Student Profile for Enhancing Engineering Tutoring

Erasmus+

Student Profile for Enhancing Engineering Tutoring

European Commission

sepie

SERVICIO ESPAÑOL PARA LA INTERNACIONALIZACIÓN DE LA EDUCACIÓN

UAB

Student Profile for Enhancing Engineering Tutoring

Erasmus+

The Project

This project goal emerges from the potential synergy among

a) the huge amount of academic data actually existing at the academic departments of faculties and schools, and
b) the maturity of data science in order to provide algorithms and tools to analyse and extract information from what is more commonly

A rich picture can be extracted from this data if conv
this project is to apply data mining algorithms to p
information about and to identify student profiles. An
referring to within the project scope is, for example:
Students that are blocked on a certain set of subjects
etc.

With such classification that, of course, devise a more
that will be established from the very beginning of
patterns will be depicted. Comparison among the differ
order to establish correlations and get a more complete

1.- Available data and a priori characterization of poten
2.- Pre-processing of data. First examples of processi
from partners
3.- Data analysis at partner level. Interpretation of categ
4.- Setting up the basis and definitions for an IT-based assistant tutoring system.
5.- Prototyping the developed tools into a website based tool.

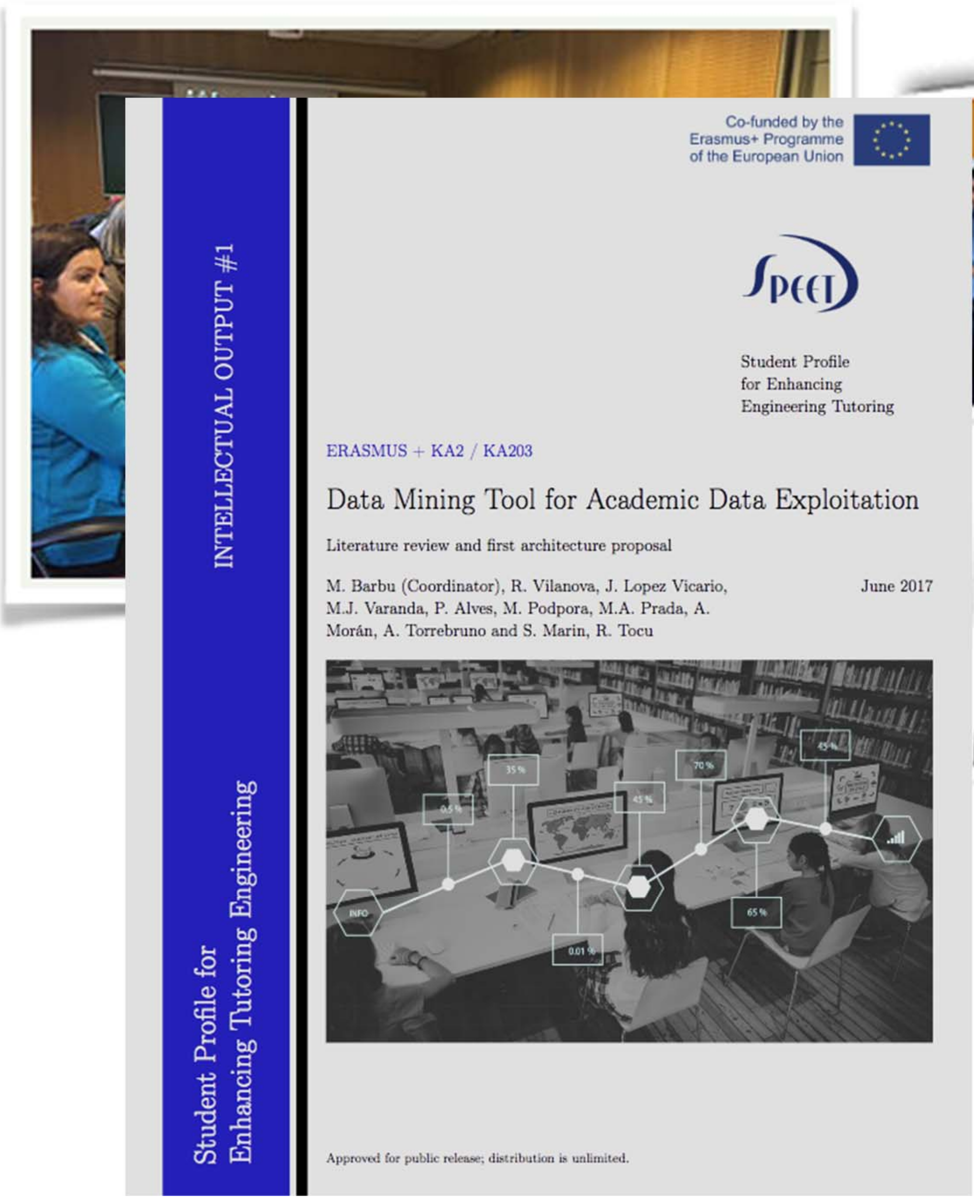
SPEET ... in short

SPEET is an ERASMUS+ project aimed to determine and categorize the different profiles for engineering students across Europe. The main rationale behind this proposal is the observation that students performance can be classified according to their behavior while conducting their studies. After years of teaching and sharing thoughts among colleagues from different EU institutions it seems students could obey to some classification according to the way they face their studies. Therefore, if it would be possible to know what kind of student one student is, this may be of valuable help for tutors.

[www.speet-project.com]

Student Profile for Enhancing Engineering Tutoring

Erasmus+


Co-funded by the Erasmus+ Programme of the European Union

Student Profile for Enhancing Engineering Tutoring

ERASMUS + KA2 / KA203

Data Mining Tool for Academic Data Exploitation

Literature review and first architecture proposal

M. Barbu (Coordinator), R. Vilanova, J. Lopez Vicario, M.J. Varanda, P. Alves, M. Podpora, M.A. Prada, A. Morán, A. Torrebruno and S. Marin, R. Tocu

June 2017

Approved for public release; distribution is unlimited.

INTELLECTUAL OUTPUT #1

Student Profile for Enhancing Tutoring Engineering

[www.speet-project.com]

Student Profile for Enhancing Engineering Tutoring

Erasmus+

Co-funded by the Erasmus+ Programme of the European Union

Student Profile for Enhancing Engineering Tutoring

INTELLECTUAL OUTPUT #1

Student Profile for Enhancing Tutoring Engineering

ERASMUS + KA2 / KA203

Data Mining Tool for Academic Data Exploitation

Literature review and first architecture proposal

M. Barbu (Coordinator), R. Vilanova, J. Lopez Vicario, M.J. Varanda, P. Alves, M. Podpora, M.A. Prada, A. Morán, A. Torrebruno and S. Marin, R. Tocu

June 2017

Approved for public release; distribution is unlimited.

SPEET: AN INTERNATIONAL COLLABORATIVE EXPERIENCE IN DATA MINING FOR EDUCATION

R. VIlanova[1], J. Vicario[1],M. Prada[2] ,M. Barbu[3],M. Dominguez[3],M. J. Varanda[4],M. Podpora[5],U. Spagnolini[6]

[1]*Universitat Autonoma Barcelona (SPAIN)*
[2]*Universidad de Leon (SPAIN)*
[3]*University Dunarea de Jos, Galati (ROMANIA)*
[4]*Instituto Politecnico de Bragança (PORTUGAL)*
[5]*Opole University of Technology (POLAND)*
[6]*Politecnico de Milano (ITALY)*

Abstract

This paper presents the collaborative experience that is under development as the European ERASMUS+ project SPEET (Student Profile for Enhancing Engineering Tutoring). This project goal emerges from the potential synergy among a) the huge amount of academic data actually existing at the academic departments of faculties and schools, and b) the maturity of data science in order to provide algorithms and tools to analyse and extract information from what is more commonly referred to Big Data. A rich picture can be extracted from this data if conveniently processed. The main purpose of this project is to apply data mining algorithms to process this data in order to extract information about and to identify student profiles. An idea of the student profile we are referring to within the project scope is, for example: Students that finish degree on time, Students that are blocked on a certain set of subjects, Students that leave degree earlier, etc

Keywords: Educational Data Mining, drop-off, tutoring action support.

1 INTRODUCTION

Data has always been a significant asset for institutions, and has been used to inform their day-to-day operational decisions as well as longer-term business and strategic decisions. On a strategic scale, data is used to inform senior management's business planning and overall strategy for their institutions. Student enrolment data, both historical and projected, as well as estates data, will influence the plans institutions make to build new buildings or refit current buildings to meet projected need. Financial data influences strategic decisions on expanding or reducing particular faculties or services provided.

From a more purely educational point of view, the available academic data can be collected, linked together and analysed to provide insights into student behaviours and identify patterns to potentially predict future outcomes. In this paper, usually available data will be described as well as its potential use for the benefit of students. The use of academic data for supporting tutoring action is where we will put the focus on.

Higher education institutions are not an exception and the use of analytics in education has grown in recent years for four primary reasons [1]: a substantial increase in data quantity, improved data formats, advances in computing, and increased sophistication of tools available for analytics. In recent years, the sophistication and ease of use of tools for analyzing data make it possible for an increasing range of researchers to apply data mining methodology without needing extensive experience in computer programming. Many of these tools are adapted from the business intelligence field. Higher education institutions have always operated in an information-rich landscape, generating and collecting vast amounts of data each day. A coarse classification of the types of data that higher education institutions deal with every day: Student record data, Staff data, Admissions and applications data, Financial data, Alumni data, Course data, Facilities data.

In commercial fields, business and organizations are deploying sophisticated analytic techniques to evaluate rich data sources, identify patterns within the data and exploit these patterns in decision making. Recently researchers and developers from the educational community started exploring the

[www.speet-project.com]

Student Explanatory Information

Attribute	Type of Variable
Student ID	Natural Number
YearOfBirth	Natural Number
PlaceOfBirth	Factor
Sex	Factor (Female, Male)
ResidenceCity	Factor
ResidenceCityDuringStudies	Factor
AccessToStudiesAge	Natural Number
Nationality	Factor
PreviousStudies	Factor (SciencesSecondary, TechnologicalSecondary, LiteratureSecondary,ProfessionalStudies)
PreviousStudiesCenter	Factor
FatherEducationLevel	Factor (PrimaryLevel, SecondaryLevel, UniversityLevel, DoctorateLevel)
MotherEducationLevel	Factor (PrimaryLevel, SecondaryLevel, UniversityLevel, DoctorateLevel)
AdmissionScore	Real Number

Degree Information

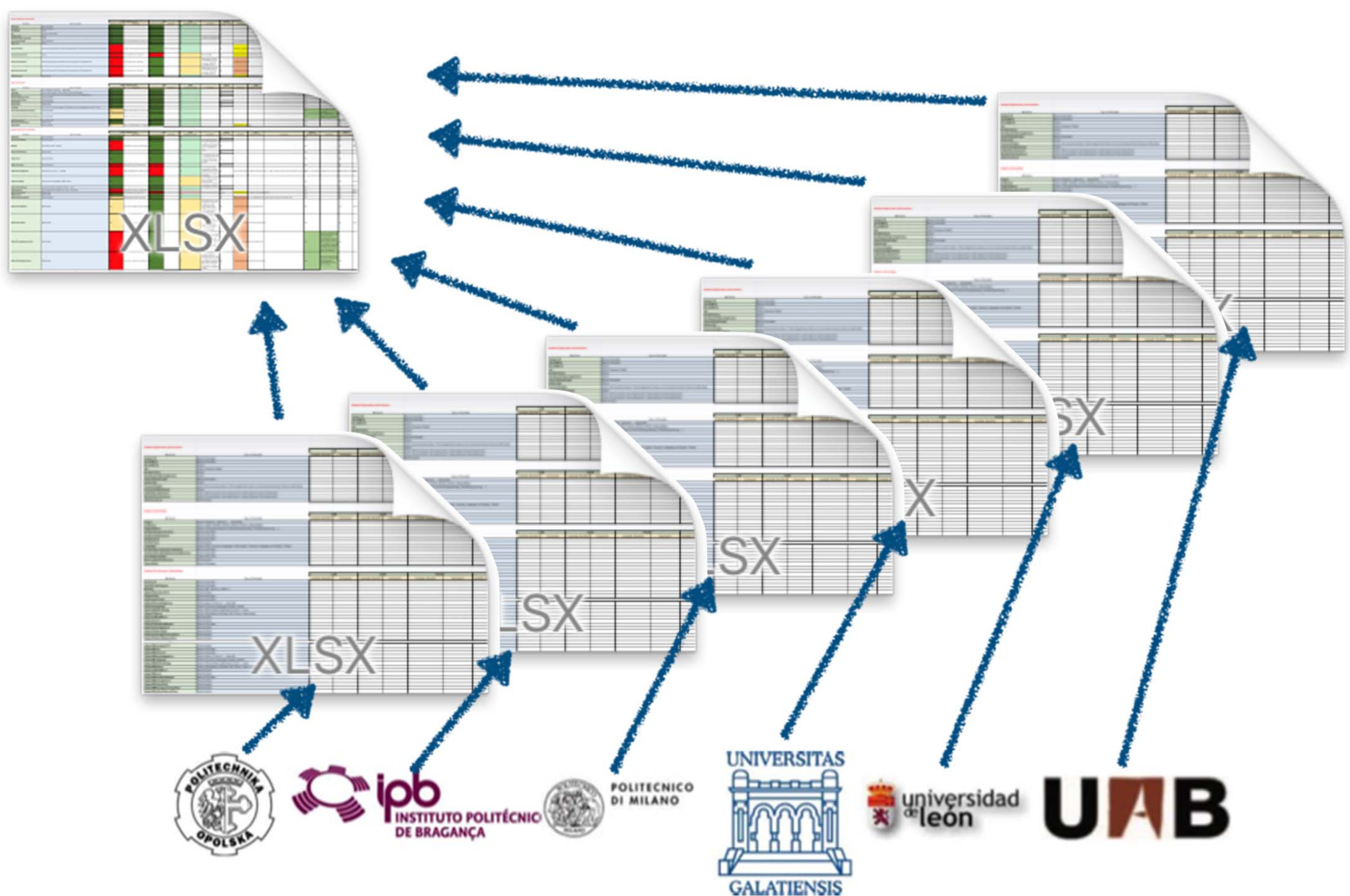
Attribute	Type of Variable
Degree	Factor (Degree1, Degree2, ..., DegreeM)
Institution	Factor (UAB, POLIMI, ULEON, GALATI, OPOLE, BRAGANÇA)
DegreeNature	Factor (ComputerScience, ElectricalEngineering, ControlEngineering, ...)
NumberStudentsFirstYear	Natural Number
NumberTotalStudents	Natural Number
NumberECTS	Natural Number
NumberYears	Real Number
Languages	Factor (Only Country Language, Only English, Country Language and English, Other)
NumberAttemptsToEnroleSubject	Natural Number
NumberAttemptsToBeEvaluatedOneYear	Natural Number
ScoreImprovement	Factor (YES, NO)
MinimumECTSToPassYear	Real Number
Especialities	Natural Number

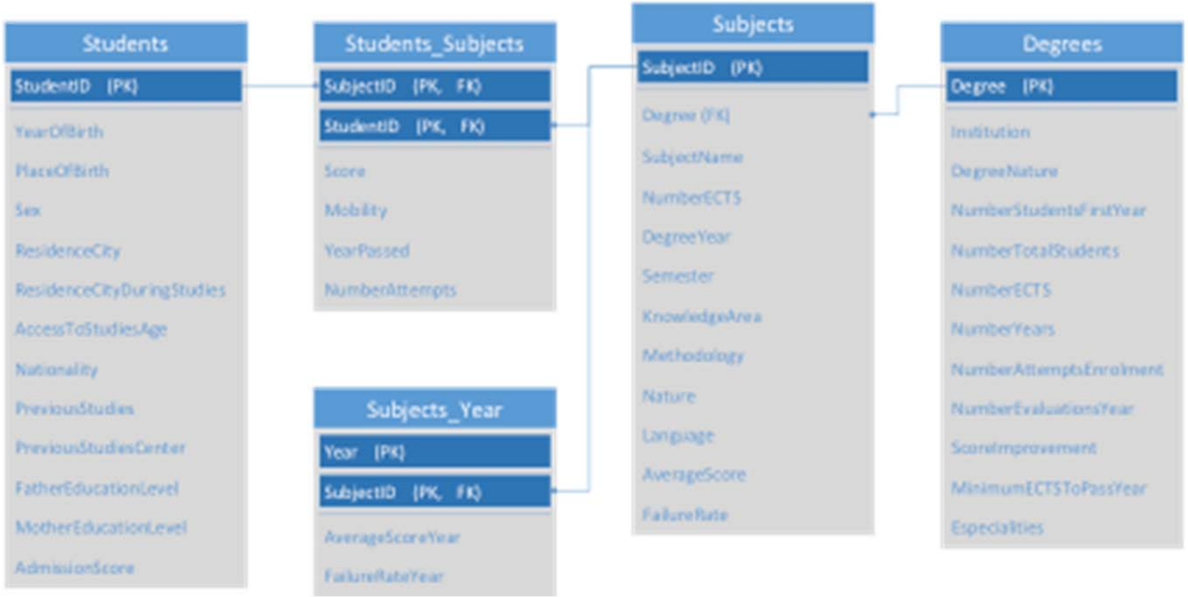
Student Performance Information

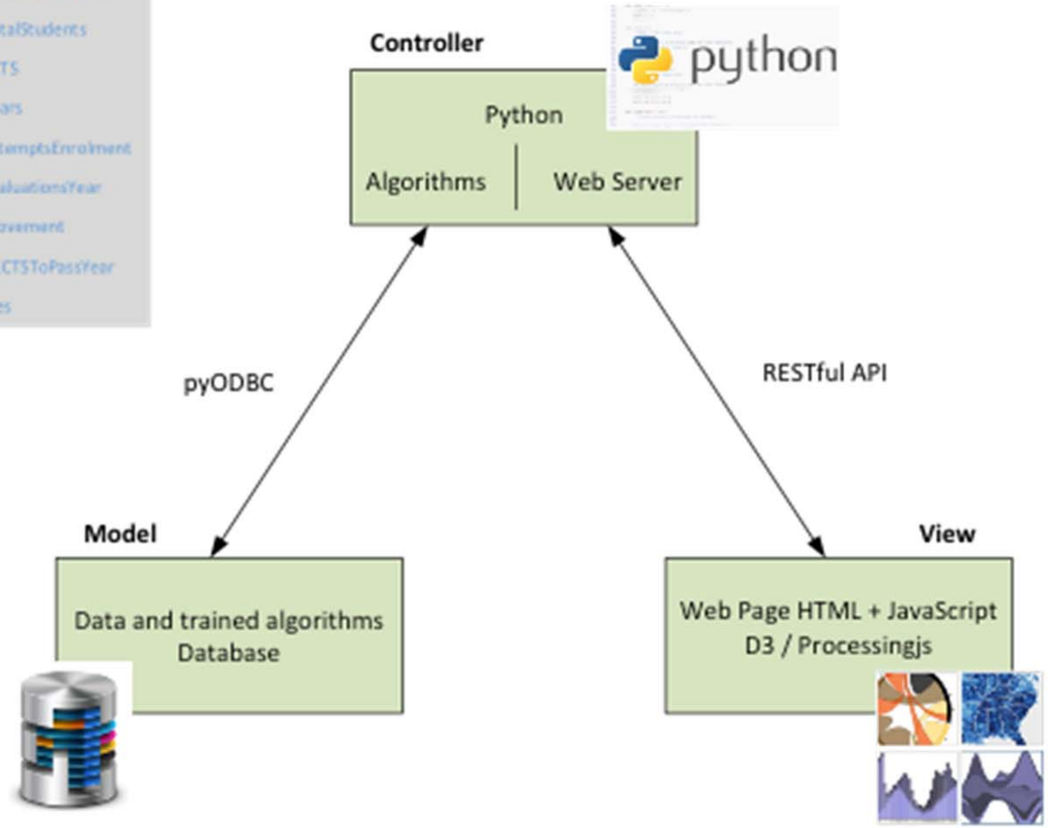
Attribute	Type of Variable
Student ID	Natural Number
YearsToFinishDegree	Natural Number
Mobility	Factor (No, Erasmus, Other)
Subject1NumberECTS	Real Number
Subject1Year	Natural Number
Subject1Semester	Natural Number
Subject1KnowledgeArea	Factor (Area 1, Area 2, ..., Area M)
Subject1Language	Factor (Country Language, English, Other)
Subject1Methodology	Factor (Theoretical, Laboratory, Theor. + Lab.)
Subject1Nature	Factor (Mandatory, Elective, BsC Thesis, Internship)
Subject1WeekHours	Real Number
Subject1Score	Real Number
Subject1NumberAttemps	Natural Number
Subject1AverageScore	Real Number
Subject1FailureRate	Real Number
Subject1AverageScoreLastYear	Real Number
Subject1FailureRateLastYear	Real Number
---	---
SubjectMNumberECTS	Real Number
SubjectMYear	Natural Number
SubjectMSemester	Natural Number
SubjectMKnowledgeArea	Factor (Area 1, Area 2, ..., Area N)
SubjectMLanguage	Factor (Country Language, English, Other)
SubjectMMethodology	Factor (Theoretical, Laboratory, Theor. + Lab.)
SubjectMNature	Factor (Mandatory, Elective, BsC Thesis, Internship)
Subject1WeekHours	Real Number
SubjectMScore	Real Number
SubjectMNumberAttemps	Natural Number
SubjectMAverageScore	Real Number
SubjectMFailureRate	Real Number
SubjectMAverageScoreLastYear	Real Number
SubjectMFailureRateLastYear	Real Number

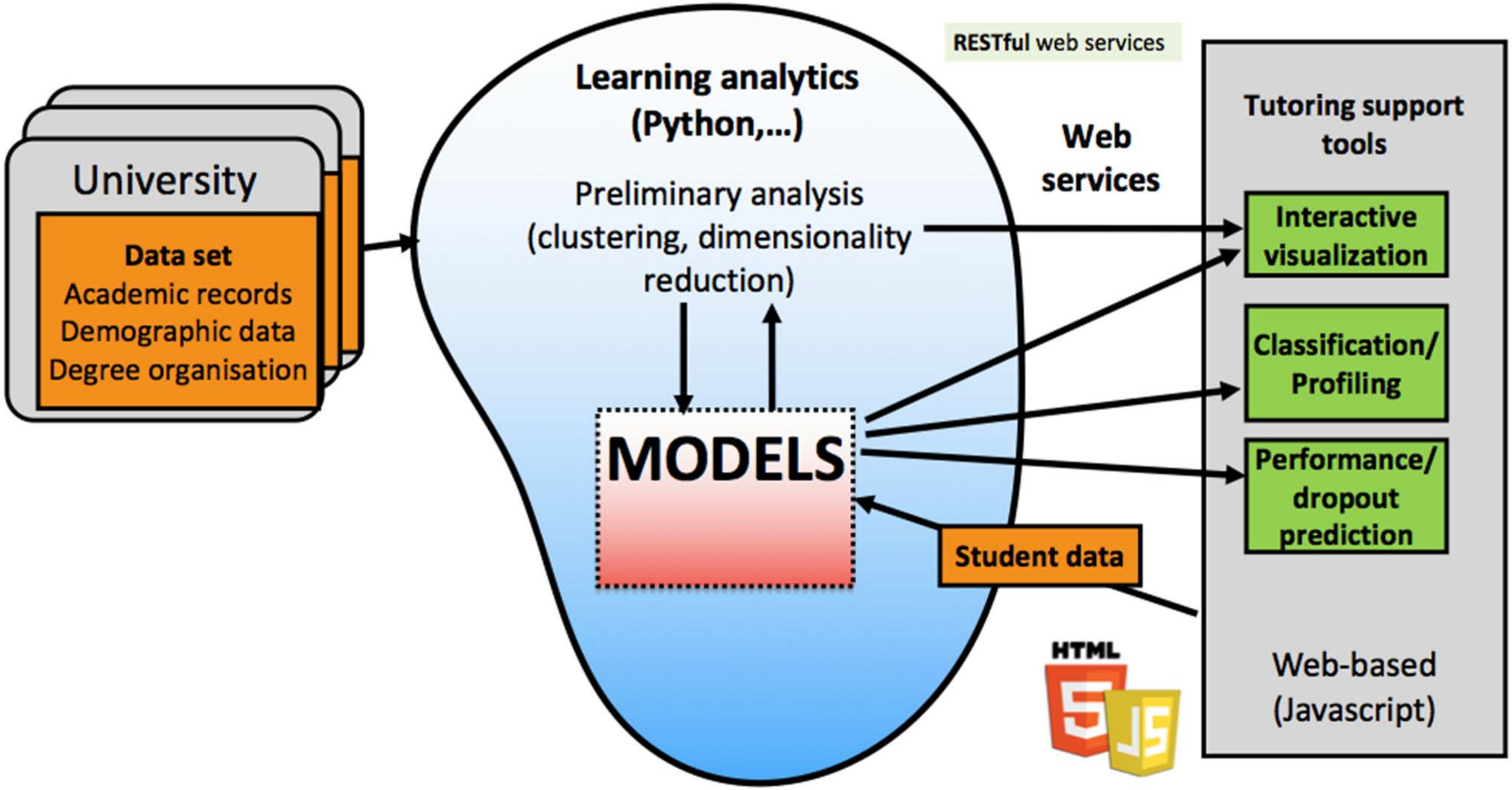
Student Profile for Enhancing Engineering Tutoring

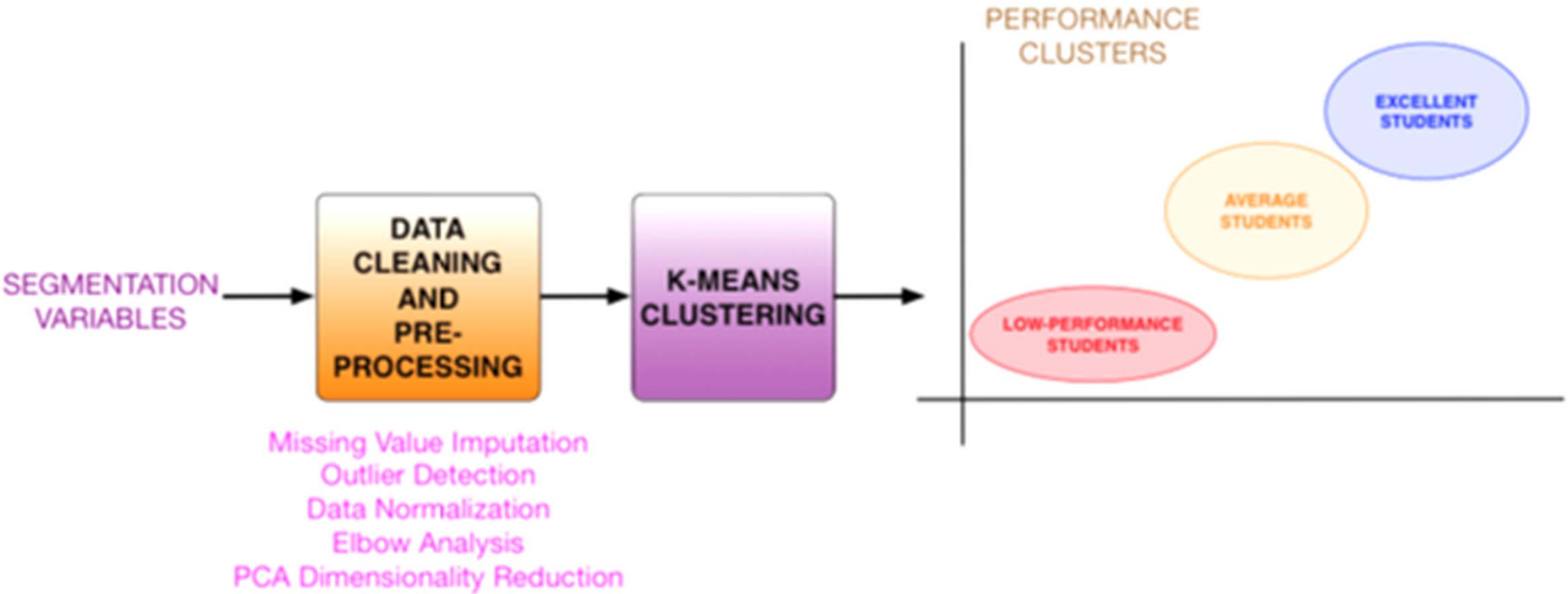
Erasmus+

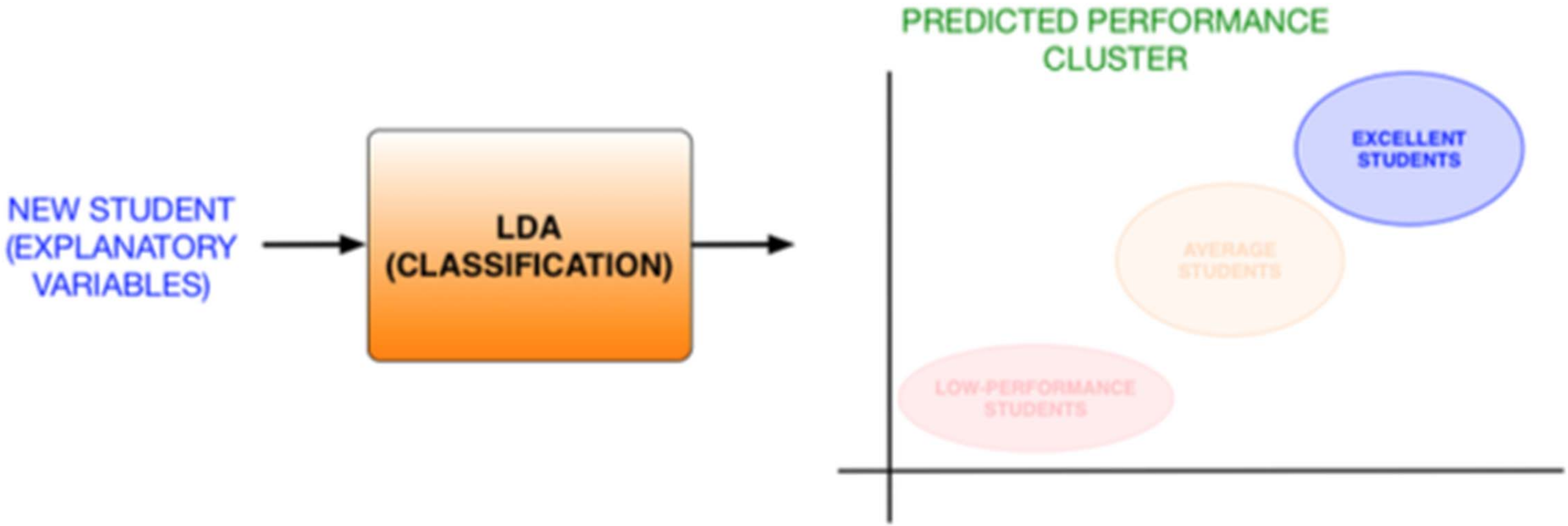



Students
StudentID (PK)
YearOfBirth
PlaceOfBirth
Sex
ResidenceCity
ResidenceCityDuringStudies
AccessToStudiesAge
Nationality
PreviousStudies
PreviousStudiesCenter
FatherEducationLevel
MotherEducationLevel
AdmissionScore
Students_Subjects
SubjectID (PK, FK)
StudentID (PK, FK)
Score
Mobility
YearPassed
NumberAttempts
Subjects_Year
Year (PK)
SubjectID (PK, FK)
AverageScoreYear
FailureRateYear
Subjects
SubjectID (PK)
Degree (FK)
SubjectName
NumberECTS
DegreeYear
Semester
KnowledgeArea
Methodology
Nature
Language
AverageScore
FailureRate
Degrees
Degree (PK)
Institution
DegreeNature
NumberStudentsFirstYear
NumberTotalStudents
NumberECTS
NumberYears
NumberAttemptsEnrolment
NumberEvaluationsYear
ScoreImprovement
MinimumECTSToPassYear
Especialities


Controller
python
Python
Algorithms
Web Server
pyODBC
RESTful API
Model
Data and trained algorithms
Database
View
Web Page HTML + JavaScript
D3 / Processingjs

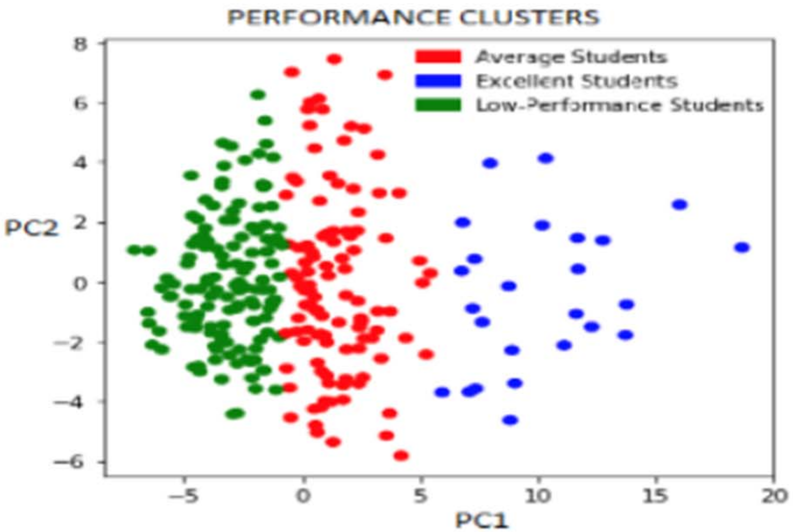

University
Data set
Academic records
Demographic data
Degree organisation
Learning analytics
(Python,...)
Preliminary analysis
(clustering, dimensionality
reduction)
MODELS
RESTful web services
Web
services
Student data
Tutoring support
tools
Interactive
visualization
Classification/
Profiling
Performance/
dropout
prediction
Web-based
(Javascript)
HTML

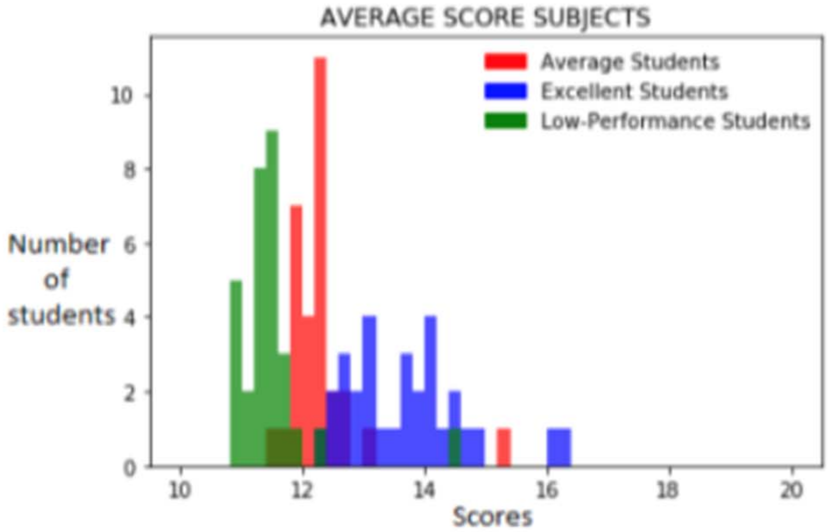

SEGMENTATION VARIABLES
DATA CLEANING AND PRE-PROCESSING
K-MEANS CLUSTERING
PERFORMANCE CLUSTERS
EXCELLENT STUDENTS
AVERAGE STUDENTS
LOW-PERFORMANCE STUDENTS
Missing Value Imputation
Outlier Detection
Data Normalization
Elbow Analysis
PCA Dimensionality Reduction

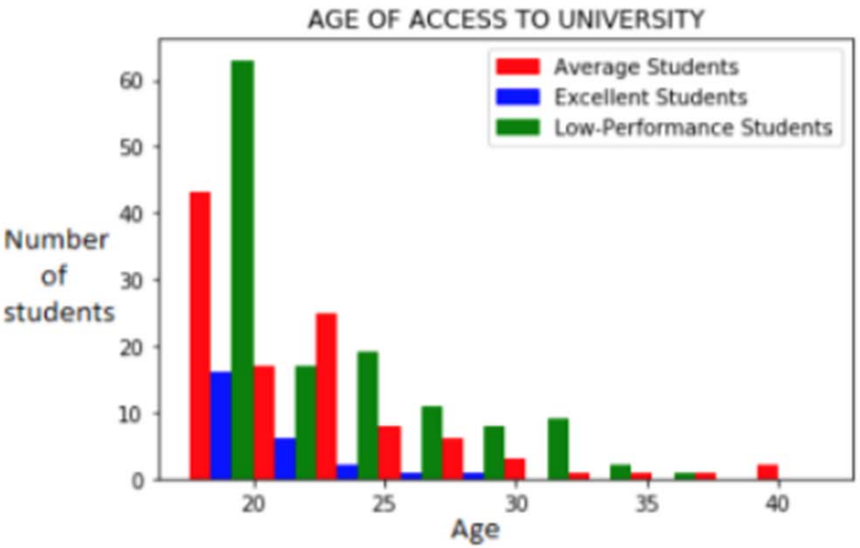

PREDICTED PERFORMANCE CLUSTER
NEW STUDENT (EXPLANATORY VARIABLES)
LDA (CLASSIFICATION)
EXCELLENT STUDENTS
AVERAGE STUDENTS
LOW-PERFORMANCE STUDENTS

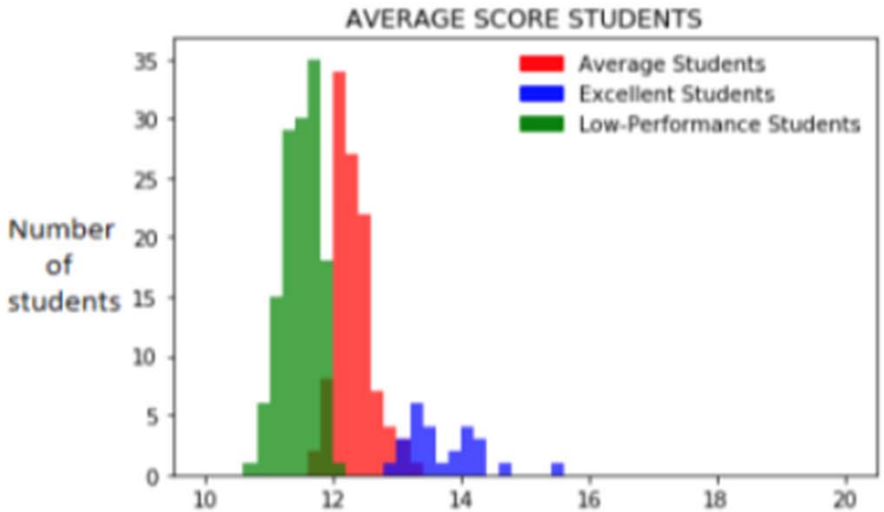
Student Profile for Enhancing Engineering Tutoring

Erasmus+



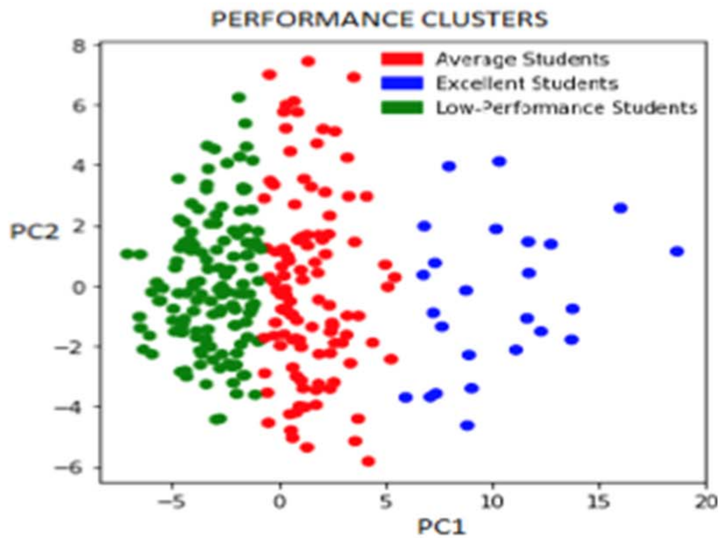


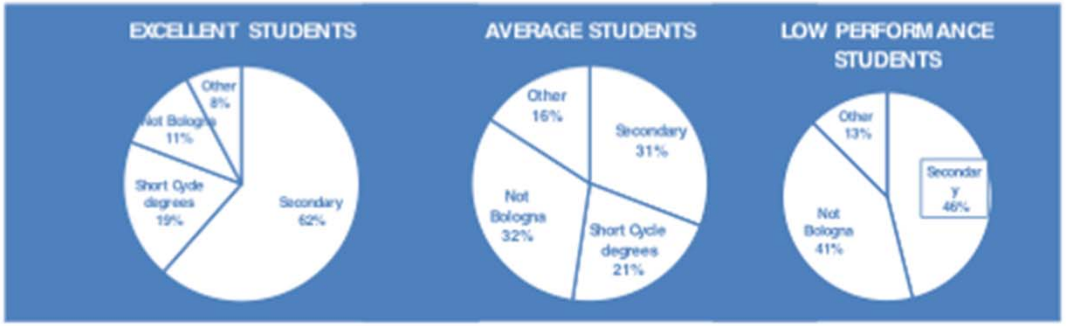


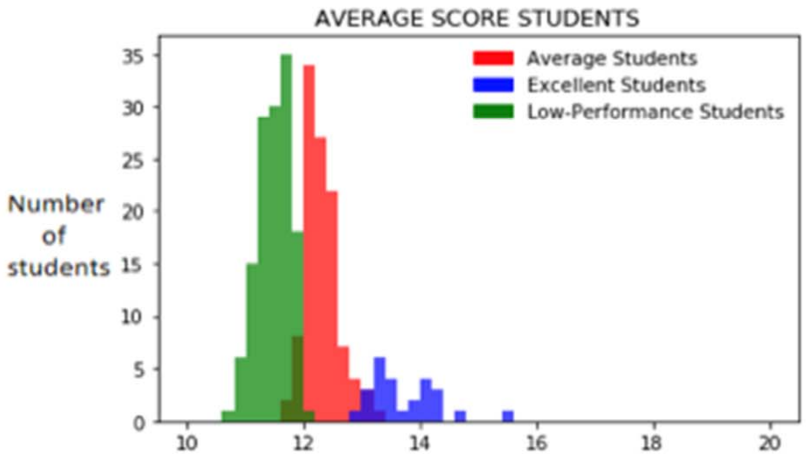


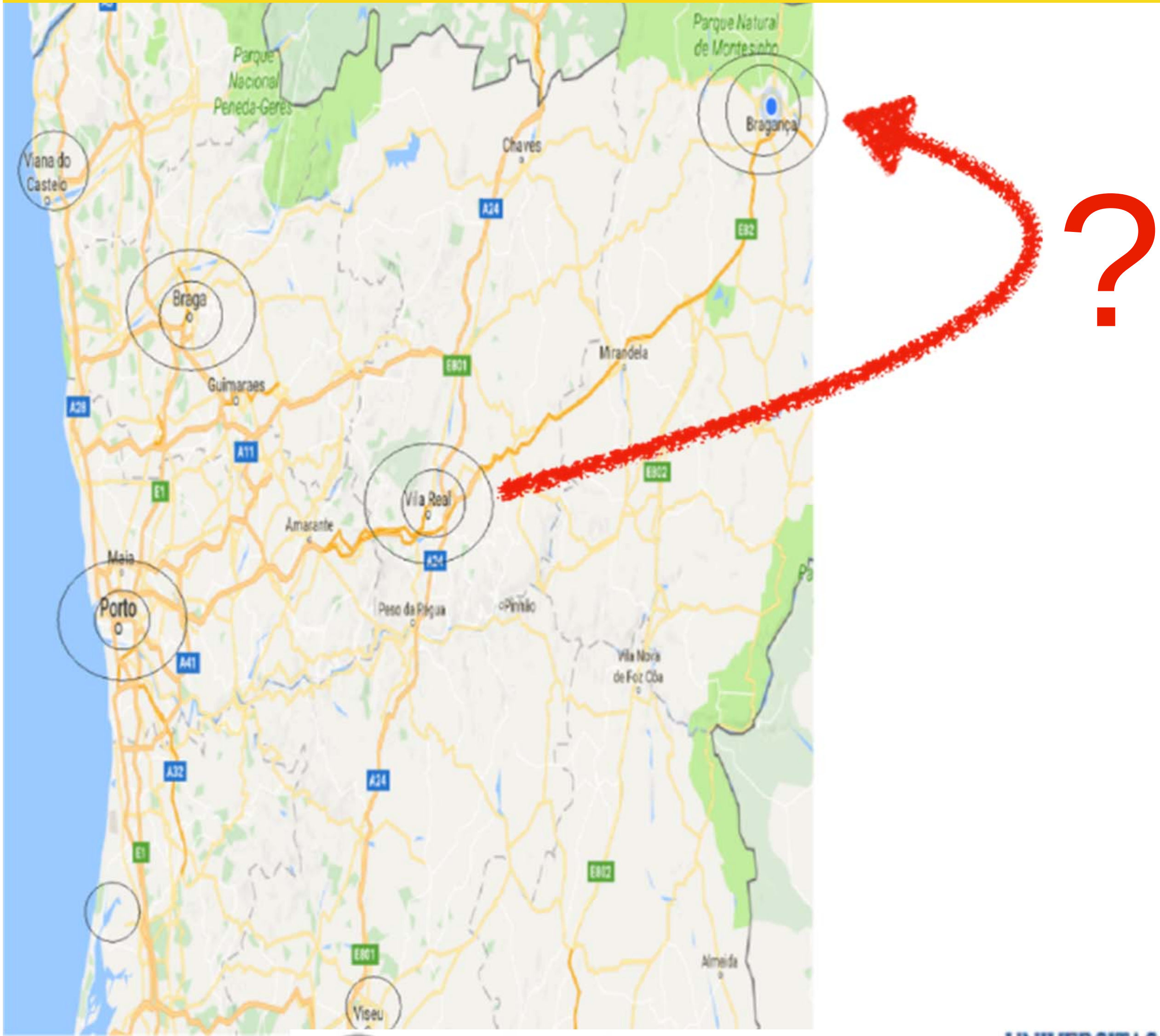
[IO#1.pdf]

Clustering Algorithms





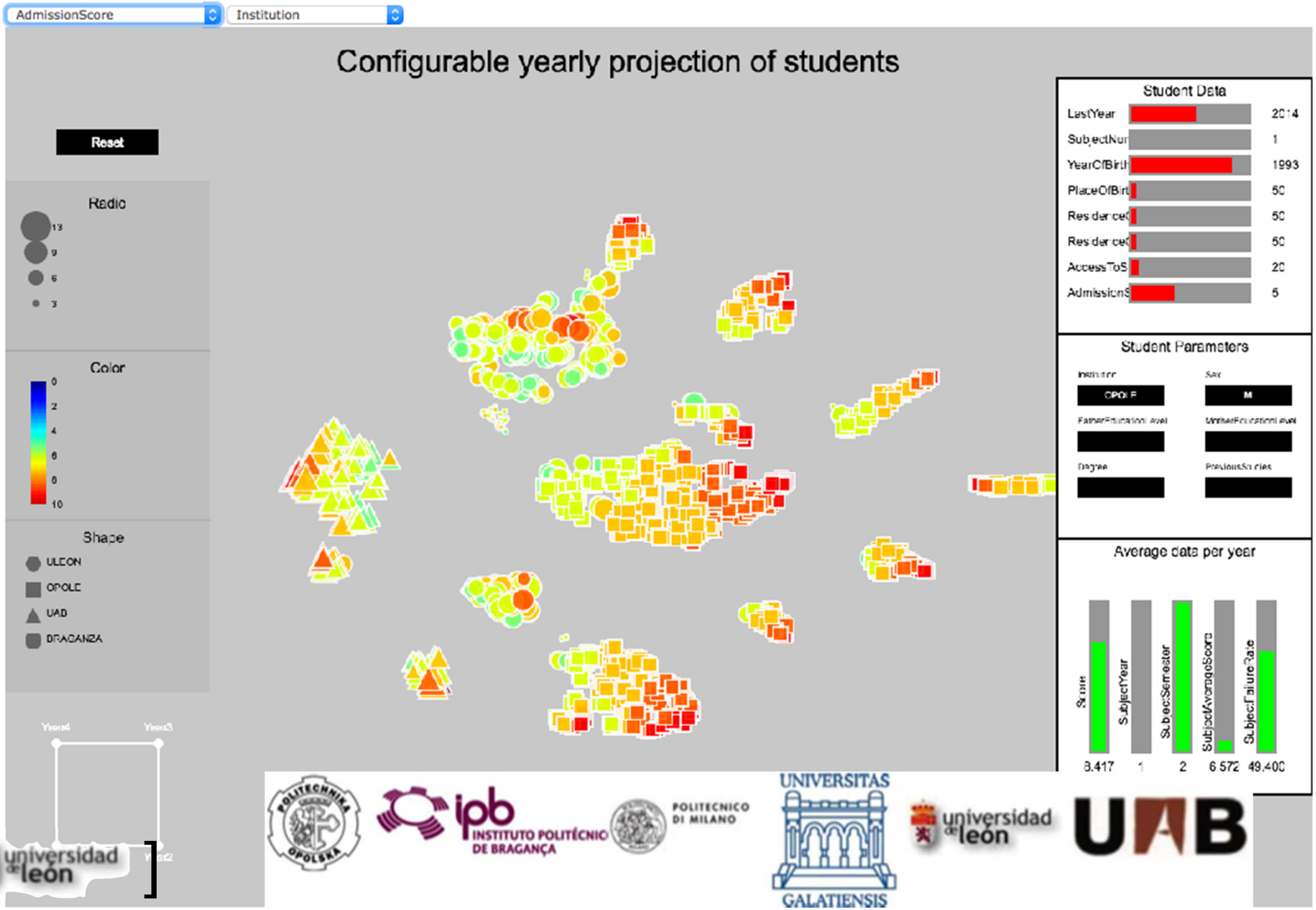


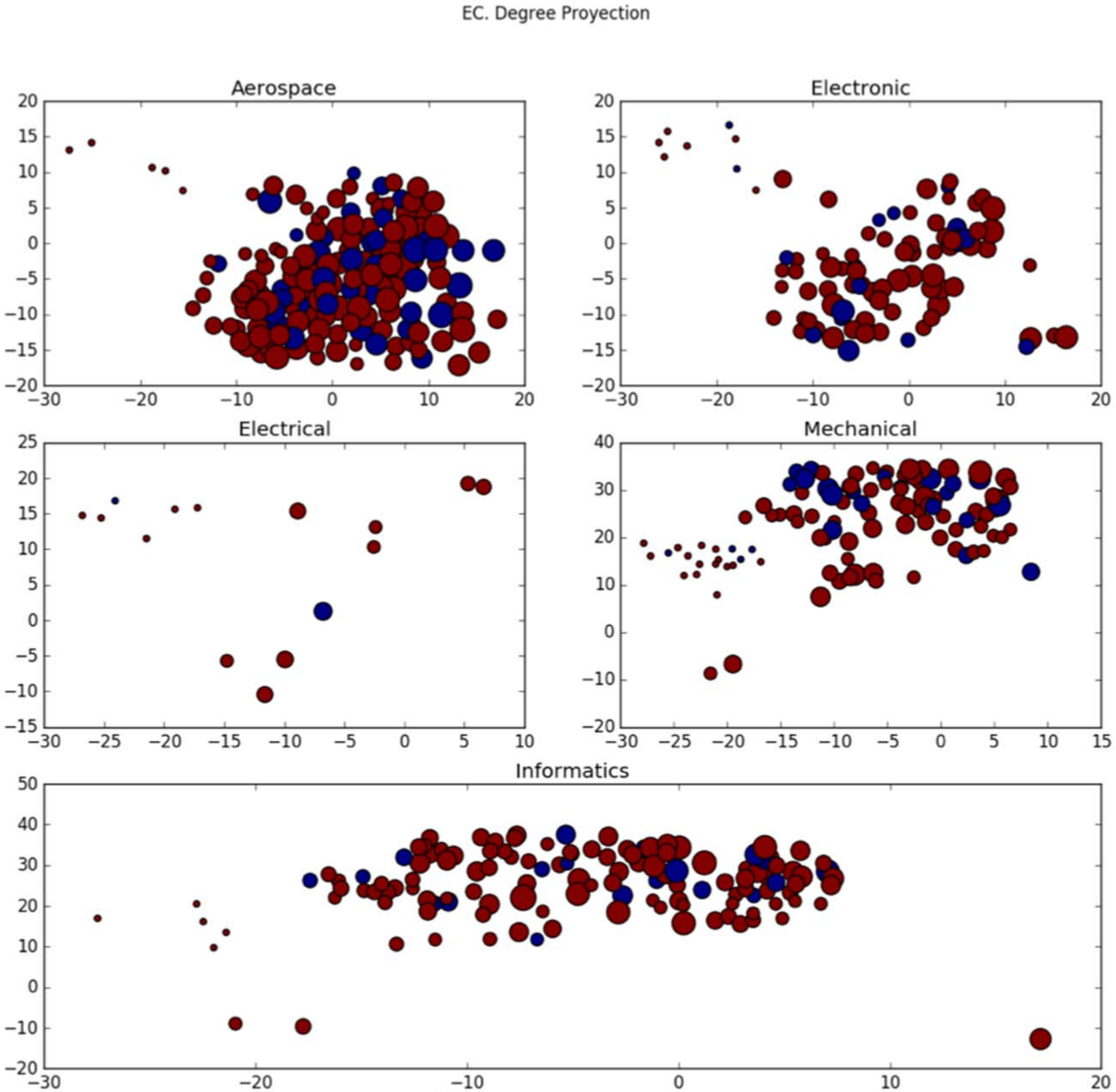

Parque Nacional Peneda-Gerês
Parque Natural de Montesinho
Viana do Castelo
Braga
Guimaraes
Chaves
Bragança
Mirandela
Vila Real
Amarante
Maia
Porto
Peso da Régua
Pinhão
Vila Nova de Foz Côa
Almeida
Viseu
?

[IO#1.pdf]

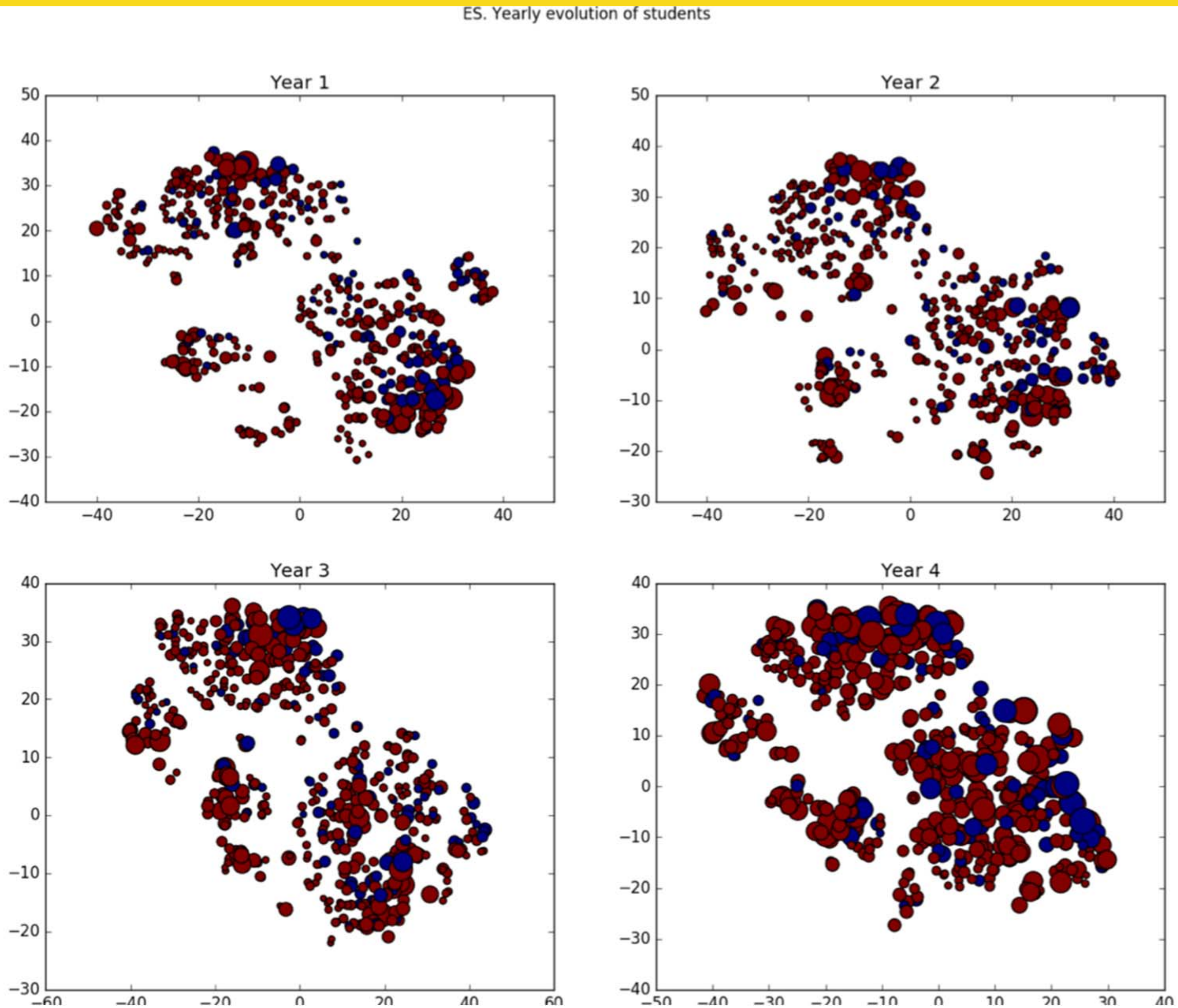
Student Profile for Enhancing Engineering Tutoring

Erasmus+





Data projection through Manifold Learning

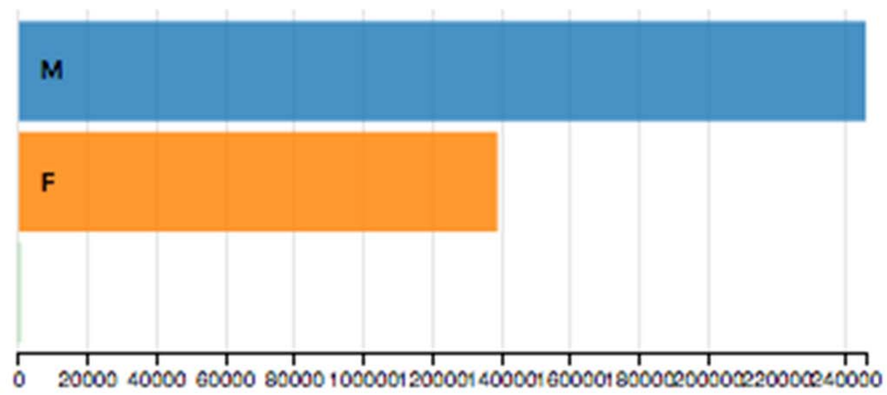


Data projection through Manifold Learning

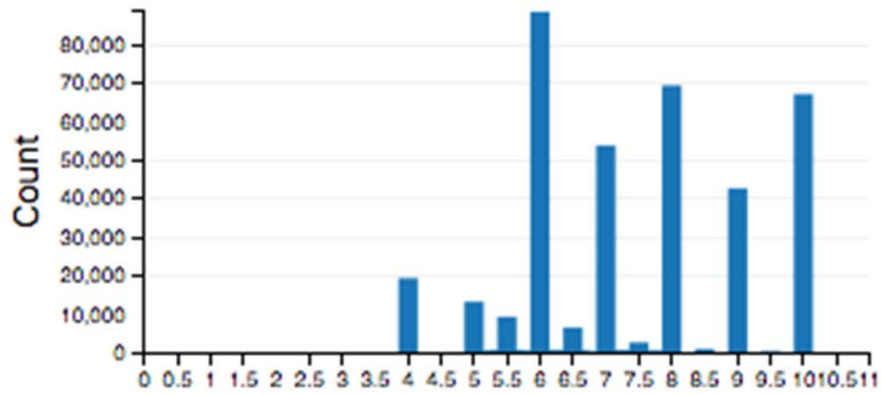
Student Profile for Enhancing Engineering Tutoring

Erasmus+

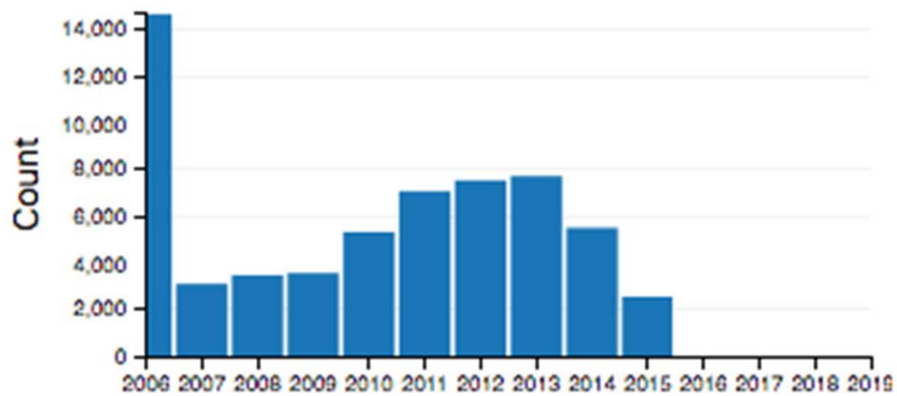
By Sex



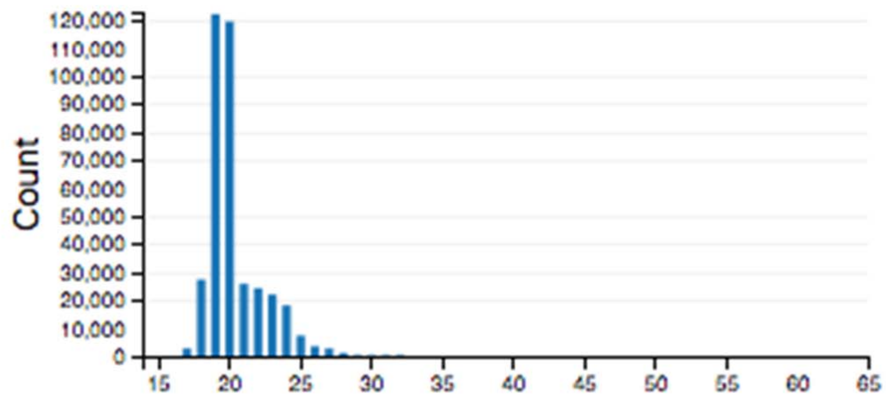
By Score



By Last Year of Studies



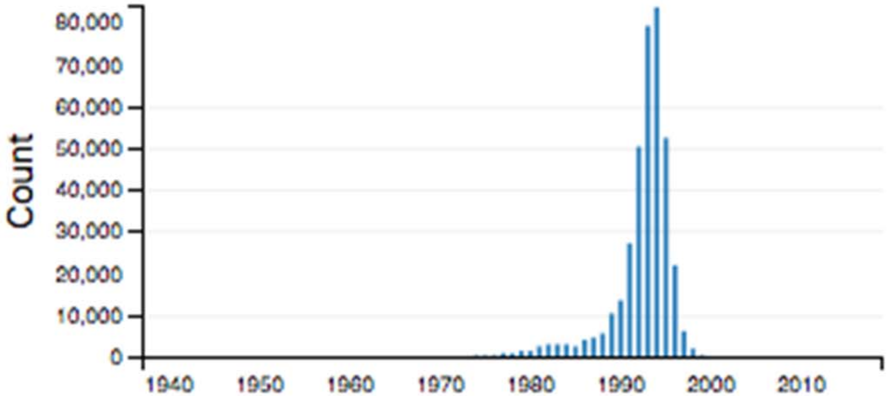
By Access to studies Age



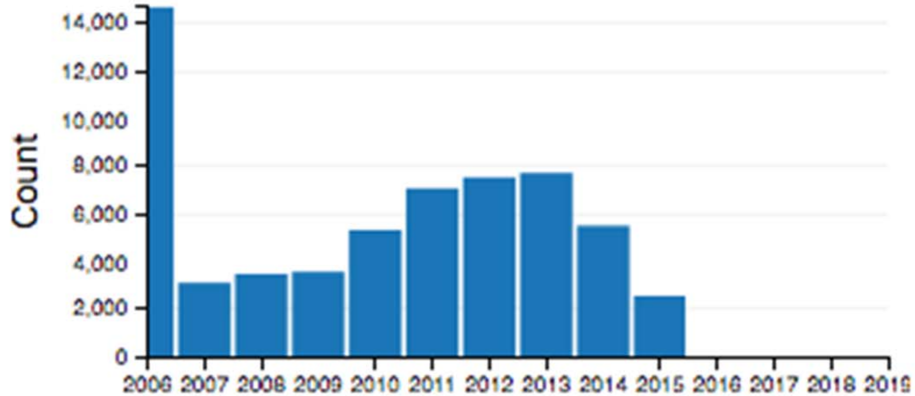
SPEET Student Profile for Enhancing Engineering Tutoring

Erasmus+

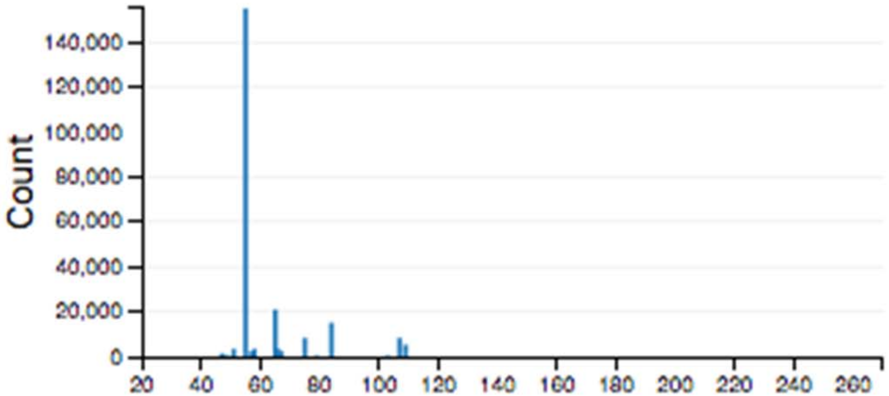
By Year of Birth



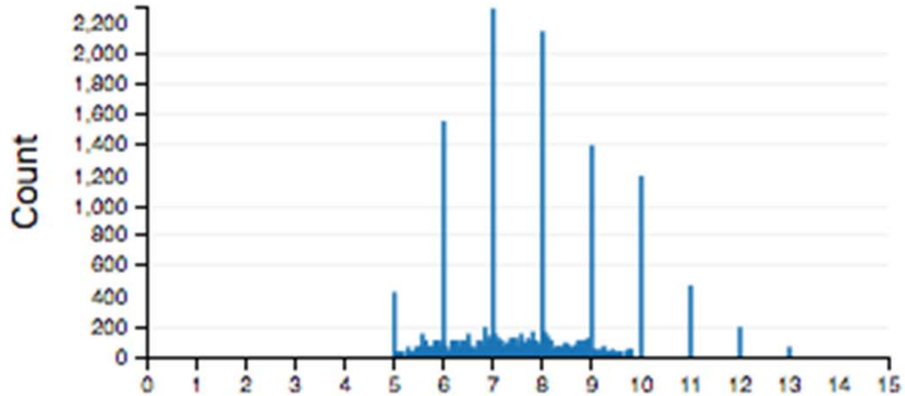
By Last Year of Studies



By Regional gross domestic product per capita



By Admission Score



Implementation

Significant Explanatory Covariates (FIXED EFFECTS)

- Sex
- Nationality
- Access to Studies Year
- Access to Studies Age
- Change of faculty
- Weighted Average Evaluations in the 1° semester of the 1° year
- Number of Subjects Passed in the 1° semester of the 1° year
- Average number of attempts per exam in the 1° semester of the 1° year

RANDOM EFFECT only on the intercept

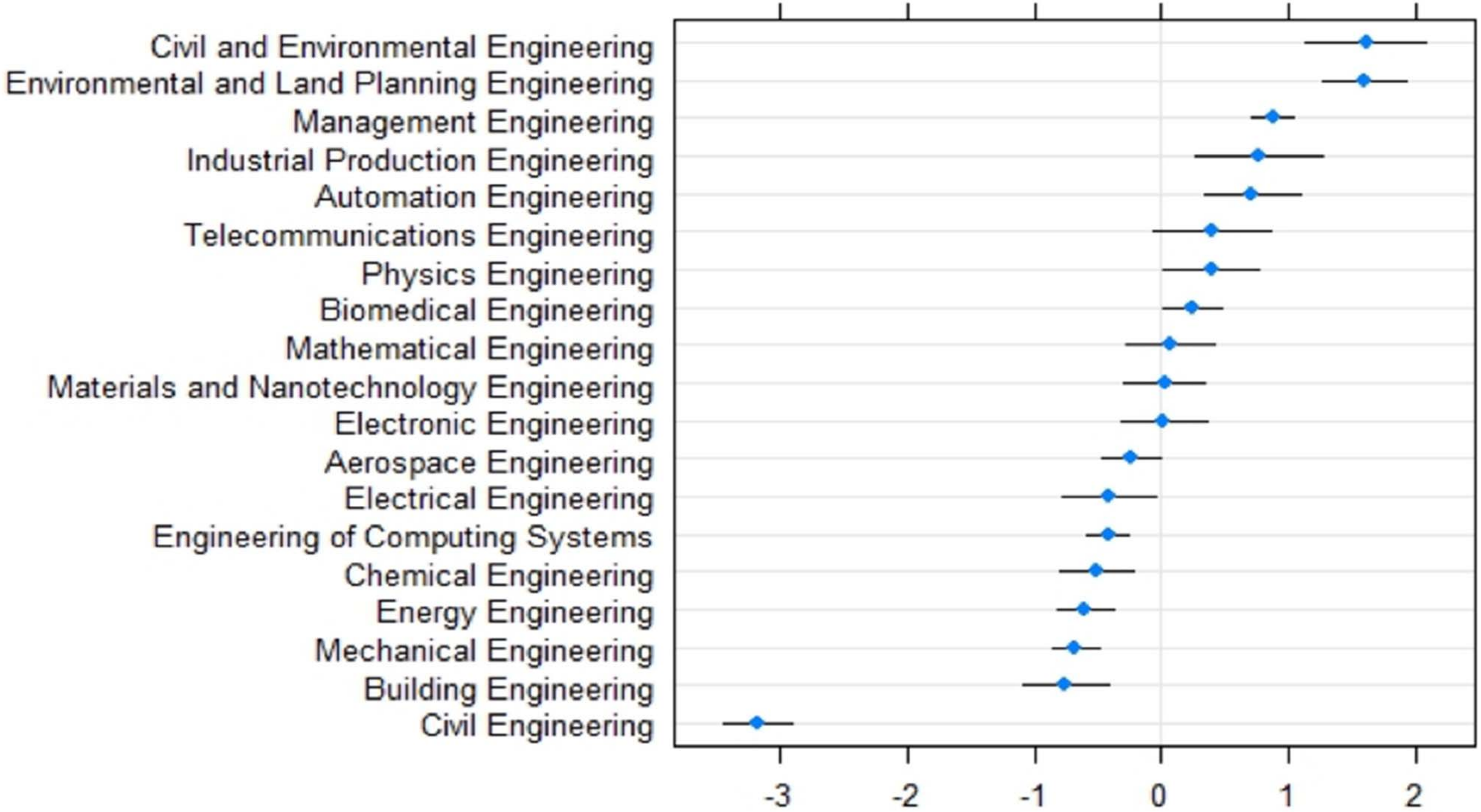
TOOLS

www.r-statistics.com

```
>> package(lme4)
>> glmer( y ~ ( 1 | DegreeNature ) + Sex
     + Nationality + … )
```

Student Profile for Enhancing Engineering Tutoring

Erasmus+



Fixed Effects

$\boldsymbol{\beta} = [\beta_0 \ \beta_1 \dots \ \beta_{p-1}]^T$

Variable	Estimate	P-value
(Intercept)	-2.322716	2.47e-05
Sex(male)	-0.292086	0.000447
Nationality(other)	-0.423296	0.020026
AccessToStudiesAge	-0.054090	0.029240
AccessToStudiesYear(2010)	-0.022170	0.801236
AccessToStudiesYear(2011)	-0.344638	8.32e-05
AccessToStudiesYear(2012)	-0.844056	< 2e-16
WeightedAverageEvaluations_11	0.060766	< 2e-16
AverageNumbAttemptsPerExam_11	0.028752	0.562176
NumbSubjectsPassed_11	1.709591	< 2e-16
Change(yes)	-0.373339	0.011962

data:
global

research
done by:

[POLITECNICO MILANO]

Student Profile for Enhancing Engineering Tutoring

INSTITUTION	DEGREE NAME	Students	Accuracy Level	1st Course Info	3rd Course Contribution
ULEON	Aerospace Engineering	166			
	Electronics Industrial Engineering	88			
	Mechanics Engineering	48			
	Computer Engineering	107			
UAB	Computer Engineering	197			
	Telecomunications Systems Engineering	25			
	Telecomunications Electronics Engineering	28			
	Chemical Engineering	65			
IPB	Mechanics Engineering	266			
	Civil Engineering	346			
	Electrotechnics Engineering	126			
	Computer Engineering	236			
	Computer Electrotechnics Engineering	67			
	Chemical Engineering	83			
GAL	Automation and Applied Informatics	63			
	Computer Science	17			
POL	Architecture	180			
	Civil Engineering	548			
	Automatic Control	50			

INSTITUTION	DEGREE NAME	Students	Silhouette value	Clustering Quality	Score Students Separation
ULEON	**Aerospace Engineering**	**166**	**0,1**		
	Electronics Industrial Engineering	**88**	**0,12**		
	Mechanics Engineering	**48**	**0,18**		
	Computer Engineering	**107**	**0,13**		
UAB	**Computer Engineering**	**197**	**0,15**		
	Telecomunications Systems Engineering	**25**	**0,27**		
	Telecomunications Electronics Engineering	**28**	**0,17**		
	Chemical Engineering	**65**	**0,3**		
IPB	**Mechanics Engineering**	**266**	**0,08**		
	Civil Engineering	**346**	**0,05**		
	Electrotechnics Engineering	**126**	**0,09**		
	Computer Engineering	**236**	**0,13**		
	Computer Electrotechnics Engineering	**67**	**0,08**		
	Chemical Engineering	**83**	**0,11**		
GAL	**Automation and Applied Informatics**	**17**	**0,29**		
	Computer Science	**63**	**0,17**		
POL	**Architecture**	**30**	**0,17**		
	Civil Engineering	**62**	**0,19**		
	Automatic Control	**27**	**0,14**		

GALATI

Automation and Applied Informatics

Input Data	No Categorical variables	Categorical Variables	Training Time
Categorical		6 %	0.4 s
1st Course	99 %	99 %	0.4 s
2nd Course	99 %	99 %	0.4 s
3rd Course			

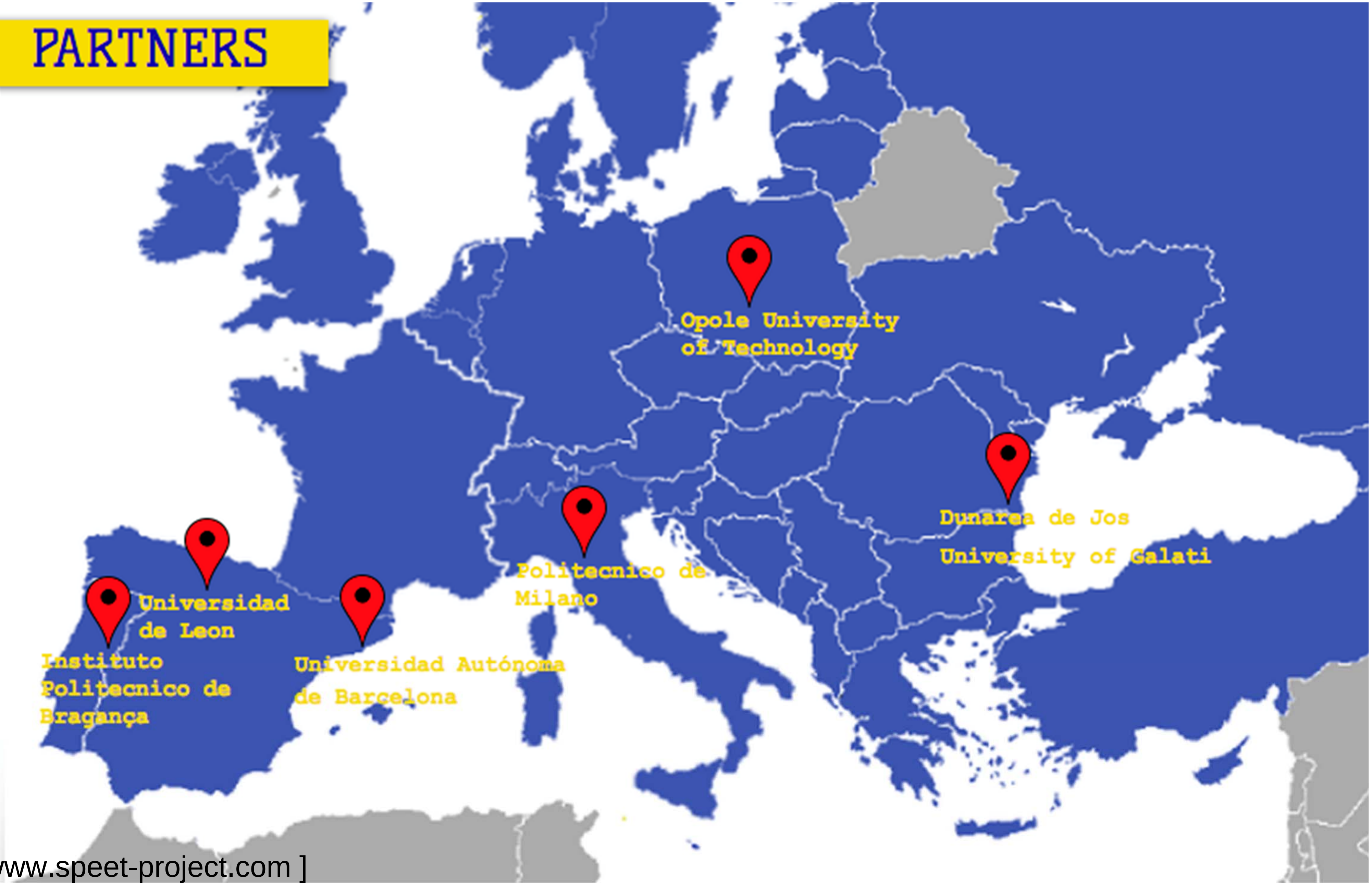
Computer Science

Input Data	No Categorical variables	Categorical Variables	Training Time
Categorical		69 %	0.4 s
1st Course	90 %	88 %	0.5 s
2nd Course	96 %	93 %	0.5 s
3rd Course			

Student Profile for Enhancing Engineering Tutoring

Erasmus+

PARTNERS



[www.speet-project.com]