

# Predicción de delitos mediante un enfoque espaciotemporal

Rafael Zambrano

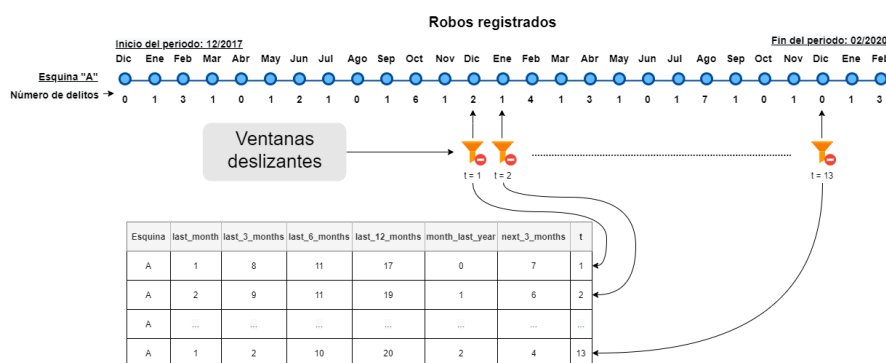
**Abstract** En esta presentación se propone mostrar un caso uso del ecosistema {tidymodels} referente a la predicción de delitos, mediante una propuesta metodológica espaciotemporal. Se implementan modelos predictivos que permiten relacionar intersecciones de calles, elementos del entorno delictivo y el momento de ocurrencia del delito, tomando como base los robos mensuales registrados entre los años 2017 y 2020 en la Ciudad Autónoma de Buenos Aires (CABA). Se compara el nivel predictivo de modelos que contemplan únicamente los delitos registrados en cada esquina, con modelos que adicionalmente capturan factores espaciales referentes al entorno delictivo de las esquinas.

**Palabras clave:** espaciotemporal - delitos - tidymodels - espacial - series de tiempo

## Metodología

Las especificidades técnicas del caso de uso y la metodología implementada pueden encontrarse en (Zambrano 2021), a continuación, se presenta un resumen de la propuesta metodológica espaciotemporal, partiendo de la problemática referente a la no existencia de un conjunto de datos preconstruido con la estructura de modelado requerida. Por ello, como punto de partida se describe el proceso de captura, ingesta, limpieza y consolidación de datos. Las unidades espaciales de conteo de delitos son las esquinas de la ciudad con un radio de metros determinado, para obtener las esquinas se procede a interceptar las calles y avenidas con el objetivo de obtener la geolocalización de las intersecciones, los objetos espaciales se obtienen mediante {osmdata}, (Padgham et al. 2017). Posteriormente, se implementa un *filtro de proximidad*<sup>1</sup> para eliminar aquellas esquinas que están muy cercanas entre sí según una distancia definida.

Definidas las esquinas a considerar, se procede a realizar el conteo de robos ocurridos en las cercanías de las esquinas para cada mes del periodo de estudio. Seguidamente, se busca recrear el entorno delictivo, para esto se agregan elementos espaciales presentes en las cercanías de las esquinas, por ejemplo, el número de comisarias, hoteles, paradas de autobús, entre otros. Todas las fuentes de datos son de acceso público, almacenadas en el repositorio del Gobierno de la Ciudad Autónoma de Buenos Aires (Buenos Aires Data, s.f.)<sup>2</sup>. Finalmente, son indexadas variables referentes al registro histórico de delitos de las esquinas, pero aplicando una *técnica de ventanas deslizantes* disponible en {sknifedatar} (Zambrano and Bartolome 2021) para capturar la incidencia delictiva con rezagos en distintos intervalos de tiempo, a continuación, se muestra un esquema del funcionamiento de esta técnica.

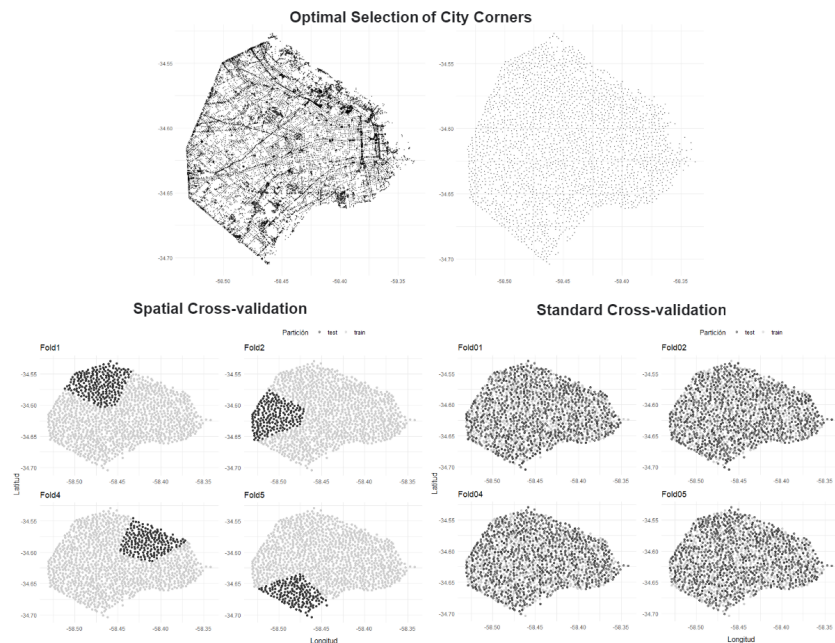


Al momento de implementar diversos modelos predictivos, se busca predecir el número de delitos que ocurran en el último trimestre del periodo de estudio, utilizando información espacial y temporal. Se implementa un enfoque de series de tiempo, mediante los modelos ARIMA y Croston, también se aplica un enfoque de corte transversal utilizando modelos XGBoost, para estos últimos se aplican técnicas de validación cruzada para el ajuste hiperparámetros y evaluación de las métricas de los modelos. Debido a la posible presencia de autocorrelación espacial, se compara la validación cruzada estándar con la aplicación de una *variante adaptada para datos espaciales*<sup>3</sup>, utilizando un modelo de k-means que genera las particiones en los datos como se muestra a continuación.

<sup>1</sup>El algoritmo para la implementación del filtro de proximidad está disponible en {rangeBuilder}

<sup>2</sup>Buenos Aires Data. (s.f.). <https://data.buenosaires.gob.ar/>

<sup>3</sup>Mediante el paquete {spatialsample} fue posible aplicar esta variación, (Silge 2021).



Las principales bibliotecas utilizadas del ecosistema {tidymodels} fueron {recipes}, {workflows}, {rsample}, {parsnip}, {tune}, además de extensiones especializadas como {sknifedatar}, {spatialsample} y {modeltime}.

Resultados

La captura de los factores espaciales en conjunto con la metodología de ventanas deslizantes, generaron mejores predicciones en todas las métricas de evaluación consideradas en comparación con los modelos de series temporales. Sin embargo, al comparar los modelos de aprendizaje automático con las 2 tipos de validación cruzada, se observaron diferencias que pueden responder a la presencia de factores latente de autocorrelación espacial, en la siguiente tabla se puede observar el comportamiento de las métricas de evaluación.

Validación	mae	sd mae	rmse	sd rmse	rmsle	sd rmsle
Particiones estándar						
CV estándar	2.479	0.030	3.654	0.059	0.387	0.008
Conjunto de validación	2.761		3.899		0.442	
Particiones espaciales						
CV espacial	3.114	0.325	4.414	0.530	0.470	0.045
Conjunto de validación	2.800		3.922		0.473	

Referencias

10 Padgham, Mark, Bob Rudis, Robin Lovelace, and Maëlle Salmon. 2017. "Osmdata." *The Journal of Open Source Software* 2 (14). <https://doi.org/10.21105/joss.00305>.

Silge, Julia. 2021. *Spatialsample: Spatial Resampling Infrastructure*. <https://CRAN.R-project.org/package=spatialsample>.

Zambrano, Rafael. 2021. "Un enfoque espaciotemporal para la predicción de delitos en la ciudad de Buenos Aires." *Revista de investigación en modelos matematicos aplicados a la gestion y la economia* II (November): 38–62. <http://www.economicas.uba.ar/wp-content/uploads/2016/04/Zambrano-Rafael-1.pdf>.

Zambrano, Rafael, and Karina Bartolome. 2021. *Sknifedatar: Swiss Knife of Data*. <https://github.com/rafzamb/sknifedatar>.