

FORMATO PRESENTACIÓN JORNADAS DE SEGUIMIENTO 2018

SUBDIVISIÓN DE PROGRAMAS TEMÁTICOS CIENTÍFICO-TÉCNICOS

Máximo número de transparencias 18.

Tamaño máximo del archivo 4 Megas

Sugerencias de presentación:

- Tamaño recomendable de letra a emplear: 24
- Encabezados: 28 (el tamaño superior al texto)
- Tamaño mínimo a emplear: 20
- Emplear negrita
- No disminuir de 0.85 el espacio entre líneas.
- Respetar tamaño máximo de los recuadros



INFER.java: Un lenguaje de programación probabilístico para el desarrollo de aplicaciones inteligentes sobre grandes volúmenes de datos

DATOS BÁSICOS DEL PROYECTO

Referencia: TIN2015-74368-JIN

Investigador Principal : Andrés R. Masegosa Arredondo

Tutor: Antonio Salmerón Cedrán

Organismo: Universidad de Almería

Centro: Departamento de Matemáticas

Subvención concedida (Costes directos): 165.200 euros

Fecha inicio: 1/1/2017 **Fecha finalización (prevista):** 31/12/2017

Presenta: [Nombre de la persona que presenta si es distinto del IP](#)



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

PARTICIPANTES

Entidades participantes:

Universidad de Almería

Investigadores participantes:

Andrés R. Masegosa (IP)

Antonio Salmerón (Tutor)

Rafael Cabañas (Contratado)



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

CONTEXTO

- El **aprendizaje automático** (ML, siglas en inglés) está siendo ampliamente adoptado por empresas e instituciones:
 - Aplicaciones Software, Ciencias de la Salud, Coches Autónomos, etc



- Mi grupo de investigación tiene una larga trayectoria en el campo del **ML** (modelos gráficos probabilísticos).

Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.



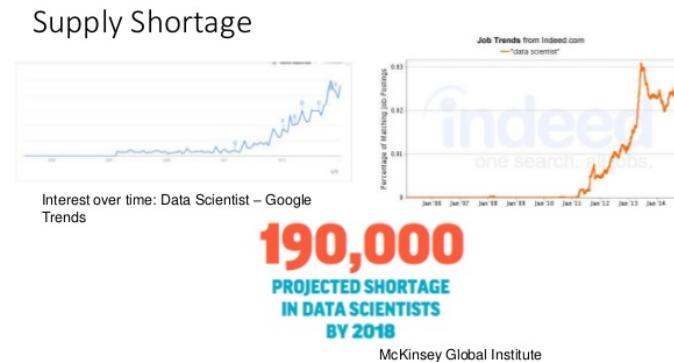
MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

MOTIVACIÓN

- **Desarrollo Efectivo de Soluciones de ML requiere:**
 - Expertos altamente cualificados
 - Algoritmos de ML son complejos
 - Modelos de ML necesitan ser personalizados.



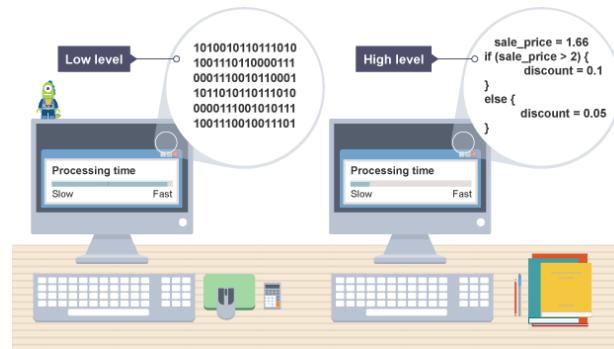
- Tendencia que beneficia a las **grandes corporaciones**, por la necesidad de **capacidad técnica y financiera** de llevar a cabo proyectos de ML.

Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.



HIPÓTESIS DE PARTIDA

- Programa de la Agencia americana DARPA:
 - Probabilistic Programming for Advance Machine Learning



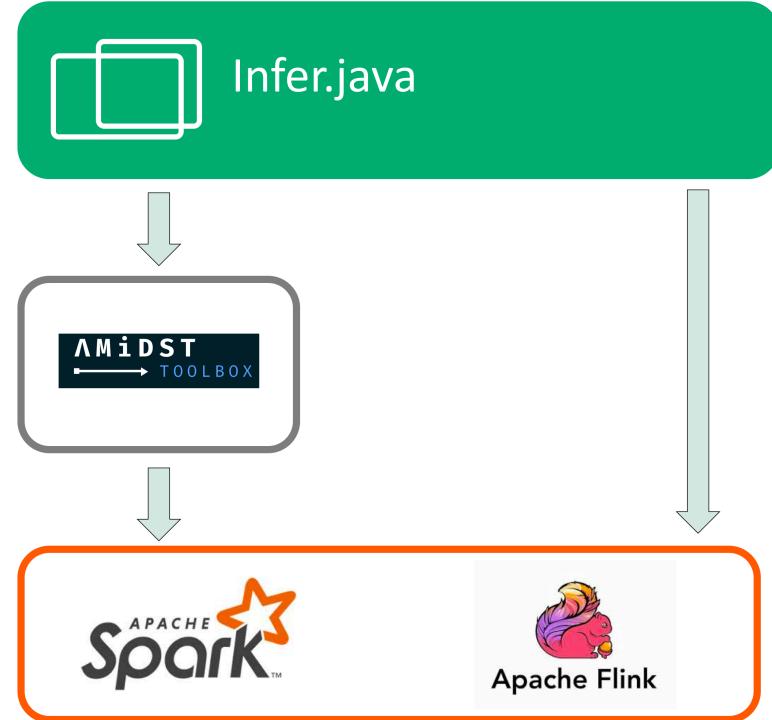
- Lenguajes de Programación Probabilísticos:
 - Incrementar enormemente el número de personas capaces de construir un sistema de ML.
 - Hacer que los expertos en ML sean mucho más productivos.
 - Permitir del desarrollo de nuevas aplicaciones de ML

Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.



ESTRATEGIA GENERAL DEL PROYECTO

```
//Firstly, the probabilistic model is defined.  
Variable fire = Engine.newRareEventVariable();  
Variable temperature = Engine.newGaussianVariable().dependsPositivelyOf(fire);  
Variable smoke = Engine.newEventVariable().dependsPositivelyOf(fire);  
Variable tempSensor = Engine.newGaussianVariable().dependsNoisilyOf(temperature);  
Variable smokeSensor = Engine.newEventVariable().dependsNoisilyOf(smoke);  
  
//Secondly, the probabilistic model is learnt with historical data  
Stream<SensorMeasures> historicalData = Engine.loadData("./historicalData.csv");  
  
historicData.forEach(sensorMeasure -> {  
    tempSensor.observe(sensorMeasure.getTempSensor());  
    smokeSensor.observe(sensorMeasure.getSmokeSensor());  
    Engine.updateModel();  
});  
  
//Finally, the model is used to monitor the presence of a fire, by producing a data stream  
//with the probability of a fire.  
Stream<SensorMeasures> streamOfMeasurements = Engine.connectDataStream("SensorMeasurements");  
  
Stream<Double> probsOfFire = streamOfMeasurements.map(sensorMeasure -> {  
    tempSensor.observe(sensorMeasure.getTempSensor());  
    smokeSensor.observe(sensorMeasure.getSmokeSensor());  
    Engine.updateModel();  
    return Engine.probabilityOf("Fire==True");  
});
```



Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

OBJETIVOS PROPUESTOS Y ALCANZADOS

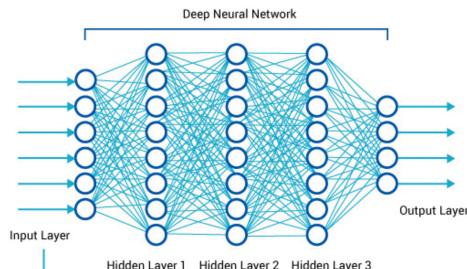
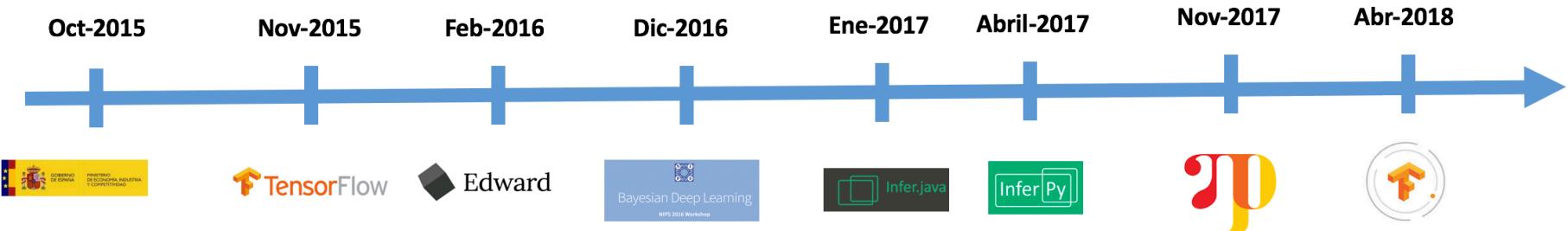
Paquete Trabajo	Grado Consecución	Grado Esperado
PT0. Gestión Proyecto	50%	100%
PT1. Estado Del Arte Y Requerimientos	90%	95%

En esta diapositiva se indicarán de forma sucinta los objetivos propuestos en la memoria original, los posibles cambios como consecuencia del progreso de la investigación; indicar el grado de consecución de los mismos (indicar porcentaje aproximado) y los que se prevén alcanzar en la duración prevista del proyecto. **En caso de estimarse la necesidad de prórroga, debe indicarse aquí. NO hay prórroga en los JIN, salvo por maternidad.**

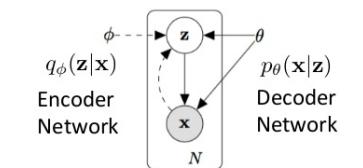
Explique la relación de los objetivos con la de otros proyectos del tutor o del equipo con el que se relaciona o en el que se ha integrado, si procede.



ESTRATEGIA GENERAL DEL PROYECTO



Variational Autoencoder



Minimize: $D_{KL}[q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z}|\mathbf{x})]$

Intractable: $p_\theta(\mathbf{z}|\mathbf{x}) = \frac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{p_\theta(\mathbf{x})}$

Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.

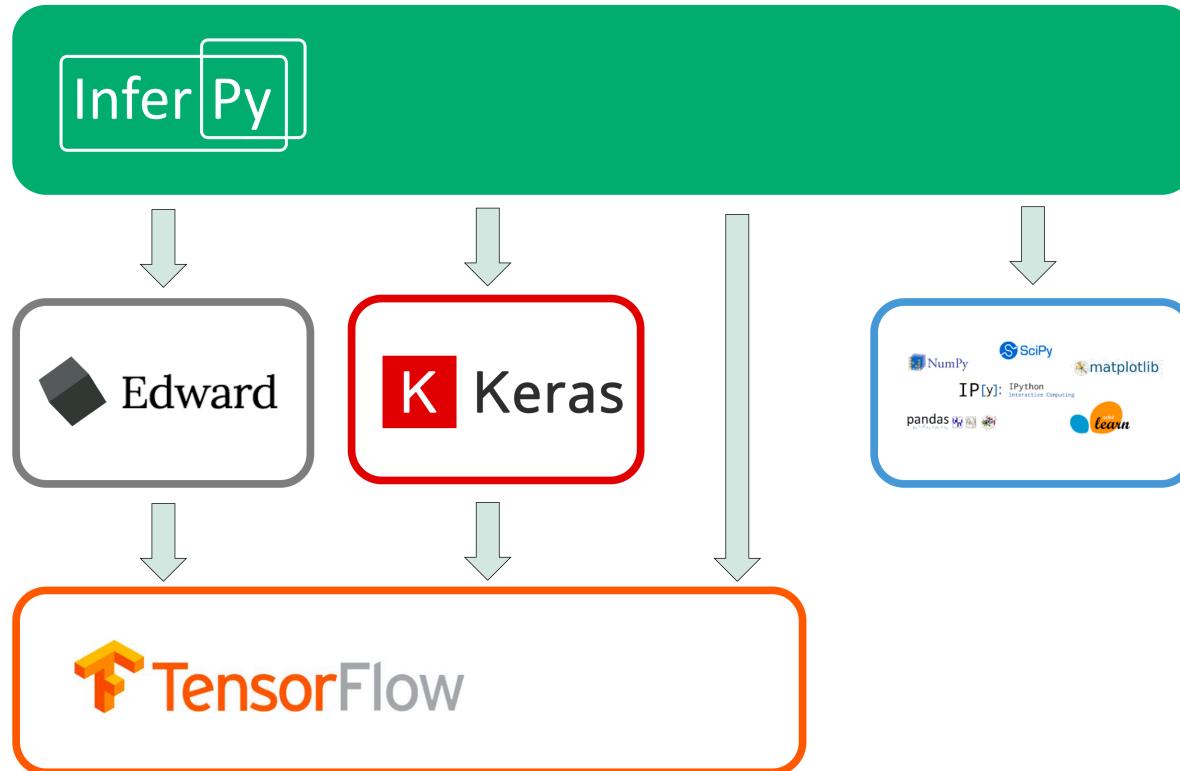


MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

ESTRATEGIA GENERAL DEL PROYECTO



Motivar científicamente el proyecto en el contexto de la línea de trabajo del equipo y del tutor y de los resultados previos disponibles, describir la hipótesis de partida y la estrategia general del proyecto.



OBJETIVOS PROPUESTOS Y ALCANZADOS

Paquete de Trabajo	Grado Consecución	Grado Esperado
PT2. Sintaxis y Semántica	90%	95%
PT3. Compilador	90%	95%
PT4. Motor de Inferencia	40%	95%
PT5. Librería de Modelos	20%	95%
PT6. Difusión y Explotación	40%	100%

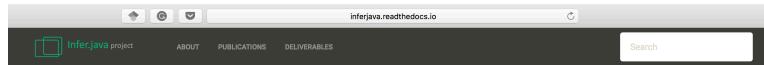
En esta diapositiva se indicarán de forma sucinta los objetivos propuestos en la memoria original, los posibles cambios como consecuencia del progreso de la investigación; **indicar el grado de consecución de los mismos** (indicar porcentaje aproximado) y los que se prevén alcanzar en la duración prevista del proyecto. **En caso de estimarse la necesidad de prórroga, debe indicarse aquí. NO hay prórroga en los JIN, salvo por maternidad.**

Explique la relación de los objetivos con la de otros proyectos del tutor o del equipo con el que se relaciona o en el que se ha integrado, si procede.



RESULTADOS CIENTÍFICO-TÉCNICOS

PT0. Gestión del Proyecto



In this project we aim to develop a probabilistic programming language that supports a distributed computational model. This programming language will be developed as an open software project freely available to the scientific and professional community interested in the use of machine learning techniques (ML) for big data problems.

This project has been developed by the University of Almería and has received funding from the Spanish Ministry of Economy, Industry and Competitiveness under the project TIN2015-74368-JIN.



© Copyright 2017, Andrés R. Masegosa, Rafael Cabañas.
Created using Sphinx 1.6.5.

[Back to top](#)



Rafael Cabañas de Paz
Dóctor en Ciencias Computación

Describir en 3-4 diapositivas las principales actividades realizadas y los resultados obtenidos en cada uno de los objetivos del proyecto, así como las dificultades encontradas o los resultados no previstos. **Valore la relevancia de los mismos. Indique publicaciones representativas derivadas de cada uno de ellos con sus indicadores habituales (IF, cuartil, GII-GRIN-SCIE, etc.). Si se planteó un análisis de riesgos en la solicitud, mencionar si se ha aplicado algún punto del plan de contingencia.**



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

RESULTADOS CIENTÍFICO-TÉCNICOS

PT1. Estado del Arte y Requerimientos Funcionales

An Introduction to Probabilistic Modeling with Deep Neural Networks

Andrés R. Masegosa¹ Rafael Cabañas de Paz¹ Antonio Salmerón¹

Abstract

Recent advances in variational inference are significantly expanding the toolbox of probabilistic modeling. Historically, variational inference (and probabilistic modeling) has been restricted to small or medium data sets which fit within the main memory of the computer, and to distributions families belonging to the conjugate exponential family, where variational updating equations can be computed in closed-form. Two main advances are helping variational methods

At the same time, the inference problem, as the problem of computing the posterior probability over hidden quantities given the known evidence, has been the corner-stone (and the bottleneck) of the feasibility and applicability of PGMs. There have been many advances in the inference part during this time.

At the beginning, the first proposed inference algorithms were able to compute this posterior in a exact way by exploiting the conditional independence relationships encoded by the graphical structure of the model. Even

Cabañas-Paz R., Masegosa A.R, Salmerón A. An introduction to Probabilistic Modelling with Deep Neural Networks. Submitted to the International Journal of Approximate Reasoning. Q1.

Describir en 3-4 diapositivas las principales actividades realizadas y los resultados obtenidos en cada uno de los objetivos del proyecto, así como las dificultades encontradas o los resultados no previstos . Valore la relevancia de los mismos. Indique publicaciones representativas derivadas de cada uno de ellos con sus indicadores habituales (IF, cuartil, GII-GRIN-SCIE, etc.). Si se planteó un análisis de riesgos en la solicitud, mencionar si se ha aplicado algún punto del plan de contingencia.



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

RESULTADOS CIENTÍFICO-TÉCNICOS

PT2. Sintáxis y Semántica, PT3. Compilador, PT4. Motor Inferencia y PT5. Librería de Modelos

The screenshot shows the GitHub repository interface for 'PGM-Lab / InferPy'. It displays the 'Releases' tab for version 0.1.0, which was released by 'rcabanadepaz' 24 days ago. The release notes mention that InferPy is a high-level API for probabilistic modeling written in Python and capable of running on top of Edward and Tensorflow. It is strongly inspired by Keras and has a focus on enabling flexible data processing, easy-to-code probabilistic modeling, scalable inference and robust model validation. The changes log includes fixes for bugs and additions of new distributions like Gamma, Bernoulli, inverseGamma, Laplace.

The screenshot shows the official InferPy documentation website at inferpy.readthedocs.io. The main page features a large 'InferPy' logo. The navigation menu includes links to Documentation, Quick Start, Getting Started, Guiding Principles, Guides, Model Zoo, Package Reference, inferpy package, and inferpy.models package. A sidebar for 'Rollbar' provides information about error monitoring and reporting. The main content area describes InferPy as a high-level API for probabilistic modeling written in Python, capable of running on top of Edward and Tensorflow. It is strongly inspired by Keras and focuses on enabling flexible data processing, easy-to-code probabilistic modeling, scalable inference, and robust model validation. The page also highlights that InferPy is a probabilistic programming language that allows for fast prototyping of hierarchical probabilistic models, creates computational efficient batched models, and runs seamlessly on CPU and GPU.

- **Masegosa A.R , Cabañas-Paz R., Salmerón A. Deep Probabilistic Modelling Made Easy. BayesComp Conference 2018.**

Describir en 3-4 diapositivas las principales actividades realizadas y los resultados obtenidos en cada uno de los objetivos del proyecto, así como las dificultades encontradas o los resultados no previstos . **Valore la relevancia de los mismos. Indique publicaciones representativas derivadas de cada uno de ellos con sus indicadores habituales (IF, cuartil, GII-GRIN-SCIE, etc.). Si se planteó un análisis de riesgos en la solicitud, mencionar si se ha aplicado algún punto del plan de contingencia.**

RESULTADOS CIENTÍFICO-TÉCNICOS

• PT4. Motor de Inferencia

- **Masegosa AR et al.** Bayesian models of data streams with hierarchical power priors. **ICML 207. A++.**
- **Ramos-López D.** et al. Scalable importance sampling estimation of Gaussian mixture posteriors in Bayesian networks. Accepted for publication at **International Journal of Approximate Reasoning. 2018. Q1**
- **Dogadov S.** et al. Variational Robust Subspace Clustering with Mean Update Algorithm. Second International Workshop on Robust Subspace Learning and Applications in Computer Vision. 2017.

Describir en 3-4 diapositivas las principales actividades realizadas y los resultados obtenidos en cada uno de los objetivos del proyecto, así como las dificultades encontradas o los resultados no previstos. Valore la relevancia de los mismos. Indique publicaciones representativas derivadas de cada uno de ellos con sus indicadores habituales (IF, cuartil, GII-GRIN-SCIE, etc.). Si se planteó un análisis de riesgos en la solicitud, mencionar si se ha aplicado algún punto del plan de contingencia.



RESULTADOS CIENTÍFICO-TÉCNICOS

- **PT4. Motor de Inferencia**
 - Submitted
 - **Dogadov S**, et al. Global, Efficient, and Scalable Solvers for Variational Bayesian Low-rank Subspace Clustering. Submitted to **Journal of Selected Subspace Clustering. Q1.**
 - **Masegosa AR** et al. AMIDST: A Java Toolbox for Probabilistic Machine Learning. Submitted to **Knowledge Based Systems. Q1.**
 - **Masegosa AR** et al. Analysing Concept Drift: A case study in the financial sector. Submitted to **Plos-One. Q1.**
 - About to be submitted
 - **Masegosa AR et al.** Learning Latent Variable Models from non-stationary data streams. To be submitted to **IEEE PAMI. Q1.**
 - **Masegosa AR** et al. Representing Bayesian Networks as Conjugate Exponential Family Models.
 - **Masegosa AR** et al. Data reduction for Variational Inference in Latent Variable Models over Massive Datasets.

Describir en 3-4 diapositivas las principales actividades realizadas y los resultados obtenidos en cada uno de los objetivos del proyecto, así como las dificultades encontradas o los resultados no previstos. Valore la relevancia de los mismos. Indique publicaciones representativas derivadas de cada uno de ellos con sus indicadores habituales (IF, cuartil, GII-GRIN-SCIE, etc.). Si se planteó un análisis de riesgos en la solicitud, mencionar si se ha aplicado algún punto del plan de contingencia.



RESUMEN DE LOS RESULTADOS DEL PROYECTO

	Número	Indicios de calidad
Artículos científicos derivados del proyecto en revistas JCR ó Scopus (de ellos cuantos en acceso abierto)	1 (1?)	Q1
	4 Submitted	4 Q1
Revisiones (surveys), editoriales y otros artículos científicos (de ellos cuantos en acceso abierto)		
Libros, capítulos de libros y monografías (nac/internac)		
Conferencias en congresos (nacionales/internac, indicando cuántas por invitación).	3 Internac.	1 A++
Patentes/Registros Software (indicar estado)		
Otros resultados		

Indique únicamente aquellos resultados que derivan directamente del presente proyecto. Puede indicar por separado los derivados de colaboraciones, revisiones, etc.



INTERNACIONALIZACIÓN DE LA INVESTIGACIÓN

1. Colaboración Científica con Investigadores Internacionales



2. Estancias de Investigación (2 Meses en 2017 + 1.5 meses 2018)



3. Propuesta de Proyecto H2020: ANODYNE

AnoDyne



Vestas



Relacionar las colaboraciones internacionales relacionadas con el proyecto y la relevancia para el mismo y para el equipo investigador
Participación en proyectos europeos e internacionales, indicando Programa, convocatoria, tipo de proyecto y resultado de la propuesta presentada. Indicar financiación recibida en el grupo.



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

Otros aspectos destacables relacionados con el proyecto



DataBeers Almería

BayesComp 2018

Seminario Mayo 2018

P.ej. Organización de eventos, actividades de divulgación, etc.



MINISTERIO DE ECONOMIA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.

EJECUCION DEL PRESUPUESTO

Concepto	Ejecutado: Cantidad y (%) respecto de la solicitud presentada	Existen cambios relevantes respecto a solicitud original? (*)
Personal (indicar)		SI
Inventariable (indicar principales equipos)		NO
Fungible		SI
Viajes y dietas (indicar principales conceptos)		NO

En esta diapositiva se indicará cómo se ha ejecutado (o está ejecutando) el gasto en relación con la solicitud presentada. No hace falta que las cantidades sean exactas, sino indicativas de la ejecución del gasto realizado hasta el momento. Indique el remanente no gastado hasta la fecha.

(*) Gastos no contemplados en la solicitud original: explique la justificación del gasto en relación con el presupuesto inicial y justifique su necesidad por razones científico-técnicas



PLANTEAMIENTO FUTURO

1. Propuesta de Proyecto H2020: ANODYNE

ANODYNE



AALBORG UNIVERSITET



Norwegian University of
Science and Technology



HUGINEXPERT

Vestas



1. Propuesta de Proyecto H2020: ANODYNE

Si tiene previsto solicitar proyecto en una próxima convocatoria, explique la posible relación (continuidad de la investigación) y diferencias con el presente proyecto.



MINISTERIO DE ECONOMÍA,
INDUSTRIA Y
COMPETITIVIDAD



JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnicos.