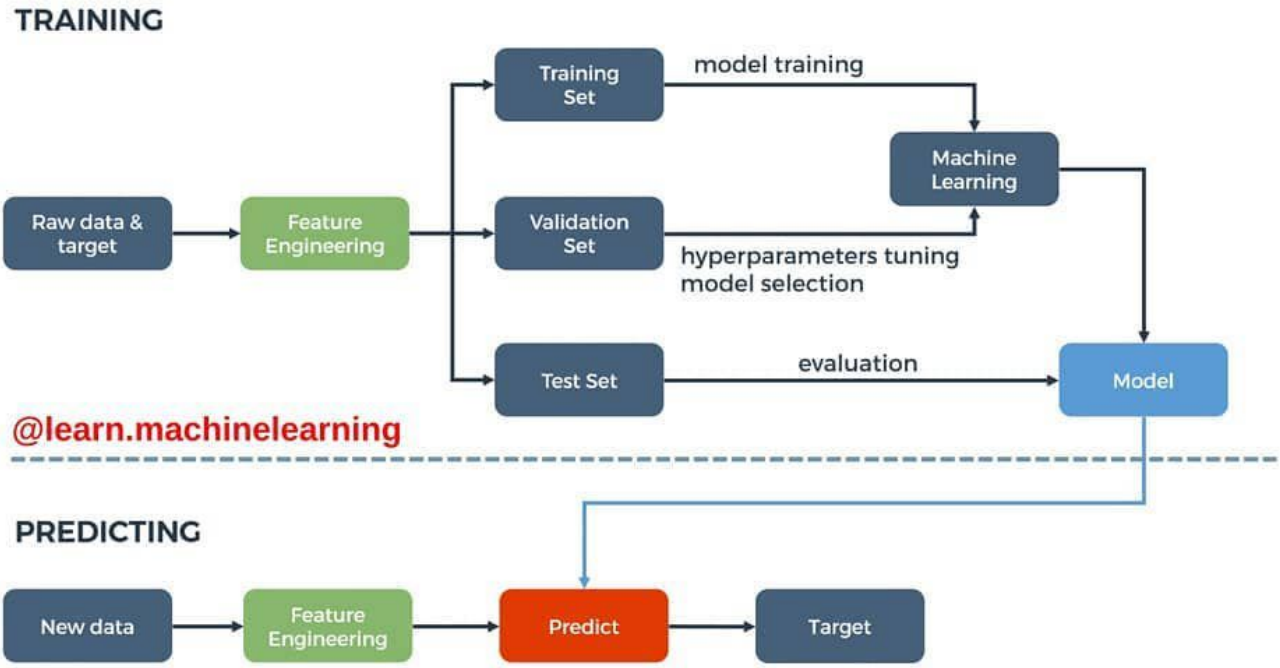
EVERYTHING ABOUT AUTOML

@learn.machinelearning

ABOUT AUTOML



- **“AutoML is not complete Data science automation, it only tries to automate the ML tuning part of data science ” that means AutoML trying to automate parts of the data science workflow.**
- Traditional machine learning model development is resource-intensive, requiring significant domain knowledge and time to produce and compare dozens of models. With automated machine learning, you'll accelerate the time it takes to get production-ready ML models with great ease and efficiency.

AutoML categories

- AutoML for automated parameter tuning (a relatively basic type)
- AutoML for non-deep learning, for example, AutoSKlearn. This type is mainly applied in data pre-processing, automated feature analysis, automated feature detection, automated feature selection, and automated model selection. @learn.machinelearning
- AutoML for deep learning/neural networks, including NAS and ENAS as well as Auto-Keras for frameworks.

WHY AutoML??

- Improve efficiency by automatically running repetitive tasks. This allows data scientists to focus more on problems instead of models.
- Automated ML pipelines also help avoid potential errors caused by manual work.
- AutoML is a big step toward the democratization of machine learning and allows everyone to use ML features.

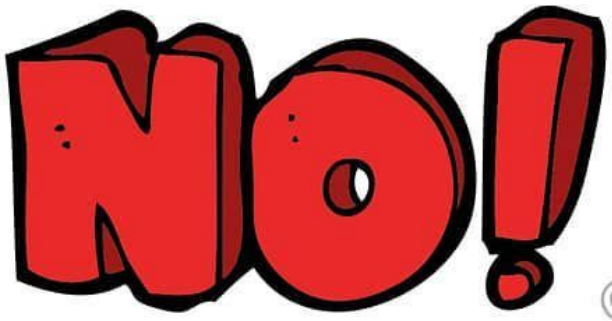
Hyperparameter tuning

AUTO ML SOLUTIONS

SOLUTIONS	OPEN SOURCE ?	# CONTRIBUTORS	# STARS	BACKEND FRAMEWORKS	OPTIMIZATION ALGORITHM
Amazon AWS Machine learning	No	-	-	TensorFlow, PyTorch...	-
Autosklearn	Yes	25	2,218	Sklearn	bayesian
AutoWeka	Yes	4	124	Java WEKA ML library	bayesian grid search
DataRobot	No	-	-	-	-
Google Cloud HyperTune	No	-	-	Tensorflow	bayesian
H2O	Yes	103	3,075	H2O	grid search
H2O Driverless	No	-	-	H2O	-
Hyperopt	Yes	25	2,078	-	random tree parzen estimator
IBM Watson	No	-	-	-	-
MLBox	Yes	3	482	Sklearn, Keras, XGBoost...	tree parzen estimator
PyBrain	Yes	32	2,544	homemade + libraries (LIBSVM)	metaheuristics grid search
TPOT	Yes	32	4,002	sklearn	genetic

source - hackernoon.com

Do you need to worry??



- Building models out of data is only one of the tasks that a data scientist does. Most of the work done by data scientists involves collecting, cleaning, formatting and analyzing data. As of the time of this answer, AutoML can’t do any of these tasks.
- AutoML can help you find the best possible model (out of the ones we’ve discovered so far) for a given problem but it can’t come up with a new solution altogether. You can view AutoML as just a slightly smarter model selection algorithm. If the model for the given problem doesn’t exist yet AutoML won’t be able to find it. Humans on the other hand might.

Do you need to worry??

- Coming back to the analytics world, there are plenty of activities that lie at the heart of data science where human influence, intervention, and oversight are vital. Here is a tentative list of some of the crucial but often overlooked activities that a data scientist may need to perform in his/her role and that are less prone to automation
  - **Identifying problems in the real-world that can be analysed through the lens of data science**
  - **Framing the problem as a data science problem (e.g., shall the problem be addressed as a supervised, unsupervised or reinforcement learning task? Or do traditional statistics suffice?)**
  - **Anticipating risks and devising strategies to manage them**
  - **Designing the data collection methodology, performing data annotation, assessing data quality, if no labelled data are available**

Do you need to worry??

- Identifying and controlling human biases, especially if relying on external data
- Incorporating domain knowledge into the process, for instance, via feature engineering
- Critically analysing and evaluating the results of a model
- Explaining model decisions in a human-interpretable way
- Analysing ethical issues and assessing the impact of the project output in society
- Effectively communicating the results to stakeholders

Missing pieces of AUTOML

- **Unsupervised Learning & Reinforcement Learning**
- **Complex Data Types**
- **Feature Engineering embedded with Domain Knowledge**

AutoML tools

- **AutoML libraries**
  - hyperopt
  - scikit-optimization
  - optuna
  - SMAC3
  - pycma
  - GPyOpt
  - botorch
  - spearmint
  - nevergrad
  - MOE
  - dragonfly
  - dask_ml
  - aws
  - google cloud
  - azure

- **AutoML systems**
  - AutoWEKA
  - Auto-sklearn
  - TPOT
  - H2O AutoML
  - TransmogrifAI
  - MLBoX
  - ROBO
  - AUTOFOLIO

- **Architecture Search**
  - Auto-PyTorch
  - AutoKeras
  - DEvol
  - HyperAS
  - talos
  - FLEXFOLIO

Thank You.

Please support us by

Like, Comment, Share and Save it for Later

Happy Machine Learning

@learn.machinelearning