



AUTOAI: THE SECRET SAUCE

JORGE CASTAÑÓN, PH.D. | LEAD DATA SCIENTIST @ IBM



@DeveloperWeek @castanan #DEVWEEK2020

NOW: CLIENTS / PRODUCTS @ IBM DATA AND AI



A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or a neural network diagram. The lines are vertical and horizontal, with some diagonal connections, and the circles are placed at various points along these lines.

AUTOAI

AutoAI is a new tool within IBM's Watson Studio that utilizes sophisticated training features to automate many of the complicated and time consuming tasks of feature engineering and building machine learning models, without the need to be a pro at data science.

\$2,900,000,000,000

The business value generated by AI augmentation globally in 2021

6,200,000,000 hrs

Worker hours saved by AI augmentation globally by 2021

How can we
accelerate the
path to AI?

IBM Research

IBM Research + Watson Studio



AutoAI

AutoAI automatically
prepares data, applies
algorithms, and builds
model pipelines best suited
for your data and use case.

What used to take days or
weeks, only takes minutes.

What is behind **AutoAI**?

AutoAI

Upload data in
a CSV file

Prepare data

Select model

Generate ranked
model pipelines

Save and deploy
a model

Feature type detection
Missing values
imputation
Feature encoding and
scaling

Selection of the best
estimator for the data

Hyper-paramater
optimization (HPO)
Optimized feature
engineering

Model Selection

- Test estimators against small subsets of the data
- Increase subset size gradually
- Save time without losing quality
- Ranks estimators and selects the best

“Selecting Near-Optimal Learners via Incremental Data Allocation”, AAAI Conference in AI, 2016

Dataset	Application Area	Full Training		DAUB				
		Allocation	Time (s)	Iterations	Allocation	Time (s)	Speedup	Loss
Buzz	social media	1,578k	56,519	57	302k	5,872	10x	0.0%
Cover Type	forestry	1,578k	43,578	13	160k	3,848	11x	1.1%
HIGGS	signal processing	1,578k	49,905	56	372k	2,001	25x	0.0%
Million Songs	music	1,578k	115,911	53	333k	17,208	7x	0.6%
SUSY	high-energy physics	1,578k	26,438	31	214k	837	31x	0.9%
Vehicle Sensing	vehicle management	1,578k	68,139	50	296k	5,603	12x	0.0%

Results from paper¹ of applying Model Selection method (DAUB) vs full training.

Why Feature Engineering?

“Coming up with features is difficult, time-consuming, and requires expert knowledge. Applied machine learning is basically Feature Engineering.”

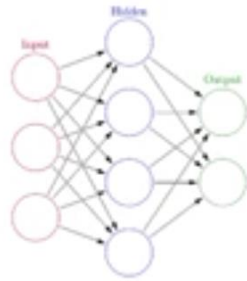
-Andrew Ng

“... some Machine Learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.”

-Pedro Domingos

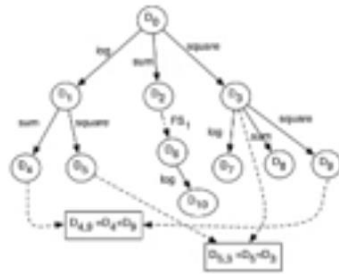
Auto Feature Engineering

(1) Pattern Learning



Learn patterns between classification data and transforms from 1000s of data offline.

(2) Exploration



Hierarchical exploration driven by performance numbers.

“Feature Engineering for Predictive Modeling Using Reinforcement Learning.”, AAAI Conference in AI, 2018

Hyperparameter Optimization

- Refines best performing pipeline
- Optimized for costly function evaluations
- Enables fast convergence to a good solution

“RBFOpt: an open-source library for black-box optimization with costly function evaluations.”, *Mathematical Programming Computation*, 2018.

Why AutoAI is different than other AutoML libraries?

1. **Flexible.** Outcome pipelines can be deployed as web services or exported as a Python scripts.
2. **Overfitting** is one of the most common problems observed in trained models. An overfitted model is not able to generalize to perform well on new unseen data. AutoAI applies robust methodologies based on cross-validation to avoid overfitting.
3. **Ease of use.** Choosing the data and the column to predict is the only actions needed by the user.
4. **Always improving.** Several teams are working in parallel to improve the performance of the AutoAI tool. Keep tuned for the new contributions!

AutoAI Questions?

References:

“Selecting Near-Optimal Learners via Incremental Data Allocation”, AAAI Conference in AI, 2016

“Feature Engineering for Predictive Modeling Using Reinforcement Learning.”, AAAI Conference in AI, 2018

“RBFOpt: an open-source library for black-box optimization with costly function evaluations.”, *Mathematical Programming Computation*, 2018.



Schedule Your Visit Here:
<https://ibm.co/talktothemlhub>