

Modelos de Predicción de Tiempos de Viaje Vehicular

Presentado por:

Nicolás Escobar, Leonela Pereira

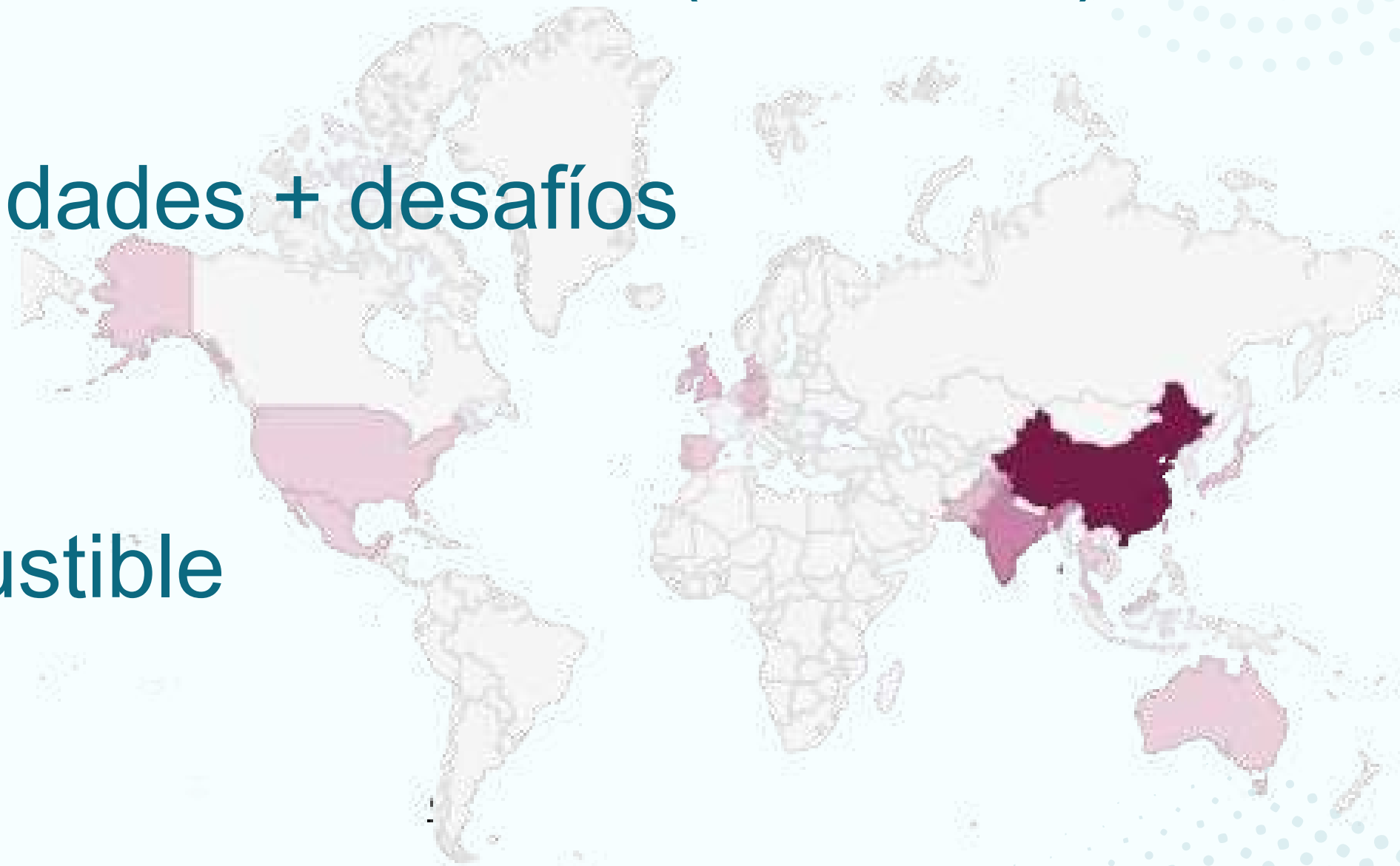
Tutores:

Libertad Tansini

Pedro Piñeyro, Carlos Testuri

Motivación | Contexto Actual

- 55% de la población mundial vive en ciudades (2018, NU)
- Se prevé 2/3 para el 2050
- Rápida urbanización: oportunidades + desafíos
- Desafíos: congestión
- Problemas:
 - Consumo excesivo combustible
 - Contaminación
 - Pérdida calidad de vida
- Transporte: facilita acceso a empleos, educación, servicios, contribuye al crecimiento económico



Agenda

Revisión Sistemática

Definición del Problema

Datos

Métodos Implementados

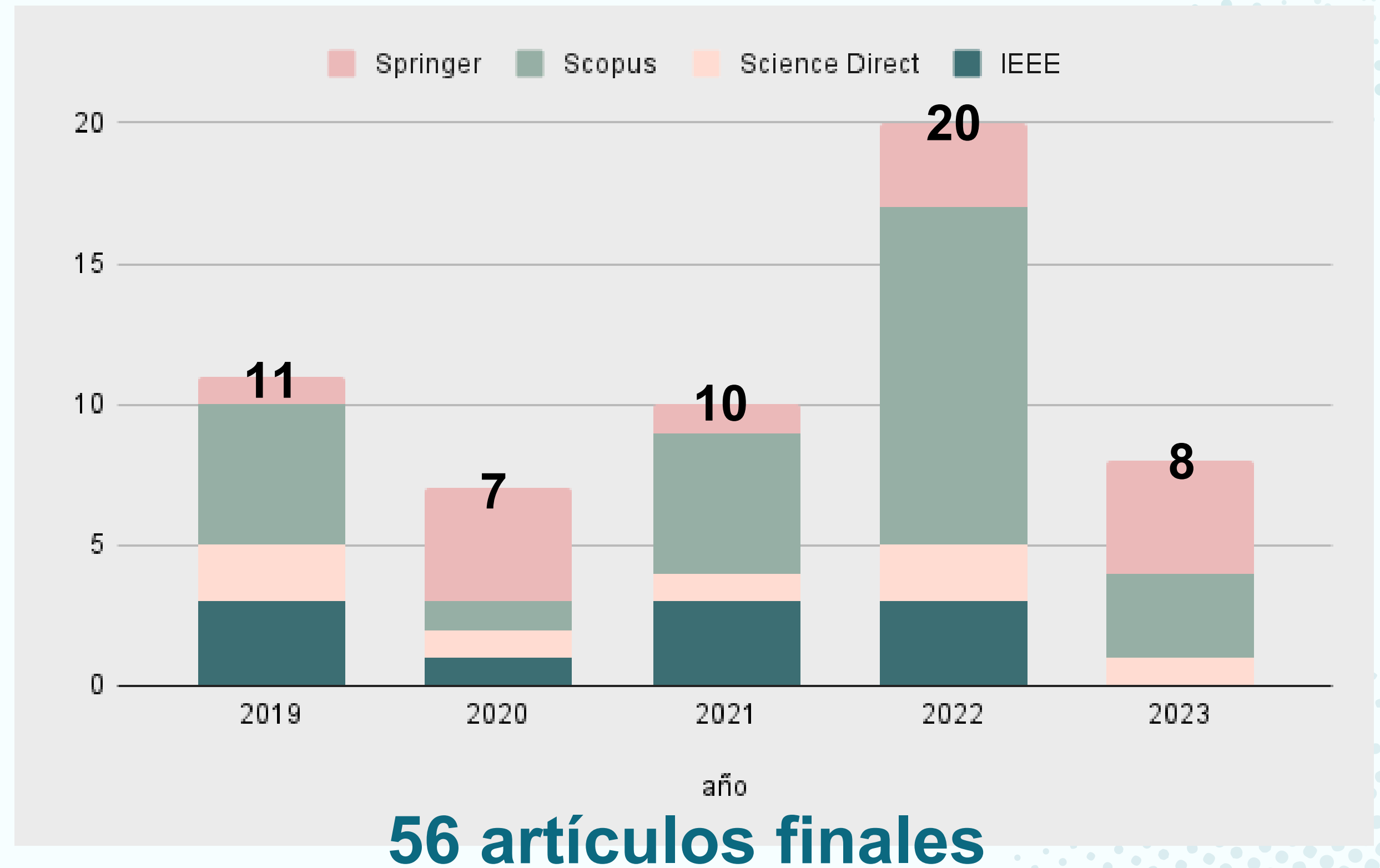
Evaluación Desempeño

Conclusiones

Trabajo Futuro

Revisión Sistemática - Artículos Recuperados

- Formulación string
- Criterios exclusión
- 17 set 2023



RS- Elementos a considerar en TTP

Elementos a tener en cuenta en el **proceso** de predicción:

- Factores: temporal, espacial, exógeno
- Datos: recolección, depuración
- Método de predicción: ingenuos, teóricos, basados en datos

Elementos **transversales** al problema de predicción:

- Tipo de Vehículo: particular, taxi, ómnibus
- Entorno: urbano, autopista
- Enfoque: trayectoria conocida, OD



Agenda

Revisión Sistemática

Definición del Problema

Datos

Métodos Implementados

Evaluación Desempeño

Conclusiones

Trabajo Futuro

Definición del Problema

- Predicción de tiempos de viaje (TTP) para **ómnibus urbano**
- Interés: mejorar confiabilidad del servicio



- El recorrido del ómnibus tiene n paradas
- t_i es el instante en que el ómnibus pasa por la parada i
- El ómnibus se encuentra en la parada intermedia k

Objetivo: predecir el tiempo restante hasta llegar al destino

Agenda

Revisión Sistemática

Definición del Problema

Datos

Métodos Implementados

Evaluación Desempeño

Conclusiones

Trabajo Futuro

Datos - Presentación

Línea 117 CUTCSA
Enero 2021 - Agosto 2024

Dataset: GPS. información sobre cada instancia de recorrido

Variante	Dirección Origen	Cantidad Paradas	Dirección Destino
635	Juncal y Sarandí	24	Bv. Gral. Artigas y Parva Domus
2779	Ciudadela y Rincón	35	Bv. Gral. Artigas y Parva Domus
2778	Bv. Gral Artigas y Tabaré	21	Av. 18 de Julio y Convención
634	Bv. Gral Artigas y Tabaré	22	Juncal y Sarandí

Datos - Análisis Exploratorio Datos (EDA)

- Exploración general inicial
- Nulos
- Análisis:
 - formato
 - tipos de datos
 - coherencia valores
 - columnas
- Mapas: recorridos de variantes y detección de fechas no crecientes

matplotlib

NumPy

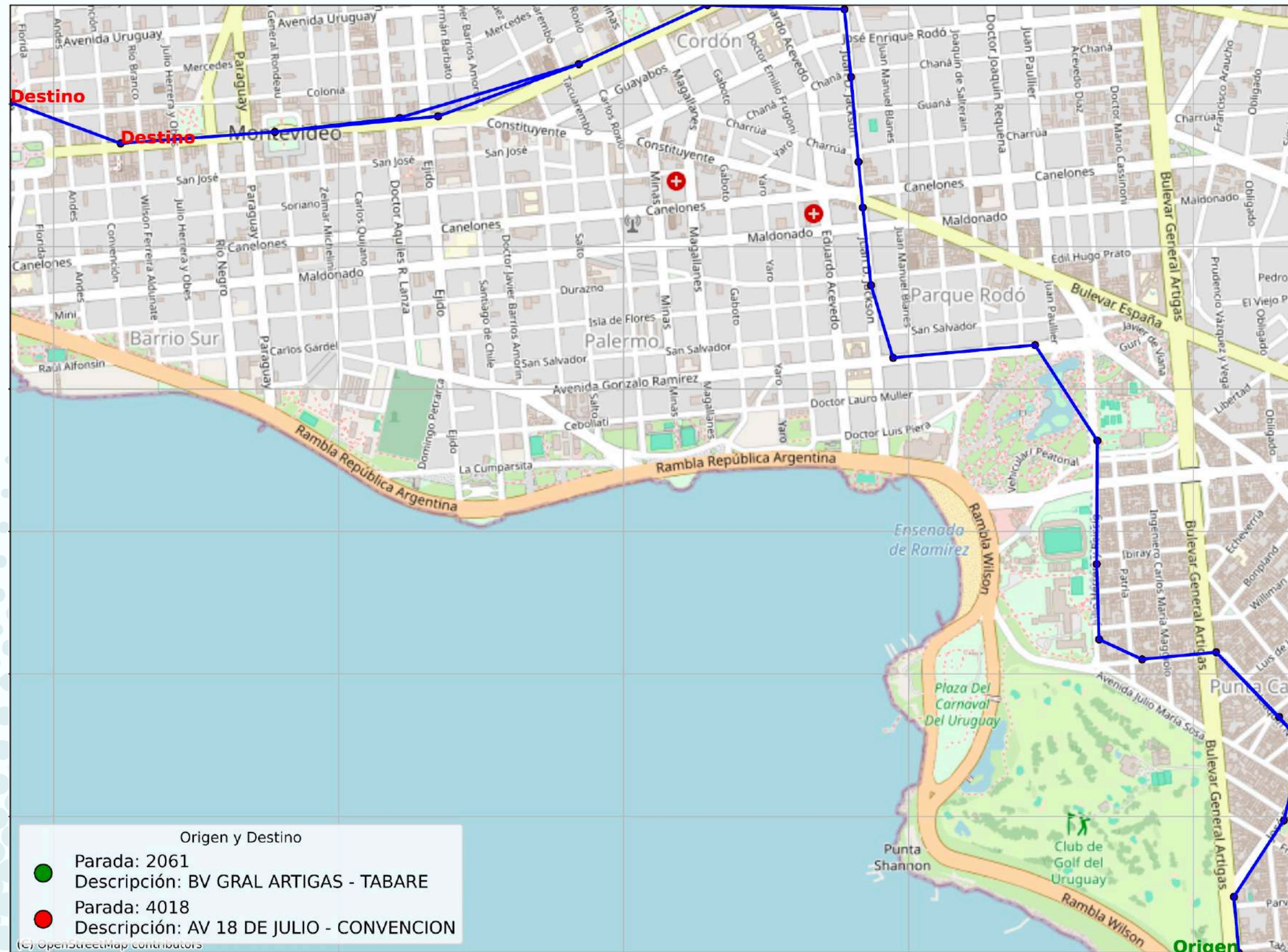


pandas

Languages

Jupyter Notebook 79.0% Python 21.0%

Datos - Análisis Exploratorio de datos



Problema

- Desplazamiento de paradas a partir de julio 2021
- Destinos distintos en variante 2778

Datos - Preprocesamiento

Limpieza

- Eliminar recorridos con:
 - Nulos
 - Datos previos a julio 2021
 - Duplicados
 - Recorridos con fecha no creciente
 - Outliers: IQR
- Eliminar columnas irrelevantes
- Unificar formato fecha
- Variante 2778: nos quedamos con el 1er destino

Creación

- Nuevas características:
 - día, mes, año, hora, minuto, segundo, ordinal para día de la semana, tipo de día
- Nuevo campo: travel_time

Modificación

- Renombrar campos

Datos - Metadata Final

Externo: id_recorrido, coche, variante, travel_time (min)

Espacial: parada (ordinal de la parada)

Temporal:

- fecha_pasada
- tipo_día: semana, fin de semana, feriado (categórico)
- día(1-7), dia_numero, mes, año
- hora, minuto, segundo

Datos - Dataset Final

Variante	Trayectorias Iniciales	Trayectorias finales
634	55648	25783
635	52678	29787
2778	8248	5118
2779	8298	4749

Entrenamiento (70%)

Validación (15%)

Testing (15%)

Agenda

Revisión Sistemática

Definición del Problema

Datos

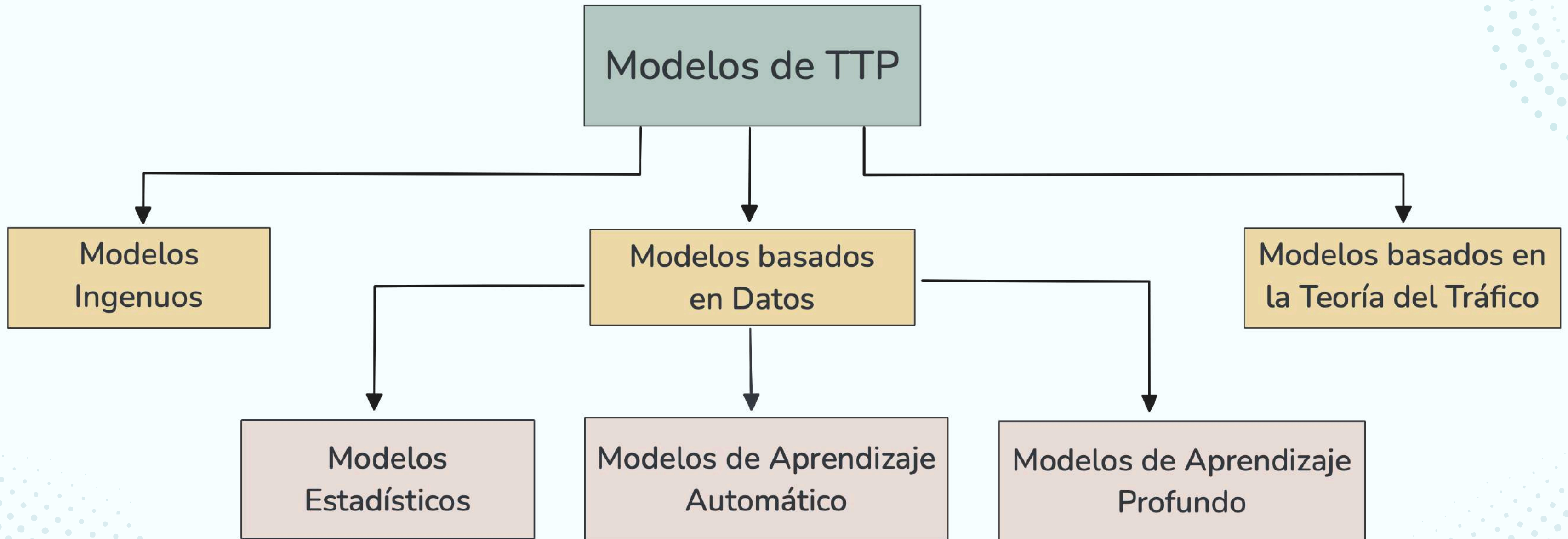
Métodos Implementados

Evaluación Desempeño

Conclusiones

Trabajo Futuro

Métodos Implementados- Introducción (RS)



Métodos Implementados- ARIMA (baseline)

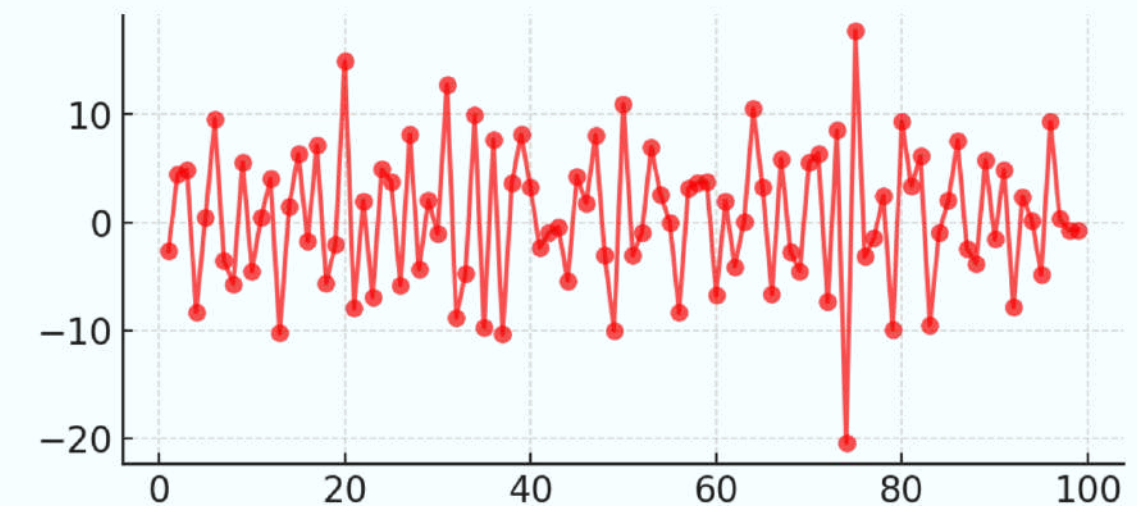
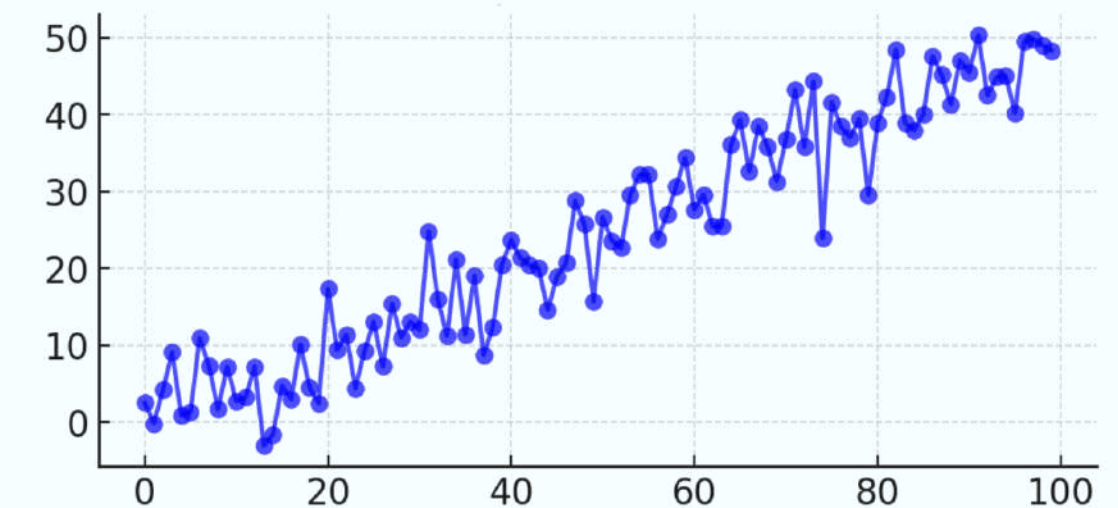
Modelo Estadístico

- Series temporales, incluye **AR**, **MA** y **ARMA**
- Ejemplo: predicción de temperatura máxima
- Restricción para ARMA → **serie estacionaria**
- ARIMA: **diferenciación** → aplicar ARMA

Implementación

- Statsmodels, Test ADF, modelado por parada
- Búsqueda en cuadrícula, MAPE

serie temporal original



serie temporal diferenciada

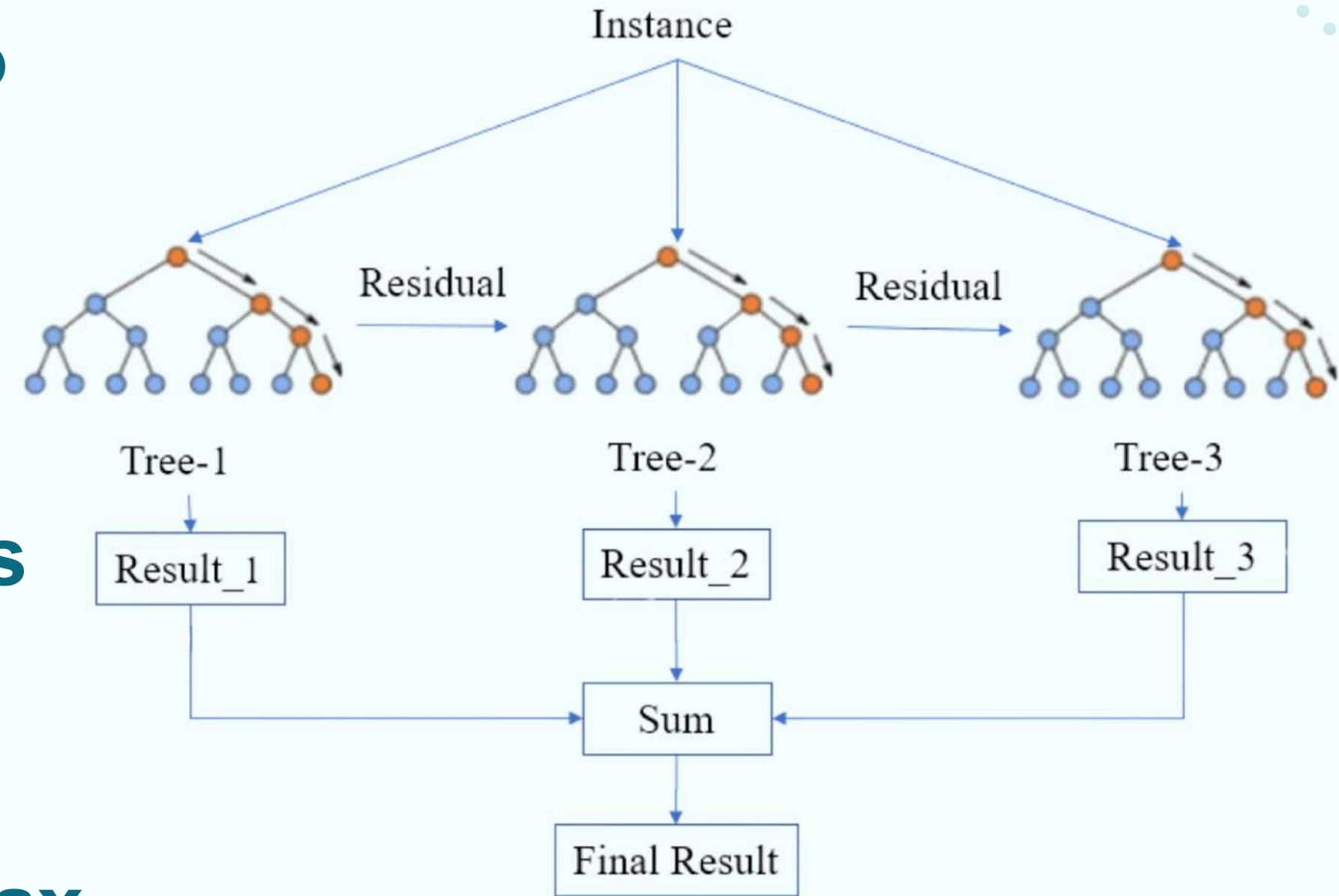
Métodos Implementados- XGBoost

Modelo de Aprendizaje Automático

- Eficiente, escalable, no lineal
- Árboles de decisión, **Boosting**
- Utiliza **residuo** para entrenar
- Restricción → **datos estructurados**

Implementación

- XGBoost, 3 modelos: **min, med, max**
- Búsqueda RandomizedSearchCV, MSE

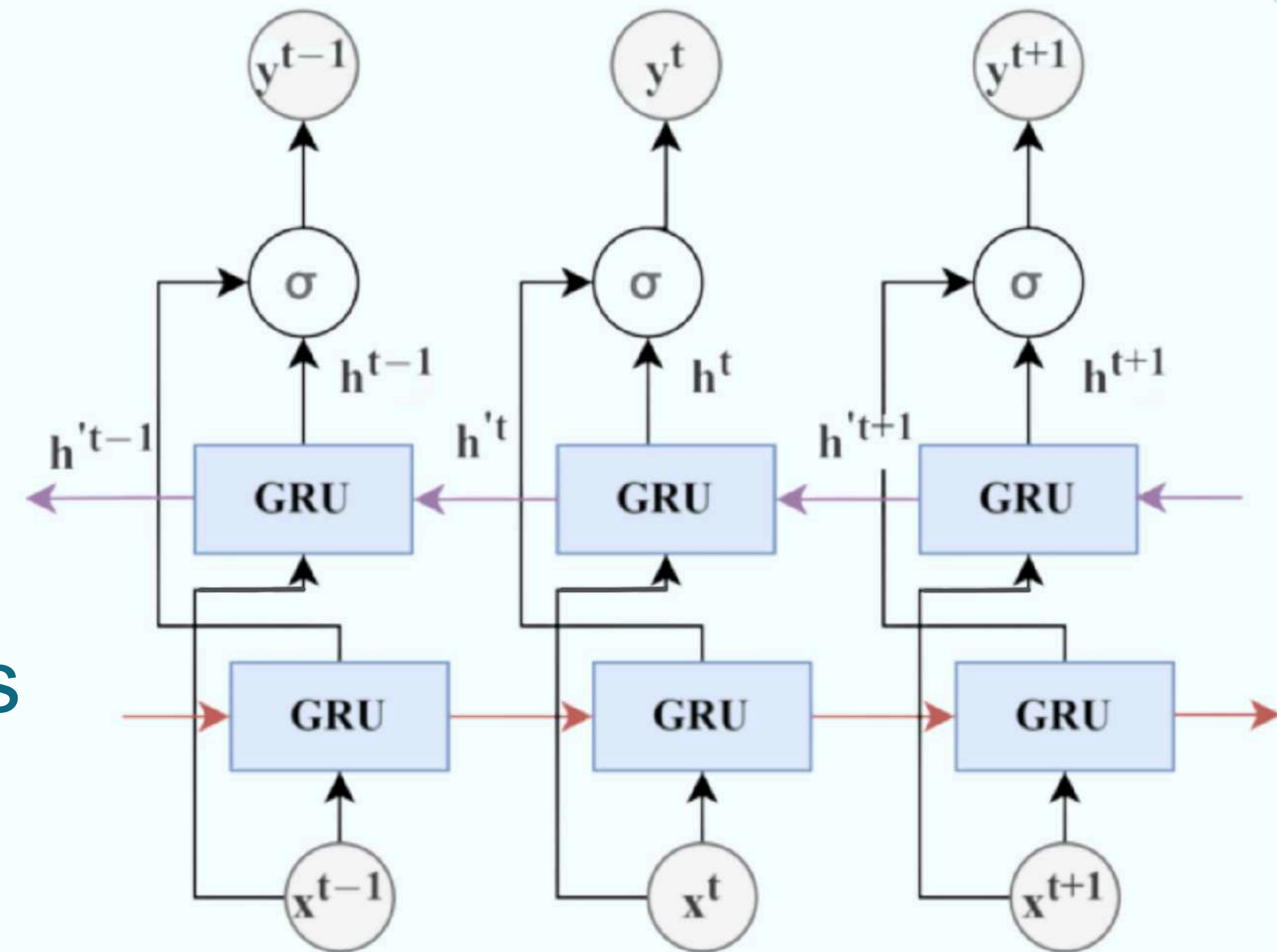


Wang y cols., Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm

Métodos Implementados- Bi-GRU

Modelo de Aprendizaje Profundo

- Red neuronal recurrente
- Captura **dependencias** temporales
- Almacena información histórica
- **Alto** costo computacional
- **Bi-GRU** procesa en dos direcciones
- Mejora el **contexto** del modelo
- Restricción → **datos secuenciales**



Chughtai y cols., A heterogeneous ensemble approach for travel time prediction using hybridized feature spaces and support vector regression

Métodos Implementados- Bi-GRU

Secuencias de Entrada

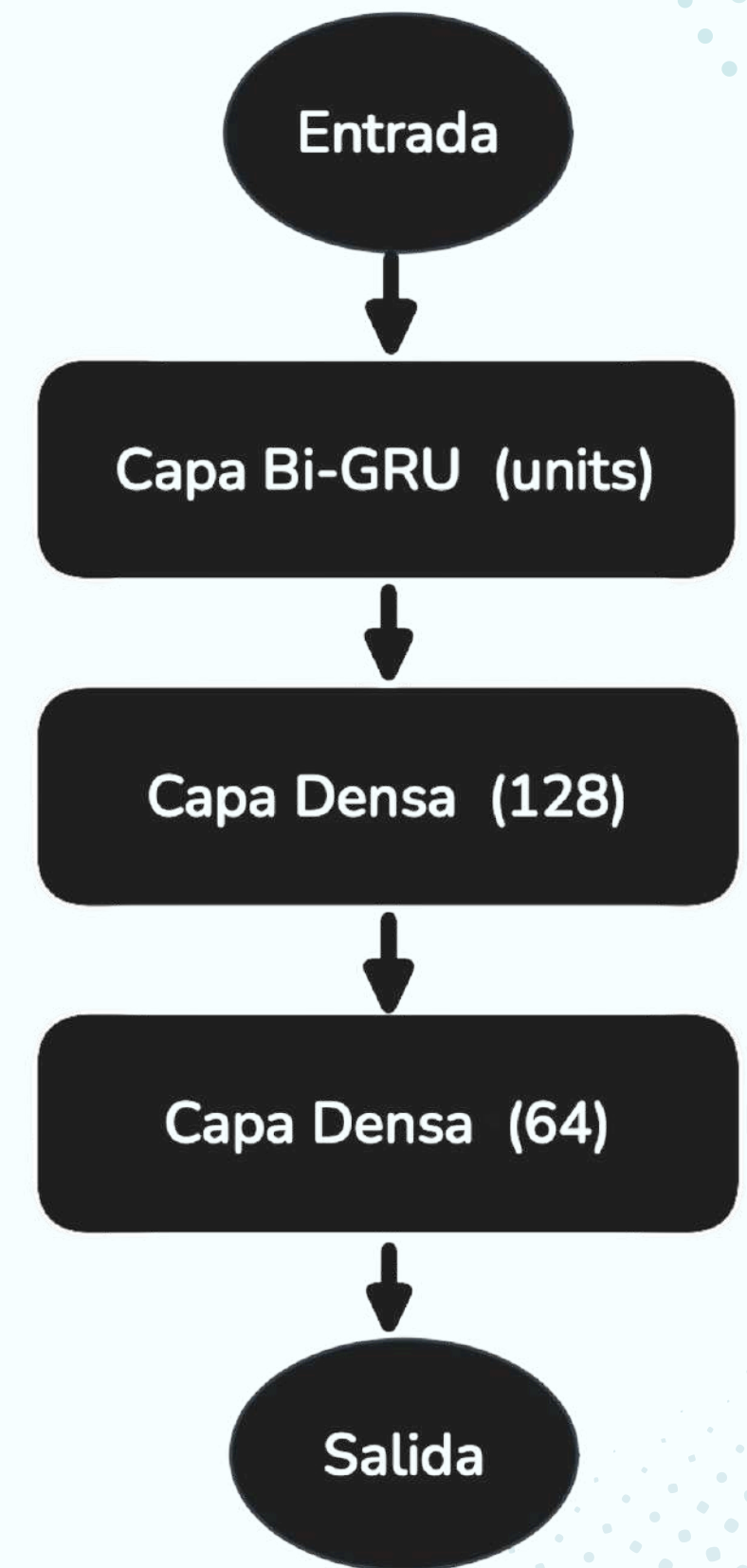
- (viajes, paradas, características)

Arquitectura del Modelo

- Capas: Entrada, Bi-GRU, 2 Capas Densas, Salida

Implementación

- TensorFlow / **Keras API**
- Búsqueda Ray Tune / HyperBand, MSE
- Para evitar sobreajuste → **Early Stopping**



Métodos Implementados - Tiempos Entrenamiento

Tiempos de entrenamiento (en minutos) por modelo y variante				
Modelo	Variante 634	Variante 635	Variante 2778	Variante 2779
ARIMA	6.05	8.45	1.65	2.73
XGBoost_min	0.45	0.55	0.08	0.13
XGBoost_med	0.50	0.67	0.08	0.17
XGBoost_max	0.85	1.13	0.15	0.25
Bi-GRU	359.55 (≈ 6 hs)	285.88 (≈ 4:45 hs)	87.28 (≈ 1:27 hs)	52.03

Equipo dedicado:
Mac mini M4 Pro
12 núcleos CPU - 48GB RAM

Agenda

Revisión Sistemática

Definición del Problema

Datos

Métodos Implementados

Evaluación Desempeño

Conclusiones

Trabajo Futuro

Evaluación Desempeño - Experimentos | Análisis

Experimentos:

- Standard test
- Fin de Semana
- Vacaciones

Análisis:

- Comparación dentro de una misma variante
- Comparación del desempeño de un mismo modelo en distintas variantes
- Comparación entre experimentos



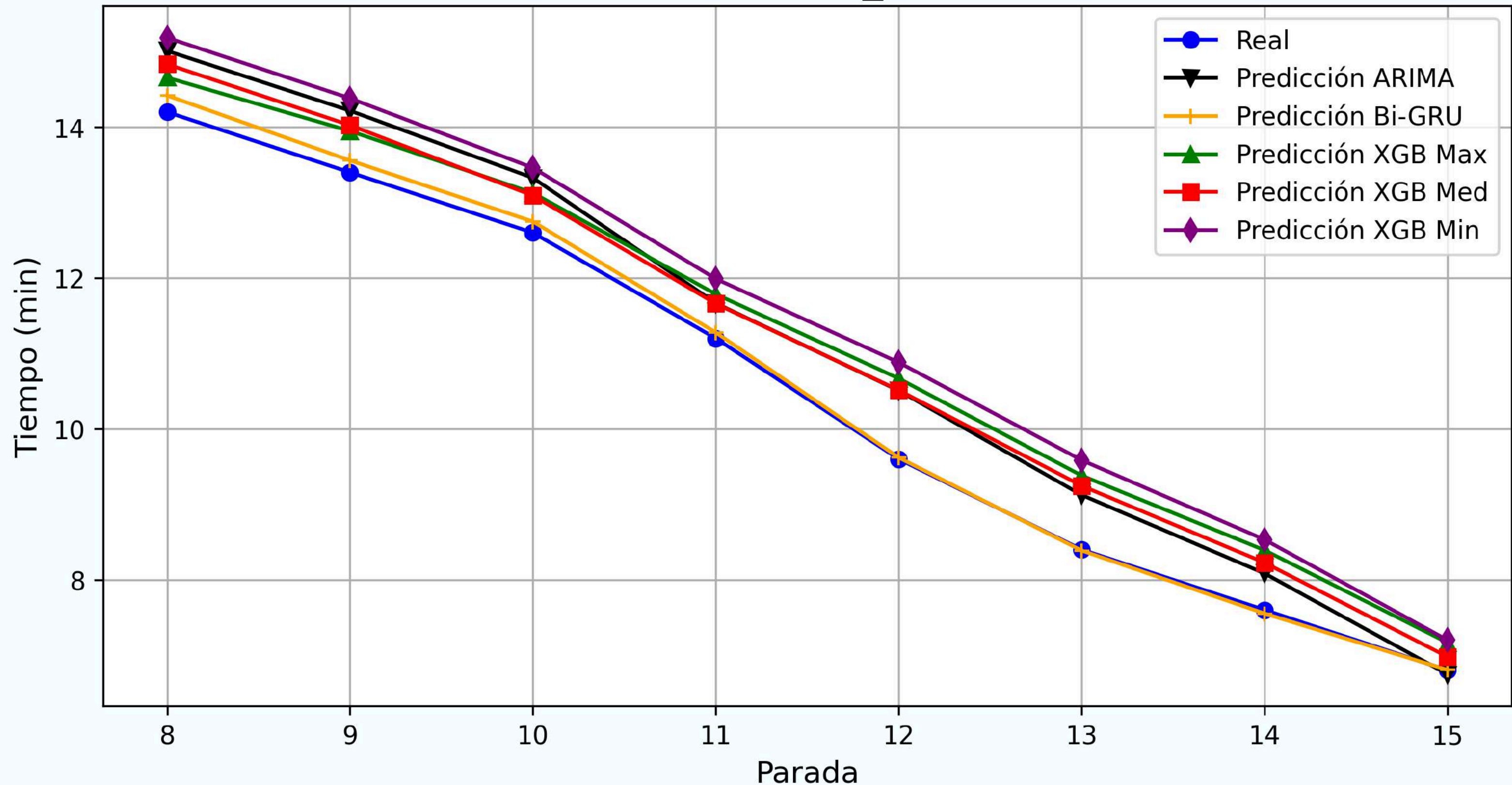
Evaluación Desempeño - Resultados Standard Test

Variante	Modelo	MSE(m^2)	RMSE(m)	MAE(m)	MAPE(%)	R^2
635	ARIMA	1.946	1.395	1.000	10.578	0.973
	Bi-GRU	0.081	0.284	0.216	3.786	0.999
	XGBoost_max	1.376	1.173	0.845	10.146	0.981
	XGBoost_medium	1.575	1.255	0.897	9.734	0.978
	XGBoost_min	1.922	1.386	0.987	10.354	0.973

Variante	Modelo	MSE(m^2)	RMSE(m)	MAE(m)	MAPE(%)	R^2
2779	ARIMA	4.234	2.058	1.485	10.542	0.969
	Bi-GRU	0.440	0.663	0.493	4.611	0.997
	XGBoost_max	2.578	1.606	1.147	9.922	0.981
	XGBoost_medium	2.831	1.682	1.190	8.812	0.979
	XGBoost_min	3.800	1.949	1.377	10.075	0.972

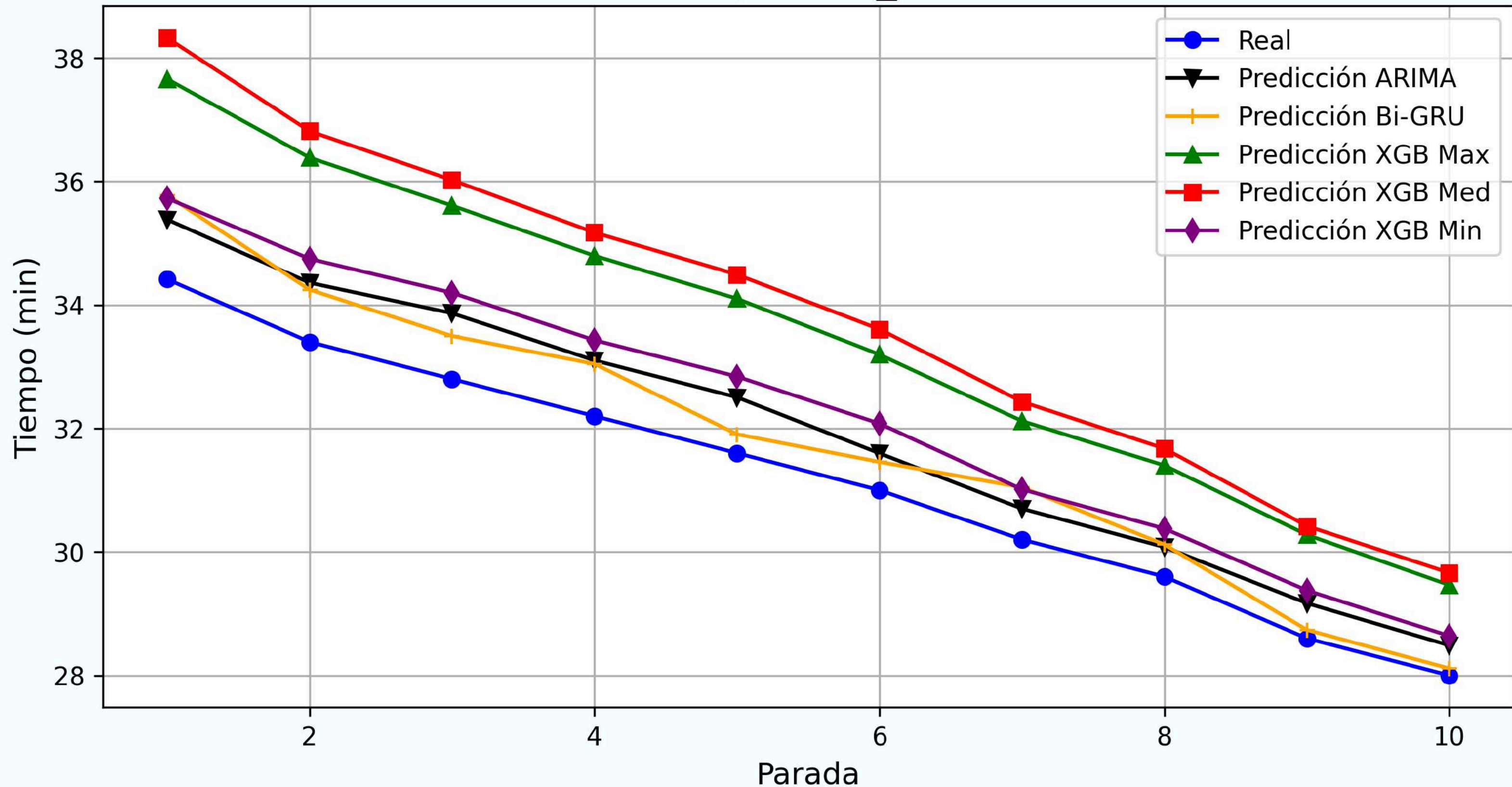
Evaluación Desempeño - Predicciones

635: Predicción vs Real para id_recorrido: 29333704362



Evaluación Desempeño - Predicciones

2779: Predicción vs Real para id_recorrido: 29354976906



Agenda

Revisión Sistemática

Definición del Problema

Datos

Métodos Implementados

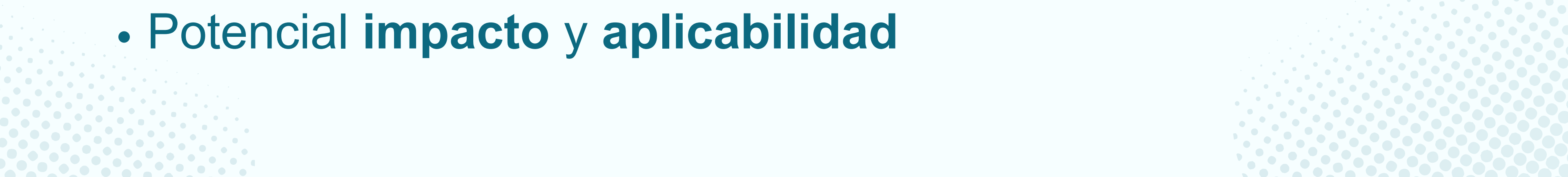
Evaluación Desempeño

Conclusiones

Trabajo Futuro



Conclusiones

- **RS** → alcance y contextualización
 - **EDA** y calidad de datos → impacto en desempeño
 - **Bi-GRU** → **excelente precisión** / costo computacional alto
 - **XGBoost** → **equilibrio** entre precisión y eficiencia
 - Conjunto de **características** influye en desempeño
 - **ARIMA** → **dificultad** para capturar relaciones complejas
 - **Resultados** → alineados con la literatura
 - Potencial **impacto** y **aplicabilidad**
- 

Agenda

Revisión Sistemática

Definición del Problema

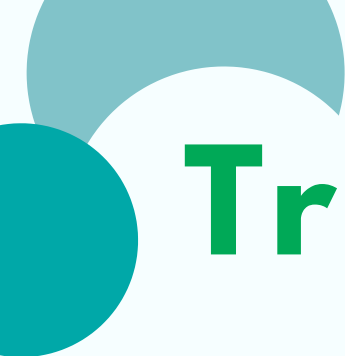
Datos

Métodos Implementados

Evaluación Desempeño

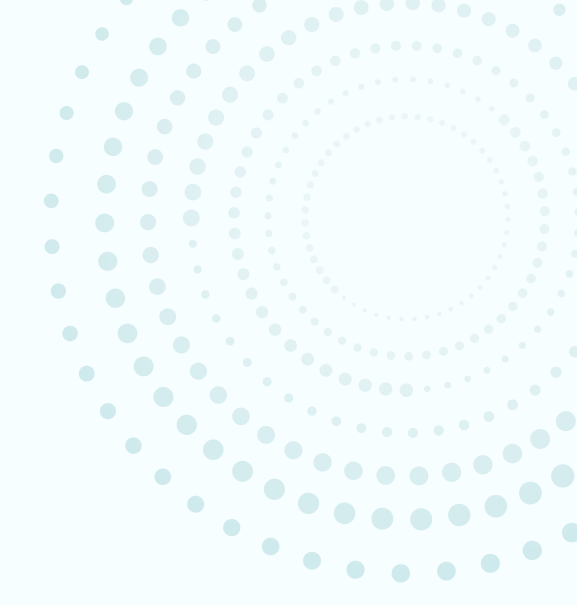
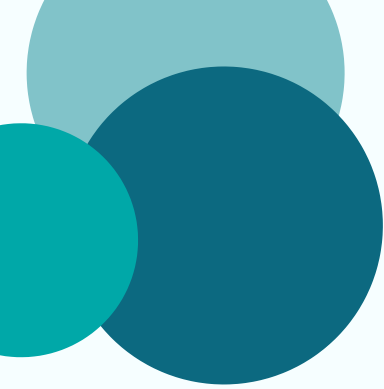
Conclusiones

Trabajo Futuro



Trabajo Futuro

- **Unificar** búsqueda de hiperparámetros
 - Incorporar características **exógenas**
 - TTP entre **paradas específicas**
 - Modelar coche en **marcha** / coche **detenido**
 - Arquitecturas híbridas (CNN, GCN, atención, encoder-decoder)
 - Predecir TTP en **tiempo real** (Filtro de Kalman)
- 



¡Gracias!