Tool	Platform	Input data sources		Data pre-processing	Data types detected					Feature engineering				ML Tasks		Model selection and Hyperparameter optimization					Quick start / early stop			Model evaluation / Result analysis/ Visualization		
		Spreadsheet datasets	Image, text		Numerical	Categorical	Datetime	Time-series	Other (Hierarchical types) (7*)	Datetime, categorical processing	Imbalance, missing values	Feature selection, reduction	Advanced feature extraction (8*)	Supervised learning (9*)	Unsupervised learning (10*)	Ensemble	Genetic algorithm	Random search	Bayesian search	Neural architecture search	Quick finding of starting model	Allow maximum limit search time	Restrict time consuming combination of components	Model dashboard	Feature importance	Model explainability and interpretation, and reason code (11*)
TransmogrifAI	Apache Spark	Y	N	Y(*)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	N	N			Y	Y	
H2O-AutoML	H2O clusters	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y	N	Y	N	N	N	Y	Y	Y	Y	Y
Darwin (+)	GCP	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	N	Y	Y	Y
DataRobot (+)	Datarobot & AWS	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y		Y	Y	Y
Google AutoML (+)	Google Cloud	N	Y	Y						N	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Auto-sklearn		Y	N	N	N	N	N	N	N	Y(2*)	Y	Y	Y	Y	N	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y
MLjar (+)	MLJAR Cloud	Y(3*)	N	Y	Y	Y	N	N	N	Y	Y(4*)	N	N	Y(5*)	N	Y	N	Y	N	N	N	N	N	Y	Y	N
Auto_ml		Y	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	N	N	N	N	Y	Y	Y
TPOT		Y	N	N	N	N	N	N	N	N	Y	N	Y	Y	N	Y	Y	N	N	N	N	Y	N	Y	Y	N
Auto-keras		Y	Y	N	N	N	N	N	N	N	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	N	Y	N	Y
Ludwig		Y	Y	Y(*)	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	N	N	Y	Y	N
Auto-Weka		Y	N	N	Y	Y	N	N	N	N	Y	Y	N	Y	N	Y	N	Y	Y	N	N	Y	Y	Y	N	N
Azure ML (+)	Azure	Y	Y	Y(6*)	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y	Y	N		Y	Y	Y	Y	
Sagemaker (+)	AWS	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N		Y	N	Y	Y	Y
H2O-Driverless AI (+)	H2O clusters	Y(3*)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y

Fig. 2. Comparison table of functionality for AutoML tools. $(+)$: commercialized tools; $(*)$: the function is not very stable, it fails for some datasets; $(2*)$: categorical input must be converted into integers; $(3*)$: datasets have to include headers; $(4*)$: missing values must be represented as NA; $(5*)$: multiclass classification not provided; $(6*)$: need some users' input for dataset description such as column types; $(7*)$: ability to detect primitive data types and rich data types such as: text (id, url, phone), numerical (integer, real); $(8*)$: advanced feature processing: bucketing of values, removing features with zero variance or features with drift over time; $(9*)$: supervised learning includes binary classification, multiclass classification, regression; $(10*)$: unsupervised learning includes clustering and anomaly detection; $(11*)$: model interpretation and explainability refers to techniques such as LIME, Shapley, Decision Tree Surrogate, Partial Dependence, Individual Conditional Expectation, Lift chart, feature fit, prediction distribution plot, accuracy over time, hot spot and reason codes; In a few empty cells, it is not clear that the functionality is provided from documentations of the tools, to the best of our knowledge.