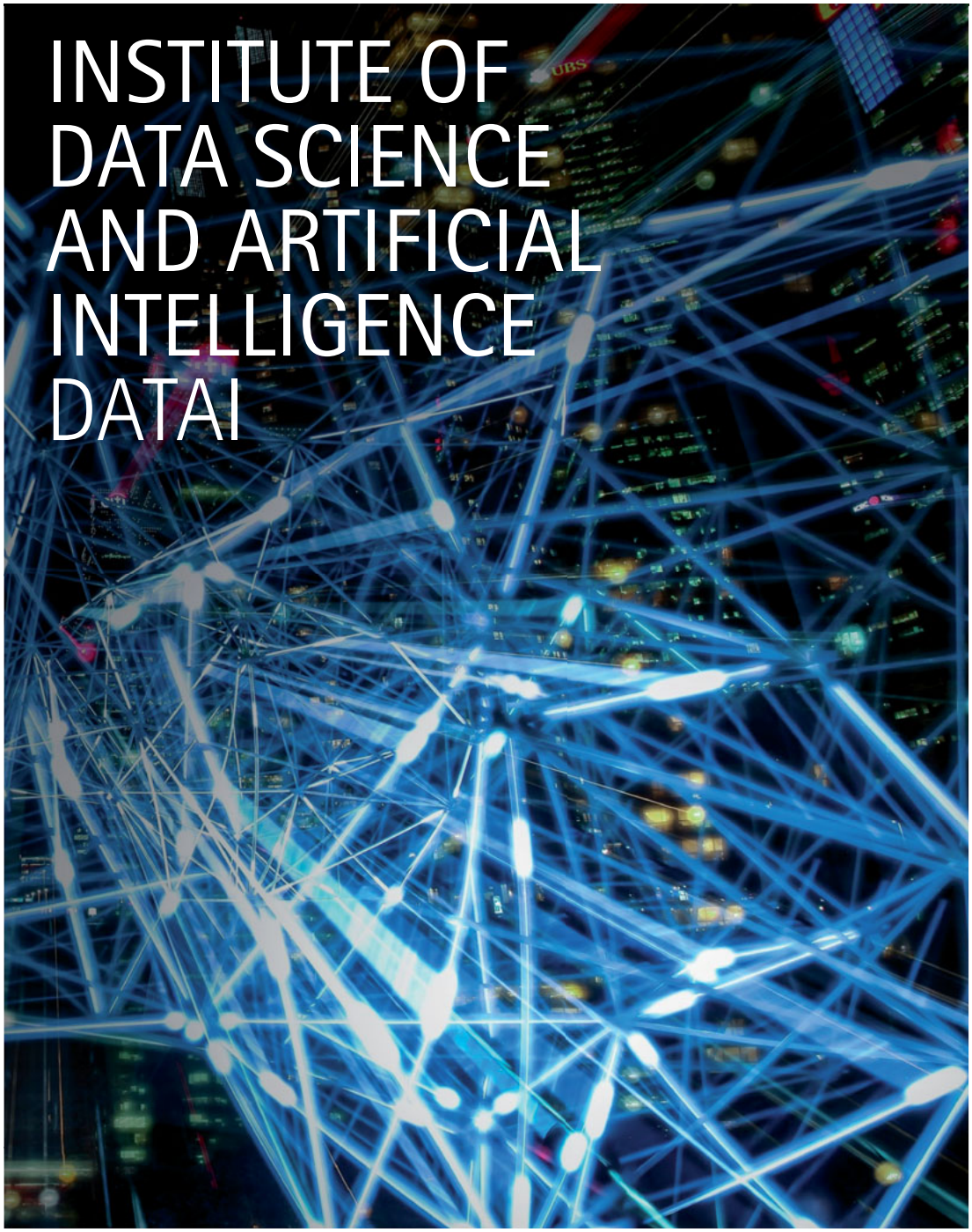
ANNUAL REPORT. 2020 AND 2021

INSTITUTE OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE DATAI



Universidad de Navarra

DATAI IS A CROSS-CUTTING RESEARCH, INNOVATION AND TRAINING CENTER BELONGING TO THE UNIVERSITY OF NAVARRA THAT PROMOTES COLLABORATIVE WORK BETWEEN THE UNIVERSITY'S DIFFERENT GROUPS AND INDIVIDUALS WITH THE AIM OF ACHIEVING EXCELLENCE IN **DATA SCIENCE AND ARTIFICIAL INTELLIGENCE**

CONTENTS

0. INTRODUCTION

INTRODUCTION

The Institute of Data Science and Artificial Intelligence (DATAI) was launched on December 11, 2019. This cross-cutting research, innovation and training center belonging to the University of Navarra promotes collaborative work between the University's different groups and individuals with the aim of achieving excellence in data science and artificial intelligence.

The Institute, which is part of the Technological Campus of the University of Navarra (TECNUN) and is located in Pamplona, aims to serve as a cross-cutting entity to bring together statistical analysis and data processing projects that are already being carried out by the University's different schools, groups and services. It was conceived with a clear collaboration-based vocation to integrate training and research activity in big data and artificial intelligence.

The Board of Directors is made up of Director Jesús López Fidalgo and three deputy directors: Elisabeth Viles (TECNUN School of Engineering), Stella Maris Salvatierra (School of Economics) and Sergio Ardanza-Trevijano Moras (School of Sciences); manager Fernando de la Puente García-Ganges and secretary José María González Gullón.

MASTER'S PROGRAM BOARD OF DIRECTORS AT THE PRESENTATION OF THE INSTITUTE (FEBRUARY 2020)

DATAI has members on all university campuses: Pamplona, Madrid, San Sebastián and Barcelona. In addition to collaborating with the University of Navarra's different schools, the Institute also works with staff from research and assistance centers such as CEIT, CIMA and CUN.

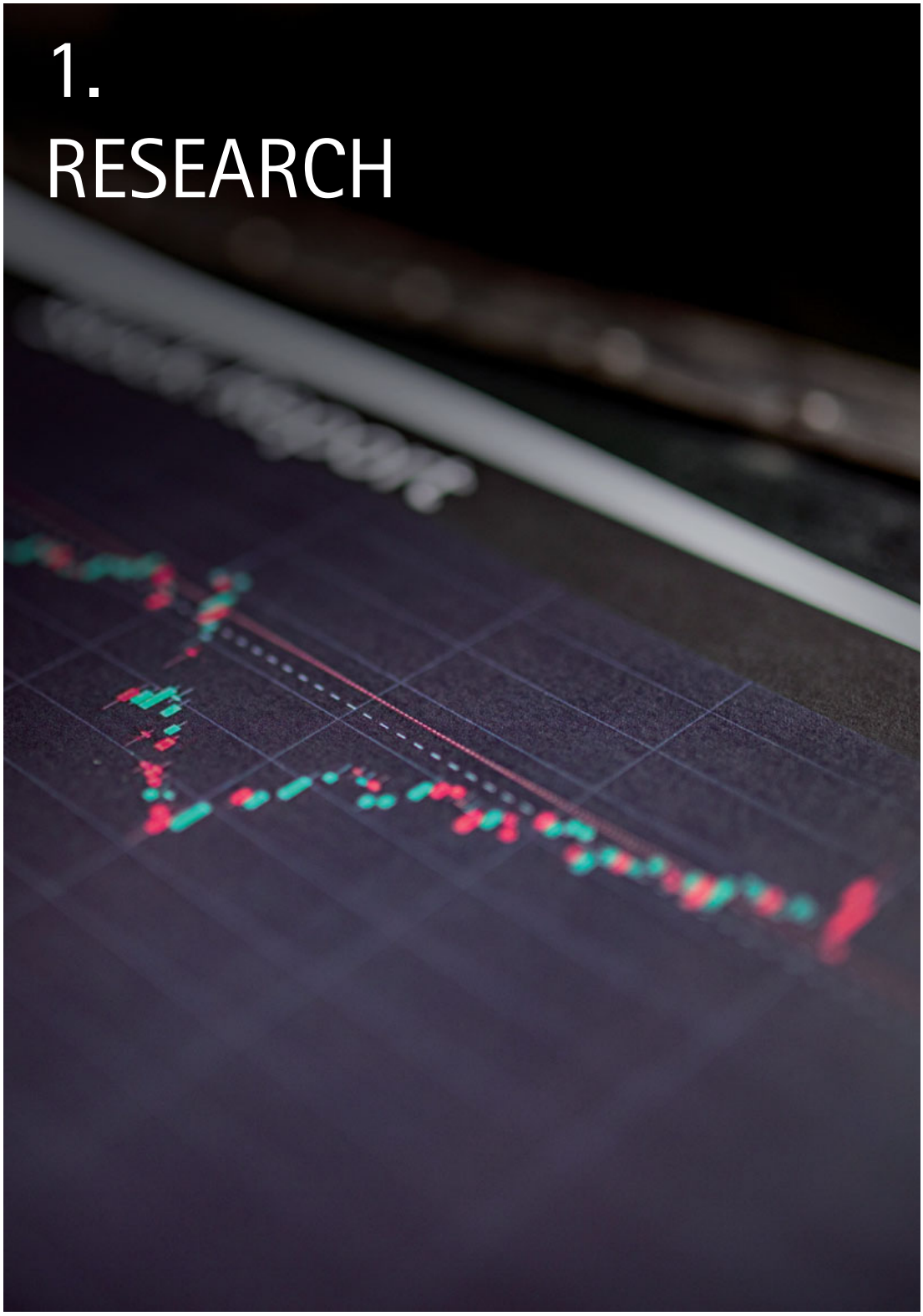
Many external researchers from companies such as BBVA and HSBC International Banking also participate in the Institute's training and teaching activities. Its goals include providing extensive support in **research** and **innovation,** and **transferring knowledge** to industry, business and society.

It is also closely involved in **training** researchers, students and professionals. For this reason, the Institute runs a university-accredited Master's Degree in Big Data Science aimed primarily at young professionals and doctoral students in different areas with significant data analysis needs.

The Institute's **Advisory Committee** is made up of representatives from all University schools: *Architecture*, *Sciences*, *Communication*, *Law*, *Economics*, *Education and Psychology*, *Nursing*, *Pharmacy and Nutrition*, *Humanities and Social Sciences*, *Medicine*, *and Theology.* It also has representatives from *CEIT*, *CIMA* and *CUN*, as well as collaborators from external institutions such as *BBVA*.

SCHOOLS	REPRESENTATIVE
ARCHITECTURE	EDUARDO BAYO PÉREZ
SCIENCES	DIEGO MAZA OZCOIDI
COMMUNICATION	IDOIA PORTILLA MANJÓN
LAW	JUAN CARLOS HERNÁNDEZ PEÑA
ECONOMICS	ALEJO SISON GALSIM
EDUCATION AND PSYCHOLOGY	MAITE AZNÁREZ SANADO
NURSING	MARÍA PUEYO GARRIGUES
PHARMACY AND NUTRITION	IÑAKI FERNÁNDEZ DE TROCÓNIZ
HUMANITIES AND SOCIAL SCIENCES	JUAN JOSÉ PONS IZQUIERDO
MEDICINE	ALFREDO GEA SÁNCHEZ
TECHNOLOGICAL CAMPUS OF THE UNIVERSITY OF NAVARRA (TECNUN)	ÁNGEL RUBIO DÍAZ-CORDOVÉS
THEOLOGY	RUBÉN HERCE
INSTITUTES AND OTHER CENTERS	
CEIT	SAIOA ARRIZABALAGA JUARISTI
CIMA	MIKEL HERNÁEZ ARRAZOLA
CUN	MARÍA A. FERNÁNDEZ SEARA
IESE	SAMPSA SAMILA
EXTERNAL INSTITUTIONS	
BBVA	MARCO BONILLA

1. RESEARCH



RESEARCH

The Institute promotes and disseminates research findings in data science, big data and artificial intelligence. It aims to provide a high-quality forum for researchers, professionals and policy-makers at the local, national and international level.

Since the Institute was founded, it has taken a cross-cutting approach to research, which involves the collaboration of experts from different areas and professionals from the fields of data science, big data and artificial intelligence, as well as statistics and computer science.

The Institute seeks to carry out quality research that can be transferred to society and the business sector.

1.1 SCIENTIFIC COMMITTEE

DATAI has a **Scientific Committee** made up of five prominent figures from the field of data science and artificial Intelligence: Amparo Alonso Vetanzos, former president of AEPIA (Spanish Association for Artificial Intelligence) and member of the Spanish Higher Council for Artificial Intelligence; Enrique del Castillo, distinguished professor at the University of Pennsylvania; Nuria Oliver, Director of Data Science at Vodafone; John Stufken, Director of Informatics and Analytics at the Bank of America; and Trevor Hastie, honorary member of the committee, of the Department of Statistics and the Department of Biomedical Data Science at Stanford University.

Amparo Alonso Betanzos

President of AEPIA (Spanish Association for Artificial Intelligence)

Full Professor of Computational Science

→ See CV

Enrique del Castillo

Distinguished Professor of Industrial Engineering and Professor of Statistics

The Pennsylvania State University

→ See CV

Nuria Oliver

PhD Director of Research in Data Science

Vodafone Chief Data Scientist | Data-Pop Alliance

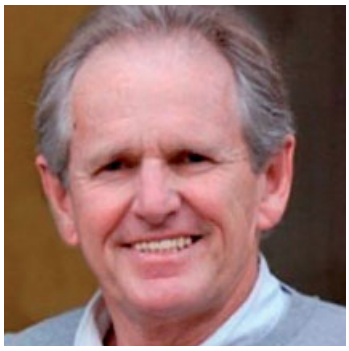
→ See CV

John Stufken

Director, Informatics and Analytics

Bank of America Excellence Professor | UNC Greensboro

→ See CV



Trevor Hastie (Honorary Member)

Professor, Department of Statistics and Department of Biomedical Data Science

Stanford University

→ See CV

1.2 RESEARCH AREAS

DATAI has 73 members, including seconded members (12), associate members (43) and guests (18); they are organized into the nine research areas that comprise DATAI. In turn, each area has its own lines of research.

1.2.1 STRATEGIC LINES OF RESEARCH

1

Optimal Experimental Design: Modeling studies used in various areas, such as tumor growth, clinical trials in the first three phases, retention of radiation in the human body and pharmacokinetics. Intelligent subsampling (active learning) in big data. Application of optimal design techniques in the selection of the initial parameters of an algorithm and in computational experiments (in silico). Personalized medicine.

2

Complex Networks and Social Network Analysis: Network segmentation and detection of influential or singular nodes. This involves studying the indices for measuring the properties of the network and its nodes.

3

Machine Learning in Bioinformatics and Computational Neuroscience: Processing of a large number of variables generated by genetic analysis, usually with few subjects. In particular, it seeks to detect reliable gene interactions by means of classifiers based on hierarchical Bayesian networks. MicroRNA expression in patients with various pathologies. Identification of biomarkers in colorectal cancer. Automatic classification of neurons.

4

Natural Language Processing and Sentiment Analysis: This is conducted mainly through deep learning and parameter regularization.

5

Fair Learning: Modeling and development of algorithms that are not biased by prejudice or unfair assessments of certain variables. The fundamental problem with developing a fair model is that many variables contain a part that causes the model to use social biases to make decisions; at the same time, however, those variables contain important information for making decisions.

To differentiate between the two, path analysis models are used in the context of causal inference, which helps extract the most informative part from each variable that does not produce bias.

1
OPTIMAL EXPERIMENTAL DESIGN

Modeling in various areas, such as tumor growth, clinical trials in the first three phases, retention of radiation in the human body and pharmacokinetics. This is a line carried out by the Institute's core team, which has been working for more than 30 years and has obtained relevant results. Given the strategic nature of this line, it is therefore necessary to establish a strong working team at DATAI. There is already a consolidated group, especially at the Universidad de Salamanca and the Universidad de Castilla-La Mancha, but it also includes researchers from the Universidad Pública de Navarra, the Universidad de Almería and Universidad de La Laguna. More recently, efforts have been made within this line to apply optimal experimental design techniques to the fundamental problems emerging in data science, big data and artificial intelligence. Specifically:

A. **Active learning** (intelligent subsampling) in big data. One of the main issues with processing large amounts of data is the computational cost of model analysis and fitting, which increases exponentially in some cases, thus making them NP problems. A common technique, based on the central limit theorem or the law of large numbers, is to take a random subsample and perform the calculations with a reasonable amount of data. This line of research seeks to improve the sampling method so that the most informative data within the large sample are captured. This is especially relevant when the distribution of explanatory variables is highly asymmetric or presents extreme kurtosis. The latter topic is extremely underrepresented in the literature.
B. **Green algorithms**. The rise in computational capacity, together with the possibility of collecting large amounts of data, has led us to develop and implement algorithms that require a lot of computational time, which results in high energy consumption. In part, so-called green algorithms seek to optimize this energy cost through more efficient algorithms that require fewer execution resources. Optimal design techniques can be applied to the optimal selection of the initial parameters of an algorithm, so that the speed of convergence is higher and the computational time is thus lower.
C. **In silico experiments**. Also called computer experiments, these are becoming increasingly common through so-called digital twin models. They are simulators constructed with complex systems of partial derivative equations, which usually require a high computational cost. The theory of ***computer experiments*** is that they

seek to simplify these models so that, with minimal loss of accuracy, results can be achieved in a much shorter time as the model acquires greater ***explainability***.

D. **Model selection**. Typically, there are many possible rival models for the same phenomenon. In addition, there are a large number of variables, from which it is necessary to select a few that are sufficient to explain the response and make predictions. Model discrimination plays an essential role in this respect. The team has been working on this topic for years and has made some key contributions.

E. **Personalized medicine**. The team has developed customized experimental designs, so that the experiment performed responds to the circumstances of the subject, instead of conducting full randomizations for all subjects. This makes it possible to obtain better model fits with fewer experiments. The team seeks to use this technique to propose treatment tailored to each patient in light of their particular circumstances.

2
COMPLEX NETWORKS AND SOCIAL NETWORK ANALYSIS

The fundamentals of network science are increasingly being applied to study a variety of real-world phenomena, many of which are complex systems by their very nature. Researchers in various fields consider the formulation and study of complex systems a crucial topic. These complex systems can be modeled by means of networks (also called graphs), which are represented by nodes and links (network components) that interact with each other to define difficult-to-identify patterns and structures. In general, the nodes represent the network entities and the links represent the connections established between them. Examples of complex networks include brain structures, social relationships, cell phone communications, molecular biological interactions, banking transactions and transport networks.

Network data are complex in many of these fields of application in terms of both the large volume of information to be processed and the wide range of data, thereby making analysis a challenge. Therefore, given the levels of information, studying the structures and influence of the different network components (nodes and links) leads to the study of new and more efficient models and techniques.

In this context, complex network analysis is a research field related to the design, development and application of mathematical and computational models for the detection and analysis of the relevant underlying structures and patterns in these networks.

The tasks that should be studied in these types of network are: detection of groups of nodes (entities) with similar behavior (e.g., groups of people with common interests in a social network, a set of similar proteins in a biological network); identification of influential nodes in the network (e.g., users in a social network with a high number of followers, accounts with a high number of transactions in a banking network); the analysis of "information" dispersion (e.g., identification of network entities that maximize dispersion in an epidemiological scenario); prediction of future relationships between nodes (e.g., identification of "beneficial" future relationships in a network of entrepreneurs); identification of anomalous nodes (e.g., detection of sensors with atypical measurements with respect to the norm in networks of wireless sensors), etc.

The wide range of possible scenarios for network modeling means that the analysis of complex networks is a cross-cutting field of research with great applicability in the current context.

3
MACHINE LEARNING IN BIOINFORMATICS AND COMPUTATIONAL NEUROSCIENCE

Translational medicine seeks to combine various scientific disciplines to improve the prevention, diagnosis and treatment of clinical conditions. By definition, it is a highly interdisciplinary field whose main objective is to improve the global healthcare system. Although most biomedical disciplines have evolved to incorporate digital devices, the algorithms that collect process and analyze data are overwhelmingly unfamiliar and/or inaccessible to most professionals.

What is now known as the digital medicine revolution is nothing more than the natural evolution of information technologies to aid with clinical and medical evaluations. Machine learning is already making a difference in this context.

A prime example is the use of basic neuroscience to personalize the medical evaluations of patients with brain disorders. More specifically, a program that takes basic neuroscience research from clinical validation to hospital protocols could lead to the development of personalized treatment.

In this context, DATAI plans to work on a line focused on machine learning in bioinformatics and computational neuroscience. This will specifically involve processing a large number of variables generated by genetic analysis, usually with few subjects. In particular, it seeks to detect reliable gene interactions by means of classifiers based on hierarchical Bayesian networks, microRNA expression in patients with various pathologies, identification of biomarkers in colorectal cancer and automatic classification of neurons.

4
FAIR LEARNING

In the context of AI, fairness refers to the various attempts to correct for algorithmic bias. Although definitions of fairness are always controversial, the results of an algorithm can be considered fair if they are independent of a given variable, especially those considered sensitive, such as individual traits that should not correlate with the outcome (e.g., gender, ethnicity, sexual orientation and disabilities). In machine learning, the problem of algorithmic bias is well known and has been widely studied.

Debates about fairness in machine learning tend to focus on the impact of different models on socially sensitive groups. However, impartiality is a broad topic and is relevant to almost all contexts. Which users are shown certain ads, offered certain prices, given rewards, given lower call waiting times and identified based on different propensity models?

Various fairness criteria have been proposed in recent years, but two approaches stand out: **demographic fairness** (equivalent to eliminating disparate impact) and **equal opportunities.**

Demographic fairness ensures that any decision made by an AI model is not related to a sensitive attribute such as race, gender or age. In other words, being male or female should not determine whether someone will commit a crime.

Equality of opportunity is a bit more subtle: it requires that individuals who qualify for a good outcome obtain that outcome with the same probability, regardless of whether they are members of the sensitive group.

For example, the percentage of individuals who qualify for a loan and end up receiving it should not differ between racial groups.

Research on fairness in machine learning is a relatively new topic but one that is becoming highly relevant due to the rise in AI and the use of data. It is a cross-cutting field that has ethical and legal implications and is already the focus of the European Commission's AI strategy, which highlights fairness as one of the seven key aspects when developing AI models. Spain's artificial intelligence strategy also highlights the importance of the sustainable development of AI.

In this context, DATAI plans to develop a fairness research line focused on:

- The study of algorithm fairness, measurement and mitigation of bias, and the development of new mathematical tools and models to improve fairness.
- The application of fairness for the development of sustainable and ethical AI to improve society.

5
NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a subfield of AI concerned primarily with the processing and understanding of human language by machines. By enabling machines to understand human language, NLP improves the accuracy and efficiency of processes. This is achieved by automating a range of repetitive tasks. Examples of NLP applications include classification of entries, machine translation, spell checking and summarization.

This area of study emerged in the 1950s and is now regaining relevance due to the large amount of text currently generated and the capacity to process it. Generically, text analysis can be divided into two main components:

- **Syntax analysis:** this is used to establish meaning by studying the grammar of a sentence. It is the process of structuring the text using the grammatical conventions of language. Essentially, it consists of analyzing sentences by dividing them into groups of words and phrases that create a correct sentence. This does not take account of the fact that sentences can be meaningless, which is where semantic analysis comes in.
- **Semantic analysis:** Our understanding of language is based on years of listening and learning the context and meaning. Human language is dynamic. The invention of machine learning algorithms means that computers can understand the meaning and logic behind our expressions, at least to some extent.

With the development of NLP, we are now able to perform **sentiment analysis** for human language. This makes it possible to detect emotions in the text, which is one of the applications of NLP. Thus, an algorithm can read a text and label the underlying emotion (whether it is positive, negative or neutral). This, together with the growth of social networks, makes it possible to conduct research that would have been impossible just a few years ago. For example, one can objectively measure how the population feels about a particular event.

NLP is a field that has been developed mainly for text analysis in English and has started to be extended to Spanish only in recent years. Its application to Spanish is therefore one of the newest lines to be developed in the near future in AI-related research. This has been reinforced by the strategy of the Spanish government, which is in the process of publishing the Strategic Project for Economic Recovery and Transformation (PERTE) associated with the Spanish language and AI. For these reasons, DATAI is interested in developing a line of research based on:

- Sentiment analysis through NLP, given its high development and impact potential for society in the future, especially Spanish-speaking societies.

Moreover, the fact that it applies to almost all fields underlines DATAI's vocation to become an entity of excellence that promotes cross-cutting research throughout the University.

The Institute has already carried out collaborations with the School of Communication in the field of NLP and is currently in talks with the Justice Department to form part of a working group that will focus on the application of NLP to improve the management of justice in Navarre.

1.2.2 RESEARCH AREAS OF DATAI RESEARCHERS

1
Fundamentals

2
Advanced Analytics
BBVA

3
AI and Management
IESE

4
Computational Biology
CIMA

5
Computational Biology
TECNUN

6
Computer Science

7
Ethics and Law

8
Health Analytics

9
Industrial Data Analysis & Information Management

1
FUNDAMENTALS

This is a multidisciplinary group with a solid background in applied and methodological research in data science. Its research is based primarily on modeling and experimental designs, including active learning.

The main lines of research are as follows:

- Experimental designs: active learning.
- Natural language processing and sentiment analysis.
- Complex networks: data mining and dynamics.
- Fractional integration time series and long-memory processes.

2
ADVANCED ANALYTICS – BBVA

BBVA believes that knowledge derived from financial data can transform the banking industry and its role in the world. For this reason, the Advanced Analytics area is researching the application of AI to the banking sector. The lines of research include the following:

- Optimization and machine learning techniques for banking portfolios.
- Advanced dynamic pricing methods.
- Recommendation engines.

3
AI AND MANAGEMENT – IESE

The new Initiative on Artificial Intelligence and the Future of Management is a multidisciplinary project that analyzes the impact of AI on business management and helps train managers in the ethical and socially responsible use of AI in their companies.

Artificial intelligence will affect all spheres of economic activity in the coming years, so managers must learn to adapt to the evolution of this sector. They must also be able to transform their companies and ensure that they and their employees have the necessary skills to continue delivering value in this new context. Consequently, the purpose of IESE's new AI initiative is to respond to these needs for knowledge generation and management training.

The areas of research are as follows:

- AI skills and the labor market.
- The impact of industrial automation.
- The development and use of AI in businesses, especially in Europe.
- The future of work and organizations.

4
COMPUTATIONAL BIOLOGY – CIMA

Molecular biology has undergone a revolution due to the capacity to simultaneously study the function and expression of thousands of genes and proteins in the patient's body.

Through the use of information technology, databases and statistical analysis, we can accurately and rapidly analyze large amounts of data to allow us to understand the complexity of the mechanisms that cause disease.

The Computational Biology Program at CIMA - University of Navarra currently has two main lines of research:

- Analysis of transcriptomic data, both on a large scale and at single-cell resolution.
- Development of new file formats for storage of and access to omics data.
- Machine learning methods to solve biomedical problems and their translation to clinical practice.

5
COMPUTATIONAL BIOLOGY – TECNUN

The Computational Biology Group at TECNUN is a multidisciplinary team with extensive experience in the development of optimization algorithms, statistical analysis, and machine learning and deep learning methods applied to different biological problems. The group's main focus is on human health through the integration of high-throughput data (genomics, transcriptomics, proteomics, metabolomics, etc.) and biological databases (genomic, pharmacological, metabolic, etc.). The main lines of research are as follows:

- Precision oncology.
- Predictive models of drug toxicity based on structural features of molecules.
- Compression schemes adapted to different omics data.
- DNA sequencing analysis.

6
COMPUTATIONAL SCIENCE

Computational science covers a wide range of topics, including theoretical and algorithmic foundations, as well as cutting-edge developments in computer vision, intelligent systems, bioinformatics and other exciting fields. The roles of computer scientists can be grouped into the following categories: software design and implementation, the development of efficient ways to solve computer problems, and the design of new ways to use computers.

Information technology is essentially the study of how to map machine language into a more understandable and usable language to help people solve complex problems. This field uses the theory of computation in the calculation of information. The lines of research are as follows:

- Climate change.
- Image analysis.
- Modeling and simulation.
- Topological data analysis.
- Blockchain and authenticated data structures.
- Soft computing and resilience.

7
ETHICS AND LAW

The challenge for the area of Ethics and Law is to ensure that new technologies that affect society as a whole seek not only to improve predictions, but also to guarantee that the factors underlying these predictions are clear. Legislation should be promoted, without halting progress in the development of these technologies, to gain a better understanding of the social or biological functions involved. Ultimately, knowledge of social phenomena should be obtained before predictions are made. The lines of research are as follows:

- Ethics and AI.
- Legal regulation of big data, AI and data protection.
- Virtue ethics and AI.

8
HEALTH ANALYTICS

Health analytics refers to the use of large amounts of data collected to provide organizations with useful information and improve patient care. This knowledge is developed in analytical disciplines to foster data-driven decision-making. In turn, these decisions improve the planning, management, measurement, learning and, ultimately, quality of patient care.

As healthcare organizations around the world strive to improve patient care and do more with less, analytics is gaining prominence. Developing analytics skills can help healthcare organizations leverage big data to improve their bottom line and provide patients with added value. The lines of research associated with this area are as follows:

- Analysis of biomedical images.
- Systems neuroscience.
- Medical imaging.
- Causal inference in biomedicine.
- Epidemiology.
- Clinical research.

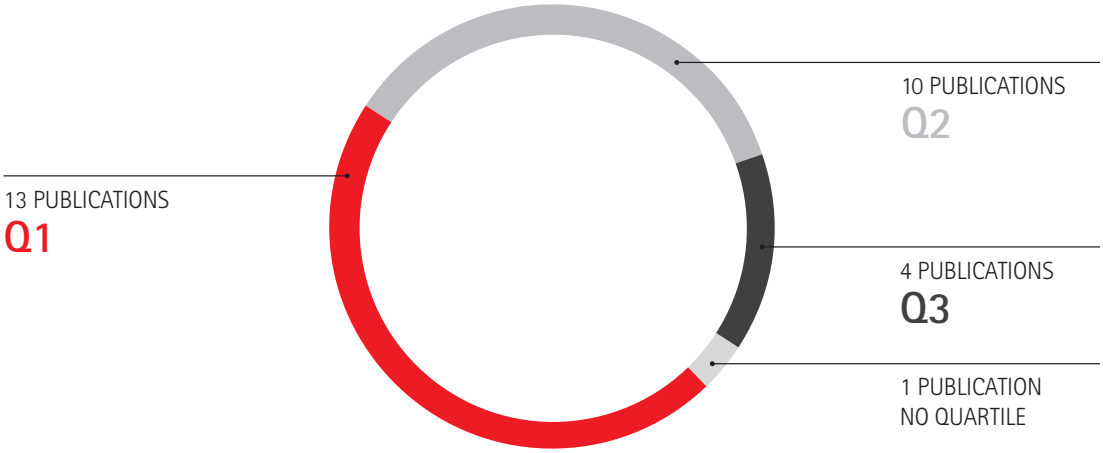
9
INDUSTRIAL DATA ANALYSIS & INFORMATION MANAGEMENT

The group's research focuses primarily on measuring the impact of big data on the high-tech industry. Our research has two main areas of specialization for the development of industrial processes: cybersecurity analytics and big data. The main areas of activity are as follows:

- Communication and security protocols.
- Identification of anomalies and significant parameters in processes.
- Predictive and preventive maintenance models.

1.3 PUBLICATIONS

In 2020 and 2021, the scientific activity of DATAI's seconded members and Board of Management (five researchers) resulted in 28 publications (13 Q1, 10 Q2, 4 Q3 and one with no quartile), thus reflecting the Institute's cross-cutting character:



- De La Calle-Arroyo, Carlos; López-Fidalgo, Jesús; Rodríguez-Aragon, Licesio J. **Optimal designs for Antoine Equation**. CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS. 214, ELSEVIER, 15/07/2021. ISSN 0169-7439, ISSN 1873-3239. Q1.

- Belar, Alazne; Arantzamendi, María; Santesteban, Yolanda; López-Fidalgo, Jesús; Martínez, Marina; Lama, Marcos; Rullan, María; Olza, Ines; Breeze, Ruth; Centeno, Carlos. **Cross-sectional survey of the wish to die among palliative patients in Spain: one phenomenon, different experiences**. BMJ SUPPORTIVE & PALLIATIVE CARE. 11- 2, pp. 156 - 162. BMJ PUBLISHING GROUP, 01/06/2021. ISSN 2045-435X, ISSN 2045-4368. Q2
- Pallotti, María Caterina; López-Fidalgo, Jesús; Centeno, Carlos; Celin, Daniela; Biasco, Guido; Giovannini, Maddalena; Maltoni, Marco; Noguera, Antonio. **Does Delirium Phenomenology in Persons with Advanced Cancer follow a Specifi c Pattern?**. JOURNAL OF PALLIATIVE MEDICINE. 24 - 7, pp. 1061 - 1066. MARY ANN LIEBERT, INC, 16/04/2021. ISSN 1096-6218, ISSN 1557-7740. Q2
- Gamero-Salinas, Juan; Kishnani, Nirmal; Monge-Barrio, Aurora; López-Fidalgo, Jesús; Sanchez-Ostiz, Ana. **The influence of building form variables on the environmental performance of semi-outdoor spaces. A study in mid-rise and high-rise buildings of Singapore**. ENERGY AND BUILDINGS. 230, ELSEVIER SCIENCE SA, 01/01/2021. Q1
- Landecho, Manuel F.; Alegría-Murillo, Leire; López-Fidalgo, Jesús; Colina, Inmaculada; Santesteban, Virginia; García-Unciti, Marisol; Beloqui, Oscar; Fruhbeck, Gema; Cuervo, Marta. **Unravelling gender-specific factors that link obesity to albuminuria**. EUROPEAN JOURNAL OF CLINICAL INVESTIGATION. 50 - 11, WILEY, 01/11/2020. ISSN 0014-2972, ISSN 1365-2362. Q1.
- López-Fidalgo, Jesús; Amo-Salas, Mariano. **Optimal dose calibration in radiotherapy**. RADIATION PHYSICS AND CHEMISTRY. 174, PERGAMON-ELSEVIER SCIENCE LTD, 01/09/2020. ISSN 0969-806X. Q1.
- Higueras, Manuel; Howes, Adam; López-Fidalgo, Jesús. **Optimal experimental design for cytogenetic dose-response calibration curves**. INTERNATIONAL JOURNAL OF RADIATION BIOLOGY. 96 - 7, pp. 894 - 902. TAYLOR & FRANCIS LTD, 02/07/2020. ISSN 0955-3002, ISSN 1362-3095. Q1.
- Gamero-Salinas, Juan; Monge-Barrio, Aurora; Kishnani, Nirmal; López-Fidalgo, Jesús; Sánchez-Ostiz, Ana. **Passive cooling design strategies as adaptation measures for lowering the indoor overheating risk in tropical climates**. ENERGY AND BUILDINGS. 252, ELSEVIER SCIENCE SA, 01/12/2021. ISSN 0378-7788, ISSN 1872-6178. Q1.
- Gamero-Salinas, Juan; Kishnani, Nirmal; Monge-Barrio, Aurora; López-Fidalgo, Jesús; Sánchez-Ostiz, Ana. **Evaluation of thermal comfort and building form attributes in different semi-outdoor environments in a high-density tropical setting**. BUILDING AND ENVIRONMENT. 205, PERGAMON-ELSEVIER SCIENCE LTD, 01/11/2021. ISSN 0360-1323, ISSN 1873-684X. Q1.

- Belar, Alazne; Martínez, Marina; Centeno, Carlos; López-Fidalgo, Jesús; Santesteban, Yolanda; Lama, Marcos; Arantzamendi, Maria. **Wish to die and hasten death in palliative care: a cross-sectional study factor analysis**. BMJ SUPPORTIVE & PALLIATIVE CARE. BMJ PUBLISHING GROUP, 14/10/2021. ISSN 2045-435X, ISSN 2045-4368. Q2.
- Berni, Rossella; López Fidalgo, Jesús Fernando; Vining, Geoff G. **Foreword to the Special Issue on "Statistics, Statistical Engineering, and Innovation for Industry 4.0"** FOREWORD. APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY. 37 - 5, pp. 837 - 838. 23/09/2021. ISSN 1524-1904, ISSN 1526-4025. Q3.
- Arias-Casais, Natalia; López-Fidalgo, Jesús; Garralda, Eduardo; Jose Pons, Juan; Rhee, John Y.; Lukas, Radbruch; de Lima, Liliana; Centeno, Carlos. **Trends analysis of specialized palliative care services in 51 countries of the WHO European region in the last 14 years**. PALLIATIVE MEDICINE. 34 - 8, pp. 1044 - 1056. SAGE PUBLICATIONS LTD, 10/06/2020. ISSN 0269-2163, ISSN 1477-030X. Q1.
- Clark, David; Baur, Nicole; Clelland, David; Garralda, Eduardo; López-Fidalgo, Jesús; Connor, Stephen; Centeno, Carlos. **Mapping Levels of Palliative Care Development in 198 Countries: The Situation in 2017**. JOURNAL OF PAIN AND SYMPTOM MANAGEMENT. 59 - 4, pp. 794 - +. ELSEVIER SCIENCE INC, 01/04/2020. ISSN 0885-3924, ISSN 1873-6513. Q2.
- García-Rodenas, Ricardo; Carlos García-García, Jose; López-Fidalgo, Jesús; Angel Martin-Baos, Jose; Wong, Weng Kee. **A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs**. COMPUTATIONAL STATISTICS & DATA ANALYSIS. 144, ELSEVIER, 01/04/2020. ISSN 0167-9473, ISSN 1872-7352. Q2.
- Moreno-Ajona, David; Irimia, Pablo; Rodríguez, Jose Antonio; García-Velloso, Maria Jose; López-Fidalgo, Jesús; Fernandez-Alonso, Leopoldo; Grochowitz, Lukasz; Munoz, Roberto; Dominguez, Pablo; Gallego-Cullere, Jaime; Martínez-Vila, Eduardo. **Elevated circulating metalloproteinase 7 predicts recurrent cardiovascular events in patients with carotid stenosis: a prospective cohort study**. BMC CARDIOVASCULAR DISORDERS. 20 - 1, BMC, 26/02/2020. ISSN 1471-2261. Q3.
- Sánchez-Cárdenas, M. A., Garralda, E., Arias-Casais, N. S., Benitez, E., Van Steijn, D., Moine, S., ... & Centeno, C. (2021). **Palliative care integration indicators: An European regional analysis.** BMJ Supportive & Palliative Care. Q2.
- Aguilar, C., Serna-Jiménez, J., Benitez, E., Valencia, V., Ochoa, O., & Sotelo, L. I. (2021). **Influence of high power ultrasound on natural microflora, pathogen and lactic acid bacteria in a raw meat emulsion**. Ultrasonics Sonochemistry, 72, 105415. Q1.

- Balaguer, Á., Benítez, E., De la Fuente, J., & Osorio, A. (2021). **Maternal and paternal parenting styles as a whole: validation of the simple form of the parenting style evaluation scale**. Anales de Psicología/Annals of Psychology, 37(1), 77-87. Q3.
- Balaguer, Á., Benítez, E., Albertos, A., & Lara, S. (2020). **Not everything helps the same for everyone: relevance of extracurricular activities for academic achievement**. Humanities and Social Sciences Communications, 7(1), 1-8. Q2.
- Sánchez-Cárdenas MA, Garralda E, Benítez E, Arias-Casais N, van Steijn D, Centeno C. **Palliative Care Coverage across European National Health Systems: Proposal of a Synthetic Indicator**. International Journal of Environmental Research and Public Health. 2021; 18(20):10753. Q2.
- Balaguer, Á., Benitez, E., de la Fuente, J., & Osorio, A. (2021). **Structural empirical model of personal positive youth development, parenting, and school climate**. Psychology in the Schools. Q3.
- Calatrava, M., de Irala, J., Osorio, A., Benitez, E., & López-del Burgo, C. (2021). **Matched and Fully Private? A New Self-Generated Identification Code for School-Based Cohort Studies to Increase Perceived Anonymity**. Educational and Psychological Measurement, 00131644211035436. Q2.
- Albertos, A., Koning, I., Benítez, E., & De Irala, J. (2021). **Adolescents' Alcohol Use: Does the Type of Leisure Activity Matter? A Cross-National Study.** International journal of environmental research and public health, 18(21), 11477. Q1.
- Martínez, I., Viles, E., & Olaizola, I. G. (2021). **Data Science Methodologies: Current Challenges and Future Approaches**. Big Data Research, 24, 100183. Q1.
- Otamendi, U., Martínez, I., Quartulli, M., Olaizola, I. G., Viles, E., & Cambarau, W. (2021). **Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules**. Solar Energy, 220, 914-926. Q2.
- I. Martínez, E. Viles and I. G. Olaizola, **A survey study of success factors in data science projects**, *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 2313-2318, doi: 10.1109/BigData52589.2021.9671588. (Sin cuartil)
- Bejines, C.; Ardanza-Trevijano, S.; Elorza, J. **On Self-Aggregations of Min-Subgroups**. Axioms 2021, 10, 201. Q1.
- Bejines, C; Ardanza-Trevijano, S; Chasco, MJ; Elorza, J. **Aggregation of indistinguishability operators**. Fuzzy Sets and Systems, 2021. Q1.

1.4 INDUSTRIAL DOCTORAL STUDENTS

In the 2020-2021 academic year, seven industrial doctoral students from BBVA began their industrial dissertations at the University of Navarra. Each has a supervisor from the bank itself and a supervisor at the University of Navarra.

These dissertations are being carried out as part of the following doctoral programs: Natural and Applied Sciences (3 doctoral students), Applied Engineering (1) and Economics and Business Administration (3).

	FIRST NAME AND SURNAMES	DOCTORAL PROGRAM	DISSERTATION TITLE
	PALOMA MARÍN MARTÍNEZ	Natural and Applied Sciences	Predicción de rentabilidad y flujo adverso en mercados financieros y plataformas multi-contribuidas
	JESÚS SALVADOR RENERO QUINTERO	Applied Engineering	Adding Causal Inference to Reinforcement Learning learning methods
	JUAN JOSÉ FERNÁNDEZ TEBAR	Economics and Business Administration	Multi-product approach to pricing and customer-relationship management
	CHRISTIAN OJEDA TREJO	Natural and Applied Sciences	Algorithmic market-making strategies with stochastic optimization techniques and reinforcement learning
	AINHOA GUERRERO SAN MARTÍN	Natural and Applied Sciences	Sistemas de Recomendación Conversacionales y su aplicación en el ámbito financiero
	CONRADO GARCÍA MONTIEL	Economics and Business Administration	"Multi-product dynamic pricing strategies with unlimited inventories"
	GÖKNIL PENCIK	Economics and Business Administration	Reshaping Quantitative Risk Management Metrics in Financial Industries with Deep Learning Applications

In the 2021-2022 academic year, four other students are scheduled to begin their dissertations.

1.5 MEETINGS ORGANIZED BY DATAI

In 2020 and 2021, DATAI organized the following activities:

- **1st Advisory Committee Meeting:** December 14, 2020. Representatives from almost all centers attended. Director Jesús López Fidalgo gave a presentation on the Institute. Following that, proposals were made to further DATAI's growth.
- **1st Scientific Conference:** December 16, 2020. Iciar Astiasarán, Vice President for Research, opened the conference. Afterwards, several researchers from centers such as CIMA, TECNUN, CUN, the School of Law, the School of Economics, DATAI and CEIT gave presentations.
- **Visit from Carme Artigas,** Secretary of State for Digitalization and Artificial Intelligence, and Laura Flores, General Director of AI: March 23, 2021.
- **Meeting with the team of Guzmán Garmendia, Director General of Telecommunications and Digitalization** (DGTT) of the Regional Government of Navarre: June 24, 2021.

1.6 SEMINARS

In 2020 and 2021, DATAI held 17 seminars. They were held every two weeks by professionals from the University's schools, as well as guest speakers:

DATA SCIENCE APPLICATIONS IN CARDIOLOGY

ALBERTO GARCÍA GALINDO

UNIVERSIDAD DE SALAMANCA –DATAI

COMPUTING CONFIDENCE INTERVALS FROM MASSIVE DATA VIA PENALIZED QUANTILE SMOOTHING SPLINES

ENRIQUE DEL CASTILLO

UNIVERSIDAD DE PENNSYLVANIA-COMITÉ CIENTÍFICO DATAI

SPACE FILLING: WEAK COVERING AND QUANTIZATION OF HIGH DIMENSIONAL SETS

JACK NOONAN

SEMINAR 07 — 01.03 2021

SEMINAR 08 — 10.03 2021

SEMINAR 09 — 24.03 2021

SEMINAR 10 — 28.04 2021

SEMINAR 11 — 12.05 2021

SEMINAR 12 — 26.05 2021

THE DIGITAL NEWS REPORT SURVEY. METHODOLOGY, CONSULTATION AND APPLICATIONS

SAMUEL NEGREDO BRUNA

SCHOOL OF COMMUNICATION

FLORENCE NIGHTINGALE: A ROLE MODEL IN THE 21ST CENTURY

GUADALUPE GÓMEZ MELIS

UPC

FAIR MACHINE LEARNING: AN APPROACH BASED ON OPTIMAL TRANSPORT

PAULA GORDALIZA

BCAM

METHODOLOGY AND MANAGEMENT OF APPLIED RESEARCH IN INDUSTRY: THE WORK OF THE DOCTOR OF ENGINEERING

JESÚS MONGE

BAYESIAN SPATIAL CONDITIONAL OVERDISPERSION MODELS FOR COUNT DATA: APPLICATIONS TO INFANT MORTALITY AND TO COVID-19 INCIDENCE

MABEL MORALES OTERO

UPV / EHS

DIFFERENTIAL REPLICATION AS A TOOL FOR MACHINE LEARNING ACCOUNTABILITY IN PRACTICE

IRENE UNCETA MENDIETA

ESADE. RAMÓN LLUL UNIVERSITY

SEMINARIO 13 — 09.06 2021

SEMINARIO 14 — 15.09 2021

SEMINARIO 15 — 19.10 2021

SEMINARIO 16 — 15.11 2021

SEMINARIO 17 — 22.12 2021

SEMI-OUTDOOR SPACE AS A STRATEGY TO ADAPT TO THE RISK OF OVERHEATING IN A TROPICAL, HIGH URBAN DENSITY CONTEXT. SOCIAL AND ENVIRONMENTAL BENEFITS

JUAN CARLOS GAMERO SALINAS

PREDOCTORAL RESEARCHER - SCHOOL OF ARCHITECTURE (UNIV. OF NAVARRA)

A BAYESIAN NONPARAMETRIC MODEL FOR CLASSIFICATION OF LONGITUDINAL PROFILES

ROLANDO DE LA CRUZ

UIA

2. TRANSFER AND INNOVATION

TRANSFER AND INNOVATION

DATAI collaborates with a number of research groups, university centers, companies and other institutions. The cross-cutting and interdisciplinary nature of the team of researchers and experts in data science, statistics and artificial intelligence allows us to work on data projects with high analytical complexity and provide value in each phase.

2.1 INTERNAL COLLABORATIONS

COLLABORATIONS AND PROJECTS

Scientific collaborations are carried out between members of the Institute and researchers from other research groups at the University and external groups that perform significant scientific work and participate in research projects funded by competitive calls for proposals, contracts and other types of calls.

CONSULTANCY AND SERVICES

The Institute also offers **advice** and support for ad hoc work.

In the 2020-2021 academic year, the Institute received 24 requests to provide consultancy services (either from a group or department at the University of Navarra, a research group from another university or center, or from a company or public/private institution).

Several members of the Institute have been involved in these collaborations in different professional categories: analyst, senior analyst and expert.

Some of the centers and entities requesting these collaborations: ICS (Atlantes, EASH, Public Discourse, Creativity), the School of NURSING, the School of COMMUNICATION, the School of PSYCHOLOGY, Clínica Universidad de Navarra (CUN), the ADMISSIONS office, the OFFICE OF THE VICE PRESIDENT OF STUDENTS and the ALUMNI ASSOCIATION.

2.2 COLLABORATIONS WITH INDUSTRY

La misión del DATAI es funcionar como el motor de investigación y educación traslacional en Big Data, Ciencia de Datos e Inteligencia Artificial desarrollando proyectos de alto impacto en la industria. Para lograrlo, el Instituto colabora activamente con socios de la industria en un amplio espectro para desarrollar interacciones mutuamente beneficiosas.

2.3 STRATEGIC PARTNERSHIPS

The University of Navarra and BBVA are strategic partners in the field of advanced data analytics and artificial intelligence. They collaborate to train their employees and drive data science research.

This collaboration was made official in July 2020 during a virtual event attended by Iciar Astiasarán, the Vice President for Research, and Jesús López Fidalgo, the Director of the Institute of Data Science and Artificial Intelligence; and, from BBVA, Carlos Casas, Head of Talent and Culture, and Ricardo Martín, Head of Data.

COLLABORATION AGREEMENT BETWEEN THE UNIVERSITY OF NAVARRA AND BBVA

This agreement focuses mainly on three collaboration areas: industrial dissertations, mentioned above, a group within the master's degree, which will be mentioned in the next chapter, and professional certification for data scientists:

PROFESSIONAL CERTIFICATION FOR DATA SCIENTISTS

This globally recognized certification, issued by the University of Navarra and BBVA, is awarded to BBVA professionals who complete the Data Scientist Fundamentals program, one of the training programs offered by the BBVA Campus, and who complete at least 400 hours of work on the bank's big data platform. The work is supervised by academics from various centers at the University with proven experience in advanced data analytics.

The mission of the certification is to provide a standard in the field of data science through which organizations can identify professionals with skills in the field of advanced analytics. In addition, the certification helps strengthen the credibility and visibility of the profession of data scientist.

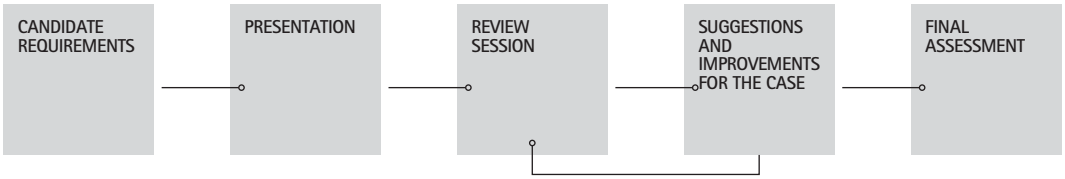
The process consists of three distinct phases:

a. Preparatory training
b. Implementation of a use case
c. The certification exam

In the 2020-2021 academic year, 45 candidates applied for professional certification.

a. Use case review

The use case review is a process to assess data science projects carried out by candidates as part of their work at BBVA to obtaining certification. The procedure is as follows:



b. The certification exam

The certification exam measures performance in five specific areas: formulation of problems in the framework of data science, machine learning, statistics, mathematics and programming. The exam assesses general knowledge in these areas and application in science projects.

BBVA makes the most of this close relationship with the University of Navarra to attract talent on campus. These students will have access to job listings for positions at the bank. Meanwhile, data scientists in the group with an interest in academia will be able to teach classes and perform other teaching activities at the University of Navarra.

2.4 *SURVEY AND EXPERIMENT INCUBATOR* CONTEST

This contest is organized by the Institute of Data Science and Artificial Intelligence (DATAI) of the University of Navarra and the Department of Statistics, Computer Science and Mathematics of the Universidad Pública de Navarra (UPNA), with the collaboration of the Navarran Society of Mathematics Teachers TORNAMIRA; the Navarre Institute of Statistics; the Department of Education of the Regional Government of Navarre; and the provincial delegation of the National Institute of Statistics.

The contest is held annually and is aimed at high school students and students carrying out basic or intermediate vocational training in Navarre who have an interest in statistics and data science. The participants must carry out a statistics and probability project on a topic of interest chosen by the team, such as society, health, ecology, economics and consumption and the environment. The contest has a local phase, the winners of which go on to participate in the national phase along with the winning teams from other provinces.

The aim of this contest is:

- To promote the teaching and learning of statistics at non-university educational levels.
- To complement knowledge and skills and to promote the importance and usefulness of statistics, probability and operations research in real life and in all academic disciplines.
- To generate an interest in statistics, probability and operations research among students as a fundamental tool in virtually all sciences.

- To familiarize students with the different stages of a statistical project.
- To teach students how to interpret the different statistical information that surrounds us and to encourage them to adopt a critical spirit.

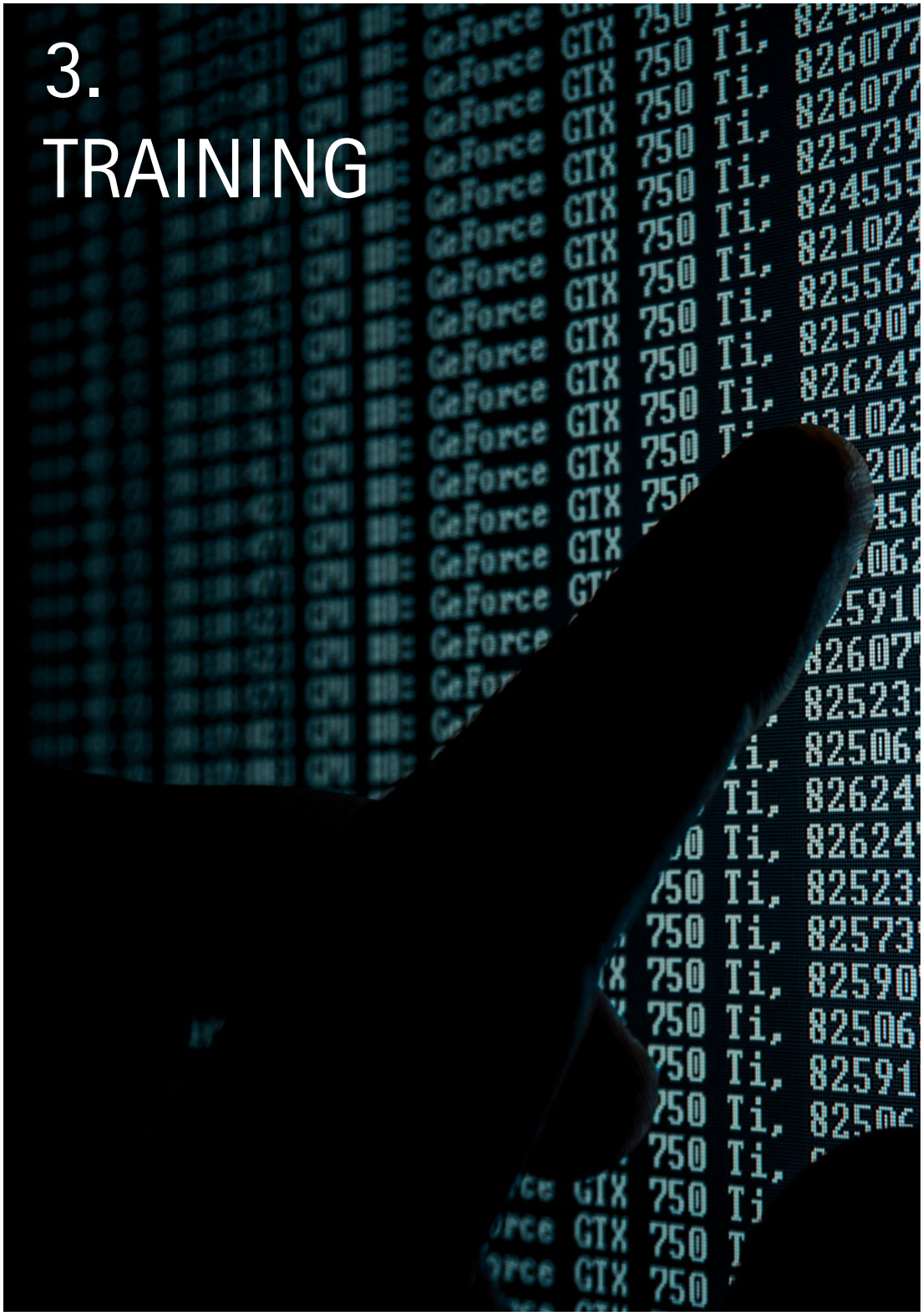
On June 18, 2021, the final Navarran phase of the 2020-2021 Survey and Experiment Incubator contest was held online. A total of 139 students from nine different centers participated in the program, including high school and vocational training students. This contest seeks to awaken students' curiosity for data science as a basic tool for research.

Jesús López Fidalgo, Director of the Institute of Data Science and Artificial Intelligence, began the session with an informative talk in which an online contest was also held.

The participants presented projects on statistics, probability and operations research that explored aspects such as career paths, social networks, the effects of Covid-19 in the areas of health and education, video games and the living conditions of homes.

Sponsors:

3. TRAINING



TRAINING

DATAI's training activities in the 2020-2021 academic year included the university-accredited Master's Degree in Big Data Science, micro-courses, in-company programs and other programs.

3.1 MASTER'S DEGREE IN BIG DATA SCIENCE

The Master's Degree in Big Data Science is a university-accredited degree that was created to meet the demands of companies and organizations for knowledge of big data and data science.

It offers a technical and specific but highly practical program.

The teaching staff members are all from prestigious companies and institutions and have extensive experience in these tools. Students are taught how to use these tools in real-life situations and how to apply them in various fields.

The Executive format, which has on-campus sessions on Friday afternoons and Saturday mornings, makes it easier for students to combine their professional work with class attendance, and encourages networking between professionals and companies.

In the 2017-2018 and 2018-2019 academic years, a big data specialist program was taught on the Pamplona campus (in the Siemens Gamesa classroom at the ICS). In the 2019-2020 academic year, the master's degree became a university-accredited master's degree and, since then, has been taught entirely on the Madrid postgraduate campus (Alumni Building).

In the 2020-2021 academic year, there were two groups:

- 39 students in an open group
- 25 students in the BBVA group

The main employers of our students:

Master's Thesis Projects have been carried out at companies such as BBVA, PwC, Deloitte and EY.

3.2 MICRO-COURSES

The need has become increasingly relevant for professionals in many disciplines to enter the job market, including the field of research, with knowledge and skills in handling data for value creation.

In the 2020-2021 academic year, therefore, the aim of these programs was to allow professionals with any degree, even those in clearly related areas such as mathematics and statistics, to refresh their knowledge and gain some measure of how to use the main methodologies and tools available in the market to solve problems associated with data management and analysis.

These programs are carried out online as an advanced training course.

The Institute of Data Science and Artificial Intelligence (DATAI) provides advanced training courses (2 ECTS credits each) in:

- Data Management with Python
- Data Management with R
- Data Science Statistics
- Machine Learning
- Data Visualization
- Spreadsheets for Data Analysis
- Social Network Analysis

To offer more comprehensive training, the individual courses can be grouped into the following advanced training courses, each worth 2 ECTS credits:

Advanced Training Course in Data Management Tools TOTAL: 6 ECTS CREDITS	DATA MANAGEMENT WITH PYTHON DATA MANAGEMENT WITH R SPREADSHEETS FOR DATA ANALYSIS
Advanced Training Course in Data Science Techniques TOTAL: 6 ECTS CREDITS	DATA SCIENCE STATISTICS MACHINE LEARNING DATA VISUALIZATION SOCIAL NETWORK ANALYSIS

The courses are delivered in livestreaming. Each course lasts three weeks, with two hours of theory and practical classes followed by one hour of tutoring. All explanatory sessions will be conducted in Zoom, recorded and made available to students. Complementary materials will include academic articles and theoretical and applied scientific communication videos.

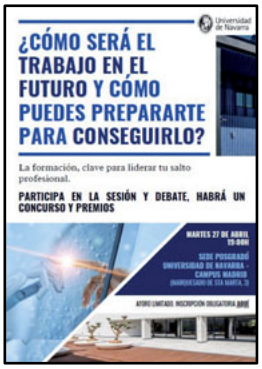
They are aimed at university students and professionals interested in exploring the field of data science and with basic knowledge of computers and spreadsheets.

Micro-courses taught in the 2020-2021 academic year

- **Data Science Statistics**
- **Social Network Analysis**
- **Data Management with Python**

  Due to high demand, this micro-course was offered twice. The first course was aimed at students and the second at university professionals and doctoral students.

3.3 EVENTS AND ACTIVITIES

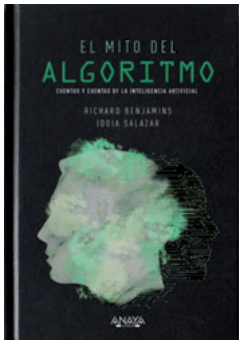


WORKSHOP

The future of DS and AI. Aimed at students and academic staff of the Master's Degree in Big Data Science.
On-campus program.
April 16, 2021

CURSO FORMATIVO

Introductory Course to GIT. HERIBERT VALERO (Data scientist at Ferrovial).
Online.
March 15 and 22, 2021



PRESENTACIÓN DEL LIBRO

El Mito del Algoritmo. With the participation of the authors, RICHARD BENJAMINS and IDOIA SALAZAR, and JESÚS LÓPEZ. Jesús López Fidalgo acted as moderator. On-campus and streaming. February 25, 2021

WORKSHOP

Data Visualization. SOLEDAD GUARCH (Senior Digital Strategy Analyst, British Airways).
On-campus program.
February 19, 2021

WEBINAR

Data Visualization Strategies. Panel discussion with MIGUEL ÁNGEL CASARES (Adjunct Professor, University of Navarra); MARÍA MANSO (Global Head of People Analytics, BBVA); OMAR ARCE TORREBLANCA (Business Analytics Manager, Axesor). On-campus and streaming.
February 19, 2021

INFORMATION SESSION

Entrevistas y procesos de selección de Data Science / Analytics. MERCEDES SÁNCHEZ LOR (Technology Manager at the recruitment company Page Personnel). On-campus program. February 5, 2021

INFORMATION SESSION AND DISCUSSION

What will the job of the future be like and how can we prepare for it? On-campus program. April 27, 2021

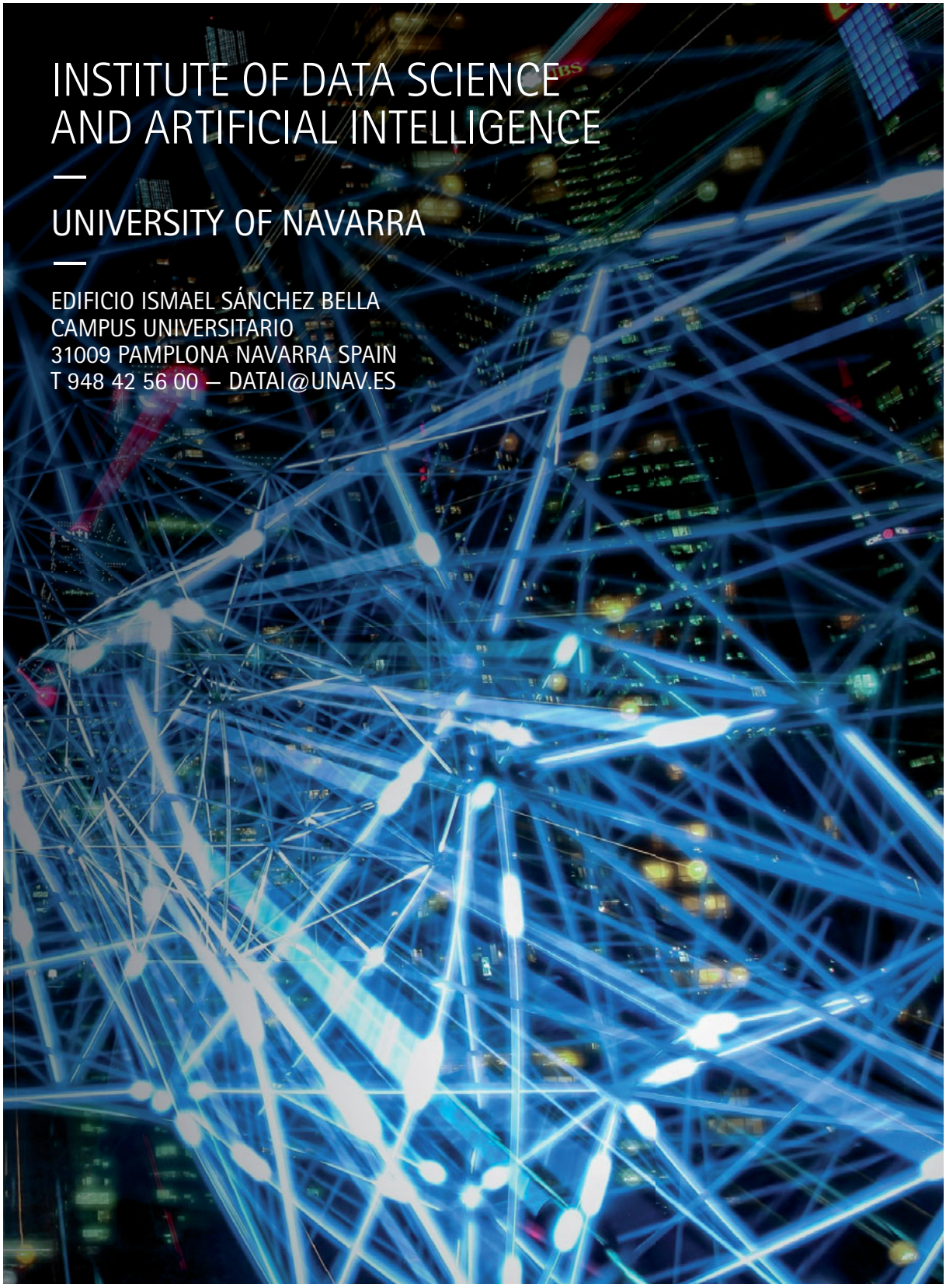
BOOTCAMP

Salidas profesionales Data Science / Big Data. MARTA ARDÍZONE (Talent Brand Manager at ETS Factory); KRISTINA MONSON (Deputy Director of Career Services, University of Navarra); MATÍAS ÁVILA CLEMENTE (Data Scientist at BBVA Data & Analytics). On-campus and streaming. October 29, 2020

3.4 IN-COMPANY PROGRAMS

A schedule of tailor-made courses, workshops and seminars for companies, research groups, services and centers looking to update their knowledge and specialize in data analysis topics related to real needs.

It may also be possible to design introductory and specialized software courses.



INSTITUTE OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

—

UNIVERSITY OF NAVARRA

—

EDIFICIO ISMAEL SÁNCHEZ BELLA
CAMPUS UNIVERSITARIO
31009 PAMPLONA NAVARRA SPAIN
T 948 42 56 00 — DATAI@UNAV.ES

Universidad de Navarra | DATAI
INSTITUTO DE CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL