

## NUESTRO MODELO – ARIMA

Para predecir en esta serie temporal de contaminantes en el aire, hemos utilizado el modelo ARIMA:

**ARIMA** tiene tres componentes: AR (término Autorregresivo), I (término diferencial) y MA (término promedio móvil). Entre ellos:

**AR:** El valor pasado usado para predecir el siguiente valor. El término AR se define por el parámetro "p" en ARIMA. El valor de p está determinado por el gráfico PACF.

**MA:** El número de errores de predicción anteriores utilizados para predecir valores futuros. El parámetro q en ARIMA representa el elemento MA. El gráfico ACF se utiliza para identificar el valor "q" correcto.

**Tiempos de diferencia:** El número de veces que se realiza una operación de diferencia en una serie para que sea estable. Se pueden usar pruebas como ADF y KPSS para determinar si la secuencia es estacionaria y ayudar a identificar el valor d.

El **dataset** utilizado es un csv de Alcobendas que muestra las partículas contaminantes del aire. Se ha transformado y el resultado es una serie temporal con datos cada hora de 9 variables diferentes. Se han eliminado los nan antes de pasárselo al modelo.

Previo paso a entrenar el modelo, hemos dividido **train y test** a través de slicing para alterar el orden.

Hemos utilizado el método **auto-arima** para ajustar automáticamente los parámetros para obtener un mejor rendimiento del modelo.

Una vez obtenidos éstos, llamamos al **modelo ARIMA** con esos parámetros y lo entrenamos.

Como output obtenemos el modelo entrenado, listo para ser guardado en un **pickle** y poder hacer predicciones futuras.

## DOCUMENTACION DE LA FUNCION

"""

Function description:

The function trains an ARIMA model for a given Time Series.

We split train and test using the parameter train\_size (from 0 to 1 where 0.5 is 50%)

Then we call the auto-arima function to get the best parameters

Finally we use those best params to train our ARIMA model

ARIMA Model:

We can combine both Autoregressive (AR) and Moving Average (MA) models together to create an ARMA model. The time series is regressed on the previous values and the previous errors.

For ARMA(p, q), p is the order of the AR process and q is the order of the MA process.

Parameters:

df (Pandas DataFrame type): is the given dataframe

column\_name (string): the chosen column to create new columns from

train\_size (float): value from 0 to 1 where 0.5 is 50% representing the % of our data that will be used to train the model

start\_p (int): initial value for p parameter. By default 1

start\_q (int): initial value for q parameter. By default 1

max\_p (int): max value for p parameter. By default 5

max\_q (int): max value for q parameter. By default 5

max\_d (int): max value for d parameter. By default 3

return: our trained ARIMA model

"""