# 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.**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**Tabla

Descripción generada automáticamentePara 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.

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.

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

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.

* **Tabla

  Descripción generada automáticamente 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%).**

Para la regresión lineal tenemos la formula:

Texto

Descripción generada automáticamenteDonde:

\*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.

Imagen que contiene objeto, antena, reloj

Descripción generada automáticamenteEn la media móvil tenemos:

Donde equivale a que es el valor que queremos predecir, por lo que el numerador de la fracción es la suma de los (en nuestro caso 4) últimos valores antes de la predicción.

Finalmente, para el suavizado exponencial la ecuación es:

Tabla

Descripción generada automáticamenteDonde 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:

Gráfico, Gráfico de líneas

Descripción generada automáticamenteSIZE Regresión Lineal:

Gráfico, Gráfico de líneas

Descripción generada automáticamenteSIZE Media Móvil:

Gráfico, Gráfico de líneas

Descripción generada automáticamenteSIZE Suavizado Exponencial:



Gráfico, Gráfico de líneas

Descripción generada automáticamenteMB/s Regresión Lineal:

MB/s Media Móvil:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Gráfico, Gráfico de líneas

Descripción generada automáticamenteMB/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 para ambos casos ya 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.