

Inspira Crea Transforma

Predicción de eventos de poda en redes de distribución de energía

Diana Lisette Arango Cañas

Andrés Alonso Ocampo Dávila

Juan Felipe Renza Chavarría

Daniel Alonso Sánchez Echeverri

Proyecto Integrador 2

Maestría en Ciencias de los Datos y Analítica

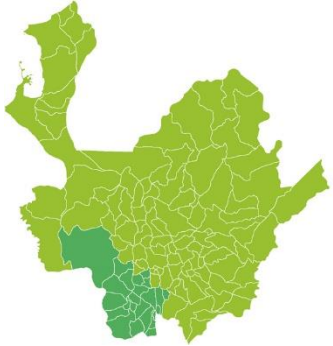
Junio de 2023

Contexto – EPM

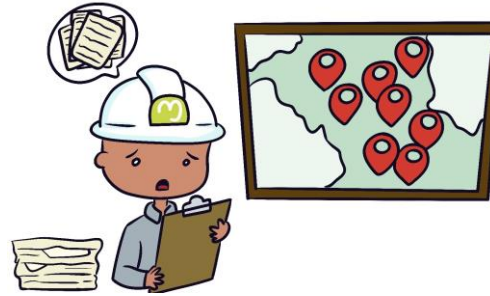
Empresa industrial y comercial del Estado, de propiedad del Municipio de Medellín, que presta los servicios de: generación, transmisión y **distribución de energía eléctrica**, gas por red, agua y saneamiento.



Contexto – Trabajos de poda



Se busca desarrollar un modelo basado en series de tiempo que permita estimar la cantidad de eventos de poda en una región para una ventana de tiempo determinada

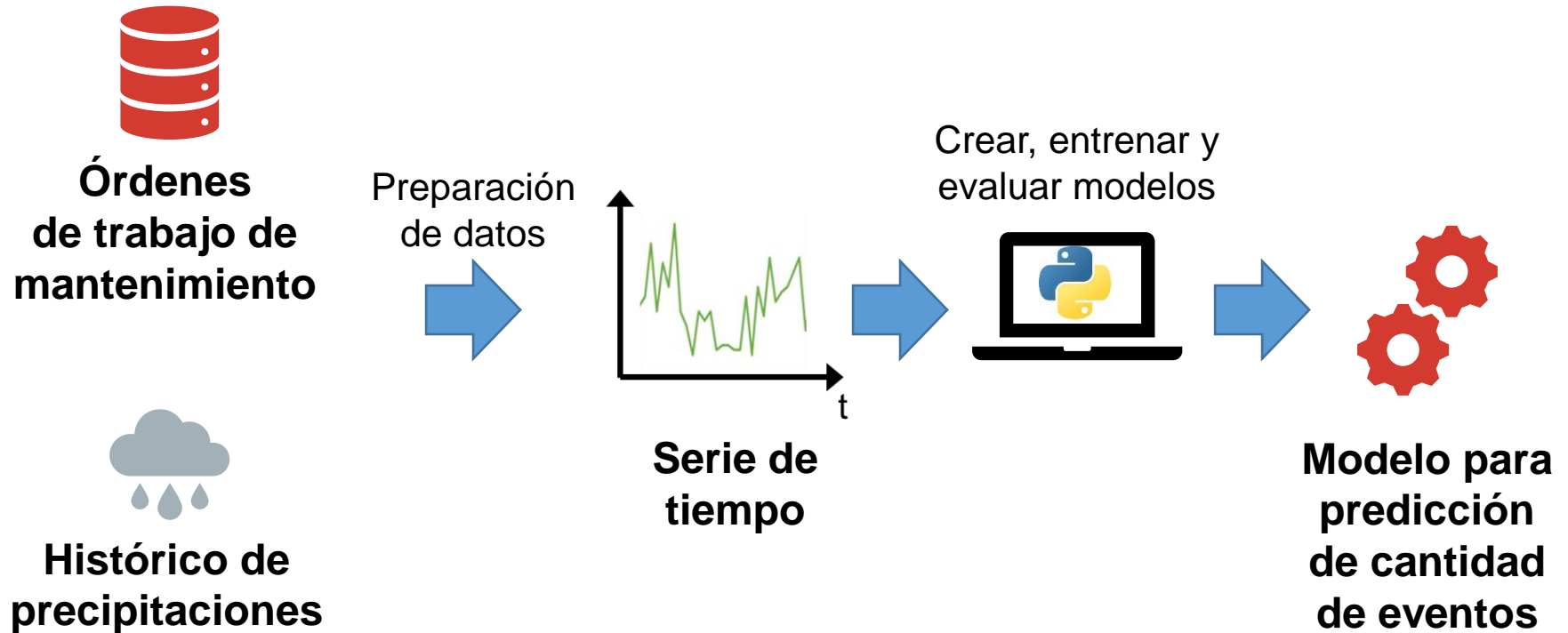


Existe un alto grado de incertidumbre acerca de la cantidad de eventos de poda que se van a generar en un periodo de tiempo



Disponer de una predicción de la cantidad de eventos que deberán atenderse posibilita la planificación y asignación de los recursos requeridos

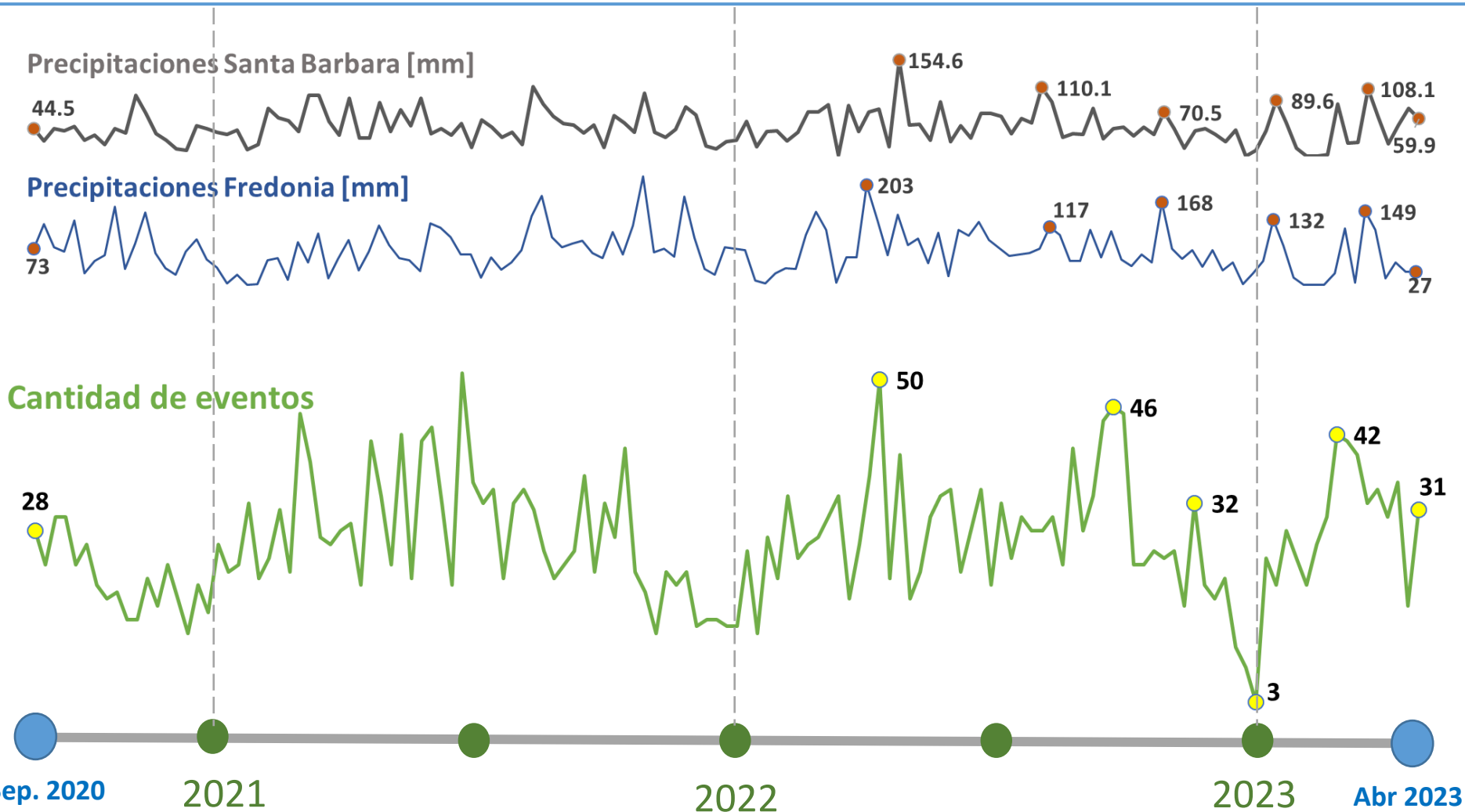
Metodología



Datos

137 datos con frecuencia semanal con la cantidad de eventos de poda

5 variables exógenas: 4 sobre precipitación + 1 de disponibilidad del sistema



Solución propuesta - Modelos

Métodos clásicos de pronóstico en series de tiempo

ARIMA

Auto **ARIMA**

ARIMAX

Métodos de machine learning usados para pronóstico en series de tiempo

Regresión lineal

Random Forest

XGBoost

RN

Aproximación por métodos de series de tiempo

Análisis de la serie de tiempo

- Prueba de estacionaridad

La serie de tiempo es estacionaria

Prueba de Dickey-Fuller (ADH Test):

p-value = **0.0000864 < 0.05**

Prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS test):

p-value = **0.100000 > 0.05**

- La serie de tiempo **no tiene componente de tendencia**
- La serie de tiempo **tiene componente estacional**

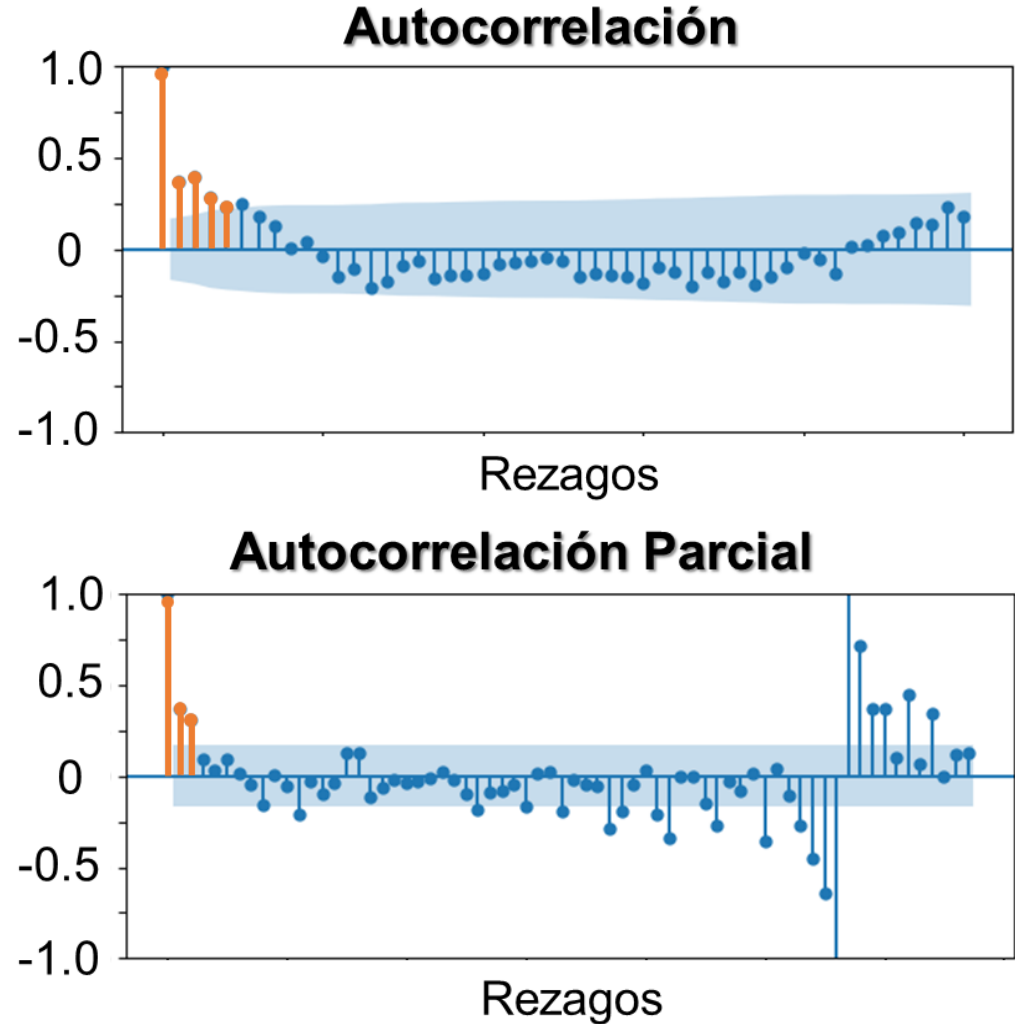
Análisis de la serie de tiempo

Las funciones de autocorrelación y autocorrelación parcial indican que

los valores previos de la serie (rezagos) **pueden ayudar a predecir el valor actual.**

Los datos de la serie de tiempo no se transforman. Al eliminar la componente estacional se pierde la correlación de la serie con sus rezagos.

Se trabajará con los datos en su formato original.



Modelo ARIMA (2,0,3)



Modelo ARIMAX (1,0,1)



Aproximación por métodos de machine learning

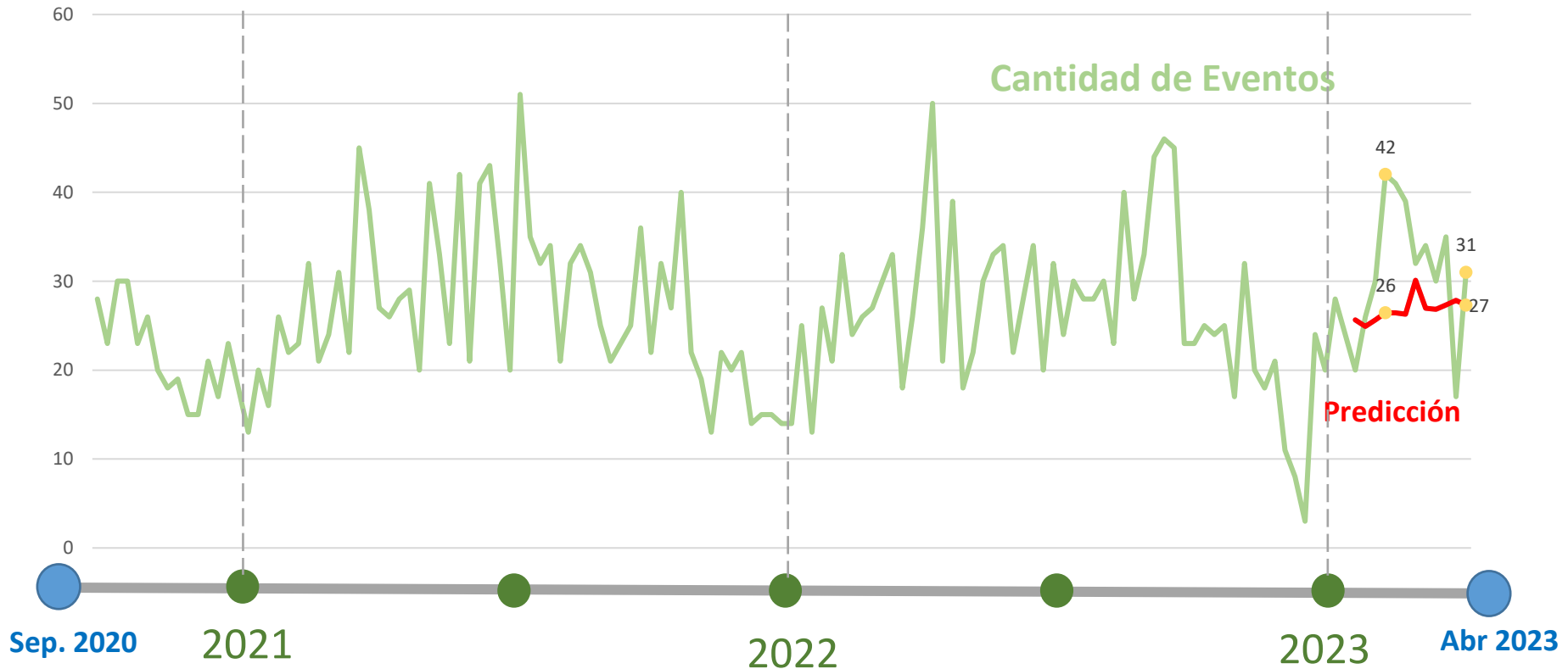
Modelos Random Forest

Hiperparametrización

Se ejecuta una función de *gridsearch*, la cual compara los resultados obtenidos con cada configuración presentada usando *backtesting*.

Random Forest		
• lags	3, 4, 5	3
• n_estimators	50, 100, 200	50
• criterion	squared_error, friedman_mse	squared_error
• max_depth	3, 5, 7	3

Modelo Random Forest



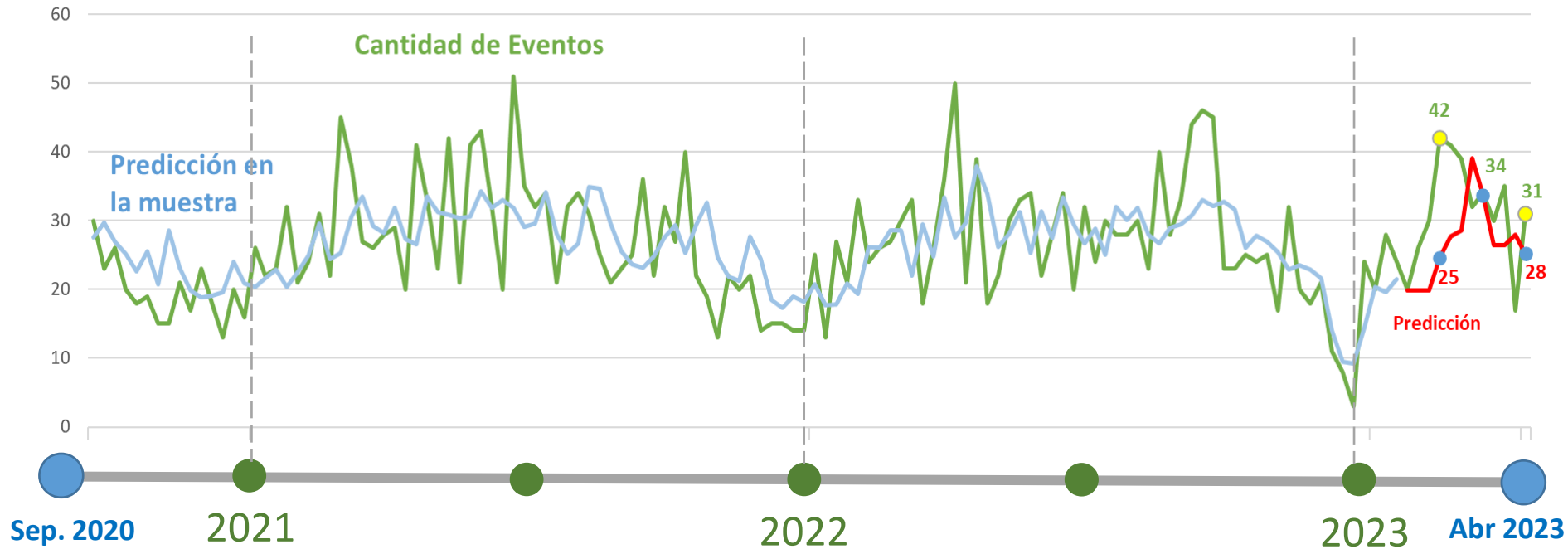
Modelo Red Neuronal (RN)

Hiperparametrización

Se ejecuta una función de *gridsearch*, la cual compara los resultados obtenidos con cada configuración presentada usando *backtesting*.

Red Neuronal		
• capa 1	32, 64	32
• capa 2	128, 256	256
• criterio	mae, mse	mse
• optimizador	Adam, RMSprop	Adam

Modelo Red Neuronal (RN)



Resultados - modelado de serie de tiempo

Comparación de métricas entre modelos

Modelo	RMSE en datos de prueba	Observaciones
ARIMA (2,0,3)	9.635	Datos de la serie de tiempo únicamente
AUTOARIMA (1,0,1)	9.909	Datos de la serie de tiempo únicamente
ARIMAX	11.154	Datos de la serie de tiempo más variable dummy de disponibilidad del sistema
ARIMAX	9.232	Datos de la serie de tiempo, más variable dummy, más datos de precipitaciones, más variable trimestre.

Resultados - modelado con ML

Comparación de métricas entre modelos

En todos los casos los modelos se entrenan con los datos de la serie de tiempo, más variable dummy, más datos de precipitaciones.

Modelo	RMSE en datos de prueba	Observaciones
Regresión Lineal	8.175	R^2 ajustado = 0.320
Random Forest	8.694	{'criterion': 'squared_error', 'max_depth': 5, 'n_estimators': 50, lags: 5}
XGboost	8.214	{'learning_rate': 0.05, 'max_depth': 4, 'n_estimators': 100, lags: 4}
Red Neuronal	7.772	{'optimizer': adam, 'hidden_units_l1': 32, 'hidden_units_l2': 256,}

Predicción

Se usa el modelo de **Red Neuronal** para **predecir la cantidad de eventos de poda que se generarán en la semana N** a partir de:

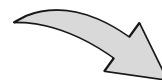
- la cantidad de eventos ocurridos en las **semanas N – 3 a N – 1**
- las precipitaciones en las zonas monitoreadas en la semana **N – 1**
- la indisponibilidad del sistema de información en la semana **N – 1**
- el trimestre correspondiente a la semana a predecir

Trimestre		Cantidad de eventos de poda		Precipitaciones (mm)	
Trimestre 2	1	2023/03/20	30	Bolombolo	20.8
Indisponibilidad		2023/03/27	35	Fredonia	27
Dummy	0	2023/04/03	17	Sta. Barbara	59.9
				Urrao	30.2

Predicción de eventos de poda semana N

2023/04/10

28



Valor real

31

Conclusiones

- El modelo de Red Neuronal entrega una estimación de la cantidad de eventos de poda que se deben atender en la semana siguiente con RMSE de 7.77 lo cual se considera aceptable, teniendo en cuenta el contexto de negocio, aportando así a una planeación óptima de los recursos en campo.
- Las variables exógenas de precipitación de las estaciones Bolombolo, Fredonia, Santa Barbara y Urrao no resultaron estadísticamente significativas para los modelos clásicos de series de tiempo. Por su parte la variable “Trimestre” resultó estadísticamente significativa y aporta a la capacidad predictiva del modelo.
- La metodología desarrollada en este proyecto integrador entrega predicciones que se desvían en ocho unidades del valor real de la cantidad de trabajos de poda por ejecutar en una semana, por lo cual se considera que la metodología puede ser replicada en las otras regiones del departamento de Antioquia.

¡Muchas gracias!