

Modelo Predictivo de Frecuencia de Siniestros por Agua

Desarrollo de un modelo para estimar la frecuencia de siniestros por agua, optimizando la gestión de riesgos y la tarificación.



by Aleida Aponte



Objetivo y Enfoque Metodológico

Objetivo

Estimar la frecuencia de siniestros por agua con variables estructurales, climáticas y demográficas.

Enfoque

Modelo Poisson con offset y modelo Lasso para selección de variables y regularización.

El conjunto de datos incluye 87,862 observaciones con más de 50 variables predictoras disponibles.

Modelo Poisson: Variables Significativas

stro_1a_AGUAACAGBC

Historial previo de siniestros.

Exp(coef): 1.112. Mayor historial, mayor riesgo.

anualidad_seguro

Gasto anual en seguro.

Exp(coef): 0.973. Mayor gasto, menor riesgo.

antigedif

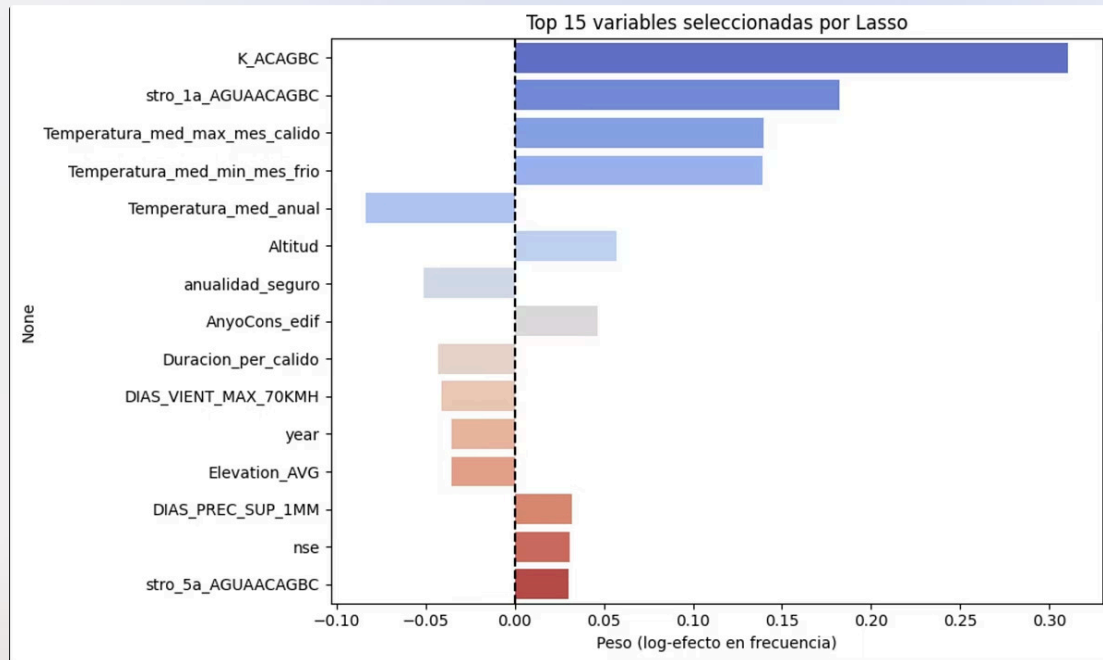
Antigüedad del edificio. Exp(coef): 0.983. Edificios más antiguos, menor riesgo.

Pseudo R²: 0.9999. Deviance alta indica posible sobre-dispersión.

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	resp	No. Observations:	87862			
Model:	GLM	Df Residuals:	87854			
Model Family:	Poisson	Df Model:	7			
Link Function:	Log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-3.8052e+06			
Date:	Sun, 13 Apr 2025	Deviance:	7.5133e+06			
Time:	15:40:37	Pearson chi2:	1.84e+09			
No. Iterations:	13	Pseudo R-squ. (CS):	0.9999			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	3.6559	0.007	490.298	0.000	3.641	3.670
Altitud	0.0005	4.06e-06	123.056	0.000	0.000	0.001
antigedif	-0.0168	7.7e-05	-217.878	0.000	-0.017	-0.017
antigref	-0.0006	8.32e-05	-6.693	0.000	-0.001	-0.000
ANTIGUEDAD_VIVIENDA	-0.0006	8.32e-05	-6.693	0.000	-0.001	-0.000
anualidad_reemplazo	-0.0023	5.35e-05	-43.619	0.000	-0.002	-0.002
anualidad_seguro	-0.0267	0.000	-68.950	0.000	-0.027	-0.026
stro_1a_AGUAACAGBC	0.1061	9.13e-05	1161.988	0.000	0.106	0.106
CUPD_CAP_Corr_aguaacagbc	0.0010	1.57e-06	616.619	0.000	0.001	0.001
=====						

Modelo Lasso: Variables Seleccionadas



1

K_ACAGBC

Factor correctivo clave. Peso log: +0.31.

2

Temperatura_med_max_mes_calido

Climas extremos. Peso log: +0.14. Mayor riesgo.

3

anualidad_seguro

Gasto en seguro. Peso log: -0.05. Reduce frecuencia.

Lasso eliminó más de 40 variables, previniendo sobreajuste y favoreciendo la interpretabilidad.

Poisson vs. Lasso: Comparación

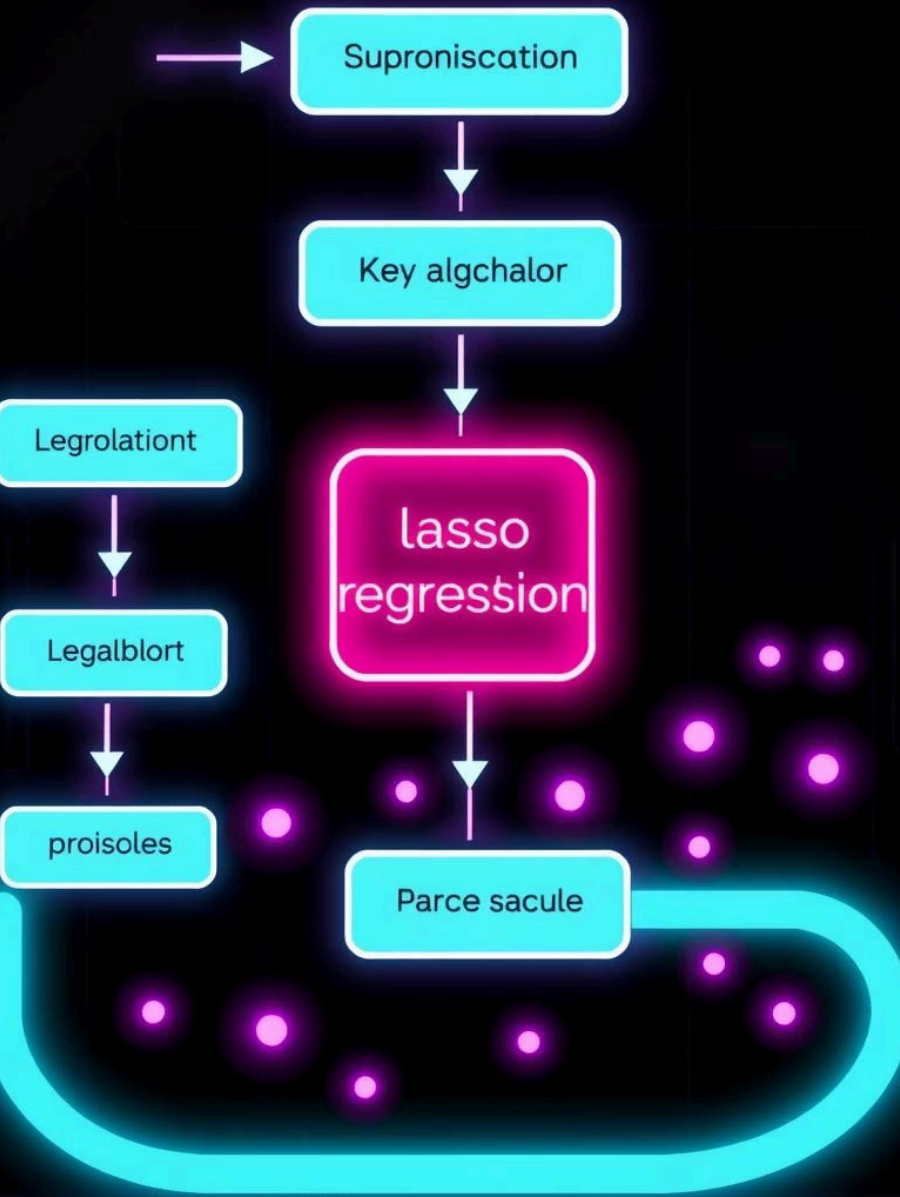
Poisson GLM

- Selección manual de variables
- $\exp(\text{coef})$ = cambio multiplicativo
- Riesgo de sobreajuste alto

Lasso (L1)

- Selección automática
- Coeficiente logarítmico
- Riesgo de sobreajuste bajo

Variables coincidentes: `stro_1a_AGUACAGBC`, `Altitud`, `anualidad_seguro`.



Ventajas del Modelo Lasso



Análisis Simplificado

Selección automática de variables predictivas.



Prevención de Sobreajuste

Reduce la dimensionalidad, eliminando ruido.



Identificación de Factores Clave

Enfoque en los predictores más relevantes.

Modelo Lasso aporta parsimonia y ayuda a reducir el modelo, simplificando el análisis de riesgos.

Recomendaciones



Priorizar Monitoreo

stro_1a_AGUACAGBC, K_ACAGBC, anualidad_seguro.



Modelo Negativo Binomial

Capturar sobre-dispersión.



Estrategias Preventivas

Antigüedad del edificio y clima.

Optimizar la gestión de riesgos basándose en los factores más influyentes identificados por los modelos.

