



**WOMEN IN DATA SCIENCE**  
@ STANFORD UNIVERSITY



**FACULTAD  
DE INGENIERIA**  
Universidad de Buenos Aires

# Mujeres en la Ciencia de Datos

# Sobre teras y semillas

## La ciencia de datos al servicio del agro

**Yanina Bellini Saibene**

[bellini.yanina@inta.gob.ar](mailto:bellini.yanina@inta.gob.ar)

@yabellini

# Mucho gusto...

Licenciada en Sistemas de Información

Magister en Explotación de Datos y  
Gestión del conocimiento

**Investigadora** en el Instituto Nacional de  
Tecnología Agropecuaria (21 años)  
Socia Activa de SADIO (11 años)  
Organizadora Capítulo **R-Ladies** Santa  
Rosa (2 años)



# AgTech AgroTICs



*Enmarca aplicaciones de  
**Ciencia de Datos** en el sector  
**Agropecuario** (y una serie de otras  
disciplinas relacionadas a lo digital-  
electrónico).*

# Data Mining Machine Learning



**Aprendizaje  
Supervisado**

**Aprendizaje  
NO  
Supervisado**

**Variables de  
entrada**



**Modelo**

**Algoritmo**

# Aprendizaje Supervisado



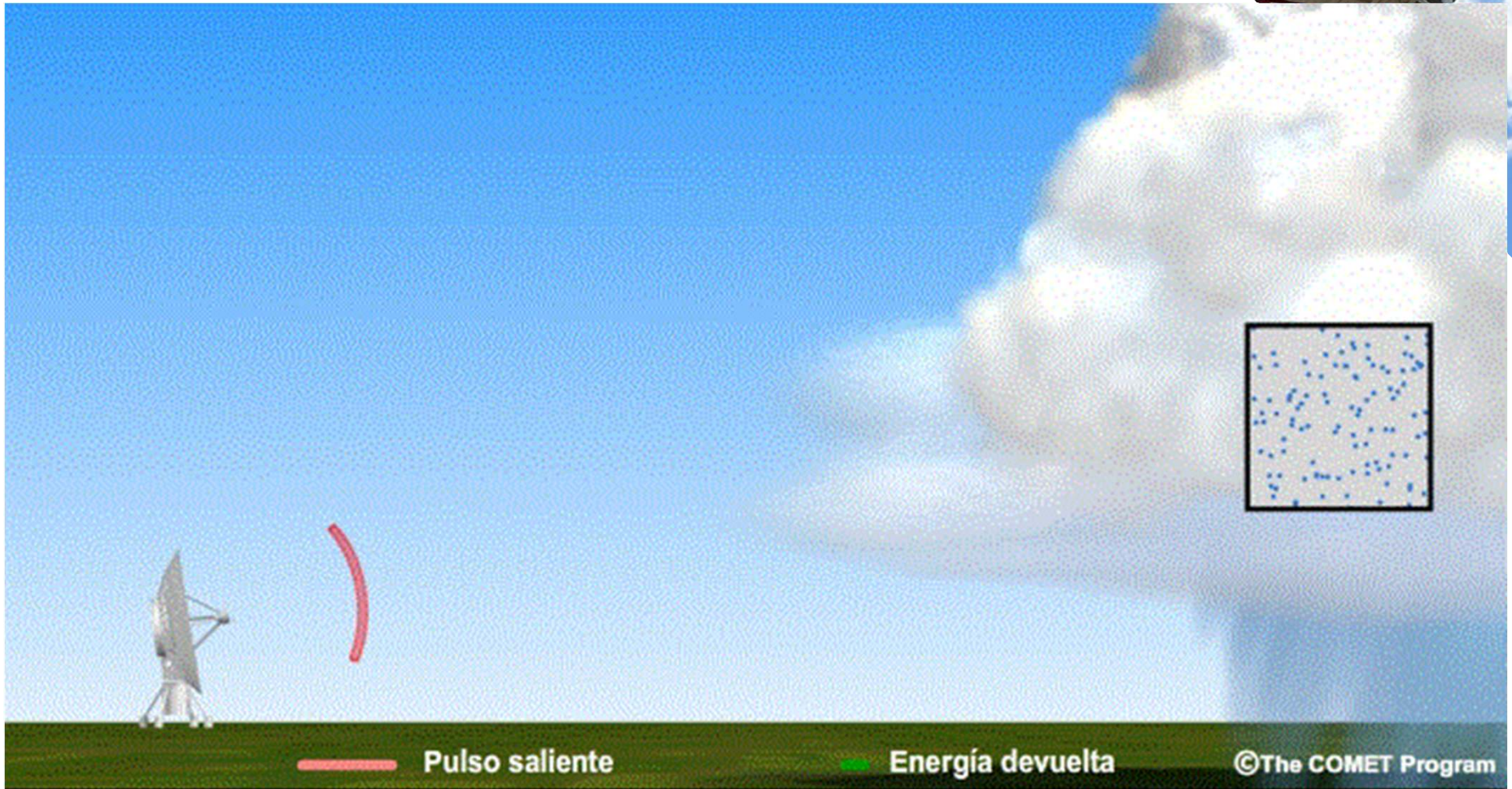
Estimación de ocurrencia  
de granizo en superficie,

mediante datos de radar meteorológico  
utilizando técnicas de **Data Mining**.



# Aprendizaje supervisado

*Modelo de granizo: resultados*



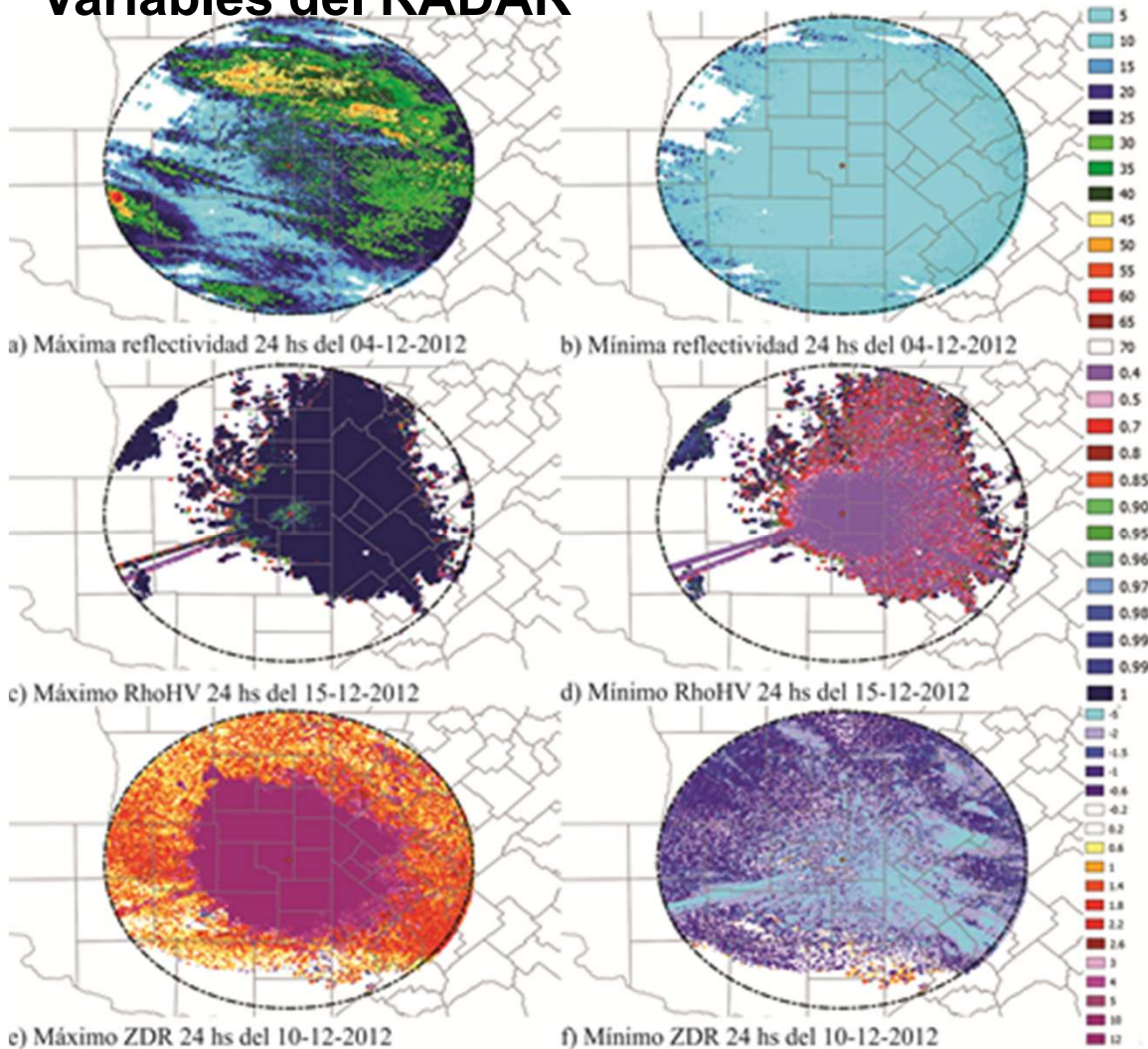


# Aprendizaje supervisado

## *Ocurrencia de granizo: variables*



### Variables del RADAR



### Variables del RADAR

5 variables

144 tomas al día

480x480 pixeles

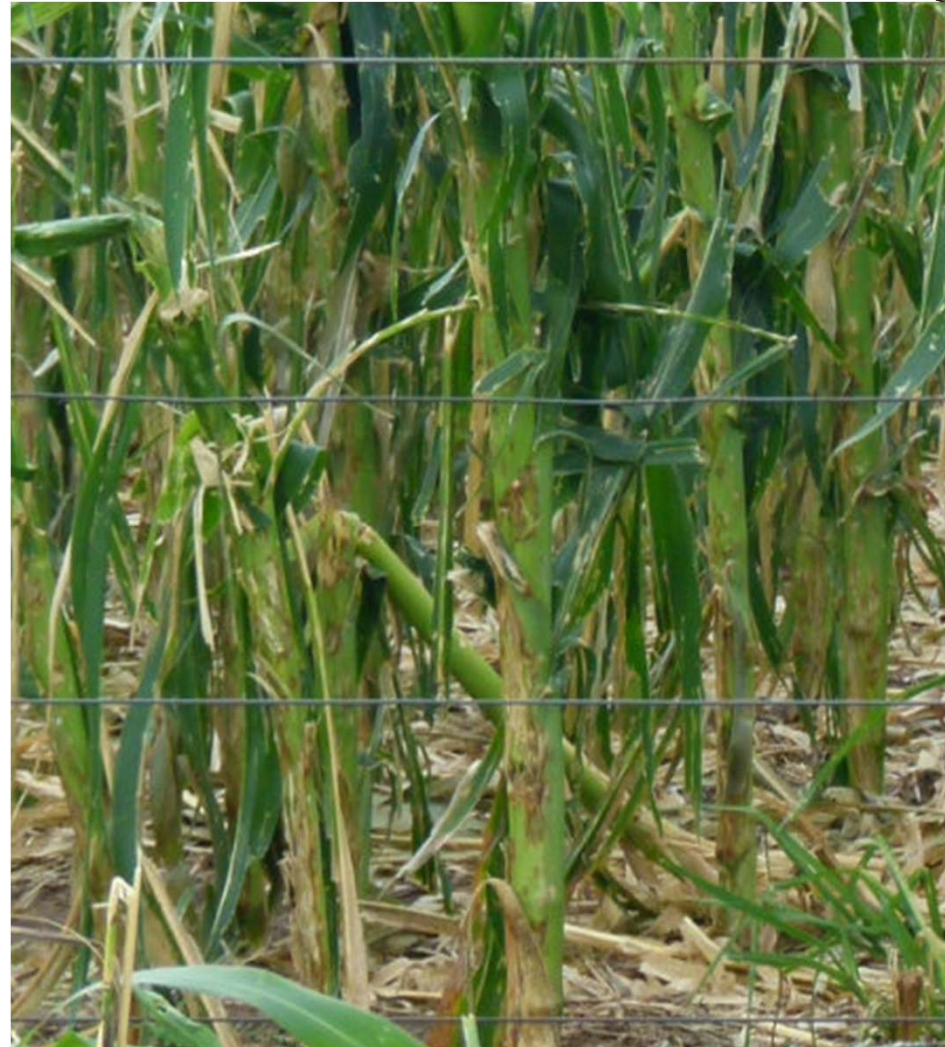
165.888.000 datos  
diarios



# Aprendizaje supervisado

## *Ocurrencia de granizo: variables*

Lotes con y sin daño por granizo





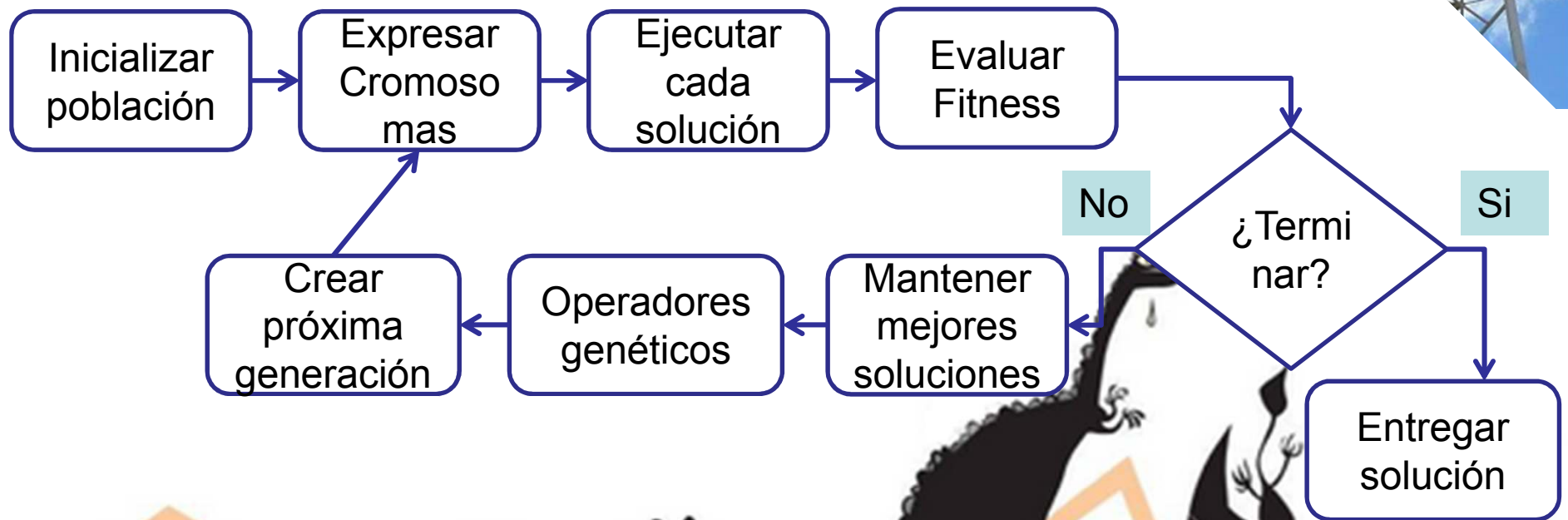
# Aprendizaje supervisado

## Ocurrencia de granizo

Metodología: Gene Expression Programing

Cromosomas

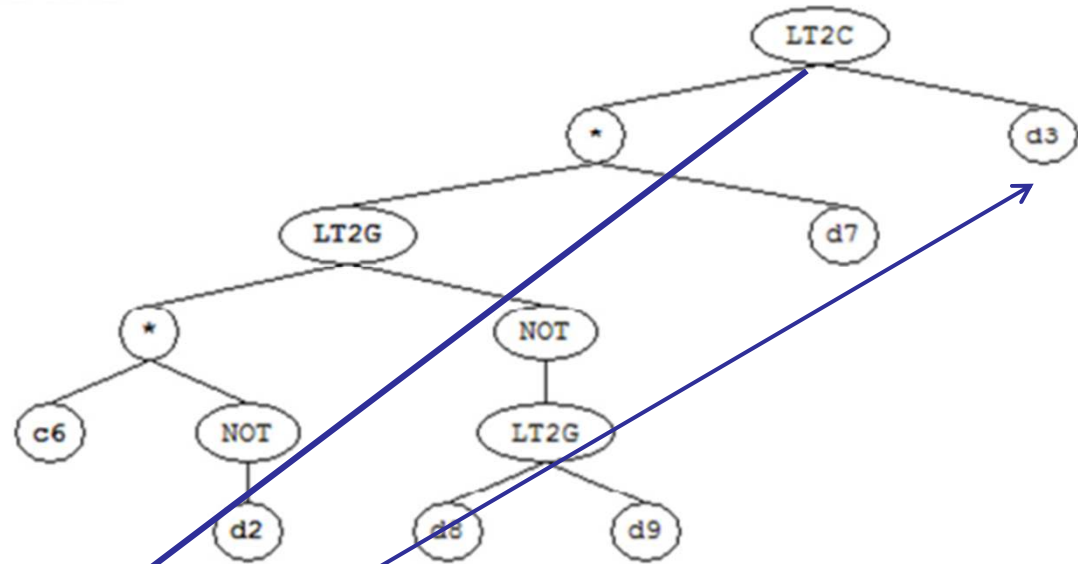
Árboles de Expresión



# Aprendizaje

## Modelo de granizo

Sub-ET 1



y = 0.0

```
y = (gepGOE2A((d[AvZDR]*gepGOE2C(G1(
y = y + (G2C3+(((gepGOE2G(gepLT2A(d
y = y + (((((gepAND1(G3C6,d[MxRho]))+((G3C0+G3C5)/2.0))/2.0)-gepGOE2G(gepAND1(d[MxZDR],d[MnZDR])).
y = y + (d[MnRho]-(gepGOE2E(G4C6,exp(d[MnRho])))*d[MnRho])))
```

SLOPE = 0.112107575724697  
INTERCEPT = -6.45746730193226

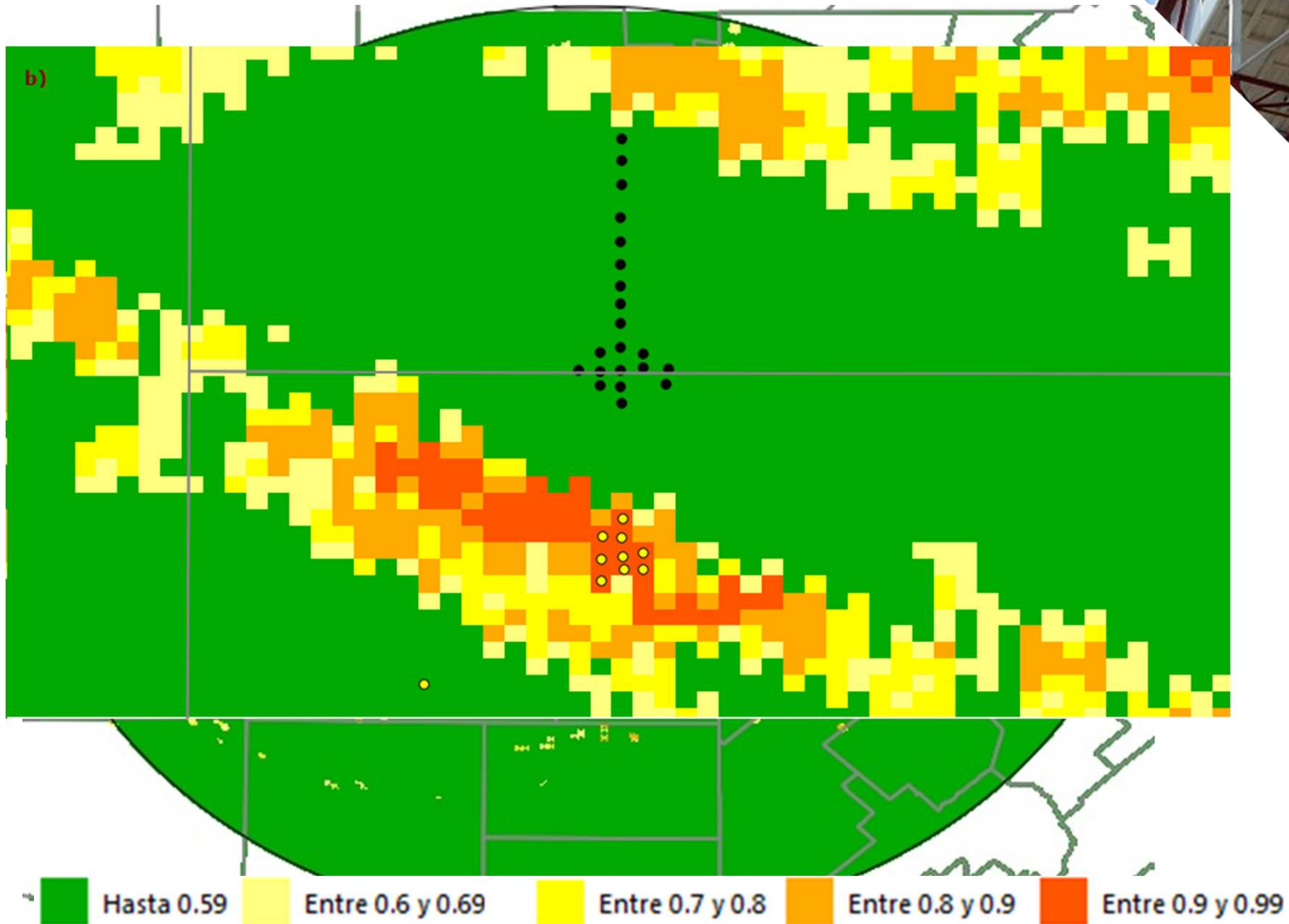
```
probabilityOne = 1.0 / (1.0 + exp(-(SLOPE * y + INTERCEPT)))
return probabilityOne
```

```
def gepLT2C(x, y):
    if (x < y):
        return (x+y)
    else:
        return (x-y)
```

**Figura 34.** Árbol de Expresión del modelo Pire (*Variables:* MxdBZ = d0, MndBZ = d1, TotdBZ = d2, AvdBZ = d3, MxRho = d4, TotRho = d6, AvRho = d7, MxZDR = d8, MnZDR = d9, TotZDR = d10, AvZDR = d11. *Constantes:* Sub-ET 1: C6 = 9.64666513565478, Sub-ET 2: C4 = -4.10822324594867, Sub-ET 4: C6 = 5.10116885891293).

# Aprendizaje supervisado

## *Modelo de granizo: resultados*





# Aprendizaje supervisado

## *Modelo de granizo y daño en cultivos*



### Herramientas



The Python ARM Radar Toolkit



### Equipo de trabajo



Romina  
MEZHER



Santiago  
BANCHERO



Laura  
BELMONTE



Juan  
CALDERA

<https://github.com/INTA-Radar>



# Aprendizaje NO Supervisado

**Clasificación de  
Sistemas Productivos  
Preponderantes  
utilizando técnicas de  
agrupamiento para la  
estimación de emisiones  
de Gases de Efecto  
Invernadero.**







# Aprendizaje NO supervisado

*Sistemas productivos preponderantes*



## **Metodología: Clustering**

¿por qué  
atributo de  
similitud se  
podrían  
agrupar estos  
“casos”?



K-means

K-medoids



# Aprendizaje NO supervisado

*Sistemas productivos preponderantes*

*Resultados*

13 grupos

Indice Jaccard:  $> 0.80$  😊

Silueta: entre 0.1 😞 y 0.57 😊

Cantidad de casos: entre 64 y 1207 🤔



*La pregunta del millón:  
¿cómo sabemos si esos grupos  
representan grupos reales?*

# **Aprendizaje NO supervisado**

*Sistemas productivos preponderantes*



***1. Miramos los grupos con expertos del negocio***

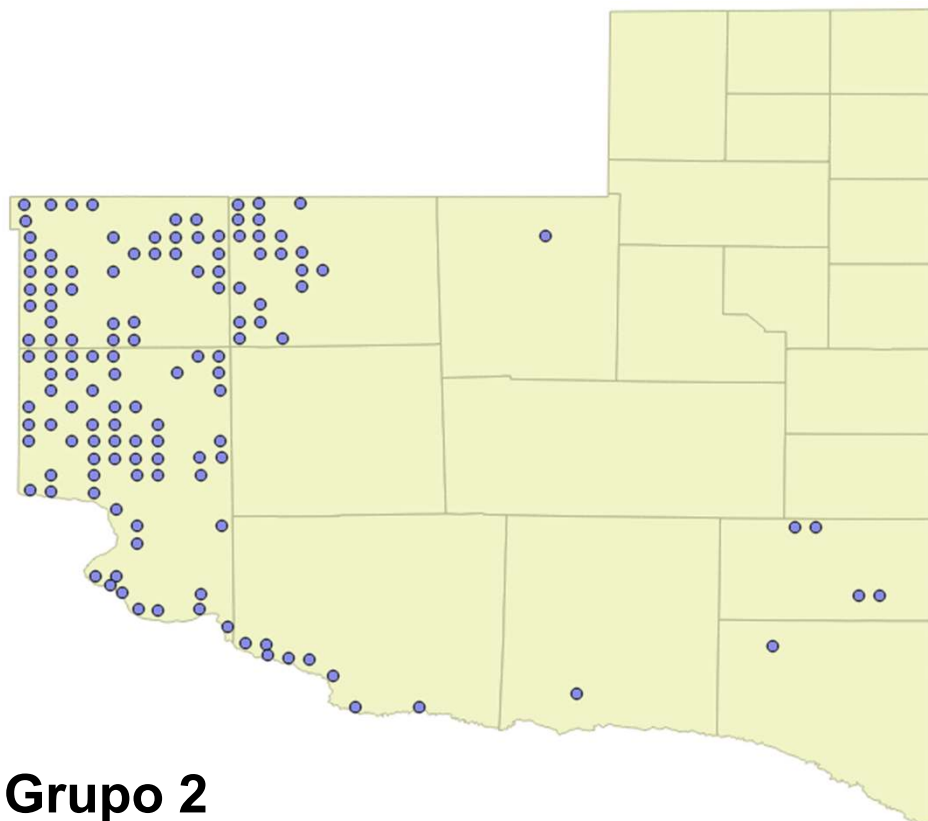
***2. Buscamos una coherencia de:***

- Tamaños***
- Actividades***
- Distribución espacial***

# Aprendizaje NO supervisado

*Sistemas productivos preponderantes*

*Algunos resultados:*



**Grupo 2**

**N 213**

**Silueta 0.31**

**Jaccard 0.91**

## Características

Superficie (ha)	2803
% Area Cultivada	1%
<b>% Monte natural</b>	<b>87%</b>
<b>Ganadería</b>	
% Bovinos	26%
% Ovinos	6%
% Porcinos	1%
<b>% Caprinos</b>	<b>56%</b>
% Equinos	11%

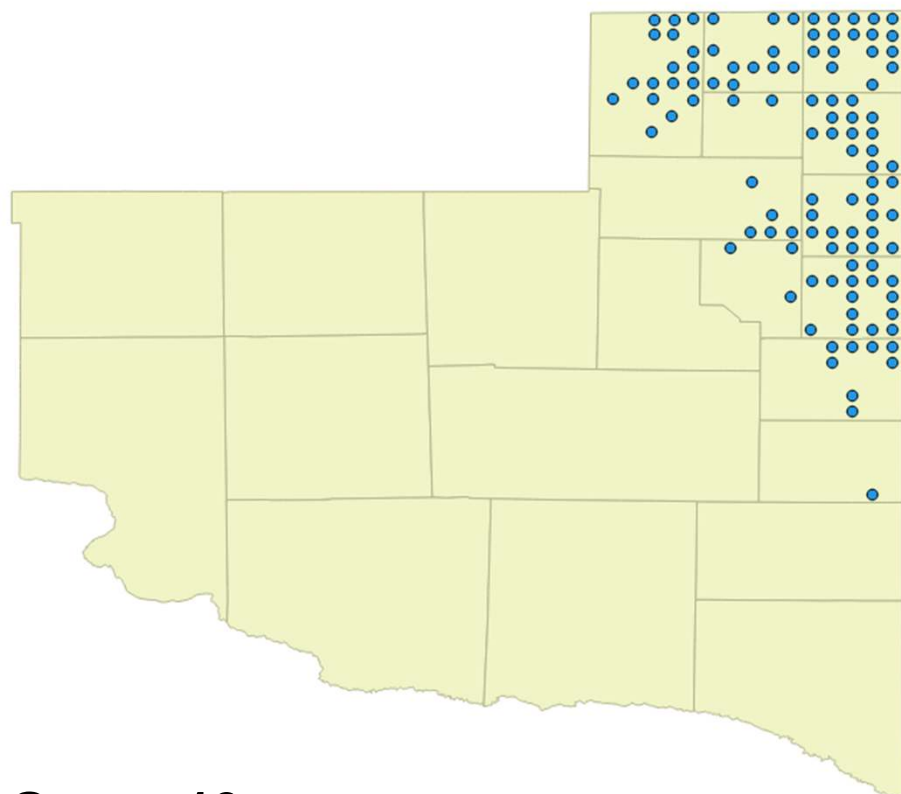


***Sistema cría caprina y bovina***

# Aprendizaje NO supervisado

*Sistemas productivos preponderantes*

*Algunos resultados:*



**Grupo 13**

**N 175**

**Silueta 0.52**

**Jaccard 0.92**

## Características

Superficie (ha) 403

## Agricultura

% Cosecha Fina 5%

**% Cosecha Gruesa 87%**

% Forrajeras Anuales 4%

% Forrajeras Perenne 1%

## Ganadería

% Bovinos 2%

% Otras producciones 4%



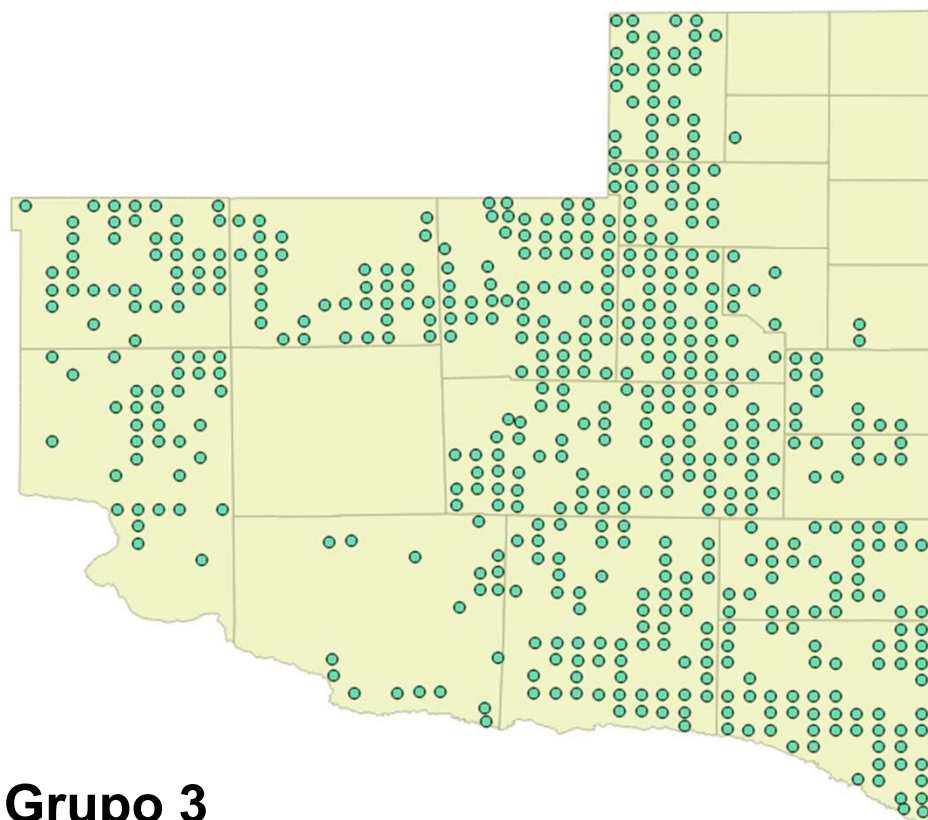
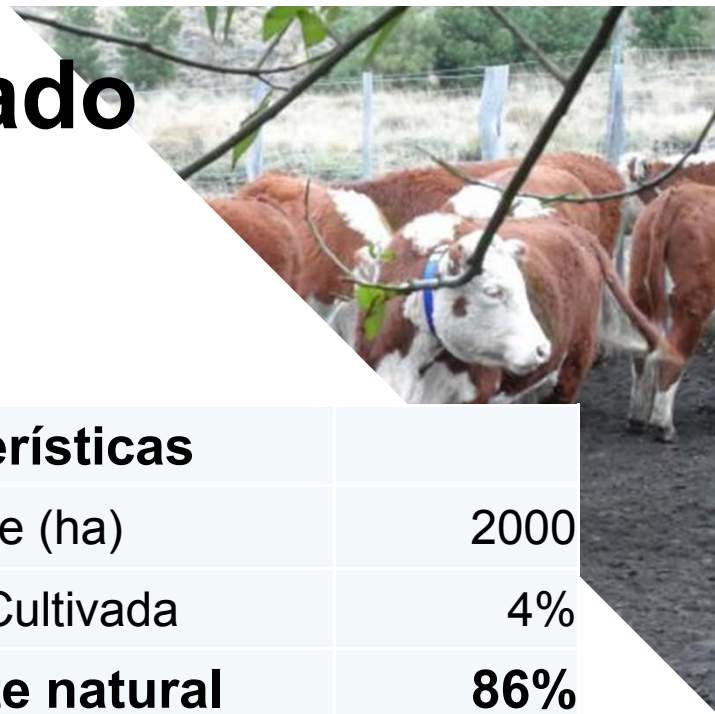
***Sistema agricultura de verano***



# Aprendizaje NO supervisado

*Sistemas productivos preponderantes*

*Algunos resultados:*



**Grupo 3**

**N 1065**

**Silueta 0.36**

**Jaccard 0.95**

## Características

Superficie (ha) 2000

% Area Cultivada 4%

**% Monte natural 86%**

## Ganadería

**% Bovinos 93%**

% Ovinos 3%

% Porcinos 0%

% Caprinos 1%

% Equinos 3%

***Sistema Cría Bovina***

# Aprendizaje NO supervisado

## *Sistemas productivos preponderantes*



### Herramientas



**fpc** Flexible Procedures for Clustering

by ORPHANED

[View Source](#)

### Equipo de trabajo



**Anabella  
LOZZA**

**Héctor  
LORDA**



# Aprendizaje Supervisado

Geotecnologías en  
la nube y su  
aplicación en  
emergencias  
agropecuarias

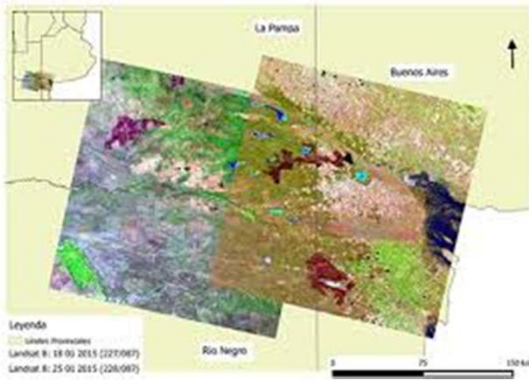




# Aprendizaje supervisado

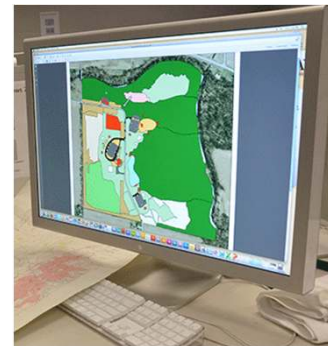
## Teledetección

Un poco de historia: 1998

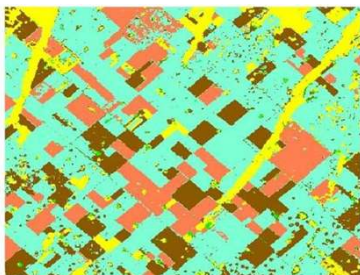


COMISIÓN NACIONAL DE  
ACTIVIDADES ESPACIALES

\$\$



\$\$





# Aprendizaje supervisado

## *Teledetección*

Obteniendo información desde datos remotos: Anguil



1984

1997

2000

2002

# Aprendizaje supervisado

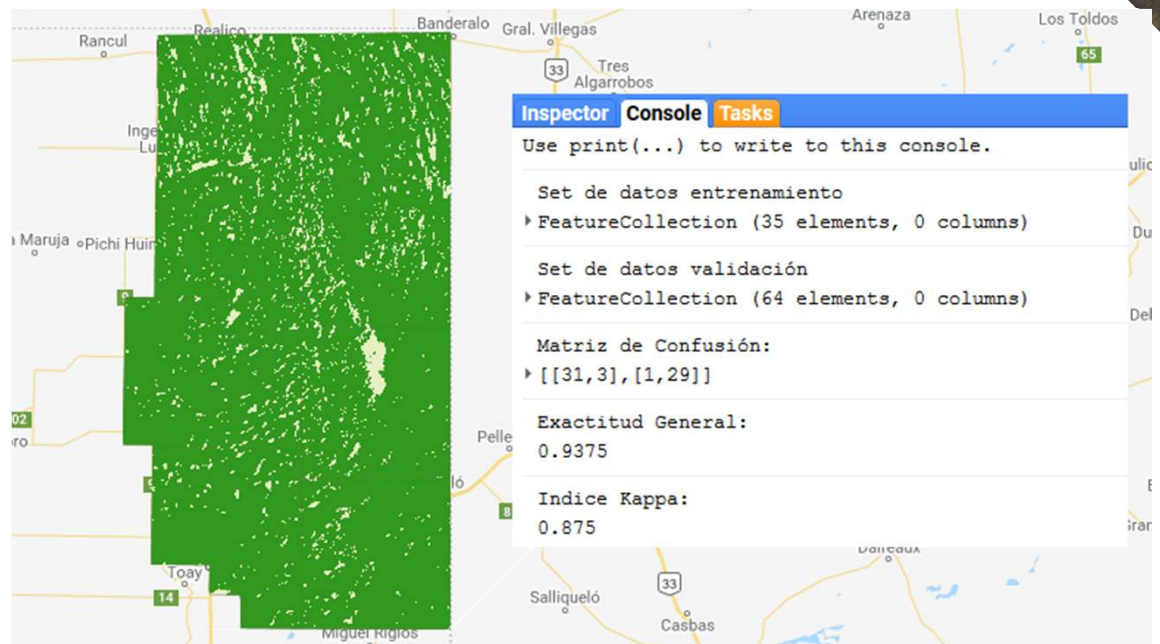
## Teledetección



Obtener una imagen



Filtro una imagen



Aplico un algoritmo de clasificacion



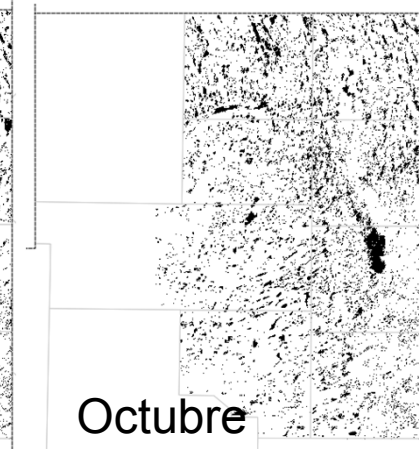
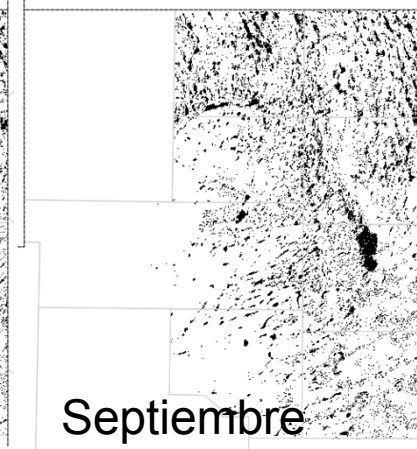
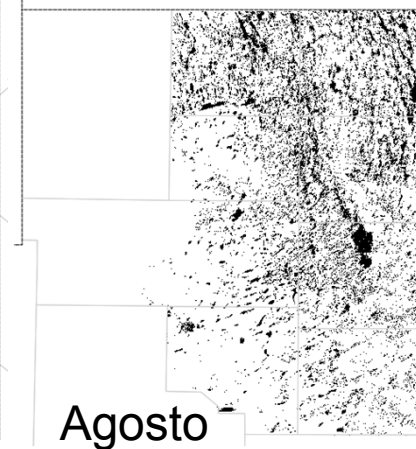
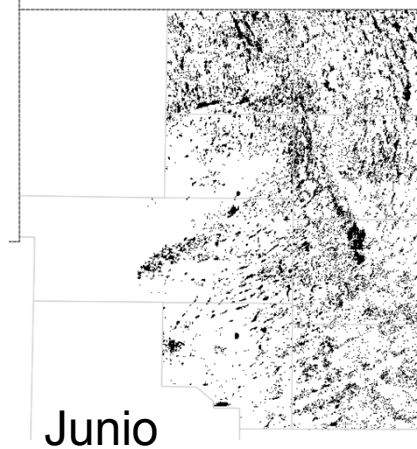
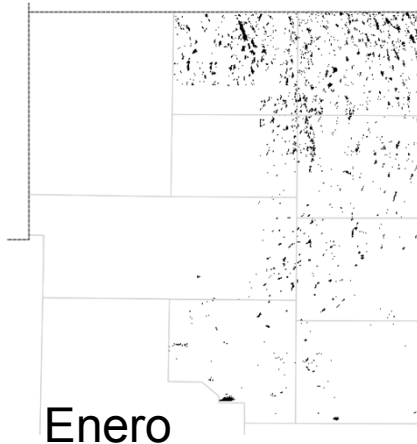
# Aprendizaje supervisado

## Teledetección



2017

Mes	Exactitud	Índice Kappa	Sup. afectada	Sup. analizada	%	Obs.
Enero	0,98	0,95	46.011	1.498.237	3	Nubes y humo
Agosto	0,99	0,97	300.337	2.219.654	14	



# Aprendizaje supervisado

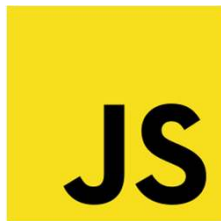
## Teledetección



### Herramientas



Google Earth Engine



### Equipo de trabajo



Julio  
FERNANDEZ



Laura  
BELMONTE

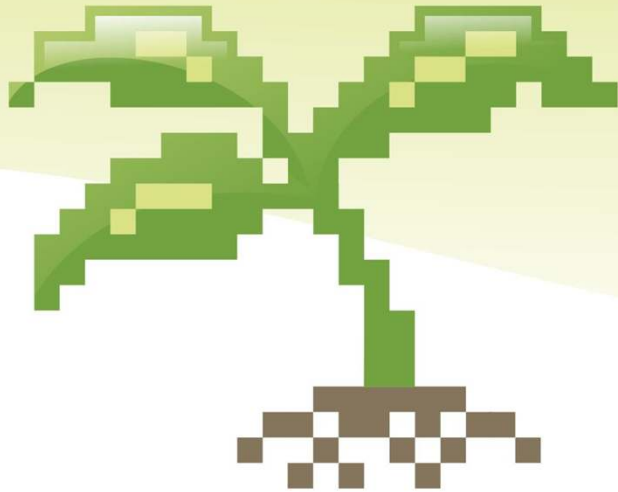


Mariela  
FUENTES



Universidad Nacional de La Pampa





# 11<sup>o</sup> CONGRESO de AGRO INFORMATICA

## Llamado a presentación de trabajos #CAI2019

Cierre de recepción de trabajos:  
26 de abril de 2019

Más detalles:

<http://48jaiio.sadio.org.ar/simposios/cai>

### Temas:

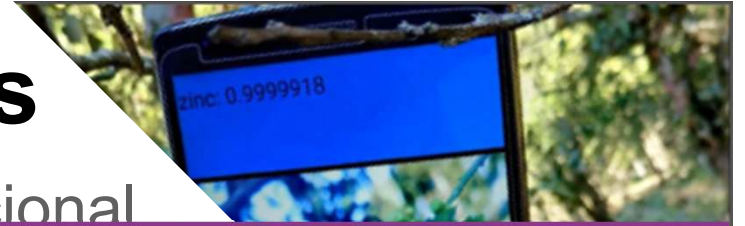
- Software y sistemas de información agropecuarios.
- Modelización de sistemas de producción.
- Integración y trazabilidad de cadenas agro-industriales.
- Geomática, Sistemas de Información Geográficos, IDEs, Teledetección y Observación Terrestre.
- Robótica agro-industrial.
- Agricultura y ganadería de precisión.
- Redes de sensores en cultivos, tambos, feed-lots y plantas de procesamiento.
- Sistemas embebidos y desarrollos electrónicos en la agro-industria.
- Monitoreo y control medio ambiental.
- Ontologías, Big Data, Open Data y DataMining e inteligencia artificial aplicadas al agro.
- Bioinformática y registros biológicos.
- Servicios Web Agroindustriales y Web 2.0.
- Nuevos desarrollos y experimentos en AgroTICs.
- Aplicaciones móviles.
- Internet de las cosas aplicadas al agro.
- Experiencias educativas en TICs aplicadas al agro.





# Vinculaciones y espacios

R-Ladies es una organización internacional



## R-Ladies Argentina



R-Ladies Bariloche

R-Ladies Buenos Aires

R-Ladies Córdoba

R-Ladies La Plata

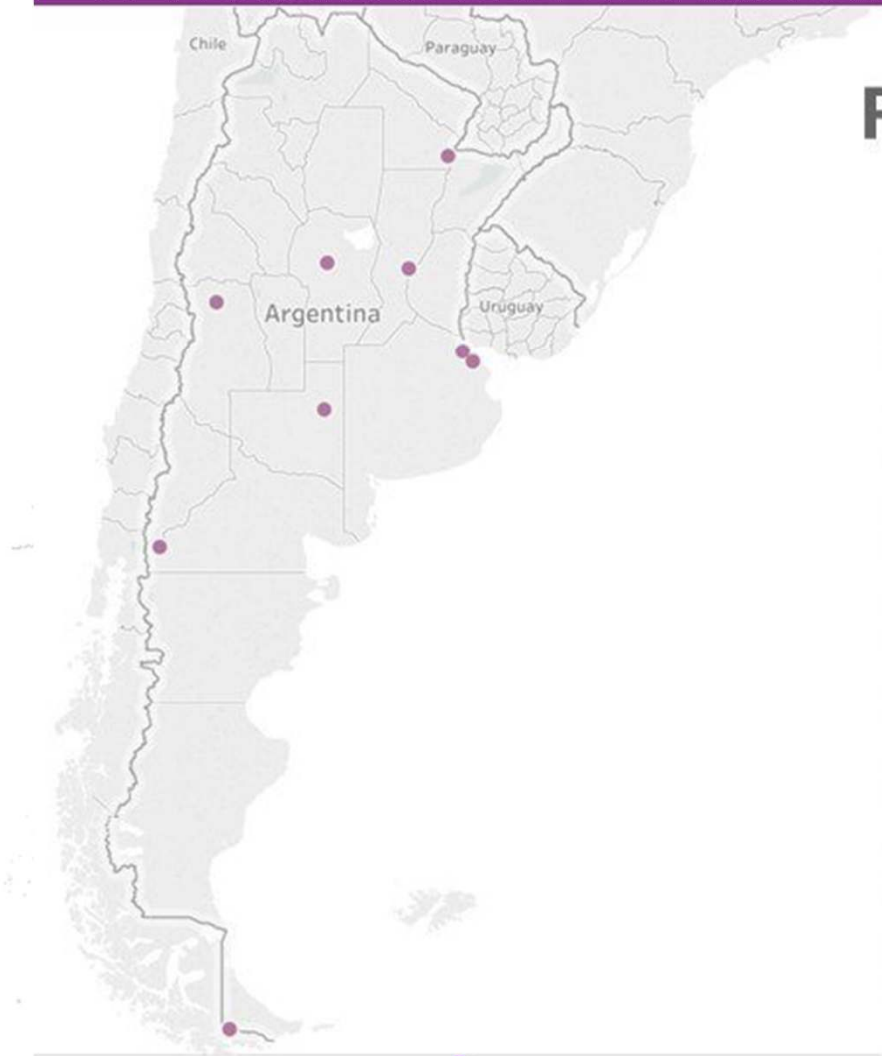
R-Ladies Mendoza

R-Ladies Resistencia - Corrientes

R-Ladies Santa Fe

R-Ladies Santa Rosa

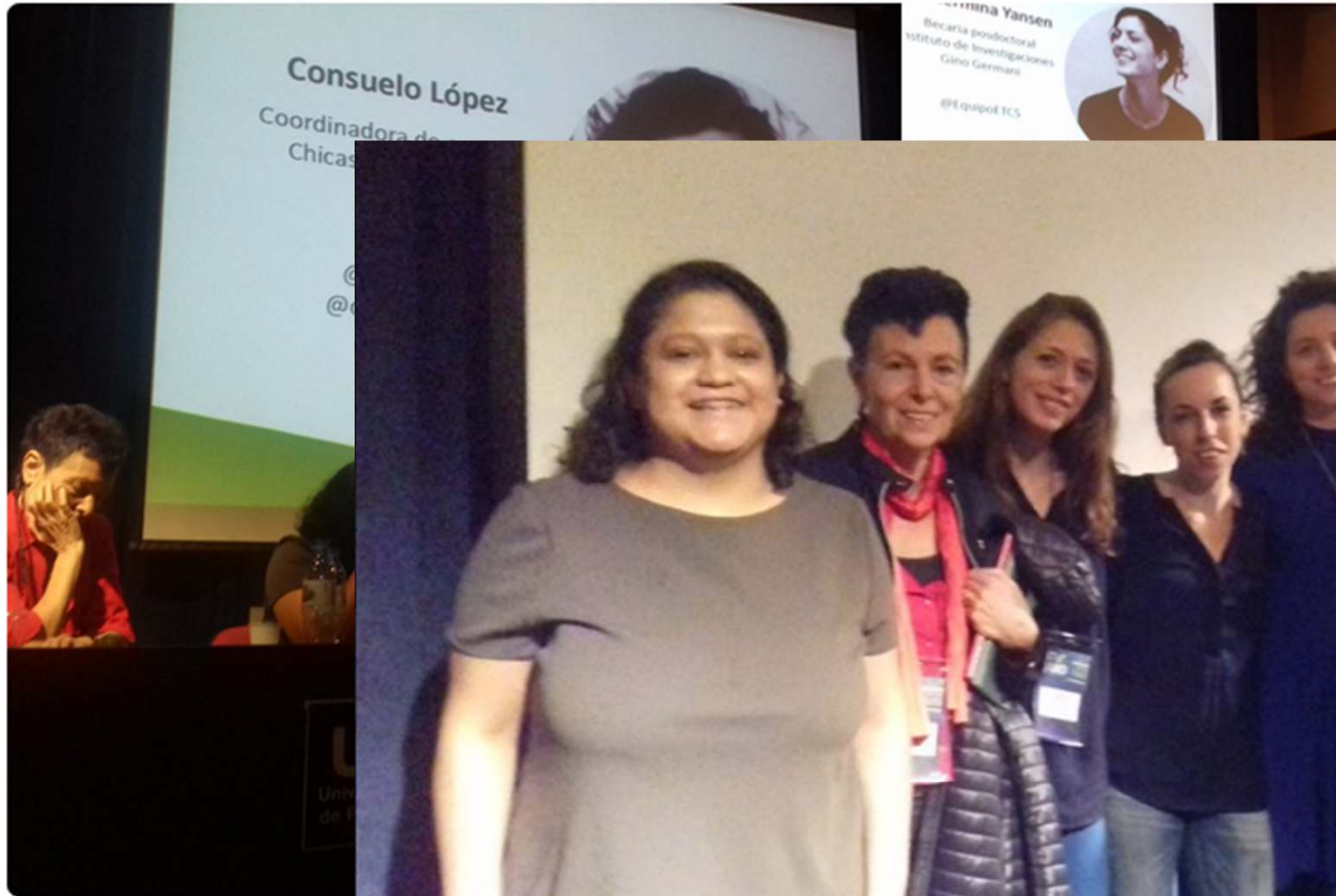
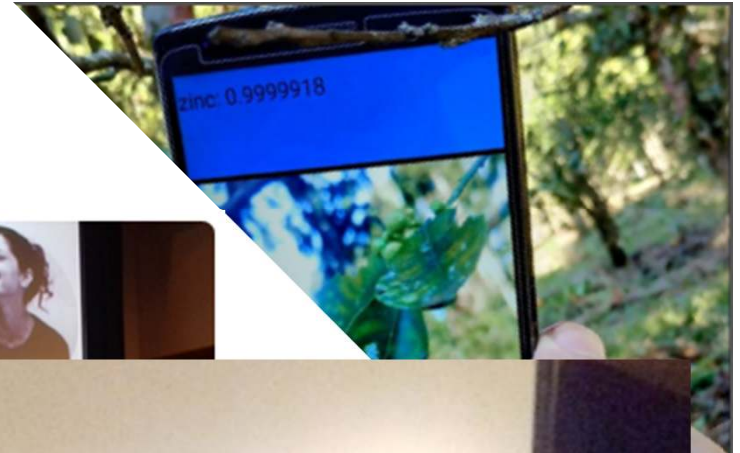
R-Ladies Ushuaia



<https://rladies.org/>



# Enfoque de género





# Eventos regionales



R-Ladies BuenosAires @RLadiesBA · 4 sept.

Presentación de #rladies en #LatinR2018 de la mano de @Chucheria , @\_lacion\_ y @yabellini ! :)





# Traducción colaborativa

R for Data Science: In

<https://github.com/cienciadedatos>



**Muchas gracias**  
**@yabellini**

**<http://48jaiio.sadio.org.ar/simposios/CAI>**

**<https://rladies.org/>**

**<https://github.com/cienciadedatos>**

**<https://github.com/INTARadar>**

**<https://github.com/yabellini>**



**WOMEN IN DATA SCIENCE**  
@ STANFORD UNIVERSITY



**FACULTAD  
DE INGENIERIA**  
Universidad de Buenos Aires