

MMF1922 Presentation

Auto Machine Learning (AutoML)

Group #3: Shuqi (Serena) Chang; Samantha MacPherson; Xinyi (Cynthia) Shen; Jinzhu (Gloria) Yang



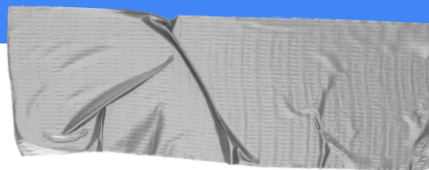


Table of Content

- Introduction of AutoML
- Why AutoML is Important
- Popular AutoML Tools
- Comparisons between Different Tools
- Conclusions

slido

Please download and install the
Slido app on all computers you
use



What is AutoML?

① Start presenting to display the poll results on this slide.

Introduction

What is AutoML?

- **Definition:** Automated machine learning (AutoML) is the process of automating the tasks of applying machine learning to real-world problems [Wikipedia]. *It is the combination of automation and ML.*
- **Differences between AutoML and Traditional Machine Learning:**
 - ☐ Automation vs. Manual Effort
 - ☐ Standardization vs. Flexibility in Deployment
 - ☐ Expertise and Cost (Low vs. High)

The AutoML Process

1. Data Preparation:

- ☐ Detect data types
- ☐ Suggest suitable transformations

2. Feature Engineering:

- ☐ Algorithms, e.g., PCA
- ☐ Prior knowledge

3. Model Selection:

- ☐ Perform ensemble learning
- ☐ Test multiple algorithms

4. Hyperparameter Tuning:

- ☐ Grid search
- ☐ Bayesian optimization
- ☐ Genetic algorithms

5. Deployment:

- ☐ Built-in deployment tools, e.g., API
- ☐ Monitor performance over time

Why AutoML is Important

➤ Democratization of Machine Learning and Reducing Skill Barriers

- ❑ Making Machine Learning Easier
- ❑ Broader Access for Businesses

➤ Driving Innovation

- ❑ Encouraging New Ideas
- ❑ Leveraging Expertise

➤ Efficiency

- ❑ Streamlining Processes
- ❑ Saving Time

Run AutoML

Project Name:

Training Frame:

Seed:

Max models to build:

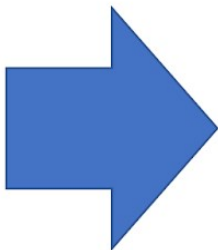
Max Run Time (sec):

Early stopping metric:

Early stopping rounds:

Stopping Tolerance:

nfolds:



	model_id	auc	logloss	mean_per_class_error	rmse	mse
	GBM_grid_1_AutoML_20190506_000950_model_8	0.981646	0.00279929	0.105491	0.0217763	0.000474206
	GBM_grid_1_AutoML_20190506_000950_model_13	0.97855	0.00278203	0.106747	0.0212029	0.000449564
	GLM_grid_1_AutoML_20190506_000950_model_1	0.977505	0.00403073	0.111985	0.0260326	0.000677699
	GBM_grid_1_AutoML_20190506_000950_model_10	0.97727	0.0027134	0.0977369	0.0202304	0.000409271
	GBM_grid_1_AutoML_20190506_000950_model_16	0.976829	0.00273863	0.0977566	0.0208157	0.000433294
	GBM_grid_1_AutoML_20190506_000950_model_5	0.969094	0.00313117	0.100327	0.0217719	0.000474017
	GBM_grid_1_AutoML_20190506_000950_model_17	0.963728	0.00304207	0.101602	0.0220067	0.000484294
	DeepLearning_grid_1_AutoML_20190506_000950_model_6	0.95929	0.00727195	0.105513	0.0240526	0.000578527
	DeepLearning_grid_1_AutoML_20190506_000950_model_4	0.954595	0.00835674	0.114497	0.0240522	0.00057851
	DeepLearning_grid_1_AutoML_20190506_000950_model_3	0.952148	0.0103404	0.106811	0.0246377	0.000607014

Popular AutoML Tools



Google AutoML

➤ Strengths:

- ☐ User-Friendly
- ☐ Cloud-Based
- ☐ Integration with Google Services
- ☐ Pre-trained Models

➤ Weaknesses:

- ☐ Limited Customizations
- ☐ Cost
- ☐ Primarily Cloud-Dependent



H2O.ai AutoML

➤ Strengths:

- ☐ Open-Source
- ☐ Wide Range of Algorithms
- ☐ Scalability
- ☐ Flexibility

➤ Weaknesses:

- ☐ Complexity
- ☐ Less Polished UI
- ☐ No Built-In Deployment Support



Microsoft Azure AutoML

➤ Strengths:

- ☐ End-to-End Automation
- ☐ Great for Time Series
- ☐ Enterprise-Ready
- ☐ Multiple Language Support

➤ Weaknesses:

- ☐ Cost
- ☐ Learning Curve
- ☐ Limited Open-Source Integration

Popular AutoML Tools



auto-sklearn

➤ Strengths:

- ☐ Open-Source
- ☐ Meta-Learning
- ☐ Great for Small Datasets
- ☐ Ease of Use

➤ Weaknesses:

- ☐ Scalability
- ☐ Limited to Scikit-learn Models
- ☐ Manual Deployment



TPOT

Tree-based Pipeline Optimization Tool

➤ Strengths:

- ☐ Genetic Programming
- ☐ Pipeline Focus
- ☐ Open-Source

➤ Weaknesses:

- ☐ Slow for Large Datasets
- ☐ Limited Algorithm Choices
- ☐ No Native Deployment Support



DataRobot

➤ Strengths:

- ☐ End-to-End Automation
- ☐ Great for Time Series
- ☐ Enterprise-Ready
- ☐ Multiple Language Support

➤ Weaknesses:

- ☐ Cost
- ☐ Learning Curve
- ☐ Limited Open-Source Integration

Comparisons between Different Tools

- Open-source libraries like auto-sklearn and TPOT are great for smaller or academic projects
- Cloud-based solutions like Google AutoML and Azure AutoML are preferred for enterprise-scale applications due to their robust infrastructure.

AutoML Tool	Open Source	Algorithms Supported	Deployment Support	Ease of Use	Scalability	Cost
Google AutoML	×	Pre-trained & Custom Models	Easy, Cloud-Based	High	High	Expensive
H2O.ai AutoML	✓	GBM, DL, RF, XGBoost, etc.	Manual	Moderate	High	Free (OSS)
MS Azure AutoML	×	Wide Range (good for time series)	Easy, Cloud-Based	Moderate	High	Expensive
auto-sklearn	✓	Scikit-Learn Algorithms	Manual	Moderate	Limited	Free
TPOT	✓	Scikit-Learn Algorithms	Manual	Moderate	Limited	Free
DataRobot	×	Wide Range	Easy, No-Code	High	High	Very Expensive



Which AutoML platform offers a no- code interface for beginners?

① Start presenting to display the poll results on this slide.



Which platform is ideal for experimentation using genetic programming?

① Start presenting to display the poll results on this slide.



When should businesses prefer H2O.ai over Google AutoML?

① Start presenting to display the poll results on this slide.

Conclusions



Key Takeaways

- **Democratization of ML:** AutoML bridges the gap between technical and non-technical users.
- **Platform Selection:** Choose based on your needs—user-friendliness vs. flexibility vs. enterprise scalability.
- **Future Outlook:** AutoML will continue evolving, making advanced ML more accessible to businesses and individuals.

- ❖ **Encourage participants to explore AutoML platforms for their projects.**
- ❖ **Highlight the importance of aligning AutoML solutions with business goals.**

THANK YOU