META LEARNING

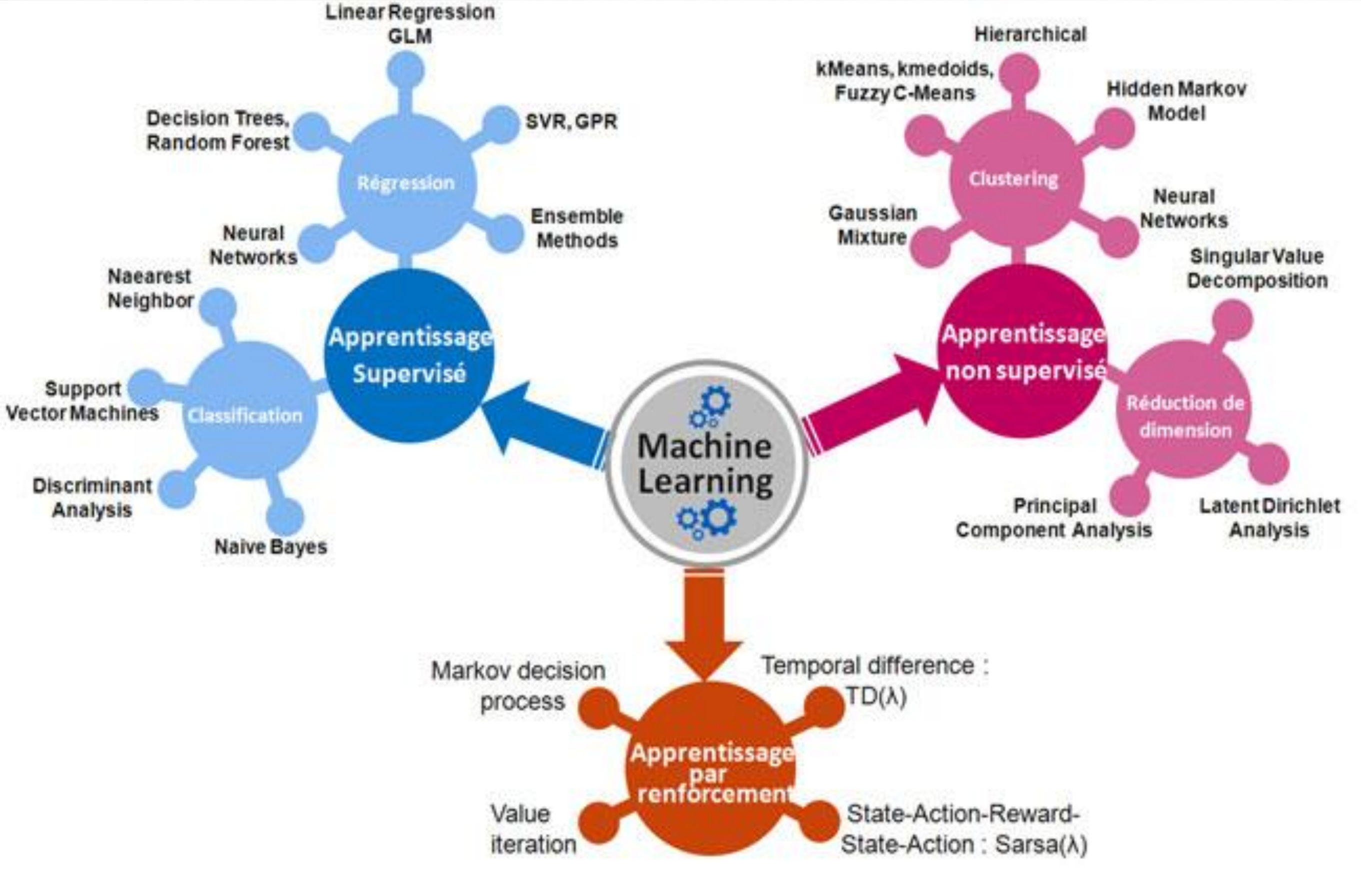
WEI-PENG CAO

2018.6.25

CONTENT

- Background
- Concept of Meta Learning
- A typical case of Meta Learning
- Considerations
- An example: Clustering algorithm selection

META LEARNING - BACKGROUND



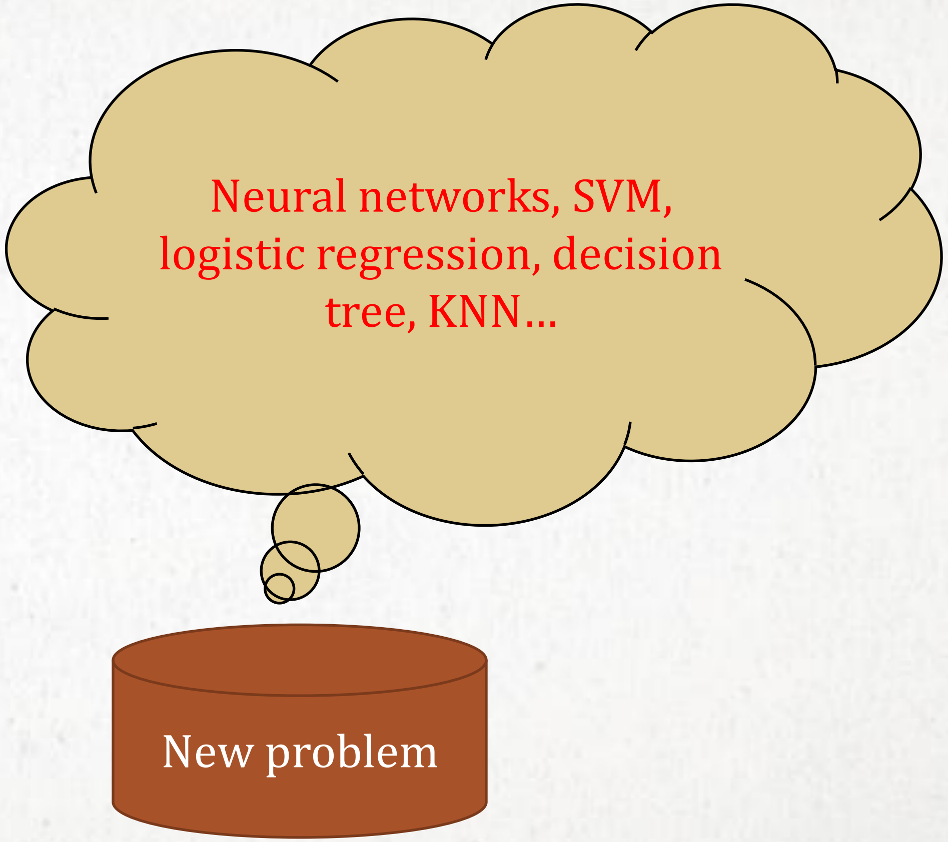
Why do we need so many algorithms?

META LEARNING - BACKGROUND

Different datasets have different inherent characteristics (e.g., data distribution) and each algorithm can only learn well if its bias matches the learning problem.

A learning algorithm may perform very well in one domain, but not on the next, which leads researchers to create a large number of algorithms.

META LEARNING - BACKGROUND



META LEARNING - BACKGROUND

trial and error

⬇

a high computational cost

⬇

This cost could be reduced if the most suitable algorithm(s) could be recommended.

Meta learning

META LEARNING - CONCEPT

- **The core issue of meta learning**

to study the relationship between the learning problem and the effectiveness of different learning algorithms.

- **Goal**

Algorithm Recommendation and Hyperparameter Recommendation

META LEARNING - CONCEPT

- **Meta data :**

The characterization of the datasets and the performance of the ML algorithms.

- **Meta dataset:**

Each sample corresponds to one of the original datasets;

The attributes of each sample are the meta-features of a dataset;

The label is the predictive performance of the candidate algorithms when applied to a dataset ;

META LEARNING - METHODS

- **Meta features** are able to describe the main aspects of a dataset and usually extracted by two approaches: Statistics and Model-based properties.

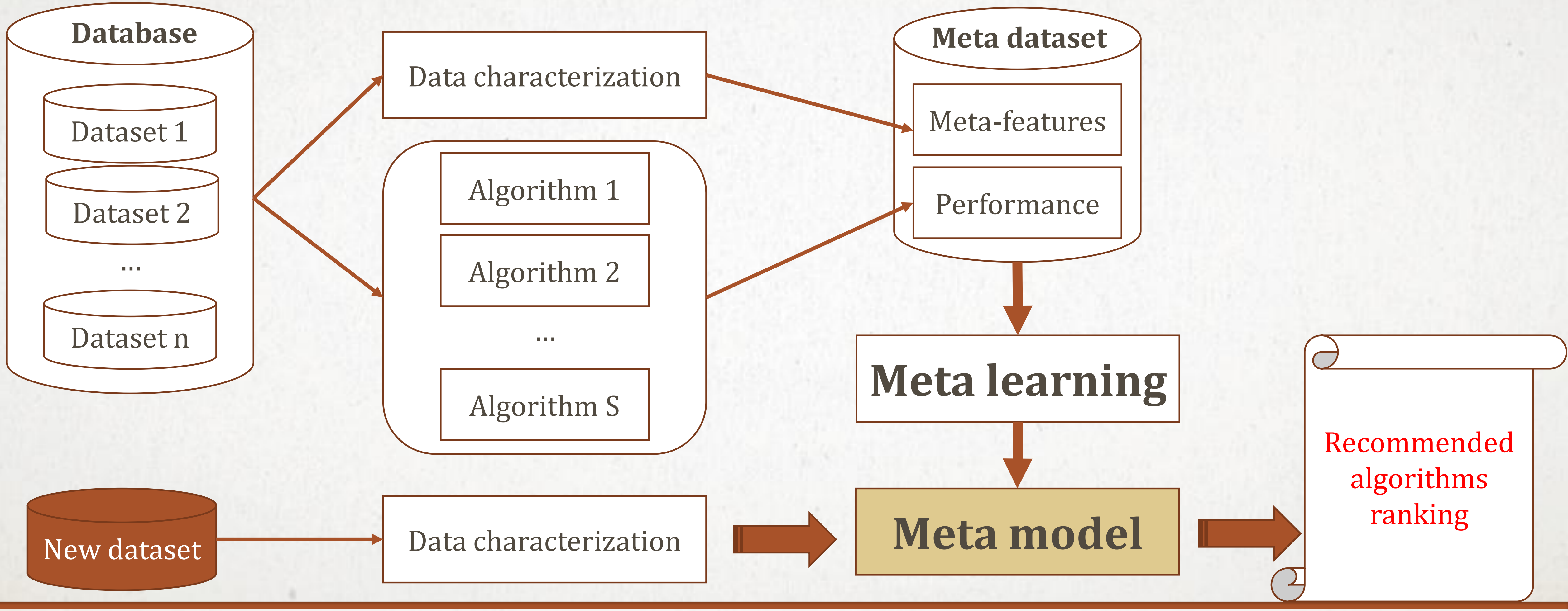
Statistics:

① A function of the dataset size, uses equation LgE = $\log_{10}(\boldsymbol{n})$, where $\boldsymbol{n}$ is the number of samples;
② The ration between the number of samples ($\boldsymbol{n}$) and the number of attributes ($\boldsymbol{p}$): LgREA = $\log_{10}(n/p)$;
③ The percentage of missing values;
④ The complexity of a problem;
⑤ …

Model-based properties: a set of properties of a model

① For example, if a decision tree algorithm is applied to the dataset, statistics about nodes, leaves and branches can be used to describe the dataset.
② For example, if a neural network algorithm is applied to the dataset, statistics about the number of hidden layer and the number of hidden nodes in each layer can be used to describe the dataset.

META LEARNING - METHODS



META LEARNING –CONSIDERATIONS

The construction of Meta dataset

is the key to

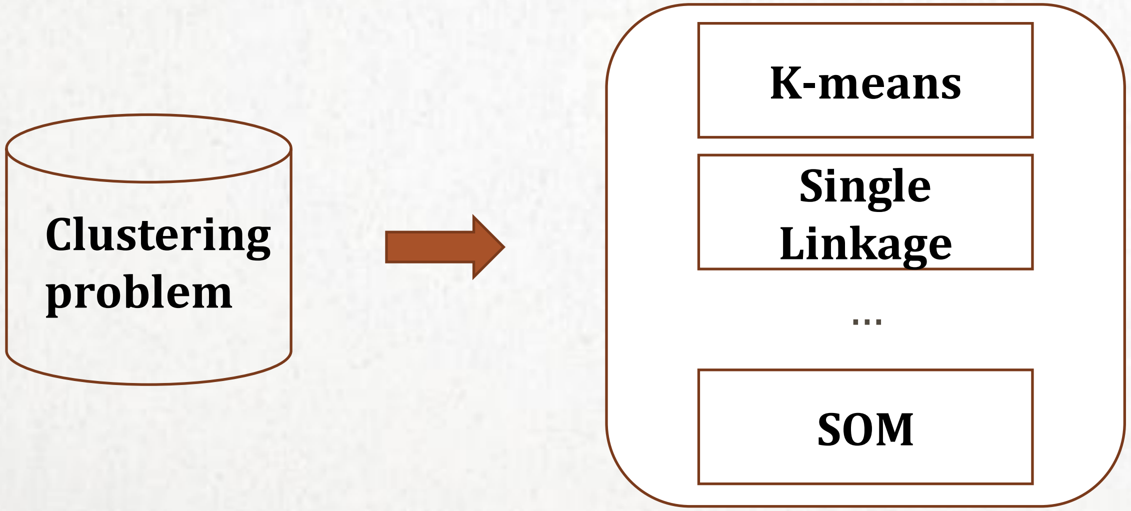
the successful use of meta-learning.

CONTENT

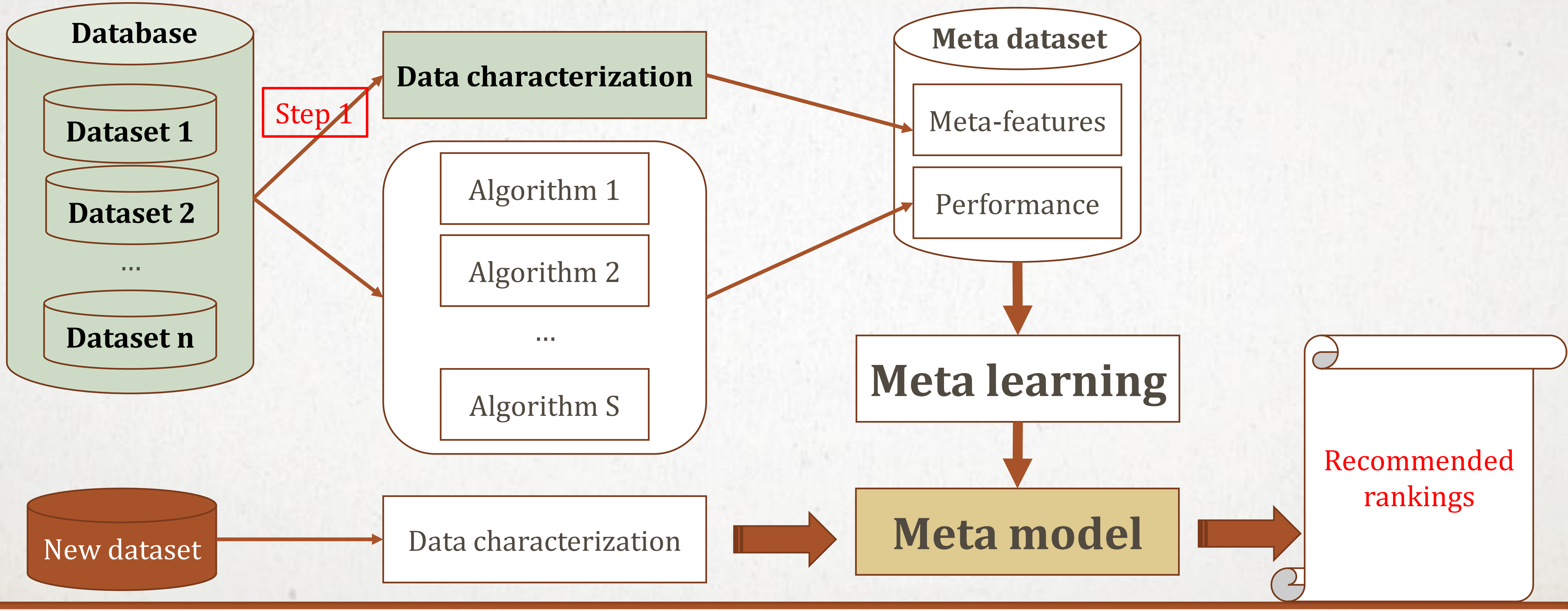
- Background
- Concept of Meta Learning
- A typical case of Meta Learning
- Considerations
- An example: Clustering algorithm selection

META LEARNING –EXAMPLE

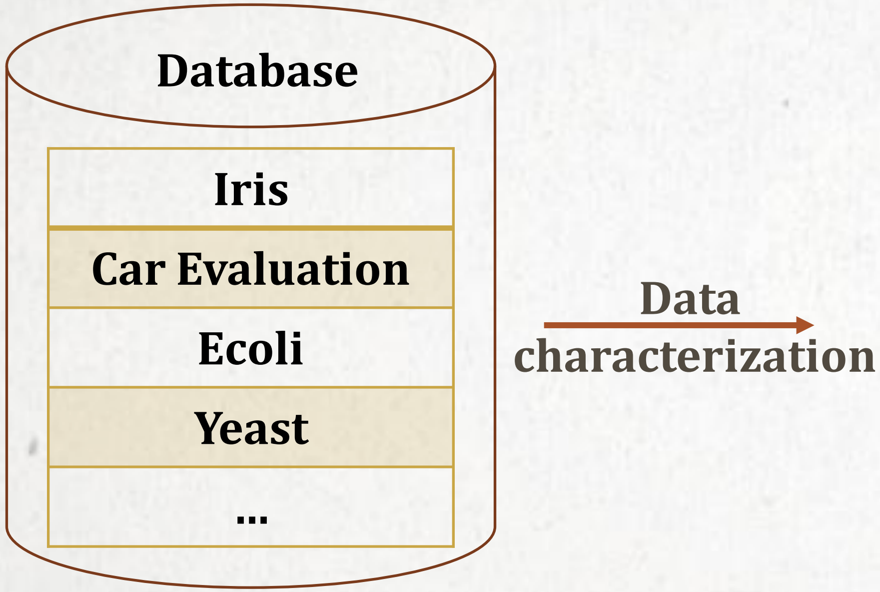
Motivation:



META LEARNING - METHODS



META LEARNING –EXAMPLE



Total: 100

Table 1. Meta-attribute values for some datasets

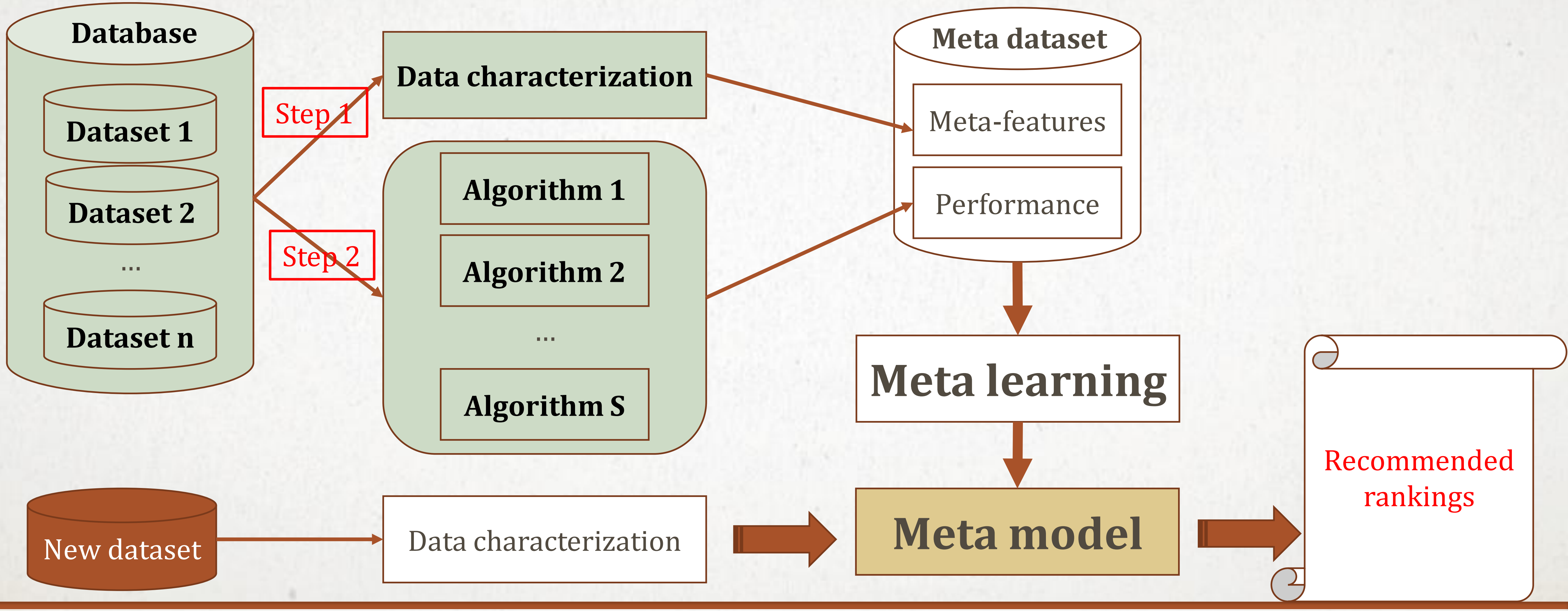
Dataset/Value	1	2	3	4	5	6	7	8	9	10
Iris	7.23	2.00	0.00	0.00	1.00	0.59	0.07	2.23	0.00	0.00
Car Eval.	10.75	2.58	0.00	1.00	0.00	0.00	0.00	0.00	0.00	5.38
Ecoli	8.39	2.81	0.00	0.00	1.00	0.18	3.59	54.19	0.00	0.00
Yeast	10.54	3.00	0.00	0.00	1.00	0.09	2.91	31.56	0.00	0.00
Wine	7.48	3.70	0.00	0.00	1.00	0.30	0.35	2.97	0.00	0.00
Tic-Tac-Toe	9.90	3.17	0.00	1.00	0.00	0.00	0.00	0.00	0.01	3.91

Notes:

(1) Log_2 of the number of samples;
(2) Log_2 of the number of attributes;
(3) Proportion of binary attributes;
(4) Proportion of discrete attributes;
(5) Proportion of continuous attributes;
(6) Mean absolute correlation between continuous attributes;
(7) Mean skewness of continuous attributes;
(8) Mean kurtosis of continuous attributes;
(9) Mean absolute concentration between discrete attributes;
(10) Mean entropy of discrete attributes.

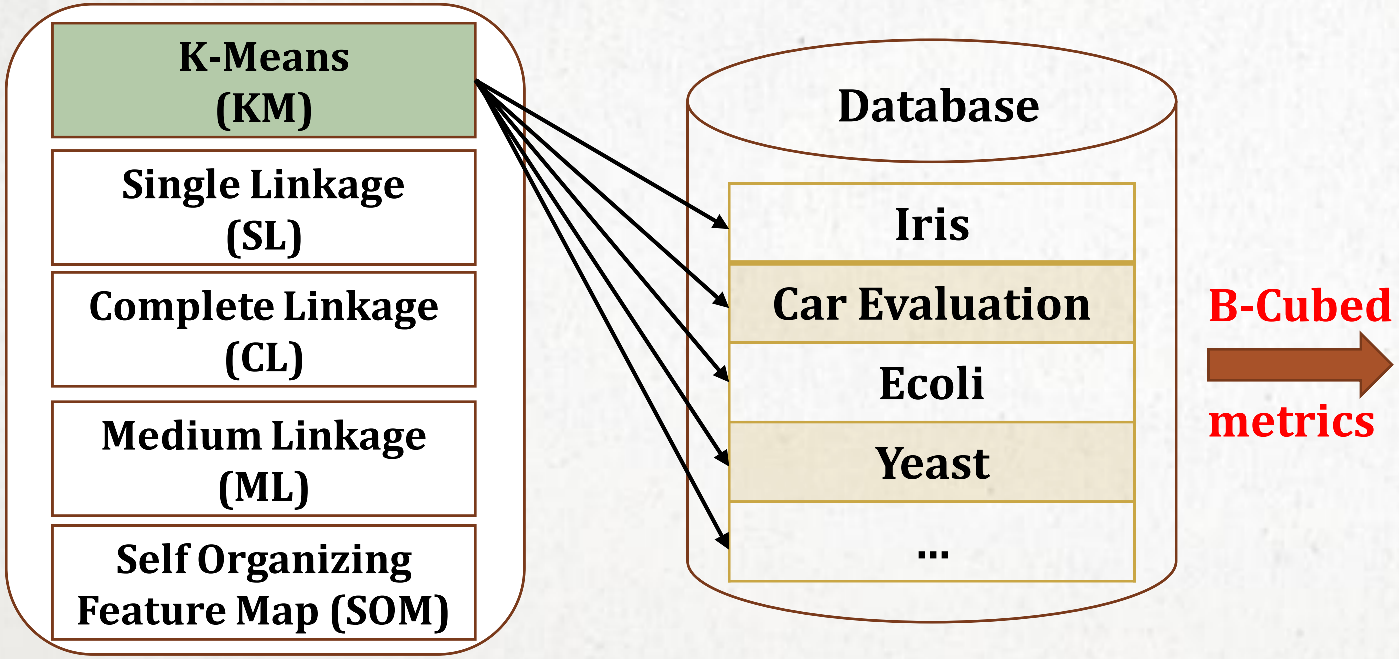
The StatLog and METAL project~

META LEARNING - METHODS



META LEARNING –EXAMPLE

Candidate algorithms



Total: 5

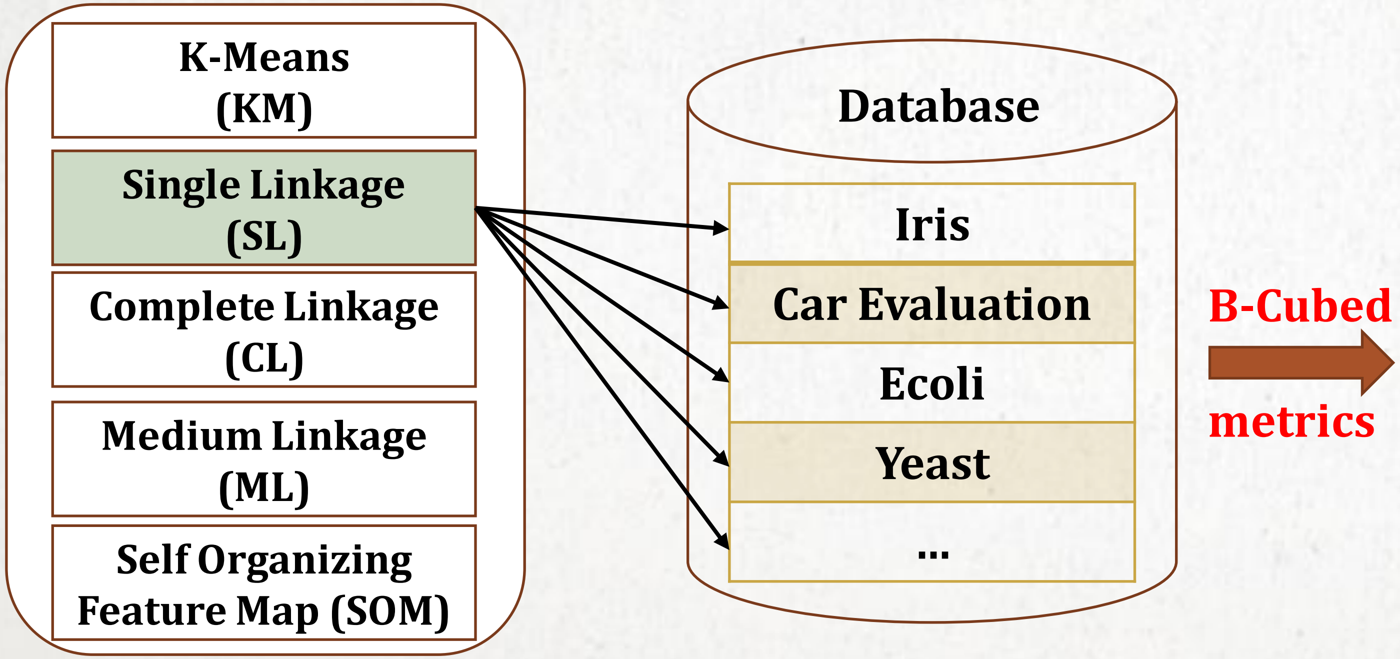
Total: 100

Table 2. Best *FBC* resulting values for the clustering solution

Dataset	KM	SL	CL	ML	SOM
Iris	0.75109	0.79579	0.75899	0.79484	0.75075
Car Eval.	0.42647	0.59709	0.36724	0.41249	0.38790
Haberman	0,70548	0,75917	0,72637	0,75512	0,55101
Ecoli	0.66051	0.46530	0.78537	0.68680	0.58789
Yeast	0.37361	0.38382	0.35382	0.38464	0.37607
Wine	0.93653	0.51704	0.74443	0.50954	0.94668
Tic-Tac-Toe	0.53386	0.70747	0.55915	0.61800	0.53404

META LEARNING –EXAMPLE

Candidate algorithms



Total: 5

Total: 100

Table 2. Best *FBC* resulting values for the clustering solution

Dataset	KM	SL	CL	ML	SOM
Iris	0.75109	0.79579	0.75899	0.79484	0.75075
Car Eval.	0.42647	0.59709	0.36724	0.41249	0.38790
Haberman	0,70548	0,75917	0,72637	0,75512	0,55101
Ecoli	0.66051	0.46530	0.78537	0.68680	0.58789
Yeast	0.37361	0.38382	0.35382	0.38464	0.37607
Wine	0.93653	0.51704	0.74443	0.50954	0.94668
Tic-Tac-Toe	0.53386	0.70747	0.55915	0.61800	0.53404

META LEARNING –EXAMPLE

Candidate algorithms

- K-Means (KM)
- Single Linkage (SL)
- Complete Linkage (CL)
- Medium Linkage (ML)
- Self Organizing Feature Map (SOM)

Total: 5

Database

- Iris
- Car Evaluation
- Ecoli
- Yeast
- ...

Total: 100

B-Cubed metrics

Table 2. Best *FBC* resulting values for the clustering solution

Dataset	KM	SL	CL	ML	SOM
Iris	0.75109	0.79579	0.75899	0.79484	0.75075
Car Eval.	0.42647	0.59709	0.36724	0.41249	0.38790
Haberman	0,70548	0,75917	0,72637	0,75512	0,55101
Ecoli	0.66051	0.46530	0.78537	0.68680	0.58789
Yeast	0.37361	0.38382	0.35382	0.38464	0.37607
Wine	0.93653	0.51704	0.74443	0.50954	0.94668
Tic-Tac-Toe	0.53386	0.70747	0.55915	0.61800	0.53404

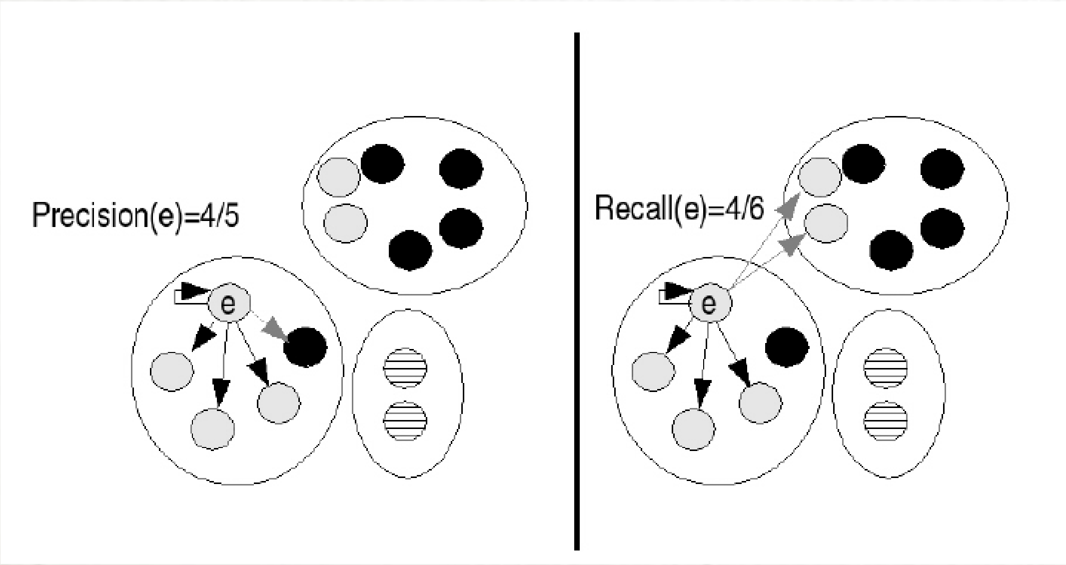
META LEARNING –EXAMPLE

B-Cubed Metrics:

$$FBC = \left(0.5 * \left(\frac{1}{\frac{1}{K}\sum_{i=1}^{K}\frac{CL(i)}{Cluster(i)}}\right) + 0.5 * \left(\frac{1}{\frac{1}{K}\sum_{i=1}^{K}\frac{CL(i)}{Label(i)}}\right)\right)^{-1}$$

The precision of each cluster

The recall rate of each class



$$FBC \in (0,1]$$

The closer to 1, the better!

META LEARNING –EXAMPLE

Transfer the FBC resulting values to ranking values

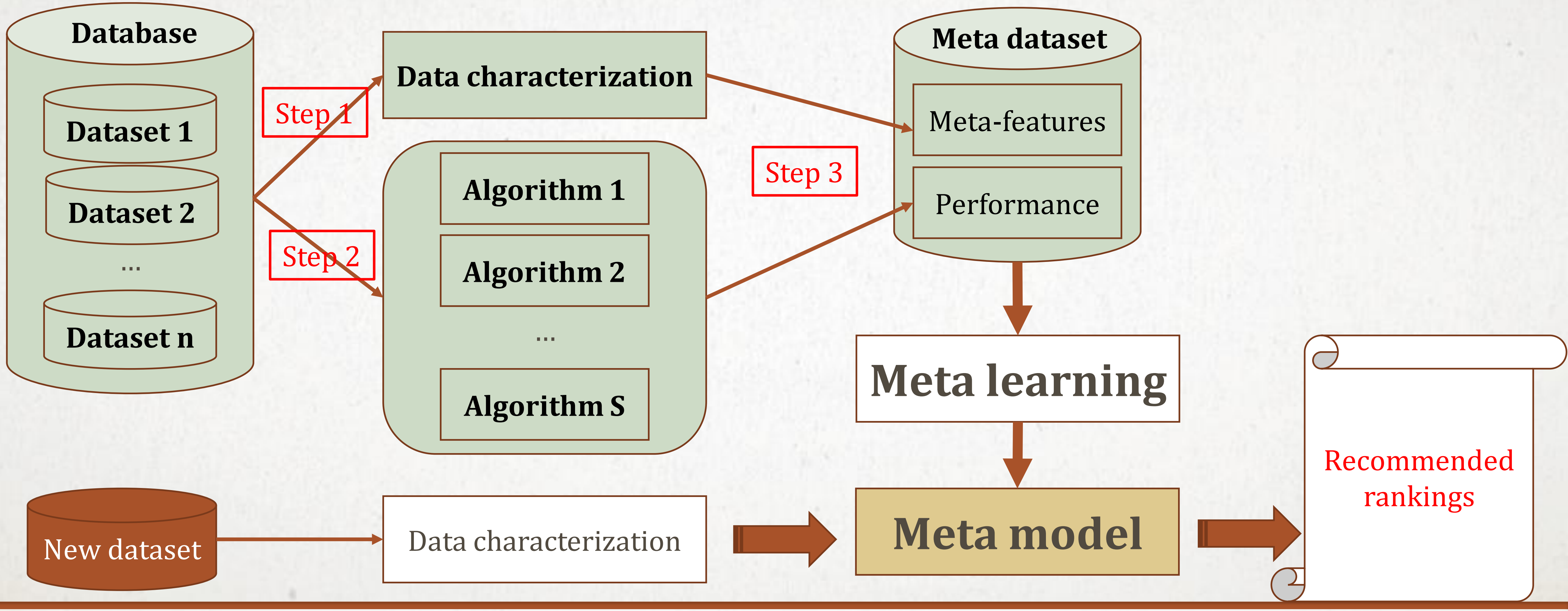
Table 2. Best *FBC* resulting values for the clustering solution

Dataset	KM	SL	CL	ML	SOM
Iris	0.75109	0.79579	0.75899	0.79484	0.75075
Car Eval.	0.42647	0.59709	0.36724	0.41249	0.38790
Haberman	0,70548	0,75917	0,72637	0,75512	0,55101
Ecoli	0.66051	0.46530	0.78537	0.68680	0.58789
Yeast	0.37361	0.38382	0.35382	0.38464	0.37607
Wine	0.93653	0.51704	0.74443	0.50954	0.94668
Tic-Tac-Toe	0.53386	0.70747	0.55915	0.61800	0.53404

Table 3. Predictive table built with ranking values

Dataset	KM	SL	CL	ML	SOM
Iris	4	1	3	2	5
Car Eval.	2	1	5	3	4
Haberman	4	1	3	2	5
Ecoli	3	5	1	2	4
Yeast	4	2	5	1	3
Wine	2	4	3	5	1
Tic-Tac-Toe	5	1	3	2	4

META LEARNING - METHODS



META LEARNING –EXAMPLE

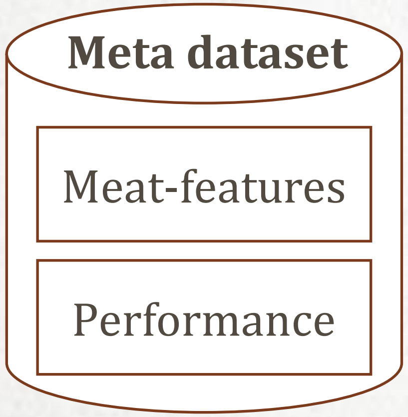


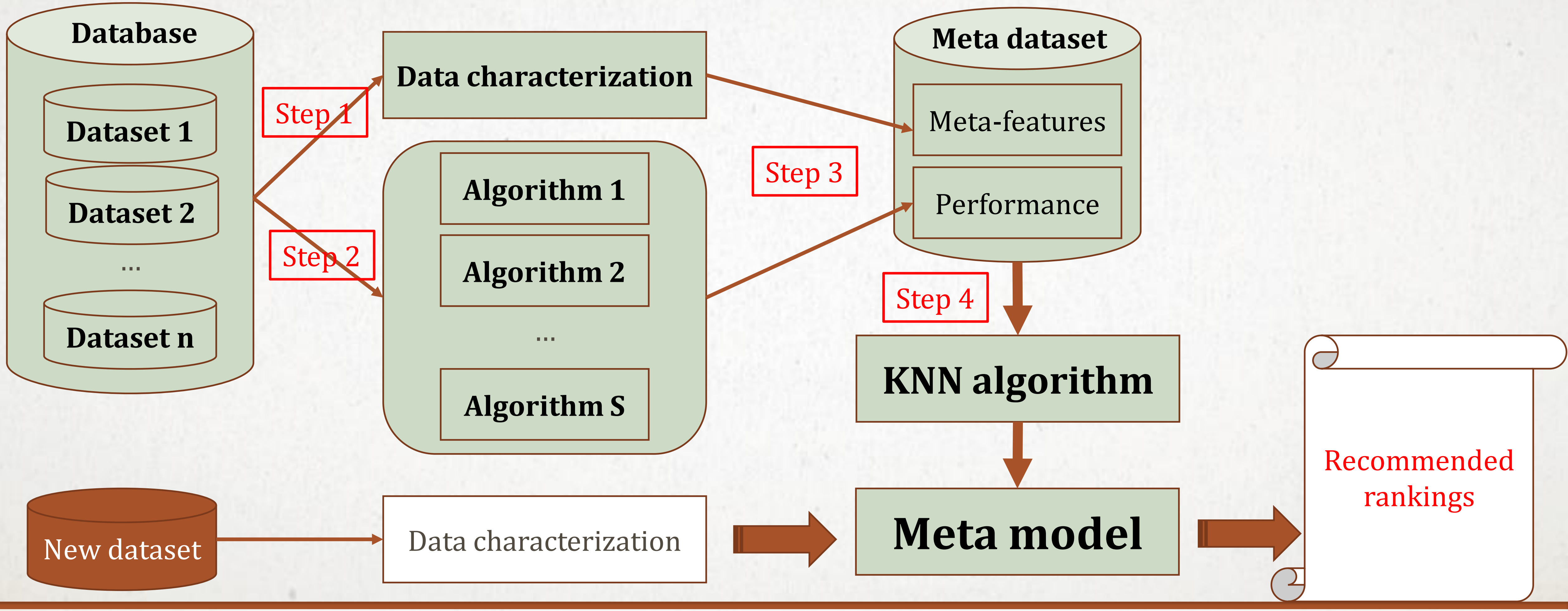
Table 1. Meta-attribute values for some datasets

Dataset/Value	1	2	3	4	5	6	7	8	9	10
Iris	7.23	2.00	0.00	0.00	1.00	0.59	0.07	2.23	0.00	0.00
Car Eval.	10.75	2.58	0.00	1.00	0.00	0.00	0.00	0.00	0.00	5.38
Ecoli	8.39	2.81	0.00	0.00	1.00	0.18	3.59	54.19	0.00	0.00
Yeast	10.54	3.00	0.00	0.00	1.00	0.09	2.91	31.56	0.00	0.00
Wine	7.48	3.70	0.00	0.00	1.00	0.30	0.35	2.97	0.00	0.00
Tic-Tac-Toe	9.90	3.17	0.00	1.00	0.00	0.00	0.00	0.00	0.01	3.91

Table 3. Predictive table built with ranking values

Dataset	KM	SL	CL	ML	SOM
Iris	4	1	3	2	5
Car Eval.	2	1	5	3	4
Haberman	4	1	3	2	5
Ecoli	3	5	1	2	4
Yeast	4	2	5	1	3
Wine	2	4	3	5	1
Tic-Tac-Toe	5	1	3	2	4

META LEARNING - METHODS



META LEARNING –EXAMPLE

New problem

Haberman's Survival dataset

Meta dataset

Meta-features

Performance

KNN algorithm

3-nearest neighbors

- **Iris**
- **Car Eval.**
- **Haberman**

Table 5. Meta-attributes extracted from the new problem

1	2	3	4	5	6	7	8	9	10
8.26	1.58	0.00	0.00	1.00	0.05	1.06	6.27	0.00	0.00

META LEARNING –EXAMPLE

New problem

Haberman's Survival dataset

KNN

3-nearest neighbors

- **Iris**
- **Car Eval.**
- **Haberman**

Recommended rankings

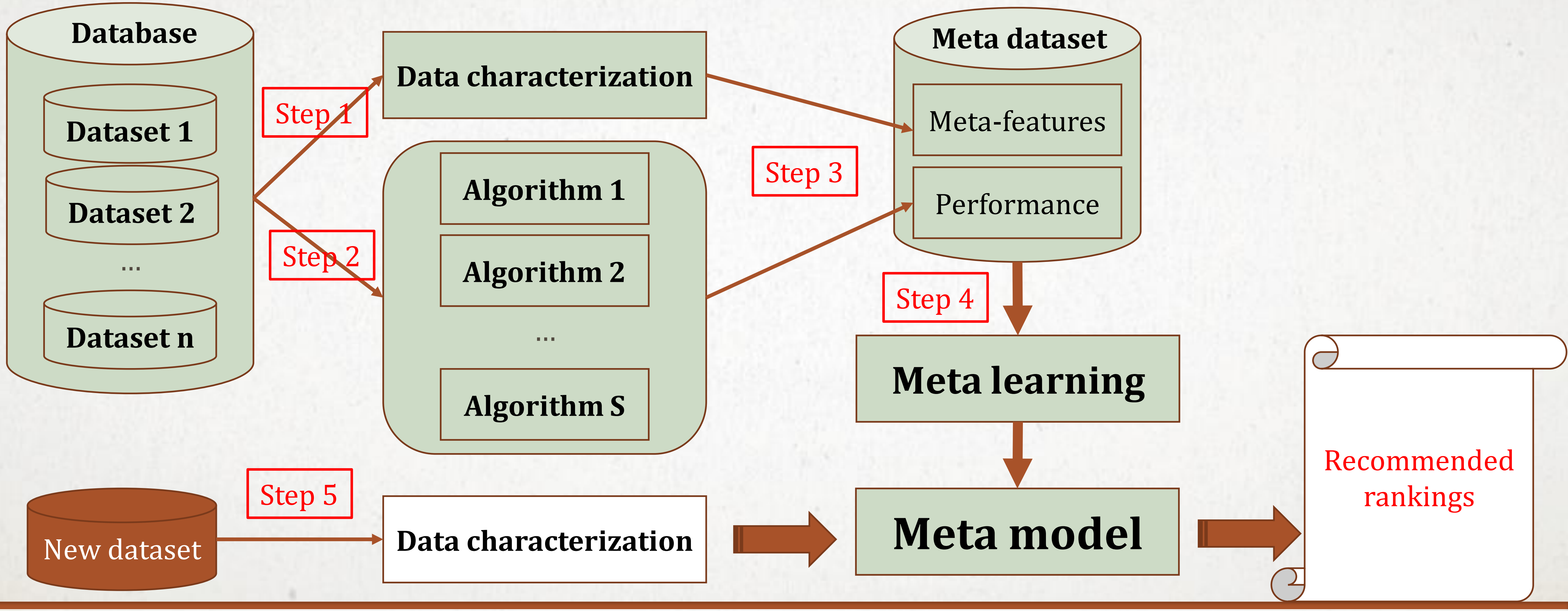
KM	SL	CL	ML	SOM
3	1	4	2	5

Average ranking

Table 3. Predictive table built with ranking values

Dataset	KM	SL	CL	ML	SOM
Iris	4	1	3	2	5
Car Eval.	2	1	5	3	4
Haberman	4	1	3	2	5

META LEARNING - METHODS



META LEARNING

References:

[1] Giraud-Carrier, et al. "Introduction to the special issue on meta-learning." Machine learning 54, no. 3 (2004): 187-193.

[2] De Souto, et al. "Ranking and selecting clustering algorithms using a meta-learning approach." In Neural Networks, 2008. IJCNN 2008.

[3] Amigó, et al. "A comparison of extrinsic clustering evaluation metrics based on formal constraints." *Information retrieval* 12, no. 4 (2009): 461-486.

[4] Ferrari, et al. "Clustering algorithm recommendation: a meta-learning approach." In International Conference on Swarm, Evolutionary, and Memetic Computing, pp. 143-150, 2012.

[5] Bruno, et al. “A New Data Characterization for Selecting Clustering Algorithms Using Meta-Learning.” Information Sciences, unpublished, 2018.

Q & A