



**MÁSTER EN CIENCIA DE DATOS E  
INGENIERÍA DE COMPUTADORES**

# **Uso de Machine Learning en la Determinación de Niveles de Suciedad en Sistemas Fotovoltaicos a partir de parámetros ambientales**

*Autora : Luíza Araujo Costa Silva*

*Tutora: Prof<sup>a</sup>. Dr<sup>a</sup>. María del Carmen Pegalajar Jiménez*

*Julio/2021*

# Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

# Contenido

- **Introducción**
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

# Energía Fotovoltaica

- Radiación / Luz Solar → Energía Eléctrica
- Paneles o módulos solares
- España → El **7,5%**  
de la energía generada <sup>1</sup>



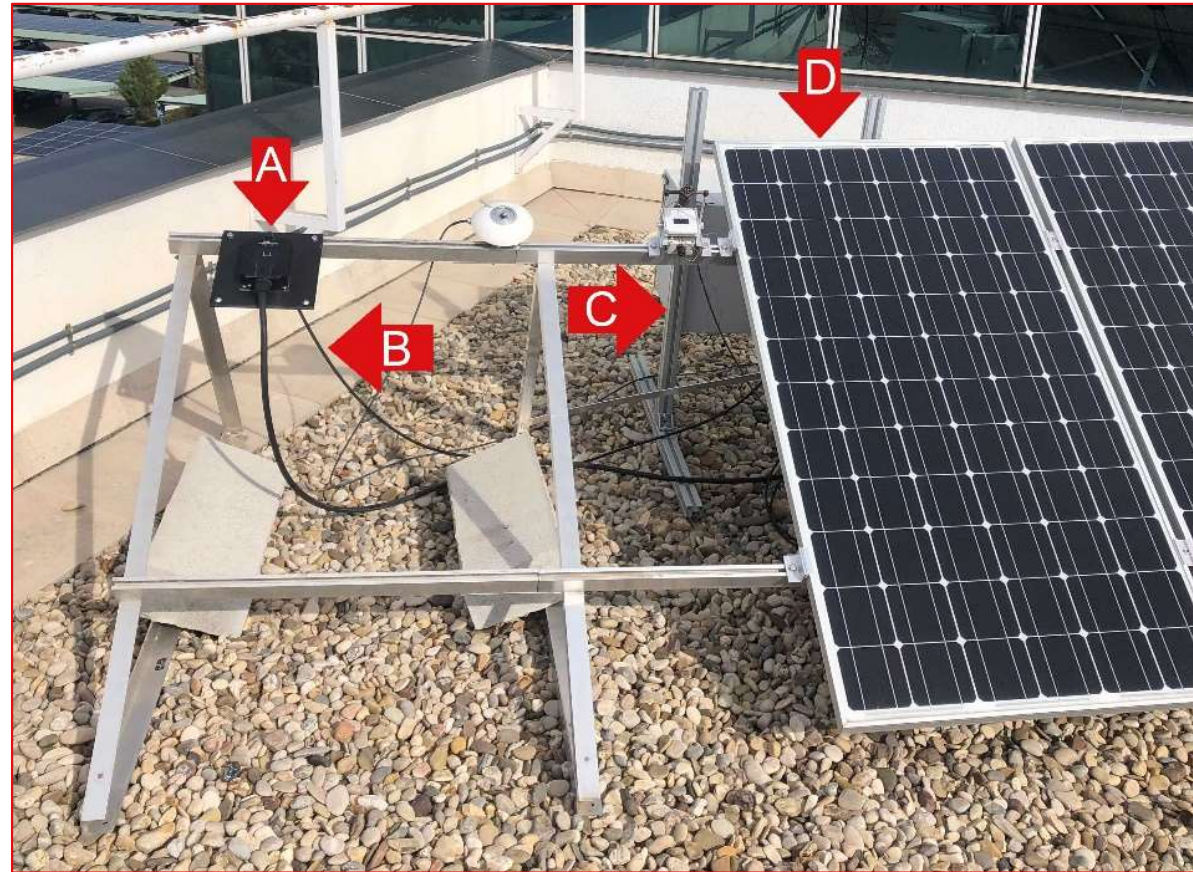
## Soiling: La suciedad en los módulos fotovoltaicos

- Impide la transmisión de luz en el módulo
- Reducción de hasta un **25%** del rendimiento energético<sup>2</sup>
- Pérdidas causadas<sup>3</sup>:
  - 2018 → € 3 a 5 mil millones
  - 2023 → € 7 mil millones
- Parámetros ambientales:
  - lluvia, temperatura, contaminación atmosférica...

## Nivel de suciedad

- **SR** – *Soiling Ratio*
  - Energía no generada debido al soiling
  - Variación de 0 a 1

$$SR = \frac{\text{Salida Eléctrica}^{(D)}}{\text{Salida Eléctrica}^{(A)}}$$



## El problema propuesto

- Actualmente:
  - Modelos matemáticos, utilizándose datos de mediciones en suelo
- Propuesta:
  - Utilizar técnicas de *Machine Learning* para determinar el *Soiling Ratio* (**SR**), utilizándose datos de parámetros ambientales procedentes de satélites



# Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión



# Contenido

- Introducción
- **Análisis Exploratorio y Preprocesamiento**
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

## Procedencia de los datos

### CEACTEMA

Centro de Estudios Avanzados en  
Ciencias de la Tierra, Energía y Medio  
Ambiente



Marzo/2019 a  
Marzo/2021

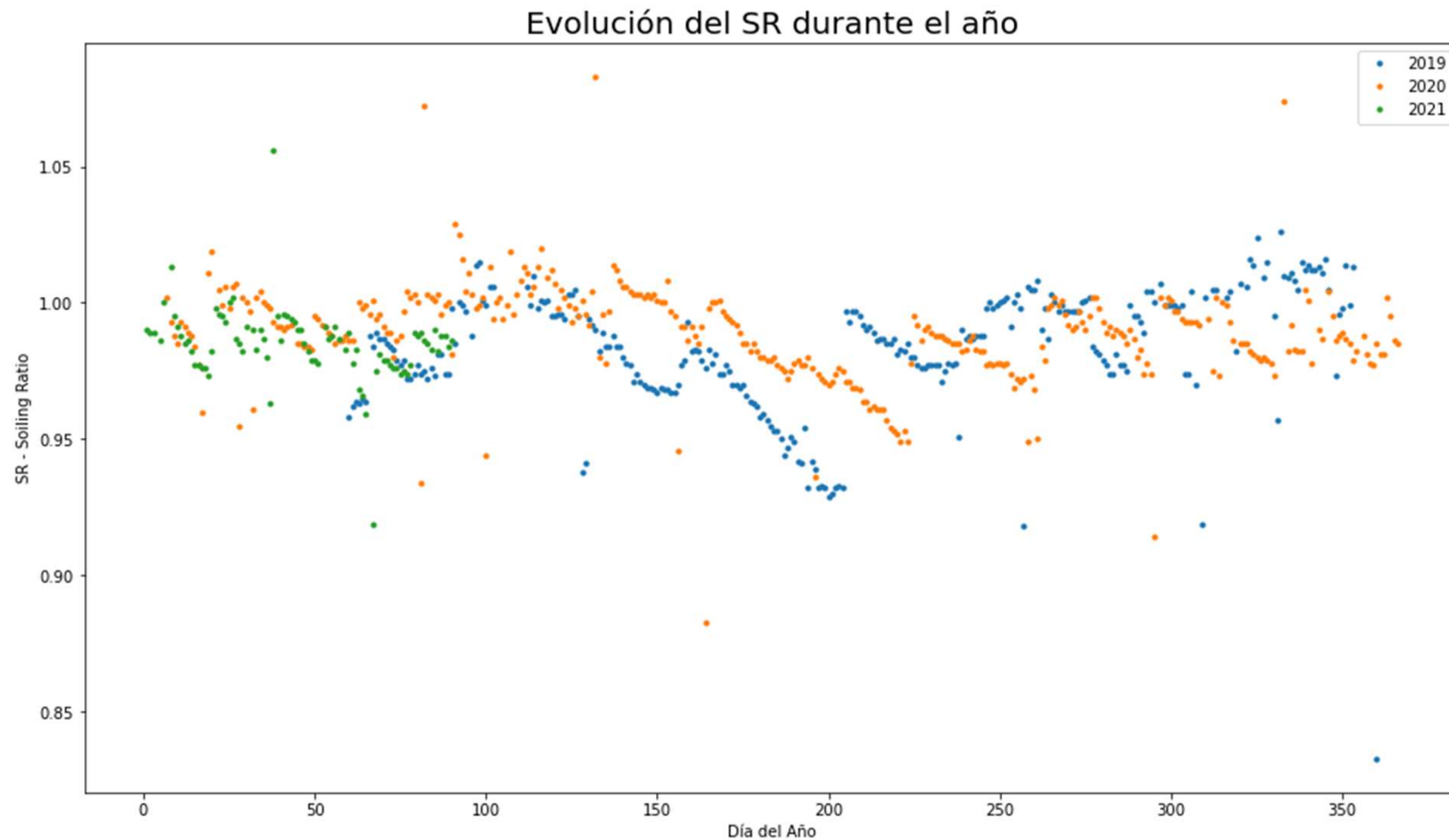


### MERRA-2

*Modern-Era Retrospective analysis for  
Research and Applications*



# Análisis de los datos



## Análisis y Preprocesamiento de los datos

- *Outliers* → Método Z-Score
- Imputación de datos nulos → Media entre sus vecinos
- Estacionariedad → Prueba de *Dickey-Fuller* aumentada
- Normalización → RobustScaler
- Ventanas deslizantes → 1, 2, 3, 5, 7, 10, 14 días anteriores

## Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

## Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- **Experimentos Realizados**
- Predicciones Secuenciales
- Conclusión

## Métodos utilizados

- Scikit-learn
  - Regresión Lineal
  - Árbol de Decisión
  - Random Forest
  - MLP – Perceptrón Multi-Capa
- Keras
  - LSTM





## Modelos Matemáticos

- Coello y Boyle <sup>4</sup>
- You <sup>5</sup>

'Temperature',  
'Wind\_speed',  
'PM10',  
'DUSMASS25',  
'Rainfall'

Modelo	RMSE	MAE	MAPE	R <sup>2</sup>	Origen
Coello y Boyle	0.016876	0.011128	1.136918	0.341808	Datos de Suelo
	0.038538	0.023643	2.420922	0.205066	Datos de Satelites
You	0.019935	0.013792	1.418861	0.347129	Datos de Suelo
	0.030156	0.01736	1.777113	0.185304	Datos de Satelites
Media:	0.02637625	0.01648075	1.6825	0.2697185	

## Selección de Características

- **Input\_1** → Algoritmo Boruta
  - 'Temperature', 'Relative\_Humidity', 'Pressure', 'Wind\_speed', 'Air\_Density', 'Short\_wave\_irradiation', 'Rainfall'
  - (+) 'Ventana\_d-1', 'Ventana\_d-2',... 'Ventana\_d-*n*']
- **Input\_2** → Modelos Matemáticos
  - 'Temperature', 'Wind\_speed', 'PM10', 'DUSMASS25', 'Rainfall'
  - (+) 'Ventana\_d-1', 'Ventana\_d-2',... 'Ventana\_d-*n*']
- **Input\_3** → Sólo ventanas deslizantes
  - 'Ventana\_d-1', 'Ventana\_d-2',... 'Ventana\_d-*n*']

## Partición Entrenamiento / Test

- Entrenamiento → 70%
- Test → 30%

## Métricas Utilizadas

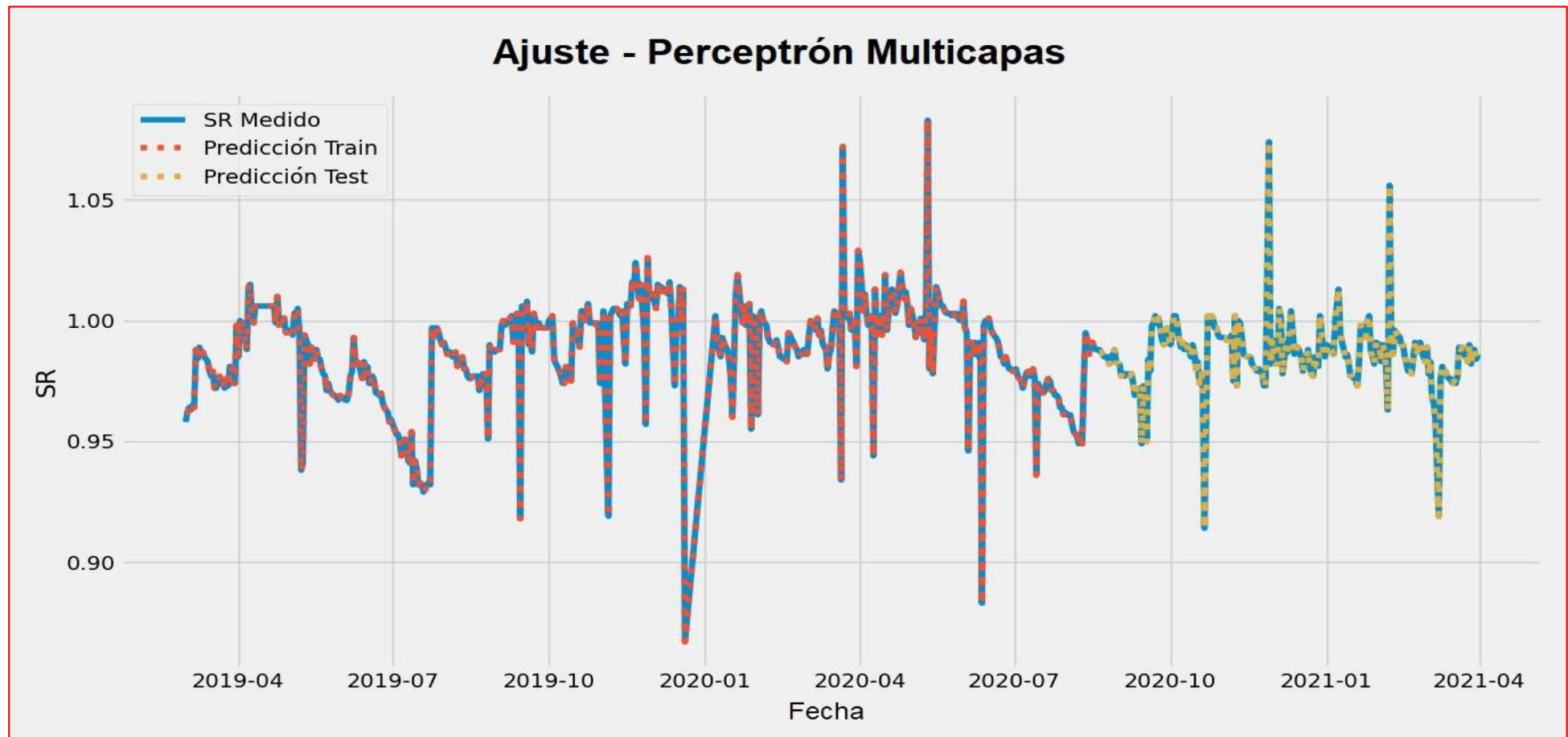
- RMSE
- MAE
- MAPE
- $R^2$

## Resultados Obtenidos por Algoritmo

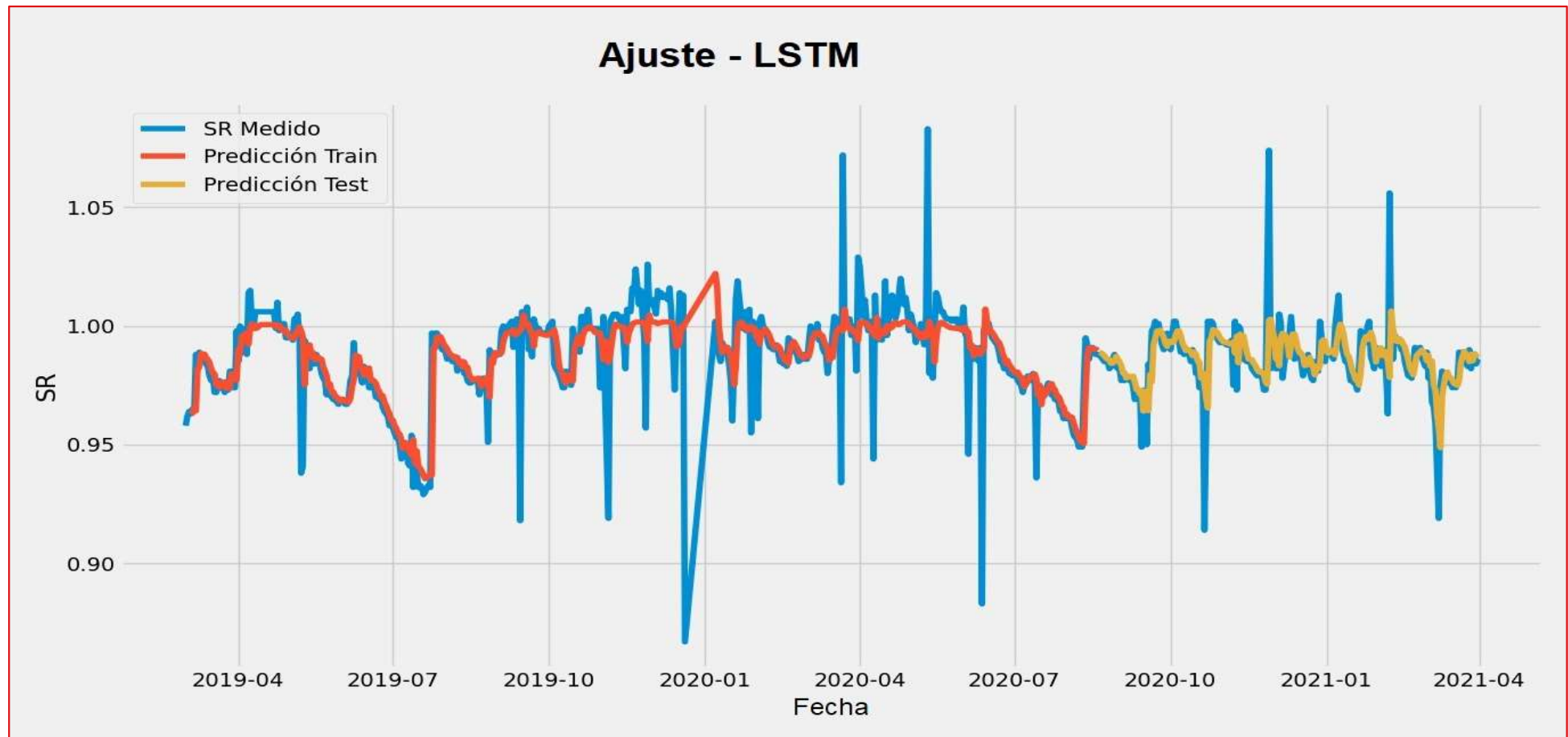
Métodos	Mejor Modelo			Peor Modelo			Media
	Input	Ventana	RMSE	Input	Ventana	RMSE	RMSE
Regresión Lineal	Input_3	1	0.000068	Input_1	14	0.003456	0.0010227
MLP	Input_3	1	0.000068	Input_2	5	0.000908	0.0003173
Random Forest	Input_3	1	0.000576	Input_2	5	0.002139	0.0010653
Árbol de Decisión	Input_3	1	0.001117	Input_2	5	0.003177	0.0017348
LSTM	Input_3	3	0.012109	Input_1	2	0.016131	0.0131012

Modelo Matemático	Mejor RMSE	Peor RMSE	Media RMSE
Coello y Boyle	0.016876	0.038538	0.027707
You	0.019935	0.030156	0.025046

## Ajuste MLP



## Ajuste LSTM



## Los 5 Mejores Resultados Obtenidos

#	Modelo	Input	Vent	Test			
				RMSE	MAE	MAPE	R <sup>2</sup>
1	Regresión Lineal	Input_3	1	0.000068	0.000013	0.001361	0.999977
2	MLP (1 capa con 30 neuronas)	Input_3	1	0.000068	0.000013	0.001365	0.999977
3	Regresión Lineal	Input_2	1	0.000073	0.000031	0.003113	0.999974
4	MLP (1 capa con 30 neuronas)	Input_2	1	0.000073	0.000031	0.003119	0.999974
5	MLP (1 capa con 30 neuronas)	Input_1	1	0.000078	0.000036	0.003712	0.999971

#	Modelo Matemático	RMSE	MAE	MAPE	R <sup>2</sup>
6	Coello y Boyle	0.016876	0.011128	1.136918	0.341808
7	You	0.019935	0.013792	1.418861	0.347129



## Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

# Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- **Predicciones Secuenciales**
- Conclusión

## Predicciones secuenciales

- Predicciones **futuras** en la serie temporal
  - Se entrena el modelo con los valores medidos
  - Se predice utilizando las salidas del modelo como entrada para la siguiente predicción
  - Hasta el horizonte temporal futuro que se desee

## Ejemplo:

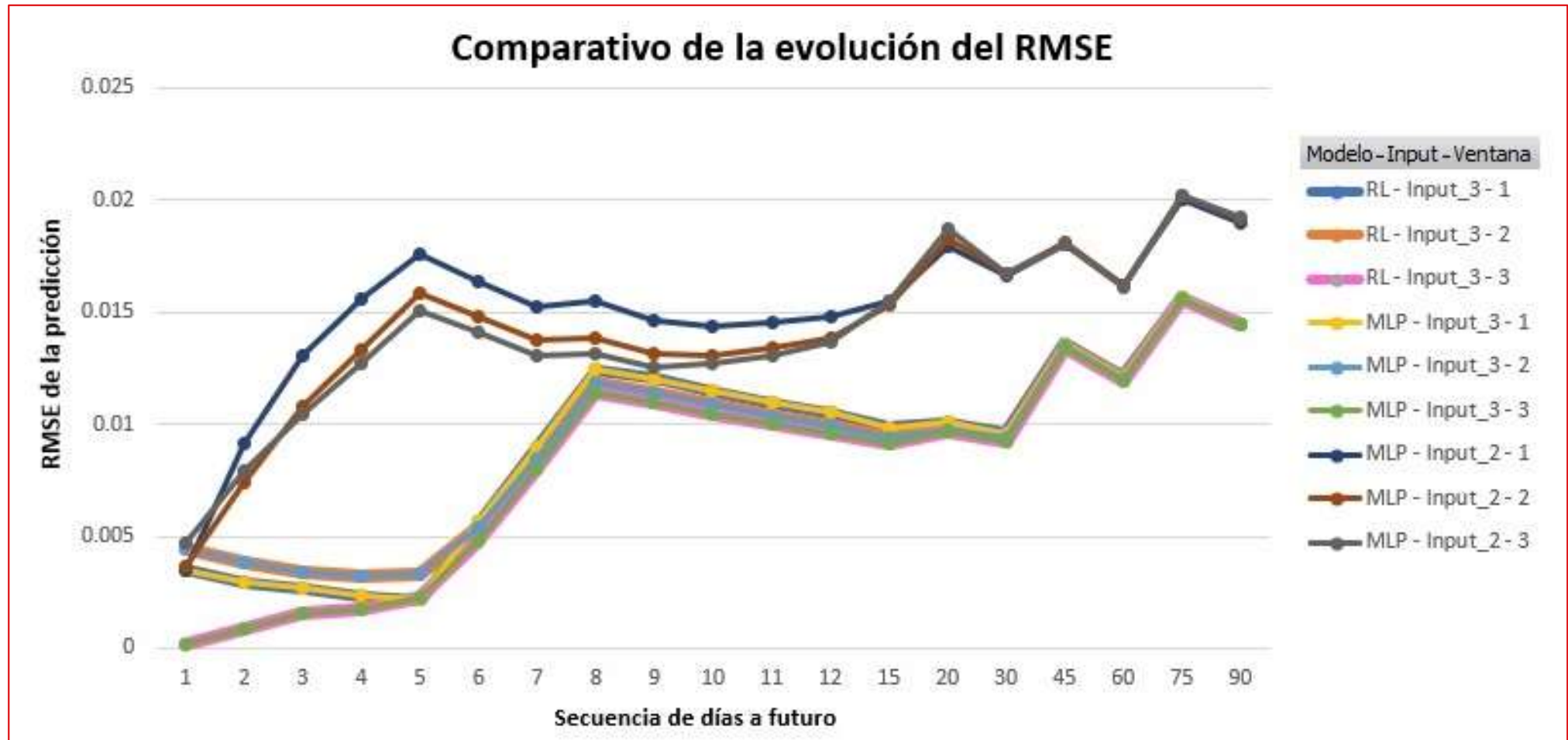
- Ventanas:
  - 2 días
- Predicción:
  - 90 días futuros

	Fecha	SRIndex	Entrada: Ventanas (2 días)	Predicción	Validación
Datos de Entrenamiento	01/01/2019	SR <sub>0</sub>			SR <sub>0</sub>
	02/01/2019	SR <sub>1</sub>			SR <sub>1</sub>
	03/01/2019	SR <sub>2</sub>	V <sub>0</sub> = SR <sub>0</sub> , SR <sub>1</sub>	y_train <sub>0</sub>	SR <sub>2</sub>
	04/01/2019	SR <sub>3</sub>	V <sub>1</sub> = SR <sub>1</sub> , SR <sub>2</sub>	y_train <sub>1</sub>	SR <sub>3</sub>
	05/01/2019	SR <sub>4</sub>	V <sub>2</sub> = SR <sub>2</sub> , SR <sub>3</sub>	y_train <sub>2</sub>	SR <sub>4</sub>
	06/01/2019	SR <sub>5</sub>	V <sub>3</sub> = SR <sub>3</sub> , SR <sub>4</sub>	y_train <sub>3</sub>	SR <sub>5</sub>
	( ... )	( ... )	( ... )	( ... )	( ... )
	30/12/2020	SR <sub>650</sub>	V <sub>648</sub> = SR <sub>648</sub> , SR <sub>649</sub>	y_train <sub>648</sub>	SR <sub>650</sub>
Datos de Test	31/12/2020	SR <sub>651</sub>	V <sub>649</sub> = SR <sub>649</sub> , SR <sub>650</sub>	y_train <sub>649</sub>	SR <sub>651</sub>
	01/01/2021	SR <sub>652</sub>	V_test <sub>0</sub> = SR <sub>650</sub> , SR <sub>651</sub>	y_test <sub>0</sub>	SR <sub>652</sub>
	02/01/2021	SR <sub>653</sub>	V_test <sub>1</sub> = SR <sub>651</sub> , y_test <sub>0</sub>	y_test <sub>1</sub>	SR <sub>653</sub>
	03/01/2021	SR <sub>654</sub>	V_test <sub>2</sub> = y_test <sub>0</sub> , y_test <sub>1</sub>	y_test <sub>2</sub>	SR <sub>654</sub>
	04/01/2021	SR <sub>655</sub>	V_test <sub>3</sub> = y_test <sub>1</sub> , y_test <sub>2</sub>	y_test <sub>3</sub>	SR <sub>655</sub>
	05/01/2021	SR <sub>656</sub>	V_test <sub>4</sub> = y_test <sub>2</sub> , y_test <sub>3</sub>	y_test <sub>4</sub>	SR <sub>656</sub>
	06/01/2021	SR <sub>657</sub>	V_test <sub>5</sub> = y_test <sub>3</sub> , y_test <sub>4</sub>	y_test <sub>5</sub>	SR <sub>657</sub>
	( ... )	( ... )	( ... )	( ... )	( ... )
	31/03/2021	SR <sub>741</sub>	V_test <sub>5</sub> = y_test <sub>88</sub> , y_test <sub>89</sub>	y_test <sub>89</sub>	SR <sub>741</sub>

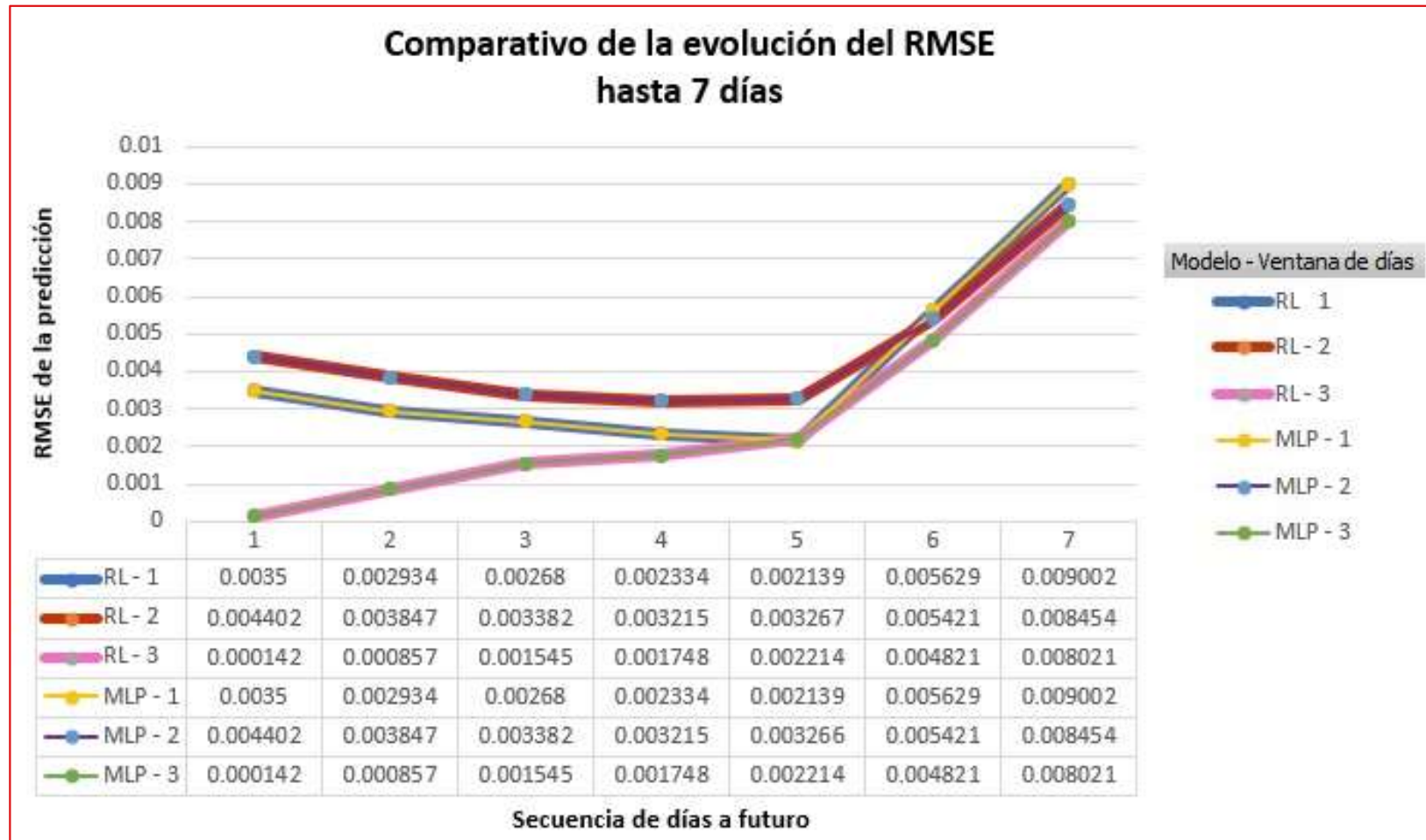
## Predicciones secuenciales - Experimentos

- **Regresión Lineal y MLP**
  - Input\_3: sólo las ventanas
- **MLP**
  - Input\_2: Mismos datos de los modelos matemáticos + las ventanas
- Ventanas de **1, 2 y 3** días
- Predecir hasta **90** días futuros

# Resultados



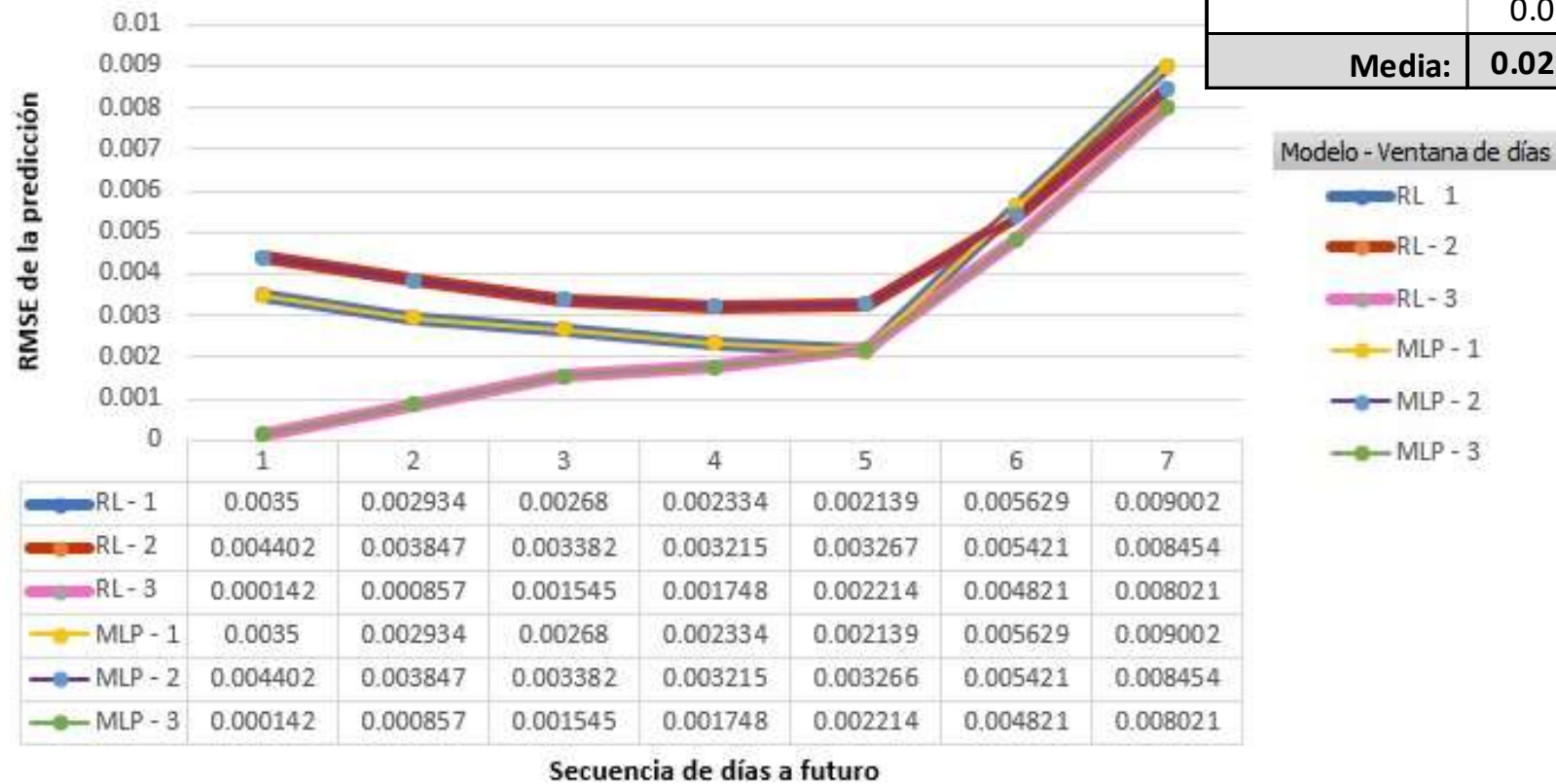
# Resultados





# Resultados

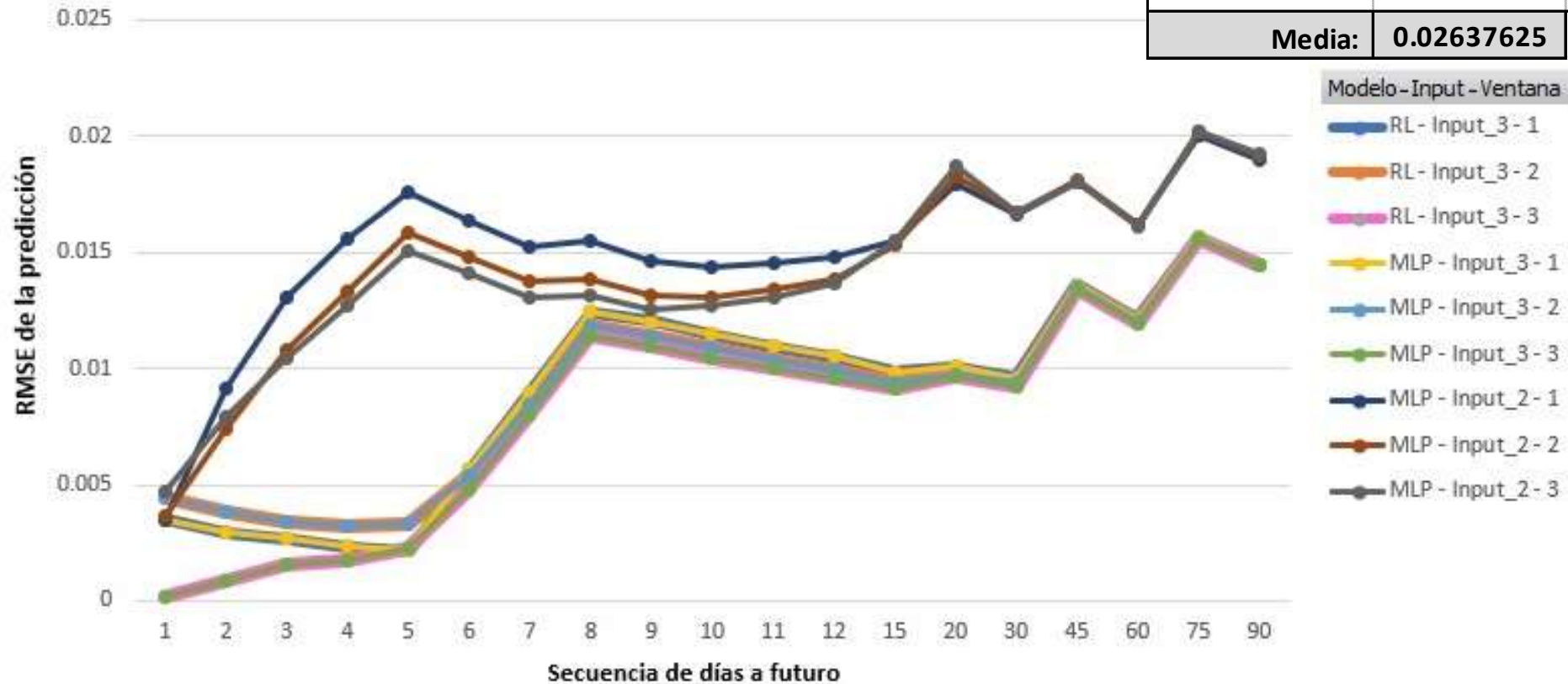
Comparativo de la evolución del RMSE hasta 7 días



Modelo	RMSE	Origen
Coello y Boyle	0.016876	Datos de Suelo
	0.038538	Datos de Satelites
You	0.019935	Datos de Suelo
	0.030156	Datos de Satelites
Media:	0.02637625	

# Resultados

Comparativo de la evolución del RMSE



# Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- Conclusión

## Contenido

- Introducción
- Análisis Exploratorio y Preprocesamiento
- Experimentos Realizados
- Predicciones Secuenciales
- **Conclusión**

## Conclusión

- Resultados muy **prometedores**
  - Regresión Lineal, MLP, Random Forest y Árbol de Decisión
  - LSTM: no tan eficaz
- Predicciones Secuenciales:
  - Error muy bajo **hasta 5 días**: Input\_3
  - Error más bajo que los modelos matemáticos hasta 90 días

## Conclusión

- Contexto de los datos:
  - **Procedencia geográfica** de las mediciones:
    - Jaén no sufre efectos severos del soiling
    - Bajos niveles → estimativas menos precisas <sup>6</sup>
  - Conjunto de datos relativamente **pequeño**
- Parámetros Ambientales: **más investigación**
  - Más **errores** en datos de satélite <sup>7</sup>
  - **Ruido** para el predictor

## Líneas de Futuro

- Construir modelos con **datos ambientales de suelo** y **compararlos** a los resultados obtenidos en este trabajo
- Validar los modelos con mediciones de **SR** procedentes de **otras ubicaciones**, principalmente que sufran de manera **más severa** los efectos del **soiling** y con **más mediciones** disponibles.



## Líneas de Futuro

- Extender la **colaboración** con el CEACTEMA/UJA:
  - **nuevos modelos** matemáticos actualmente desarrollados por estos investigadores
- Publicación de un **artículo científico**
- Utilización de **pronósticos meteorológicos** futuros en las predicciones secuenciales como **entrada externa**.

## Referencias

1. Red Electrica de España, «*La demanda de energía eléctrica de España aumenta un 4,8 % en marzo*», (2021).
2. A. Comerio, «*Avaliação do impacto de sujidade de módulos fotovoltaicos na geração de energia elétrica*», (2019).
3. J. G. Bessa, L. Micheli, F. Almonacid, y E. F. Fernández, «*Monitoring photovoltaic soiling: assessment, challenges, and perspectives of current and potential strategies*», (2021).
4. M. Coello y L. Boyle, «*Simple Model for Predicting Time Series Soiling of Photovoltaic Panels*», (2019).
5. S. You, Y. J. Lim, Y. Dai, y C. H. Wang, «*On the temporal modelling of solar photovoltaic soiling: Energy and economic impacts in seven cities*», (2018).
6. Å. Skomedal y M. Deceglie, «*Combined Estimation of Degradation and Soiling Losses in Photovoltaic Systems*», (2020).
7. J. M. Carmona, P. Gupta, D. F. Lozano-García, A. Y. Vanoye, F. D. Yépez, y A. Mendoza, «*Spatial and Temporal Distribution of PM<sub>2.5</sub> Pollution over Northeastern Mexico: Application of MERRA-2 Reanalysis Datasets*», (2020).



Uso de Machine Learning en la Determinación de Niveles de Suciedad en Sistemas Fotovoltaicos a partir de parámetros ambientales

***¡Muchas Gracias!***

