



**EL IMPACTO DEL MACHINE LEARNING EN LA  
MEDICINA ACTUAL**  
- UNA VERSIÓN DE BOLSILLO-

*20 febrero, 2018*

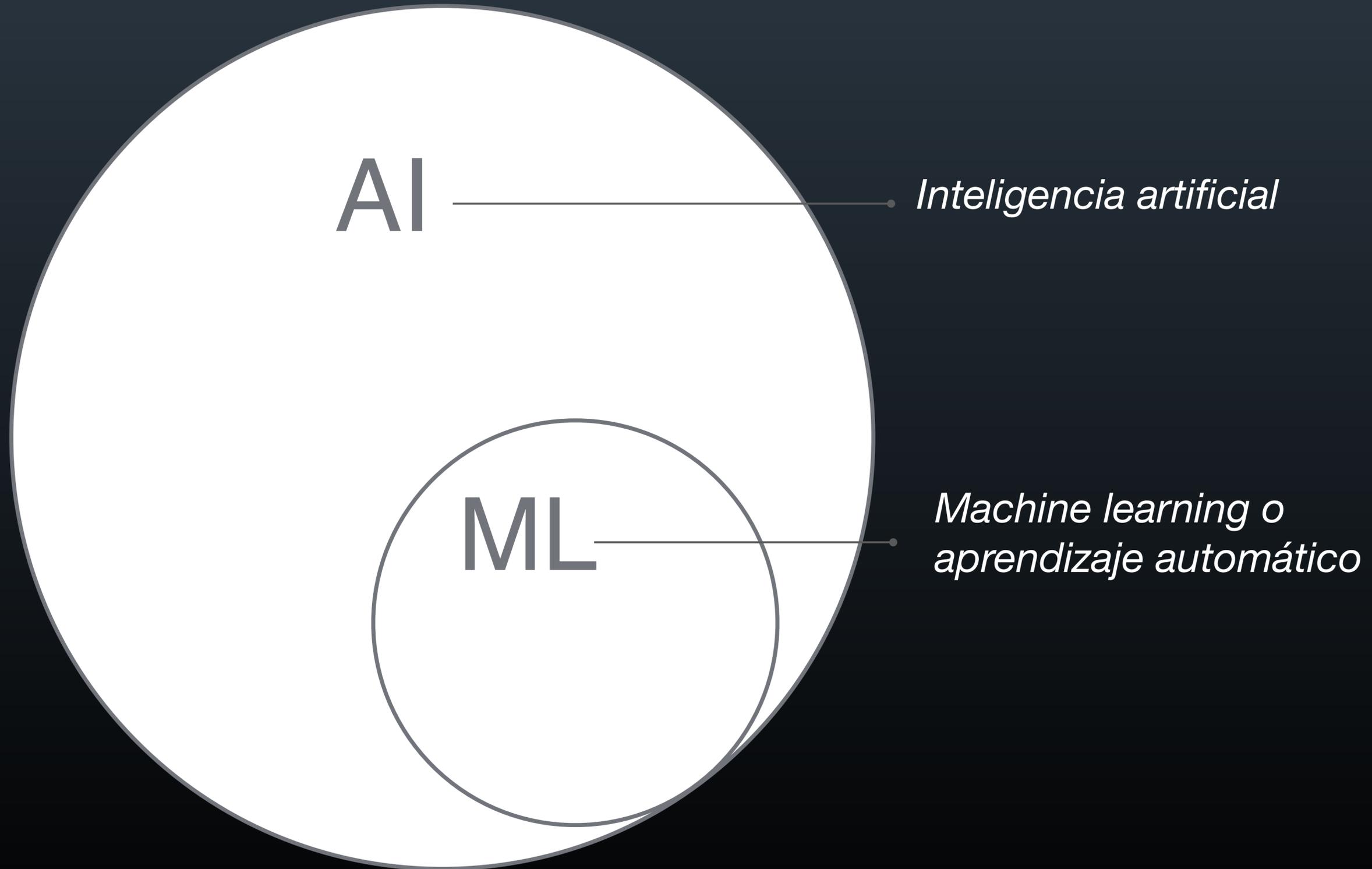
Irene Tortosa, Machine Learning Analyst

# Agenda

---

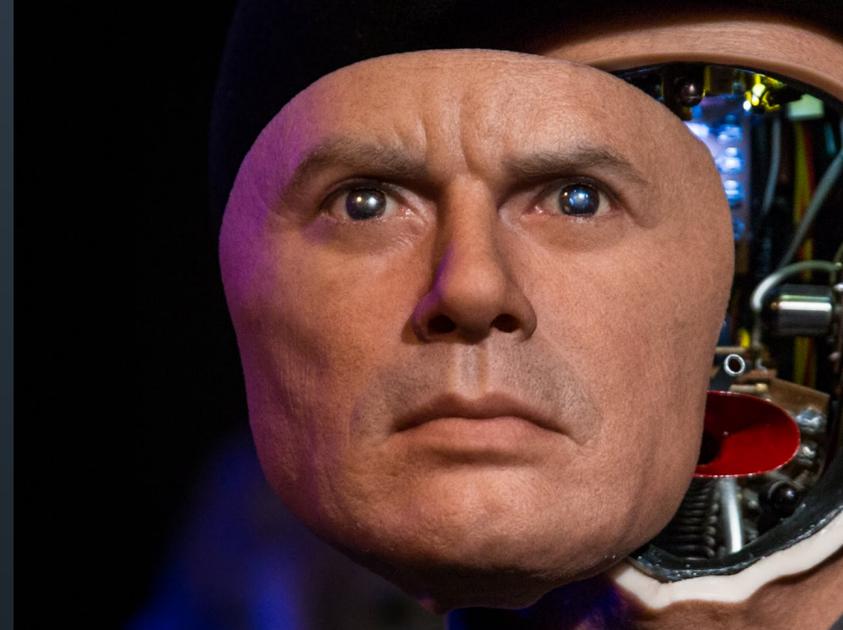
<b>1</b>	¿Qué es el <i>Machine Learning</i> ?
<b>2</b>	El Machine Learning en la medicina actual
<b>3</b>	Introducción a BigML
<b>4</b>	Casos de uso

1	¿Qué es el <i>Machine Learning</i> ?
2	El Machine Learning en la medicina actual
3	Introducción a BigML
4	Casos de uso



# AI

Películas

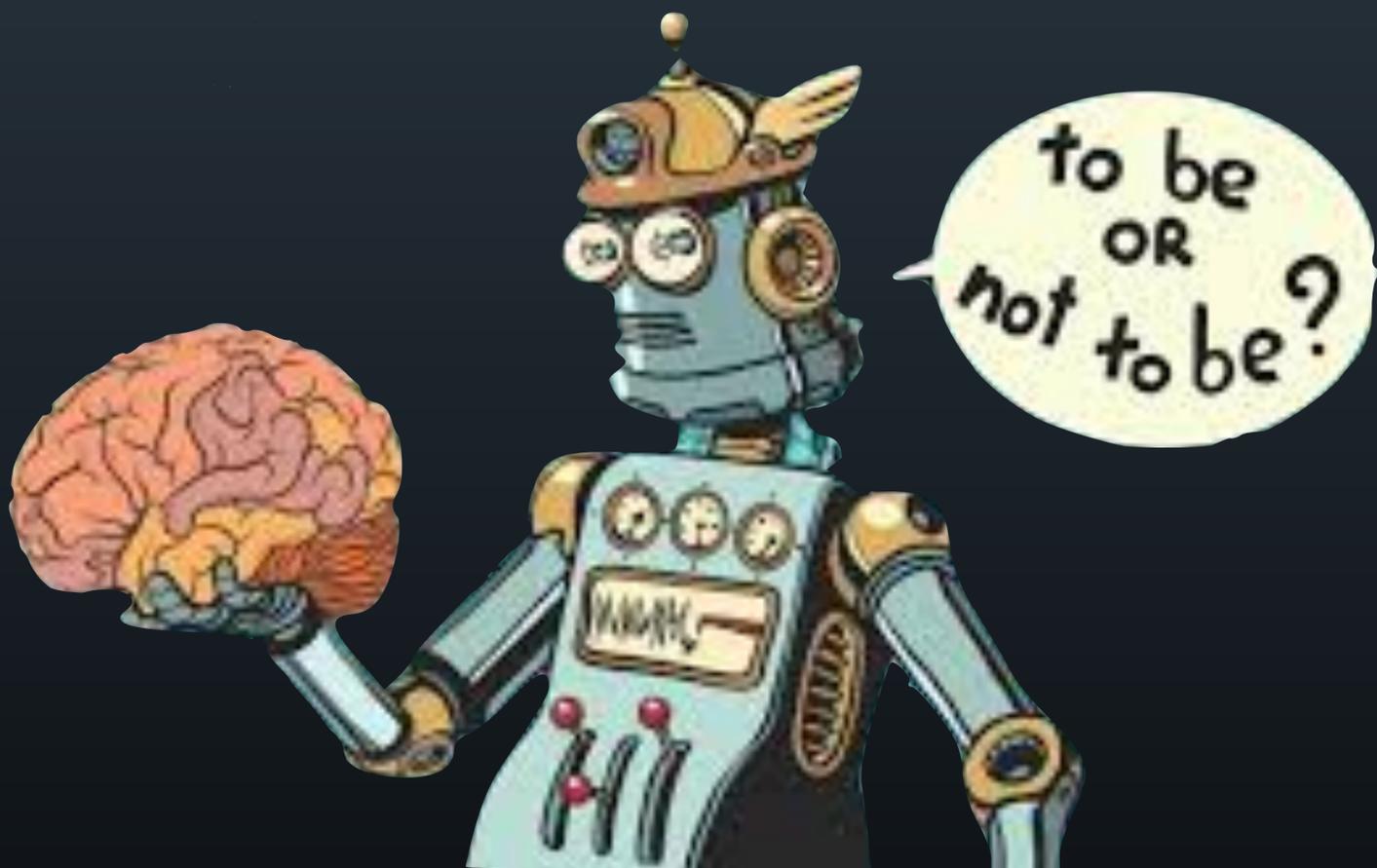


Juegos



Política





***THE HYPE***

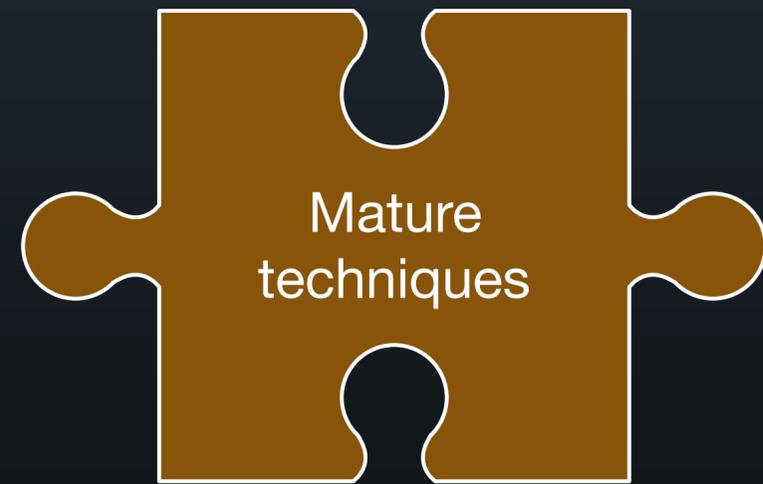


*“A field of study that gives computers the ability to learn without being explicitly programmed”*

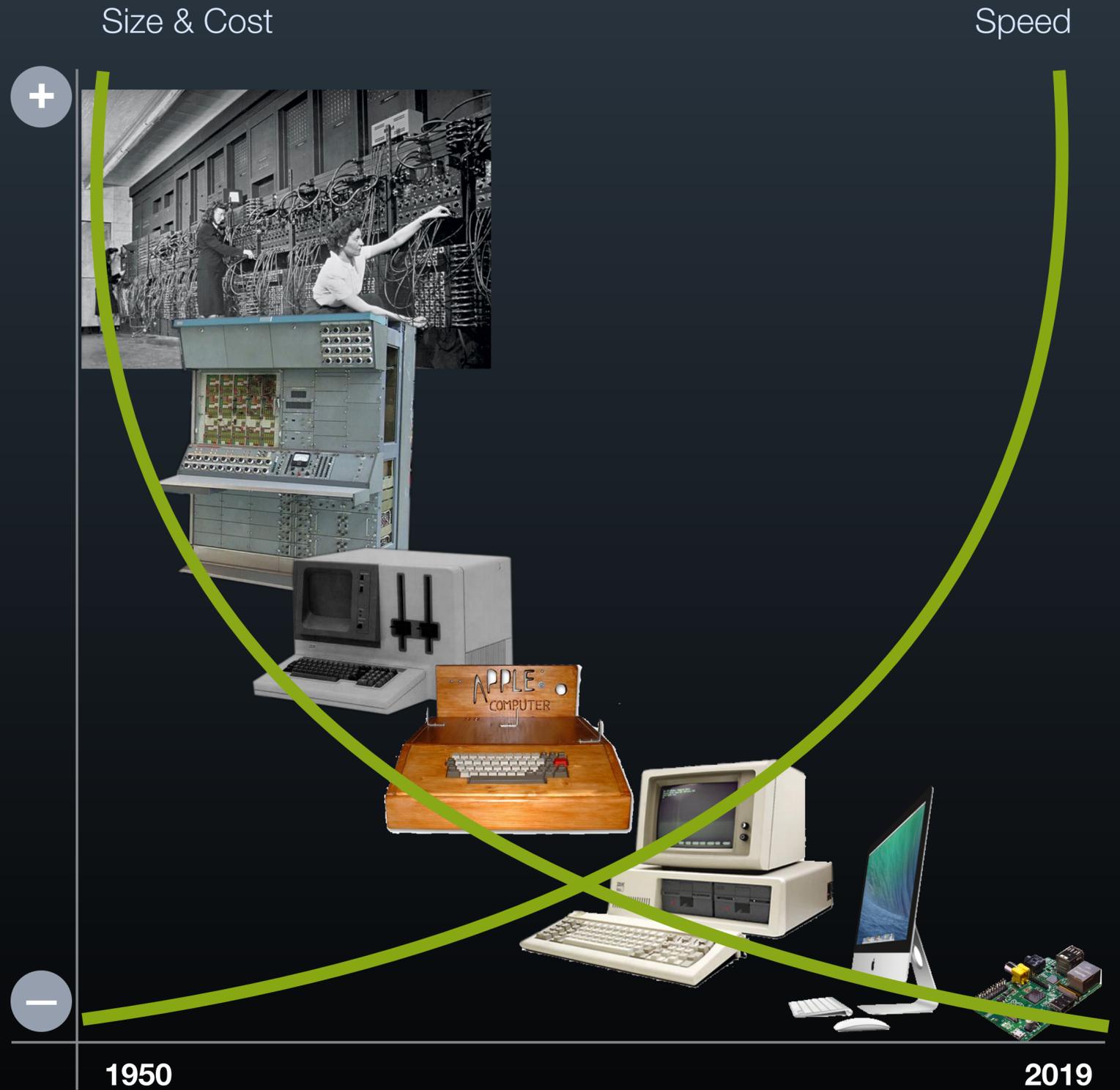
Profesor Arthur Samuel

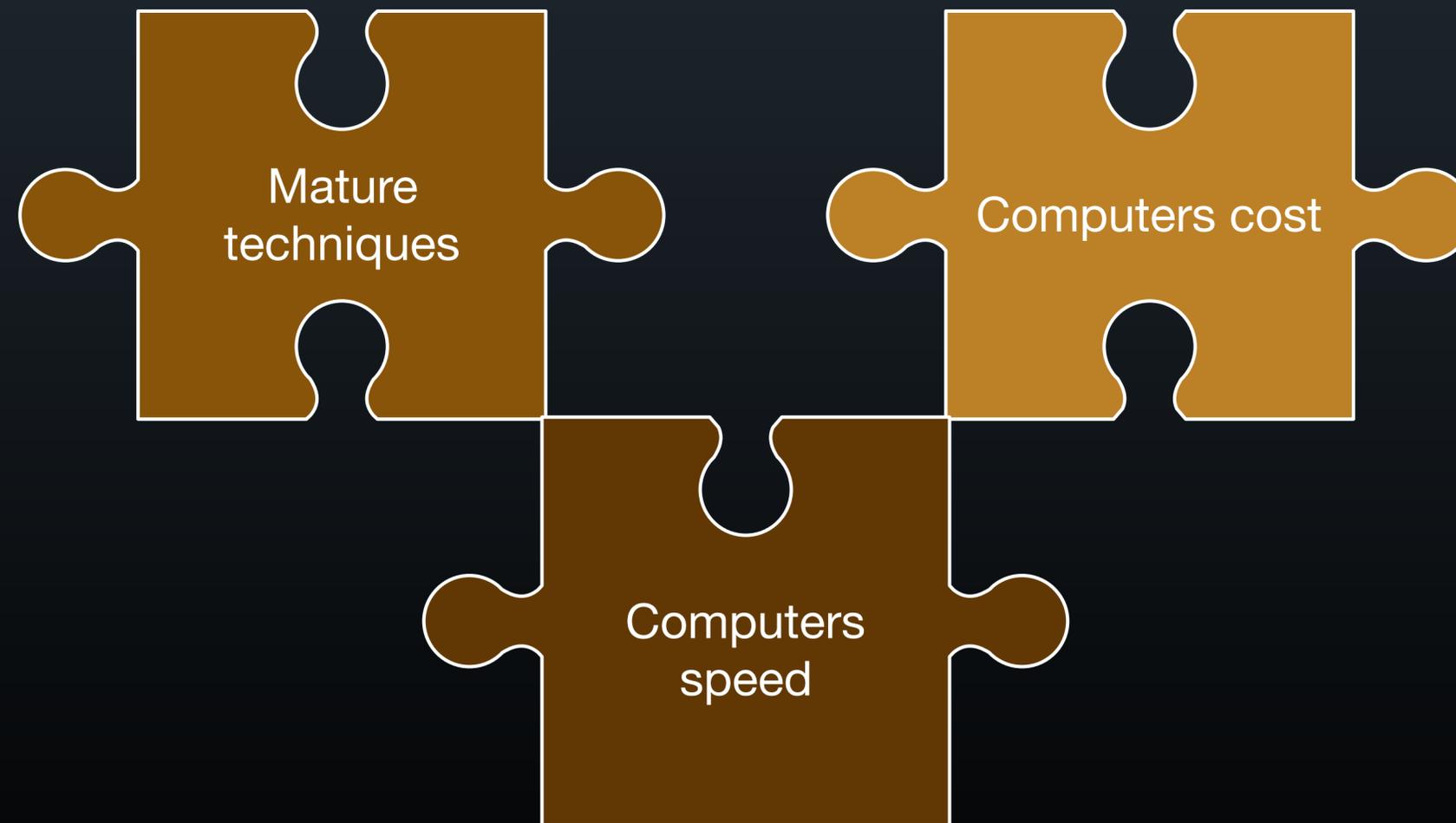
- Primer programa de aprendizaje automático desarrollado por Arthur Samuel para IBM en 1952 para jugar a las damas
- La primera conferencia de *Machine Learning* fue en Pittsburg en 1980

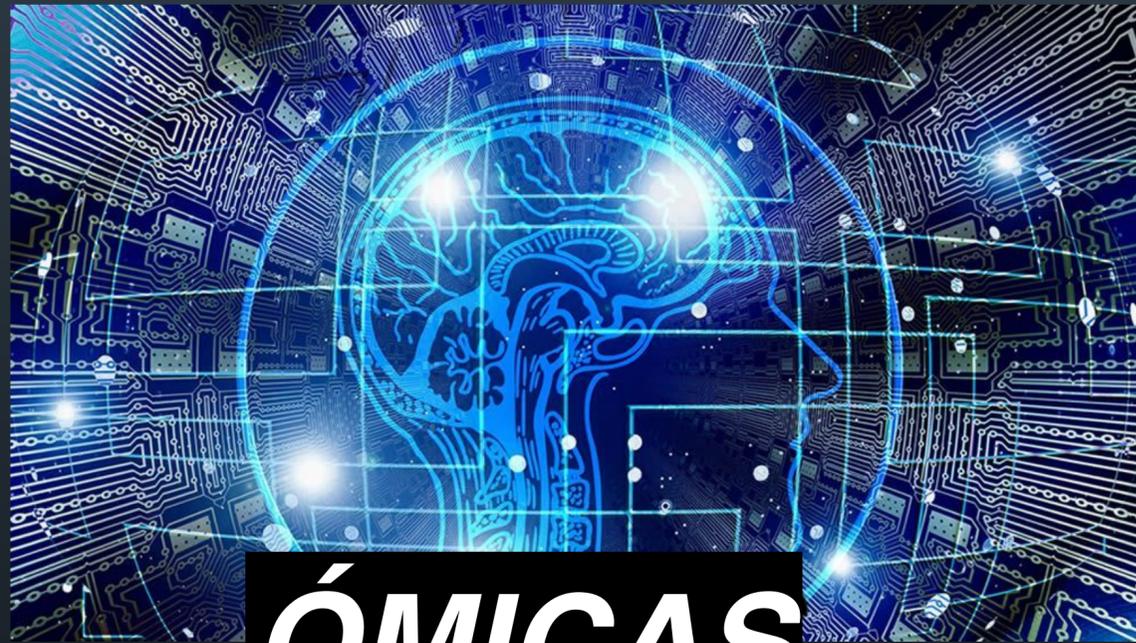
**¿Porqué ahora?**



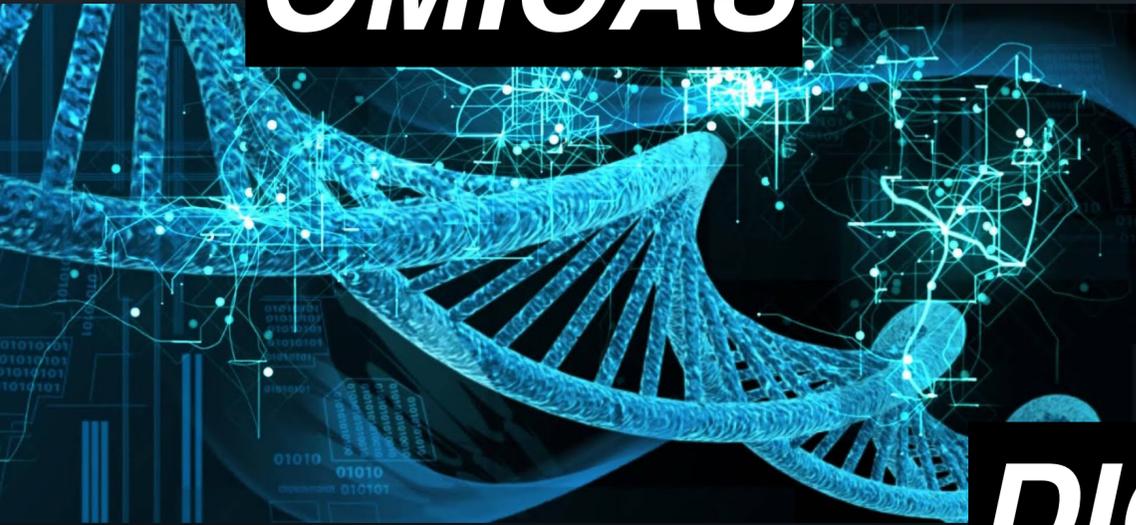
Abaratamiento del **coste** de los ordenadores y aumento del **poder de computación**







# ÓMICAS



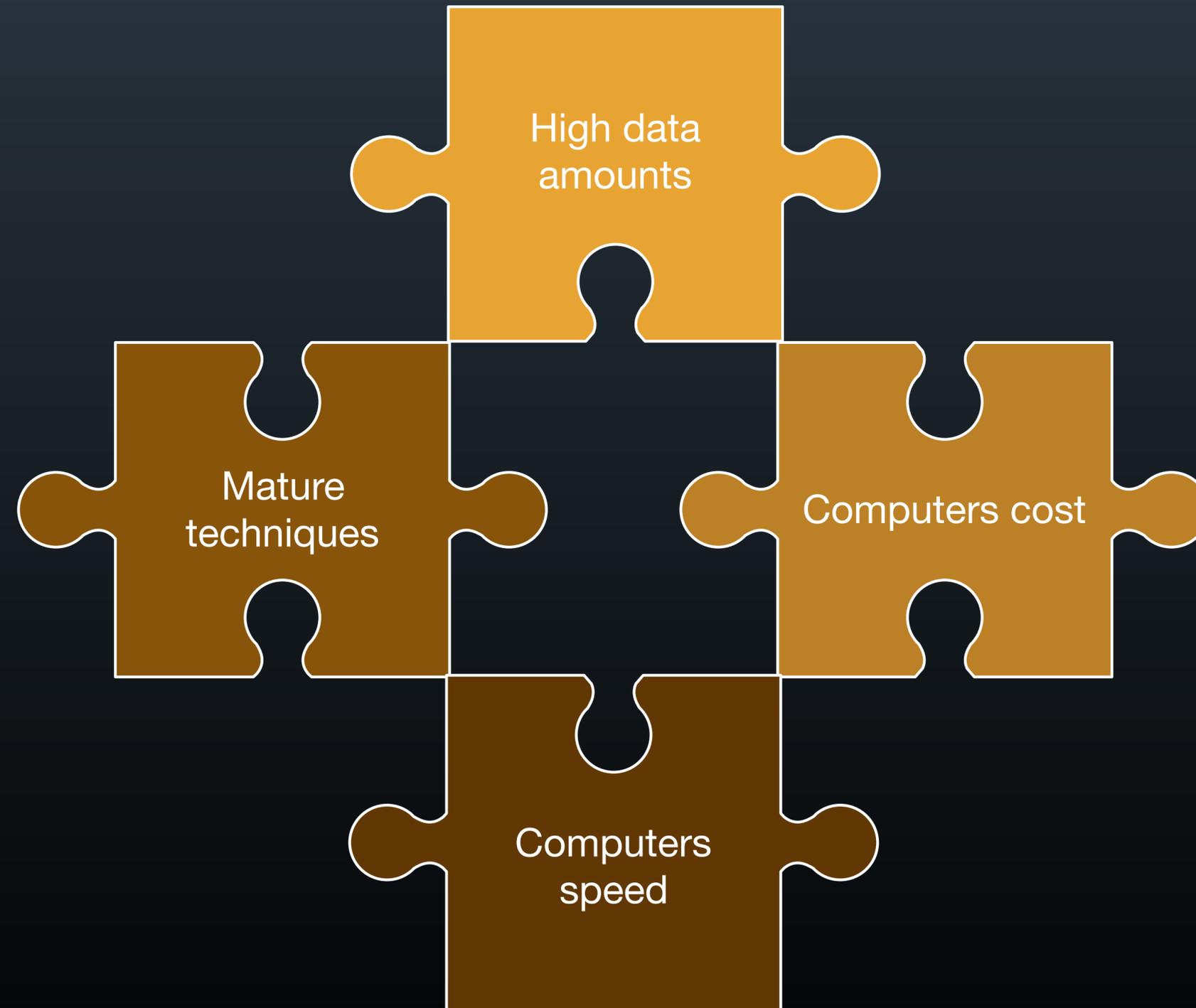
# DIGITALIZACIÓN



# WEARABLES

La digitalización de la historia clínica, un nuevo impulso a la investigación





La información esta siendo desaprovechada, tantos datos **exceden las capacidades humanas**





El ML encuentra **patrones** en grandes volúmenes de **datos**, generando modelos con una gran **capacidad predictiva**

# Agenda

---

1	¿Qué es el <i>Machine Learning</i> ?
2	El Machine Learning en la medicina actual
3	Introducción a BigML
4	Casos de uso

# Machine learning helps UI Health Care reduce surgical site infection by 74%, save \$1.2 million

And that's not counting the value-based savings for reducing those infections, says the machine learning co-developer the University of Iowa Hospitals & Clinics.

By [Bill Siwicki](#) | September 10, 2018 | 12:31 PM



*Herramienta para ayudar en la toma de decisiones, abaratamiento de costes*

- 3 años de estudio (finalizado 2016)
- El modelo calcula el **riesgo de infección** y lo relaciona con intervenciones específicas previniendo al doctor pre -y post-quirúrgico.
- **↓74% del riesgo de infección** en pacientes operados (general y colorectal)

## Johns Hopkins researchers use deep learning to combat pancreatic cancer

Early detection is key to treatment, and with AI-enabled detection methods nearly a third of pancreatic cancer cases could be found four to 12 months sooner, they say.

***Diagnóstico temprano y mayor precisión***

- Problema:
  - Dificultad en el diagnóstico - **diagnóstico tardío**
  - **Incremento de casos** que tiene que ver al radiólogo - **incremento de errores**
- Solución → ML Deep Learning o aprendizaje profundo
  - Miles de **TACs** de pacientes con cáncer de páncreas para **entrenar el modelo**
  - Aprende a detectar **mínimos cambios de textura** en páncreas y alrededores
  - Estiman un 30% de casos se podían haber detectado 4-12 meses antes

# PadChest: A large chest x-ray image dataset with multi-label annotated reports

Aurelia Bustos, Antonio Pertusa, Jose-Maria Salinas, Maria de la Iglesia-Vayá

(Submitted on 22 Jan 2019)

We present a labeled large-scale, high resolution chest x-ray dataset for the automated exploration of medical images along with their associated reports. This dataset includes more than 160,000 images obtained from 67,000 patients that were interpreted and reported by radiologists at Hospital San Juan Hospital (Spain) from 2009 to 2017, covering six different position views and additional information on image acquisition and patient demography. The reports were labeled with 174 different radiographic findings, 19 differential diagnoses and 104 anatomic locations organized as a hierarchical taxonomy and mapped onto standard Unified Medical Language System (UMLS) terminology. Of these reports, 27% were manually annotated by trained physicians and the remaining set was labeled using a supervised method based on a recurrent neural network with attention mechanisms. The labels generated were then validated in an independent test set achieving a 0.93 Micro-F1 score. To the best of our knowledge, this is the largest public chest x-ray database suitable for training supervised models concerning radiographs, and the first to contain radiographic reports in Spanish. The PadChest dataset can be downloaded from [this http URL](http://bimcv.cipf.es/bimcv-projects/padchest/).

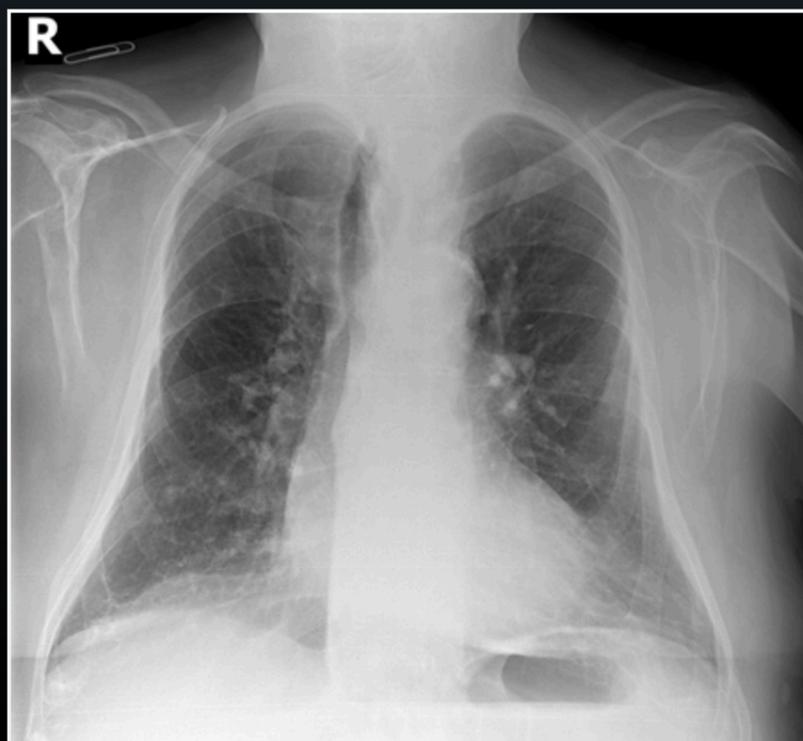
Subjects: **Image and Video Processing (eess.IV)**; Computer Vision and Pattern Recognition (cs.CV)

MSC classes: 92B20, 92C50, 68T50, 92B10

Cite as: [arXiv:1901.07441](https://arxiv.org/abs/1901.07441) [eess.IV]

(or [arXiv:1901.07441v1](https://arxiv.org/abs/1901.07441v1) [eess.IV] for this version)

<http://bimcv.cipf.es/bimcv-projects/padchest/>

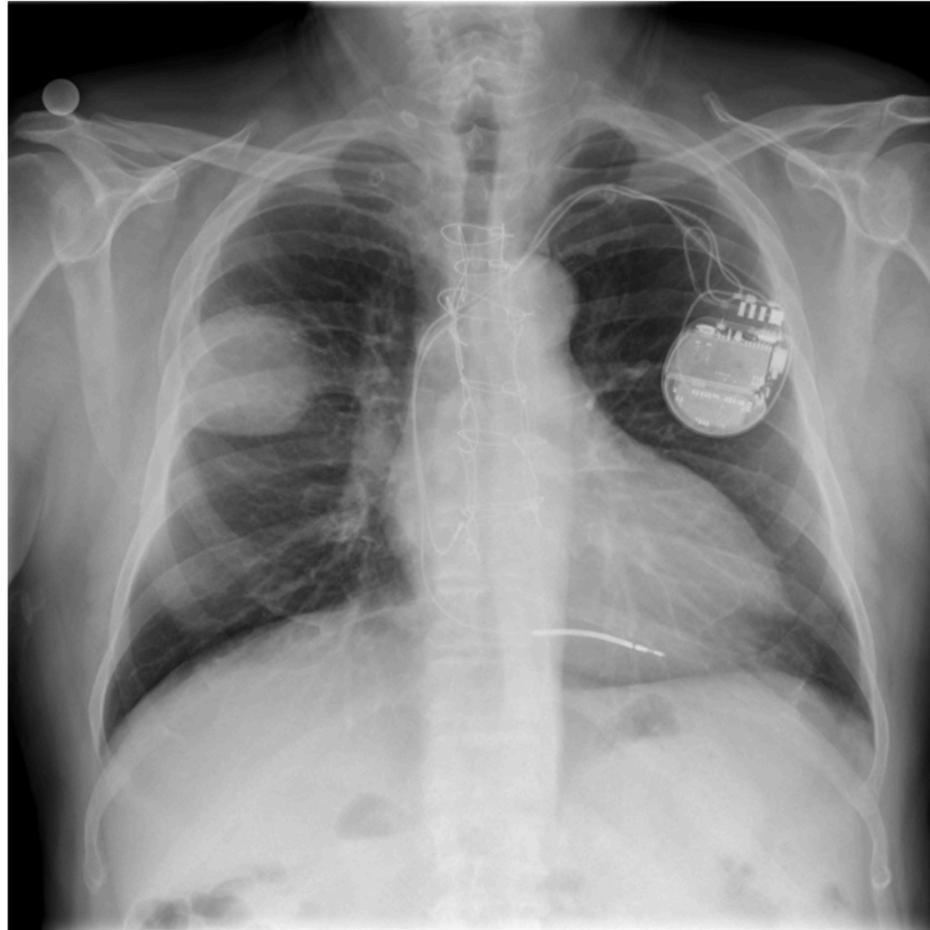


- +160.000 radiografías de 67.000 pacientes
- (2009-2017) Hospital San Juan (Alicante)
- Radiólogos 27% diagnóstico manual + **red neuronal supervisada**

→ 0.93 F-measure = medida del rendimiento

# Etiqueta imágenes de acuerdo con lo que ha aprendido

PA



L



*Optimización de tiempo, herramienta de ayuda*

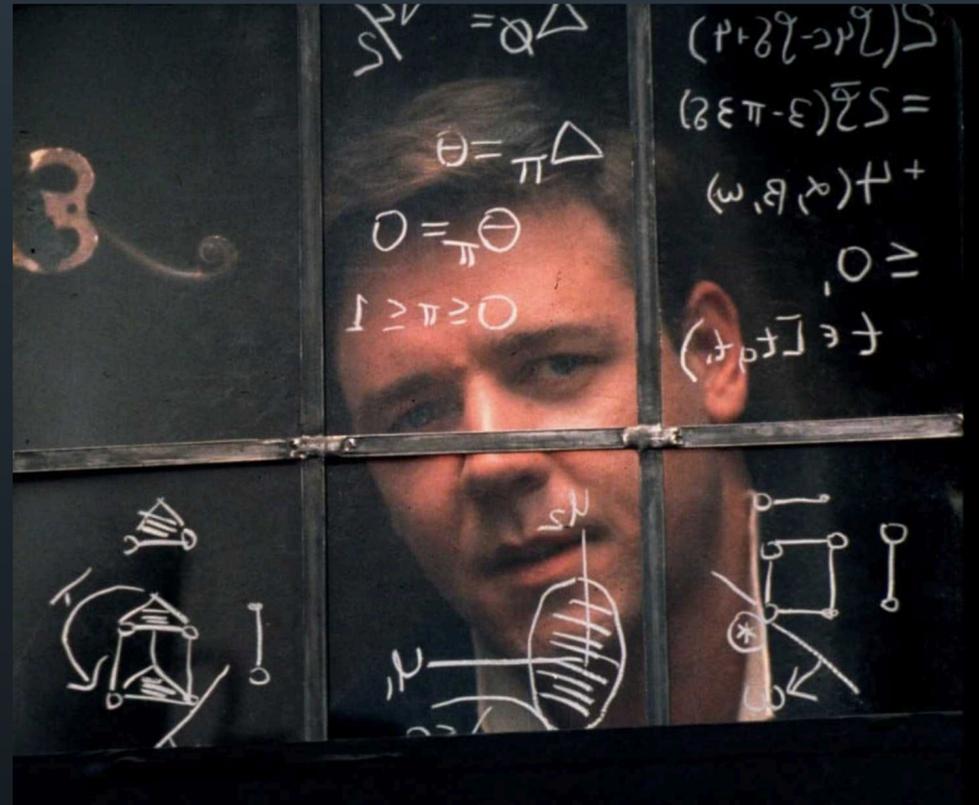
**Labels** ['pulmonary mass', 'pacemaker', 'cardiomegaly', 'vascular hilar enlargement', 'sternotomy', 'dual chamber device', 'suture material']

**Localizations** ['loc lung field', 'loc right', 'loc hemithorax', 'loc hilar', 'loc cardiac', 'loc middle lung field']

**LabelsLocalizationsBySentence** [['pulmonary mass', 'loc right', 'loc lung field', 'loc middle lung field', 'loc hemithorax'], ['pacemaker', 'dual chamber device'], ['cardiomegaly', 'loc cardiac'], ['cardiomegaly', 'loc cardiac'], ['vascular hilar enlargement', 'loc hilar'], ['sternotomy', 'suture material']]

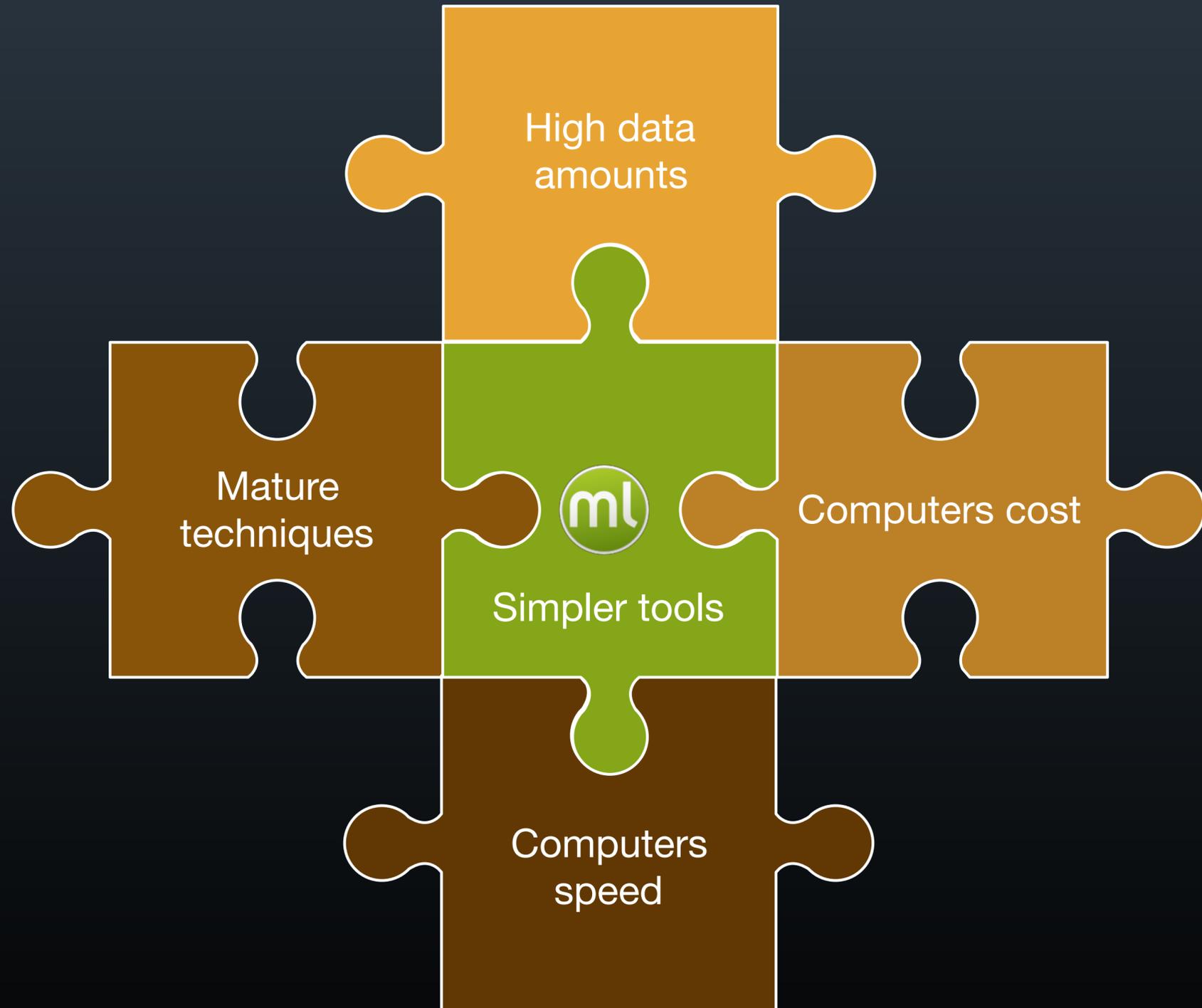
**labelCUIS** ['C0149726' 'C0030163' 'C0018800' 'C0185792' 'C2732817' 'C4305366']

**LocalizationsCUIS** ['C0225759' 'C0444532' 'C0934569' 'C0205150' 'C1522601' 'C0929434']



Las herramientas de *Machine Learning* son difíciles de utilizar, barrera técnica de entrada





# Agenda

---

1	¿Qué es el <i>Machine Learning</i> ?
2	El Machine Learning en la medicina actual
<b>3</b>	<b>Introducción a BigML</b>
4	Casos de uso



BigML es una  
plataforma online  
creada con el objetivo  
de democratizar el ML

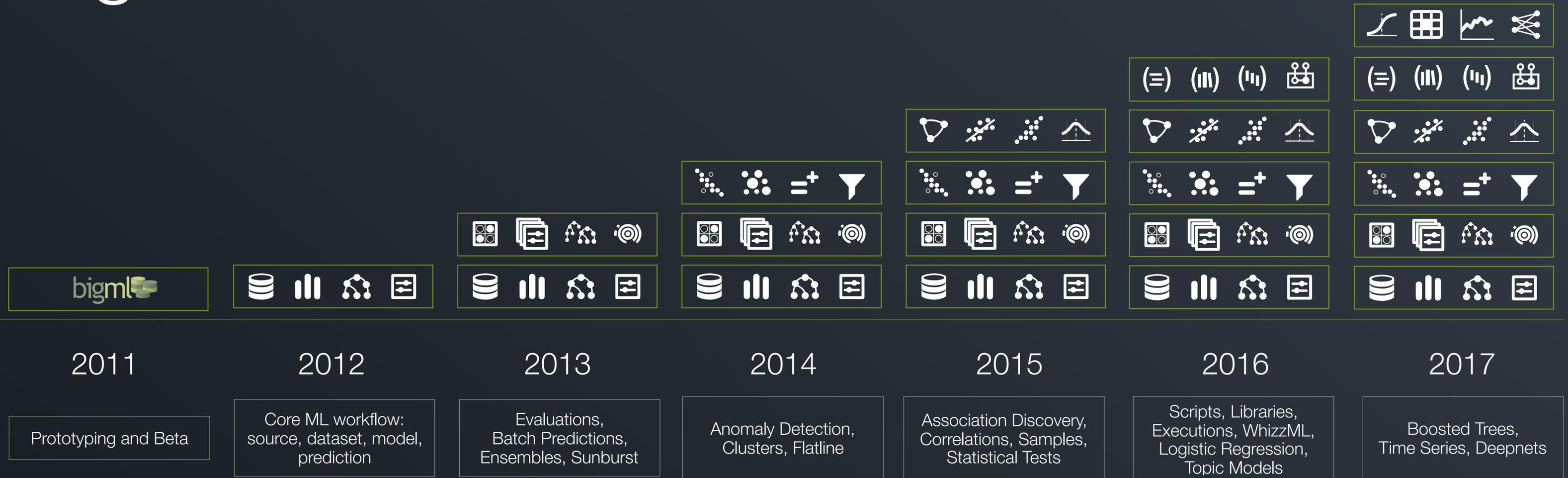
# Fundada en Corvallis, Oregon, US, en 2011





Con sede europea  
en Valencia desde  
2015

# BigML, Inc



6 patentes otorgadas y 12 patentes solicitadas

**Pionera en Machine Learning as a Service desde 2011**



# 85,000+

## REGISTERED CUSTOMERS



COVINGTON



juriblox

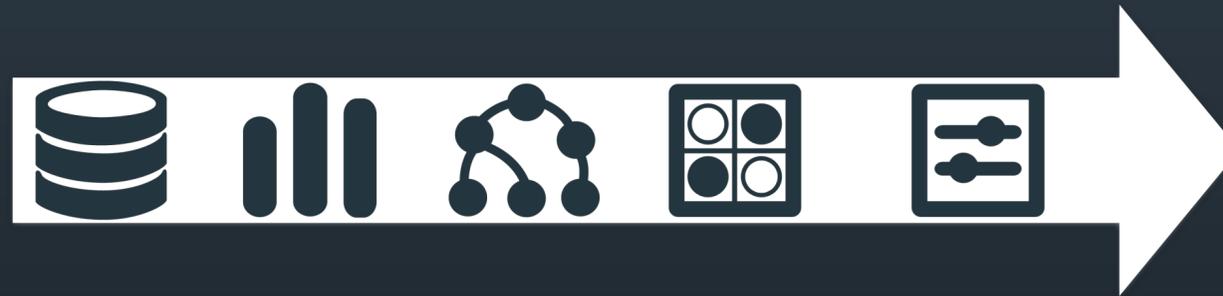


realmatch



SEAGATE



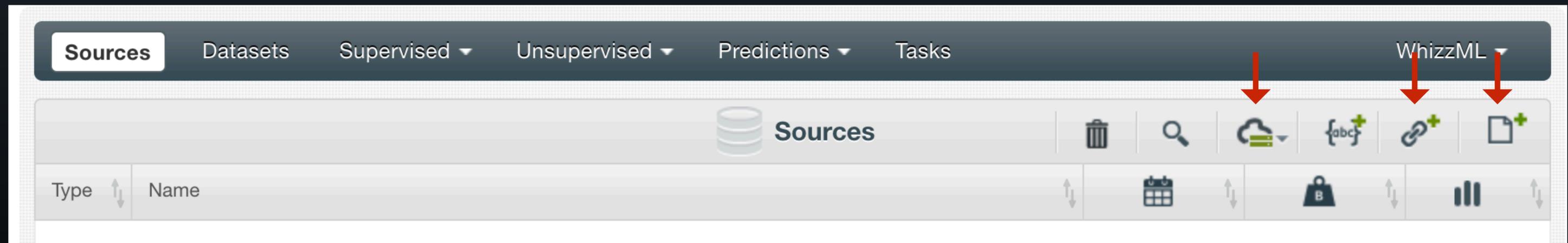


1. Crea una **f fuente** o *source*
2. Entiende tu **tabla de datos**
3. Entrena un **modelo**
4. **Evalúa** el modelo
5. Obtén una **predicción**

# 1. Crea una fuente o *source*

## Sube tus datos a BigML

- Drag & drop → CSV o Excel
- Google Drive, Google Storage, Dropbox o Azure
- URL remota



# 1. Crea una fuente o *source*

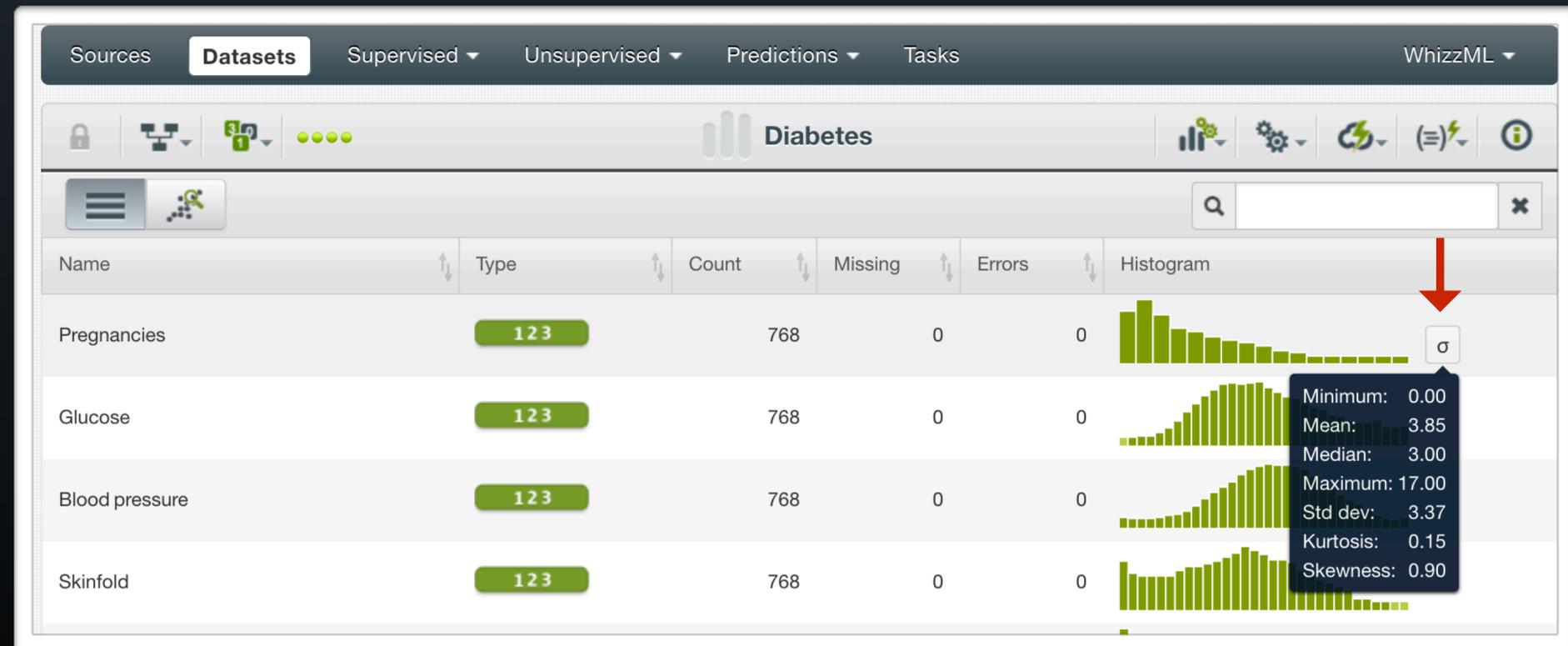
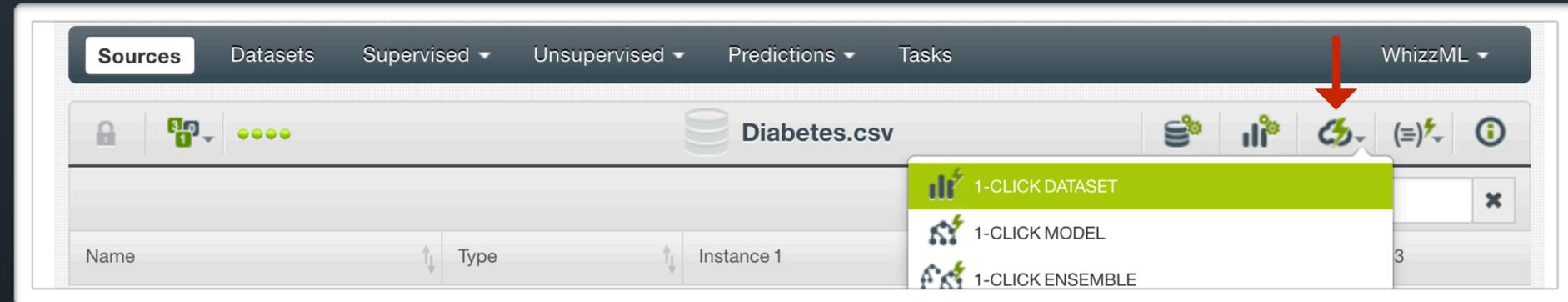
## Análisis automático de los datos

- Detección de **tipo de variable**
- Análisis de texto (NLP)
- Detección de fecha

Name	Type	Instance 1
Pregnancies	123	6
Glucose	123	148
Blood pressure	123	72
Skinfold	123	35
Insulin	123	0
BMI	123	33.6
Diabetes pedigree	123	0.627
Age	123	50
Diabetes	ABC	true

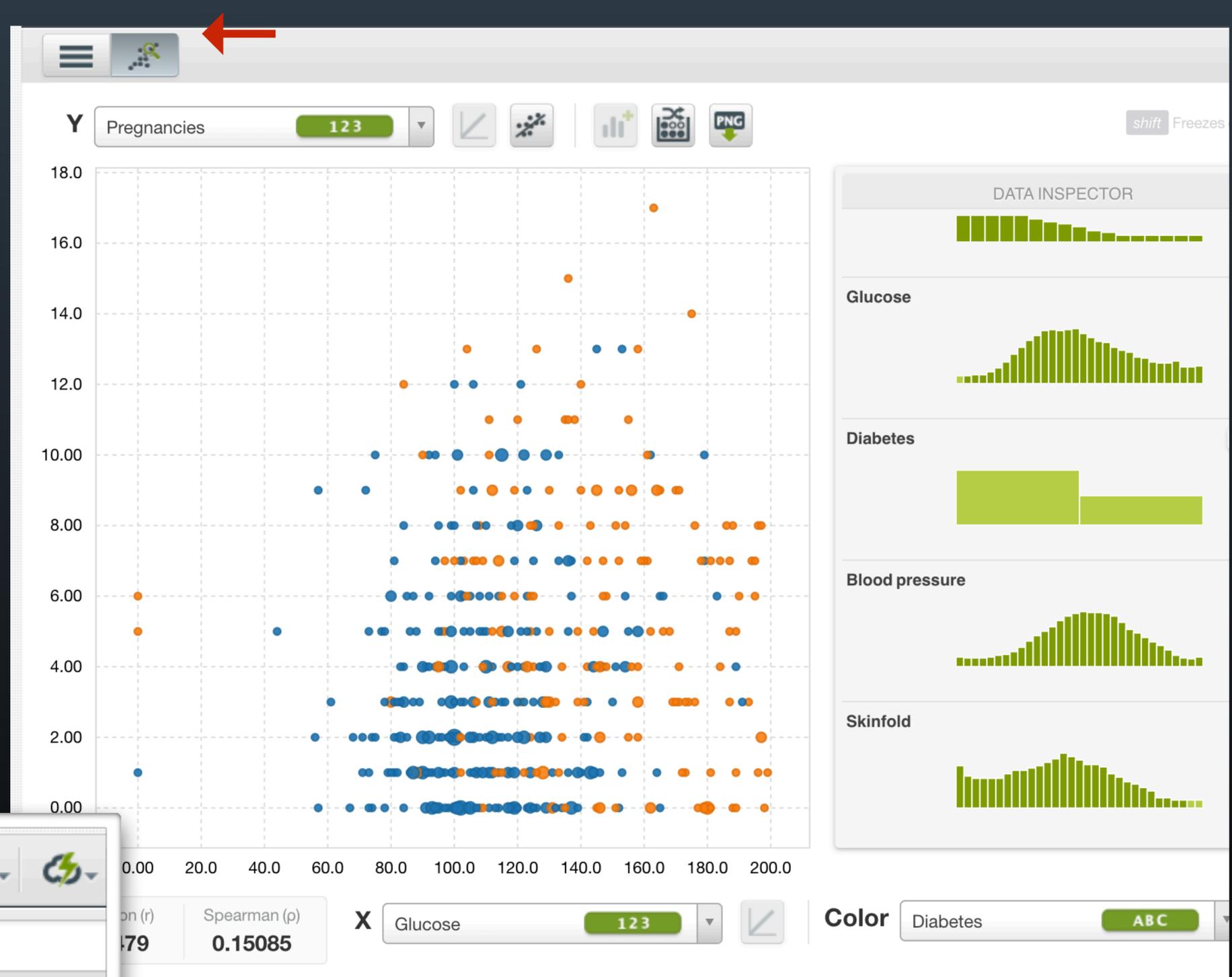
## 2. Entiende tu tabla de datos

- 1-Click dataset
- **Distribución** de parámetros
- **Métricas** estadísticas de cada variable



## 2. Entiende tu tabla de datos

- Visualización *scatterplot*:
- Separación datos (1-Click):
  - 80% → entrenamiento
  - 20% → evaluación



Diabetes

SAMPLE AND FILTER DATASET

TRAINING | TEST SPLIT

SAMPLE

FILTER

TRANSFORM DATASET

ADD FIELDS

AGGREGATE INSTANCES

JOIN DATASETS

## 2. Prepara una tabla de datos

- Transformación de datos: filtrado de datos y transformaciones de tabla

The screenshot displays the WhizzML interface for a dataset named "Diabetes". The top navigation bar includes "Datasets", "Supervised", "Unsupervised", "Predictions", "Tasks", and "WhizzML". Below the navigation bar, the dataset name "Diabetes" is prominently displayed. A context menu is open, showing two main categories of actions: "SAMPLE AND FILTER DATASET" and "TRANSFORM DATASET".

SAMPLE AND FILTER DATASET	TRANSFORM DATASET
TRAINING   TEST SPLIT	ADD FIELDS
SAMPLE	AGGREGATE INSTANCES
FILTER	JOIN DATASETS
REMOVE DUPLICATES	MERGE DATASETS
	ORDER INSTANCES

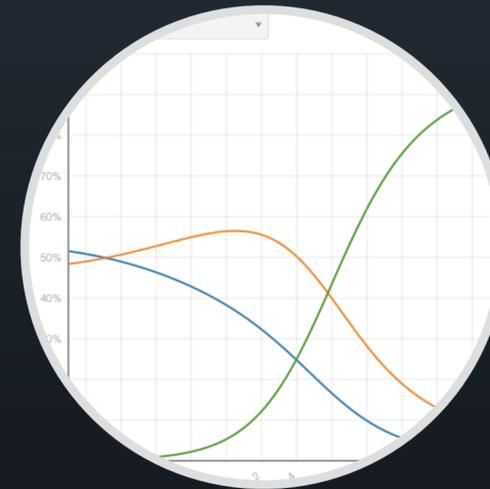
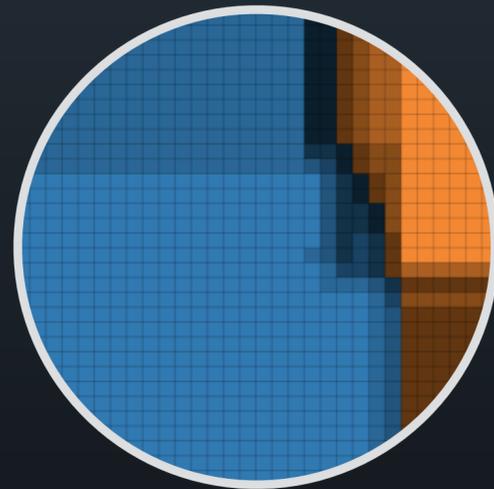
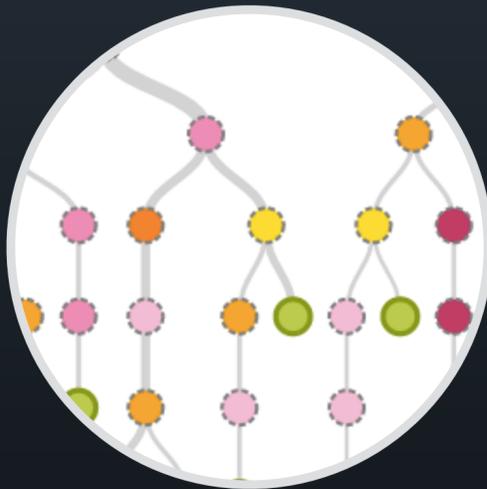
The background interface shows a table with a "Type" column and three rows, each containing the value "123". Below the table, there are numerical values: "768", "0", and "0". To the right, a histogram is visible, showing a distribution of data points.

# 3. Entrena un modelo

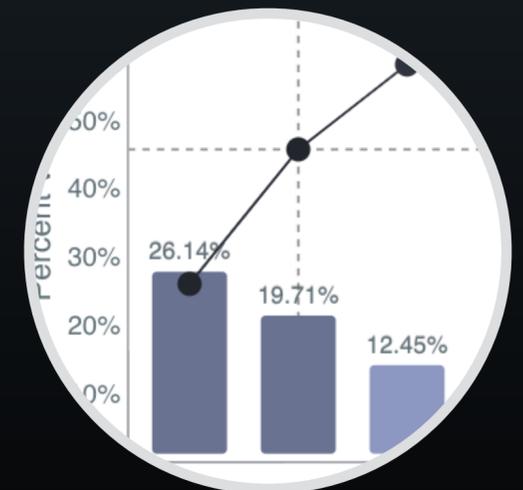
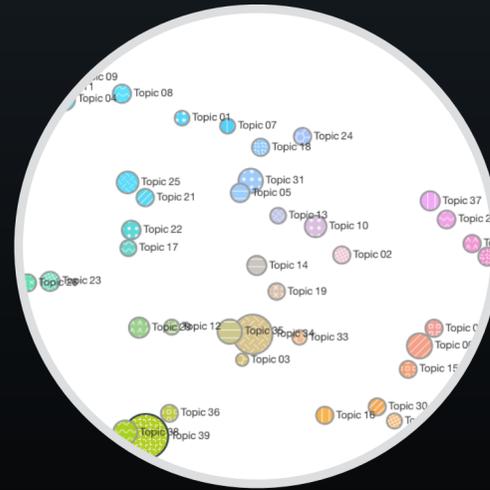
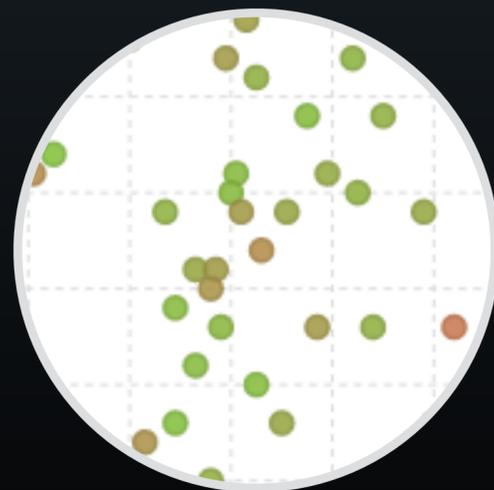
CLASSIFICATION AND REGRESSION

TIME SERIES

SUPERVISED



UNSUPERVISED



CLUSTER ANALYSIS

ANOMALY DETECTION

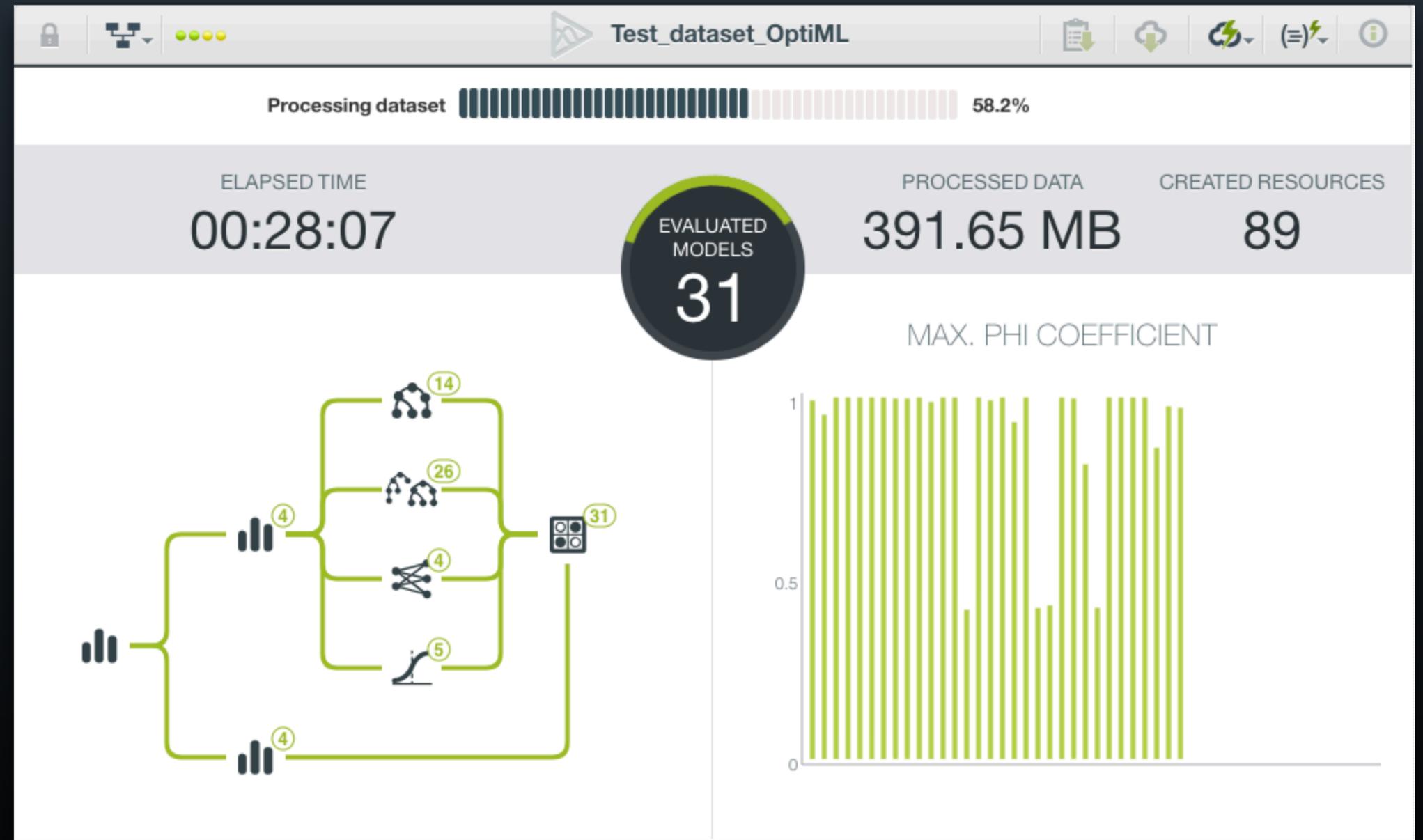
ASSOCIATION DISCOVERY

TOPIC MODELING

PCA

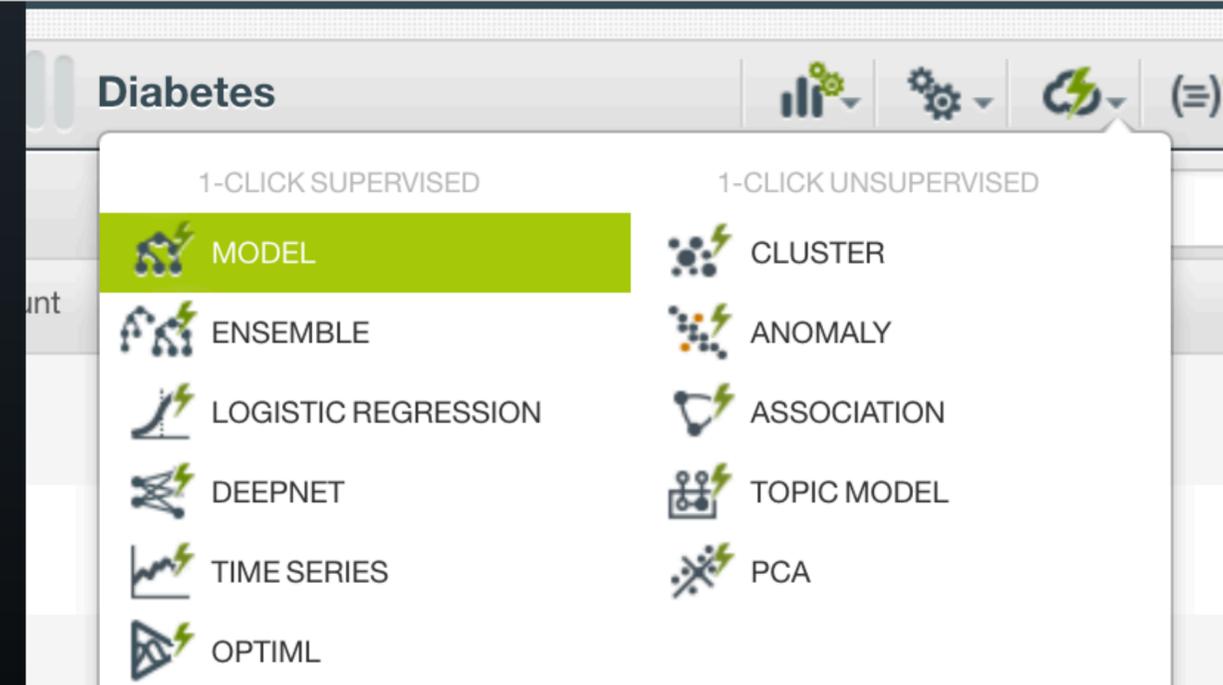
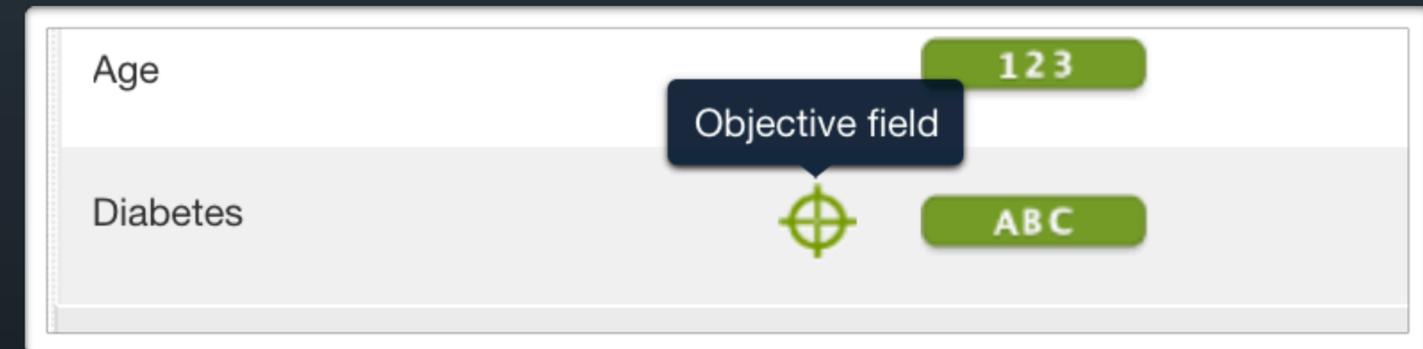
### 3. Entrena un modelo → OptiML

- Machine learning sobre machine learning
- **Optimización automática** considerando todas las combinaciones posibles de algoritmos y parámetros



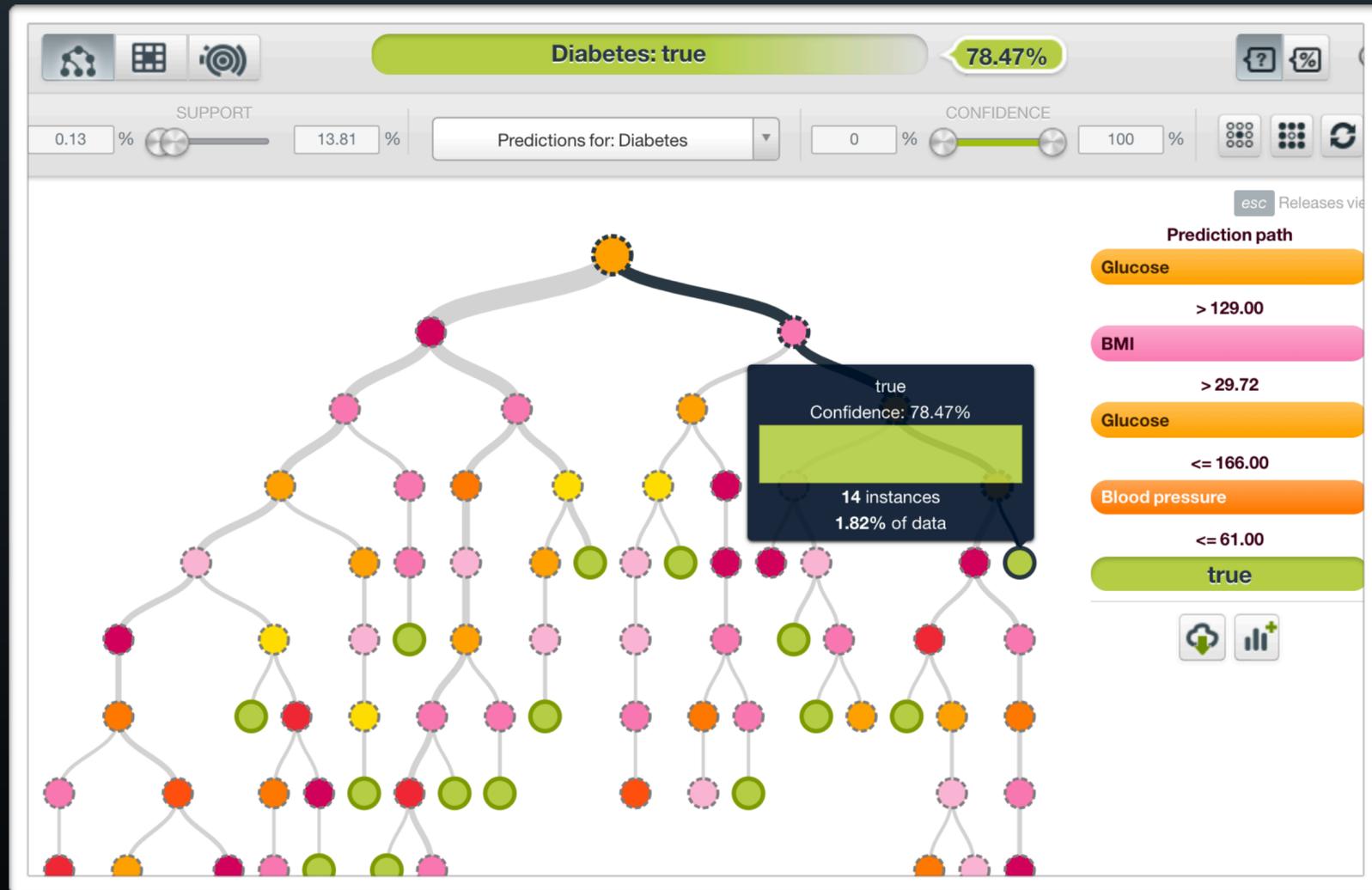
### 3. Entrena un modelo

- Selecciona la **variable objetivo**
- 2 opciones para crear el modelo:
  - Utiliza la opción **1-click** (automatizada y optimizada)
  - Configurar parámetros

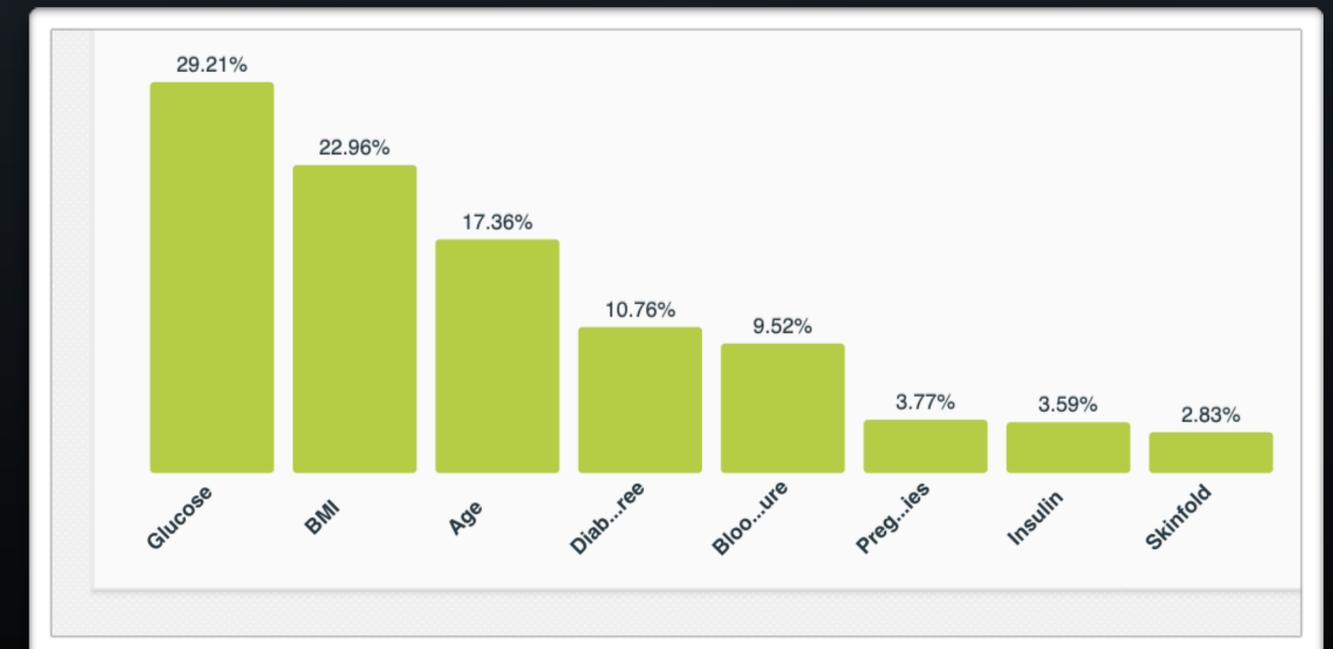
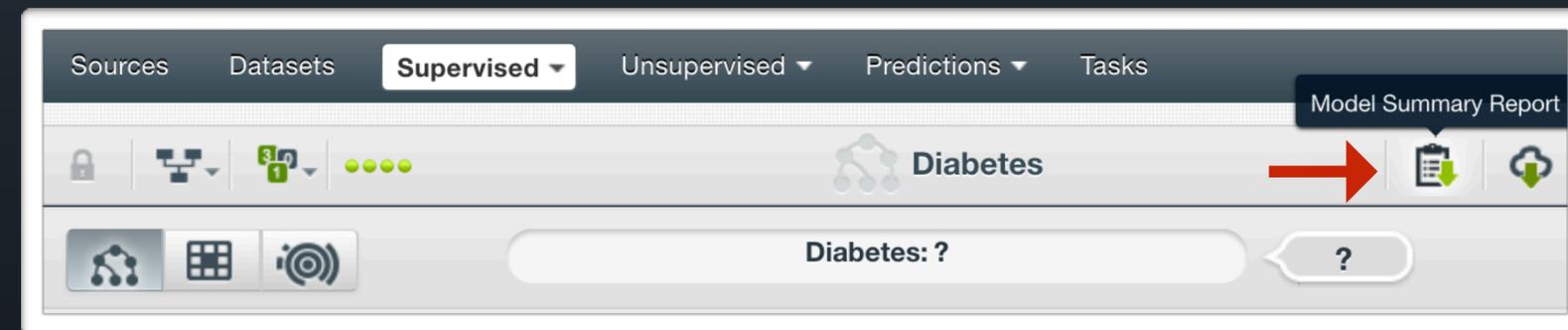


# 3. Entrena un modelo

- Visualización del modelo



- Visualiza qué campos tienen más impacto en el modelo



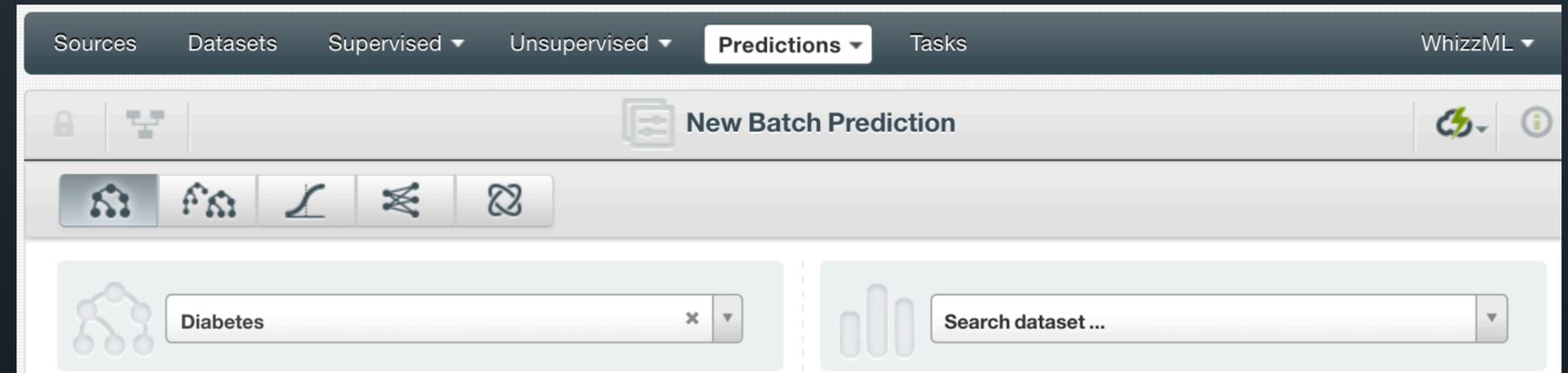
## 4. Evalúa el modelo

- Visualiza la **matriz de confusión**
- **Métricas de evaluación**
  - Exactitud
  - Precisión
  - Recall
  - F-measure
  - Coeficiente-Phi



## 5. Obtén una predicción

- **Predicciones individuales**
- Predicciones en serie de varias instancias o *batch predictions*



# Agenda

---

1	¿Qué es el <i>Machine Learning</i> ?
2	El Machine Learning en la medicina actual
3	Introducción a BigML
4	Casos de uso

# #1 Acra & BigML



## Problema



**Reingresos en la UCI** en menos de 72h de dar el alta

- tiempo medio estancia: 21 días
- pacientes reingresados: **30 días +**  
**↑ riesgo de muerte**

# #1 Acra & BigML

Problema



Datos



**Reingresos en la UCI** en menos de 72h de dar el alta

- tiempo medio estancia: 21 días
- pacientes reingresados: **30 días +**  
**↑ riesgo de muerte**

- **Constantes** monitorizadas
- **Medicación** administrada
- **Pruebas medicas**
- **Historia clínica** del paciente



# #1 Acra & BigML

## Problema



Reingresos en la UCI en menos de 72h de dar el alta

- tiempo medio estancia: 21 días
- pacientes reingresados: **30 días +**  
↑ riesgo de muerte



## Datos

- **Constantes** monitorizadas
- **Medicación** administrada
- **Pruebas medicas**
- **Historia clínica** del paciente



## Objetivo

Predecir la **probabilidad de reingreso**



¿Alargar la estancia?

***Ayuda en la toma de decisiones,  
herramienta de prevención***

#2

# Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191921#sec012>

Kalluri Thishya, Kiran Kumar Vattam, Shaik Mohammad Naushad, Shree Bhushan Raju, Vijay Kumar Kutala 

Published: April 5, 2018 • <https://doi.org/10.1371/journal.pone.0191921>

Problema



1. ↑ **variabilidad farmacocinética** del tacrolimus
2. **Diabetes post-transplante**

#2

# Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191921#sec012>

Kalluri Thishya, Kiran Kumar Vattam, Shaik Mohammad Naushad, Shree Bhushan Raju, Vijay Kumar Kutala 

Published: April 5, 2018 • <https://doi.org/10.1371/journal.pone.0191921>

Problema



1. ↑ **variabilidad farmacocinética** del tacrolimus

2. **Diabetes post-transplante**



Datos



- **Historial clínico**
- **Polimorfismos** asociados a actividad enzimática: CYP3A5 y ABCB1

Objetivos

1. Predecir dosis terapéutica

2. Predecir riesgo de diabetes



#2

# Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191921#sec012>

Kalluri Thishya, Kiran Kumar Vattam, Shaik Mohammad Naushad, Shree Bhushan Raju, Vijay Kumar Kutala 

Published: April 5, 2018 • <https://doi.org/10.1371/journal.pone.0191921>

Problema



- 1. ↑ variabilidad farmacocinética del tacrolimus
- 2. Diabetes post-transplante



Datos

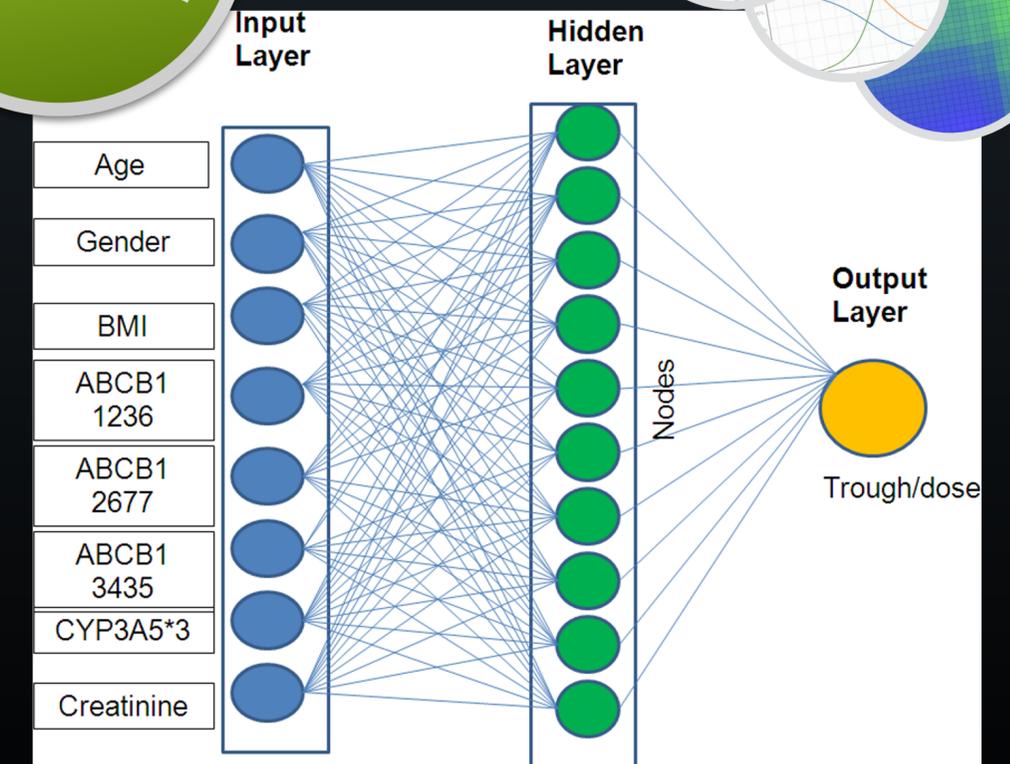
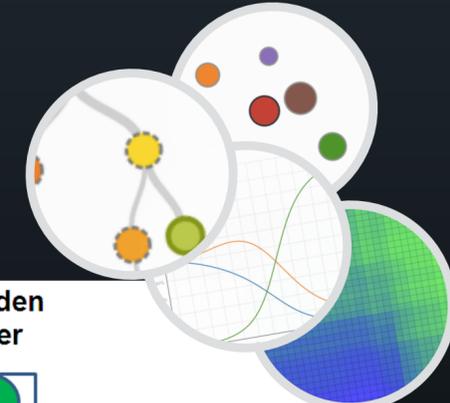


- Historial clínico
- Polimorfismos asociados a actividad enzimática: CYP3A5 y ABCB1

Objetivos

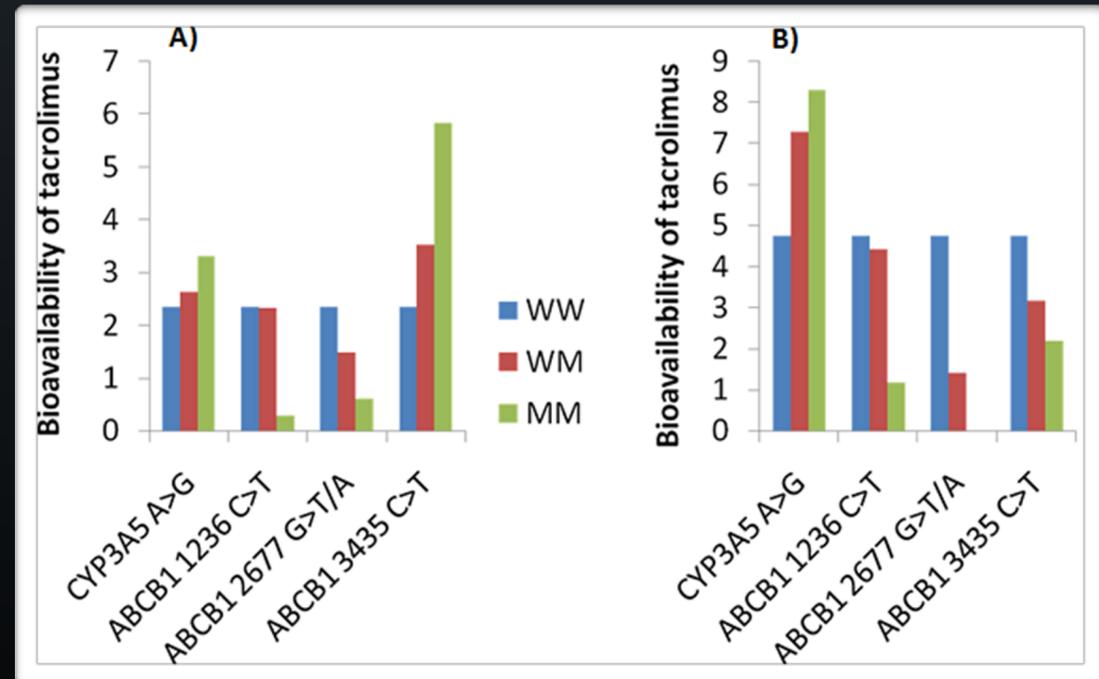
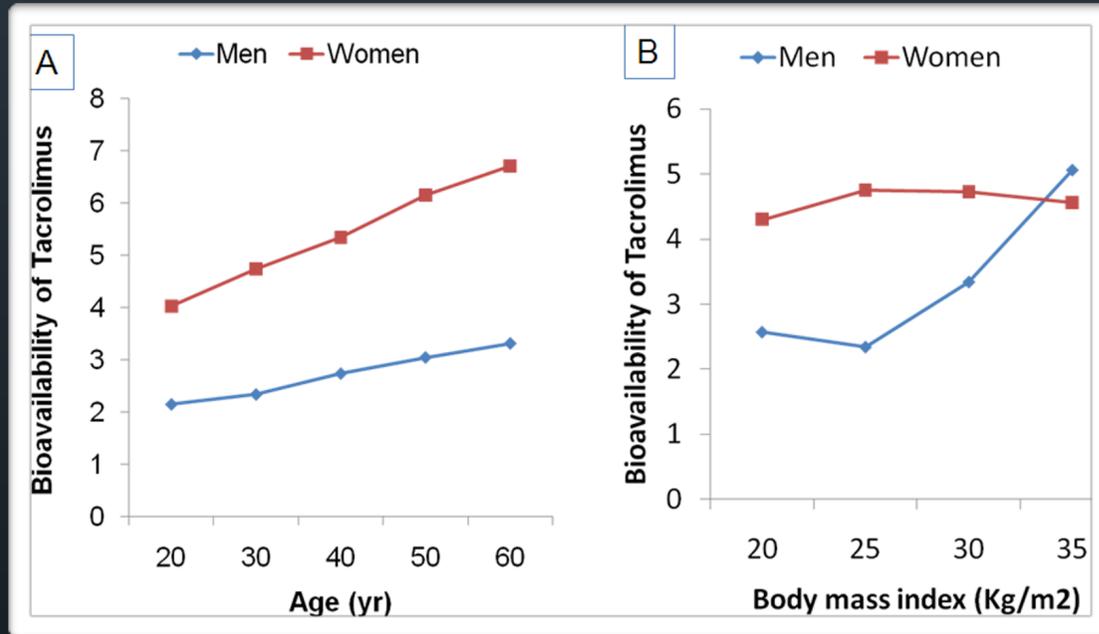
- 1. Predecir dosis terapéutica
- 2. Predecir riesgo de diabetes

Modelo





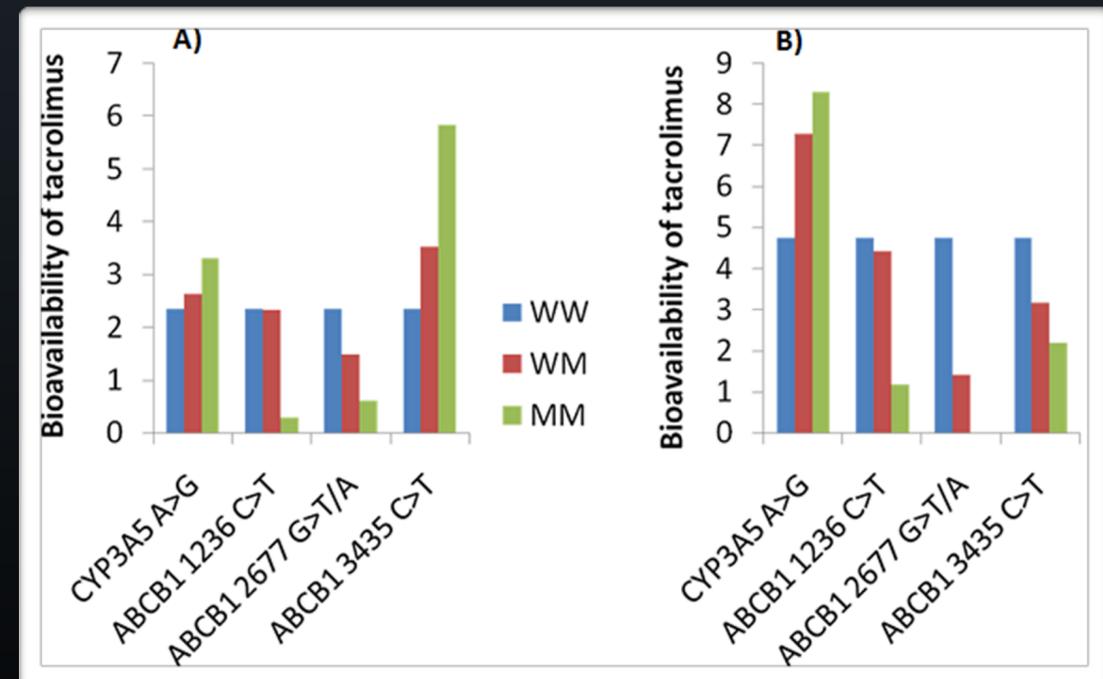
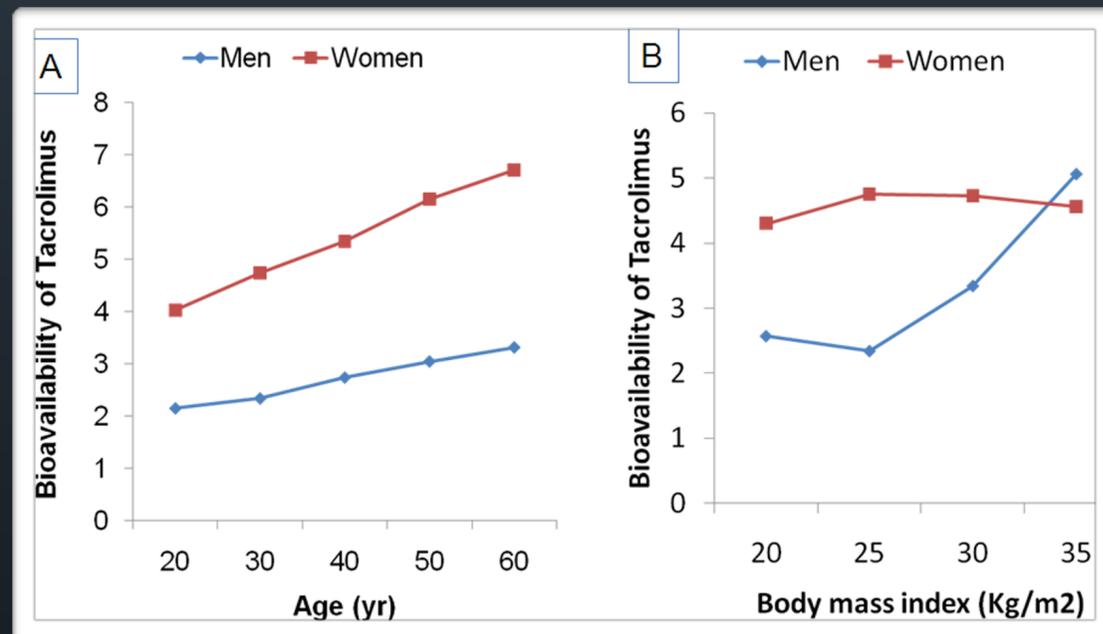
# Resultado 1



- Dimorfismo sexual: edad, IMC, factores genéticos
- Interacción genética entre CYP3A5\*3 con 2 variantes de ABCB1



## Resultado 1



- Dimorfismo sexual: edad, IMC, factores genéticos
- Interacción genética entre CYP3A5\*3 con 2 variantes de ABCB1



## Resultado 2

Variable	Odds ratio	95% CI	P value
Age	1.02	0.95–1.09	0.59
Gender	1.81	0.33–10.07	0.50
Body mass index	1.06	0.86–1.30	0.61
CYP3A5*3	1.73	0.70–4.28	0.23
ABCB1 1236 C>T	2.59	0.79–8.50	0.12
<b>ABCB1 2677 G&gt;T/A</b>	<b>4.83</b>	<b>1.22–19.03</b>	<b>0.02</b>
ABCB1 3435 C>T	0.86	0.34–2.19	0.76

- Revela **asociación** entre ABCB1 2677 y el riesgo de padecer diabetes

***Herramienta para implementar la medicina de precisión***

#3

## Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns

 **Adrián M. Orgueira**<sup>1, 2\*</sup>,  **Beatriz A. Rodriguez**<sup>1, 2</sup>, **Natalia A. Vence**<sup>1</sup>, **Ángeles B. López**<sup>1, 2</sup>,   
**José Ángel D. Arias**<sup>1</sup>, **Nicolás D. Varela**<sup>1</sup>, **Marta Sonia G. Pérez**<sup>1</sup>, **Manuel Mateo P. Encinas**<sup>1, 2</sup> and **José Luis B. López**<sup>1, 2</sup>

<https://www.frontiersin.org/articles/10.3389/fonc.2019.00079/abstract>

<sup>1</sup>Complejo Hospitalario Universitario de Santiago, Spain

### Problema



Comienzo del **tto** en la **Leucemia Linfocítica Crónica (LLC)**

### Objetivo

Identificar si la **progresión de LLC** esta ligada a un **patrón de expresión** concreto

# #3

## Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns

 **Adrián M. Orgueira**<sup>1, 2\*</sup>,  **Beatriz A. Rodriguez**<sup>1, 2</sup>, **Natalia A. Vence**<sup>1</sup>, **Ángeles B. López**<sup>1, 2</sup>,  **José Ángel D. Arias**<sup>1</sup>, **Nicolás D. Varela**<sup>1</sup>, **Marta Sonia G. Pérez**<sup>1</sup>, **Manuel Mateo P. Encinas**<sup>1, 2</sup> and **José Luis B. López**<sup>1, 2</sup>

<https://www.frontiersin.org/articles/10.3389/fonc.2019.00079/abstract>

<sup>1</sup>Complejo Hospitalario Universitario de Santiago, Spain

### Problema

Comienzo del **tto** en la **Leucemia Linfocítica Crónica (LLC)**

### Objetivo

Identificar si la **progresión de LLC** esta ligada a un **patrón de expresión** concreto

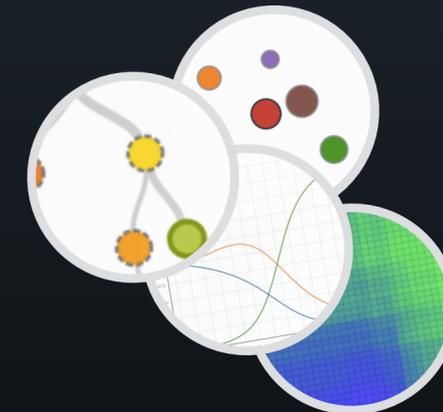


### Datos

- Expresión de RNA (290 genes)
- Estado de la mutación en la cadena pesada de la Ig



### Modelo



#3

# Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns

 **Adrián M. Orgueira**<sup>1, 2\*</sup>,  **Beatriz A. Rodriguez**<sup>1, 2</sup>, **Natalia A. Vence**<sup>1</sup>, **Ángeles B. López**<sup>1, 2</sup>,   
**José Ángel D. Arias**<sup>1</sup>, **Nicolás D. Varela**<sup>1</sup>, **Marta Sonia G. Pérez**<sup>1</sup>, **Manuel Mateo P. Encinas**<sup>1, 2</sup> and **José Luis B. López**<sup>1, 2</sup>

<https://www.frontiersin.org/articles/10.3389/fonc.2019.00079/abstract>

<sup>1</sup>Complejo Hospitalario Universitario de Santiago, Spain

## Problema

Comienzo del tto en la Leucemia Linfocítica Crónica (LLC)

## Objetivo

Identificar si la progresión de LLC esta ligada a un patrón de expresión concreto

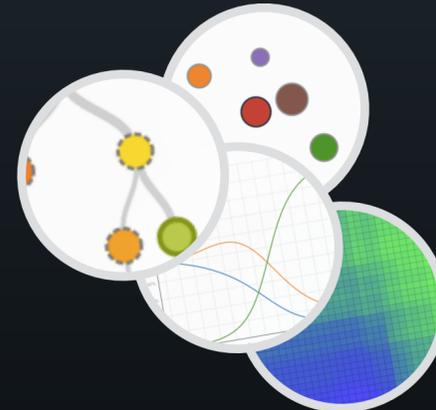


## Datos

- Expresión de RNA (290 genes)
- Estado de la mutación en la cadena pesada de la Ig



## Modelo



## Resultado

**Estratificación de pacientes**  
4 grupos, diferentes comienzos de tratamiento

***Progreso de la enfermedad en diferentes grupos de pacientes***

# #4 BigML - Pfizer

Datos



Modelo



Predicción



- Múltiples aplicaciones
- ¿Se aprobará el fármaco?
- ¿Se rechazará el fármaco?



Fase Pre-clínica



Fase Clínica

Segmentación y predicción de la aceptación del fármaco

Optimización I+D

big ml<sup>®</sup>

# Enlaces de utilidad



Qué es **bigml**<sup>®</sup>

<https://bigml.com/about>



Entrenamiento

<https://bigml.com/pricing#training>

Videos de Educación

<https://bigml.com/education/videos>



Certificaciones

<https://bigml.com/certifications>



Documentación

<https://bigml.com/documentation/dashboard/>



Versiones Privadas

<https://bigml.com/private-deployments>



Tutoriales

<https://bigml.com/tutorials/>



Clientes

<https://bigml.com/customers>

Introducción a ML

<https://bigml.com/ml101/>

101



Precios

<https://bigml.com/pricing>

RESEARCH ARTICLE

## Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation

Kalluri Thishya, Kiran Kumar Vattam, Shaik Mohammad Naushad, Shree Bhushan Raju, Vijay Kumar Kutala 

Published: April 5, 2018 • <https://doi.org/10.1371/journal.pone.0191921>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0191921>

RESEARCH ARTICLE

## Designing a more efficient, effective and safe Medical Emergency Team (MET) service using data analysis

Christoph Bergmeir , Irma Bilgrami, Christopher Bain, Geoffrey I. Webb, Judit Orosz, David Pilcher

Published: December 27, 2017 • <https://doi.org/10.1371/journal.pone.0188688>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188688>

RESEARCH ARTICLE

## ECG-derived spatial QRS-T angle is associated with ICD implantation, mortality and heart failure admissions in patients with LV systolic dysfunction

Sarah Gleeson , Yi-Wen Liao , Clementina Dugo , Andrew Cave , Lifeng Zhou , Zina Ayar , Jonathan Christiansen , Tony Scott , Liane Dawson , Andrew Gavin , Todd T. Schlegel , Patrick Gladding 

Published: March 30, 2017 • <https://doi.org/10.1371/journal.pone.0171069>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0171069>

Front. Oncol. | doi: 10.3389/fonc.2019.00079

## Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns

 Adrián M. Orgueira<sup>1, 2\*</sup>,  Beatriz A. Rodríguez<sup>1, 2</sup>, Natalia A. Vence<sup>1</sup>, Ángeles B. López<sup>1, 2</sup>,  José Ángel D. Arias<sup>1</sup>, Nicolás D. Varela<sup>1</sup>, Marta Sonia G. Pérez<sup>1</sup>, Manuel Mateo P. Encinas<sup>1, 2</sup> and José Luis B. López<sup>1, 2</sup>

<sup>1</sup>Complejo Hospitalario Universitario de Santiago, Spain

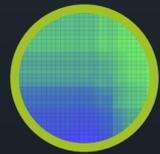
<sup>2</sup>University of Santiago de Compostela, Spain

<https://www.frontiersin.org/articles/10.3389/fonc.2019.00079/abstract>



## CLASSIFICATION

- ¿Tendrá este paciente que someterse a diálisis?



## REGRESSION

- ¿Cuántos días estará este paciente en el hospital?



## TIME SERIES FORECASTING

- Disponibilidad en lista de espera de los próximos meses



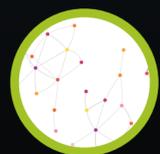
## CLUSTER ANALYSIS

- Estratificación de pacientes, grupos de pacientes que tienen características similares



## ANOMALY DETECTION

- ¿Está funcionando un aparato médico de forma inusual?



## ASSOCIATION DISCOVERY

- ¿Qué características están asociadas con una determinada enfermedad?