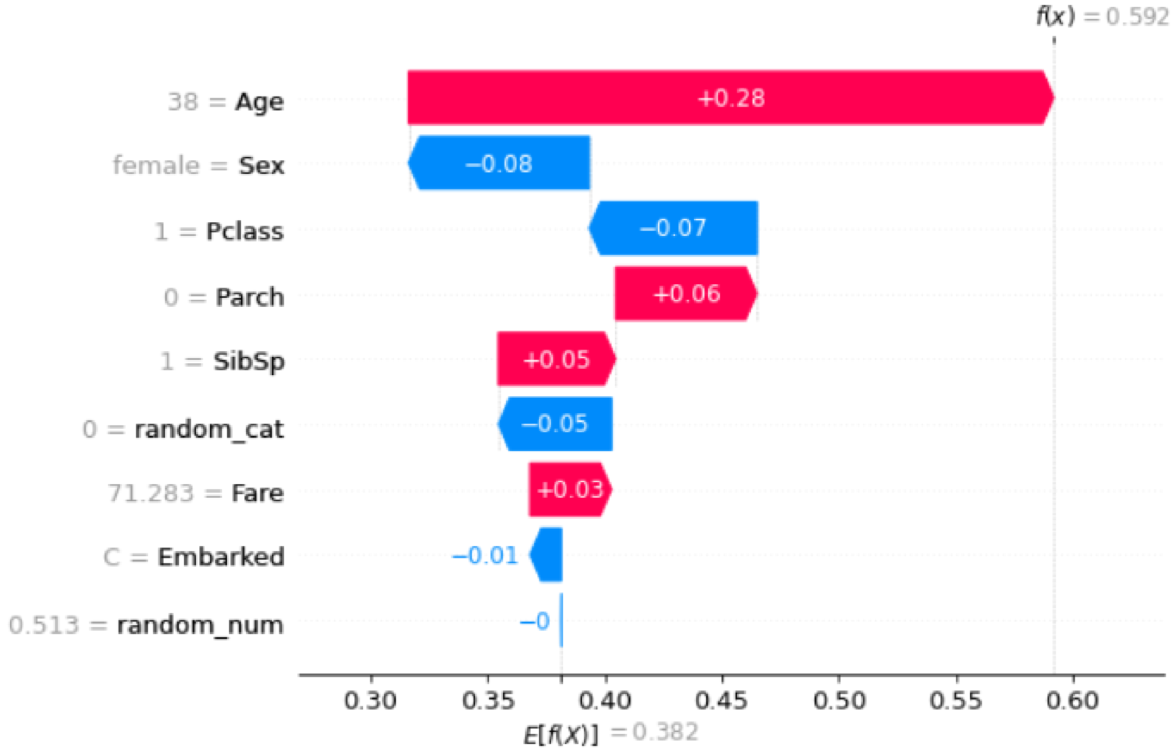
How to Explain ML models?

Example: Titanic Data Set – Local Explanations

Passenger 288
Random Forest Prediction: Did not survive



Passenger 869
Random Forest Prediction: Survived

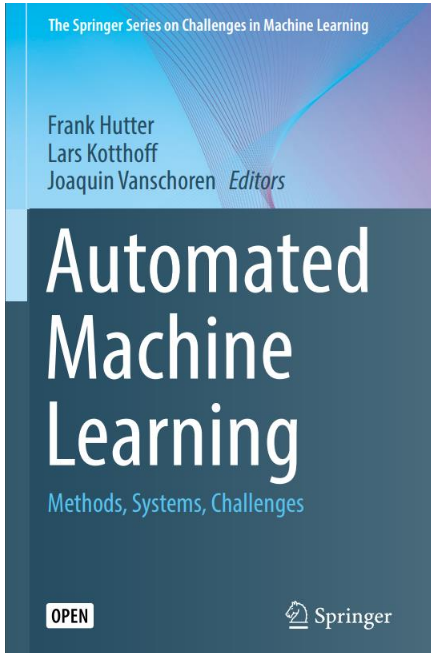


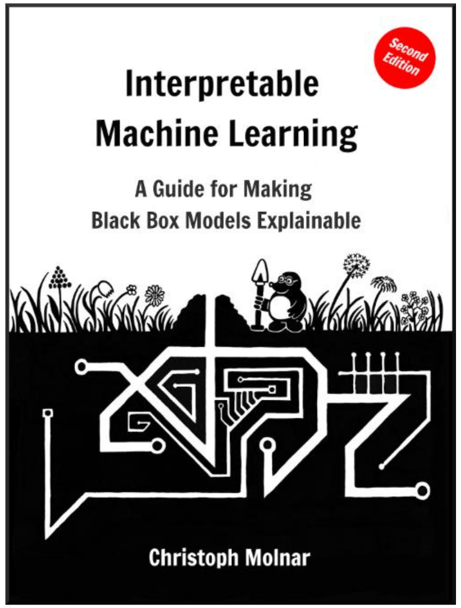
Outline

- AutoML Packages
  - Lazy Predict
  - Auto-sklearn
  - Optuna
  - TPOT
  - PyCaret

- Explainable AI (XAI)
  - Definitions and Concepts
  - Permutation Feature Importance
  - Drop-column Feature Importance
  - Mean-Decrease-in-Impurity
  - Shapley Additive Explanations

Hutter, Kotthoff,
Vanschoren (2019)





Molnar (2022)

Further Reading

- https://www.automl.org/automl/
- https://www.automl.org/wp-content/uploads/2019/05/AutoML_Book.pdf
- https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html
- https://machinelearningmastery.com/auto-sklearn-for-automated-machine-learning-in-python
- Feurer et al. (2015). Efficient and Robust Automated Machine Learning. Advances in Neural Information Processing Systems 28 (NIPS 2015).
- Feurer et al. (2022). Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. https://arxiv.org/abs/2007.04074
- Olson and Moore (2016). TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning. http://proceedings.mlr.press/v64/olson_tpot_2016.pdf
- Lundberg and Lee (2017). A Unified Approach to Interpreting Model Predictions. https://arxiv.org/abs/1705.07874
- Ribeiro et al. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. https://arxiv.org/abs/1602.04938
- Jang, **Pilario**, Lee, Na. (2023). Explainable Artificial Intelligence for Fault Diagnosis of Industrial Processes. IEEE Trans. On Industrial Informatics. doi: 10.1109/TII.2023.3240601
- Khan, Pao, **Pilario**, Sallih, Rehan (2023). Two-phase flow regime identification using multi-method feature extraction and explainable kernel Fisher discriminant analysis. International Journal of Numerical Methods for Heat & Fluid Flow. Emerald Publishing, Ltd. doi: 10.1108/HFF-09-2023-0526
- Clement, T.; Kemmerzell, N.; Abdelaal, M.; Amberg, M. XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process. Mach. Learn. Knowl. Extr. 2023, 5, 78-108. https://doi.org/10.3390/make5010006
- Arrieta et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, Vol. 58, June 2020, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012