

Test de modulo datamining

Puntos totales 11/11 ?

Contesta a todas las preguntas, siempre hay una opción correcta. Contesta según lo has hecho en la práctica entregada.

0 de 0 puntos

correo electrónico *

dianamtlopez@hotmail.com

Nombre *

DIANA MARÍA

Apellidos *

TORO LÓPEZ

Pregunta nº 1

1 de 1 puntos

✓ ¿Cuántas filas y cuántas columnas tiene el dataframe df_seeds? * 1/1

- ☐ 200 filas y 8 columnas
- ☐ 210 columnas y 8 filas
- ☐ 21 columnas y 10 filas



☒ 210 filas y 8 columnas



Pregunta nº 2

1 de 1 puntos

✓ Convierte en factor la columna seeds, respetando las etiquetas: *

1/1

- ☐ `df_seeds.tipo <- factor(df_seeds$tipo)`
- ☐ `df_seeds$tipo <- factor(df_seeds$tipo, levels=c("Kama", "Rosa", "Canadian"))`
- ☐ `df_seeds$tipo <- factor(tipo, labels=c("Kama", "Rosa", "Canadian"))`
- ☒ `df_seeds$tipo <- factor(df_seeds$tipo, labels=c("Kama", "Rosa", "Canadian"))` ✓

Pregunta nº 3

1 de 1 puntos

✓ ¿Cual es la media del area de cada uno de los tipos? *

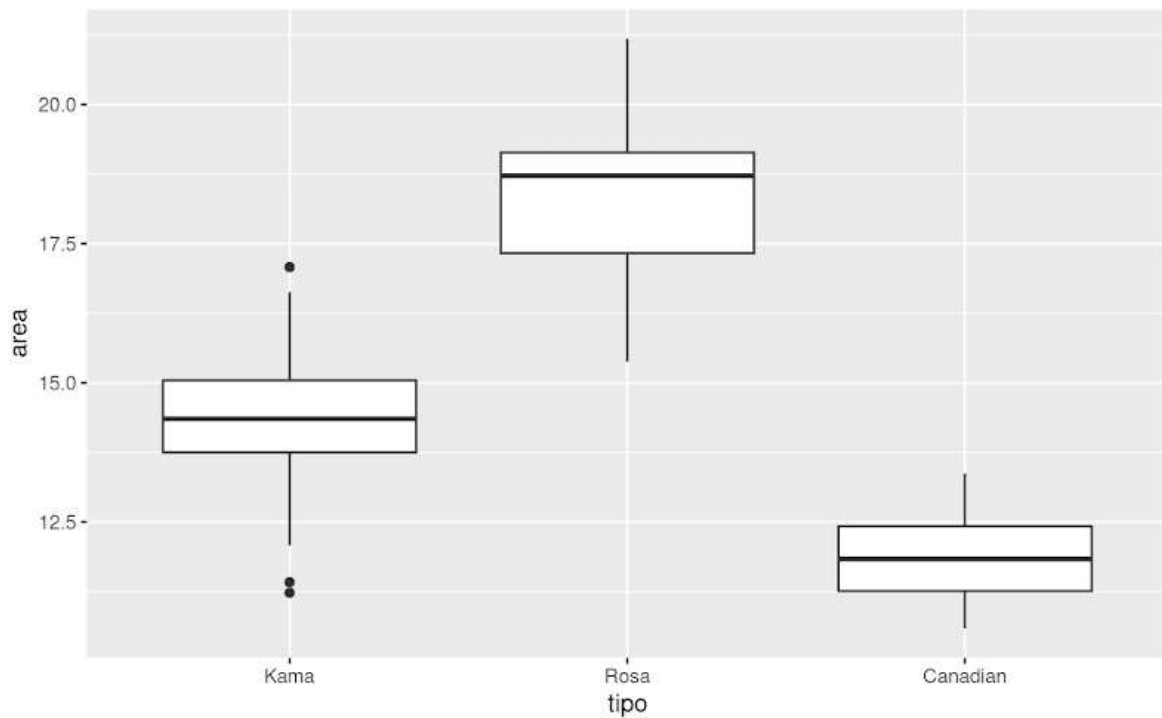
1/1

- ☒ Kama: 14.33 Rosa 18.33 Canadian: 11.87 ✓
- ☐ Kama: 14.4 Rosa 18.7 Canadian: 11.8
- ☐ Kama: 1.48 Rosa 2.07 Canadian: 11
- ☐ Kama: 14 Rosa 38 Canadian: 10.8

Pregunta nº 4

1 de 1 puntos





✓ ¿Como se llama el siguiente tipo de gráfico?. ¿Qué representa la línea del centro de la caja? *1/1

- ☒ Diagrama de cajas o boxplot. La línea del centro de cada caja representa la mediana. ✓
- ☐ Diagrama de cajas o boxplot. La línea del centro de cada caja representa la media.
- ☐ Diagrama de puntos o scatter plot. La línea del centro de cada caja representa la mediana.
- ☐ Diagrama violin plot. La línea del centro de cada caja representa la mediana.

Pregunta nº 5

1 de 1 puntos



✓ ¿Como pintarías un diagrama de puntos (o scatterplot) con ggplot con las siguientes características? *1/1

- En el eje X la variable compacto
- En el eje Y la variable area
- Cada tipo de semilla debería tener un color diferente

- ☐ ggplot(df_seeds, aes(x=compacto , y=area, color=tipo)) + geom_line()
- ☒ ggplot(df_seeds, aes(x=compacto , y=area, color=tipo)) + geom_point() ✓
- ☐ ggplot(df_seeds, aes(x=compacto , y=area)) + geom_point(aes(color="tipo"))
- ☐ ggplot(df_seeds, c(x=compacto , y=area, color=tipo)) + geom_point()

Pregunta nº 6

1 de 1 puntos

✓ ¿Qué hace la siguiente línea?: * 1/1
`df_seeds |> mutate(is_kama = tipo=='Kama') -> df_seeds`

- ☐ Crea una nueva columna llamada tipo que tiene siempre el valor 'Kama'
- ☐ Filtra y nos deja en el dataframe solo los tipos de semilla 'Kama'
- ☒ Crea una nueva columna llamada is_kama que tiene el valor TRUE, cuando el tipo de semilla es Kama y FALSE en caso contrario. ✓
- ☐ Crea una nueva columna llamada is_kama que viene del filtrado de valores consecutivos del dataframe correspondiente

Pregunta nº 7

1 de 1 puntos



- ✓ Vamos a dividir el conjunto de datos en test y training porque vamos a entrenar un modelo que me permita diferenciar si una semilla es de tipo Kama o no. *1/1

¿Por qué es aconsejable dividir el dataset en los grupos de train y test?

- ☒ Utilizaremos el dataset de train para entrenar el modelo. El dataset de test nos dará la precisión del modelo con muestras que no ha visto durante el entrenamiento. Esto permite obtener una evaluación de la calidad del modelo cuando lleguen datos nuevos. ✓
- ☐ Utilizaremos el dataset de test para entrenar el modelo. El dataset de train nos dará la precisión del modelo con muestras que no ha visto durante el entrenamiento. Esto permite obtener una evaluación de la calidad del modelo cuando lleguen datos nuevos.
- ☐ Nunca es aconsejable dividir el dataset en train y test porque nos quedamos con menos muestras
- ☐ Es aconsejable dividir el dataset en los grupos de train y test para aumentar la precisión del modelo en la clasificación de semillas de tipo Kama. Esto permite obtener una evaluación de la calidad del modelo cuando lleguen datos nuevos.

Pregunta nº 8

1 de 1 puntos

- ✓ Vamos a crear un modelo para realizar una clasificación binaria, donde le pasaremos como entrada las columnas: area, perimetro, compacto, longitud, coeficient.asimetria y longitud.ranura *1/1

¿Qué tipo de algoritmo o modelo debería usar?

- ☐ Debería usar un modelo de regresión lineal para crear una clasificación binaria, ya que este tipo de modelo es adecuado para hacer una predicción numérica a partir de un conjunto de variables.

Debería usar un algoritmo de clustering para crear una clasificación binaria, ya que

- ☐ este tipo de algoritmo es adecuado para agrupar datos similares en conjuntos distintos.
- ☒ Debería usar un modelo de regresión logística. ✓
- ☐ Debería utilizar un modelo de regresión por refuerzo

Pregunta nº 9

1 de 1 puntos

✓ Crea un modelo que me permita clasificar si una semilla es de tipo Kama *1/1
o no con las siguientes columnas:
area, perimetro, compacto, longitud, coeficient.asimetria, longitud.ranura

- ☐ `model <- glm(df_seeds_train,
formula=is_kama~area+perimetro+compacto+longitud+coeficient.asimetria+longitud.ranura)`
- ☐ `model <- glm(df_seeds_train,
formula=is_kama~area+perimetro+compacto+longitud+coeficient.asimetria+longitud.ranura, family='gaussian')`
- ☐ `model <- lm(df_seeds_train,
formula=tipo~area+perimetro+compacto+longitud+coeficient.asimetria+longitud.ranura)`
- ☒ `model <- glm(df_seeds_train,
formula=is_kama~area+perimetro+compacto+longitud+coeficient.asimetria+longitud.ranura, family='binomial')` ✓

Pregunta nº 10

1 de 1 puntos

✓ Si usamos un umbral de 0 en la salida del modelo (lo que equivale a probabilidad de 0.5 cuando usamos el predict con type='response') *1/1
¿Cuales son los valores de precisión y exhaustividad?

- ☒ Precision: 0.95 Exhaustividad: 1 ✓
- ☐ Precision: 0.9337 Exhaustividad: 0.957
- ☐ Precision: 1 Exhaustividad: 1
- ☐ Precision: 0 Exhaustividad: 0



Pregunta nº 11

1 de 1 puntos

✓ ¿Qué están haciendo las siguientes líneas? *

1/1

- ☐ Las siguientes líneas están dividiendo el conjunto de datos en una serie de subconjuntos utilizando un algoritmo de ordenamiento para clasificar las semillas en diferentes grupos.
- ☒ Estamos seleccionando varias columnas para realizar una clusterización por kmeans. El resultado son 3 clusteres que comparamos con el tipo original del dataframe ✓
- ☐ Las siguientes líneas están creando un gráfico de barras para mostrar los diferentes tipos de semillas y su relación con los valores de las características del conjunto de datos.
- ☐ Las siguientes líneas están utilizando un modelo de regresión lineal para predecir el tipo de semilla a partir de las características del conjunto de datos, y compararlo con el tipo real de las semillas para medir la precisión del modelo.

Este formulario se creó en KeepCoding.

Google Formularios

