

GCE-THESIS: An Information System for Students

EINIS - Prof. M. Murasziewicz





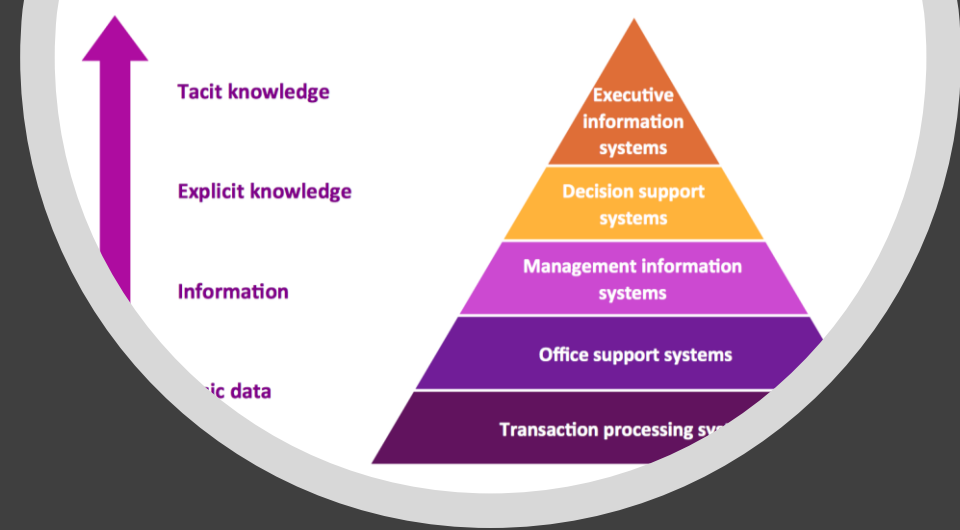
TÉCNICO
LISBOA



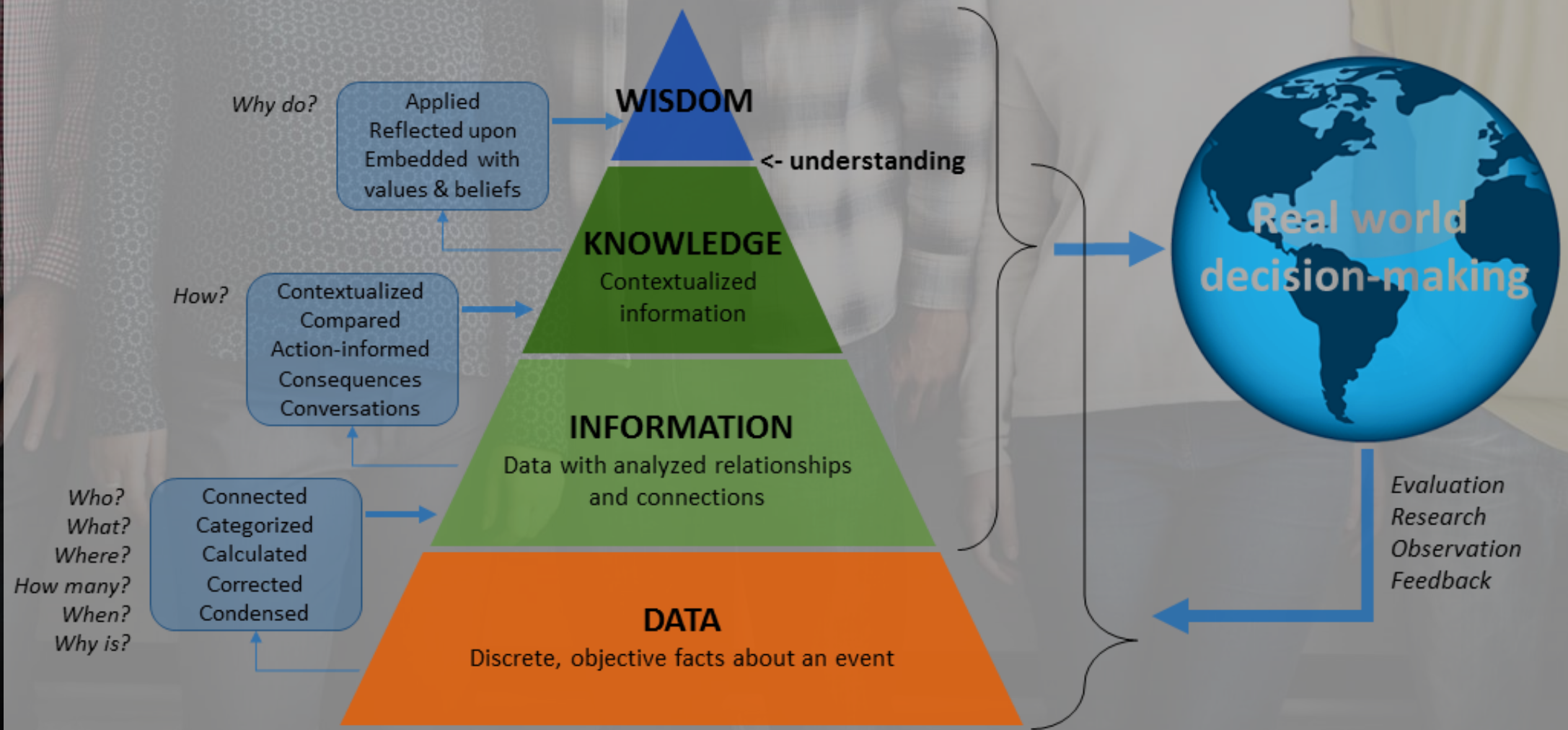
GRUPO DE CONTACTO
COM EMPRESAS

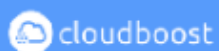
The Problem

- Storing information
- Answering questions.
- Knowledge discovery



Thesis Proposed Identifier	Title	Guidance	Number of Applications	state	
8117	Dissertation: Large scale membership and consistency	Miguel Matos (50%) - Teacher Rodrigo Seromenho Miragaia Rodrigues (50%) - Teacher	1	Assigned	Apply Details
9723	PROJECT: Identification of fixed expressions and popular sayings in text	Jorge Manuel Evangelista Baptista jbaptis@ualg.pt (50%) - Nuno João Neves Mamede (50%) - Teacher	1	Assigned	Apply Details
9778	Project - Data processing in sensor networks in urban transport.	Pedro Manuel Moreira Vaz Antunes de Sousa (100%) - Teacher	1	Not assigned	Apply Details
9691	Dissertation: Custom Content Placement for Digital Games	Carlos António Roque Martinho (100%) - Teacher	1	Assigned	Apply Details
7690	The Cunning Thralls: tactically smart NPCs in Conan Exiles.	João Miguel De Sousa de Assis Dias (50%) - Teacher Pedro Alexandre Simões dos Santos (50%) - Lecturer	1	Assigned	Apply Details
7630	Project: Game for Stimulation of Social Cognition	Rui Filipe Fernandes Prada (70%) - Lecturer Ana Ribeiro Moreira ana.moreira@oninetspeed.pt (15%) - Daniel Neto danielcarvalhoneto@gmail.com (15%) - Susana Vieira (60%) - Teacher	4	Assigned	Apply Details
7788	Design: Evolving flocking behavior for robots operating in natural unstructured environments for long durations	Anders Christensen Anders.Christensen@iscta.pt (20%) - Sancho Oliveira Sancho.Oliveira@iscte.pt (20%) -	1	Assigned	Apply Details
7707	Dissertation: Visual Language to Orchestrate Containers on the Cloud	Rui Maranhão (100%) - Teacher	1	Assigned	Apply Details
8076	Dissertation: Change Detection on Frequent Patterns	Claudia Martins Antunes (100%) - Teacher	1	Not assigned	Apply Details
8072	[Dissertation] Creating Recipes using Machine Learning and Computational Creativity	Francisco António Chaves Saraiva de Melo (50%) - Teacher Helena Sofia Andrade Nunes Pereira Pinto (50%) - Teacher	1	Assigned	Apply Details
7714	[Dissertation] Research in Advanced Techniques of Semi Automatic Modeling and Production of Responsive Web Applications	Alberto Manuel Rodrigues da Silva (100%) - Teacher	1	Assigned	Apply Details
7982	Project - Development of a mobile application for speech therapy	Alberto Abad (50%) - Teacher Anna Maria Pompili (50%) -	2	Assigned	Apply Details
7883	Dissertation: Personality-based storytelling through companion interaction	Carlos António Roque Martinho (100%) - Teacher	5	Assigned	Apply Details
8164	Discussion: ARrow: Navigation via AR on Smartphones	Joaquim Jorge (60%) - Teacher Marco Ferreira marco.ferreira@thalesgroup.com (40%) -	3	Not assigned	Apply Details
9188	Binge-Watching & Fandom. Exploring engagement and emotional states among TV fiction fans	Nuno Jardim Nunes (100%) - Teacher	1	Assigned	Apply Details
8132	Project: VRE for medicine - PRECISE	José Borbinha (100%) - Teacher	1	Assigned	Apply Details
8003	Dissertation: Implementation and Parallelization of a Calculation Algorithm for Dedicated Functions	José Carlos Alves Pereira Monteiro (50%) - Teacher Juan Acebron juan.acebron@gmail.com (50%) -	1	Not assigned	Apply Details
7975	Thesis: Global Illumination with GPU Vertex Connection and Merging	Joao Antonio Madeiras Pereira (80%) - Lecturer Vasco Costa vasco.costa@gmail.com (20%) -	1	Assigned	Apply Details





Writer



Edit



HOME ANNOUNCEMENTS TUTORIALS LEARN JAVASCRIPT ABOUT | SIGN UP



Still trying to find a decent catch-phrase
Jan 8 · 5 min read

<https://goo.gl/RTWy27>

The MEAN Stack: A Practical Example—GCE-Thesis

This article aims to give you very high-level insight into two topics: firstly, introducing GCE-Thesis, an open-source project based on the MEAN stack.



Login Register

GRUPO DE CONTACTO COM EMPRESAS

Login

G Waiting List

G FAQ

1.2.3-Beta

MEAN STACK



Mongo DB
(database system)



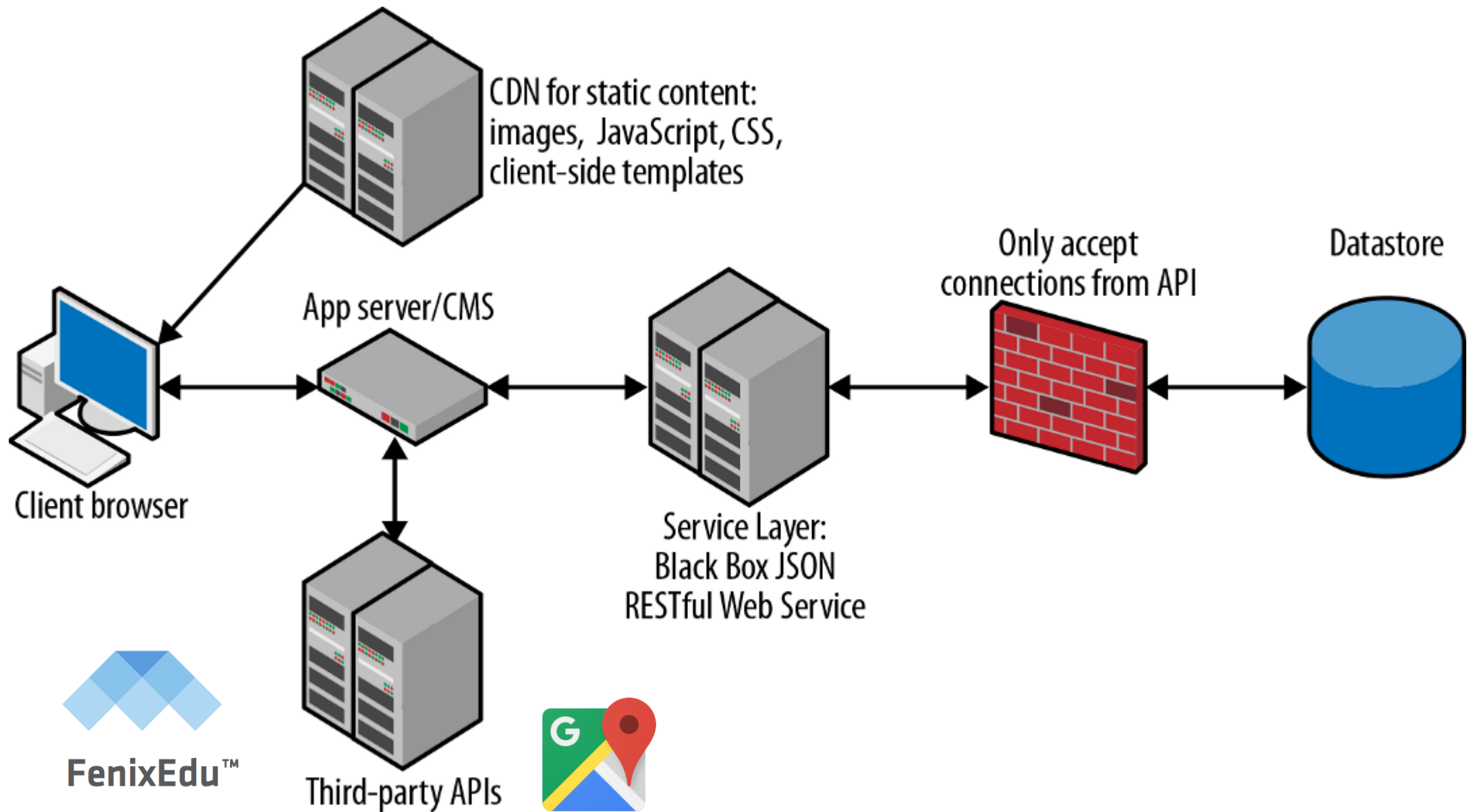
Express
(back-end web
framework)



Angular.js
(front-end
framework)



Node.js
(back-end runtime
environment)



GAUSSIAN NAIVE BAYES CLASSIFIER

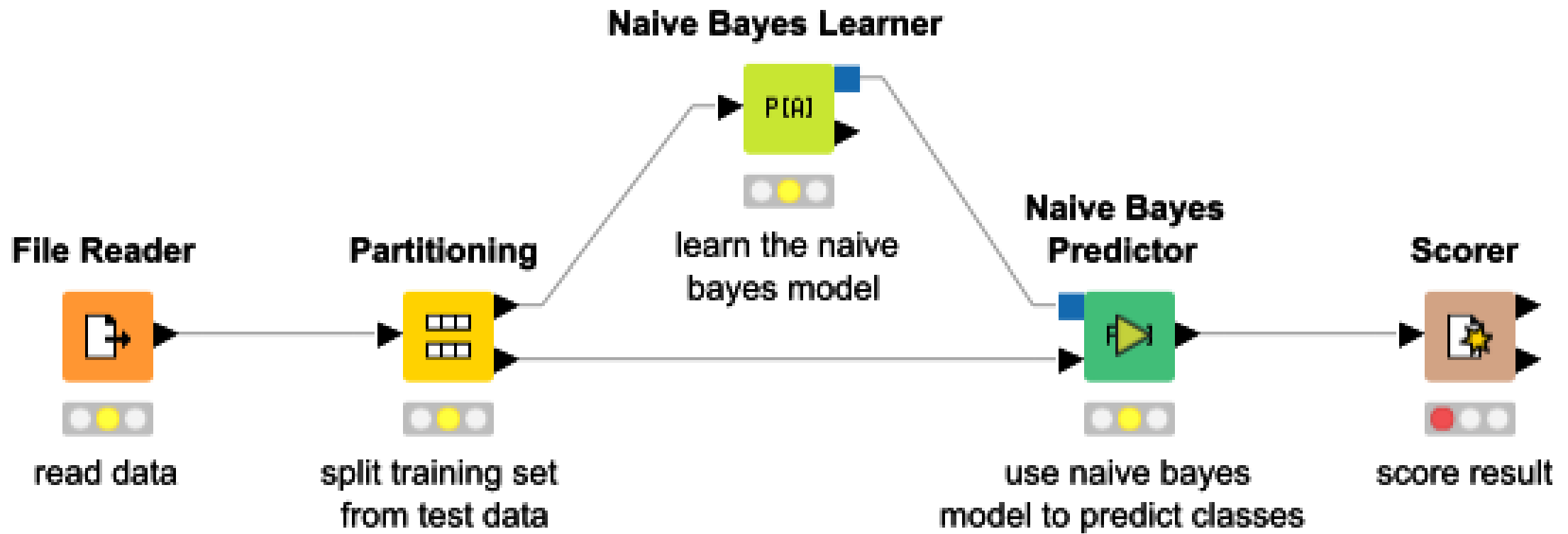
"Gaussian" because this is a normal distribution

This is our prior belief

$$P(\text{class} | \text{data}) = \frac{P(\text{data} | \text{class}) \times P(\text{class})}{P(\text{data})}$$

We don't calculate this in naive bayes classifiers

Chris Albon



Ex:

Formal Software Specification using B - Requisite Specification: Representing States and Invariants

['formal', 'softwar', 'specif', 'us', 'requisit', 'specif', 'repres', 'state', 'invari']

```
const ThesisSchema = mongoose.Schema({
  id: {
    type: String,
    required: true
  },
  title: {
    type: String,
    required: true
  },
  supervisors: {
    type: [String],
    required: true
  },
  vacancies: {
    type: String,
    required: true
  },
  status: {
    type: String,
    enum: STATUS,
    default: "Not assigned",
    required: true
  },
  location: {
    type: String,
    required: false
  },
  courses: {
    type: [String],
    required: true
  },
  observations: {
    type: String,
    required: false
  },
  objectives: {
    type: String,
    required: false
  },
  requirements: {
    type: String,
    required: false
  },
  areas: {
    type: [String],
    required: false
  },
  clicks: {
    type: Number,
    default: 1,
    required: false
  },
  type: {
    type: Number,
    enum: TYPE,
    default: 1,
    required: false
  },
  lastModified: {
    type: Date,
    required: false
  }
});
```



Manual classification of 427 theses

The Experience

What can vary?

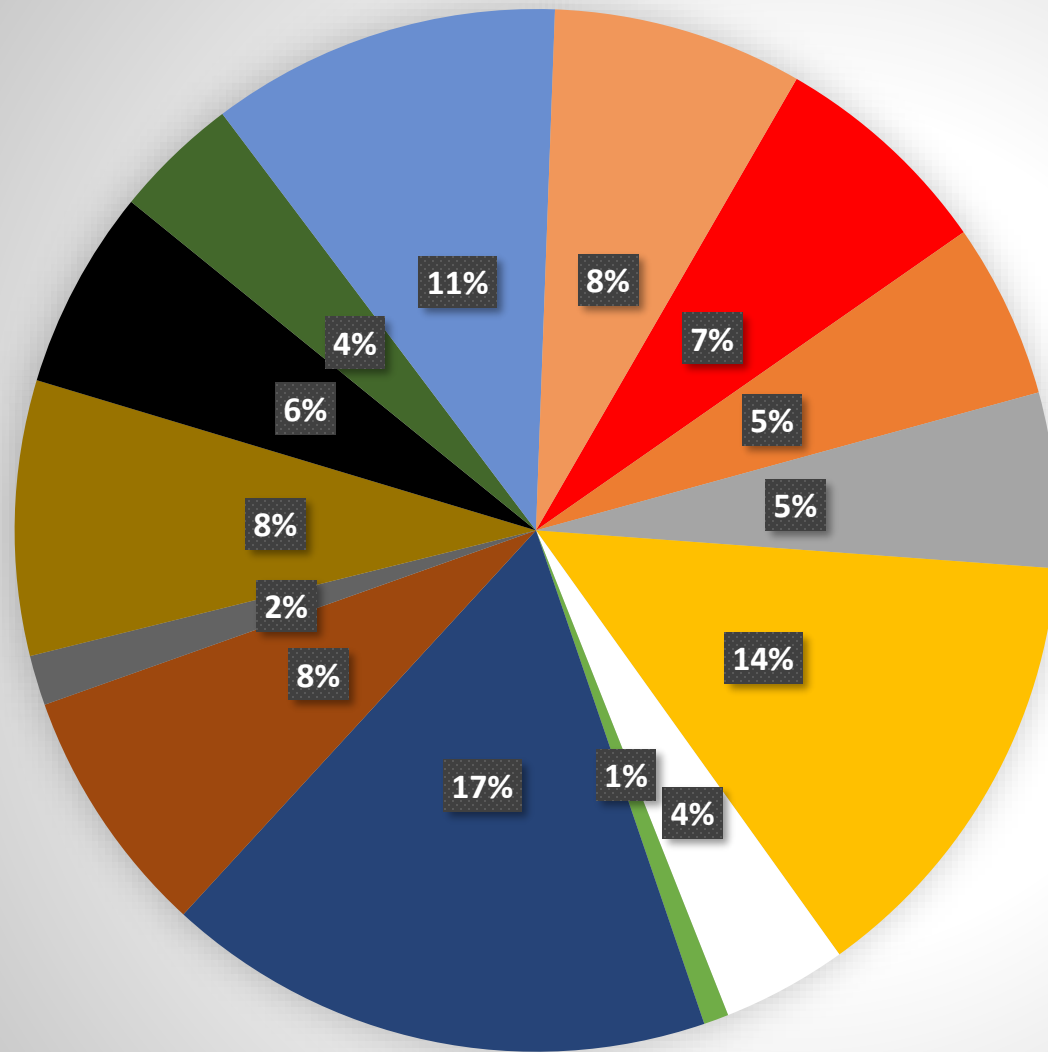
Input (theses pre-processing)

Classifier (training data)

Classification (classifier's input)

Distribution of Theses according to its official areas (manual classification)

133 Theses



- Algorithms and Applications
- Bioinformatics and Computational Biology
- Cyber-Security
- Software Engineering
- Data Science and Engineering
- It Service Management
- Interaction and Visualization
- Games
- Intelligent Robotics
- Cyberphysical Systems
- Information Systems
- Distributed Systems
- Enterprise Systems
- Intelligent Systems

Experience 1 -Base

Classifier trained with objectives and programs of each course related to each area (**15** different scientific areas).

Each area may have repeated courses.

Input of the classifier is the **title of the thesis**.

Results:

53/422=

12,6%

```
Belchior\Desktop>cd GCE-NEIIST
Belchior\Desktop\GCE-NEIIST>npm test
test C:\Users\Rafael Belchior\Desktop\GCE-NEIIST
reporter dot --timeout 10000 "server/tests/*.spec.js"
er initiated
.....
53 passing (689ms)
369 failing

1) Test 1: NB Classifier, all courses from all areas
   Test1:

   AssertionError: expected 'Information Systems' to equal 'Language a
+ expected - actual
-Information Systems
+Language and Information Technologies
at Context.<anonymous> (server\tests\test-classifier.spec.js:45:97)

Test 1: NB Classifier, all courses from all areas
Test2:

   AssertionError: expected 'Cyberphysical Systems' to equal 'Data Sci
ted - actual
al Systems
and Engineering
```

Problems?

Raw data is not uniform (PT/EN)

Courses (blue) which have multiple scientific areas in common

Needs manual corrections

Optimal classification criteria input

Manual classification may be wrong



Experience 2 - Input

Classifier trained with objectives and programs of each course related to each area (15 different scientific areas).

Each area may have repeated courses.

Input of the classifier is the title of the thesis (all in English).

Results:

64/422=

15,2% (20,63% increase)

```
C:\Users\Rafael Belchior\Desktop\GCE-NEIIST>npm test
> GCE@1.0.0 test C:\Users\Rafael Belchior\Desktop\GCE-NEIIST
> mocha --reporter dot --timeout 10000 "server/tests/*.spec.js"

BA logger initiated

.....
64 passing (602ms)
358 failing

1) Test 1: NB Classifier, all courses from all areas
   Test1:
     AssertionError: expected 'Information Systems' to equal 'Language and I
+ expected - actual
    Information Systems
    language and Information Technologies
    Context.<anonymous> (server\tests\test-classifier.spec.js:43:97)

    NB Classifier, all courses from all areas
     Error: expected 'It Systems Technology' to equal 'Data Science
- actual
    Technology
    and Engineering
    Context.<anonymous> (server\tests\test-classifier.spec.js:48:97)

    NB Classifier, all courses from all areas
     Error: expected 'Physical Systems' to equal 'Games'
```

Problems?

Raw data is not uniform (PT/EN)



Courses (blue) which have multiple scientific areas in common

Needs manual corrections

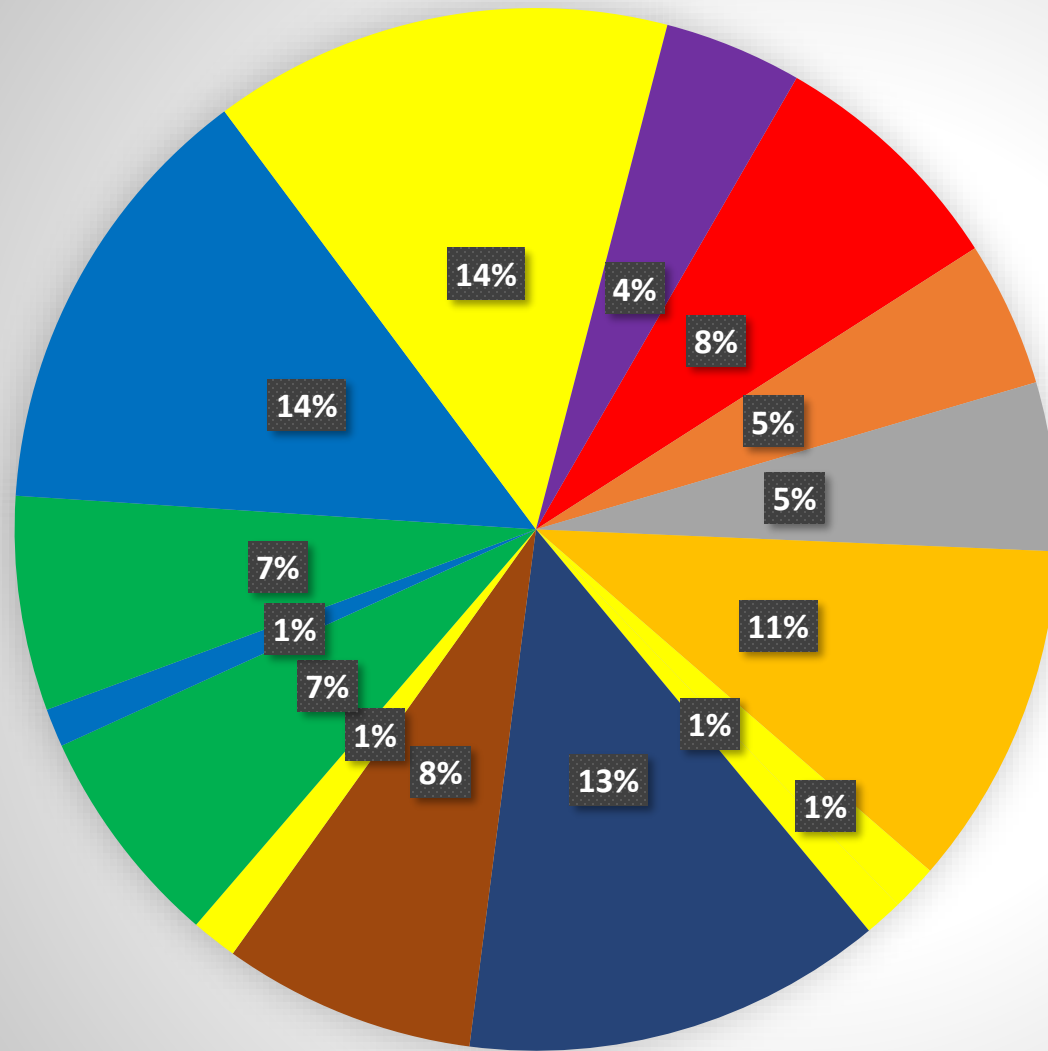
Optimal classification criteria input

Manual classification may be wrong



Distribution of Theses according to its official areas (manual classification)

421 Theses



- Algorithms and Applications
- Bioinformatics and Computational Biology
- Cyber-Security
- Software Engineering
- Data Science and Engineering
- It Service Management
- Interaction and Visualization
- Games
- Intelligent Robotics
- Cyberphysical Systems
- Information Systems
- Distributed Systems
- Enterprise Systems
- Intelligent Systems
- Language and Information Technologies

```
const AREAS = [  
  "Algorithms and Applications",  
  "Bioinformatics and Computational Biology",  
  "Cyber-Security",  
  "Software Engineering",  
  "Data Science and Engineering",  
  "It Service Management",  
  "Interaction and Visualization",  
  "Games",  
  "Intelligent Robotics",  
  "Cyberphysical Systems",  
  "Information Systems",  
  "Distributed Systems",  
  "Enterprise Systems",  
  "Intelligent Systems",  
  "Language and Information Technologies"  
];
```

15 areas => 12 areas

The problema of reducing more areas?

Courses in common between Games and
Interaction and Visualization removed.

Experience 3 - areas

Classifier trained with objectives and programs of each course related to each area (15 different scientific areas).

Each area may have repeated courses.

Input of the classifier is the title of the thesis (all in English).

Results:

84/422=

19,9% (30,9% increase)

```
84 passing (451ms)
```

```
338 failing
```

```
1) FILE_1__TEST_1:NB Classifier, First Classification
   Test1:
```

```
AssertionError: expected 'Language and Information Technologies' to equal '
+ expected - actual
```

```
-Language and Information Technologies
+Intelligent Systems
```

```
at Context.<anonymous> (server\tests\test-classifier.spec.js:63:108)
```

```
FILE_1__TEST_1:NB Classifier, First Classification
   Test2:
```

```
AssertionError: expected 'Enterprise and Information Systems' to equal 'Int
+ expected - actual
```

```
-Enterprise and Information Systems
+Intelligent Systems
```

```
at Context.<anonymous> (server\tests\test-classifier.spec.js:74:108)
```

```
FILE_1__TEST_1:NB Classifier, First Classification
```

```
AssertionError: expected 'Enterprise and Information Systems' to equal 'Gam
+ expected - actual
```

```
-Enterprise and Information Systems
```

```
at Context.<anonymous> (server\tests\test-classifier.spec.js:84:108)
```

```
FILE_1__TEST_1:NB Classifier, First Classification
```

Problems?

Raw data is not uniform (PT/EN)



Courses (blue) which have multiple scientific areas in common



Needs manual corrections

Optimal classification criteria input(title or title + supervisors or title + requirements, etc).

Manual classification may be wrong





What's next?

Student's clustering (based on their grades or courses)

Supervisor's clustering (based on the grades their students had on their theses)

Predicting a student's area of interest based on their courses (taking into account that one course may belong to different areas)

Predict how much time a certain specialization takes

Predict how many vacancies are needed for each course/specialization

Most popular supervisors



References:

EINIS slides - Prof. M. Muraszekiewicz

EDAMI slides – Prof. Marzena Kryszkiewicz

www.gce-neiist.org

<https://en.wikipedia.org/wiki/Classifier>

https://en.wikipedia.org/wiki/Information_system

<http://www.infogineering.net/data-information-knowledge.htm>



Thank you!
Pytanie?

Rafael Belchior
rafael.belchior@tecnico.ulisboa.pt