Universitat de les Illes Balears

Escola Politècnica Superior

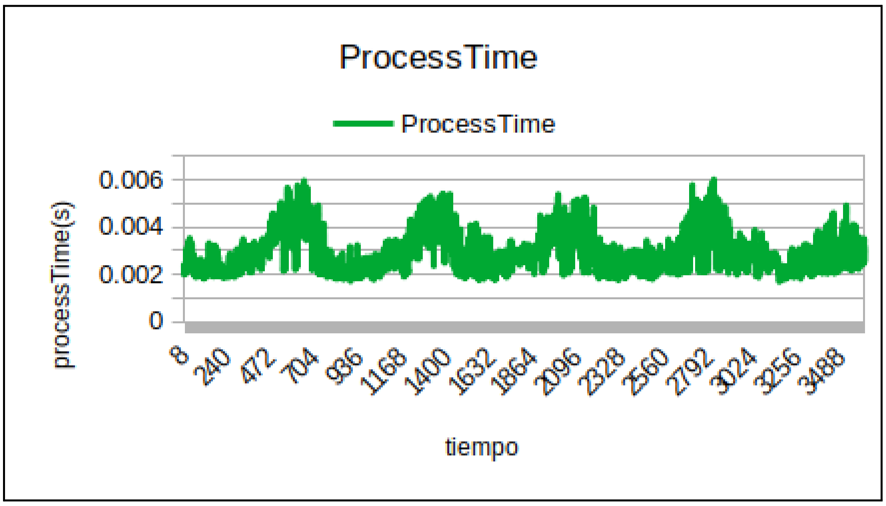
21719 - Avaluació del Comportament de Sistemes Informàtics.

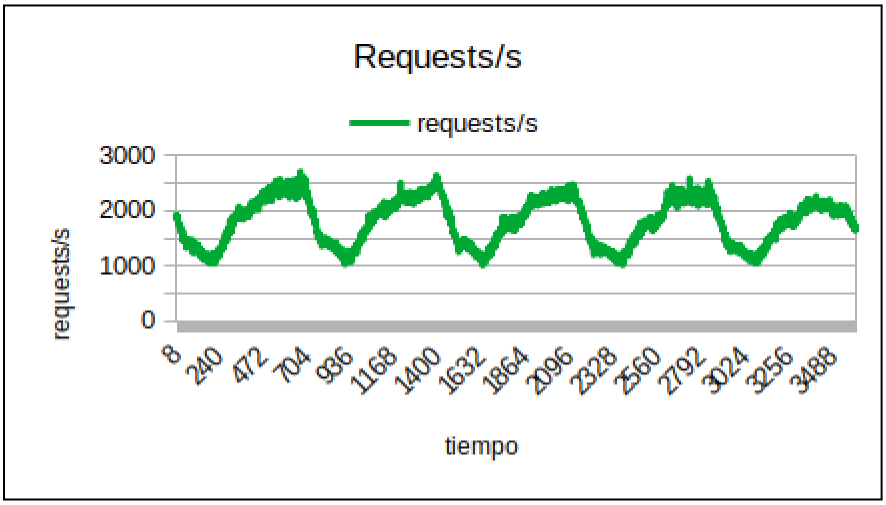
Práctica 4: Tema 7. Planificación de la capacidad

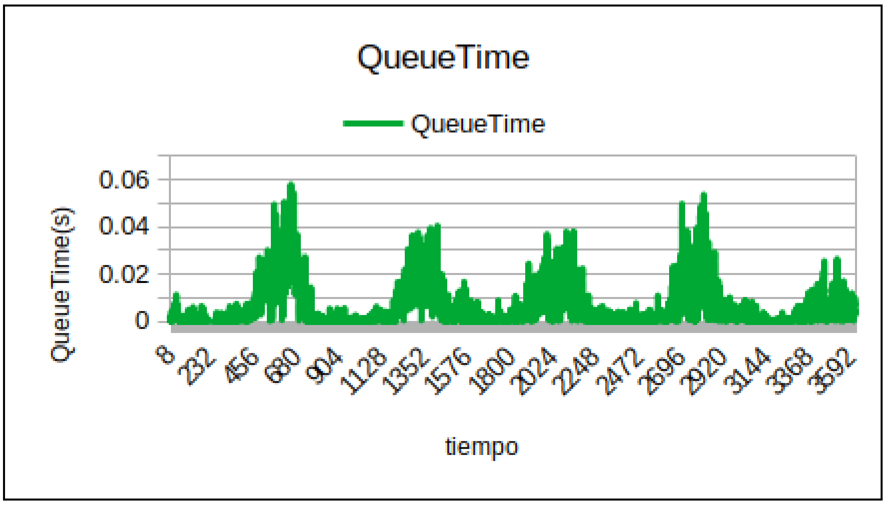
Khaoula Ikkene
Grupo 102
khaoula.ikkene1@estudiant.uib.cat

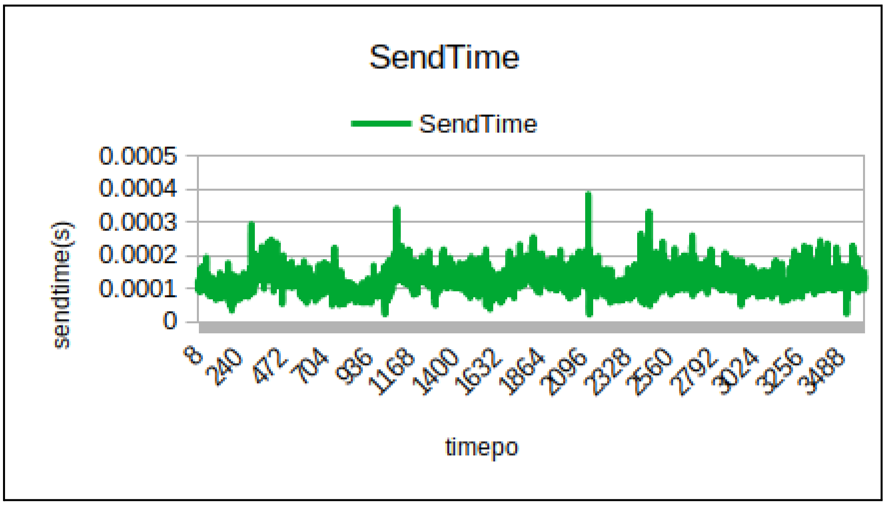
1. ¿Qué patrón siguen los datos monitorizados? Justifica la respuesta y proporciona una representación gráfica.

Al representar los datos de nuestras cuatro columnas (requests/s, ProcessTime, QueueTime, SendTime) con el tiempo podemos observar claramente que siguen patrones cíclicos, salvo el SendTime que es tiene una tendencia estacionaria. Las siguientes gráficas muestras dicho comportamiento.









Los patrones observados en los datos (requests/s, processtime, queuetime)son cíclicos y no estacionales, dado que el periodo de repetición de cada patrón se mide en segundos (las muestras se recogen cada dos segundos).

En detalle:
El comportamiento del sendtime tiene un patrón que se repite aproximadamente cada 23 minutos.
El queuetime muestra un patrón repetitivo aproximadamente cada 12 minutos.
El processtime presenta un patrón recurrente cada 38 minutos.
Se observa también que los incrementos en todas las gráficas coinciden aproximadamente. Es decir, cuando aumenta el queuetime, también aumentan el processtime y el sendtime.

Mientras que el Sendtime tiene una tendencia estacionaria.

El comportamiento del queuetime es notablemente volátil, ya que alcanza valores máximos muy rápidamente y luego disminuye a valores mínimos en poco tiempo.

El processtime tiene un comportamiento estale. Cuando aumenta el tiempo de espera en la cola (queuetime), también aumenta el tiempo de procesamiento de las peticiones (processtime). Esto es razonable, ya que un aumento en las peticiones entrantes incrementa el tiempo de espera en la cola, lo que conduce a una mayor competencia por los recursos y a la sobrecarga del sistema, y esto puede disminuir el tiempo de procesamiento de las peticiones.

2. Calcula los valores solicitados haciendo uso de la regresión lineal, medias móviles (últimos 50 valores), medias móviles (últimos 200 valores), medias móviles (últimos 1000 valores) y suavizado exponencial (α = 0.7).

La ecuación general para calcular la línea de regresión viene dada por:

$y = a + bx$

y: la variable dependiente *x: la variable independiente* *a: el corte con el eje de ordenadas* *b: la pendiente de la línea de regresión que representa la relación entre las dos variables.*

Los parámetros a y b se calculan según el método de los mínimos cuadrados de la siguiente manera:

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

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

Para el cálculo de las medias, es importante tener en cuenta que para el tiempo es preferible utilizar medias aritméticas, mientras que para las frecuencias (como solicitudes por segundo) es más útil emplear la media armónica.

Las medias que usaremos, por lo tanto, serán, la aritmética y la harmónica que se calculan de acuerdo con la siguiente fórmula

$$x_a = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$x_h = \frac{n}{\sum_{i=1}^{n} x_i}$$

En cuanto al tiempo, dado que las muestras se toman cada dos segundos, podemos suponer que la primera muestra se toma en el segundo 1, la siguiente después de 3 segundos, y así sucesivamente.

Los resultados obtenidos son los siguientes

REGRESIÓN LINEAL				
	requests/s	***Processtime***	***QueueTime***	***SendTime***
b	0.093510817626646	-1.13480700533046E-09	-1.80507409572467E-07	1.66055452965941E-09
a	1361.54631943417	0.002928235002164	0.00639257738985	0.000118127862582
media x	3.60E+03	3.60E+03	3600	3600
media y	1698.18526289009	2.92E-03	5.74E-03	1.24E-04
Valor esperado	***2034.91771716364***	***0.002920063256918***	***0.005092743533518***	***0.00013008551575***

Para las medias móviles tenemos

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

donde f_{t+1} es el valor de la predicción, y_t es el valor observado hasta el instante t, y n es el número de observaciones utilizadas para calcular f_{t+1}.

MEDIAS MÓVILES				
	requests/s	Processtime	QueueTime	SendTime
50	1838.20598966	8.27138888977958E-05	2855.70740473665	2.34801691828731E-06
200	1997.17873467	0.003085739545	0.005978187905	0.000133555441
1000	1851.423686869	0.002969863078	0.0060415707373	0.0001286483489

En cuanto al suavizado exponencial usaremos la siguiente fórmula

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

donde f_{t+1} es el valor esperado del periodo t+1, y_{t+1} es el valor observado en el instante t+1, f_t es el valor estimado en el instante t, y α es el peso que se le otorga al valor observado más reciente.

SUAVIZADO EXPONENCIAL CON α = 0.7			
requests/s	processTime	QueueTime	SendTime
1659.87238713321	0.003101739801289	0.00469532160698	0.000143260238274

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

Para determinar cuál técnica de predicción es más adecuada, es fundamental considerar las situaciones en las que cada una resulta útil y el comportamiento de los datos monitorizados.

Comenzando con la regresión lineal, esta técnica es útil cuando la variable a predecir es una función lineal de las variables independientes, es decir, cuando los datos presentan una distribución o tendencia lineal no estacional. Sin embargo, esto no se ajusta a nuestros datos. Por lo tanto, la regresión lineal queda descartada.
Las medias móviles consisten en calcular la media de las n observaciones previas, lo que hace que la predicción sea bastante precisa cuando se trabaja con datos estacionarios o casi estacionarios, debido a la repetición del patrón. Además, dependiendo del número de observaciones n que se elija, las predicciones pueden ser más o menos exactas, y se busca minimizar el error cuadrático de nuestras predicciones, que se muestra en la siguiente tabla.

Error cuadratico	requests/s	Processtime	QueueTime	SendTime
50	1182032.22732618	2.12128955954679E-05	0.002270968957504	5.75426687062092E-08
200	2104880.97566919	9.46224561072874E-06	0.001092581303767	1.77797176226727E-08
1000	378653.650172959	1.46074980524468E-06	0.000152178347273	2.75535086574148E-09

Por otro lado, el suavizado exponencial, como se mencionó anteriormente, asigna un mayor peso a los valores observados más recientes de la variable que queremos predecir. Por lo tanto, no es una técnica adecuada para datos estacionarios o con una tendencia definida, ya que no incorpora un mecanismo para ajustar la tendencia en los cálculos.

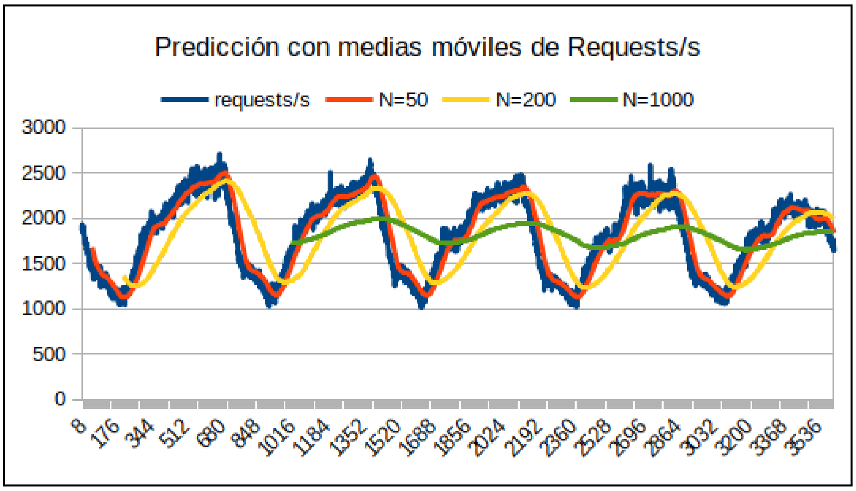
Sin embargo, los resultados obtenidos muestran que el error cuadrático al usar el suavizado exponencial es considerablemente menor que con las medias móviles. En la gráfica se observa claramente que las predicciones son casi exactas, excepto en los valores atípicos.

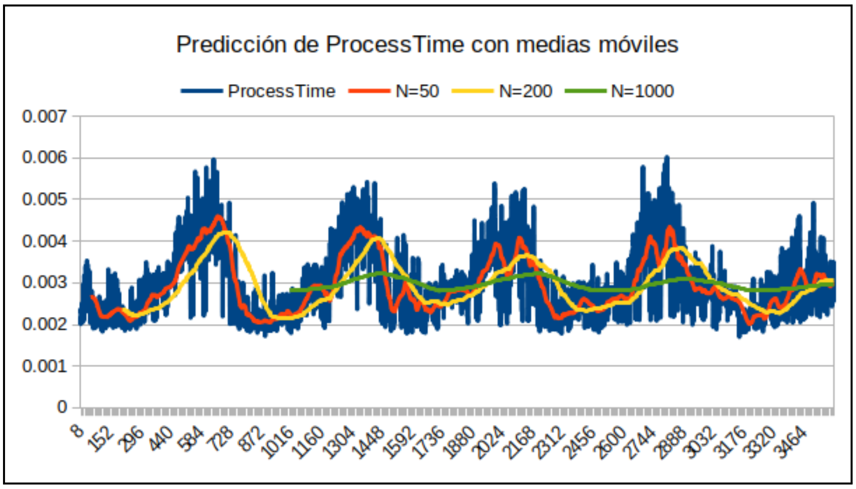
Un factor adicional que explica la precisión de esta técnica es el peso asignado a las observaciones más recientes. Dado que los cambios en el comportamiento de nuestros datos ocurren en intervalos relativamente cortos, esta técnica ha logrado capturar estos cambios y reflejarlos en el cálculo del valor predicho. Esto es particularmente relevante porque estamos trabajando con muestras recolectadas cada dos segundos y se producen cambios significativos en intervalos de minutos.

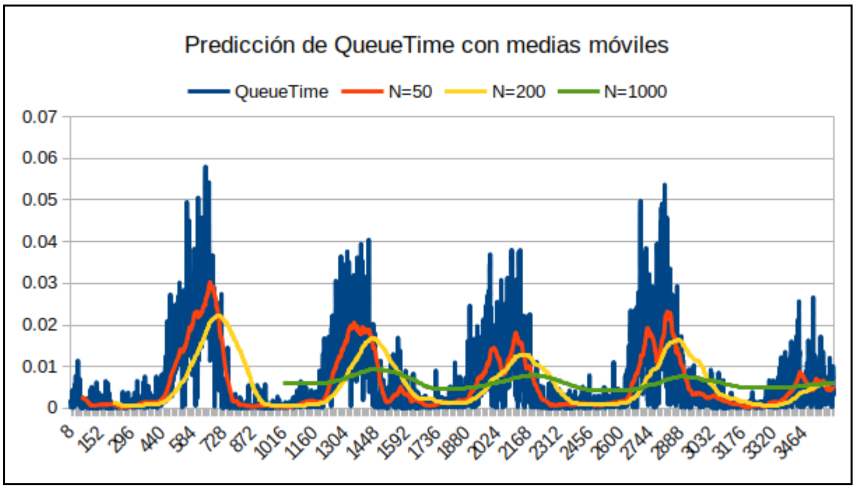
Columna	Requests/s	ProcessTime	QueueTime	SendTime
Error cuadratico	294.16836533244	1.89065561561682E-08	2.42552578432128E-06	1.89065561561682E-08

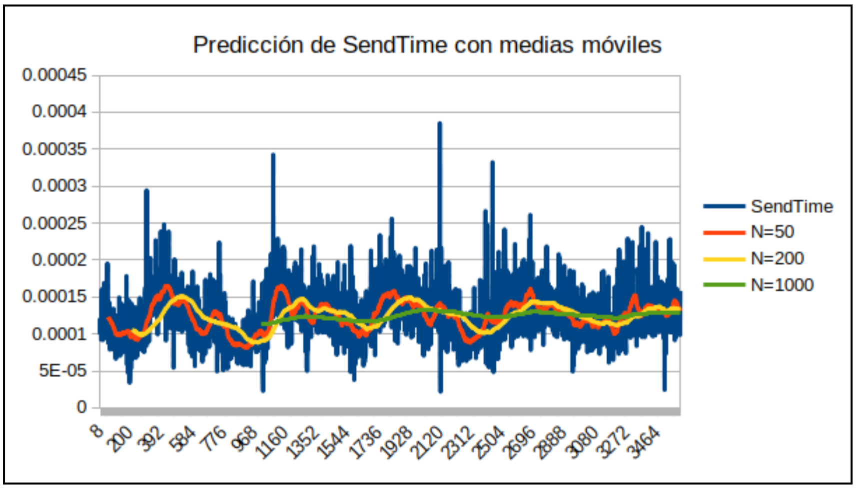
En conclusión, para nuestros datos, la mejor técnica de predicción ha sido el suavizado exponencial, aunque teóricamente esperábamos que sería la de medias móviles.

A continuación, adjuntaré las gráficas representativas de las predicciones comentadas anteriormente para mejor visualización

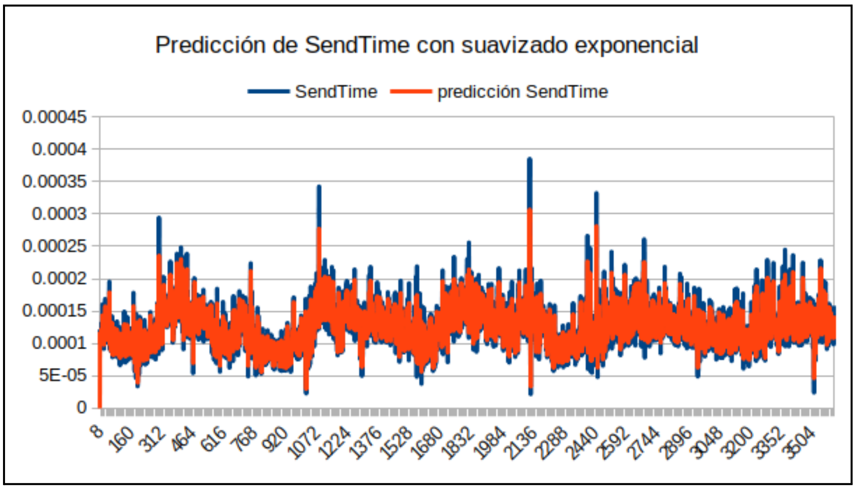




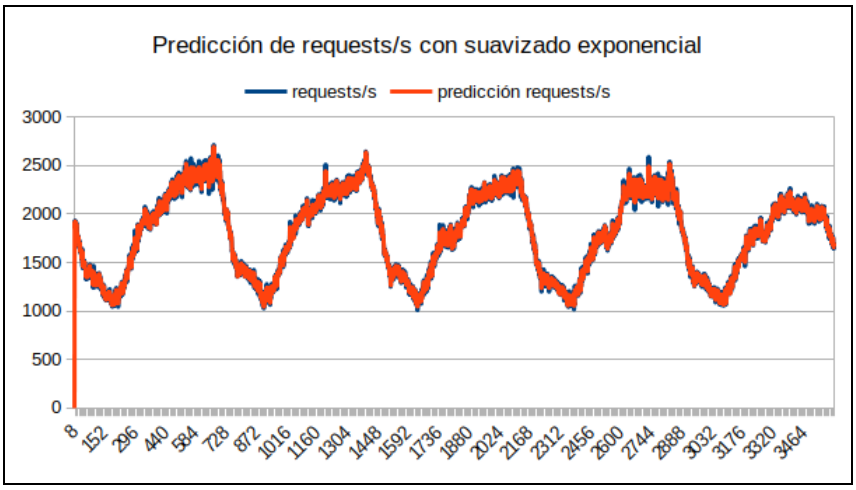




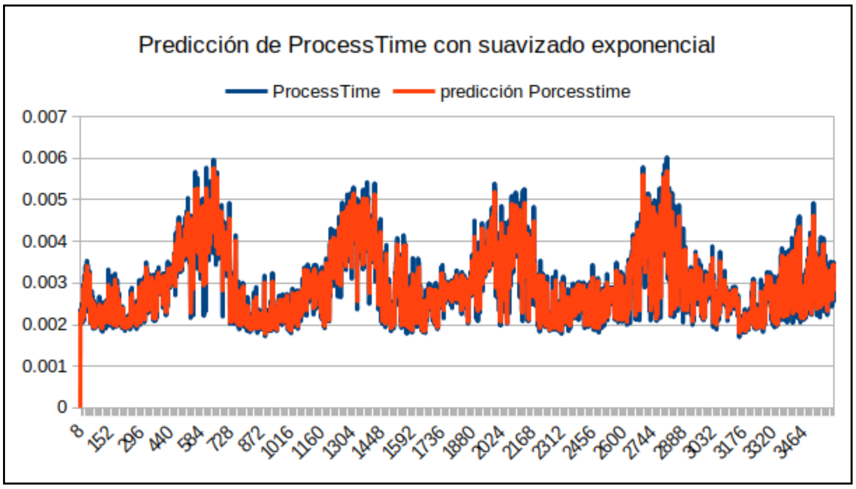
Predicción de SendTime con medias móviles
SendTime
N=50
N=200
N=1000

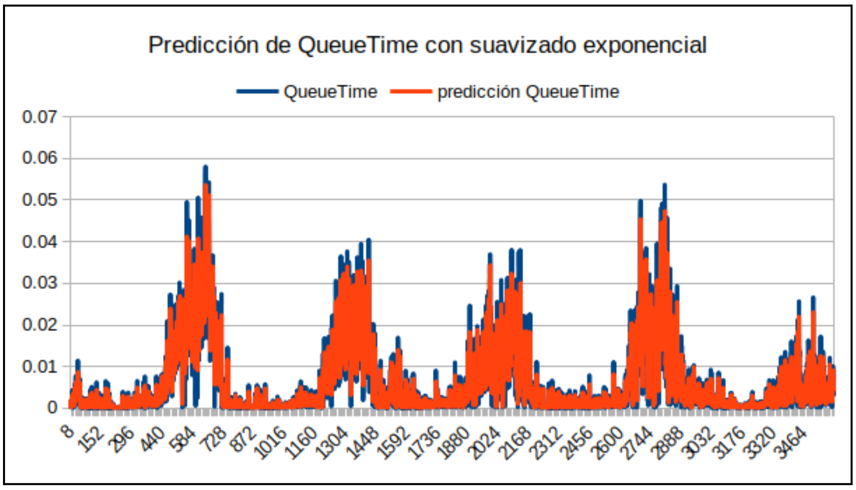


Predicción de SendTime con suavizado exponencial
SendTime
predicción SendTime



Predicción de requests/s con suavizado exponencial
requests/s
predicción requests/s


Predicción de ProcessTime con suavizado exponencial
ProcessTime
predicción Porcesstime
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
8
152
296
440
584
728
872
1016
1160
1304
1448
1592
1736
1880
2024
2168
2312
2456
2600
2744
2888
3032
3176
3320
3464


Predicción de QueueTime con suavizado exponencial
QueueTime
predicción QueueTime
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
8
152
296
440
584
728
872
1016
1160
1304
1448
1592
1736
1880
2024
2168
2312
2456
2600
2744
2888
3032
3176
3320
3464