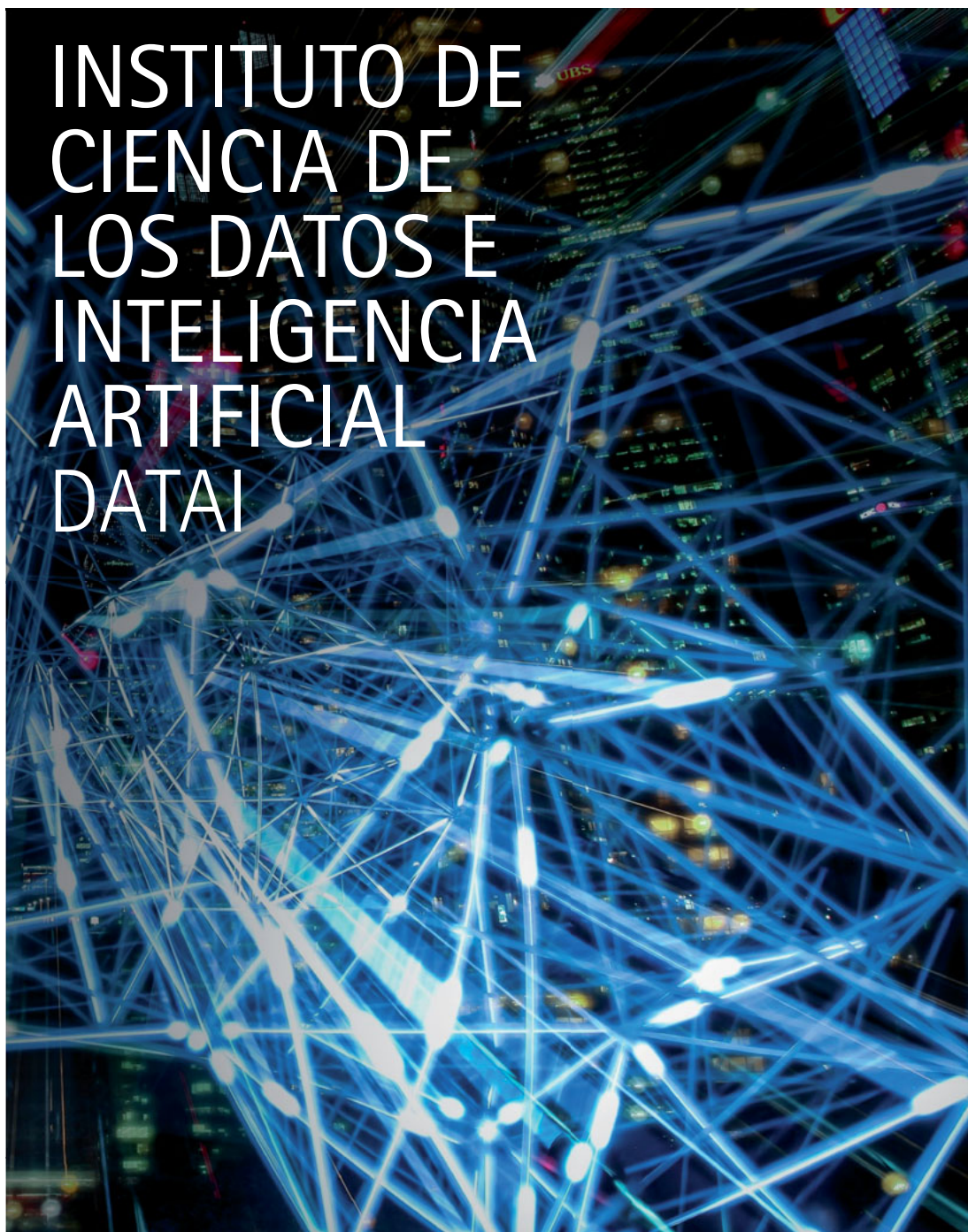

MEMORIA ACTIVIDADES. AÑOS 2020 Y 2021



Universidad
de Navarra

**DATAI ES UN CENTRO
DE INVESTIGACIÓN
TRANSVERSAL, DE
INNOVACIÓN Y FORMACIÓN
DE LA UNIVERSIDAD DE
NAVARRA QUE BUSCA UN
TRABAJO COLABORATIVO
DE DIFERENTES GRUPOS
Y PERSONAS DE LA
UNIVERSIDAD CON EL
OBJETIVO DE ALCANZAR
LA EXCELENCIA EN
CIENCIA DE DATOS E
INTELIGENCIA ARTIFICIAL**

CONTENIDOS

0. INTRODUCCIÓN	04
1. INVESTIGACIÓN	07
1.1 Comité Científico	08
1.2 Áreas de Investigación	10
1.3 Publicaciones	21
1.4 Doctorados industriales	25
1.5 Reuniones organizadas	26
1.6 Seminarios	26
2. TRANSFERENCIA E INNOVACIÓN	28
2.1 Colaboraciones internas	29
2.2 Empresa y entidades	30
2.3 Acuerdos estratégicos	30
2.4 Concurso Incubadora de sondeos y experimentos	32
3. FORMACIÓN	34
3.1 Máster Oficial en Big Data Science	35
3.2 Microcursos	36
3.3 Eventos y Actividades	38
3.4 Programas In Company	40

0. INTRODUCCIÓN



INTRODUCCIÓN

El 11 Diciembre de 2019 se puso en marcha el Instituto de Ciencia de los Datos e Inteligencia Artificial (en adelante DATAI). Es un centro de investigación transversal, de innovación y formación de la Universidad de Navarra que busca un trabajo colaborativo de diferentes grupos y personas de la Universidad con el objetivo de alcanzar la excelencia en Ciencia de Datos e Inteligencia Artificial.

El instituto, con dependencia orgánica de TECNUN y sede en Pamplona, pretende actuar como órgano transversal que aglutina proyectos de análisis estadístico y procesamiento de datos que ya se están realizando en distintas facultades, grupos y servicios de la Universidad. Nace con una clara vocación de colaboración e integración de la actividad investigadora en Big Data e Inteligencia Artificial y de formación.

La junta directiva está formada por su Director D. Jesús López Fidalgo, tres subdirectores: Dña. Elisabeth Viles (Tecnun - Escuela de Ingeniería), Dña. Stella Maris Salvatierra (Facultad de Económicas) y D. Sergio Ardanza-Trevijano Moras (Facultad de Ciencias); el Gerente D. Fernando de la Puente García-Ganges y el secretario D. José María González Gullón.



JUNTA DIRECTIVA DEL MÁSTER EN LA PRESENTACIÓN DEL INSTITUTO (FEBRERO 2020)

DATAI tiene miembros en todos los campus de la universidad: Pamplona, Madrid, San Sebastián y Barcelona. Además de las distintas facultades y escuelas de la Universidad de Navarra, el instituto trabaja con personal de los centros de investigación y asistencia como son el CEIT, el CIMA y la CUN, entre otros.

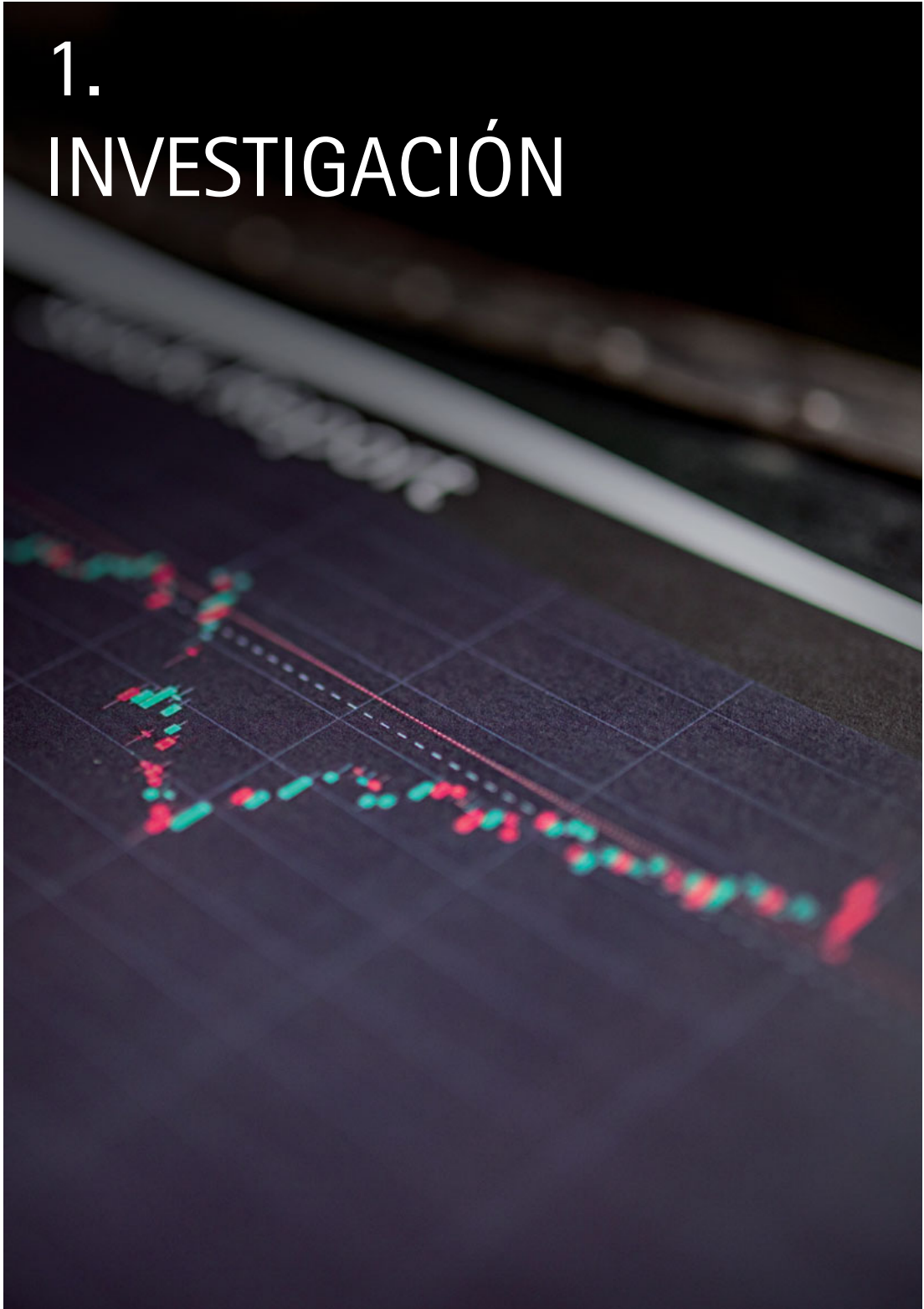
Muchos de los investigadores externos de empresas como BBVA y HSBC Banca Internacional también participan en las labores de formación y docencia del instituto.

Entre sus objetivos existe un importante apoyo en la **investigación**, así como en **innovación y transferencia** de conocimientos al ámbito industrial, empresarial y social. También existe una fuerte implicación en la **formación** de investigadores, estudiantes y profesionales. Por esta razón, el Instituto dirige un Máster oficial en Big Data Science destinado principalmente a jóvenes profesionales, así como estudiantes de doctorado, en diferentes áreas con necesidades importantes de análisis de datos.

El **Consejo Asesor** del instituto está formado por representantes de todas las facultades y escuelas de la universidad: *Arquitectura, Ciencias, Comunicación, Derecho, Económicas, Educación y Psicología, Enfermería, Farmacia y Nutrición, Filosofía y Letras, Medicina, Teología*. También tiene representantes de los institutos *CEIT, CIMA y CUN*, así como de instituciones externas como *BBVA*.

FACULTADES Y ESCUELAS	REPRESENTANTE
ARQUITECTURA	EDUARDO BAYO PÉREZ
CIENCIAS	DIEGO MAZA OZCOIDI
COMUNICACIÓN	IDOIA PORTILLA MANJÓN
DERECHO	JUAN CARLOS HERNÁNDEZ PEÑA
ECONÓMICAS	ALEJO SISON GALSIM
EDUCACIÓN Y PSICOLOGÍA	MAITE AZNÁREZ SANADO
ENFERMERÍA	MARÍA PUEYO GARRIGUES
FARMACIA Y NUTRICIÓN	IÑAKI FERNÁNDEZ DE TROCÓNIZ
FILOSOFÍA Y LETRAS	JUAN JOSÉ PONS IZQUIERDO
MEDICINA	ALFREDO GEA SÁNCHEZ
TECNUN	ÁNGEL RUBIO DÍAZ-CORDOVÉS
TEOLOGÍA	RUBÉN HERCE
INSTITUTOS Y OTROS CENTROS	
CEIT	SAIOA ARRIZABALAGA JUARISTI
CIMA	MIKEL HERNÁNDEZ ARRAZOLA
CUN	MARÍA A. FERNÁNDEZ SEARA
IESE	SAMPESA SAMILA
INSTITUCIONES EXTERNAS	
BBVA	MARCO BONILLA

1. INVESTIGACIÓN



INVESTIGACIÓN

El Instituto promueve y difunde los avances en la investigación en materia de Ciencia de los Datos, Big Data e Inteligencia Artificial. Aspira a proporcionar un foro de alta calidad para investigadores, profesionales y formuladores de políticas a nivel local, nacional e internacional.

La investigación, desde la fundación del Instituto, tiene un enfoque transversal, donde colaboran expertos de distintas áreas, así como profesionales del ámbito de la Ciencia de los Datos, Big Data e Inteligencia Artificial, así como de la estadística y la informática.

El Instituto busca realizar una investigación de calidad con transferencia a la sociedad y al sector empresarial.

1.1 COMITÉ CIENTÍFICO

DATAI cuenta con un **Comité Científico** formado por 5 destacados miembros en el ámbito de la Ciencia de los Datos: Dña. Amparo Alonso Vetanzos, ex -presidenta de la AEPIA (Spanish Association for Artificial Intelligence) y miembro del Consejo Superior Español de Inteligencia Artificial, D. Enrique del Castillo, distinguished professor de la Universidad de Pensilvania, Dña. Nuria Oliver, Directora de Ciencia de Datos de Vodafone, D. John Stufken, Director de Informática y Analítica del Banco de América y Trevor Hastie, miembro honorífico procedente del Departamento de Estadística y Departamento de Ciencia de Datos Biomédicos de la Universidad de Stanford.



Amparo Alonso Betanzos

President of AEPIA (Spanish Association for Artificial Intelligence)

Full Professor of Computational Science

→ Ver CV



Enrique del Castillo

Distinguished Professor of Industrial Engineering and Professor of Statistics

The Pennsylvania State University

→ Ver CV



Nuria Oliver

PhD Director of Research in Data Science

Vodafone Chief Data Scientist | Data-Pop Alliance

→ Ver CV



John Stufken

Director, Informatics and Analytics

Bank of America Excellence Professor | UNC Greensboro

→ Ver CV



Trevor Hastie (Honorary Member)

Professor, Department of Statistics and Department of Biomedical Data Science

Stanford University

→ Ver CV

1.2 ÁREAS DE INVESTIGACIÓN

DATAI cuenta con 73 miembros entre adscritos (12), asociados (43) e invitados (18); Están organizados en 9 áreas de investigación que conforman DATAI. A su vez, cada área dispone de sus propias líneas de investigación.

1.2.1 LÍNEAS DE INVESTIGACIÓN ESTRATÉGICAS

1

Diseño Óptimo de experimentos: Estudios de modelización utilizados en diversas áreas, tales como crecimiento de tumores, ensayos clínicos en las tres primeras fases, retención de la radiación en el cuerpo humano o farmacocinética. Submuestreo inteligente (active learning) en Big Data. Aplicación de las técnicas de diseño óptimo en la elección de los parámetros iniciales de un algoritmo y en experimentos computacionales (in silico). Medicina personalizada.

2

Redes complejas y análisis de redes sociales: Segmentación de redes y detección de nodos influyentes o singulares. Para ello se estudian índices de medición de las propiedades de la red y de sus nodos.

3

Machine Learning en Bioinformática y Neurociencia Computacional: Tratamiento de gran número de variables que generan los análisis genéticos, habitualmente con pocos sujetos. En particular se buscará detector interacciones fiables de los genes utilizando clasificadores basados en redes jerárquicas bayesianas. Expresión micro RNA en pacientes con diversas patologías. Identificación de biomarcadores en cáncer colorectal. Clasificación automática de neuronas.

4

El procesamiento del lenguaje natural y análisis del sentimiento. Principalmente mediante Deep Learning y regularización de parámetros.

5

Fair learning Modelización y desarrollo de algoritmos no sesgados por prejuicios o valoraciones injustas de determinadas variables. El problema fundamental al desarrollar un modelo justo es que muchas variables contienen una parte que lleva al modelo a utilizar prejuicios sociales para tomar decisiones, pero al mismo tiempo esas variables tienen una información importante para tomar decisiones. Para diferenciar ambas cosas se utilizan modelos de path analysis en el contexto de la inferencia causal, que ayudan a extraer de cada variable la parte más informativa que no produzca sesgos.

1

DISEÑO ÓPTIMO DE EXPERIMENTOS

Modelización en diversas áreas, tales como crecimiento de tumores, ensayos clínicos en las tres primeras fases, retención de la radiación en el cuerpo humano o farmacocinética. Esta es una línea del equipo nuclear del instituto, que lleva en marcha más de 30 años y en la que se han obtenido resultados relevantes. Por tanto, siendo línea estratégica, necesita en estos momentos acabar de formar un equipo sólido de trabajo en DATAI. Ya existe un grupo consolidado, especialmente en las universidades de Salamanca y Castilla La-Mancha, pero que incluye también investigadores de la Universidad Pública de Navarra, Almería o La Laguna. Más recientemente, dentro de esta línea se está haciendo un esfuerzo por aplicar las técnicas de diseño óptimo de experimento a los problemas esenciales que están surgiendo en Ciencia de Datos, Big Data e Inteligencia artificial. En concreto:

- A. Active learning** (Submuestreo inteligente) en big data. Uno de los problemas importantes del tratamiento de grandes cantidades de datos es el coste computacional de los análisis y ajustes de modelos, que en algunos aspectos crece exponencialmente convirtiéndolos en problemas NP. Una técnica habitual, basada en el teorema central del límite o en las leyes de los grandes números, consiste en tomar una submuestra aleatoria y hacer los cálculos con una cantidad razonable de datos. Esta línea de investigación busca mejorar el método de muestreo, de modo que se capturen los datos más informativos dentro de la muestra grande. Esto adquiere especial relevancia cuando la distribución de las variables explicativas es altamente asimétrica o con curtosis extrema. Este último es un tema apenas estudiado en la literatura de modo formal.
- B. Algoritmos verdes.** El incremento de la capacidad computacional, junto con la posibilidad de recoger grandes cantidades de datos, nos ha llevado a desarrollar e implementar algoritmos que requieren mucho tiempo de computación, con el consiguiente gasto de energía. En parte, los llamados algoritmos verdes buscan optimizar ese coste energético mediante algoritmos más eficientes que requieran menos recursos de ejecución. Las técnicas de diseño óptimo pueden aplicarse a la elección óptima de los parámetros iniciales de un algoritmo, de modo que su velocidad de convergencia sea mayor y así reducir el tiempo de computación.
- C. Experimentos in silico.** También llamados computer experiments, son cada vez más frecuentes a través de los llamados modelos *digital twins*. Son simuladores contruidos con complejos sistemas de ecuaciones en derivadas parciales, que

habitualmente requieren un gran coste computacional. La teoría de **computer experiments** busca simplificar esos modelos de modo que con una mínima pérdida de precisión, puedan conseguirse resultados en un tiempo mucho menor, a la vez que se añade una mayor **explicabilidad** del modelo.

D. Selección de modelos. Típicamente para un mismo fenómeno existen muchos modelos rivales posibles, a lo que se añade una gran cantidad de variables, entre las que hay que seleccionar unas pocas que sean suficientes para explicar la respuesta y hacer predicciones. En este aspecto la discriminación de modelos juega un papel esencial en el que el equipo lleva trabajando durante años con algunas aportaciones muy relevantes.

E. Medicina personalizada. El equipo ha desarrollado diseños experimentales personalizados, de modo que dependiendo de las circunstancias del sujeto se realiza un experimento u otro, en lugar de la aleatorización plena para todos los sujetos. Esto permite obtener mejores ajustes del modelo con menos experimentos. El equipo busca usar esta idea (técnica) para proponer un tratamiento ajustado a cada paciente, teniendo en cuenta sus circunstancias particulares.

2

REDES COMPLEJAS Y ANÁLISIS DE REDES SOCIALES

Los fundamentos de la ciencia de redes (Network Science) se aplica cada vez más para estudiar una variedad de fenómenos del mundo real, muchos de los cuales constituyen sistemas complejos por su naturaleza intrínseca. Investigadores de diversos campos consideran la formulación y estudio de sistemas complejos como un tema crucial. Estos sistemas complejos pueden ser modelados mediante redes (denominadas también grafos), que están representadas por nodos y enlaces (componentes de la red), entre los cuales se establecen interacciones que definen patrones y estructuras útiles complejas de identificar. De manera general, los nodos representan las entidades de la red y los enlaces las conexiones que se establecen estas. Entre los ejemplos de redes complejas podemos mencionar: las estructuras cerebrales, las relaciones sociales, las comunicaciones en la telefonía móvil, las interacciones biológicas moleculares, las transacciones bancarias, las redes de transporte, etc.

En muchos de estos dominios de aplicación, los datos de la red resultan complejos, tanto en el gran volumen de información a procesar como en la variedad de estos, con lo cual, constituye un desafío su análisis. Por tanto, dados los niveles de información, estudiar las estructuras e influencias de los diferentes componentes de una red (nodos y enlaces), induce el estudio de nuevos modelos y técnicas eficientes.

En este contexto, el Análisis de Redes Complejas (Complex Network Analysis) constituye el área de investigación relacionada con el diseño, desarrollo y aplicación de modelos matemáticos-computacionales para la detección y el análisis de las estructuras y patrones subyacentes de interés presentes en estas redes.

Entre las tareas oportunas a estudiar en este tipo de redes cabe mencionar: la detección de grupos de nodos (entidades) con un comportamiento similar (ej. grupos de personas con intereses comunes en una red social, conjunto de proteínas similares en una red biológica), la identificación de nodos influyentes en la red (ej. usuarios en una red social con alto número de seguidores, cuentas con alto número de operaciones en una red de transacciones bancarias), el análisis de la dispersión de "información" (ej. identificación de las entidades de la red que maximizan la dispersión en un escenario epidemiológico), la predicción de futuras relaciones entre los nodos (ej. identificar futuras relaciones de "conveniencia" en una red de empresarios de negocios), la identificación de nodos anómalos (ej. detección de sensores con mediciones atípicas respecto a la norma en unas redes de sensores inalámbricos), etc.

La amplia gama de escenarios en los cuales resulta pertinente la modelación mediante redes, hace del análisis de redes complejas una línea de investigación transversal, resultando de gran aplicabilidad en el contexto actual.

3

MACHINE LEARNING EN BIOINFORMÁTICA Y NEUROCIENCIA COMPUTACIONAL

La medicina traslacional trata de combinar diversas disciplinas científicas para mejorar la prevención, el diagnóstico y el tratamiento de las afecciones clínicas. Es, por definición, un campo altamente interdisciplinar cuyo objetivo principal es mejorar el sistema sanitario mundial. Aunque la mayoría de las disciplinas biomédicas han evolucionado para incorporar dispositivos digitales, los algoritmos que recogen, procesan y analizan los datos son abrumadoramente desconocidos y/o inaccesibles para la mayoría de los profesionales.

Lo que hoy se conoce como la revolución de la medicina digital no es más que la evolución natural de las tecnologías de la información que ayudan a las evaluaciones clínicas y médicas. El aprendizaje automático ya está marcando la diferencia en este contexto.

Un ejemplo destacado es el uso de la neurociencia básica para personalizar las evaluaciones médicas de los pacientes con enfermedades cerebrales. Más concretamente, un programa que lleve la investigación neurocientífica básica hasta los protocolos

hospitalarios, pasando por validaciones clínicas podría lograr el desarrollo de tratamientos personalizados.

En este contexto DATAI planea trabajar en un línea enfocada en el aprendizaje automático en Bioinformática y Neurociencia Computacional. Concretamente se utilizará el tratamiento de gran número de variables que generan los análisis genéticos, habitualmente con pocos sujetos. En particular se buscará detectar interacciones fiables de los genes utilizando clasificadores basados en redes jerárquicas bayesianas. Expresión micro RNA en pacientes con diversas patologías. Identificación de biomarcadores en cáncer colorectal. Clasificación automática de neuronas.

4 FAIR LEARNING

Fairness (equidad en castellano) en el contexto de la IA se refiere a los diversos intentos de corregir el sesgo algorítmico. Aunque las definiciones de imparcialidad son siempre controvertidas, los resultados de un algoritmo pueden considerarse justos si son independientes de una variable determinada, especialmente las consideradas sensibles, como los rasgos de los individuos que no deberían correlacionarse con el resultado (es decir, el género, la etnia, la orientación sexual, la discapacidad, etc.). En el aprendizaje automático, el problema del sesgo algorítmico es bien conocido y estudiado.

Los debates sobre la equidad en el aprendizaje automático tienden a centrarse en el impacto de los diferentes modelos en los grupos socialmente sensibles sin embargo, la imparcialidad es un tema amplio, y es relevante para casi cualquier contexto. ¿A qué usuarios se les muestran determinados anuncios, se les ofrecen determinados precios, obtienen recompensas, obtienen menores tiempos de espera en las llamadas y se les identifica en función de distintos modelos de propensión?

En los últimos años se han propuesto varios criterios de equidad, pero predominan dos enfoques: la **equidad demográfica** (equivalente a la eliminación del impacto dispar) y la **igualdad de oportunidades**.

La equidad demográfica garantiza que cualquier decisión que tome un modelo de IA no esté relacionada con un atributo sensible (por ejemplo, la raza, el sexo o la edad). En otras palabras, ser hombre o mujer no debería determinar si una persona cometerá un crimen.

La igualdad de oportunidades es un poco más sutil: requiere que los individuos que reúnen los requisitos para un buen resultado obtengan ese resultado con la misma probabilidad,

independientemente de que sean miembros del grupo sensible. Por ejemplo, el porcentaje de individuos que reúnen los requisitos para obtener un préstamo y que acaban recibéndolo no debería diferir entre grupos raciales.

La investigación sobre la equidad en el aprendizaje automático es un tema relativamente reciente pero que está cobrando gran relevancia debido a la expansión de la IA y la utilización de los datos. Es un campo transversal que tiene implicaciones éticas y legales y sobre el que la comisión europea ya ha puesto el foco en su estrategia acerca de IA destacando como uno de los siete elementos principales la equidad a la hora de desarrollar modelos de IA. España, dentro de su estrategia de Inteligencia Artificial también remarca la importancia del desarrollo sostenible de la IA.

En este contexto DATAI planea desarrollar una línea de investigación de *fairness* centrada en:

- El estudio de la equidad algorítmica, medición y mitigación de sesgos así como el desarrollo de nuevas herramientas y modelos matemáticos para mejorar la equidad.
- La aplicación del *fairness* para el desarrollo de una IA sostenible y ética para la mejora de la sociedad.

5

PROCESAMIENTO DEL LENGUAJE NATURAL

La Procesamiento del Lenguaje Natural (PLN) es un subcampo de la IA que se ocupa principalmente del procesamiento y la comprensión del lenguaje humano por parte de las máquinas. Al permitir que las máquinas entiendan el lenguaje humano, el PNL mejora la precisión y la eficacia de los procesos. Esto se consigue automatizando diversas tareas repetitivas. Algunos ejemplos de aplicaciones de procesamiento del lenguaje natural son la clasificación de entradas, la traducción automática, la comprobación de la ortografía y el resumen.

Es un área de estudio que comenzó por los años 50s y que está recobrando relevancia debido a la gran cantidad de texto que se genera hoy en día y la capacidad para poder procesarlo. De forma genérica se puede dividir análisis de texto en dos grandes partes:

- **Análisis sintáctico:** se utiliza para establecer el significado observando la gramática de una frase. Es el proceso de estructurar el texto utilizando las convenciones gramaticales del lenguaje. Esencialmente, consiste en el análisis de oraciones dividiéndolas en grupos de palabras y frases que crean una oración correcta. Esto no tiene en cuenta el hecho de que las frases pueden carecer de sentido, que es el punto en el que el análisis semántico viene a echar una mano.

no tiene en cuenta el hecho de que las frases pueden carecer de sentido, que es el punto en el que el análisis semántico viene a echar una mano.

- **Análisis semántico:** Nuestra comprensión del lenguaje se basa en los años que llevamos escuchando y conociendo el contexto y el significado. El lenguaje humano es dinámico. Con la invención de los algoritmos de aprendizaje automático, los ordenadores son capaces de entender el significado y la lógica que hay detrás de nuestras expresiones al menos hasta cierto punto.

Con el desarrollo del PNL, hoy en día, se es capaz de realizar **análisis de sentimiento** para el lenguaje humano. Esto le permite detectar emociones en el texto, lo que constituye una de las aplicaciones de PNL. De esta forma un algoritmo puede leer un texto y etiquetar la emoción subyacente (si es positiva, negativa o neutra). Esto, junto con la proliferación de las redes sociales permite realizar investigaciones que hace pocos años hubieran sido imposibles. Por ejemplo, uno puede medir de forma objetiva cuál es la opinión que la población tiene respecto a un evento concreto.

El PNL es un campo que se ha desarrollado principalmente para el análisis de texto en inglés y solo es en estos últimos años que se ha comenzado a extender al castellano. Su aplicación por tanto a nuestra lengua es una de las líneas más emergentes a desarrollar en el futuro próximo en la investigación relacionada con IA. Esto ha quedado reforzado por la estrategia del gobierno de España que está en vías de publicar un Proyecto Estratégico para la Recuperación y Transformación Económico (PERTE) asociado a la lengua española y la IA. Por estas razones DATAI está interesado en el desarrollo de una línea de investigación basada en:

- El análisis del sentimiento a través del PNL dado su alto potencial de desarrollo e impacto que puede tener en el futuro en la sociedad especialmente en la de habla hispana.

Además su aplicación eminentemente transversal a casi cualquier dominio confirma la vocación de DATAI de convertirse en una entidad de excelencia que promueva investigación transversal a toda la universidad.

El instituto ya ha realizado colaboraciones con la Facultad de Comunicación en el ámbito del PNL y está actualmente en conversaciones con el departamento de justicia para formar parte de un grupo de trabajo que buscará la aplicación del PNL para mejorar la gestión de la justicia en Navarra.

1.2.2 ÁREAS DE INVESTIGACIÓN DE LOS INVESTIGADORES DE DATAI

1 Fundamentals	4 Computational Biology CIMA	7 Ethics and Law
2 Advanced Analytics BBVA	5 Computational Biology TECNUN	8 Health Analytics
3 AI and Management IESE	6 Computer Science	9 Industrial Data Analysis & Information Management

1 FUNDAMENTALS

Grupo multidisciplinar con una sólida formación en investigación aplicada y metodológica de Data Science. Su investigación se desarrolla fundamentalmente modelización y diseños experimentales, incluido el aprendizaje activo.

Estas son las principales líneas de investigación:

- Diseños experimentales: Active learning.
- Procesamiento del lenguaje natural y análisis de sentimientos.
- Redes complejas: minería de datos y dinámica.
- Series temporales de Integración Fraccionada y Procesos de Memoria Larga.

2 ADVANCED ANALYTICS – BBVA

BBVA cree que el conocimiento derivado de los datos financieros puede transformar la industria bancaria y su papel en el mundo. Por ello, el área de Advanced Analytics investiga en la aplicación de la IA al sector bancario. Entre las líneas de investigación destacan:

- Técnicas de optimización y aprendizaje automático para portafolios bancarios.
- Métodos avanzados de *pricing* dinámico..
- Motores de recomendación.

3

AI AND MANAGEMENT – IESE

La nueva Iniciativa sobre Inteligencia Artificial y el Futuro de la Dirección es un proyecto multidisciplinar que analiza el impacto de la IA en la dirección de empresas y que contribuye a formar a los directivos en el uso ético y socialmente responsable de la IA en sus compañías.

La inteligencia artificial afectará en los próximos años a todas las esferas de la actividad económica, por lo que los directivos deben aprender a adaptarse a la evolución de este sector. También deben ser capaces de transformar a sus empresas y asegurarse de que sus empleados, y ellos mismos, poseen las habilidades necesarias para seguir aportando valor en este nuevo contexto. En consecuencia, el propósito de la nueva iniciativa del IESE sobre IA es responder a estas necesidades de generación de conocimiento y formación directiva.

Las áreas de investigación son:

- Las habilidades en IA y el mercado laboral.
- El impacto de la automatización industrial.
- El desarrollo y el uso de la IA en las empresas, especialmente en Europa.
- El futuro del trabajo y las organizaciones.

4

BIOLOGÍA COMPUTACIONAL – CIMA

La biología molecular ha sufrido una revolución debido a la capacidad de estudiar simultáneamente el funcionamiento y la expresión de miles de genes y proteínas en el cuerpo del paciente.

Gracias al uso de la tecnología informática, las bases de datos y el análisis estadístico podemos analizar con precisión y rapidez grandes cantidades de datos que nos permiten comprender la complejidad de los mecanismos que causan las enfermedades.

El Programa de Biología Computacional del CIMA - Universidad de Navarra tiene actualmente dos líneas principales de investigación:

- Análisis de datos transcriptómicos, tanto a gran escala como a resolución de célula única.
- Desarrollo de nuevos formatos de archivo para el almacenamiento y acceso a los datos ómicos.
- Métodos de Aprendizaje Automático para resolver problemas biomédicos y su traslación a la clínica.

5

BIOLOGÍA COMPUTACIONAL – TECNUN

El Grupo de Biología Computacional de TECNUN es un equipo multidisciplinar con una amplia experiencia en el desarrollo de algoritmos de optimización, análisis estadísticos, así como métodos de aprendizaje automático y aprendizaje profundo, aplicados a diferentes problemas biológicos. Nos centramos principalmente en la salud humana mediante la integración de datos de alto rendimiento (genómica, transcriptómica, proteómica, metabolómica, etc.) y bases de datos biológicas (genómica, farmacológica, metabólica, etc.). Las principales líneas de investigación son:

- Oncología de precisión.
- Modelos predictivos de toxicidad de fármacos basados en características estructurales de las moléculas.
- Esquemas de compresión adaptados a diferentes datos ómicos.
- Análisis de secuenciación de ADN.

6

CIENCIAS DE LA COMPUTACIÓN

Las ciencias de la computación abarcan un amplio abanico, desde sus fundamentos teóricos y algorítmicos hasta los desarrollos de vanguardia en visión por ordenador, sistemas inteligentes, bioinformática y otras áreas apasionantes. El papel de los informáticos puede agruparse en las siguientes categorías: diseñar e implementar software, desarrollar formas eficaces de resolver problemas informáticos e idear nuevas formas de utilizar los ordenadores.

La informática es esencialmente el estudio de cómo "mapear" el lenguaje de las máquinas en un lenguaje más comprensible y utilizable para que las personas puedan resolver problemas complejos. Este campo utiliza la teoría de la computación en el cálculo de la información. Las líneas de investigación son:

- Cambio climático.
- Análisis de imagen.
- Modelado y simulación.
- Análisis topológico de datos.
- Blockchain y estructuras de datos autenticadas.
- Soft Computing y resiliencia.

7

ÉTICA Y DERECHO

El reto del área de Ética y Derecho, se centra en trabajar para que las nuevas tecnologías, al afectar al conjunto de la sociedad, busquen no sólo una mejora en las previsiones, sino que éstas tengan claridad en los factores que las definen. Que se promueva una legislación, sin detener el avance en el desarrollo de dichas tecnologías, que permita entender las funciones sociales o biológicas que intervienen en ella. Que en definitiva se tienda a obtener el conocimiento de los fenómenos sociales antes de su predicción. Las líneas de investigación son:

- Ética e IA.
- Regulación legal de Big Data, IA y protección de datos.
- Ética de la virtud e IA.

8

HEALTH ANALYTICS

Health Analytics se refiere al uso de grandes cantidades de datos recopilados para proporcionar a las organizaciones información útil y mejorar la atención al paciente. Estos conocimientos se desarrollan en disciplinas analíticas para impulsar la toma de decisiones basada en datos. A su vez, estas decisiones mejoran la planificación, la gestión, la medición, el aprendizaje y, en última instancia, la calidad de la atención al paciente.

A medida que las organizaciones sanitarias de todo el mundo se esfuerzan en mejorar la atención al paciente y hacer más con menos, la analítica cobra relevancia. El desarrollo de las competencias analíticas puede ayudar a las organizaciones sanitarias a aprovechar el Big Data para mejorar sus resultados y ofrecer un valor añadido a los pacientes. Las líneas de investigación asociadas a esta rama son:

- Análisis de señales biomédicas.
- Neurociencia de sistemas.
- Imagen médica.
- Inferencia causal en biomedicina.
- Epidemiología.
- Investigación clínica.

9

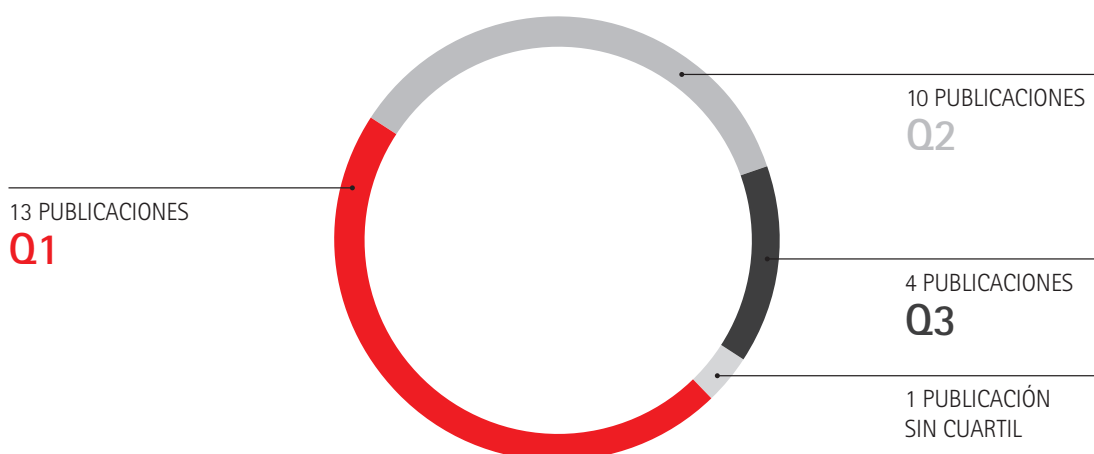
INDUSTRIAL DATA ANALYSIS & INFORMATION MANAGEMENT

La investigación del grupo se centra principalmente en la medición del impacto del Big Data en la industria de alta tecnología. Dos son las principales áreas de especialización de nuestra investigación para el desarrollo de procesos industriales, la analítica de la ciberseguridad y el Big Data. Las principales áreas de investigación son:

- Protocolos de comunicaciones y seguridad.
- Identificación de anomalías y parámetros significativos en procesos.
- Modelos de predicción y mantenimiento preventivo.

1.3 PUBLICACIONES

Durante los años 2020 y 2021 la actividad científica de los miembros adscritos y junta directiva de DATAI (5 investigadores) ha dado lugar a 28 publicaciones (13 Q1, 10 Q2, 4 Q3 y una sin cuartil), que muestran la transversalidad del instituto:



- 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 Specific 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.
- Higuera, 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; Grochowicz, 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., Benítez, 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., Benítez, 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 DOCTORADOS INDUSTRIALES

Durante el curso 2020-2021, 7 doctorados industriales del BBVA han comenzado su tesis industrial en la Universidad de Navarra. Cada uno cuenta con un supervisor del propio banco y un director de la Universidad de Navarra.

Los Programas de Doctorado bajo los que se encuadran estas tesis son: Natural and Applied Sciences (3 doctorados), Applied Engineering (1) y Economics and Business Administration (3).

	NOMBRE Y APELLIDOS	PROGRAMA DE DOCTORADO	TÍTULO DE LA TESIS
	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 PENCİK	Economics and Business Administration	Reshaping Quantitative Risk Management Metrics in Financial Industries with Deep Learning Applications

En el curso 2021-22 comenzarán la tesis otros cuatro estudiantes.

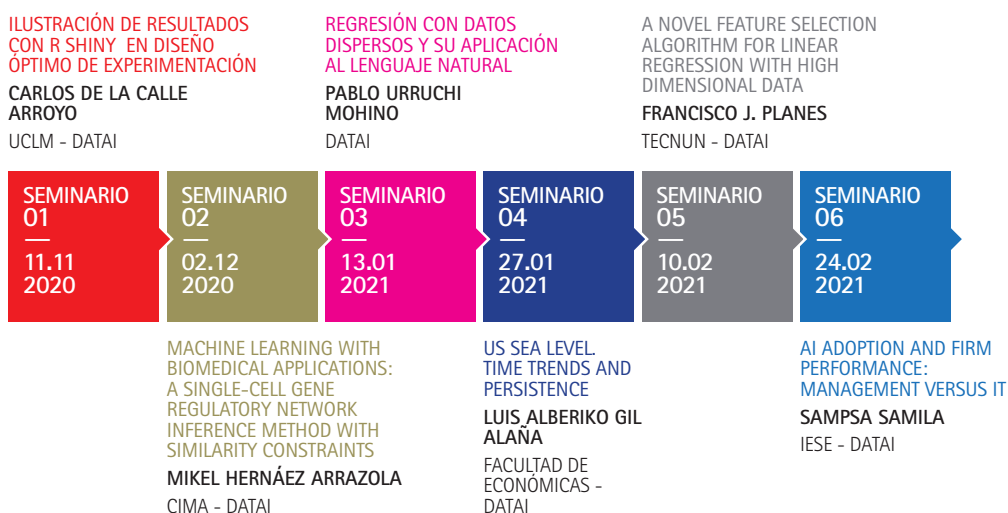
1.5 REUNIONES ORGANIZADAS

Durante los años 2020 y 2021 DATAI ha organizado las siguientes actividades:

- **I Reunión Consejo Asesor:** 14 de diciembre de 2020. Asistieron representantes de casi todos los centros. El Director del instituto, Jesús López Fidalgo, hizo una presentación del instituto, y, posteriormente, se abordaron propuestas para impulsar el crecimiento de DATAI.
- **I Jornada Científica:** 16 de diciembre de 2020. La Vicerrectora de investigación, Iciar Astiasarán, fue la encargada de hacer la apertura de esta jornada. A continuación varios investigadores de centros y facultades, como CIMA, Tecnun, CUN, Derecho, Económicas, DATAI y CEIT hicieron sus presentaciones.
- **Visita de Dña. Carme Artigas,** Secretaria de Estado de Digitalización e Inteligencia Artificial y la directora general de IA, **Laura Flores:** 23 de marzo de 2021.
- **Reunión con el equipo de Guzmán Garmendia,** Director general de Telecomunicaciones y Digitalización (DGTT) del Gobierno de Navarra: 24 Junio de 2021.

1.6 SEMINARIOS

Durante los años 2020 y 2021 DATAI ha celebrado 17 seminarios. Éstos han sido impartidos quincenalmente por profesionales de centros y facultades de la universidad, y ponentes invitados:



APLICACIONES DE DATA,
SCIENCE EN CARDIOLOGÍA
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 CIENTIFICO DATAI

SPACE FILLING: WEAK
COVERING AND
QUANTIZATION OF
HIGH DIMENSIONAL
SETS
JACK NOONAN



LA ENCUESTA DIGITAL NEWS
REPORT. METODOLOGIA,
CONSULTA Y APLICACIONES
SAMUEL NEGREDO BRUNA
FAC. COMUNICACIÓN

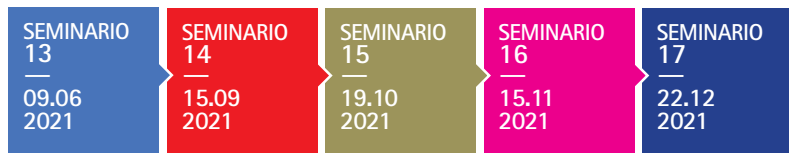
FLORENCE
NIGHTINGALE:
A ROLE MODEL IN THE
21ST CENTURY
**GUADALUPE GÓMEZ
MELIS**
UPC

APRENDIZAJE AUTOMÁTICO
EQUITATIVO: UN ENFOQUE
BASADO EN EL TRANSPORTE
ÓPTIMO
PAULA GORDALIZA
BCAM

METODOLOGÍA Y GESTIÓN DE LA
INVESTIGACIÓN APLICADA EN
LA INDUSTRIA: LA LABOR DEL
DOCTOR INGENIERO
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



EL ESPACIO SEMI EXTERIOR
COMO ESTRATEGIA DE
ADAPTACIÓN AL RIESGO DE
SOBRECALENTAMIENTO EN UN
CONTEXTO TROPICAL DE ALTA
DENSIDAD URBANA. BENEFICIOS
A NIVEL SOCIAL Y AMBIENTAL
**JUAN CARLOS GAMERO
SALINAS**
INVESTIGADOR PREDOCTORAL-
ESCUELA DE ARQUITECTURA
(UNIV. NAVARRA)

A BAYESIAN
NONPARAMETRIC MODEL
FOR CLASSIFICATION OF
LONGITUDINAL PROFILES
ROLANDO DE LA CRUZ
UIA



2. TRANSFERENCIA E INNOVACIÓN

TRANSFERENCIA E INNOVACIÓN

DATAI colabora con diferentes grupos de investigación, centros de la universidad, empresas y otras instituciones. Su carácter transversal así como la interdisciplinariedad del equipo de investigadores y expertos en Ciencia de los Datos, Estadística e Inteligencia Artificial nos permiten trabajar en proyectos de datos de alta complejidad analítica y aportar valor en cada una de sus fases.

2.1 COLABORACIONES INTERNAS

COLABORACIONES Y PROYECTOS

El instituto realiza labores de **colaboración científica** entre sus miembros e investigadores de otros grupos de investigación, tanto de la universidad como externos que llevan consigo un trabajo científico significativo y también participan en los proyectos de investigación financiados en convocatorias competitivas, contratos y otro tipo de convocatorias.

ASESORAMIENTO Y SERVICIOS

El instituto proporciona también **asesoramiento** o apoyo a trabajos puntuales y de servicio

Durante el curso 2020-2021, las colaboraciones de asesoramiento solicitadas al Instituto (ya sea por un grupo o departamento de la Universidad de Navarra, por un grupo de investigación de otras universidades o centros, así como por una empresa o institución pública o privada), han sido de 24.

En estas colaboraciones han tenido dedicación diversos miembros adscritos del Instituto en diferentes categorías profesionales: técnico, senior y experto.

Algunas de estas colaboraciones han sido con: ICS (Atlantes, EASH, Discurso público, Creatividad), la Escuela de ENFERMERÍA, la facultad de COMUNICACIÓN, la facultad de PSICOLOGÍA, la Clínica Universidad de Navarra CUN, el servicio de ADMISIÓN y VICERRECTORADO DE ALUMNOS y ALUMNI.

2.2 EMPRESA Y ENTIDADES

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 ACUERDOS ESTRATÉGICOS

La Universidad de Navarra y BBVA son socios estratégicos en el campo de la analítica avanzada de datos y la inteligencia artificial. Colaboran para formar a sus empleados e impulsar la investigación en ciencia de datos.

Esta colaboración fue formalizada en julio 2020 en un evento telemático en el que participaron por parte de la Universidad de Navarra la vicerrectora de Investigación, Iciar Astiasarán, y el director del Instituto de Ciencia de los Datos e Inteligencia Artificial, Jesús López Fidalgo; y por BBVA, Carlos Casas, responsable de Talento y Cultura; y Ricardo Martín, responsable de Data.



ACUERDO DE COLABORACIÓN ENTRE LA UNIVERSIDAD DE NAVARRA Y BBVA

Este acuerdo se centra principalmente en tres líneas de colaboración, que incluyen tesis industriales, ya mencionadas más arriba, un grupo en el máster, que se mencionará en el siguiente capítulo y la certificación profesional de científico de datos:

CERTIFICACIÓN PROFESIONAL PARA CIENTÍFICOS DE DATOS.

Esta certificación, emitida por la Universidad de Navarra y BBVA con reconocimiento a nivel global, se le otorga a los profesionales de BBVA que realicen la formación "Data Scientist Fundamentals", incluida en la oferta formativa del Campus BBVA y además realicen al menos 400 horas de trabajo en la plataforma de Big Data del banco. El trabajo está supervisado por académicos de diversos centros de la universidad que cuentan con una contrastada experiencia en analítica avanzada de datos.

La Certificación tiene como misión definir un estándar en el campo de la ciencia de datos a través del cual las organizaciones puedan identificar profesionales con dominio del campo de la analítica avanzada. Además, la Certificación refuerza la credibilidad y visibilidad de la profesión de científico de datos.

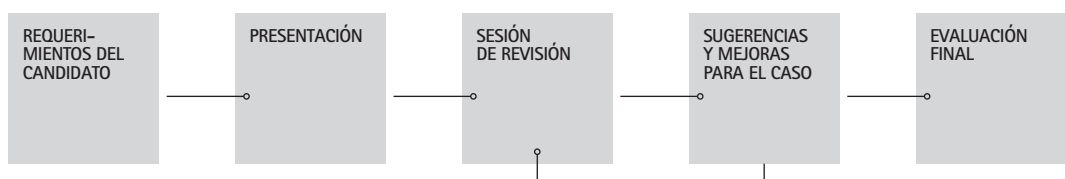
El proceso consta de tres fases bien diferenciadas:

- a. Formación recibida previamente
- b. Realización de un caso de uso
- c. El examen de certificación

Durante el curso 2020-2021, 45 candidatos se han presentado para obtener la certificación profesional.

a. La revisión del caso de uso

La revisión del caso de uso es un proceso en el que se evalúa un proyecto de ciencia de datos realizado en su trabajo en BBVA por el candidato que quiere certificarse. El procedimiento es el siguiente:



b. El examen de certificación

La Certificación mide el desempeño en 5 áreas específicas: la formulación de problemas en el marco de la ciencia de datos, machine learning, estadística, matemáticas y programación. El examen evalúa el conocimiento general de estas áreas y su aplicación en proyectos de ciencia de.

Asimismo, BBVA aprovechará esta estrecha relación con la Universidad de Navarra para captar talento entre sus alumnos, que tendrán acceso a una bolsa de trabajo para incorporarse al proyecto del banco. Mientras que aquellos científicos de datos del grupo que tengan interés en el ámbito académico, podrán impartir clases o realizar otras actividades docentes en la Universidad de Navarra.

2.4 CONCURSO INCUBADORA DE SONDEOS Y EXPERIMENTOS

Este concurso está organizado por el Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI) de la Universidad de Navarra y el Departamento de Estadística, Informática y Matemáticas de la Universidad Pública de Navarra (UPNA), con la colaboración de la Sociedad Navarra de Profesores de Matemáticas "TORNAMIRA" Matematika Irakasleen Nafar Elkartea; del Instituto de Estadística de Navarra; y del Departamento de Educación del Gobierno de Navarra y la delegación provincial del Instituto Nacional de Estadística.

El concurso se realiza anualmente, y está dirigido a alumnos de E.S.O., Bachillerato y Ciclos Formativos de grado básico o medio en Navarra con interés por la estadística y la ciencia de los datos. Los participantes deben realizar un proyecto de estadística, probabilidad sobre un tema de interés elegido por el equipo: sociedad, salud, ecología, economía y consumo, medio ambiente... Hay una fase local, y los ganadores de esta fase, pueden participar en la fase nacional junto a equipos ganadores de otras provincias.

Con este concurso se pretende:

- Fomentar la enseñanza y aprendizaje de la Estadística en los niveles educativos no universitarios.
- Complementar los conocimientos y competencias así como difundir la importancia y utilidad de la Estadística, Probabilidad e Investigación Operativa en la vida real y en todas las disciplinas académicas.
- Despertar entre los estudiantes el interés por la Estadística, Probabilidad e Investigación Operativa como herramienta fundamental en la práctica totalidad de las ciencias.

-
- Familiarizar a los estudiantes con las distintas etapas en la realización de un proyecto estadístico.
 - Aprender a interpretar las distintas informaciones estadísticas que nos rodean y fomentar el espíritu crítico.

El 18 de Junio de 2021 se celebró de manera online la Fase final de Navarra 2020-2021 del Concurso Incubadora de Sondeos y Experimentos. En él participaron 139 estudiantes de Educación Secundaria Obligatoria, Bachillerato y Formación Profesional, procedentes de nueve centros. Este certamen busca despertar en el alumnado la curiosidad por la ciencia de los datos como herramienta básica para la investigación.

Jesús López Fidalgo, director del Instituto de Ciencia de los Datos e Inteligencia Artificial, inició la sesión con una charla divulgativa en la que, además, se llevó a cabo un concurso on line.

Los participantes presentaron proyectos de estadística, probabilidad e investigación operativa sobre cuestiones como las trayectorias profesionales, las redes sociales, efectos del Covid-19 en los ámbitos de salud y educación, videojuegos, condiciones de habitabilidad en los hogares, entre otras.

Patrocinadores:



Universidad
de Navarra

DATAI
INSTITUTO DE CIENCIA DE LOS
DATOS E INTELIGENCIA ARTIFICIAL

upna

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

INĒ

Instituto Nacional de Estadística

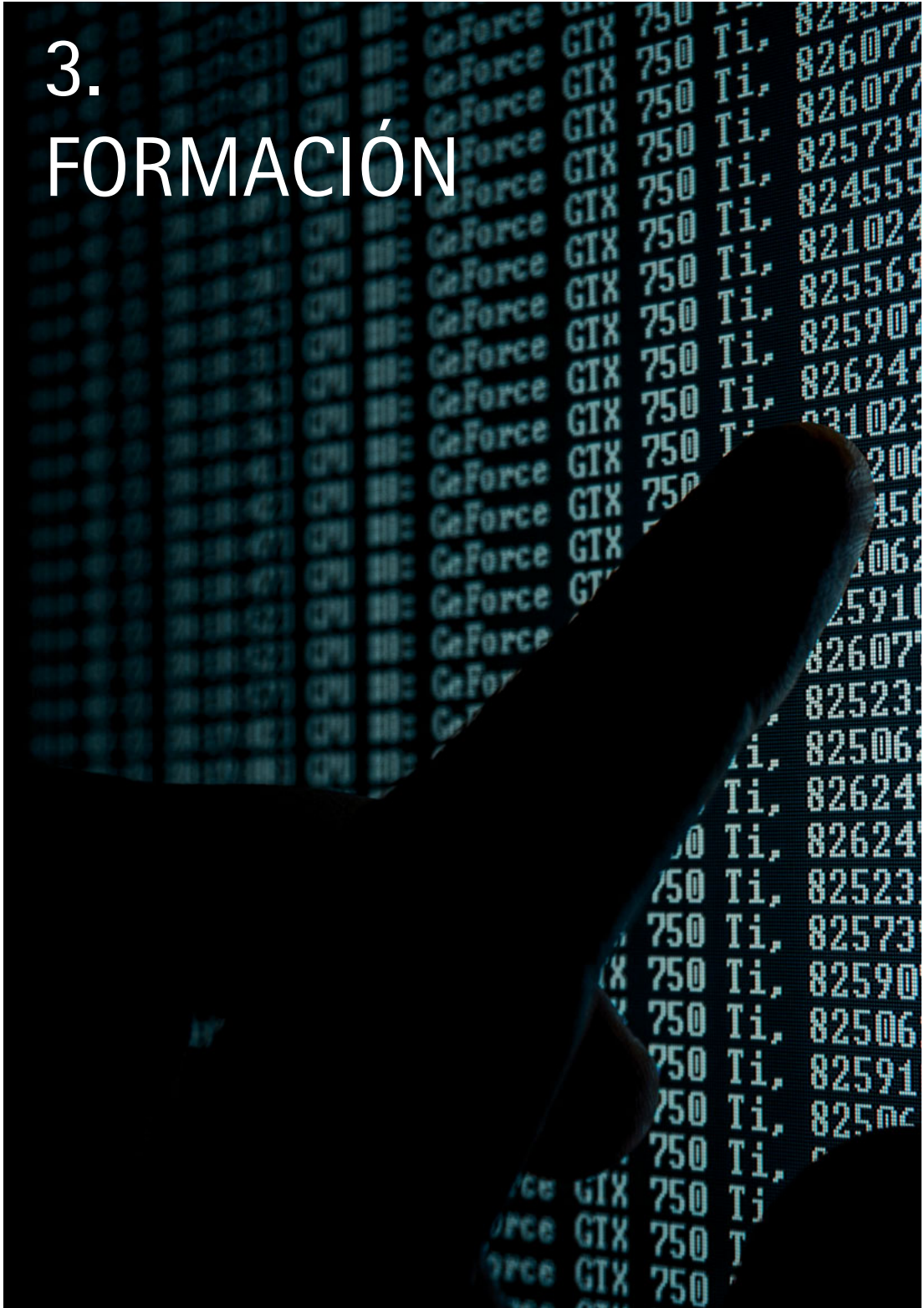
(na)stat
Instituto de Estadística de Navarra
Nafarroako Estatistika Erakundea

Gobierno
de Navarra  Nafarroako
Gobernua



3.

FORMACIÓN



FORMACIÓN

Las actividades formativas de DATAI durante el curso 2020-2021 incluyen el Máster Universitario en Big Data Science, los microcursos, programas In Company y otros programas.

3.1 MÁSTER OFICIAL EN BIG DATA SCIENCE

El Máster en Big Data Science es un título oficial que nace como respuesta a la demanda de empresas y organizaciones sobre conocimientos de Big Data y Ciencia de los Datos. Ofrece un programa técnico y específico y, a la vez, muy práctico.

El profesorado procede de diversas compañías e instituciones de prestigio, con amplia experiencia en estas herramientas. Se enseña a los alumnos cómo utilizar estas herramientas en situaciones reales y cómo aplicarlas en diversos campos.

El formato Executive, con sesiones presenciales los viernes por la tarde y sábados por la mañana, facilita compaginar el desempeño profesional con la asistencia a clase, y fomentar el networking entre profesionales y empresas.

En los cursos 2017-18 y 2018-19 se impartió un programa de especialistas en Big Data en la sede de Pamplona (en el Aula Siemens Gamesa del ICS). En 2019-2020 el máster pasó a ser un máster oficial y, desde entonces, se desarrolla íntegramente en la Sede de Posgrado de Madrid (Edificio Alumni).

En el curso 2020-2021 contamos con 2 grupos:

- 39 alumnos en grupo abierto
- 25 alumnos en grupo BBVA

Principales empleadores de nuestros alumnos:



Se han desarrollado TFM's en empresas como BBVA, PwC, Deloitte, EY...

3.2 MICROCURSOS

La necesidad de que los profesionales de cualquier disciplina lleguen al mercado laboral, incluida la investigación, con conocimientos y habilidades en el manejo de datos para la creación de valor se ha vuelto cada vez más relevante.

Para ello, durante el curso 2020-21, se ofrece este programa con la idea de permitir que profesionales de todas las titulaciones, incluso los que provienen de áreas claramente relacionadas como matemáticas y estadística, se actualicen y ganen algún grado en el uso de las principales metodologías y herramientas disponibles en el mercado para resolver problemas asociados a la gestión y análisis de datos.

Estos programas se realizan en la modalidad online bajo la tipología de Curso de Formación Superior.

El Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI) imparte cursos de formación superior (de 2 ECTS cada uno) en:

- Gestión de datos con Python
- Gestión de datos con R
- Estadística de la ciencia de los datos
- Machine Learning
- Visualización de datos
- Hojas de cálculo para el análisis de datos
- Análisis de redes sociales

Con el fin de ofrecer una formación más integral, los cursos individuales se pueden agrupar en los siguientes cursos de formación superior de 2 ECTS cada uno:

**Curso de formación superior en
Herramientas de Gestión de Datos**
TOTAL 6 ECTS

GESTIÓN DE DATOS CON PYTHON
GESTIÓN DE DATOS CON R
HOJAS DE CÁLCULO PARA EL ANÁLISIS DE DATOS

**Curso de formación superior en
Técnicas de Ciencias de Datos**
TOTAL 6 ECTS

ESTADÍSTICA DE LA CIENCIA DE LOS DATOS
MACHINE LEARNING
VISUALIZACIÓN DE DATOS
ANÁLISIS DE REDES SOCIALES

Los cursos se imparten de manera sincrónica online. Cada curso tiene una duración de 3 semanas, con clases teórico-prácticas de 2 horas de duración y a continuación una hora de tutoría. Todas las sesiones de explicación serán realizadas en la aplicación Zoom y grabadas y disponibles a los alumnos. Como material complementario se usarán artículos científicos o vídeos de divulgación científica tanto teórica como aplicada.

Dirigido a estudiantes universitarios o profesionales interesados en abordar el área de la Ciencia de Datos con conocimientos básicos de uso de ordenadores y hojas de cálculo.

Microcursos impartidos durante el curso 2020-2021

- **Estadística de la Ciencia de los Datos**
- **Análisis de redes sociales**
- **Gestión de datos con Python**

Debido a la alta demanda, este microcurso se impartió 2 veces. El primer curso se dirigió a estudiantes y el segundo a profesionales y doctorandos de la universidad.

3.3 EVENTOS Y ACTIVIDADES



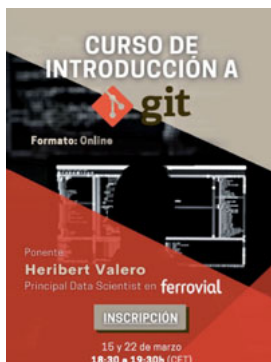
WORKSHOP
The future of DS and AI. Dirigido a alumnos y profesores del Máster en Big Data Science (MEBDS). Presencial. 16/04/2021



PRESENTACIÓN DEL LIBRO
El Mito del Algoritmo. Con la intervención de los autores, RICHARD BENJAMINS e IDOIA SALAZAR, y JESÚS LÓPEZ Fidalgo como moderador. Presencial y en streaming. 25/02/2021



WORKSHOP
Visualización de datos. SOLEDAD GUARCH (Senior Digital Strategy Analyst, British Airways). Presencial. 19/02/2021



CURSO FORMATIVO
Curso de introducción a GIT. HERIBERT VALERO (Científico de datos en Ferrovial). Online. 15 y 22/03/2021



WEBINAR
Estrategias en Visualización de datos. Mesa panel con MIGUEL ÁNGEL CASARES (Adjunct professor, Universidad de Navarra); MARÍA MANSO (Global Head of People Analytics, BBVA); OMAR ARCE TORREBLANCA (Business Analytics Manager, Axesor). Presencial y en streaming. 19/02/2021



SESIÓN INFORMATIVA

Entrevistas y procesos de selección de Data Science / Analytics. MERCEDES SÁNCHEZ LOR (Manager de Tecnología en la consultora de selección Page Personnel). Presencial. 5/02/2021



BOOTCAMP

Salidas profesionales Data Science / Big Data. MARTA ARDÍZONE (Talent Brand Manager en ETS Factory); KRISTINA MONSON (Deputy Director of Career Services, Universidad de Navarra); MATÍAS ÁVILA CLEMENTE (Data Scientist en BBVA Data Et Analytics). Presencial y en streaming. 29/10/2020

SESIÓN INFORMATIVA Y DEBATE

¿Cómo será el trabajo en el futuro y cómo puedes prepararte para conseguirlo? Presencial. 27/04/2021

3.4 PROGRAMAS IN COMPANY

Programación de cursos, talleres y seminarios a medida para empresas, grupos de investigación, servicios o centros que busquen la actualización y especialización sobre temas de análisis de datos relacionados con las necesidades reales.

También cabe la posibilidad de diseñar cursos de iniciación o especialización en software.

INSTITUTO DE CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL

UNIVERSIDAD DE NAVARRA

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



Universidad
de Navarra

DATAI
INSTITUTO DE CIENCIA DE LOS
DATOS E INTELIGENCIA ARTIFICIAL