DSLAB
DATA SCIENCE LABORATORY

HIGH PERFORMANCE RESEARCH GROUP IN FOUNDATIONS AND APPLICATIONS OF DATA SCIENCE

## DATA SCIENCE LAB PRESENTATION

The Data Science Laboratory is a high performance research group in Foundations and Applications of Data Science, at **University Rey Juan Carlos**. Its main goal is to coordinate and foster research, education, innovation and knowledge transfer on Data Science and Big Data. The main activities of the Data Science Lab are: Education, traianing, research, consulting and assessment.

**Education and training** of students, junior practitioners and senior professionals in data science and big data methods, technologies and best practices. Teach activities are mainly focused on **Máster in Data Science** at URJC.

**Research and innovation projects**, exploring the creation of novel statistical and computational methods for scalable data mining, machine learning, optimization as well as statistical modelling with complex data sets. Some of our projects are:

- **SABERMED** (Swarm Agent-Based Enviroment for Reputation in MEDicine), a tool capable of assessing the reputation of digital content on the web, which will detect fraudulent content through the application of techniques of Data Science, Big Data architectures and Artificial Intellligent.
- **DIGITANIMAL**: Modeling the behavior of cattle using Data Science techniques and IoT and Big Data tools.
- **MACHINE LEARNING & SECURITY:** CAM project to define new User behavior Analytics techniques.
- **KROOS:** management and knowledge recovering of scientific documents.

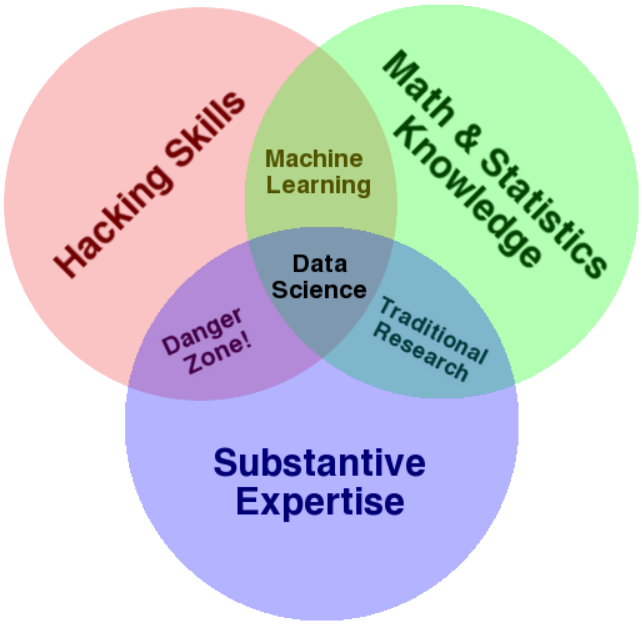
**Consulting and strategic assessment** for companies and organizations, willing to identify and integrate data science and big data technologies in their workflows, with special emphasis on performance, quality management and best practices. Some of the companies we work with are:

- **ERICSSON S.L.**: multinational networking and telecommunications company.
- **GFK**: Germany's largest market research institute, and the fourth largest market research organisation in the world.
- **LIBERTY SEGUROS S.A.**: an insurer company working all around the world.
- **MEDLAB MEDIA GROUP**: a technology start-up filled with top HR talent from Europe and many other parts of the world.
- **REPSOL S.A.**: an integrated global energy company.
- **SENSOWAVE**: solutions in several sectors such as: livestock industry, eHealth, industry, training systems special forces and bodies, etc.
- **TELEFONICA**: a Spanish multinational telecommunications company.
- **VODAFONE**: a multinational telecommunications company.

Javier M. Moguerza

Coordinator of the **Data Science Lab**.

The structure of the **DSLAB** is based on the integration of the three fundamental aspects of every Data Science project




## Data Engineering

In the area of Data Engineering, the storage, representation, transformation, computation and parallelization for large volumes of data are studied. It is responsible for the development, construction, testing and maintenance of Big Data architectures and technologies. Once continuous pipelines are available to and from the information, data analysts can carry out their analyzes. Within this area we consider two subareas:

**Computer Science & Information System**

**Process & Software Engineering Qualitys**

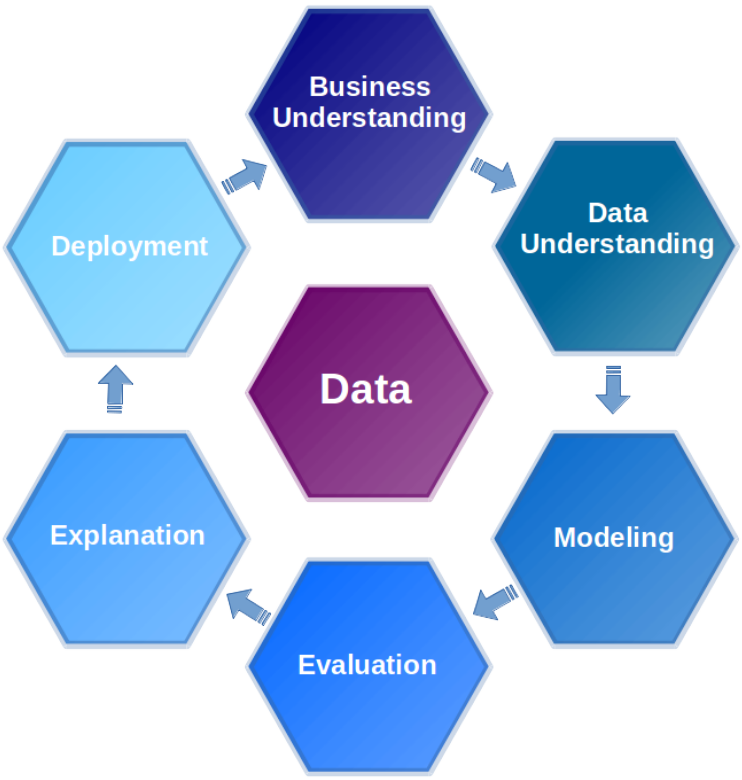
## Data Analytics

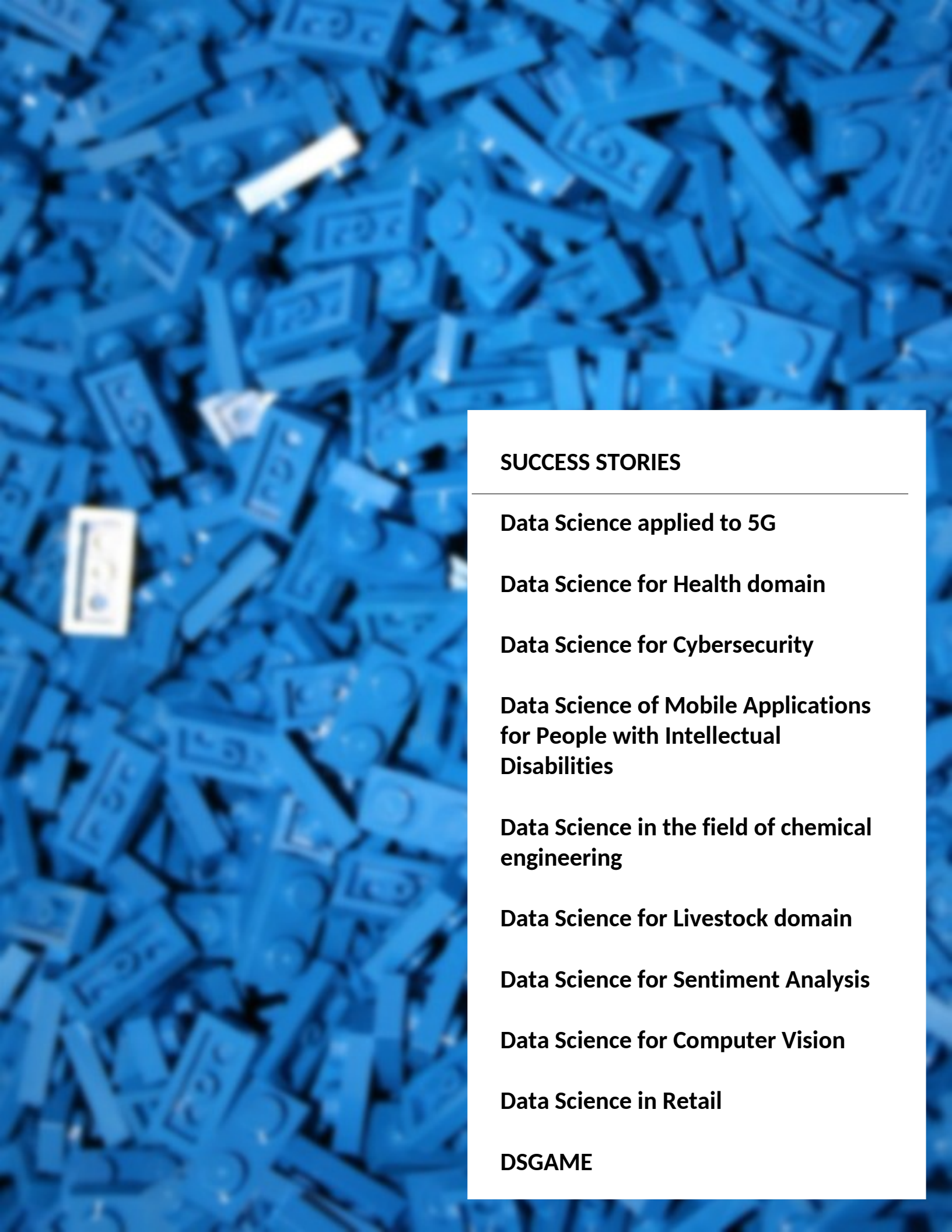
In the area of Data Analysis the models, classification, prediction, and visualization associated with the data are studied. It is the area responsible for the design and development of mathematical and statistical algorithms and models to extract valuable information from the data. Within this area we consider two subareas:

**Statistics & Machine Learning**

**Optimization & Maths**

## Data Science project Lifecycle





## SUCCESS STORIES

**Data Science applied to 5G**

**Data Science for Health domain**

**Data Science for Cybersecurity**

**Data Science of Mobile Applications for People with Intellectual Disabilities**

**Data Science in the field of chemical engineering**

**Data Science for Livestock domain**

**Data Science for Sentiment Analysis**

**Data Science for Computer Vision**

**Data Science in Retail**

**DSGAME**

# Data Science applied to 5G

## A Quality of Experience Management Framework

An automated system, based on the **COMPA** architecture, for degradation detection of video streaming is proposed.
The **QoEMU** framework has been tested with several video users in laboratory experiments. An experiment with more than 300,000 labeled traces was performed.

- Using Machine Learning techniques, the QoE Model has been able to predict the QoE.
- A degradation in the QoE has been generated based on an artificial congestion in the traffic. The **QoEMU** framework was able to detect the generated degradation and manage it through an adequate mitigation action.

## Traffic Classification in an Auto Model Evaluation MI Environment

An automated system for implementing different AI techniques and Big Data technologies has been developed. The system is able to provide all the needed tools and functionalities to perform data analysis, model creation and inline classification. It has been successfully tested on network traffic classification tasks.

## Business results

ERICSSON

**ERICSSON** company has founded a Research Chair for the implementation and development of any activities and / or research projects associated and framed in the field of Data Science applied to 5G technologies and services.

- Research: studies on specific topics will be promoted, seminars and seminars may be promoted, as well as collaboration programs with public and private entities.
- Training: organisation of courses, seminars, conferences, round tables, etc.
- Others: call for scholarships and research and training awards; agreements for the realisation of training practices, by URJC students, in companies and institutions related to the objective of the Chair; publications of joint interest, etc.

DS LAB

# Data Science for the Health domain

## SABERMED

Swarm Agent-Based Enviroment for Reputation in MEDicine is a tool capable of assessing the reputation of digital content on the web, which will detect fraudulent content through the application of techniques of Data Science, Big Data architectures and Artificial Intellligent (Deep Learning and Intelligent Agents).

## KROSS

Knowledge Recovering and Organization Semantic System is a project for the management and knowledge recovering of scientific documents.

- To obtain the relevance of documents, phrases, concepts, words... item Monitoring the dynamic evolution of relevances.
- Text generation.

## SMART BEDS

Automatic Bed Assistance based on Continous Optimization. Development of an intelligent algorithm capable of dynamically learning from different sleep sessions, obtaining the optimal configuration patterns of sleeping mattresses to improve the quality of sleep.

## RNFC

Predictive model of the ability of walking one month after suffering a hip fracture.

## Business results

MEDLAB MEDIA GROUP

**MEDLAB MEDIA GROUP**: technology start-up, highly specialised profiles in IT, research, business and journalism.

IBSAL Instituto de Investigación Biomédica de Salamanca

pixelabs

**PIXELABS S.L.**: Conversational ML learning software allows creating extremely powerful tools, including translation.

RNFC

**HOSPITAL UNIVERSITARIO FUNDACIÓN ALCORCÓN**

DS LAB

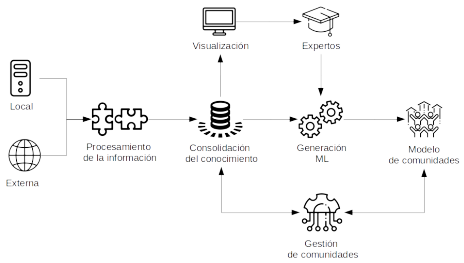
# Data Science for Cybersecurity

## UEBA

A workflow has been proposed allowing Relying Parties within identity federations to perform User and Entity Behaviour Analytics to raise their end-users security levels.

## DRACO

DRACO (Dynamic Recover and Automatic Communities Organiser) is a project that provides a software tool capable of dynamically modelling social behaviour and automatically detecting the communities in which individuals organise themselves according to their behaviour.



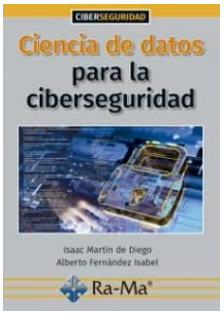

## AVATAR

Prototype for the creation and maintenance of Avatars: a solution capable of managing in a supervised (manual) or unsupervised (automatic) way the complete life cycle of a digital identity in a social network.

## Business results

Cyber security Cluster - CsC

**CsC**: professors and researchers belonging to different departments, centres and schools at Universidad Rey Juan Carlos (Madrid, Spain) collaborating in research, education, innovation and knowledge techology transfer in the fields of Cybersecurity and Privacy.




Data Science as a tool for understanding, preventing, detecting and remediating threats in the Cybersecurity domain. This book is aimed at professionals, students, engineers, mathematicians and all those interested in how to address the challenge of understanding Data Science in the Cybersecurity environment.

Dirección General de Educación Infantil, Primaria y Secundaria
CONSEJERÍA DE EDUCACIÓN E INVESTIGACIÓN
Comunidad de Madrid

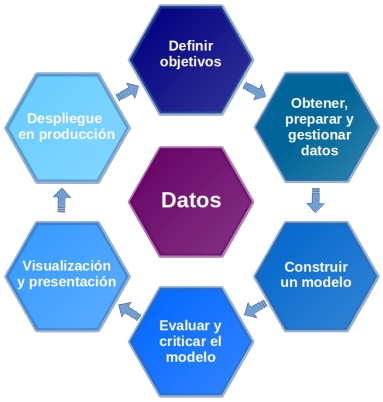
# Data Science of Mobile Applications for People with Intellectual Disabilities

## Mefacilyta: Connected by accessibility

### 1 Intro

Mefacilyta is an application for mobile phones. It serves to create supportive content that helps people with comprehension difficulties to understand daily activities.

### 2 Methods




### 3 Results

Are there any organisations, users and/or trainers whose behaviour could grouping together?

- Functional units are identified as well as the patterns of and atypical behaviours. This grouping could be dynamically done, so that changes in the pattern of behaviour could be identified.
- A Dashboard to visualize the performance of the associationsover the years:

## Business results

Fundación Vodafone España

- Is it possible to track organisations, users and/or trainers over time, with explanatory or predictive character?
- What is the average lifetime of a client? How long does a organisation, a user or a trainer on the platform. Not just in a particular session, but on a global level.
- What are the organisations, users and/or trainers most likely to abandon the platform in the near future? Data Science techniques could be used for predicting abandonment of the platform by users, organisations, and trainers. For this purpose, the history of the functional unit of interest would be studied and apply predictive methods that determine the likelihood of abandonment.
- Is there a way to recover customers when the probability of abandonment is high?

# Data Science in the field of chemical engineering

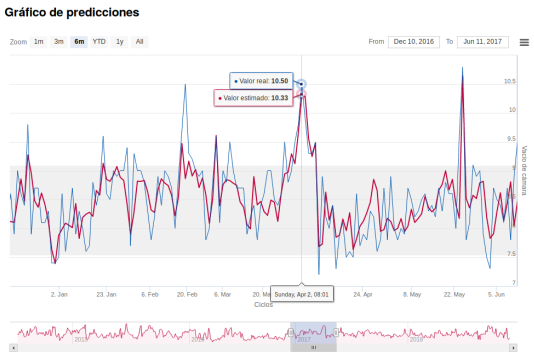
## 1 Objectives

Prediction of a chemical process at different time horizons

## 2 Methods

- Modelling of irregular time series.
- Extraction of characteristics from time series.
- Ad hoc missing data allocation.
- Implementation of new methodology for the selection of variables in high-dimensionality problems.
- Advanced Machine Learning model combining explainability with precision.

## 3 Results




## Business results

**REPSOL**:Petrochemical company in the process of digitization. Thanks to an integrated business model, REPSOL is present throughout the energy value chain and offers a complete range of products and services that make everyday life easier for people and contribute to the sustainable progress of society.

- Machine learning and Artificial Intelligence to model complex chemical processes that are impossible to solve modelling from a physicochemical point of view.

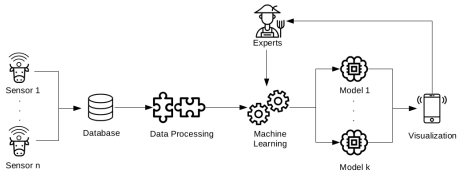
# Data Science for Livestock domain

## Animal behavior analityc using wireless sensor networks

### 1 Intro

A methodology, Animal Behavior Analytics using Wireless Sensor Networks, based on CRISP-DM for facing IoT projects related to the livestock sector.

### 2 Methods




### 3 Results

- Calving detection: $8$ of every $10$ calvings are detected
- Oestrus cycles in beef cattle: using ML techniques, when the sample of cows move away from the day of heat, it is easier to separate their behaviour from the behaviour during the heat day.
- GELOB: to prevent wolf attacks using the analysis of live data taken from livestock holdings.

## Business results

Senso Wave
Professional solutions

**SENSOWAVE**: a technological company aiming to improve safety and efficiency in sectors like industry and farming by providing innovative solutions based on IoT

IoT and Big Data solutions for livestock and industry

- Know where your animals are at all times. Forget about turning around trying to locate your cattle.**DIGITANIMAL** does it for you. Cut-off operating costs.
- Co-founded by Comunidad de Madrid, Research and innovation promotion program.

Dirección General de Educación Infantil, Primaria y Secundaria
CONSEJERÍA DE EDUCACIÓN E INVESTIGACIÓN
**Comunidad de Madrid**

DS LAB

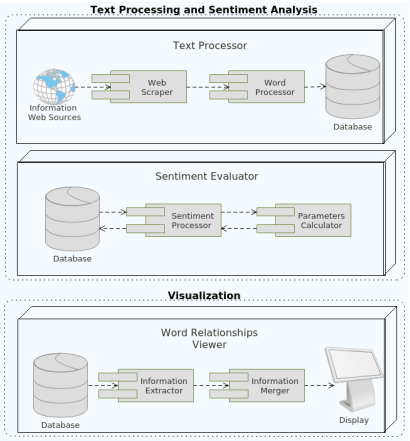
# Data Science for Sentiment Analysis

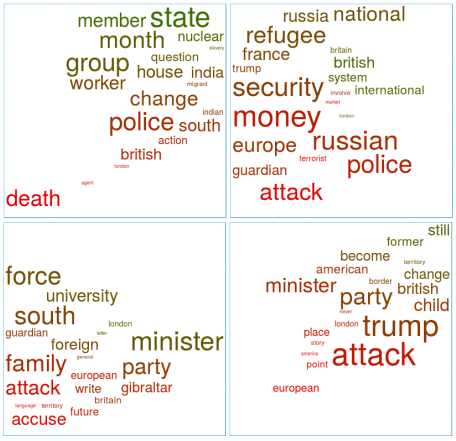
## Visual framework for dynamic emotional web analysis

### 1 Objectives

EmoWeb is a novel tool for evaluating the dynamic sentiment analysis of textual content from websites.

### 2 Methods & Results




### 3 Visualisation

http://www.datasciencelab.es/research/projects/

## Business results

EmoWeb framework is used for analyzing and representing the sentiments of texts from web sources of information.

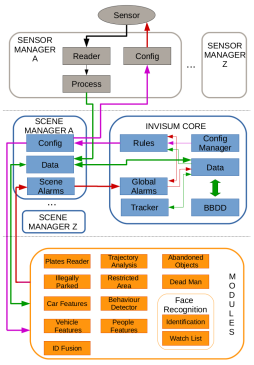
- EmoWeb consists of a hybrid approach based on a dictionary and statistical methods to produce the sentiment polarity of words.
- Words processed at lexical level through Natural Language Processing techniques.
- A web dashboard is created to manage data and present results.
- Dynamic visualization of the evolution of sentiment and occurrence of words.
- Enriched and dynamic word cloud where the position of words has meaning.
- Detecting and visualizing the most significant events related to a word of interest during a period of time.
- Visualizations are able to reflect the most significant news for its editorial trend

# Data Science for Computer Vision

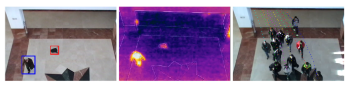
## INtelligent VideoSUveillance SysteM

INVISUM is developing an advanced and comprehensive security system, oriented to the analysis and categorisation of multimedia content that addresses the limitations of scalability and flexibility of current video surveillance systems whilst incorporating new techniques of compression, pattern detection and decision support. It uses advanced architectures that optimise the infrastructure and associated resources to maximise system efficiency

- This project forms part of the security challenge and seeks to provide a solution based on the development of a smart sensing solution together with information processing in the cloud.
- The final aim is to present to the end user correct and needed high-level information. On the business level, INVISUM seeks to provide a complete video monitoring solution that will adjust to the complexities of differing environments, providing commercial solutions for a growing marketing.



(a) Restricted Area (b) Kinect Image (c) Tampering



(d) Abandoned Object (e) Thermal Image (f) Crowd trajectories

(g) Suspect Detection (h) Restriced Area (i) Vehicle Features

## Business results

**Innovati** is an IT engineering company with nearly ten years of experience in providing its customers with creative and innovative high added value IT solutions tailored to their specific needs. Its team is composed by more than 200 degree engineers and highly qualified IT technicians. Since its foundation the company has bet for developing an enterprise model based on CONTINUOUS INNOVATION and CREATIVITY offering technological excellence and passion for work well done as the main pillars to support our customers' continuous growth and progress.

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

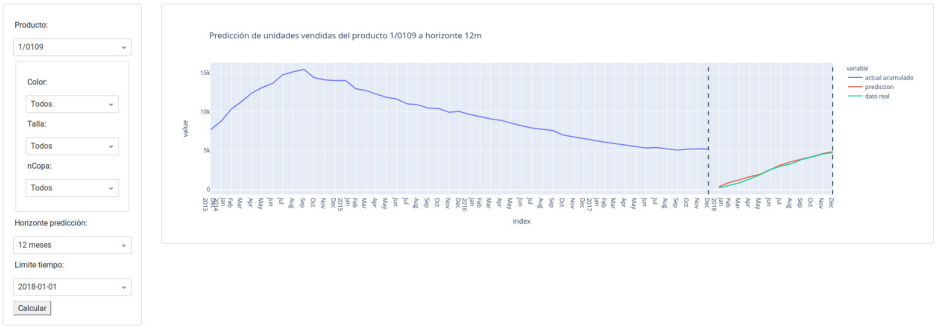
# Data Science in retail

## EUL3R

Optimisation models applied to inventory management processes in the retail industry to be integrated in the project.

- Representation of the information
- Time Series Analysis
- Forecasting
- Fusion of Information
- Visualization

**Predicción de unidades vendidas**




## Business results

autana
BUSINESS PARTNERS

**ABP**: functional and technological support to the business units in achieving their sales and customer service objectives. We help convert commercial strategy into business processes.

gisela

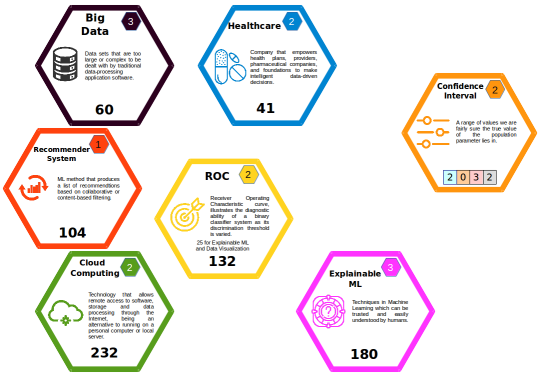
DS LAB

# Data Science GAME

## DSGAME

Promoting scientific literacy among schoolchildren and non-university youth and their interest in scientific-technological careers. Direct contact with the scientific method and research practice presented as successful case studies in the field of Data Science.

- DSGAME proposes scenarios related to success stories in the field of Data Science, Artificial Intelligence, Computer Science and Mathematics. It will pose challenges to the students orienting them towards the development of skills and values of the scientific method, helping in making decisions based on evidence.
- Board game developed in two versions, **DSGAME**, for young people from 12 years old (secondary education, high school, university) and **DSGAME KIDS** for primary school children (especially after 8 years).




## Business results

ACADEMIA JOVEN DE ESPAÑA

**Funding** oportunities.

- PRECIPITA is the **crowdfunding** platform for research and scientific dissemination projects managed by the Spanish Foundation for Science and Technology (FECYT).
- FECYT: aid for the promotion of scientific, technological and innovation culture.
- KICKSTARTER: crowdfunding platform focused on creativity and merchandising.

DSLAB
DATA SCIENCE LABORATORY