

Examen 1

Diplomado en Ciencia de Datos — Módulo 1

August 2019

Conjunto de datos

Considere el siguiente conjunto de datos.

Fecha	Cierre	Apertura	Máximo	Mínimo	Quarter	Weekday
2012-01-06	27.72	28.89	29.15	27.39	1	Friday
2012-01-08	20.88	21.50	21.58	20.84	1	Sunday
2012-01-10	21.99	22.08	22.59	21.73	1	Tuesday
2012-01-11	21.21	21.08	21.44	21.01	1	Wednesday
2012-02-07	30.77	31.25	31.73	30.55	1	Tuesday

La tabla proporcionada registra de forma transaccional los precios de las acciones de Facebook diariamente. El conjunto de datos se encuentra segmentado en dos porciones: train y test.

Diccionario de datos

- Fecha: Tipo de dato temporal que indica el día en el que se recabaron los datos
- Cierre: Precio de las acciones de Facebook al cierre del día
- Apertura: Precio de las acciones de Facebook al abrir operaciones del día
- Máximo: Costo máximo de las acciones durante el día en cuestión
- Mínimo: Costo mínimo de las acciones durante el día en cuestión
- Quarter: Cuarto de año correspondiente
- Weekday: Día de la semana

(1 punto) cuales de estas variables son discretas y cuales continuas

Ingeniería de Datos

- (1.5 puntos) Construya una función a partir del conjunto de entrenamiento (train.csv) que genere al menos tres variables en función de las existentes, recibiendo como parámetro la tabla

- (1 punto) Usando la función generada previamente, aplique el mismo proceso a los datos de test
- (1.5 puntos) Genere una función que genere las variables de: día, mes y año a partir de la columna *Fecha*
- (1 punto) Construya una función que construya al menos dos agrupaciones con respecto de las variables construidas en el punto anterior
- (1 punto) Aplique la función al conjunto de test

Calidad de datos

- (1.5 puntos) Realice las pruebas correspondientes para decir si los datos (tanto discretos como continuos) para decir si los conjuntos son estadísticamente iguales
- (1.5 puntos) Genere los gráficos de comparación para ambas muestras, tanto continuas como discretas
- (1.5 puntos) Genere las gráficas de cartas de control (en train) para detectar comportamientos anómalos. La construcción debe de hacerse a nivel mensual
- (1 punto) Aplicar los indicadores obtenidos en el punto anterior en test para decir si el proceso presenta anomalías o no

Cuestionario

Responda de forma breve y concisa a las siguientes preguntas:

- (0.5 puntos) ¿Qué es la Ciencia de Datos?
- (0.5 puntos) ¿Qué *skills* domina un Científico de Datos?
- (0.5 puntos) ¿Cuál son los pasos de un proceso de Ciencia de Datos?
- (0.5 puntos) ¿Qué diferencia hay entre un Ingeniero de Datos y un Científico de Datos?

Feedback

Por favor aporte comentarios (tanto positivos como negativos) sobre el avance del curso, el ponente y las clases. El objetivo es poder mejorar los contenidos y el desarrollo del módulo.