

# Aprendizaje Automático

## Proyecto Final

**Arturo Márquez-Flores**

Maestría en Inteligencia Artificial

Universidad Veracruzana

CIIA – Centro de Investigación en Inteligencia Artificial

Sebastián Camacho No 5, Xalapa, Ver., México 91000

<https://github.com/arturomf94/marsam>

28 de Junio del 2019

### Resumen

En este trabajo se exploran algunos de los conceptos y herramientas vistos en la clase de aprendizaje automático. El proyecto tiene como propósito implementar dos modelos de aprendizaje para un conjunto de series de tiempo.

## Datos

Los datos utilizados son series de tiempo relacionadas con un sistema de paneles solares instalados en una casa. Como se muestra en [este colab](#), las observaciones disponibles comprenden un período del 2012 al 2019, con frecuencia diaria. Las variables disponibles en la base son el uso (kW), la generación (kW), la generación por la red de electricidad (kW), la generación por los paneles solares (kW), la capacidad instalada, el voltaje de la línea 1 y línea 2 del hogar, la frecuencia de la línea 1 y 2, y el uso del calentador de agua. Sin embargo, en este trabajo nos concentraremos en el análisis del uso y la generación sin utilizar las otras variables. Las siguientes figuras muestran los valores de la serie para algunas de estas variables.

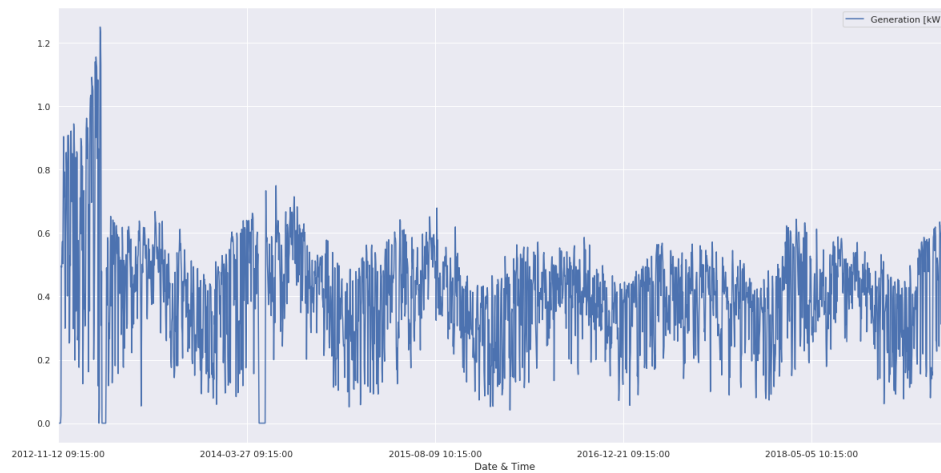


Figura 1: Generación

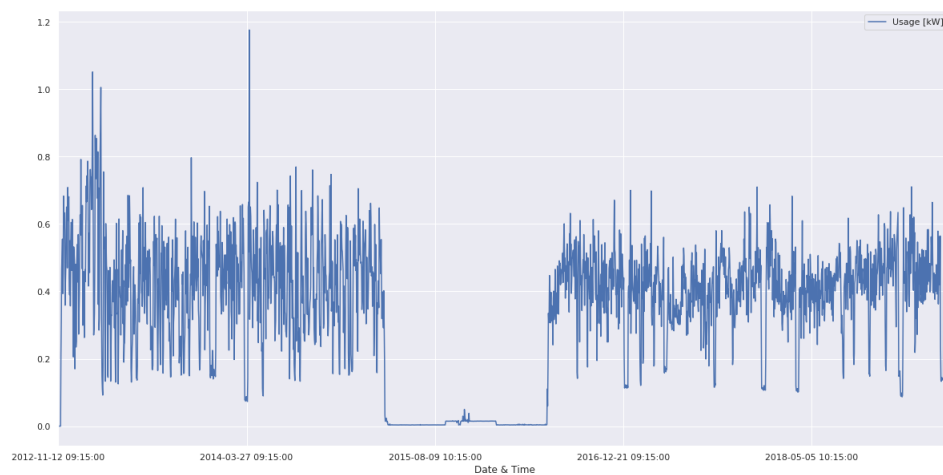


Figura 2: Uso

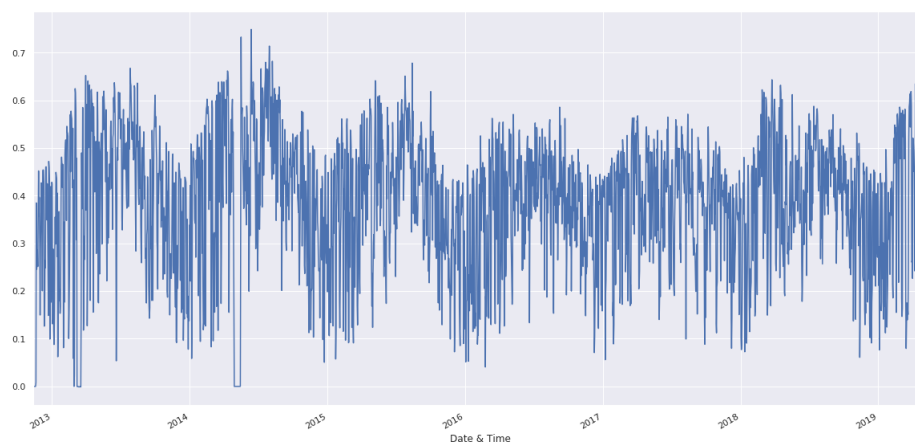


Figura 3: Generación por Páneles

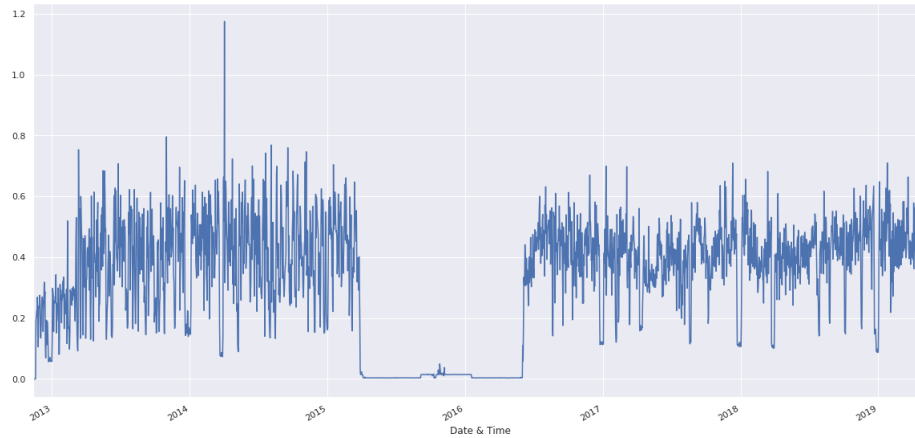


Figura 4: Generación por Red de Electricidad

En la Figura 2 podemos observar que por un período de tiempo del 2014 al 2016 se presentaron observaciones atípicas. Esto es porque durante ese período la casa no estuvo habitada. Por otro lado, es evidente por la Figura 3 que la generación por los paneles sigue un comportamiento cíclico. Para el caso del uso, mostramos que, como parte del preprocesamiento de los datos, es posible hacer una descomposición de la serie de tiempo en un elemento de tendencia, un elemento estacional y un residuo:

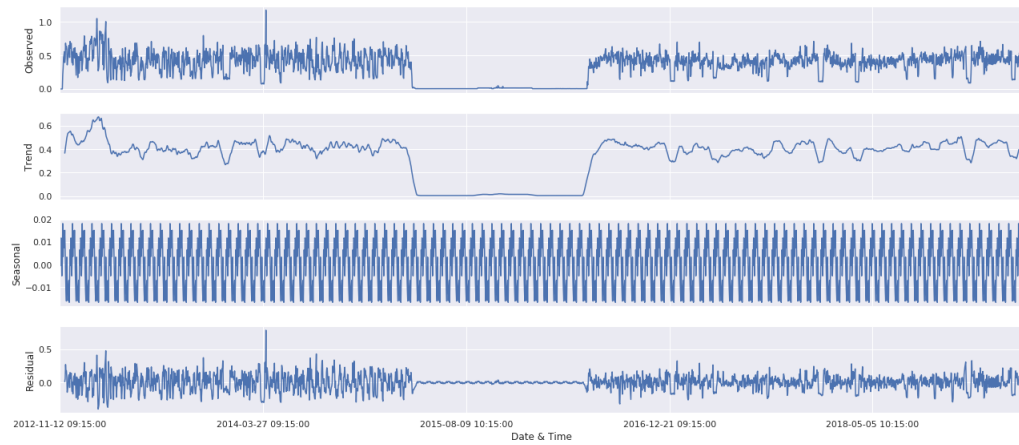


Figura 5: Descomposición

## Modelos

Los modelos de predicción utilizados fueron dos: el primero es un modelo ARIMA, cuyos parámetros fueron elegidos mediante un proceso de optimización, y un modelo de una red neuronal recurrente y autoregresiva, como se propone en [este artículo](#). El primer modelo puede replicarse en [este colab](#) y el segundo modelo puede replicarse en [este colab](#).

## Resultados

Para el primer modelo podemos observar en la Figura 6, donde se observa el modelo ajustado con sus parámetros optimizados. Este modelo presenta un error cuadrático medio de 0.03.

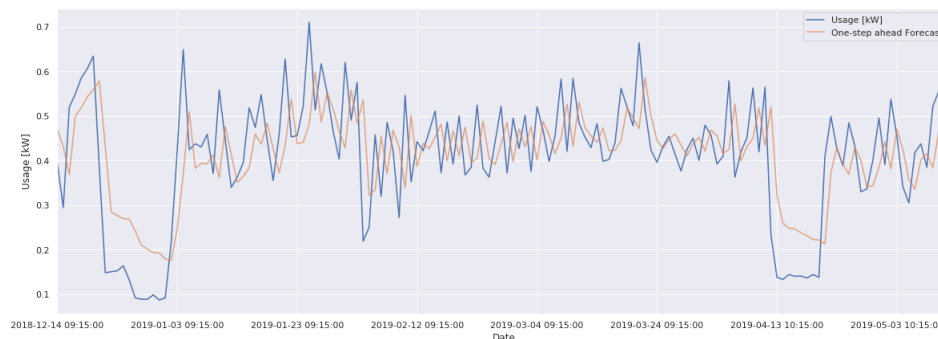


Figura 6: ARIMA

Sin embargo, aunque el modelo ajusta muy bien los datos, su predicción es, de hecho, pobre ya que no contiene la información sobre la variabilidad de la serie. Como se observa en la Figura 7, la predicción fuera del conjunto de datos de entrenamiento tiende a la media de la serie de tiempo.

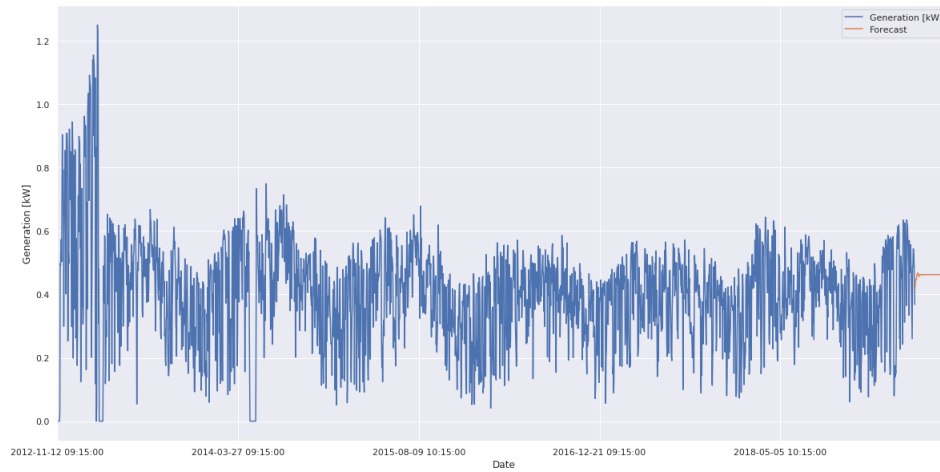


Figura 7: ARIMA Predicción

El segundo modelo se muestra en la Figura 8. Sin embargo, para este modelo en particular, utilizar la serie de tiempo con observaciones por hora fue más conveniente. En ese caso el modelo se puede observar en la Figura 9. Este modelo tiene un error cuadrático medio de 0.065.

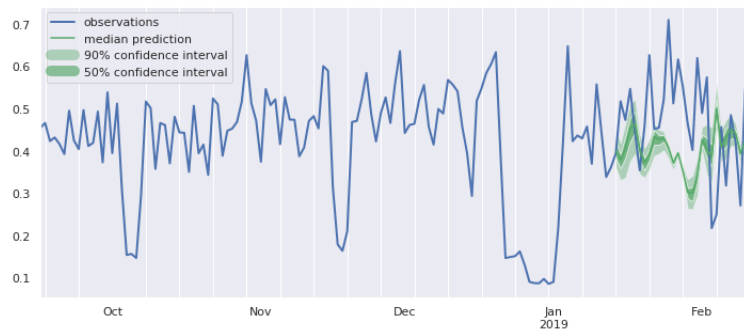


Figura 8: Red Neuronal

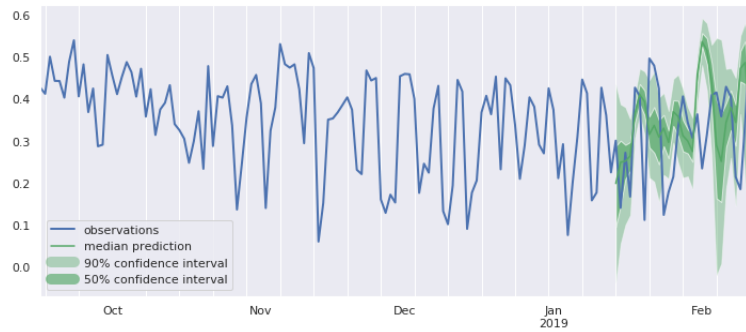


Figura 9: Red Neuronal 2

Por otro lado, es evidente que este modelo puede hacer mejor uso de la información de las fluctuaciones en los valores para hacer predicciones fuera del conjunto de datos de entrenamiento.

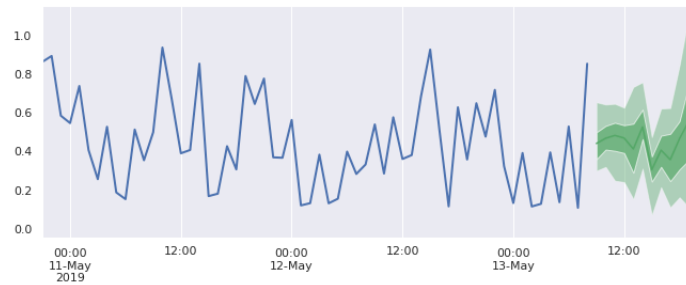


Figura 10: Predicción de Red Neuronal

Aunque ambos modelos muestran una capacidad para ajustar los datos, este estudio no es conclusivo en cuanto a la validez de sus predicciones. La primera observación es que la implementación de ellos es todavía muy rudimentaria. Una manera de complementarlos sería poder utilizar datos *dummy* para que el modelo tenga más facilidad en procesar la información de la periodicidad de la serie.