
 <b>Infer.java</b> project	<b>TIN2015-74368-JIN</b>  01/02/2017	 GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD  UNIÓN EUROPEA Fondo Europeo de Desarrollo Regional "Una manera de hacer Europa"
--	--	--

**Reference:** TIN2015-74368-JIN

**Project full title:** INFER.JAVA: UN LENGUAJE DE PROGRAMACION. PROBABILISTICO PARA EL DESARROLLO DE APLICACIONES

**Deliverable no.:** E0

**Title of the deliverable:** Informe de Seguimiento del Proyecto

<b>Author(s):</b>	Andrés R. Masegosa, Rafael Cabañas de Paz
<b>Version:</b>	0.1

**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



# PARTICIPANTES

## Entidades participantes:

**Universidad de Almería**

## Investigadores participantes:

**Andrés R. Masegosa (IP)**

**Antonio Salmerón (Tutor)**

**Rafael Cabañas (Contratado)**



# MOTIVACIÓN

- **Aprendizaje Automático (ML):**
  - Ciencias de la Salud, Internet, Robótica, Conducción autónoma, etc.
  - Tecnología disruptiva.

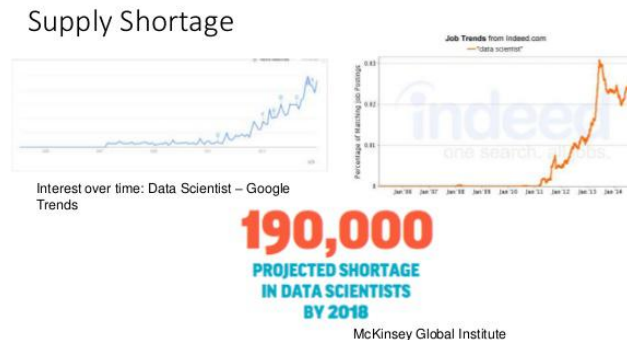


- **Grupo de Investigación:**
  - Probabilistic Machine Learning.
  - Proyecto Europeo AMIDST.



# MOTIVACIÓN

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

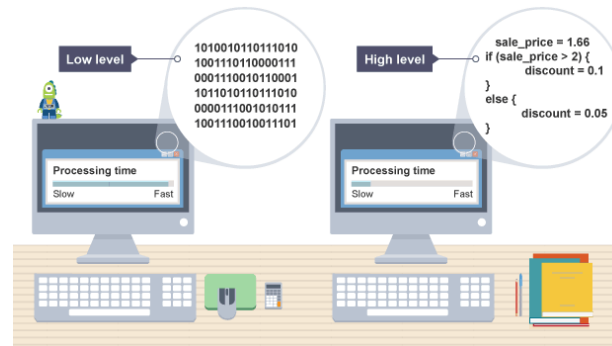


- El desarrollo de sistemas ML requiere grandes **recursos financieros y técnicos**.
- Las **grandes corporaciones** tienen ventaja competitiva.



# 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

# ESTRATEGIA GENERAL DEL PROYECTO

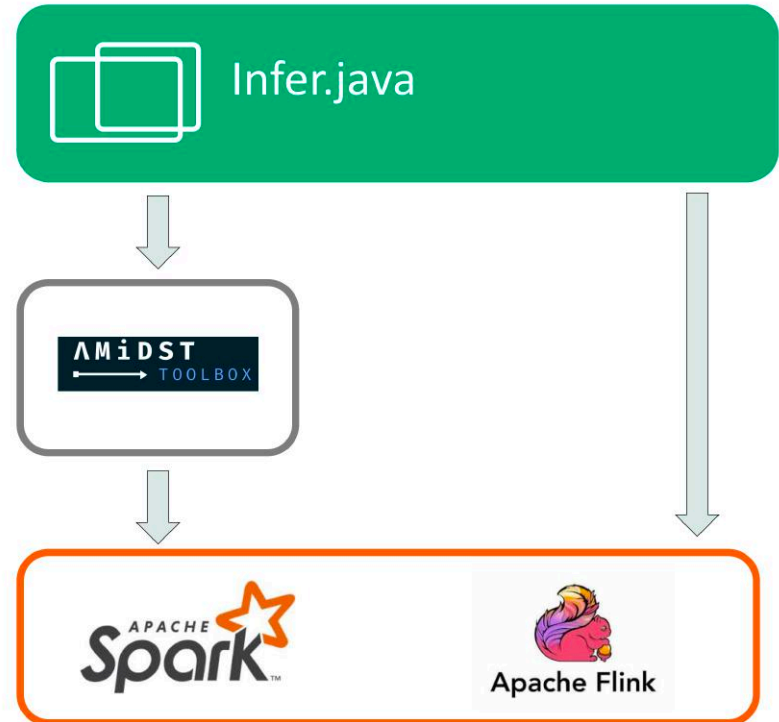
```
//Firstly, the probabilistic model is defined.
Variable fire = Engine.newRareEventVariable();
Variable temperature = Engine.newGaussianVariable().dependsPositivelyOf(fire);
Variable smoke = Engine.newEventVariable().dependsPositivelyOf(fire);
Variable tempSensor = Engine.newGaussianVariable().dependsNoisilyOf(temperature);
Variable smokeSensor = Engine.newEventVariable().dependsNoisilyOf(smoke);

//Secondly, the probabilistic model is learnt with historical data
Stream<SensorMeasures> historicalData = Engine.loadData("./historicalData.csv");

historicalData.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");
});
```

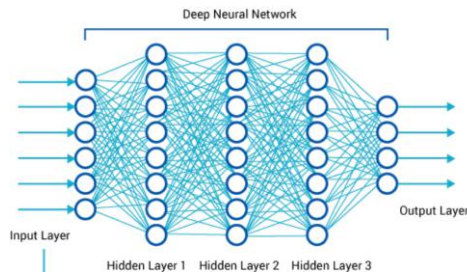
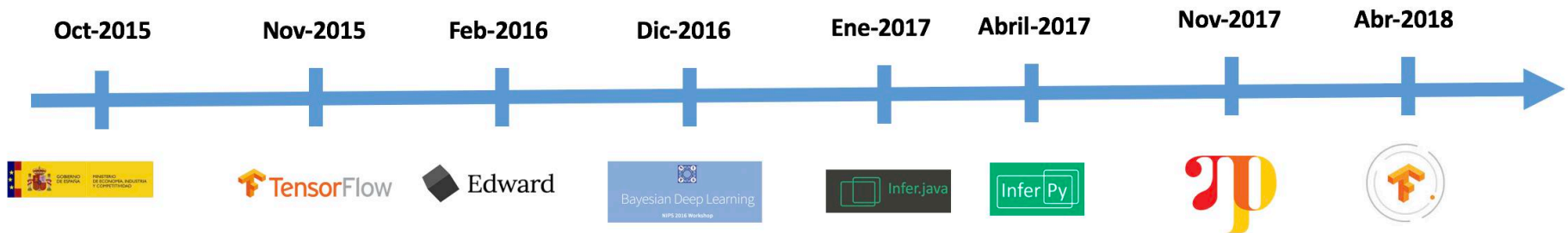


# OBJETIVOS PROPUESTOS Y ALCANZADOS

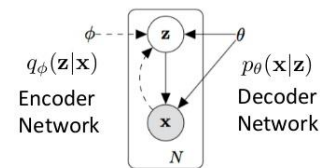
Paquete de Trabajo	Grado Consecución	Grado Esperado
PT0. Gestión Proyecto	50%	100%
PT1. Estado Del Arte Y Requerimientos	90%	95%
PT2. Sintaxis y Semántica	90%	95%
PT3. Compilador	90%	95%
PT4. Motor de Inferencia	40%	95%
PT5. Librería de Modelos	20%	95%
PT6. Difusión y Explotación	30%	100%



# ESTRATEGIA GENERAL DEL PROYECTO



## Variational Autoencoder

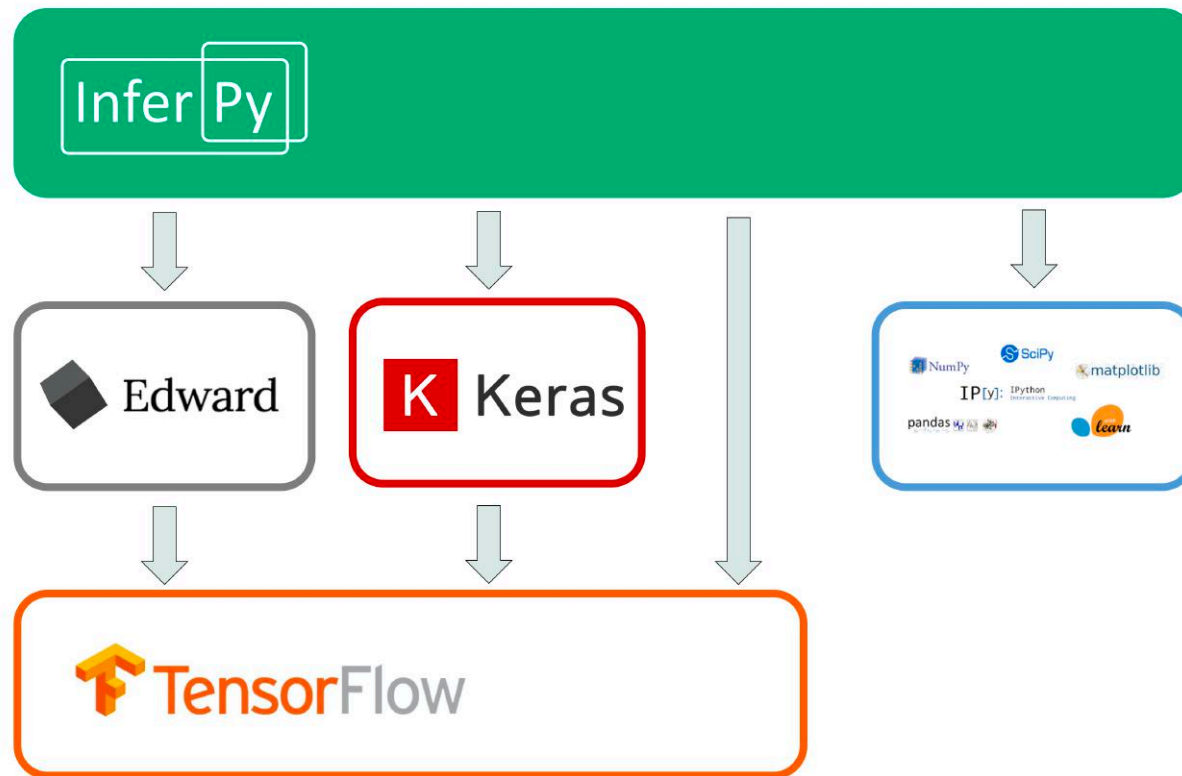


$$\text{Minimize: } D_{KL}[q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z}|\mathbf{x})]$$

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

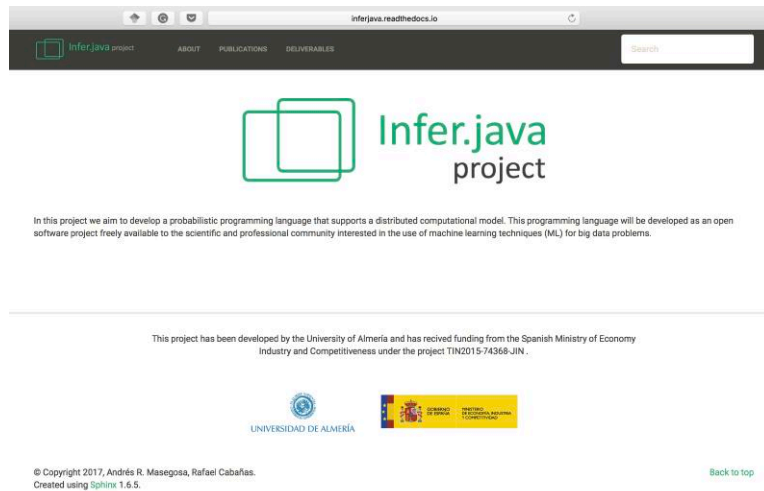


# ESTRATEGIA GENERAL DEL PROYECTO



# RESULTADOS CIENTÍFICO-TÉCNICOS

## PT0. Gestión del Proyecto



Rafael Cabañas de Paz  
Dóctor en Ciencias Computación

<http://inferjava.readthedocs.io/>



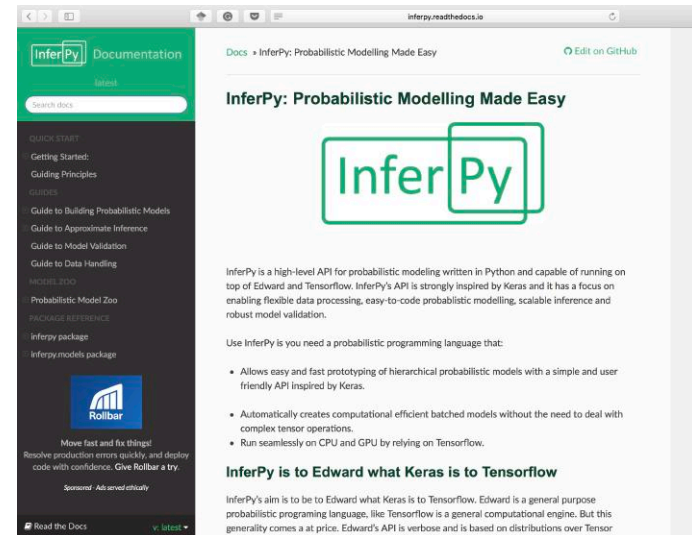
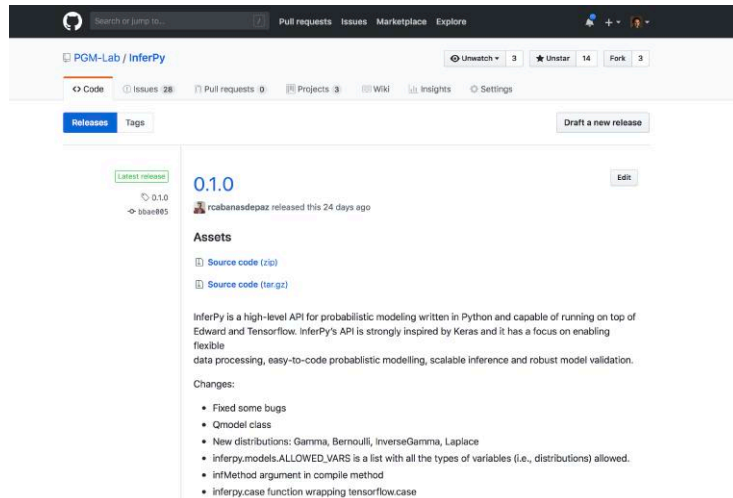
MINISTERIO DE ECONOMIA,  
INDUSTRIA Y  
COMPETITIVIDAD



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

# RESULTADOS CIENTÍFICO-TÉCNICOS

First Release on May 2018



<https://github.com/PGM-Lab/InferPy/>

<http://inferpy.readthedocs.io/>

```
$ pip install inferpy
```



MINISTERIO DE ECONOMIA,  
INDUSTRIA Y  
COMPETITIVIDAD



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

# RESULTADOS CIENTÍFICO-TÉCNICOS

- PT1. Estado del Arte
  - **Entregable E1**
  - **Publicación:** Cabañas-Paz R., Masegosa A.R, Salmerón A. An introduction to Probabilistic Modelling with Deep Neural Networks. To be submitted to the International Journal of Approximate Reasoning. Q1
- PT2 y PT3.
  - **Entregables E2 y E3.**
- PT9. Difusión y Explotación
  - **Publicación:** Masegosa A.R , Cabañas-Paz R., Salmerón A. Deep Probabilistic Modelling Made Easy. **BayesComp Conference 2018.**



# RESULTADOS CIENTÍFICO-TÉCNICOS

- PT4. Motor de Inferencia
  - 3 artículos publicados.
    - **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. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1792-1799).
  - 3 artículos enviados a revistas Q1.
  - 3 artículos que van a ser enviados a revistas antes de las vacaciones de verano.



# 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 (0)	Q1
	4 en Revisión	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



MINISTERIO DE ECONOMIA,  
INDUSTRIA Y  
COMPETITIVIDAD

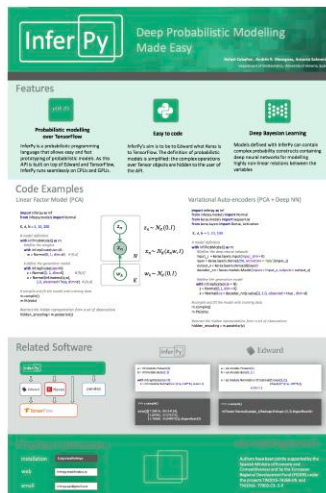


JORNADAS DE SEGUIMIENTO 2018. Subdivisión de Programas Temáticos Científico Técnico.

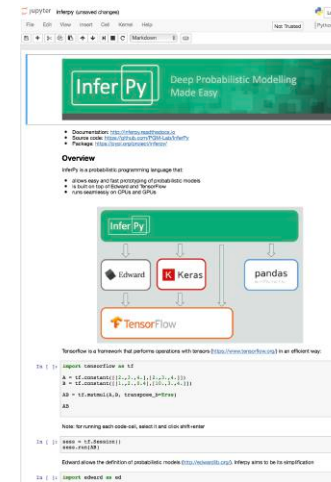
# Otros aspectos destacables relacionados con el proyecto



DataBeers Almería



BayesComp 2018



Seminario Mayo 2018

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



# EJECUCION DEL PRESUPUESTO

Concepto	Ejecutado: Cantidad y (%) respecto de la solicitud presentada	Existen cambios relevantes respecto a solicitud original? (*)
Personal (indicar)	70.000 euros (47%)	NO
Inventariable (indicar principales equipos)	3300 euros (70%) [2 Portátiles]	NO
Fungible		
Viajes y dietas (indicar principales conceptos)	7.300 euros (67%)	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 (Abril 2018)

ANODYNE



## 2. Convocatorias Naciones y Regionales:

- Extender Inferpy para IoT y “edge computing”.
- Extender Inferpy para resolver problemas de Causalidad.

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

