	Pipeline Phase	DQ Dimensions	Preparation Actions	Analysis Task	Source Type	Suggestions	Automatic	Methodology
[7]	Data Profiling	Consitency		Querying	Tabular			
[8]	Data Profiling DQ Assessment Data Cleaning	Completeness Uniqueness Duplication Accuracy	Data imputation Data filtering Data type conversion Anomaly detection	AI-based	Tabular Time series	Suggest data cleaning actions		Direct acyclic graphs composed by a set of validators, i.e., contains the couple DQ check, recommendation
[9]	DQ Assessment	Label Noise Class Overlap	Cleaning only the targeted class (novel AI-based algorithms)	Classification	Tabular	Suggest how to clean the targeted class		Ad-hoc algorithms (AI-based)
[10] [11]	Data Profiling Data Cleaning	Consistency Accuracy	Error detection and correction Data type conversion Data fusion		Tabular Semi-structured	Suggest data repairs (error correction based on integrity constrains or external sources)		Probabilistic inference methods (AI-based)
[13]	Data Cleaning	Completeness Accuracy Consistency Duplication	Data imputation Outlier detection Inconsistency detection Deduplication Normalization Feature selection	Classification Regression Clustering	Tabular		Best data cleaning pipeline	Reinforcement Learning (Q-learning) (AI-based)
[14]	Data Cleaning	Completeness Accuracy Consistency Duplication	Data imputation Outlier detection Inconsistency detection Deduplication Normalization Feature selection	Classification Regression Clustering	Tabular		Automatic best data cleaning pipeline (computed with the support of HITL)	Reinforcement Learning (Q-learning) (AI-based)
[15]	Data Profiling DQ Assessment	Accuracy	Error detection and correction		Tabular		Automatic data cleaning	Past knowledge and data profile similarity
[16]	Data Cleaning	Accuracy	Error detection and correction		Tabular	Suggest data repairs		Dataset semantic and knowledge base (containing semantic of previous dataset)
[17]	Data Cleaning		Join, pivot, unpivot, aggregation Relationalization of semi-structured		Tabular	Suggest operators		Predictive model trained with 4M Jupyter notebooks
[18]	Data Cleaning	Accuracy Consistency	Outlier detection and correction Pattern and rule violation detection and correction		Tabular	Suggest data correction		Past knowledge and data profile similarity (repository of prev. cleaned datasets)
[19]	Data Cleaning	Completeness Accuracy	Data imputation Outlier detection and correction with imputation	Classification	Tabular		Automatic cleaning	Bayesian optimization (AI-based)
[22]	Data Exploration Data Cleaning	Duplication Completeness Accuracy	Duplicates detection Data imputation Outlier detection and correction		Tabular	Suggest data repairs		Interactive and iterative process with user feedback and composite questions
[23]	DQ Assessment Data Cleaning	Completeness Consistency (Missing and overlapping timestamps, Event overlapping) Orderliness (Time ordering violation)	Data imputation Inconsistencies detection and correction	Process mining analysis	Process logs		Interactive process without suggestions	Interactive approach: iterative DQ assessment and data cleaning
[24]	Data Cleaning	Accuracy Consistency	Outlier detection and correction Inconsistencies detection and correction	Convex loss models (regression, SVM)	Tabular Images	Suggest portion of data to clean		Stochastic Gradient Descent (AI-based)
[27]	DQ Assessment	Accuracy Consistency Completeness		Classification Regression Clustering	Tabular			
[29]	Data Profiling DQ Assessment Data Cleaning	Accuracy Completeness Consistency	Outlier detection and correction Data imputation Inconsistencies detection and correction	Classification Regression Clustering	Tabular	Suggest best data preparation pipeline		Past knowledge and data profile similarity
[30]	DQ Assessment Data Cleaning	Accuracy Completeness	Outlier and noise detection and correction Data imputation	Data streams analysis	Sensor data streams		Automatic data cleaning	Predefined sequence of actions
[31]	DQ Assessment	Accuracy Completeness Consistency Timeliness Duplication Orderliness			Data streams			
[32]	DQ Assessment	Accuracy Completeness Consistency Credibility Currentness Accessibility Compliance Confidentiality Efficiency Precision Traceability Understandability Availability Portability Recoverability			IoT data			