



**FLACSO**  
ARGENTINA

Facultad  
Latinoamericana de  
Ciencias Sociales.  
Sede Argentina.

**Área Comunicación  
y Cultura.**

# CURSO DE POSGRADO **BIG DATA E INTELIGENCIA TERRITORIAL**

## Clase 1. Algunas aplicaciones de Machine Learning



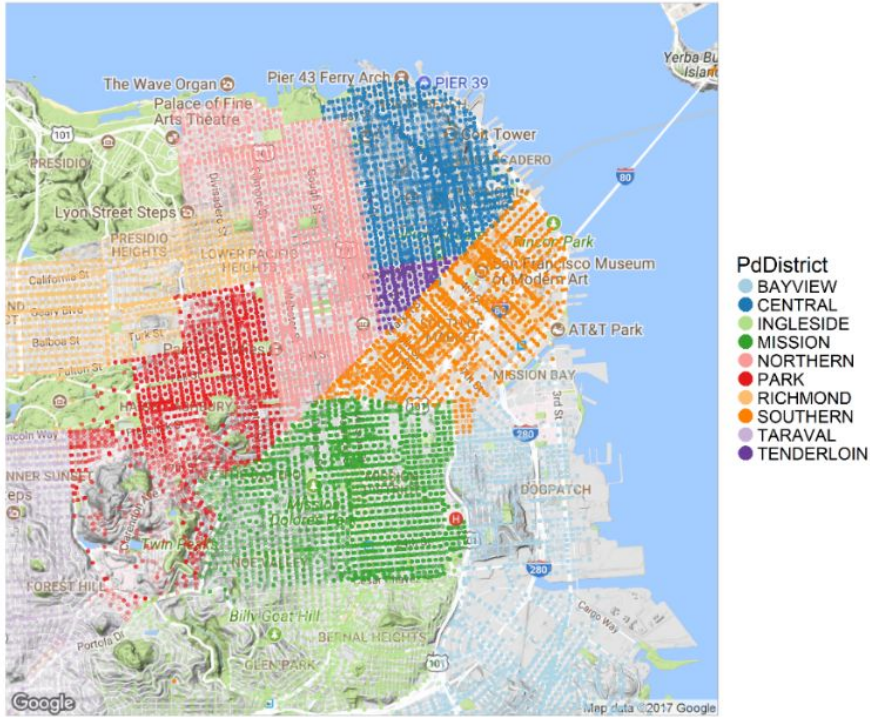
# *Para entender un poco más*

*Algunos nombres de modelos de ML que podemos encontrar en artículos científicos.*

<i>Machine Learning</i>	<i>Deep Learning</i>
árboles de decisión (decision trees)	redes neuronales
XGBOOST	redes convolucionales (CNN)
Random Forest	redes recurrentes (RNN)
Regresión logística	redes neuronales profundas (deep neural networks)
SVM	
DBSCAN	
Kmeans, hierarchical clustering	

# Predicción de crimen (Crime forecasting)

Map of PdDistricts



## Clasificación

XGBoost en este ejemplo

(<https://www.relatly.com/predicting-crime-type-in-san-francisco-using-xgboost/2960/>)

Paper (review):

<https://vciba.springeropen.com/articles/10.1186/s42492-021-00075-z>

# Predicción de crimen (Crime forecasting)

## variables

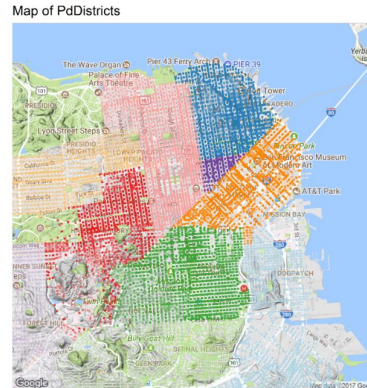
Datos de la ciudad de San Francisco, disponibles en:

<https://www.kaggle.com/c/sf-crime>

Datos: 800K observaciones, 9 variables

Modelos: DT / Random Forest / XGBoost

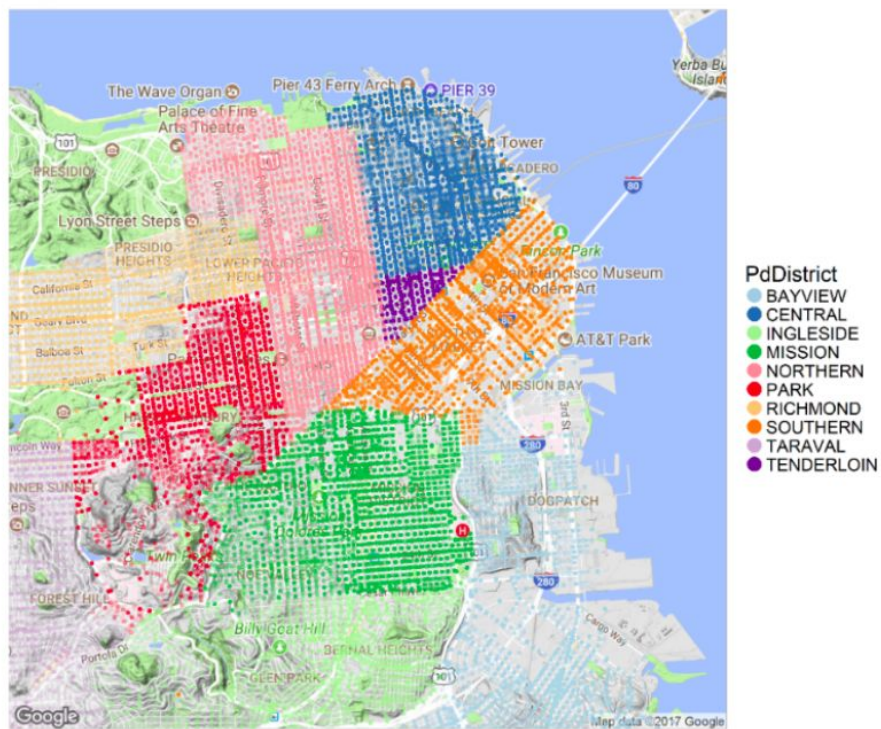
*(puede probarse cualquiera que realice clasificaciones)*



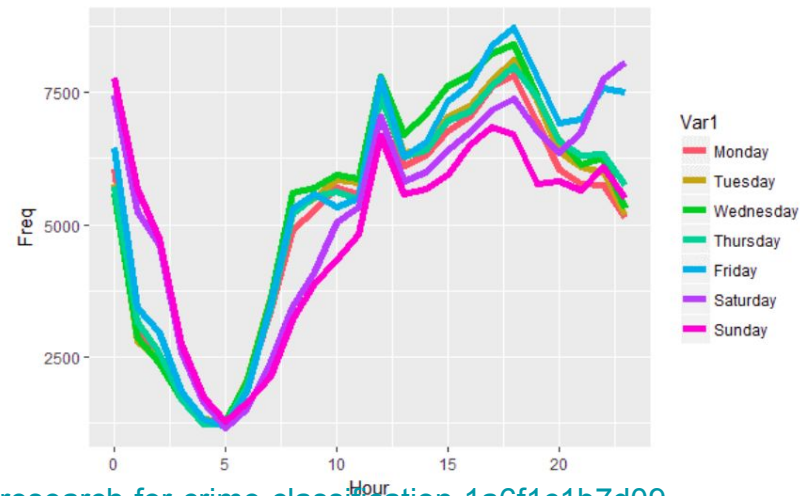
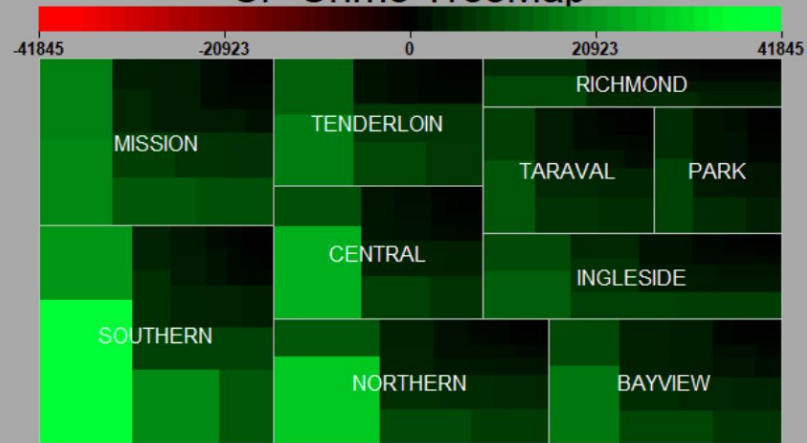
- Dates — timestamp of the crime incident
- Category — category of the crime incident
- Descript — detailed description of the crime incident
- Day of week — the day if the week
- PdDistrict — name of Police Department District
- Resolution — how the crime incident was resolved
- Address — Approximate street address of the crime incident
- X — longitude
- Y — latitude

The X and Y essentially gives the location parameter.

Map of PdDistricts

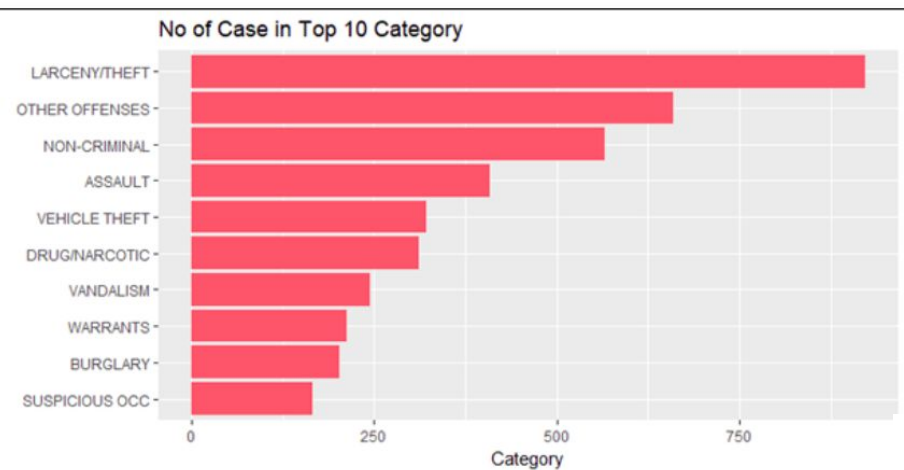


SF Crime TreeMap



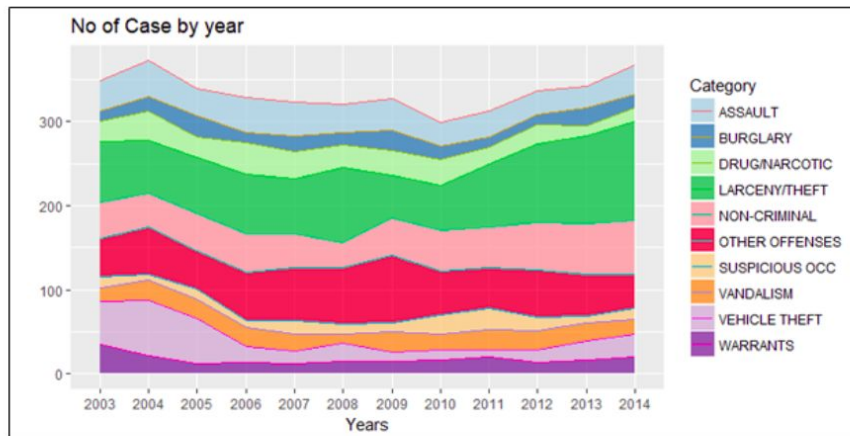


# Predicción de crimen (Crime forecasting)



Prevailing Crimes

Variables más importantes:  
Longitude, Latitude, Police department district, year of crime, week number in the entire year, day of week and month.



Prevailing Crimes with respect to Years

# Planeamiento urbano: Estacionamiento (Parking Capacity)



Figure 1. Map of predicted proportion of driveways in Somerville by block

## Clasificación

*CNN inicialmente (no funcionó).*

Modelos: Regresión logística,  
Random Forest, redes neuronales.

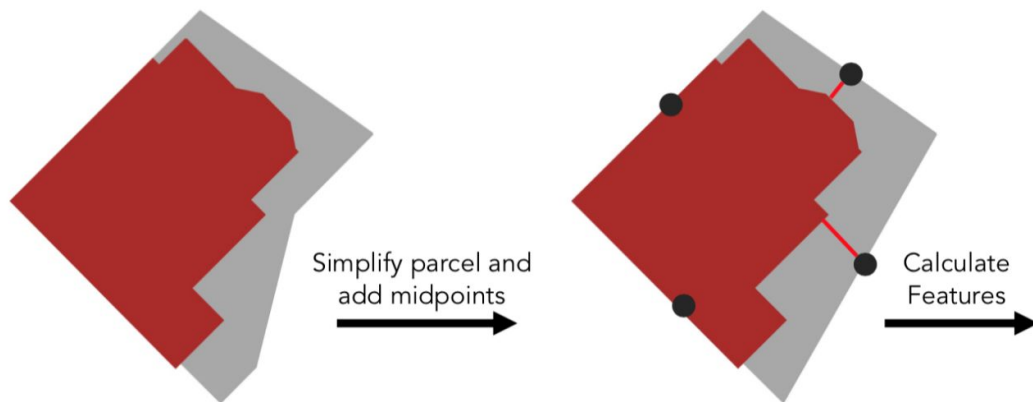
Problema desbalanceado (tengo más  
elementos de una clase que de otra),  
eso me complica a la hora de realizar  
la clasificación

Paper:

<https://www.sciencedirect.com/science/article/pii/S2542660520301335?via%3Dihub>

# Planeamiento urbano: Estacionamiento (Parking Capacity)

## Variables (Feature engineering)



- Distance from building to side 1, side 2, front, and back of parcel
- Number of buildings
- How much closer building is to one side of parcel and to front of parcel
- Number of edges in parcel and building polygons
- Distance to nearest neighbors

Finalmente se trabajó con probabilidades calibradas.



# Planeamiento urbano: Estacionamiento (Parking Capacity)

1. Se intentó utilizar redes convolucionales (CNN)
2. Luego, se adoptó una metodología más simple basada en ML.

Importancia de las variables  
obtenido con random forest



Figure 4. Feature importances from random forest model

# *Detección de edificios en ciudades*



## Segmentación de imágenes (Computer vision)

Paper (congreso CVPR):

[https://openaccess.thecvf.com/content\\_cvpr\\_2018\\_workshops/papers/w4/Hamaguchi\\_Building\\_Detection\\_From\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w4/Hamaguchi_Building_Detection_From_CVPR_2018_paper.pdf)

Fuente: <https://code.iadb.org/en/tools/building-detection>

# *Detección de edificios en ciudades*

Datos y código disponible en:

<https://github.com/EL-BID/Building-Detection>

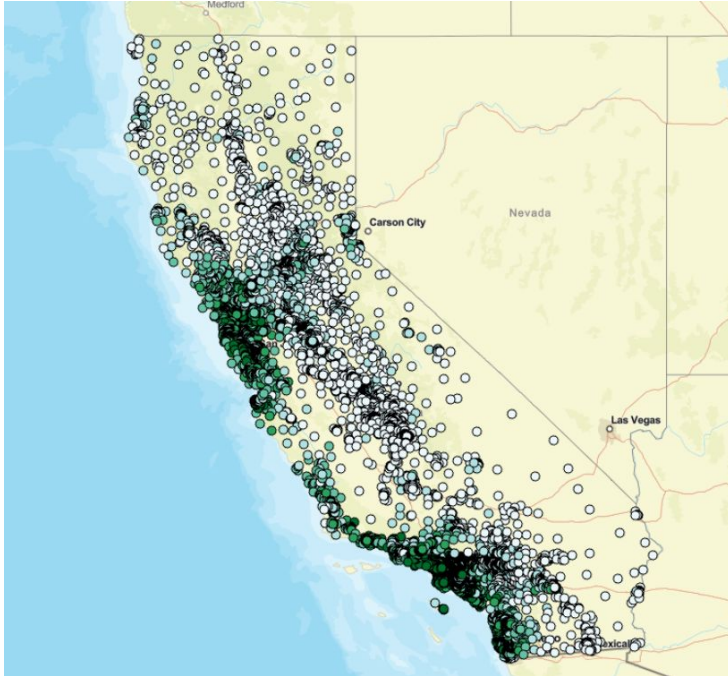
- Datos: imágenes raster

Este es un problema de segmentación de imágenes, es decir, que vamos a asignar una categoría a cada píxel de la imagen. En este caso, las categorías utilizadas son "bienes inmuebles" o "bienes no inmuebles". Al permitir una segmentación que reconoce las características únicas de cada ciudad (diferentes patrones de urbanización, diferentes accidentes geográficos, etc.), esta herramienta es aplicable en varios contextos.



Fuente: <https://code.iadb.org/en/tools/building-detection>

# Estimar precios de casas



## Regresión

Modelo: Random Forest.

## **Nota:**

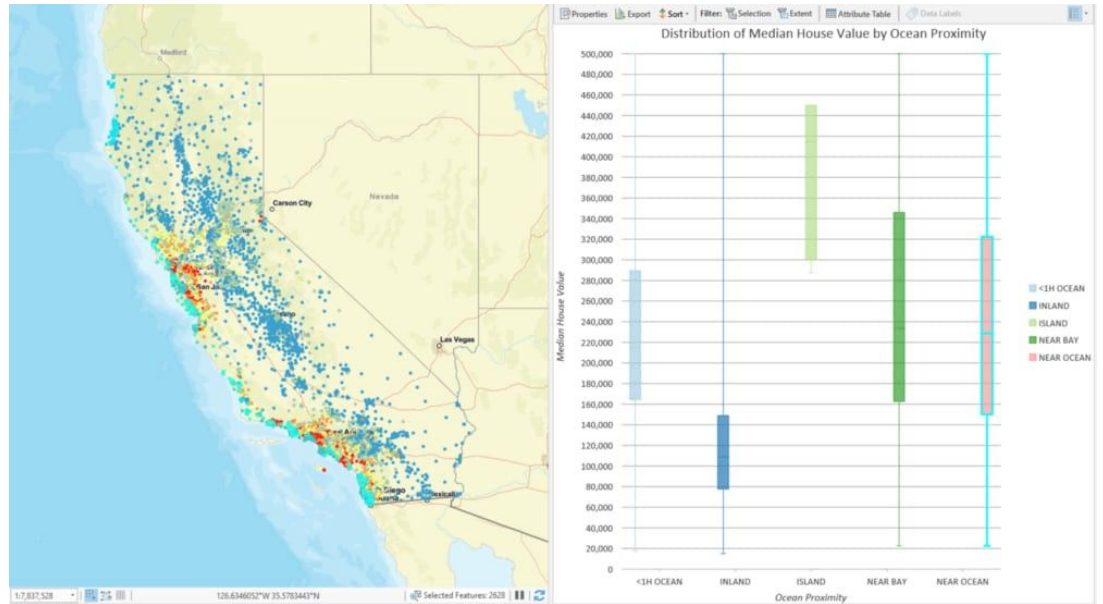
Puede utilizarse cualquier modelo que nos ayude a predecir una variable cuantitativa.

# *Estimar precios de casas*

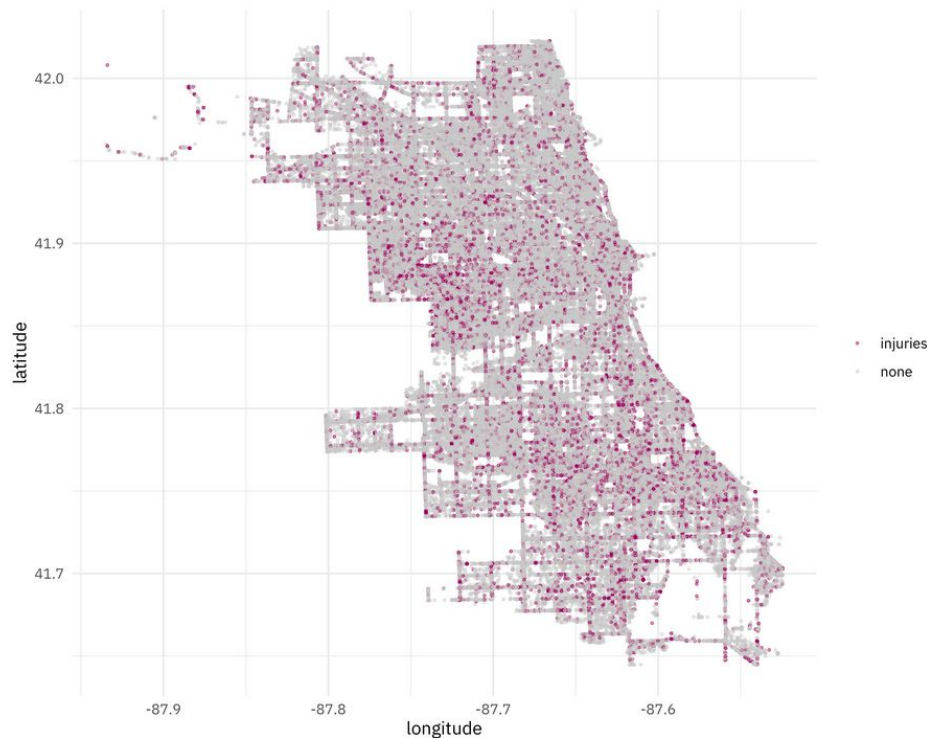
Datos: tabulares disponibles en

<https://www.kaggle.com/harrywang/housing#housing.csv>

Herramientas: ARCGis



# *Estimar lesiones a partir de accidentes automovilísticos*



## Clasificación

Modelo: bagged tree

Datos de la ciudad de Chicago

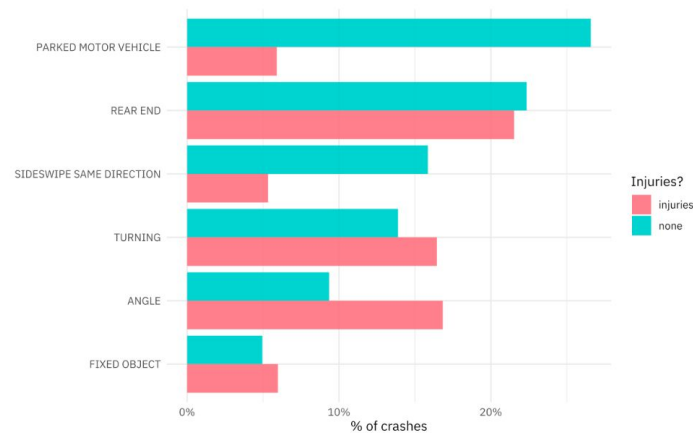
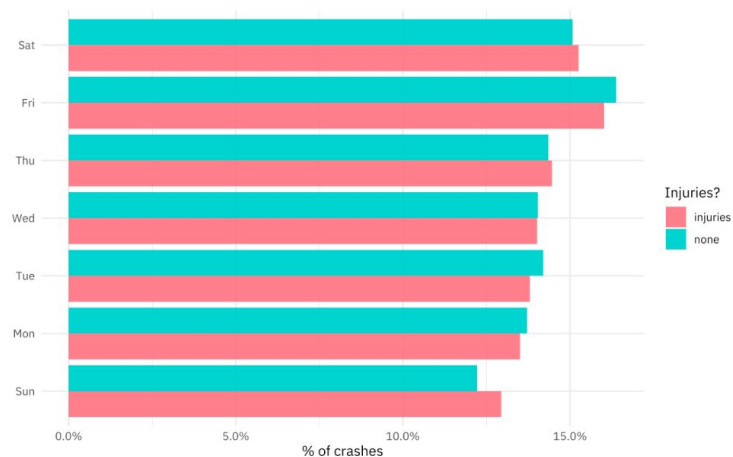
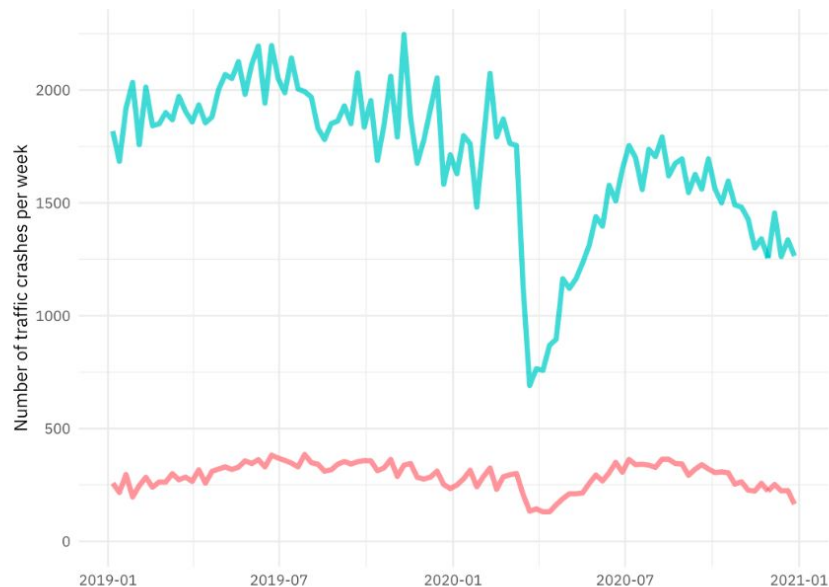
Paper:

[https://link.springer.com/chapter/10.1007%2F978-981-15-7961-5\\_96](https://link.springer.com/chapter/10.1007%2F978-981-15-7961-5_96)

Fuente: <https://juliasilge.com/blog/chicago-traffic-model/>



# Estimar lesiones a partir de accidentes automovilísticos



Fuente: <https://julasilge.com/blog/chicago-traffic-model/>

# *Estimar lesiones a partir de accidentes automovilísticos*

Datos de la ciudad de Chicago

<https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>

*Dataframe (datos puntuales)*

Modelo: bagged tree

Clasificación con clases desbalanceadas (hay mayoría de una clase).

Fuente: <https://juliasilge.com/blog/chicago-traffic-model/>



# Detectar centros de actividad urbana



## Clustering

Modelo: DBSCAN

(modelo no supervisado  
de clusterizado)

Paper (DBSCAN junto a otros  
métodos)

[https://www.researchgate.net/publication/333912582\\_Urban\\_Crowd\\_Detection\\_Using\\_SOM\\_DBSCAN\\_and\\_LBSN\\_Data\\_Entropy\\_A\\_Twitter\\_Experiment\\_in\\_New\\_York\\_and\\_Madrid](https://www.researchgate.net/publication/333912582_Urban_Crowd_Detection_Using_SOM_DBSCAN_and_LBSN_Data_Entropy_A_Twitter_Experiment_in_New_York_and_Madrid)

# *Detectar centros de actividad urbana*

Modelo no supervisado. Este modelo se

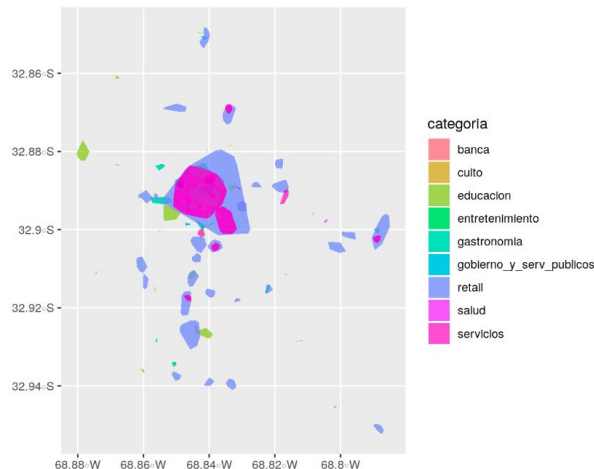
lo utiliza en este ejemplo, para identificar los centros comerciales (y de actividad) en la ciudad de Mendoza y sus alrededores

Datos: [https://bitsandbricks.github.io/data/mendoza\\_poi.csv](https://bitsandbricks.github.io/data/mendoza_poi.csv)

Luego de clusterizar (agrupar) los datos,

aplica una envolvente convexa (convex hull)

o polígono.



Fuente: <https://bitsandbricks.github.io/post/dbscan-machine-learning-para-detectar-centros-de-actividad-urbana/>

# Administración del tráfico de una ciudad

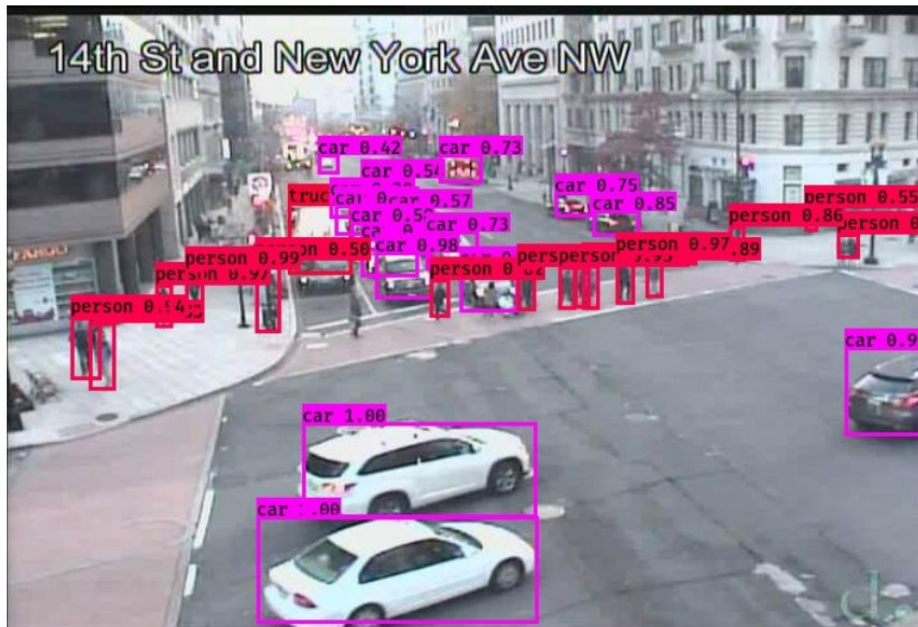


Figure 8. Anomalies of Pedestrians and Cars at One of the Intersections

## Detección de objetos en imágenes

Modelo: YOLO + DBSCAN



# Administración del tráfico de una ciudad

## Objetivos Particulares:

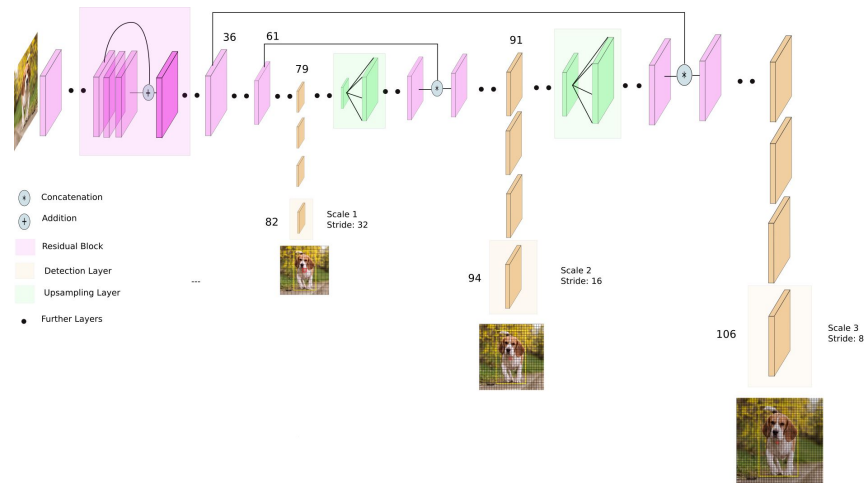
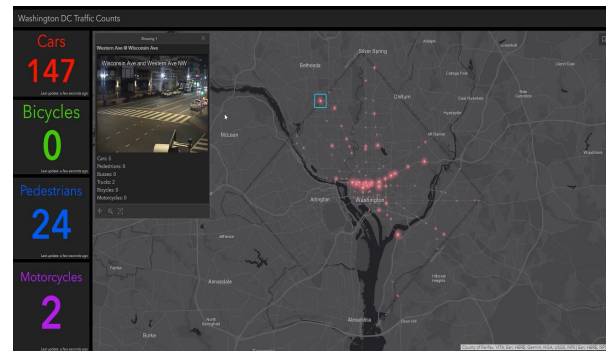
- Monitoreo de tráfico
- Detección de anomalías en el tráfico
- Detectar peatones en lugares no seguros(DBSCAN).

Datos: imágenes a partir de cámaras.

Modelo alojado en AWS (entorno cloud)

Paper:

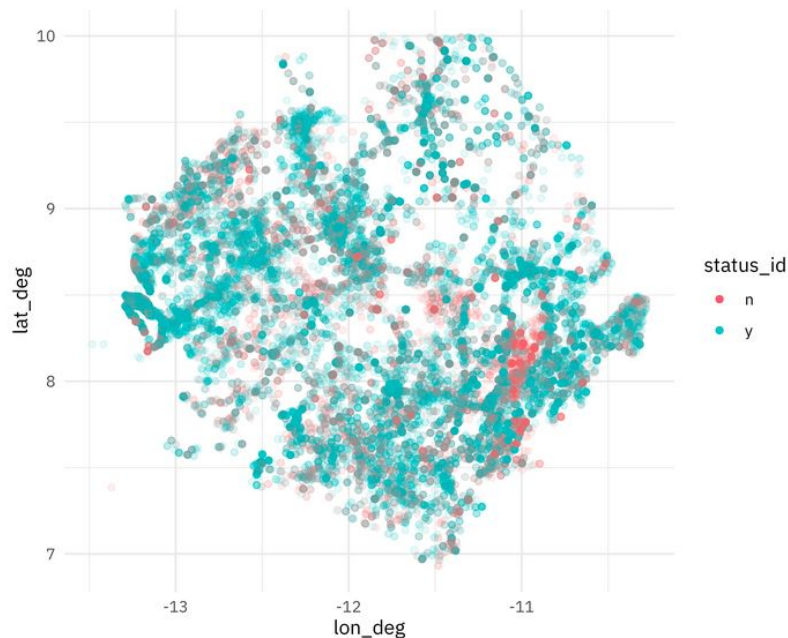
<https://ieeexplore.ieee.org/document/8611986>



Fuente: <https://medium.com/geoai/an-end-to-end-solution-on-the-cloud-to-monitor-traffic-flow-using-deep-learning-9dfdfd00b621>



# Manejo del agua (Water management)



## Clasificación

Modelo: Random Forest

Ejemplo con código en R:

<https://juliasilge.com/blog/water-sources/>

Lista de papers:

<https://www.frontiersin.org/research-topics/13328/machine-learning-for-water-resources>

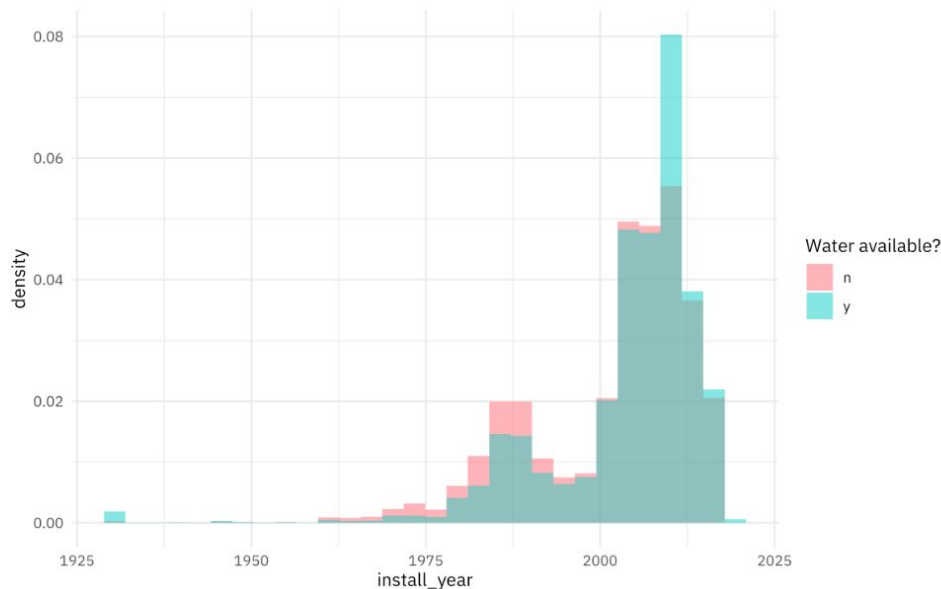
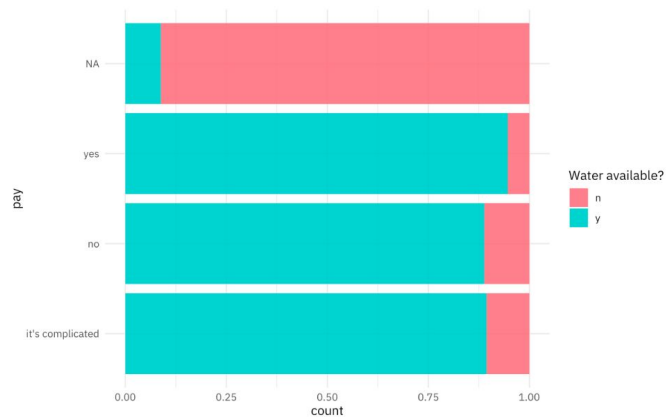
# Manejo del agua (Water management)

Datos:

<https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-05-04/readme.md>

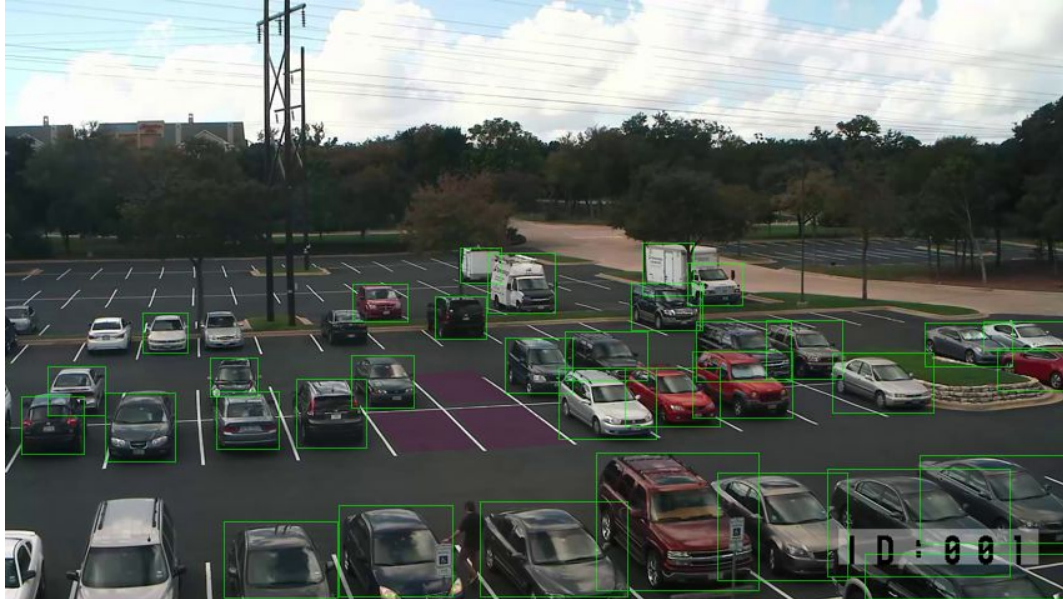
Datos puntuales (dataframe)

Modelo: Random Forest



Fuente: <https://juliasilge.com/blog/water-sources/>

# *Detección de espacios de estacionamiento*



## Detección de objetos

Modelo: Fast RCNN

(deep learning)

Paper (SVM y CNN):

<http://ceur-ws.org/Vol-2087/paper5.pdf>

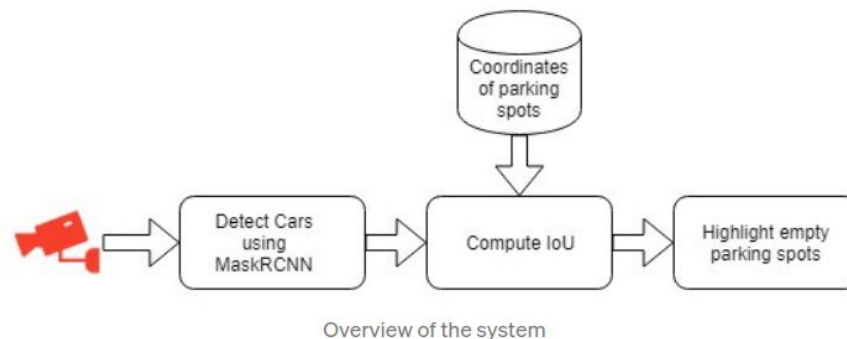
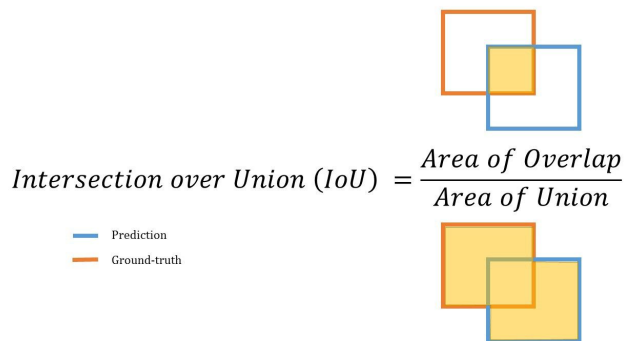
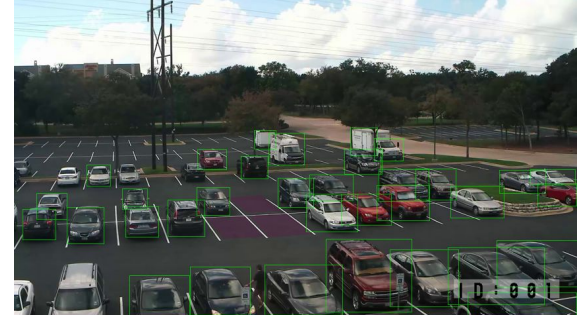
# Detección de espacios de estacionamiento

Datos: video / imágenes

Código:

<https://github.com/sainimohit23/parking-space-detection-system>

Modelo pre-entrenado: Fast R-CNN



Fuente: <https://medium.com/the-research-nest/parking-space-detection-using-deep-learning-9fc99a63875e>

# Smart Cities



Review que comenta los modelos más utilizados:

[https://www.researchgate.net/publication/345226030\\_Machine\\_Learning\\_Approaches\\_in\\_Smart\\_Cities](https://www.researchgate.net/publication/345226030_Machine_Learning_Approaches_in_Smart_Cities)

Problemas de ML a resolver con IoT & ML

<https://www.impactqa.com/blog/what-is-the-role-of-iot-machine-learning-in-smart-cities/>

# Bibliografía

- <https://medium.com/geoai/using-forest-based-classification-regression-to-model-and-estimate-house-values-5a0e26682c42>
- <https://vciba.springeropen.com/articles/10.1186/s42492-021-00075-z>
- <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229509>
- <https://www.semanticscholar.org/paper/Geographical-random-forests%3A-a-spatial-extension-of-Georganos-Grappa/ba168390bb1e762e502d8f7674fc374bef014b2a>
- <https://towardsdatascience.com/machine-learning-for-urban-planning-estimating-parking-capacity-15aab490cf8>
- <https://code.iadb.org/en/tools/building-detection>
- [https://thesai.org/Downloads/Volume12No7/Paper\\_87-Machine\\_Learning\\_Predictors\\_for\\_Sustainable\\_Urban\\_Planning.pdf](https://thesai.org/Downloads/Volume12No7/Paper_87-Machine_Learning_Predictors_for_Sustainable_Urban_Planning.pdf)
- <https://medium.com/geoai>
- [https://www.researchgate.net/publication/275220711\\_Using\\_Machine\\_Learning\\_Algorithms\\_to\\_Analyze\\_Crime\\_Data](https://www.researchgate.net/publication/275220711_Using_Machine_Learning_Algorithms_to_Analyze_Crime_Data)
- <https://www.sciencedirect.com/science/article/pii/S0198971516301831>
- [https://www.researchgate.net/publication/345226030\\_Machine\\_Learning\\_Approaches\\_in\\_Smart\\_Cities](https://www.researchgate.net/publication/345226030_Machine_Learning_Approaches_in_Smart_Cities)
- <https://www.sciencedirect.com/science/article/pii/S2542660520301335?via%3Dihub>
- [https://openaccess.thecvf.com/content\\_cvpr\\_2018\\_workshops/papers/w4/Hamaguchi\\_Building\\_Detection\\_From\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w4/Hamaguchi_Building_Detection_From_CVPR_2018_paper.pdf)
- [https://www.researchgate.net/publication/333912582\\_Urban\\_Crowd\\_Detection\\_Using\\_SOM\\_DBSCAN\\_and\\_LBSN\\_Data\\_Entropy\\_A\\_Twitter\\_Experiment\\_in\\_New\\_York\\_and\\_Madrid](https://www.researchgate.net/publication/333912582_Urban_Crowd_Detection_Using_SOM_DBSCAN_and_LBSN_Data_Entropy_A_Twitter_Experiment_in_New_York_and_Madrid)



# *Práctico 1*

# Práctico 1

- Buscar en Medium, blogs, o artículos científicos:
  1. Un ejemplo espacial de machine learning.
  2. Un ejemplo espacial de deep learning.

En los mismos determine objetivo/s del trabajo, tipo de datos, modelo de ML/DL utilizado, y enunciar si se trata de regresión/clasificación/clustering

O sea, en total, debemos buscar 2 ejemplos.

Links de búsqueda:

- <https://medium.com/>
- <https://www.sciencedirect.com/>
- <https://www.springernature.com/la>