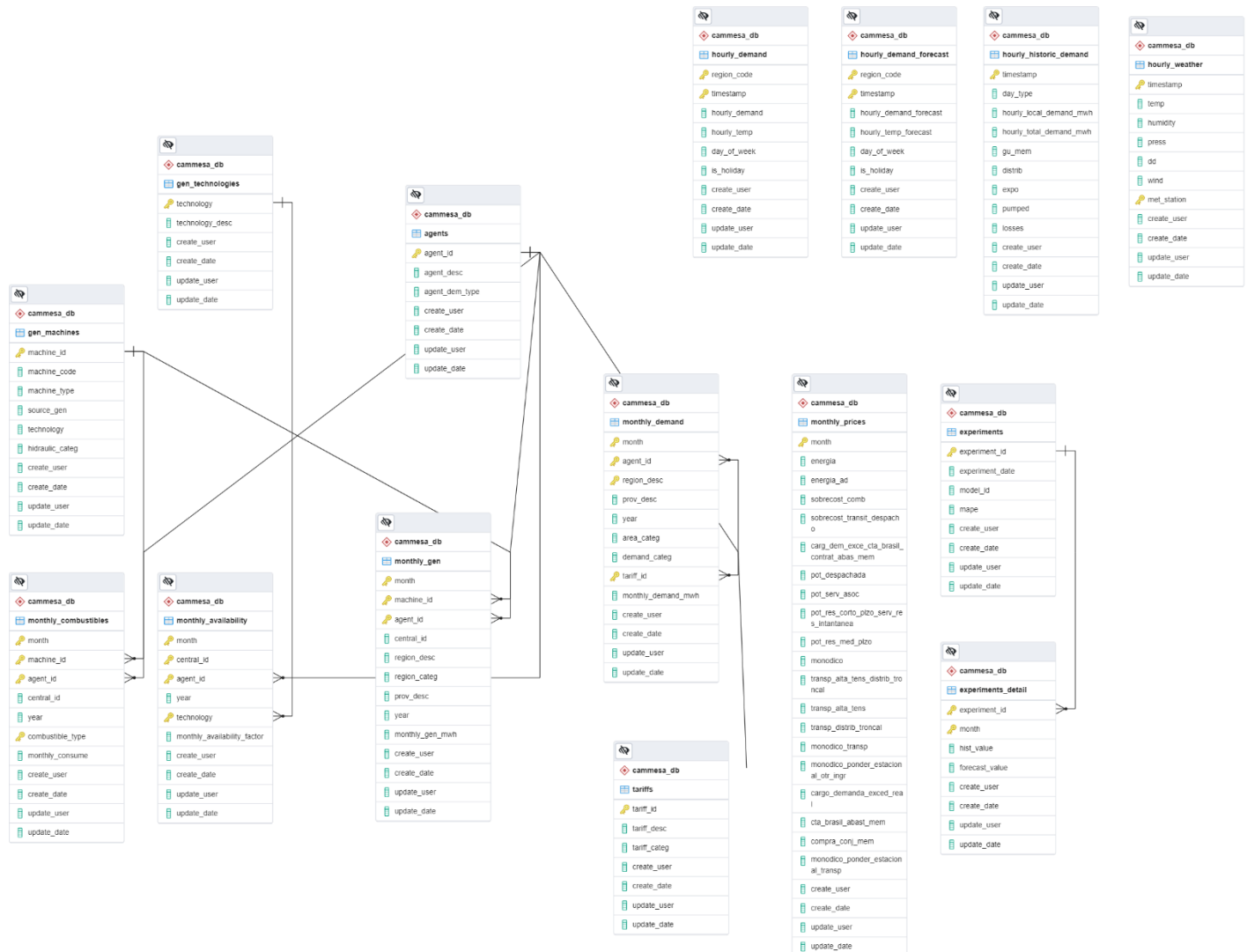


### Información sobre el documento:

<b>Año: 2023</b>	Especialización en Ciencia de Datos (2021-B)	Materia: Trabajo Final
<b>Draft</b>		Alumna: Carla Olmo

## Trabajo Técnico Realizado

Se diseñó una estructura de datos para persistir los datos input de los modelos y los resultados de los pronósticos. Previendo la evolución de la solución (ej: pronósticos diarios u horarios, incorporación de otras variables a la predicción, etc), el modelo contempla otras entidades disponibles en la fuente de datos.



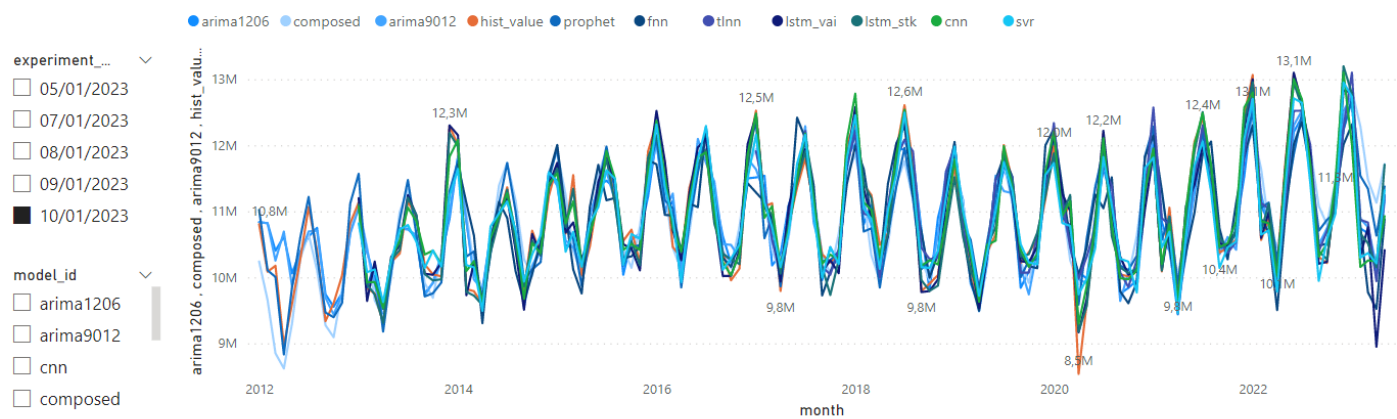
Desarrollo de autómatas/procesos en Airflow y Python que obtienen los datos, los persisten en una base de datos Postgres, re-entrenan los modelos (si es necesario) y persisten los resultados de estos pronósticos.

[https://github.com/colmo786/energy\\_demand/blob/main/python/ED\\_monthly\\_dag.ipynb](https://github.com/colmo786/energy_demand/blob/main/python/ED_monthly_dag.ipynb)

Todos los modelos han sido tunneados en sus hiperparámetros. Y sobre los hiperparámetros obtenidos, se corrieron procesos de cross validation para validar el nivel de performance de cada modelo.

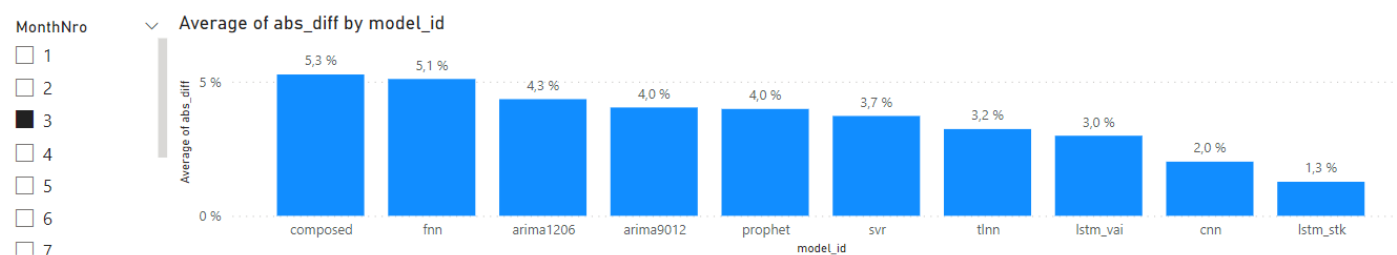
Diseño de gráficos de soporte a la selección del mejor estimador para cada mes.

arima1206 , composed , arima9012 , hist\_value, prophet , fnn , tlnn , lstm\_vai , lstm\_stk , cnn and svr by month



model_id	mape	month	hist_value	composed	composed_diff	arima1206	arima1206_diff	arima9012	arima9012_diff	prophet	prophet_diff	fnn	fnn_diff	tlnn	tlnn_diff
lstm_stk	1,01	2023-05		11716150	0,0 %	11169815	0,0 %	11226043	0,0 %	11376490	0,0 %	10879337	0,0 %	11249167	0,00
lstm_vai	1,46	2023-04		11130556	0,0 %	10025176	0,0 %	10166725	0,0 %	10633358	0,0 %	9522944	0,0 %	9942702	0,00
cnn	1,47	2023-03		11477416	0,0 %	10667986	0,0 %	10761321	0,0 %	11341879	0,0 %	9773840	0,0 %	10830482	0,00
svr	2,75	2023-02		12287854	0,0 %	10741614	0,0 %	10886256	0,0 %	11537940	0,0 %	10771677	0,0 %	10872887	0,00
prophet	2,82	2023-01		12829339	0,0 %	12676345	0,0 %	12758672	0,0 %	12717236	0,0 %	12316742	0,0 %	13100830	0,00
tlnn	2,96	2022-12		12415615	0,0 %	12249976	0,0 %	12517346	0,0 %	12233601	0,0 %	12012359	0,0 %	12270135	0,00
arima9012	3,17	2022-11	11.319.395,14	11445935	-1,1 %	11191573	1,1 %	11047598	2,4 %	10981185	3,0 %	11228695	0,8 %	11077382	0,02
arima1206	3,34	2022-10	10.217.095,55	10800307	-5,7 %	10687707	-4,6 %	10743190	-5,1 %	10871309	-6,4 %	10511795	-2,9 %	10698981	-0,05
composed	3,59														

month	composed_diff	arima1206_diff	arima9012_diff	prophet_diff	fnn_diff	tlnn_diff	lstm_vai_diff	lstm_stk_diff	cnn_diff
2022-03	0,6 %	0,3 %	1,2 %	-2,3 %	1,8 %	-0,01	-1,0 %	-2,2 %	0,5 %
2021-03	5,7 %	5,3 %	6,0 %	1,2 %	5,3 %	0,02	2,5 %	1,3 %	2,3 %
2020-03	7,9 %	6,7 %	5,4 %	3,2 %	5,2 %	0,06	1,3 %	-0,9 %	-0,1 %
2019-03	-1,4 %	2,1 %	-1,2 %	-4,4 %	1,5 %	-0,02	-1,9 %	0,2 %	-2,2 %
2018-03	8,1 %	1,3 %	4,1 %	4,5 %	3,1 %	0,03	3,1 %	0,4 %	2,0 %
2017-03	6,4 %	8,5 %	5,2 %	1,9 %	3,4 %	0,05	2,1 %	1,7 %	0,4 %
2016-03	0,1 %	-6,7 %	-7,7 %	-5,6 %	-4,5 %	0,00	-2,5 %	-3,5 %	-2,8 %
2015-03	9,9 %	9,5 %	9,5 %	8,4 %	12,4 %	0,00	6,2 %	1,8 %	3,4 %
2014-03	-3,7 %	-2,9 %	-1,8 %	-6,2 %	-8,6 %	0,00	-4,4 %	0,7 %	-4,7 %
2013-03	1,1 %	-2,3 %	-1,4 %	-4,6 %	0,0 %	0,00	-5,0 %	0,1 %	-1,8 %
2012-03	13,0 %	-2,2 %	-0,8 %	1,7 %	0,0 %	0,00	0,0 %	0,0 %	0,0 %



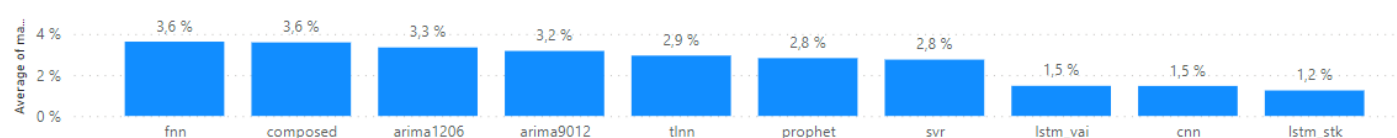
## Conclusiones

CONTEXTO: de las entrevistas con especialistas, el error promedio mensual de sus actuales estimaciones es de 5%.

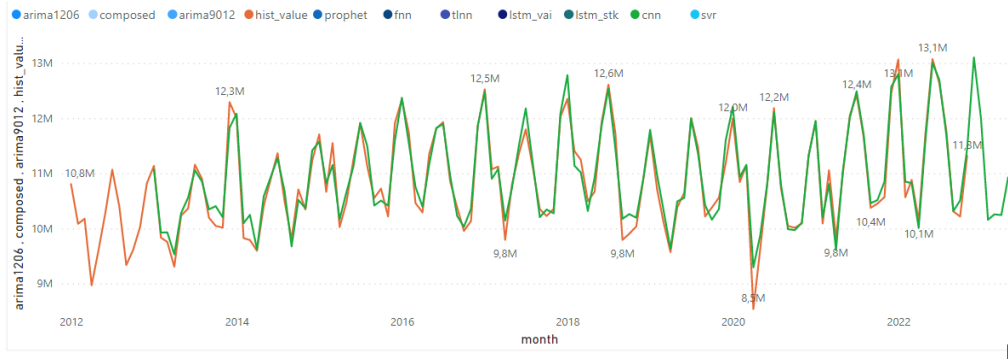
La utilización de modelos de aprendizaje basado en los datos (redes neuronales y SVM) son técnicas performantes para la predicción de series de tiempo univariadas.

Presentan mejor error porcentual absoluto medio que los modelos de regresión lineal (polinómico, arima, prophet)

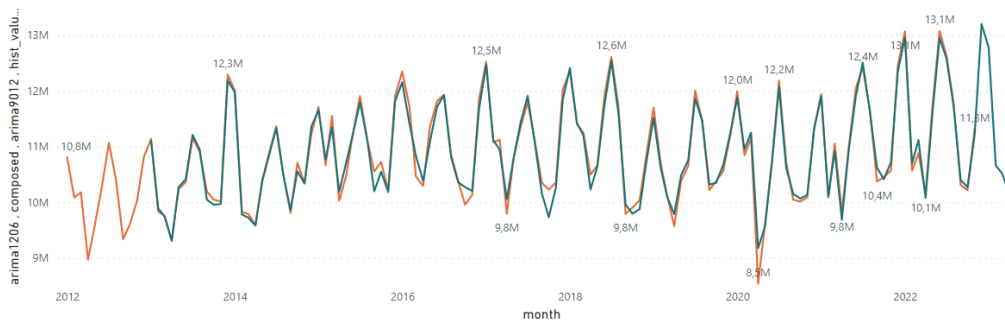
Promedio de MAPE entre todos los Experimentos



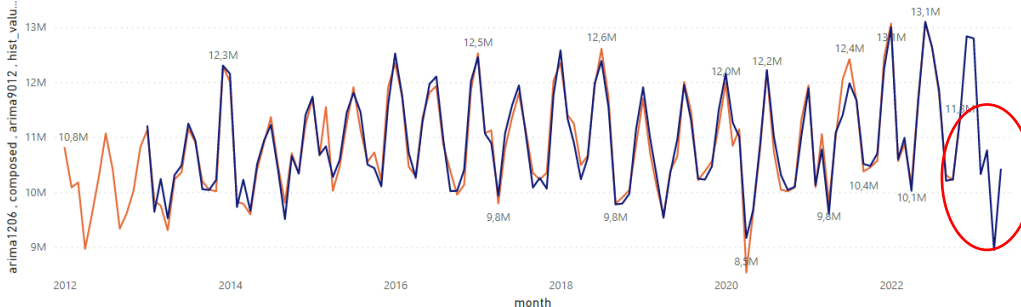
La red neuronal convolucional presenta la mayor capacidad de aprendizaje, y el menor tiempo de ejecución.



Es muy fácil caer en el overfitting con las redes neuronales.



Sin embargo, los pronósticos no siempre son aceptables, esto obliga a mantener un conjunto de modelos siempre entrenados y presentar al decisor herramientas para comparar pronósticos y seleccionar utilizando conocimiento experto en el tema.



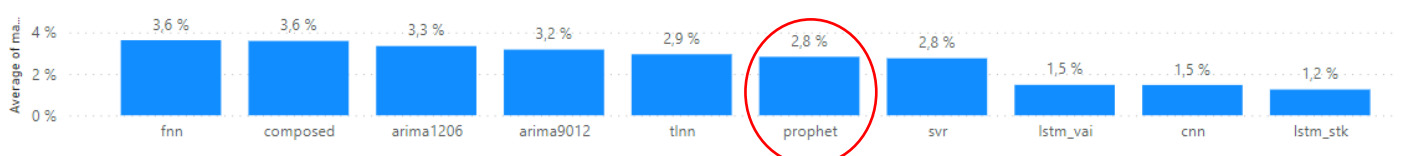
El componente aleatorio del proceso de entrenamiento de las redes neuronales hace que los resultados del entrenamiento difieran entre experimento y experimento. Ante esta problemática, se me ocurre la opción de ejecutar varios procesos de entrenamiento y tomar el modelo con menor MAPE. Sin embargo, esto no asegura que el pronóstico sea potable. Aún no encuentro propuesta contra esta situación.

Observar distintos los mape para distintos experimentos del mismo modelo (y mismos datos):

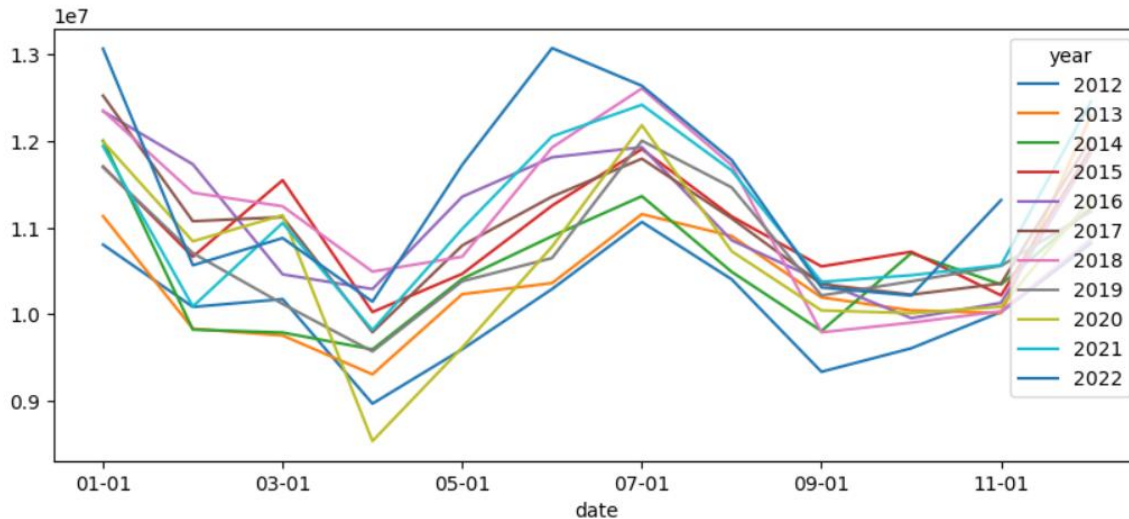
model_id	mape
fnn	3,29
fnn	3,56
fnn	4,03

El modelo de predicción de facebook performa bien y es fácil de utilizar.

Promedio de MAPE entre todos los Experimentos

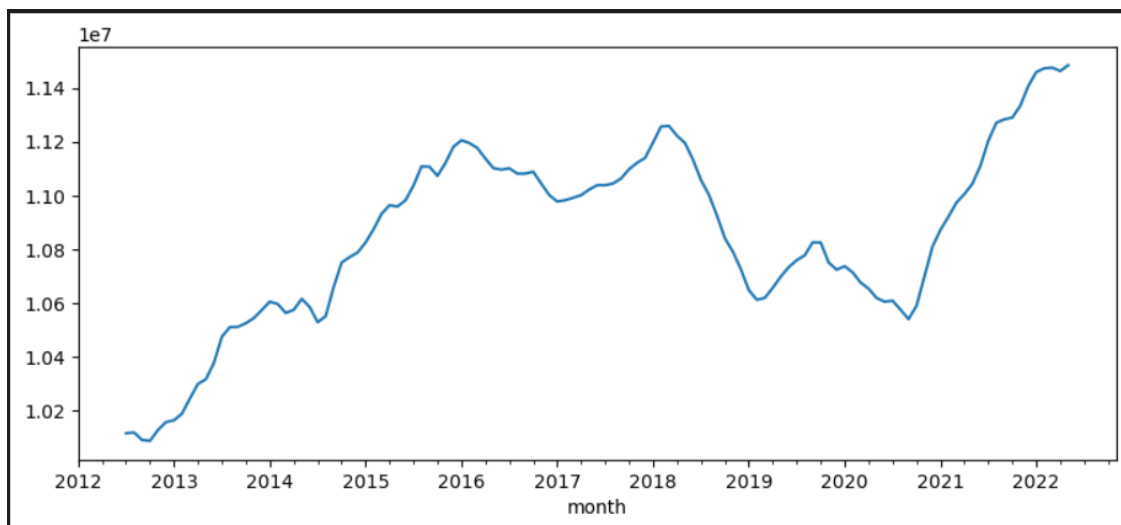


Esto muestra estacionalidad anual, no?

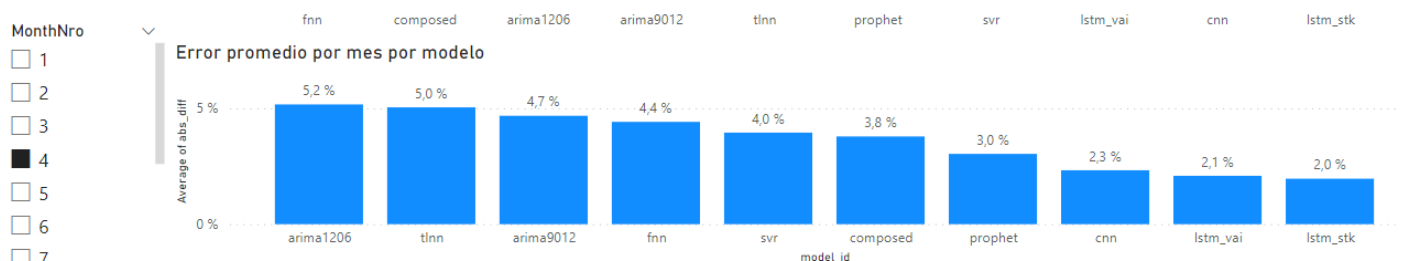


Conclusión general: no se puede seleccionar un único modelo dado que los modelos con mejor performance, al contar con procesos aleatorios, aleatoriamente puede presentar mala performance en un experimento. La conclusión a la que se llega es que se deben mantener todos los modelos y presentar los resultados de manera de facilitar la selección por expertos del valor del pronóstico para el mes a estimar.

Gráficos como el de la tendencia de la serie pueden mostrar si el valor está alcista o a la baja.



Comparar qué performance obtuvo en los 10 años anteriores cada modelo, para cada mes particular. Ejemplo Abril:



month	composed_diff	arima1206_diff	arima9012_diff	prophet_diff	fnn_diff	tlmn_diff	lstm_vai_diff	lstm_stk_diff	cnn_diff
2022-04	-3,0 %	1,5 %	0,7 %	-1,2 %	6,3 %	0,02	1,2 %	0,6 %	1,3 %
2021-04	-2,5 %	3,8 %	2,1 %	-0,9 %	1,7 %	0,04	2,2 %	1,3 %	2,1 %
2020-04	-15,9 %	-18,0 %	-16,8 %	-12,3 %	-17,5 %	-0,17	-7,4 %	-7,5 %	-8,9 %
2019-04	-3,2 %	-0,0 %	0,9 %	-2,7 %	0,8 %	0,00	0,4 %	-2,3 %	-0,6 %
2018-04	5,2 %	4,4 %	2,1 %	6,2 %	3,1 %	0,05	2,5 %	2,5 %	1,7 %
2017-04	-2,6 %	-5,7 %	-3,8 %	-0,7 %	-3,6 %	-0,03	-1,5 %	-2,7 %	-3,6 %
2016-04	1,8 %	1,3 %	2,2 %	4,3 %	-1,0 %	0,00	0,3 %	-0,9 %	-1,0 %
2015-04	-0,3 %	-1,1 %	-1,1 %	1,5 %	2,7 %	0,00	-2,5 %	-1,6 %	-1,4 %
2014-04	-2,5 %	-1,8 %	-2,3 %	0,7 %	3,0 %	0,00	-0,7 %	0,2 %	-0,3 %
2013-04	-0,8 %	-0,6 %	0,2 %	1,4 %	0,0 %	0,00	-2,4 %	-0,0 %	-2,4 %
2012-04	3,9 %	-18,4 %	-19,2 %	1,5 %	0,0 %	0,00	0,0 %	0,0 %	0,0 %

## Título: Pronóstico de Demanda Eléctrica Argentina de Mediano Plazo

### 1. Introducción

Incorporar a la intro las conclusiones de los experimentos.

### 2. Estado de la Cuestión

### 3. Definición del Problema

### 4. Justificación del estudio

### 5. Alcances del trabajo y limitaciones

### 6. Hipótesis

### 7. Desarrollo del proyecto

- Descripción de la fuente de datos y del proceso de obtención de los datos. Presentación del DER de la solución.
- Breve intro de Series de tiempo.
- Modelos de regresión para series de tiempo univariados evaluados. Teoría detrás de cada uno
- Discusión de los resultados obtenidos con cada modelo, tuneo de hiperparámetros y resultados del cross validation.
- Conclusiones generales de la performance de los modelos.
- Propuesta y discusión de gráficos e indicadores que den soporte a una elección del mejor pronóstico mensual utilizando conocimiento experto.