

# Modelos Matemáticos Predictivos de Crímenes en Zonas Urbanas

Alvaro J. Riascos Villegas y Mateo Dulce  
Universidad de los Andes y Quantil

Agosto de 2017

# Contenido

- 1 Motivación
- 2 Modelo preliminares
  - Modelo de puntos
  - Modelo de elipses espaciales
  - Modelo KDE
  - Modelo Espacio - Temporal
  - Validación
  - Aplicativo
- 3 Modelo Espacio - Temporal: Ancho de banda variable
  - Validación
- 4 Modelos Avanzados
  - ¿Por qué molestarse?
  - Trabajos previos
  - Deep Learning

# Modelos Matemáticos

- En los últimos veinte años en los círculos académicos se han venido desarrollando diversas teorías matemáticas del crimen.
  - Gorr, W. L., Lee, Y. 2015. Early Warning System for Temporary Crime Hot Spots. Journal of Quantitative Criminology
  - Chainey S., Tompson L., Uhlig S. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal.
  - Block, C. R. 1995. STAC hot-spot areas: A statistical tool for law enforcement decisions. Washington, DC: Police Executive Research Forum.

# Modelos Matemáticos

- En los últimos veinte años en los círculos académicos se han venido desarrollando diversas teorías matemáticas del crimen.
  - Gorr, W. L., Lee, Y. 2015. Early Warning System for Temporary Crime Hot Spots. Journal of Quantitative Criminology
  - Chainey S., Tompson L., Uhlig S. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal.
  - Block, C. R. 1995. STAC hot-spot areas: A statistical tool for law enforcement decisions. Washington, DC: Police Executive Research Forum.

- Esta literatura a comenzado a influenciar la administración de los recursos policiales de los principales centros urbanos: Los Ángeles CA, Atlanta GA, Chicago IL, New York NW, Alhambra CA, San Francisco CA, Modesto CA, Santa Cruz, CA.

- Una aproximación cuantitativa al problema de criminalidad permite:
  - Diseñar estrategias óptimas: priorizar y usar de forma eficiente los recursos.
  - Tomar decisiones más informadas.
  - Medir el impacto de diferentes políticas.
  - Monitorear y hacer un mejor seguimiento de las decisiones.
- En general, hacer más científica la gestión de la Policía relacionada con el combate al crimen.

- Una aproximación cuantitativa al problema de criminalidad permite:
  - Diseñar estrategias óptimas: priorizar y usar de forma eficiente los recursos.
  - Tomar decisiones más informadas.
  - Medir el impacto de diferentes políticas.
  - Monitorear y hacer un mejor seguimiento de las decisiones.
- En general, hacer más científica la gestión de la Policía relacionada con el combate al crimen.

# Contenido

- 1 Motivación
- 2 **Modelo preliminares**
  - Modelo de puntos
  - Modelo de elipses espaciales
  - Modelo KDE
  - Modelo Espacio - Temporal
  - Validación
  - Aplicativo
- 3 Modelo Espacio - Temporal: Ancho de banda variable
  - Validación
- 4 Modelos Avanzados
  - ¿Por qué molestarse?
  - Trabajos previos
  - Deep Learning



## Modelos preliminares

- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Ésta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

## Modelos preliminares

- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Ésta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

## Modelos preliminares

- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Ésta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

## Modelos preliminares

- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Ésta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

## Modelos preliminares

- Se diseñó una metodología para comparar diferentes modelos de predicción del crimen en Bogotá.
- Ésta se extiende de forma natural a otros centros urbanos con la información adecuada (Cali, Medellín).
- Datos utilizados: 329,793 crímenes ocurridos en Bogotá entre 2004 y 2014 (geolocalizados, con fecha y hora).
- Se compararon modelos de Puntos, Elipses, KDE y varias versiones de modelos espacio-temporales.

# Modelo de puntos

- Estrategia más simple: conteo en una grilla.
- Se basa en el supuesto de que los 'hotspots' son áreas de aproximadamente una cuadra de radio.
- El conteo de crímenes en cada celda nos permite detectar regiones de alta criminalidad.
- Pocos parámetros: Tamaño de la celda ( 150 Mts.) y umbral de hotspot.

# Modelo de puntos

- Estrategia más simple: conteo en una grilla.
- Se basa en el supuesto de que los 'hotspots' son áreas de aproximadamente una cuadra de radio.
- El conteo de crímenes en cada celda nos permite detectar regiones de alta criminalidad.
- Pocos parámetros: Tamaño de la celda ( 150 Mts.) y umbral de hotspot.

# Modelo de puntos

- Estrategia más simple: conteo en una grilla.
- Se basa en el supuesto de que los 'hotspots' son áreas de aproximadamente una cuadra de radio.
- El conteo de crímenes en cada celda nos permite detectar regiones de alta criminalidad.
- Pocos parámetros: Tamaño de la celda ( 150 Mts.) y umbral de hotspot.



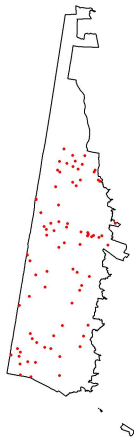
# Modelo de puntos

- Estrategia más simple: conteo en una grilla.
- Se basa en el supuesto de que los 'hotspots' son áreas de aproximadamente una cuadra de radio.
- El conteo de crímenes en cada celda nos permite detectar regiones de alta criminalidad.
- Pocos parámetros: Tamaño de la celda ( 150 Mts.) y umbral de hotspot.

- Tiene algunas limitaciones:
  - 1 El tamaño fijo de celda no permite capturar clústers más grandes.
  - 2 No modela *spill overs* de una celda a otra.
  - 3 Los *hotspots* se vuelven muy granulares y no capturan patrones generales.

# Modelo de puntos: Ejemplo

Point model illustrated (Usaquen)



## Modelo de elipses espaciales

- Se hace un clústering de los puntos en el mapa y se les asignan '*Standard Deviation Ellipses*'.
- Delimitan efectivamente regiones sin ceñirse a regiones pre existentes (cuadrantes, distritos, ...).
- Requieren muy pocos parámetros (radio de búsqueda y desviaciones).
- Sobre simplifican el análisis, no permite comparaciones al interior de las elipses.
- Distorsiona el análisis pues típicamente el crimen no se distribuye como una elipse.

## Modelo de elipses espaciales

- Se hace un clústering de los puntos en el mapa y se les asignan '*Standard Deviation Ellipses*'.
- Delimitan efectivamente regiones sin ceñirse a regiones pre existentes (cuadrantes, distritos, ...).
- Requieren muy pocos parámetros (radio de búsqueda y desviaciones).
- Sobre simplifican el análisis, no permite comparaciones al interior de las elipses.
- Distorsiona el análisis pues típicamente el crimen no se distribuye como una elipse.

## Modelo de elipses espaciales

- Se hace un clústering de los puntos en el mapa y se les asignan '*Standard Deviation Ellipses*'.
- Delimitan efectivamente regiones sin ceñirse a regiones pre existentes (cuadrantes, distritos, ...).
- Requieren muy pocos parámetros (radio de búsqueda y desviaciones).
- Sobre simplifican el análisis, no permite comparaciones al interior de las elipses.
- Distorsiona el análisis pues típicamente el crimen no se distribuye como una elipse.

## Modelo de elipses espaciales

- Se hace un clústering de los puntos en el mapa y se les asignan '*Standard Deviation Ellipses*'.
- Delimitan efectivamente regiones sin ceñirse a regiones pre existentes (cuadrantes, distritos, ...).
- Requieren muy pocos parámetros (radio de búsqueda y desviaciones).
- Sobre simplifican el análisis, no permite comparaciones al interior de las elipses.
- Distorsiona el análisis pues típicamente el crimen no se distribuye como una elipse.

## Modelo de elipses espaciales

- Se hace un clústering de los puntos en el mapa y se les asignan '*Standard Deviation Ellipses*'.
- Delimitan efectivamente regiones sin ceñirse a regiones pre existentes (cuadrantes, distritos, ...).
- Requieren muy pocos parámetros (radio de búsqueda y desviaciones).
- Sobre simplifican el análisis, no permite comparaciones al interior de las elipses.
- Distorsiona el análisis pues típicamente el crimen no se distribuye como una elipse.



# Modelo de elipses espaciales

- El radio de búsqueda determina los clústers que se forman.
- Es la distancia máxima que separa dos puntos de un mismo clúster.
- Luego los dos ejes de la elipse se calculan de acuerdo a:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}$$

- La elipse está orientada en la dirección de mayor variación.

# Modelo de elipses espaciales

- El radio de búsqueda determina los clústers que se forman.
- Es la distancia máxima que separa dos puntos de un mismo clúster.
- Luego los dos ejes de la elipse se calculan de acuerdo a:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}$$

- La elipse está orientada en la dirección de mayor variación.

# Modelo de elipses espaciales

- El radio de búsqueda determina los clústers que se forman.
- Es la distancia máxima que separa dos puntos de un mismo clúster.
- Luego los dos ejes de la elipse se calculan de acuerdo a:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}$$

- La elipse está orientada en la dirección de mayor variación.

# Modelo de elipses espaciales

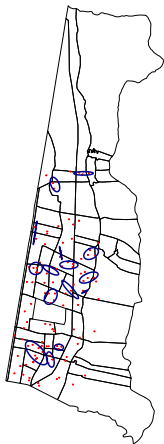
- El radio de búsqueda determina los clústers que se forman.
- Es la distancia máxima que separa dos puntos de un mismo clúster.
- Luego los dos ejes de la elipse se calculan de acuerdo a:

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}}$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}$$

- La elipse está orientada en la dirección de mayor variación.

# Modelo de Elipses: Ejemplo



# Modelo KDE

- Ajustar una densidad gaussiana a cada punto y sumar.
- Es una suavización de la distribución observada.
- Revela áreas de alta criminalidad sin ceñirse a regiones pre establecidas.
- Sólo hay que ajustar el ancho de banda, aunque no hay consenso sobre la mejor forma de hacerlo (usamos como criterio el error PAI).

# Modelo KDE

- Ajustar una densidad gaussiana a cada punto y sumar.
- Es una suavización de la distribución observada.
- Revela áreas de alta criminalidad sin ceñirse a regiones pre establecidas.
- Sólo hay que ajustar el ancho de banda, aunque no hay consenso sobre la mejor forma de hacerlo (usamos como criterio el error PAI).

# Modelo KDE

- Ajustar una densidad gaussiana a cada punto y sumar.
- Es una suavización de la distribución observada.
- Revela áreas de alta criminalidad sin ceñirse a regiones pre establecidas.
- Sólo hay que ajustar el ancho de banda, aunque no hay consenso sobre la mejor forma de hacerlo (usamos como criterio el error PAI).



# Modelo KDE

- Ajustar una densidad gaussiana a cada punto y sumar.
- Es una suavización de la distribución observada.
- Revela áreas de alta criminalidad sin ceñirse a regiones pre establecidas.
- Sólo hay que ajustar el ancho de banda, aunque no hay consenso sobre la mejor forma de hacerlo (usamos como criterio el error PAI).

- Se ajustan distribuciones gaussianas:

$$K(x, x_i, \Sigma) = \frac{1}{\sqrt{(2\pi)^3 \det(\Sigma)}} \exp \left[ -\frac{1}{2} (x - x_i)^t \Sigma^{-1} (x - x_i) \right] \quad (1)$$

- Y se suman sobre todo el mapa

$$u(x, \Sigma) = \frac{1}{n+1} \sum_{i=0}^n K(x, x_i, \Sigma) \quad (2)$$

- El parámetro  $\Sigma$  de mejor ajuste se seleccionó basado en el PAI sobre un conjunto de prueba.

- Se ajustan distribuciones gaussianas:

$$K(x, x_i, \Sigma) = \frac{1}{\sqrt{(2\pi)^3 \det(\Sigma)}} \exp \left[ -\frac{1}{2} (x - x_i)^t \Sigma^{-1} (x - x_i) \right] \quad (1)$$

- Y se suman sobre todo el mapa

$$u(x, \Sigma) = \frac{1}{n+1} \sum_{i=0}^n K(x, x_i, \Sigma) \quad (2)$$

- El parámetro  $\Sigma$  de mejor ajuste se seleccionó basado en el PAI sobre un conjunto de prueba.

- Se ajustan distribuciones gaussianas:

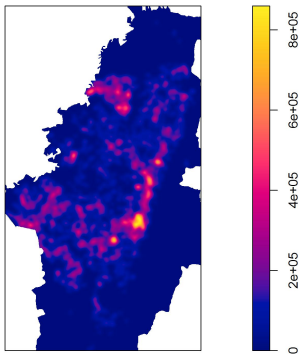
$$K(x, x_i, \Sigma) = \frac{1}{\sqrt{(2\pi)^3 \det(\Sigma)}} \exp \left[ -\frac{1}{2} (x - x_i)^t \Sigma^{-1} (x - x_i) \right] \quad (1)$$

- Y se suman sobre todo el mapa

$$u(x, \Sigma) = \frac{1}{n+1} \sum_{i=0}^n K(x, x_i, \Sigma) \quad (2)$$

- El parámetro  $\Sigma$  de mejor ajuste se seleccionó basado en el PAI sobre un conjunto de prueba.

Heat map for a 1-year period of crime



# Modelo Espacio - Temporal

- Este es un modelo basado en la clasificación de los eventos como antecedentes y réplicas.
- Es el estado del arte en modelos de predicción del crimen.

# Modelo Espacio - Temporal: Motivación

Mohler et al.: Self-Exciting Point Process Modeling of Crime

101

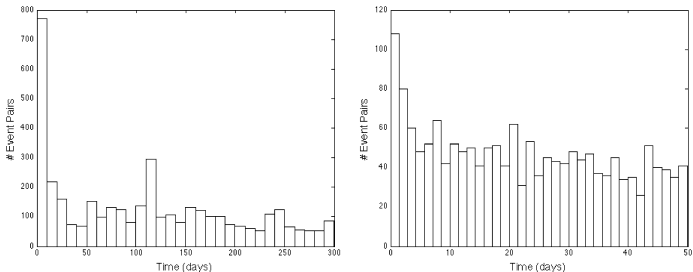


Figure 1. On the left, histogram of times (less than 300 days) between Southern California earthquake events of magnitude 3.0 or greater separated by 110 kilometers or less. On the right, histogram of times (less than 50 days) between burglary events separated by 200 meters or less.

## Modelo Espacio - Temporal: Motivación

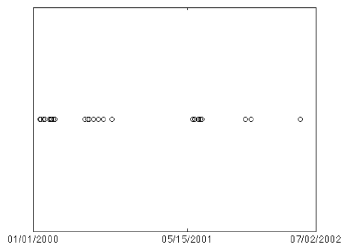


Figure 2. Times of violent crimes between two rivalry gangs in Los Angeles.



## Modelo Espacio - Temporal: especificación

- Se considera un modelo de la intensidad espacio temporal del crimen de la forma:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{k:t_k < t} g(t - t_k, x - x_k, y - y_k) \quad (3)$$

- Sea  $\mu(t, x, y) = \nu(t)\mu(x, y)$

## Modelo Espacio - Temporal: especificación

- Suponiendo que el modelo es correcto la probabilidad de que un evento  $i$  sea un evento antecedente es:

$$\pi_{ii} = \frac{\nu(t_i)\mu(x_i, y_i)}{\lambda(t_i, x_i, y_i)} \quad (4)$$

- La probabilidad de que el evento  $j$  cause el evento  $i$  (evento replica) es:

$$\pi_{ji} = \frac{g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)} \quad (5)$$

## Modelo Espacio - Temporal: estimación

- Sea  $P$  la matriz de probabilidades  $(p_{ji})$  (obsérvese que la suma de columnas da 1).
- Elija una probabilidad inicial  $P_0$ .
- Muestrear de  $P_0$  puntos de transfondo y réplica:  
 $(t_k, x_k, y_k, p_{kk})_{k=1, \dots, N}$  y  $(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})_{i > j}$  y usando esta muestra hacer una estimación inicial de:
  - 1  $\mu_1$  de  $\mu$ .
  - 2  $g_1$  de  $g$ .
  - 3  $\nu_1$  de  $\nu$usando KDE con ancho de banda variable.
- Estimar inicialmente  $\lambda_1$  usando que las columnas de  $P$  suman 1.

## Modelo Espacio - Temporal: estimación

- Fijemos  $n$  y consideremos los datos  $g_n(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})$  de  $g$ . Ahora ajustamos una un kernel del siguiente tipo:

$$g_n(t, x, y) = \frac{1}{N} \sum_{i=1}^{N_o} \frac{1}{\sigma_x \sigma_y \sigma_t (2\pi)^{(3/2)} D_i^3} \times \exp\left(-\frac{(x - x_i^o)^2}{2\sigma_x^2 D_i^2} - \frac{(y - y_i^o)^2}{2\sigma_y^2 D_i^2} - \frac{(t - t_i^o)^2}{2\sigma_t^2 D_i^2}\right).$$

- Usamos un procedimiento similar para estimar  $\nu$  y  $\mu$ . En el primer caso un Kernel Gaussiano univariado y el segundo caso un Kernel Gaussiano bivariado.

## Modelo Espacio - Temporal: estimación

- Actualizar  $P_1$ .
- Repetir los pasos anteriores hasta que la matriz  $P$  no cambie mucho.
- El número de probabilidades a estimar en cada iteración es del orden de  $N^2$ . Esto es computacionalmente costoso (típicamente  $N$  del orden de 1,000 o 30,000 - Bogotá en un año).

# Modelo Espacio - Temporal: Validación

Mohler et al.: Self-Exciting Point Process Modeling of Crime

103

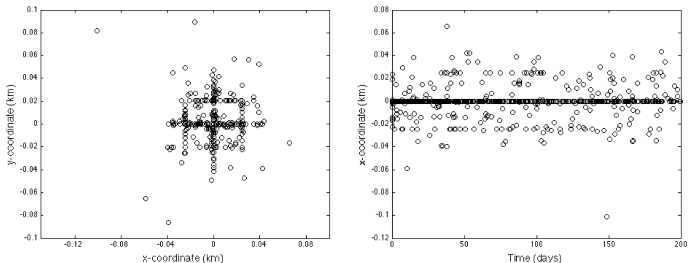


Figure 3. Spatial (left) and space-time (right) offspring/parent interpoint distances  $\{(t_i^o, x_i^o, y_i^o)\}_{i=1}^{N_o}$  sampled from  $P_{75}$ .

# Modelo Espacio - Temporal: Validación

Mohler et al.: Self-Exciting Point Process Modeling of Crime

105

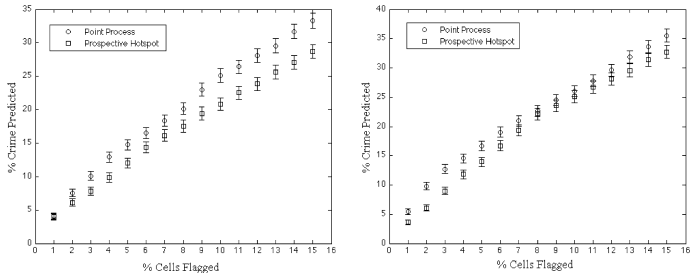


Figure 6. Forecasting strategy comparison. Average daily percentage of crimes predicted plotted against percentage of cells flagged for 2005 burglary using 200 m by 200 m cells. Error bars correspond to the standard error. Prospective hotspot cutoff parameters are 400 meters and 8 weeks (left) and optimal parameters (right) are 200 meters and 39 weeks. Spatial background intensity  $\mu(x, y)$  smoothing bandwidth for the point process is 300 meters (left) selected by cross validation and 130 meters (right) selected to optimize the number of crimes predicted.

## Validación

- Utilizamos el *Precision Accuracy Index*

$$PAI = \frac{\text{Hit Rate}}{\text{Percentage of Area}}$$
$$\text{Hit Rate} = \frac{\text{Crimes predicted in Hotspots}}{\text{Total Crimes in Hotspot}}$$
$$\text{Percentage Area} = \frac{\text{Area of Hotspots}}{\text{Total Area}}$$

- Sin embargo en modelos muy granulares no es una buena medida.



## PAI y modelo de puntos

- El modelo de puntos tiene hotspots muy reducidos y dispersos.
- Apenas el 10 % de las celdas tienen crímenes en el entrenamiento (menos en conjuntos de prueba).
- Esto hace que el PAI de 0 con frecuencia y valores muy alto en otros casos. Para este tipo de modelos es mejor utilizar el Hit Rate. Esto requiere el mismo porcentaje de área en los hotspots.

## PAI y modelo de puntos

- El modelo de puntos tiene hotspots muy reducidos y dispersos.
- Apenas el 10% de las celdas tienen crímenes en el entrenamiento (menos en conjuntos de prueba).
- Esto hace que el PAI de 0 con frecuencia y valores muy alto en otros casos. Para este tipo de modelos es mejor utilizar el Hit Rate. Esto requiere el mismo porcentaje de área en los hotspots.

## PAI y modelo de puntos

- El modelo de puntos tiene hotspots muy reducidos y dispersos.
- Apenas el 10 % de las celdas tienen crímenes en el entrenamiento (menos en conjuntos de prueba).
- Esto hace que el PAI de 0 con frecuencia y valores muy alto en otros casos. Para este tipo de modelos es mejor utilizar el Hit Rate. Esto requiere el mismo porcentaje de área en los hotspots.

- Se entrena un modelo agrupando todos los datos de una franja temporal dada para 2011.
- Se separan los datos en un conjunto de entrenamiento y uno de control.
- La validación se hace en cada muestra de la franja temporal correspondiente en 2012.
- Nuestra medida final es el promedio para 20 semanas escogidas aleatoriamente, esto resulta en el **valor esperado** del PAI.

- Se entrena un modelo agrupando todos los datos de una franja temporal dada para 2011.
- Se separan los datos en un conjunto de entrenamiento y uno de control.
- La validación se hace en cada muestra de la franja temporal correspondiente en 2012.
- Nuestra medida final es el promedio para 20 semanas escogidas aleatoriamente, esto resulta en el **valor esperado** del PAI.

- Se entrena un modelo agrupando todos los datos de una franja temporal dada para 2011.
- Se separan los datos en un conjunto de entrenamiento y uno de control.
- La validación se hace en cada muestra de la franja temporal correspondiente en 2012.
- Nuestra medida final es el promedio para 20 semanas escogidas aleatoriamente, esto resulta en el **valor esperado** del PAI.

- Se entrena un modelo agrupando todos los datos de una franja temporal dada para 2011.
- Se separan los datos en un conjunto de entrenamiento y uno de control.
- La validación se hace en cada muestra de la franja temporal correspondiente en 2012.
- Nuestra medida final es el promedio para 20 semanas escogidas aleatoriamente, esto resulta en el **valor esperado** del PAI.

# Resultados

	Ellipses	kde1	kde2	Spatiotemp
1	3.76	8.10	3.18	7.34
2	6.65	6.64	4.76	6.64
3	3.70	6.95	2.14	7.44
4	3.24	5.91	2.05	7.12
5	3.98	6.02	3.88	8.05
6	3.67	6.30	3.02	7.32
7	2.93	7.10	3.56	8.21
8	2.46	5.77	2.12	7.22
9	3.02	6.07	2.54	7.18
10	3.56	7.57	3.22	8.10
11	5.24	7.82	4.13	7.34
12	3.45	7.38	2.16	8.24
13	4.55	6.03	3.94	7.13
14	3.23	6.86	2.04	7.58
15	3.68	5.85	3.64	8.02
16	4.80	5.68	3.55	7.14
17	3.24	5.28	2.18	7.16
18	2.55	5.08	2.07	6.51
19	3.99	5.94	3.68	7.38
20	5.92	6.09	4.66	8.15
Average	3.8810	6.4220	3.1260	7.4635



# Resultados

Week	Hit Rate	
	Points	KDE
1	0.09	0.41
2	0.11	0.74
3	0.12	0.77
4	0.10	0.66
5	0.10	0.67
6	0.11	0.70
7	0.08	0.36
8	0.10	0.64
9	0.10	0.67
10	0.08	0.38
11	0.09	0.39
12	0.08	0.37
13	0.10	0.67
14	0.11	0.76
15	0.10	0.65
16	0.09	0.63
17	0.09	0.59
18	0.08	0.56
19	0.10	0.66
20	0.10	0.68
Avg.	0.09	0.59

- El ejercicio de validación sugiere que el modelo espacio temporal es el modelo más apropiado.
- Es posible hacer varias extensiones que prometen mejorar su capacidad predictiva.

- El ejercicio de validación sugiere que el modelo espacio temporal es el modelo más apropiado.
- Es posible hacer varias extensiones que prometen mejorar su capacidad predictiva.

Motivación  
**Modelo preliminares**  
Modelo Espacio - Temporal: Ancho de banda variable  
Modelos Avanzados

Modelo de puntos  
Modelo de elipses espaciales  
Modelo KDE  
Modelo Espacio - Temporal  
Validación  
**Aplicativo**

Motivación  
Modelo preliminares  
Modelo Espacio - Temporal: Ancho de banda variable  
Modelos Avanzados

Modelo de puntos  
Modelo de elipses espaciales  
Modelo KDE  
Modelo Espacio - Temporal  
Validación  
Aplicativo

~/R/Code/codigo\_crimen/App - Shiny

http://127.0.0.1:5084 | Open in Browser | Publish

### Predicción de crimen

quartil matemáticas aplicadas

Horas: 10

Día:  
 lunes  
 martes  
 miércoles  
 jueves  
 viernes  
 sábado  
 domingo

Código del cuadrante:  
E01-13

Ejecutar

Mapa de calor Patillaje

Powered by Leaflet — © OpenStreetMap contributors, CC-BY-SA

Motivación  
**Modelo preliminares**  
Modelo Espacio - Temporal: Ancho de banda variable  
Modelos Avanzados

Modelo de puntos  
Modelo de elipses espaciales  
Modelo KDE  
Modelo Espacio - Temporal  
Validación  
**Aplicativo**

Motivación  
Modelo preliminares  
Modelo Espacio - Temporal: Ancho de banda variable  
Modelos Avanzados

Modelo de puntos  
Modelo de elipses espaciales  
Modelo KDE  
Modelo Espacio - Temporal  
Validación  
Aplicativo

~/R/Code/codigo\_crimen/App - Shiny

http://127.0.0.1:5084 | Open in Browser | Publish

### Prediccion de crimen

quantil matemáticas aplicadas

**Horas:** 1 10 24

**Día:**  
 lunes  
 martes  
 miercoles  
 jueves  
 viernes  
 sabado  
 domingo

**Código del cuadrante:**  
E01-12

Ejecutar

Mapa de calor Patillaje

Powered by Leaflet — © OpenStreetMap contributors, CC-BY-SA

# Contenido

- 1 Motivación
- 2 Modelo preliminares
  - Modelo de puntos
  - Modelo de elipses espaciales
  - Modelo KDE
  - Modelo Espacio - Temporal
  - Validación
  - Aplicativo
- 3 Modelo Espacio - Temporal: Ancho de banda variable
  - Validación
- 4 Modelos Avanzados
  - ¿Por qué molestarse?
  - Trabajos previos
  - Deep Learning



## Objetivo

Evaluar la capacidad predictiva del modelo de predicción de crimen propuesto.

En particular, se compara el poder predictivo del modelo al ser entrenado con diferentes horizontes de tiempo.

## Datos

Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.

Para esta validación se usaron datos de la localidad de Santa Fé, en donde se aplicará el piloto, delimitada por latitudes entre  $[4,571, 4,629]$  y longitudes  $> -74,091$ .

## Datos

Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.

Para esta validación se usaron datos de la localidad de Santa Fé, en donde se aplicará el piloto, delimitada por latitudes entre  $[4,571, 4,629]$  y longitudes  $> -74,091$ .

## Datos

1.676 ( $\approx 10\%$ ) crímenes en la localidad de Santa Fé en el período tratado.

Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

## Datos

1.676 ( $\approx 10\%$ ) crímenes en la localidad de Santa Fé en el período tratado.

Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

# Validación

- 1 Se divide Bogotá ( $1,547\text{km}^2$ ) en 10,946 celdas de  $\approx 145\text{m}^2$  cada una; 1,019 celdas en la localidad de Santa Fé.
- 2 Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10 % de las celdas con mayor probabilidad de crimen.
- 3 Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

## Validación

- 1 Se divide Bogotá ( $1,547\text{km}^2$ ) en 10,946 celdas de  $\approx 145\text{m}^2$  cada una; 1,019 celdas en la localidad de Santa Fé.
- 2 Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- 3 Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

## Validación

- 1 Se divide Bogotá ( $1,547\text{km}^2$ ) en 10,946 celdas de  $\approx 145\text{m}^2$  cada una; 1,019 celdas en la localidad de Santa Fé.
- 2 Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- 3 Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.



## Validación

Para medir la predictabilidad del modelo se usa la medida de *Hit Rate*, la cual mide la proporción de los crímenes totales de un período que ocurrieron en los puntos calientes predichos por el modelo de crimen:

$$\text{Hit Rate} = \frac{\text{Crímenes en los hotspots predichos}}{\text{Crímenes totales}}$$

## Validación

Para medir la predictabilidad del modelo se usa la medida de *Hit Rate*, la cual mide la proporción de los crímenes totales de un período que ocurrieron en los puntos calientes predichos por el modelo de crimen:

$$\text{Hit Rate} = \frac{\text{Crímenes en los hotspots predichos}}{\text{Crímenes totales}}$$

## Validación

La siguiente tabla registra el Hit Rate del modelo al ser entrenado con datos entre 1 y 7 semanas. El Hit Rate se calculó para cada uno de los 21 turnos de 8 horas dentro de cada semana.

Entrenamiento	1	2	3	4	5	6	7
Semana 1	0,49	0,47	0,44	0,44	0,46	0,44	0,44
Semana 2	0,41	0,40	0,40	0,43	0,43	0,46	0,46
Semana 3	0,39	0,48	0,44	0,49	0,48	0,54	0,54
Promedio	0,43	0,45	0,43	0,45	0,45	0,48	0,48

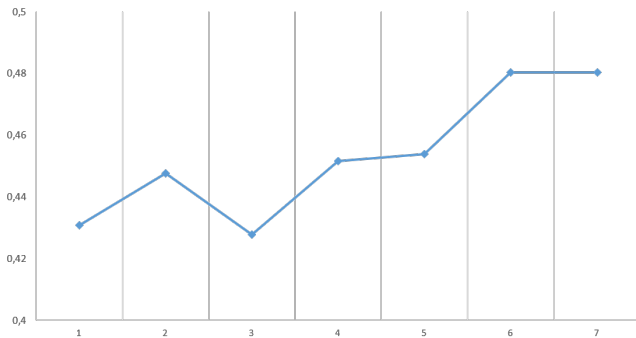
## Validación

La siguiente tabla registra el Hit Rate del modelo al ser entrenado con datos entre 1 y 7 semanas. El Hit Rate se calculó para cada uno de los 21 turnos de 8 horas dentro de cada semana.

Entrenamiento	1	2	3	4	5	6	7
Semana 1	0,49	0,47	0,44	0,44	0,46	0,44	0,44
Semana 2	0,41	0,40	0,40	0,43	0,43	0,46	0,46
Semana 3	0,39	0,48	0,44	0,49	0,48	0,54	0,54
Promedio	0,43	0,45	0,43	0,45	0,45	0,48	0,48

# Validación

HIT RATE PROMEDIO POR SEMANAS DE ENTRENAMIENTO



# Validación

Adicionalmente, para medir la eficiencia del modelo se usa la medida de *Predictive Accuracy Index*

$$PAI = \frac{\text{Hit Rate}}{\% \text{ Área de los hotspots}}$$

# Validación

Adicionalmente, para medir la eficiencia del modelo se usa la medida de *Predictive Accuracy Index*

$$PAI = \frac{Hit\ Rate}{\% \text{ \u00c1rea de los hotspots}}$$

## Validación

Al evaluar la eficiencia del modelo entrenado con 7 semanas, se encontró:

Cobertura	1 %	5 %	10 %	15 %	20 %
Semana 1	1,53	5,04	4,4	4,21	3,92
Semana 2	4,33	6,14	4,62	3,98	3,64
Semana 3	10,4	6,92	5,43	4,37	3,7
Promedio	5,33	6,02	4,8	4,19	3,75



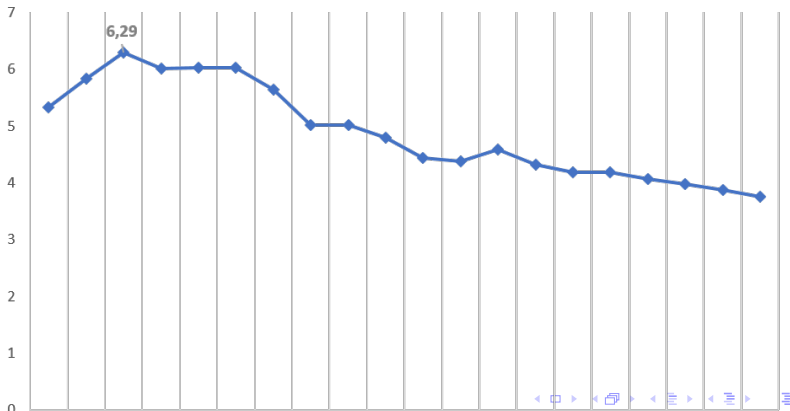
## Validación

Al evaluar la eficiencia del modelo entrenado con 7 semanas, se encontró:

Cobertura	1 %	5 %	10 %	15 %	20 %
Semana 1	1,53	5,04	4,4	4,21	3,92
Semana 2	4,33	6,14	4,62	3,98	3,64
Semana 3	10,4	6,92	5,43	4,37	3,7
Promedio	5,33	6,02	4,8	4,19	3,75

# Validación

## PAI PROMEDIO SEGÚN PORCENTAJE DE COBERTURA



# Validación

Se puede apreciar que entre mayor sea el número de datos de entrenamiento, el modelo tiene mayor capacidad predictiva.

Este método de validación se puede utilizar también para calibrar otros parámetros del modelo: tipo de kernel y anchos de banda, pesos otorgados a los datos de entrenamiento, entre otros.

# Validación

Se puede apreciar que entre mayor sea el número de datos de entrenamiento, el modelo tiene mayor capacidad predictiva.

Este método de validación se puede utilizar también para calibrar otros parámetros del modelo: tipo de kernel y anchos de banda, pesos otorgados a los datos de entrenamiento, entre otros.

## Validación

Para estimar la intensidad del modelo se muestrean crímenes según los datos de entrenamiento y se clasifican entre crímenes de trasfondo y réplicas. Con esta muestra se evalúa la precisión del modelo y se corrigen los errores generando una nueva estimación de la intensidad.

Se repite este proceso iterativamente hasta que los errores de la intensidad estimada según los datos de entrenamiento sean menores a 0,01.

## Validación

Para estimar la intensidad del modelo se muestrean crímenes según los datos de entrenamiento y se clasifican entre crímenes de trasfondo y réplicas. Con esta muestra se evalúa la precisión del modelo y se corrigen los errores generando una nueva estimación de la intensidad.

Se repite este proceso iterativamente hasta que los errores de la intensidad estimada según los datos de entrenamiento sean menores a 0,01.

## Validación

Como el número de crímenes de trasfondo y réplicas varía en cada iteración, un ancho de banda fijo puede suavizar demasiado o muy poco la intensidad, generando errores en la predicción y demoras en la convergencia de la función.

Particularmente, el ancho de banda fijo para la estimación de la función de desencadenamiento (réplicas) causaba que el modelo no caputara acertadamente esta dimensión por lo que presentaba un desempeño bajo en un horizonte de corto plazo.

## Validación

Como el número de crímenes de trasfondo y réplicas varía en cada iteración, un ancho de banda fijo puede suavizar demasiado o muy poco la intensidad, generando errores en la predicción y demoras en la convergencia de la función.

Particularmente, el ancho de banda fijo para la estimación de la función de desencadenamiento (réplicas) causaba que el modelo no caputara acertadamente esta dimensión por lo que presentaba un desempeño bajo en un horizonte de corto plazo.



# Validación

Se usa un ancho de banda variable para cada iteración que depende del número de crímenes muestreados y su concentración espacio-temporal.

Los parámetros adicionales del modelo se escogieron usando el algoritmo de esperanza-maximización.

# Validación

Se usa un ancho de banda variable para cada iteración que depende del número de crímenes muestreados y su concentración espacio-temporal.

Los parámetros adicionales del modelo se escogieron usando el algoritmo de esperanza-maximización.

## Validación

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable
Semana 1	0,44	0,57
Semana 2	0,46	0,59
Semana 3	0,54	0,62
Promedio	0,48	0,59

## Validación

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable
Semana 1	0,44	0,57
Semana 2	0,46	0,59
Semana 3	0,54	0,62
Promedio	0,48	0,59

## Validación

Otro modelo ampliamente aceptado en la literatura y con buenos resultados empíricos es el de *Kernel Density Estimation*, en el cual se estima la intensidad de crimen como:

$$\lambda(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right),$$

con  $x$  las coordenadas espaciales de los crímenes.

## Validación

Otro modelo ampliamente aceptado en la literatura y con buenos resultados empíricos es el de *Kernel Density Estimation*, en el cual se estima la intensidad de crimen como:

$$\lambda(x) = \frac{1}{Nh} \sum_{i=1}^N K \left( \frac{x - x_i}{h} \right),$$

con  $x$  las coordenadas espaciales de los crímenes.

## Validación

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable	KDE
Semana 1	0,44	0,57	0,42
Semana 2	0,46	0,59	0,44
Semana 3	0,54	0,62	0,53
Promedio	0,48	0,59	0,46

## Validación

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Predicción	bw fijo	bw variable	KDE
Semana 1	0,44	0,57	0,42
Semana 2	0,46	0,59	0,44
Semana 3	0,54	0,62	0,53
Promedio	0,48	0,59	0,46



## Validación

El *Wilcoxon signed-rank test* es una prueba no paramétrica para comparar el rango de dos muestras relacionadas y determinar si estas son estadísticamente diferentes. Se quiere determinar si los modelos probados son significativamente diferentes entre sí:

Modelo	p-value
bw fijo vs. KDE	0,061
bw fijo vs. bw variable	0,030
bw variable vs. KDE	0,016

## Validación

El *Wilcoxon signed-rank test* es una prueba no paramétrica para comparar el rango de dos muestras relacionadas y determinar si estas son estadísticamente diferentes. Se quiere determinar si los modelos probados son significativamente diferentes entre sí:

Modelo	p-value
bw fijo vs. KDE	0,061
bw fijo vs. bw variable	0,030
bw variable vs. KDE	0,016

## Conclusiones

El modelo de predicción con ancho de banda variable tiene el mayor poder de predicción de los tres modelos estimados. Esto sustenta dos de las motivaciones iniciales del proyecto:

- Existen patrones temporales en la ocurrencia de crímenes.
- El crimen se propaga espacio-temporalmente.

## Conclusiones

El modelo de predicción con ancho de banda variable tiene el mayor poder de predicción de los tres modelos estimados. Esto sustenta dos de las motivaciones iniciales del proyecto:

- Existen patrones temporales en la ocurrencia de crímenes.
- El crimen se propaga espacio-temporalmente.

## Conclusiones

El modelo de predicción con ancho de banda variable tiene el mayor poder de predicción de los tres modelos estimados. Esto sustenta dos de las motivaciones iniciales del proyecto:

- Existen patrones temporales en la ocurrencia de crímenes.
- El crimen se propaga espacio-temporalmente.

# Contenido

- 1 Motivación
- 2 Modelo preliminares
  - Modelo de puntos
  - Modelo de elipses espaciales
  - Modelo KDE
  - Modelo Espacio - Temporal
  - Validación
  - Aplicativo
- 3 Modelo Espacio - Temporal: Ancho de banda variable
  - Validación
- 4 Modelos Avanzados
  - ¿Por qué molestarse?
  - Trabajos previos
  - Deep Learning

## ¿Por qué molestarse?

Reducir la criminalidad es una de las prioridades de las políticas públicas dado sus múltiples beneficios tanto sociales como económicos.

Entre abril y junio de 2017 se reportaron 16.609 crímenes en Bogotá.

## Trabajos previos

En trabajos anteriores se ha examinado la relación entre la desigualdad (Kelly, 2000), desempleo (Hojman, 2004), privaciones económicas, densidad de población, crecimiento económico, pobreza y capital humano (Poveda, 2012), PIB, PIB per cápita, población masculina, analfabetismo (Alves et al., 2013), redes sociales (Gerber, 2014), entre otras covariables, y la ocurrencia de crímenes



## Trabajos Previos

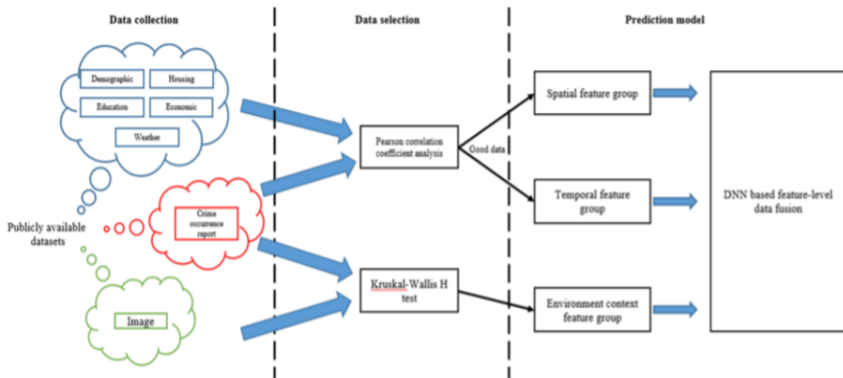
También se estudian patrones espacio-temporales de los crímenes:

- Concentración de criminalidad entre 00:00 y 3:59 cerca a lugares de entretenimiento (Cusimano et al., 2010).
- Teorías de propagación de crimen desde la sismiólogía (Mohler, 2011) o desde la epidemiología (Akers & Lanier, 2009).
- Efecto del clima en la criminalidad: hay más crímenes en verano.

## Trabajos Previos

Se ha abordado el problema de predecir la ocurrencia de crimen con diferentes metodologías entre las que se destacan análisis de regresión, estimación de densidad de kernel (KDE), y máquinas de soporte vectorial (SVM).

# Metodología



# Motivación

La mayoría de los anteriores trabajos tratan los datos de diferentes datos de la misma manera, concatenándola directamente. Sin embargo, esta metodología presenta varias limitaciones como *over-fitting*, relaciones no-lineales entre las variables, redundancia, y dependencia entre las diferentes bases de datos, que son superadas usando Redes Neuronales Profundas (DNN).

# Motivación

Relación entre la apariencia de un lugar y su índice de criminalidad:

- Teoría de la ventana rota (BWT): Un lugar "desordenado" tiene potencialmente una mayor probabilidad de ocurrencia de crimen.
- Prevención de crimen a partir del diseño ambiental.

# Motivación

Relación entre la apariencia de un lugar y su índice de criminalidad:

- Teoría de la ventana rota (BWT): Un lugar "desordenado" tiene potencialmente una mayor probabilidad de ocurrencia de crimen.
- Prevención de crimen a partir del diseño ambiental.

# Motivación

Relación entre la apariencia de un lugar y su índice de criminalidad:

- Teoría de la ventana rota (BWT): Un lugar "desordenado" tiene potencialmente una mayor probabilidad de ocurrencia de crimen.
- Prevención de crimen a partir del diseño ambiental.

## Datos

Se usan datos de múltiples dominios para la ciudad de Chicago.

- Reporte de ocurrencias de crimen (274.064 casos en 2014).
- Factores socioeconómicos.
- Información climática.
- Imágenes de *Google Street Viewer*.



# Datos

Base de datos:

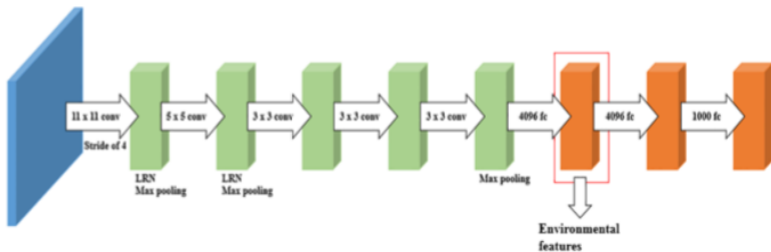
Domain	Collection	Collected data	Scale
Demographic	American FactFinder	- 10 data related to the population. - 8 data related to the race.	Census tract
Housing	American FactFinder	- 6 data related to housing.	Census tract
Education	American FactFinder	- 12 data related to education.	Census tract
Economic	American FactFinder	- 14 data related to employment. - 26 data related to industry.	Census tract
Image	Google Street View	- 60,348 images (640 x 420).	Coordinates
Weather	Weather Underground	- 20 data related to the weather. - 6 data related to weather events	Date

# Datos

Como las imagenes son datos no estructurados se extra su información usando una red convolucional ya entrenada para clasificación de imagenes conocida como *Alexnet*.

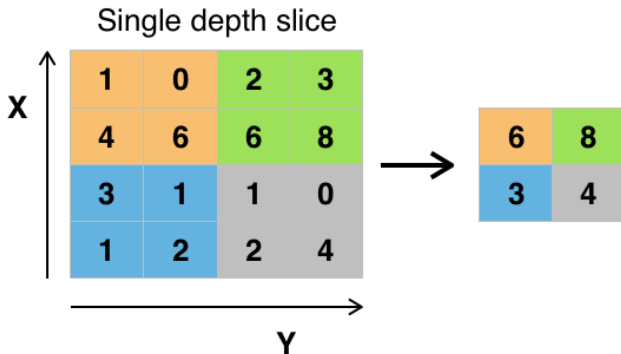
# Datos

Alexnet:



# Datos

Max Pooling:



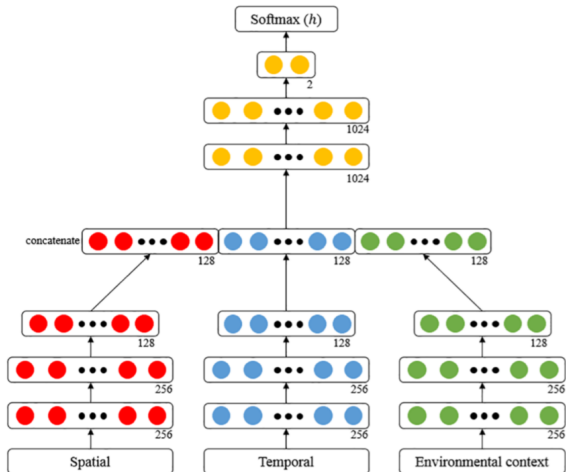
## Selección de datos

- A partir de un análisis de correlación de Pearson reducen la base de datos a 53 de las 102 variables (sin tener en cuenta las imágenes) recogidas inicialmente.
- Para los datos de apariencia del lugar se hizo un *Kruskal-Wallis H test* agrupando las imagenes por k-medias, obteniendo un  $p\text{-valor} < 0,001$ .

# Entrenamiento

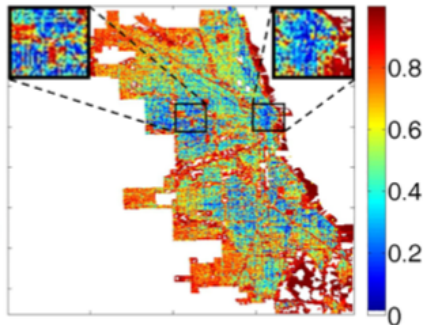
Para el modelo predictivo se configuró una DNN con cuatro capas: espacial (35 variables), temporal (15 variables), apariencia del lugar (4096 variables), y características conjuntas.

# Entrenamiento



## Resultados

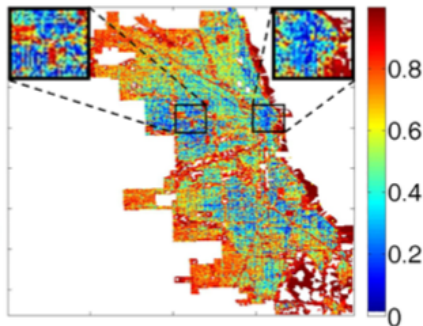
El resultado de el modelo entrenado es la probabilidad de que ocurra un crimen en un lugar.





## Resultados

El resultado de el modelo entrenado es la probabilidad de que ocurra un crimen en un lugar.



## Resultados

Se compara el poder predictivo del modelo propuesto con modelos estándar de aprendizaje de máquinas obteniendo los siguientes resultados:

Modelo	Accuracy (%)	Precision (%)	Recall (%)	AUC
SVM	67.01	98.58	1.05	0.5052
KDE	66.33	43.04	85.79	0.7258
DNN	84.25	74.35	80.55	0.8333

## Resultados

Se compara el poder predictivo del modelo propuesto con modelos estándar de aprendizaje de máquinas obteniendo los siguientes resultados:

Modelo	Accuracy (%)	Precision (%)	Recall (%)	AUC
SVM	67.01	98.58	1.05	0.5052
KDE	66.33	43.04	85.79	0.7258
DNN	84.25	74.35	80.55	0.8333

Como sustento de la metodología, se prueba el modelo sin selección de variables y sin tener en cuenta la apariencia del lugar, obteniendo los siguientes resultados:

Modelo	Accuracy (%)	Precision(%)	Recall(%)	AUC
DNN	84.25	74.35	80.55	0.833
Sin selección	77.98	70.69	57.96	0.7297
Sin apariencia	72.38	59.96	51.58	0.6718

Como sustento de la metodología, se prueba el modelo sin selección de variables y sin tener en cuenta la apariencia del lugar, obteniendo los siguientes resultados:

Modelo	Accuracy (%)	Precission(%)	Recall(%)	AUC
DNN	84.25	74.35	80.55	0.833
Sin selección	77.98	70.69	57.96	0.7297
Sin apariencia	72.38	59.96	51.58	0.6718