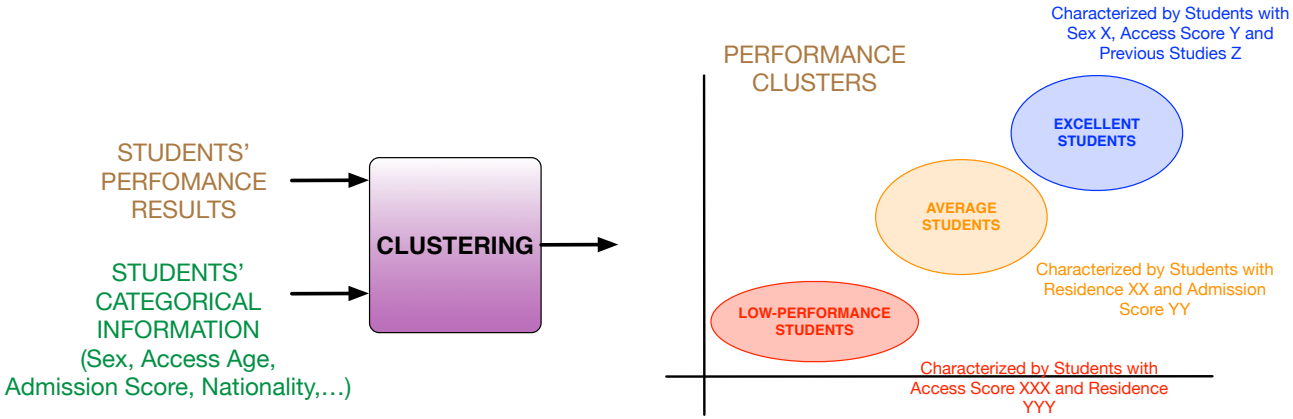
Algorithms for Student data processing

Dr. Jose Lopez Vicario,
Profesor Agregado – Universidad Autónoma de Barcelona (UAB)

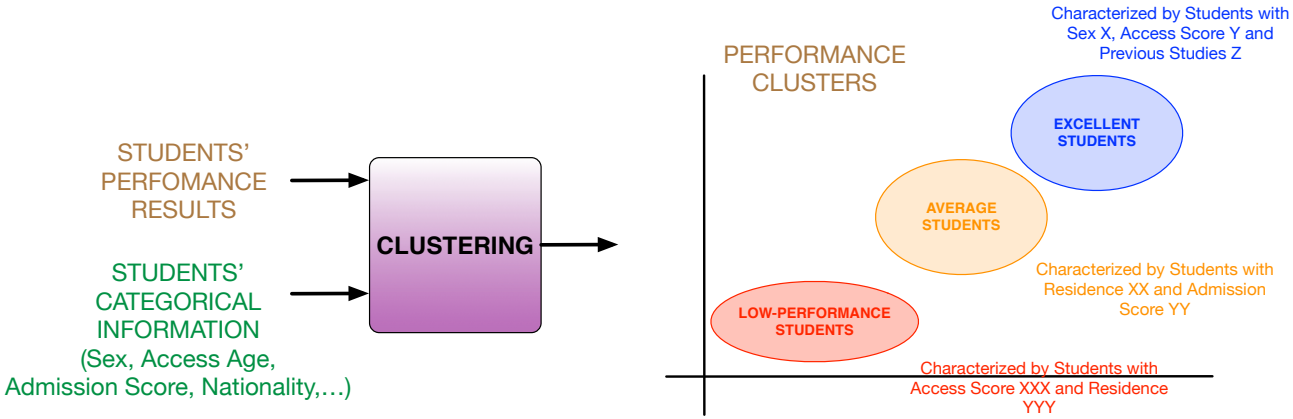
SPEET has developed two Data Processing Tools:

1. Clustering/Classification Tool:

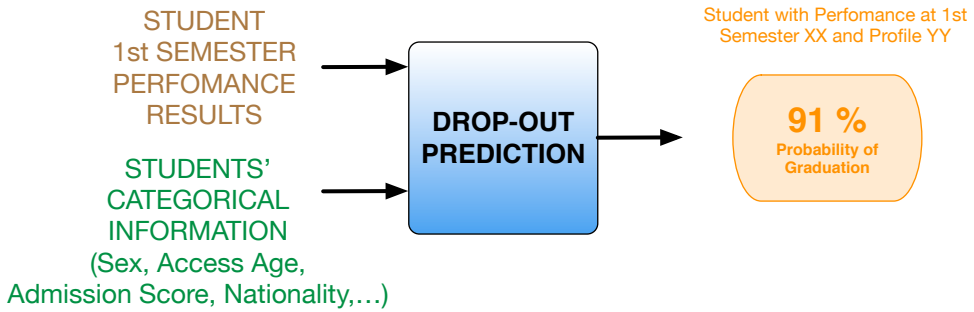


SPEET has developed two Data Processing Tools:

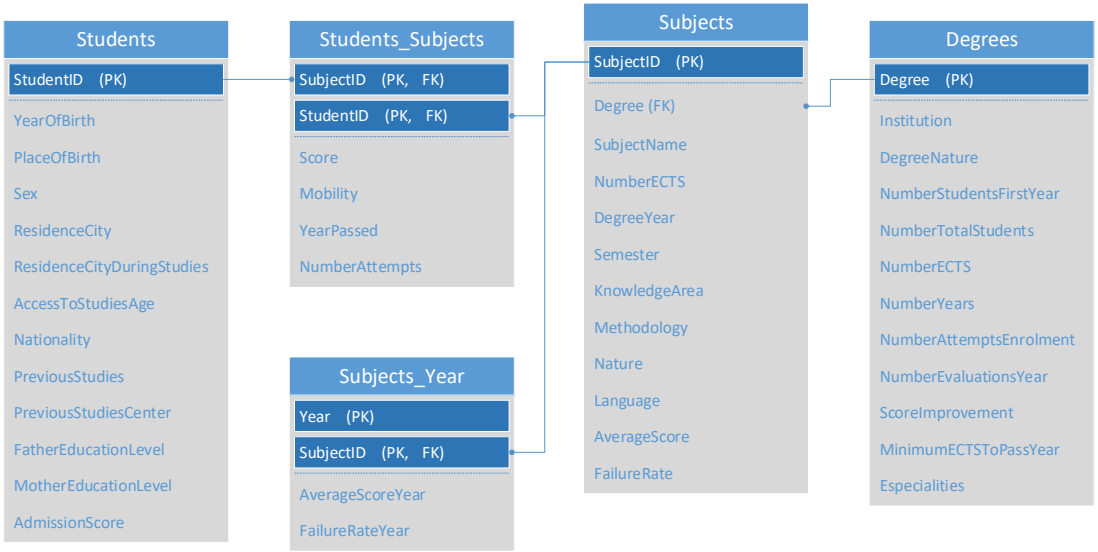
1. Clustering/Classification Tool:



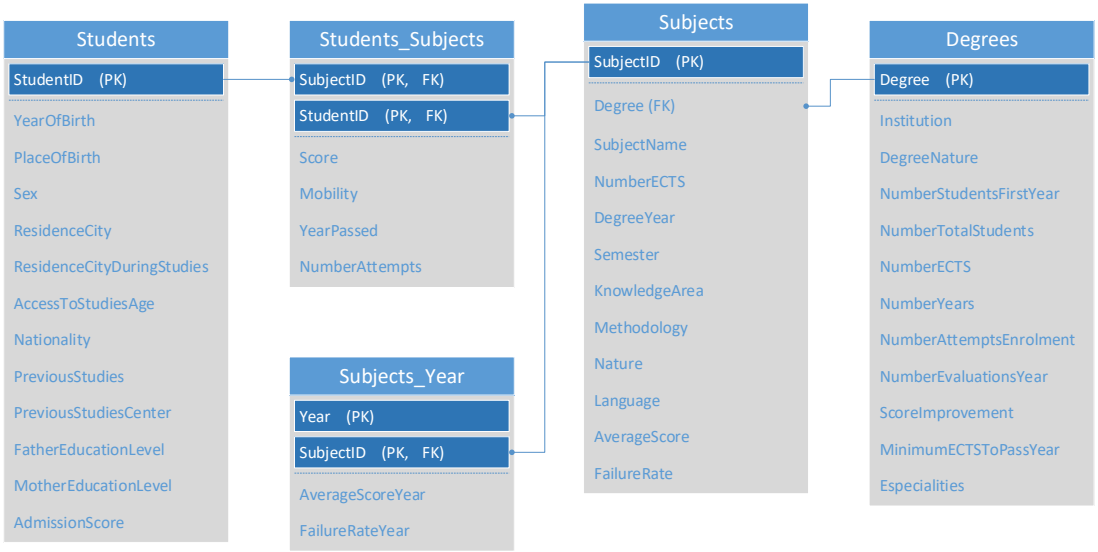
2. Drop-out Prediction Tool:



SPEET Unified Data Format



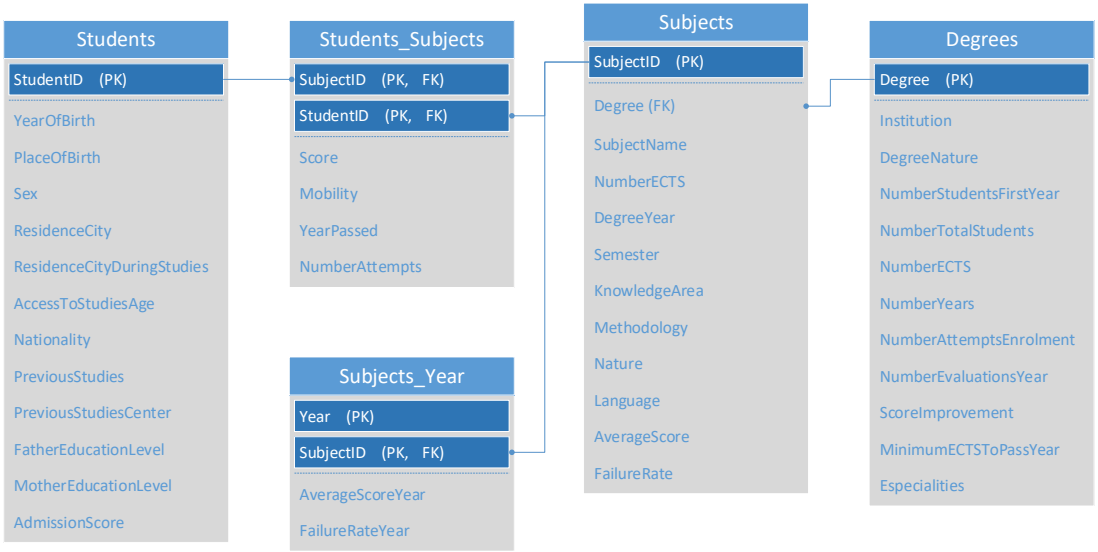
SPEET Unified Data Format



Data collection issues…

… different amount of information.

SPEET Unified Data Format

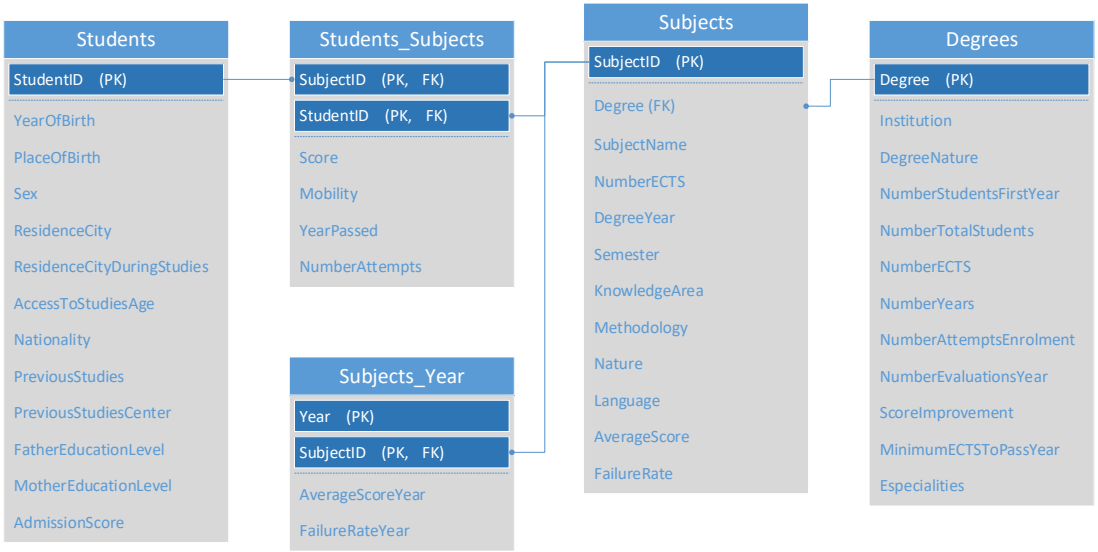


Data collection issues…

… different amount of information.

… different languages.

SPEET Unified Data Format



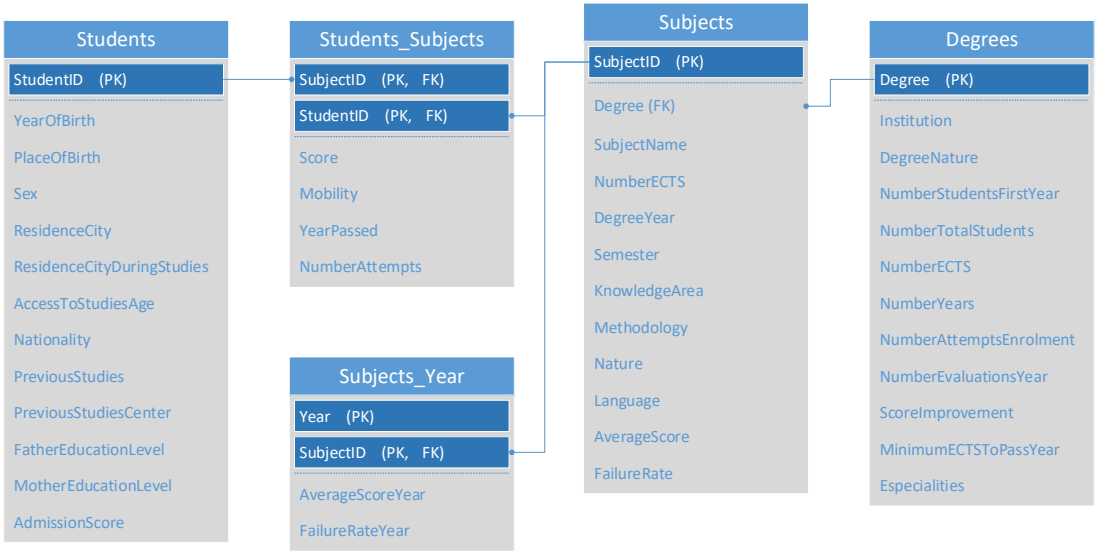
Data collection issues…

… different amount of information.

… different languages.

… different automation levels.

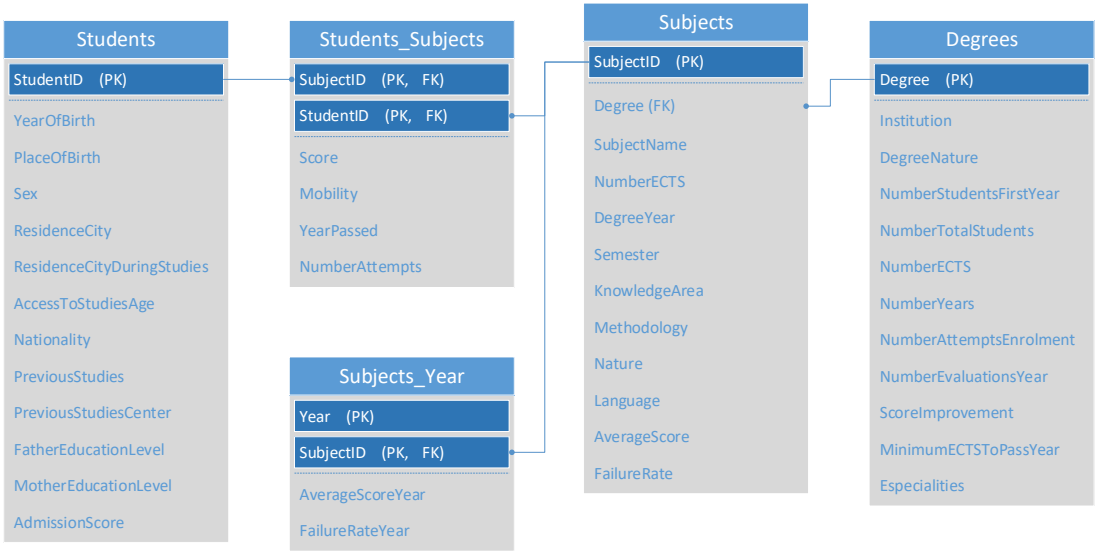
SPEET Unified Data Format



Data collection issues…

… different amount of information.
… different languages.
… different automation levels.
… different bureaucracy speeds.

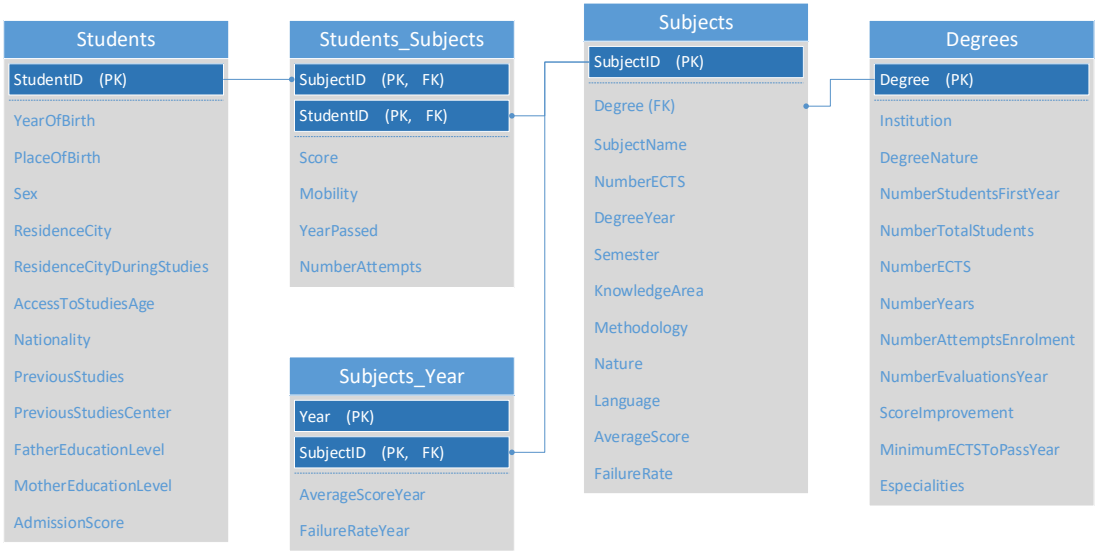
SPEET Unified Data Format



Data collection issues…

- … different amount of information.
- … different languages.
- … different automation levels.
- … different bureaucracy speeds.
- … different confidential requirements.

SPEET Unified Data Format

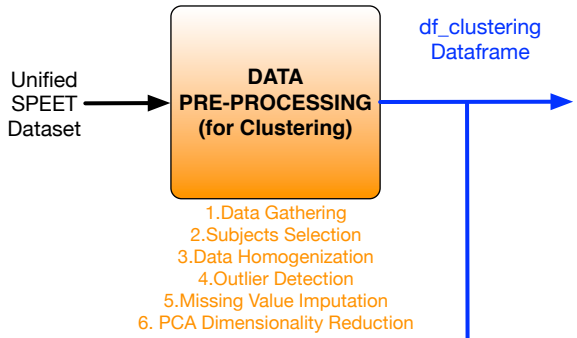


Data collection issues…

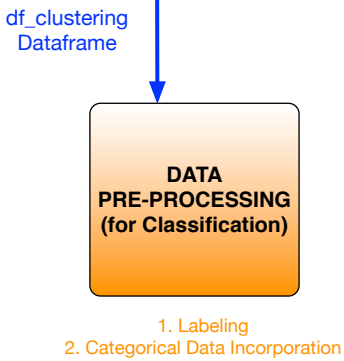
- … different amount of information.
- … different languages.
- … different automation levels.
- … different bureaucracy speeds.
- … different confidential requirements.

ONLY 6 PARTNERS FROM 4 COUNTRIES !!!!!

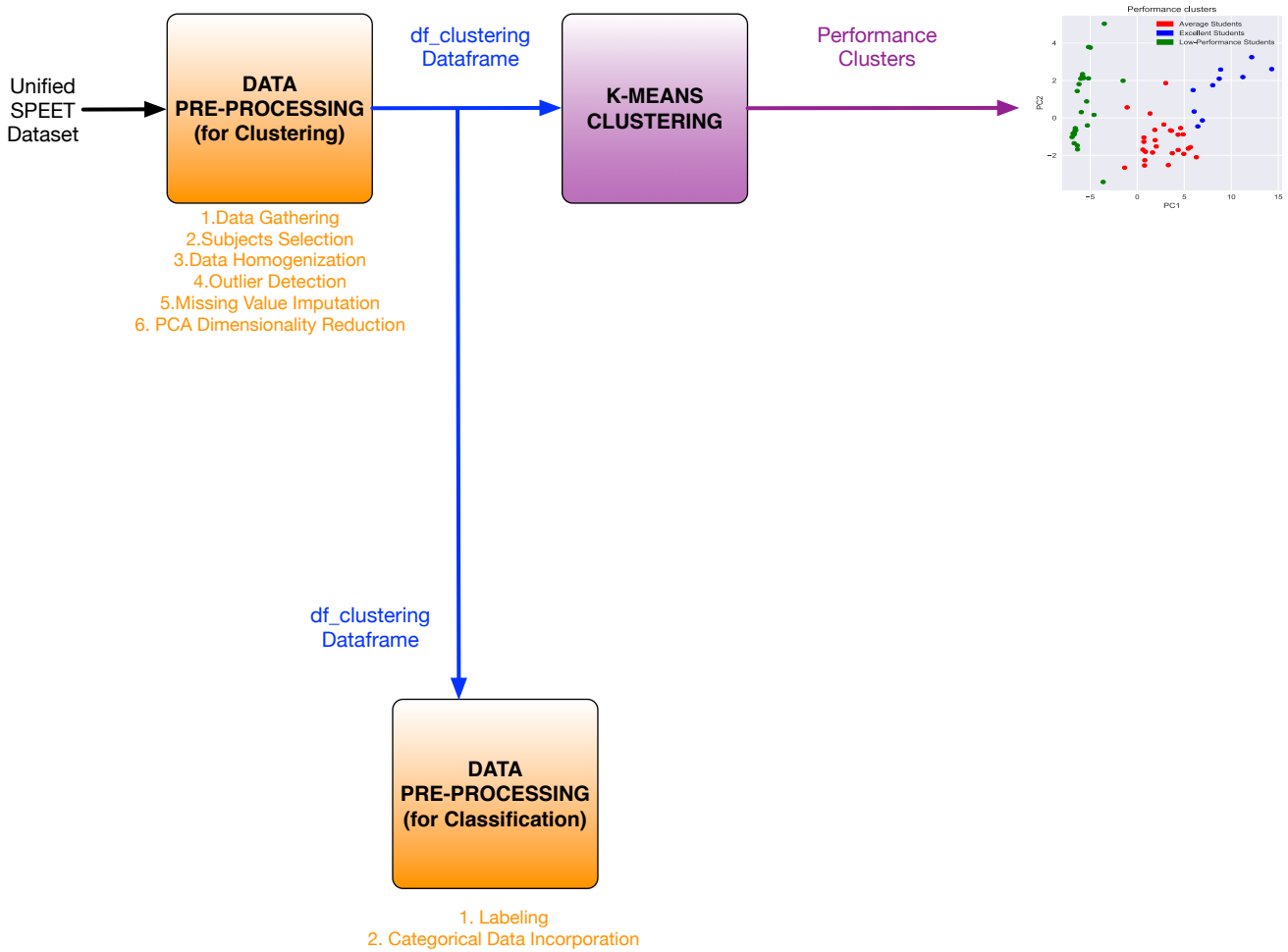
Clustering/Classification Tool



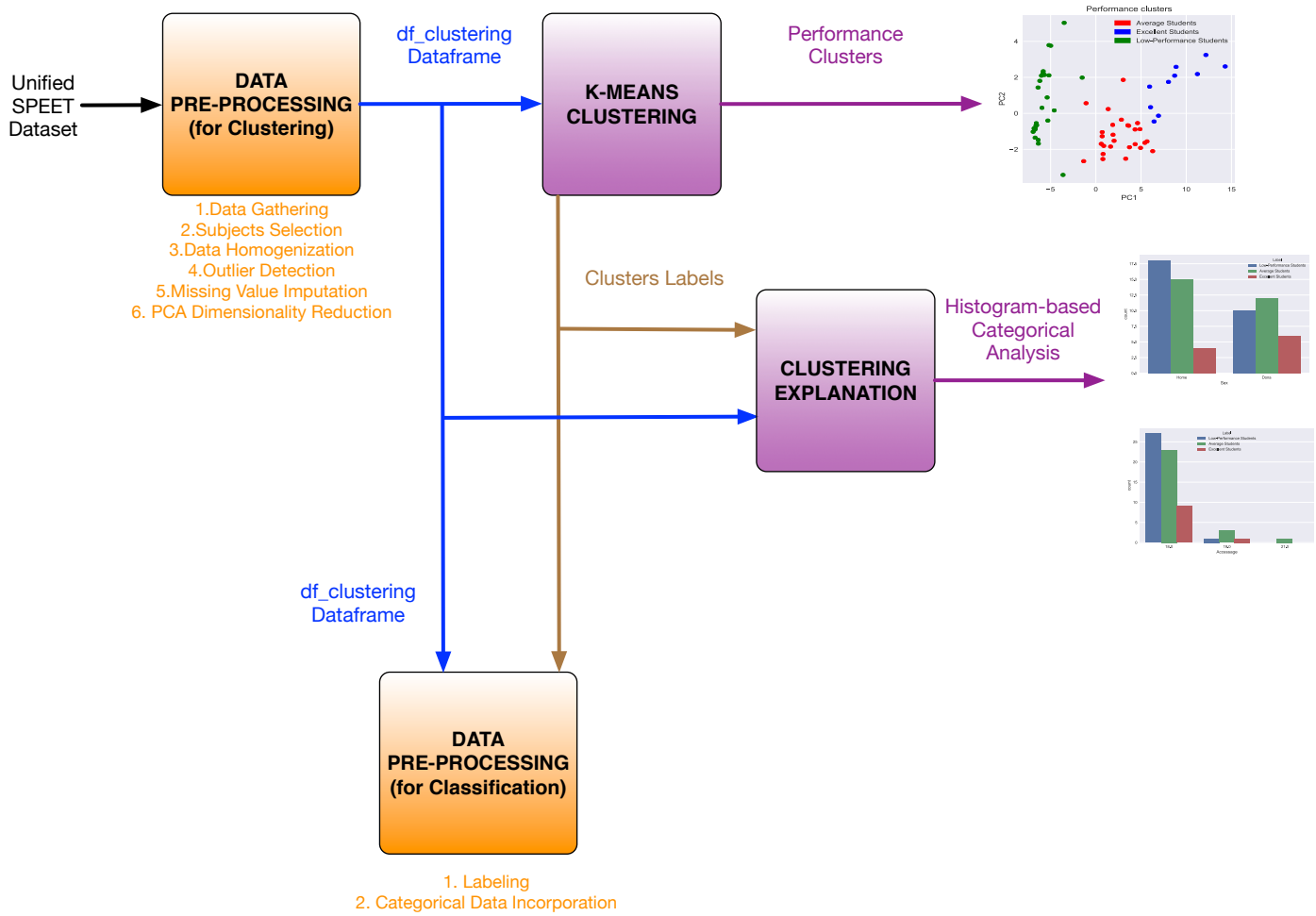
Student	Subj 1 Score	Subj 2 Score	...	Subj M Score	Sex	Access Age	Previous Studies	Admission Score	Nationality
1	8	8	...	5	M	20	Secondary	5	FR
2	6	6	...	5	F	18	ProfessionalStudies	7	SPA



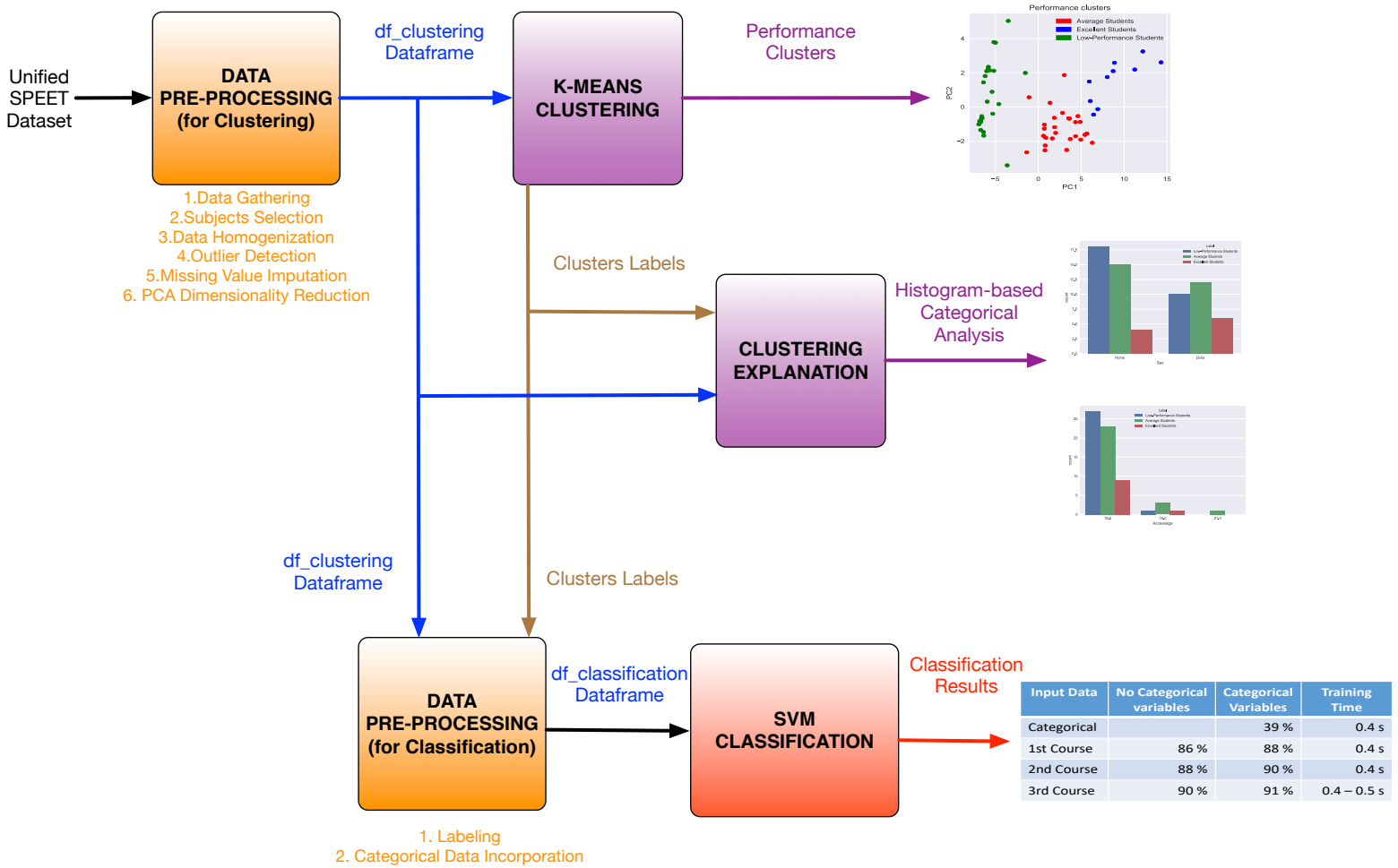
Clustering/Classification Tool



Clustering/Classification Tool

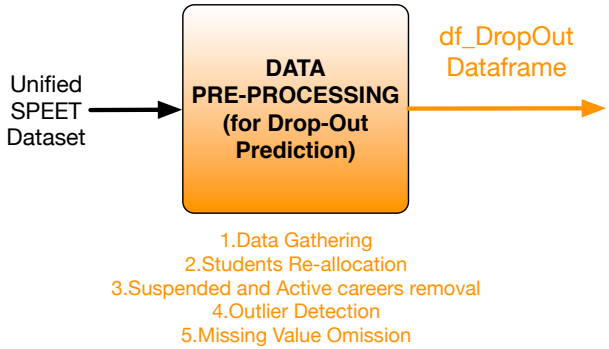


Clustering/Classification Tool



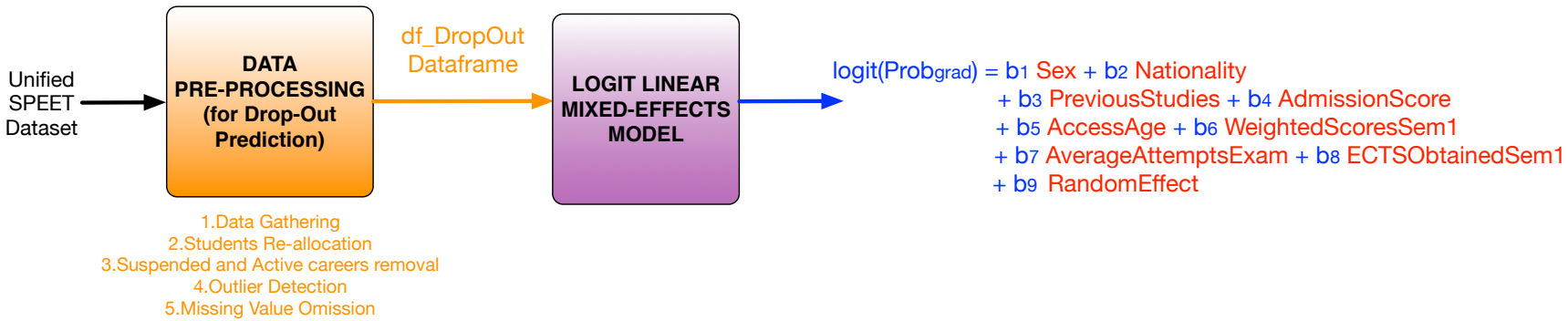
Input Data	No Categorical variables	Categorical Variables	Training Time
Categorical		39 %	0.4 s
1st Course	86 %	88 %	0.4 s
2nd Course	88 %	90 %	0.4 s
3rd Course	90 %	91 %	0.4 – 0.5 s

Drop-out Prediction Tool



Student	Sex	Nationality	Previous Studies	Admission Score	Access Age	Weighted Scores Sem1	Average Attempts Exam	ECTS Obtained Sem1	Drop-out
1	M	FR	Secondary	5	20	5	FR	24	NO
2	F	SPA	ProfessionalStudies	7	18	7	SPA	18	YES

Drop-out Prediction Tool



$Prob_{grad}$ > 0.5 is considered as Graduated Student.

Accuracy of the model: 90.87%

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

2.Could we observe clear students' profiles at these groups based on categorical variables such as (age, previous studies, nationality, etc.)? (Student-wise characterization)

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

2.Could we observe clear students' profiles at these groups based on categorical variables such as (age, previous studies, nationality, etc.)? (Student-wise characterization)

3.The quality of cluster separation (clearly or badly separated clusters) can be explained by means of the way categorical variables (age, previous studies, nationality, etc.) are distributed (homogeneous vs. heterogeneous students' profiles)?

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

2.Could we observe clear students' profiles at these groups based on categorical variables such as (age, previous studies, nationality, etc.)? (Student-wise characterization)

3.The quality of cluster separation (clearly or badly separated clusters) can be explained by means of the way categorical variables (age, previous studies, nationality, etc.) are distributed (homogeneous vs. heterogeneous students' profiles)?

4.Could we see different or similar students' characteristics at different degrees at the same institution? (Institution-wise characterization)

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

2.Could we observe clear students' profiles at these groups based on categorical variables such as (age, previous studies, nationality, etc.)? (Student-wise characterization)

3.The quality of cluster separation (clearly or badly separated clusters) can be explained by means of the way categorical variables (age, previous studies, nationality, etc.) are distributed (homogeneous vs. heterogeneous students' profiles)?

4.Could we see different or similar students' characteristics at different degrees at the same institution? (Institution-wise characterization)

5.Could we see different or similar students' characteristics at the same degree but comparing different institutions? (Degree-wise characterization)

Questions to Answer

1.Could we separate students at different groups (clusters) based on their performance behavior?

2.Could we observe clear students' profiles at these groups based on categorical variables such as (age, previous studies, nationality, etc.)? (Student-wise characterization)

3.The quality of cluster separation (clearly or badly separated clusters) can be explained by means of the way categorical variables (age, previous studies, nationality, etc.) are distributed (homogeneous vs. heterogeneous students' profiles)?

4.Could we see different or similar students' characteristics at different degrees at the same institution? (Institution-wise characterization)

5.Could we see different or similar students' characteristics at the same degree but comparing different institutions? (Degree-wise characterization)

6.Could we see if one or several courses determine the behavior of students at one degree?

Clustering/Classification Tool Demo