

Práctica 7 ACSIC

Marc Llobera Villalonga Grupo 202

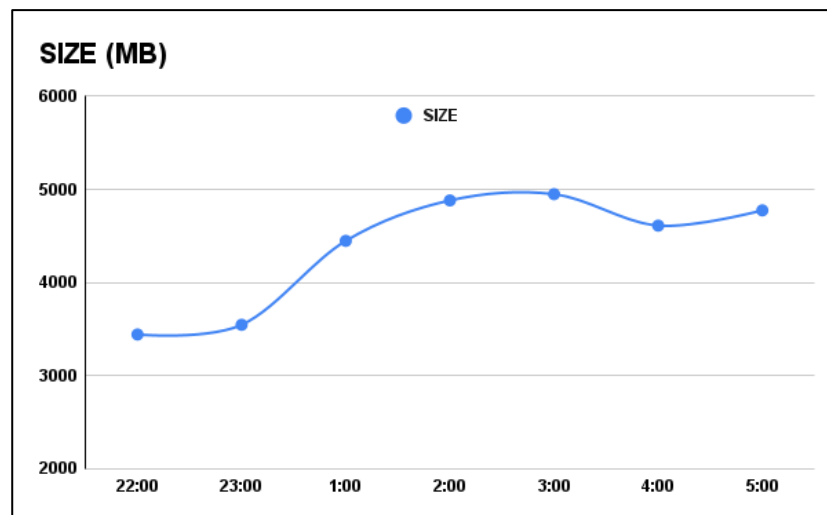
Una empresa de almacenamiento en la nube monitoriza la actividad de sus usuarios, es decir, se guarda la hora de acceso del cliente, el tamaño del fichero al que se ha accedido y la cantidad de información transmitida por unidad de tiempo (hacer uso de los datos del fichero “data.txt”, proporcionado en la práctica 6).

El director del departamento de informática de la empresa solicita calcular la cantidad de información transmitida por la red y el tamaño del fichero accedido para las 6 a.m. (recordar que la última hora monitorizada son las 5 a.m.).

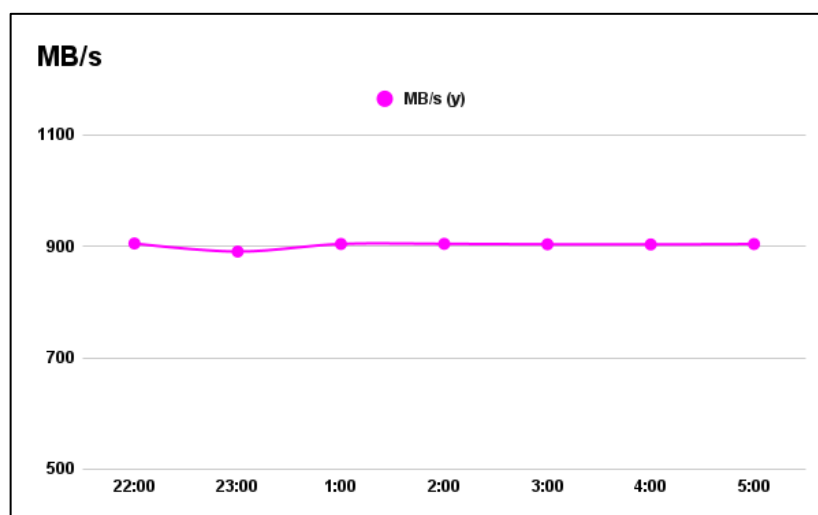
- ¿Qué patrón siguen los datos monitorizados? Proporciona una representación gráfica.

Para hacer un seguimiento claro de los datos, mostrarlo de manera gráfica y posteriormente predecir las características del fichero accedido a las 6 a.m. hemos recogido una media de los datos para cada una de las horas que tenemos.

SIZE	HOUR	MB/s
3442,893401	22:00	905,0959848
3546,785842	23:00	890,5476048
4447,973513	1:00	904,2294130
4881,890620	2:00	904,4456920
4949,224975	3:00	903,6393360
4611,401611	4:00	903,3558840
4775,229332	5:00	904,1064647



En el tamaño de los ficheros vemos que hay una tendencia a que el tamaño de estos sea mayor a medida que pasa el tiempo por lo que a primera vista podríamos decir que es un patrón de tendencia, pero sabemos por la práctica anterior que había algunos ficheros que tenían un tamaño muy superior al que vemos en estas medias por los que la variación entre estas es mínima comparado con el tamaño máximo que tiene algún fichero de los datos por lo que realmente lo más seguro es que el patrón sea estacionario ya que mantiene una media relativamente constante.



En cuanto al ancho de banda tenemos un claro patrón estacionario ya que no hay apenas variación en las medias recogidas por lo que es una media prácticamente constante durante todo el gráfico manteniéndose sobre los 900 MB/s.

- **Calcula los valores solicitados para las 6 a.m. haciendo uso de la regresión lineal, medias móviles (usar los 4 últimos valores) y suavizado exponencial (peso fijo del 60%).**

	SIZE	HOUR	MB/s
	3442,893401	22:00	905,0959848
	3546,785842	23:00	890,5476048
	4447,973513	1:00	904,2294130
	4881,890620	2:00	904,4456920
	4949,224975	3:00	903,6393360
	4611,401611	4:00	903,3558840
	4775,229332	5:00	904,1064647
Regresión Lineal	5326,127155	6:00	905,3540429
Medias Móviles (4 últimos valores)	4804,436635	6:00	903,8868442
Suavizado exponencial (60%)	4738,678600	6:00	903,8171432

Para la regresión lineal tenemos la formula:

$$y = a + bx$$

Donde:

$$b = \frac{\sum_{i=1}^n x_i \times y_i - n \times \bar{x} \times \bar{y}}{\sum_{i=1}^n x_i^2 - n \times \bar{x}^2}$$

$$a = \bar{y} - b \times \bar{x}$$

*Hay que tener en cuenta que el valor de las horas es equivalente a una enumeración que para evitar problemas la convertiremos en 1, 2, 3, 4, 5, 6, 8 donde 1 = 22:00 y 8 = 6:00.

En la media móvil tenemos:

$$f_{t+1} = \frac{y_t + y_{t-1} + \dots + y_{t-n+1}}{n},$$

Donde f_{t+1} equivale a y_{t+1} que es el valor que queremos predecir, por lo que el numerador de la fracción es la suma de los n (en nuestro caso 4) últimos valores antes de la predicción.

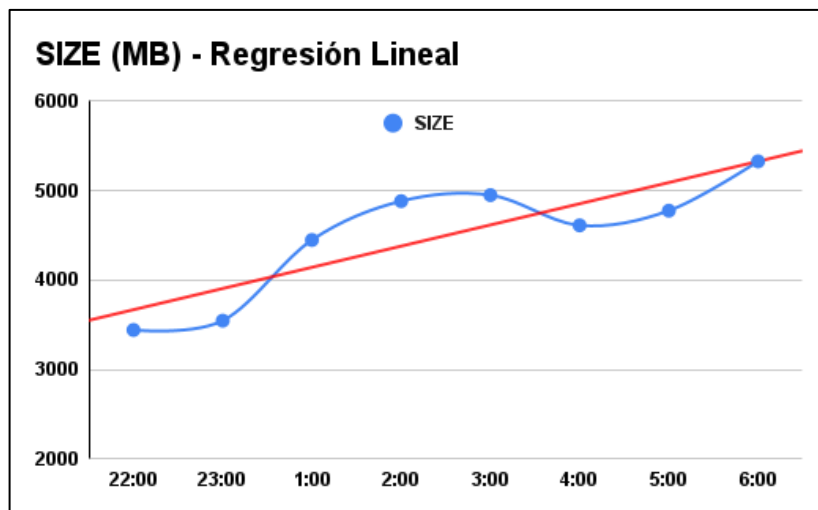
Finalmente, para el suavizado exponencial la ecuación es:

$$f_{t+1} = f_t + \alpha (y_{t+1} - f_t)$$

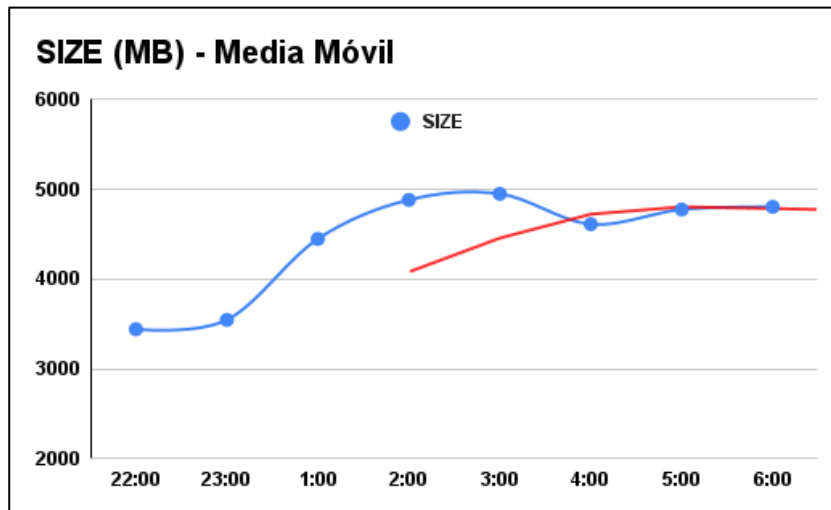
Donde tenemos que hacer la predicción de todos los valores desde el primero y la predicción del primer valor es siempre el mismo valor que este. Así nos queda esta tabla:

HOUR	SIZE	Predicción SIZE	MB/s	Predicción MB/s
22:00	3442.8934	3442.8934	905.0960	905.0960
23:00	3546.7858	3505.2289	890.5476	896.3670
1:00	4447.9735	4070.8757	904.2294	901.0844
2:00	4881.8906	4557.4846	904.4457	903.1012
3:00	4949.2250	4792.5288	903.6393	903.4241
4:00	4611.4016	4683.8525	903.3559	903.3832
5:00	4775.2293	4738.6786	904.1065	903.8171

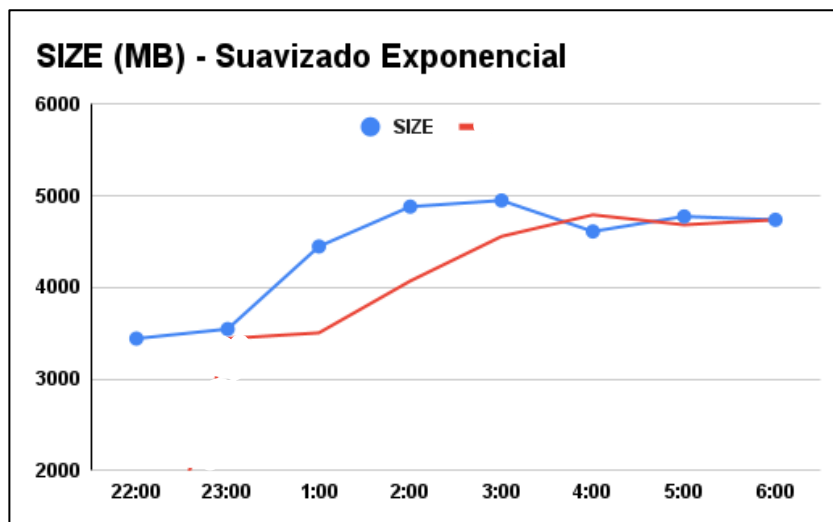
SIZE Regresión Lineal:



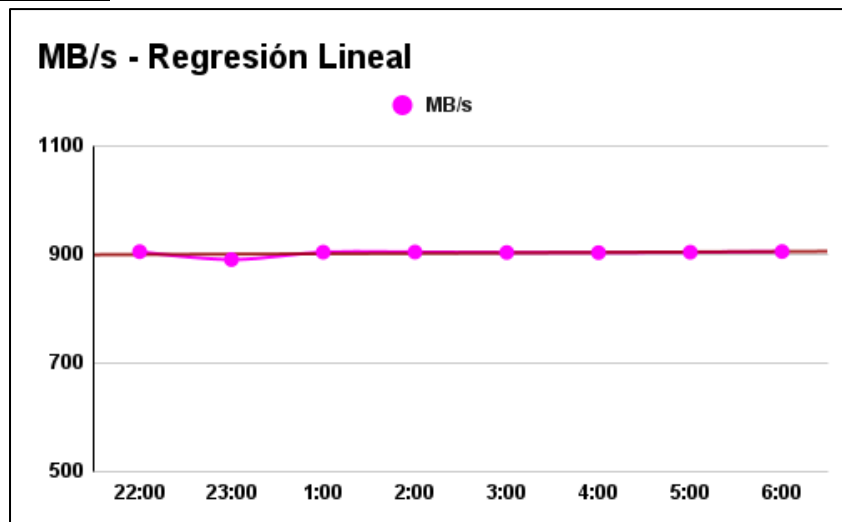
SIZE Media Móvil:



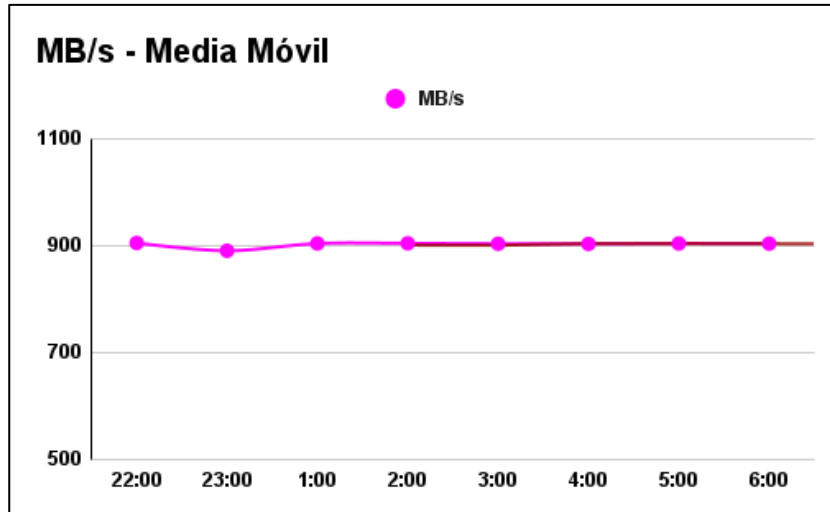
SIZE Suavizado Exponencial:



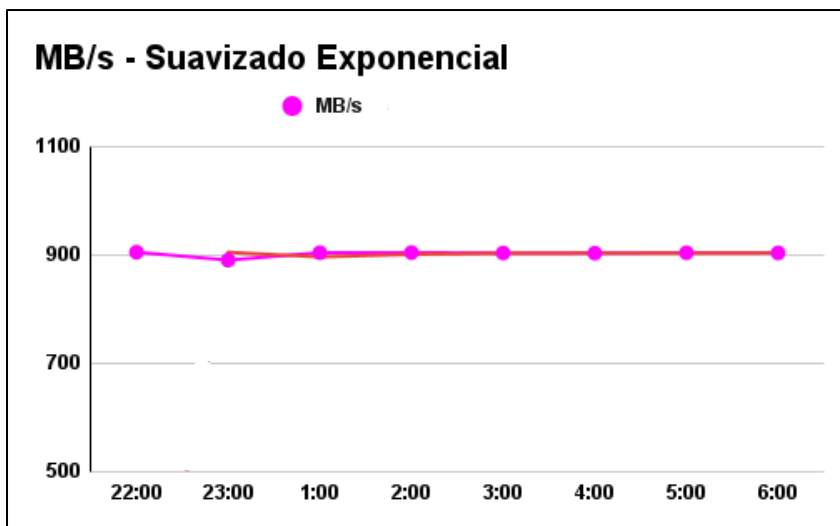
MB/s Regresión Lineal:



MB/s Media Móvil:



MB/s Suavizado Exponencial:



- **¿Qué técnica de predicción funciona mejor? ¿Por qué? ¿Cuál es la más adecuada para los datos con los que contamos?**

Al tener en cuenta que los MB/s tienen un patrón estacionario y que podríamos considerar que para el tamaño también tendríamos este patrón, y también sabiendo que tenemos un número suficiente de datos todo apunta a que la mejor predicción sería la media móvil con $n = 4$ para ambos casos ya que esta tiene una exactitud bastante alta para este tipo de gráficos y además no queremos realizar una predicción alejada en el tiempo sino que sería en la siguiente hora por lo que esta sería la más recomendable.

Otra opción sería, igualmente para ambos casos, el suavizado exponencial ya que, aunque es mejor para tendencias si que es bastante bueno para tiempos relativamente grandes y podríamos considerar que ahora los tenemos comparado con los datos originales que aunque también estaban separados por horas lo más seguro es que estaban recogidos entre ellos por diferencias de segundos o milésimas por lo que en este caso sería una buena opción.

Considerando todos los factores creo que la mejor técnica sería la de medias móviles.