

Low-Code Machine Learning on Databricks with AutoML



Stephanie Rivera Solutions Architect, Databricks



Nicolas Pelaez Technical Marketing, Databricks

ORGANIZED BY 😂 databricks

Product Safe Harbor Statement

This information is provided to outline Databricks' general product direction and is for informational purposes only. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward–looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all.



A Personal ML Journey





Credit: https://www.youtube.com/watch?v=bbHXSc2wJh0

Does applying machine learning have to be a challenge?



Common Path to Applying ML

From Data to Inference





Databricks Machine Learning

Seamlessly integrated AutoML





Databricks Lakehouse is the foundation for **all your data analytics needs**



How can I focus less on code and more on results?



Spend time on results not code

Leverage automation where it makes sense

Machine Tasks

- Tracking experiments
- Proper Syntax
- Boilerplate code
- Basic model testing
- Tuning hyperparameters

Human Tasks

- Exploring data sets for insights and building intuition
- Feature engineering
- Checking and making assumptions
- Interacting with and supervising automation
- Identifying value and interpreting results



Low-code makes life easy

AutoML, MLflow, and Bamboolib to the rescue



What is AutoML?

Automated machine learning (AutoML) is a fully-automated model development solution seeking to "democratize" machine learning. While the scope of the automation varies, AutoML technologies usually automate the ML process from data to model selection.





Integrated MLflow

Reproducible research made easier





UI-based data analysis and transformation

Integrating bamboolib into the Notebook



Prepare, transform, visualize, and explore your data—all through a UI!



Be more efficient by spending less time writing boilerplate code



Operations in bamboolib generate code so users can see, customize, and learn from what happens via the UI



DATA+AI

Enable citizen data scientists who know what they want to do to do it in Python

Titanic Voyagers Python								0	Schedule ~			
њ .	10.x Snapshot	1×	B• @• (-	⊜	8-			2	₫		
1	bam							Python	••	× -	×	
L	Load data Data											
1	5 C	History E	xport 🛛 Live Cod	e Expo	ort							
a	earch transforma	tions		or	Create plot	or Explore Data	Frame					
S	elect or drop colu	mns multiple columps										
F	ilter rows	татро солитте		U	Update							
s	elect/delete rows b	ased on a conditi	on	i	Pclass	o Name	o Sex	f Age		1		
S	Sort rows			3		Braund, Mr. Owe	male	22.0		1		
5	Sort rows beased on values in one or more columns Group by and aggregate (default) Group rows by columns and calculate MULTIPLE aggregations					Cumings, Mrs. J	female	38.0		1		
G						Heikkinen, Miss	female	26.0		0		
(no renaming possible)				1		Futrelle, Mrs. Jac	female	35.0		1		
4		5	0	3		Allen, Mr. William	male	35.0		0		
5		6	0	3		Moran, Mr. James	male	nan		0		
6		7	0	1		McCarthy, Mr. Ti	male	54.0		0		
7		8	0	3		Palsson, Master	male	2.0		3		
8		9	1	3		Johnson, Mrs. O	female	27.0		0		

Let's see it in action!



Recap



Configure AutoML experiment Preview Provide Feedback Experiments > Configure AutoML experiment	ire	"Glass-Box" AutoML with a III							
 Configure Augment Train AutoML Experiment Configuration Compute dais_mir_8_new Select an existing cluster with a Databricks Runtime for ML 8.0+ or greate a new clust ML problem type Classification Training data default.usage_logs Target column 	④ Evaluate ↓ AutoML ♥ Configure AutoML Evaluation ۞ complete. All runs have completed, and have been added to the table Model with best val_f1_score The model is ready to be registered and deployed. Or, acc Register and deploy model Edit model	& Evaluate	Custom	ize	VVILII COI © ? L ples for more thorough				
isMining • Experiment name isMining_usage_logs-2021_05_05-22_33 • Data directory ↓	Showing 16 matching runs	Vumerical columns One-hot encoding Feature standardiza Train classification mo Feature importance Inference Inference In	<pre># Use Kernel SHAP to explain predict = lambda x; model.nr. explainer = Kern shap.values = e; summary_plot(sh: except Exception as print(f"An unext 000% 25/25 _ip_address_onlineip_address_onlineip_address_on ① _ip_address_on ① </pre>	feature importance on exa edict(od_DataFrame(xcol Details Serving Status: • Ready - Stop Model Versions Model Events Model Versions Version 2 • Ready Production	Deploy Cluster: mlflow-model-dais_demo Cluster: mlflow-model-dais_demo Cluster Settings Model URL: https://dbc-60ef17e8-d99b.dev.databricks.com/model/dais_demo/2/invocations https://dbc-60ef17e8-d99b.dev.databricks.com/model/dais_demo/Production/invocations				
DATA+AI SUMMIT 2022	O 2021-05-05 1 random_f kase	Crid	ommand took 3.68 minutes → 3 25		Call the Model Browser Curl Python Request From the Model [

G

Databricks AutoML Supported Features



DATA+AI SUMMIT 2022

Thank you



Stephanie Rivera Solutions Architect, Databricks



Nicolas Pelaez Technical Marketing, Databricks "Databricks' AutoML greatly improved our time to market for our category personalisation model with ready-to-use code for quick iteration and we were able to outperform our previous model by 2-3% on the same dataset."





Allscripts saw massive improvement in their customer service operations when they put their AutoML model into production.





Fabletics was able to generate and tune models in 30 minutes (which previously had taken them days)





Customer Success with Databricks Machine Learning







"... improved accuracy of vehicle pricing, automated model updates and their frequency ..." "... increased revenue by personalizing user experience ..." "... improved developer productivity by enabling parallel training of models for different countries, types of articles, and time periods ..."



Who am I?



Stephanie Rivera

- Solutions Architect, Databricks
- **Previous Experiences:** VP Data Intelligence, Senior Architect Machine Learning, CoAuthor *Field Guide to Data Science*
- **Domain Focus:** Data Science & Machine Learning

Solution de la construction de la construction de la construcción de l



Who am I?



😂 databricks

Nicolas Pelaez

- **Technical Marketing**, Databricks
- Previous Experiences: Turbine Propulsion Engineer and Software Developer for Pratt & Whitney, Technical Marketing at Cloudera
- **Domain Focus:** Machine Learning, Data Science/Exploration, User Experiences

