

Segunda Evaluación Parcial

Para cada uno de los ejercicios:

- Suba los archivos .R/.ipynb indicados
- Adjunte el PLOT como imagen en esta hoja de trabajo
- Responda las preguntas indicadas
- Adjunte los archivos .pbix

Parte I – Teoría (25%)

En el portal encontrará el enlace al cuestionario para la parte teórica del examen. Este enlace estará abierto de 5:30pm a 6:00pm, sin embargo, al iniciar el cuestionario tendrá 12min para responder a 4 preguntas

Parte II – Clustering (30%)

Descargue la tabla de Casos de COVID-19 de México del siguiente enlace:

https://correo2urledu-my.sharepoint.com/:u:/g/personal/dsbolanosa_correo_url_edu_gt/EeiNnZIC7R9PrHveChxeY1MBT8cMG-iLb2RcQmRgeBDvPg?e=qNPNzF

Que indica si una persona falleció y sus quebrantos de salud (tales como asma, insuficiencia renal, etc..) elabore lo siguiente:

- Extraiga los datos usando “read.csv”, usando las funciones de “dplyr” para manipulación de datos cree un “data.frame” que tenga las siguientes columnas:
 - Edad del paciente
 - Conteo de intubados (Intubado=2 en la columna intado)
 - Conteo de cuidados intensivos (UCI = 1 o UCI = 2)
 - Conteo del total de casos
 - Conteo del total de fallecidos
- Cree un script de R llamado “Clustering COVID por edades” que tenga lo siguiente
 - Cree por lo menos tres modelos de K means para diferentes números de K clusters
 - Justifique porque eligió esos tres valores de K
 - Un plot que grafique los tres diferentes modelos de k means
 - ¿Cuál es el mejor valor para “K” para este problema? ¿Por qué?
 - Un plot que grafique solamente el mejor modelo de k means
 - Un plot de wss
 - Explique que interpreta del plot WSS
 - Cree una tabla o data frame que tenga los promedios de las métricas de cada cluster
 - Explique las características de los promedios encontrados de cada grupo, por ejemplo “el grupo 1 posee estas características mientras el grupo 2 posee diferentes características”

Parte III – Clasificación (30%)

Usando el archivo “hongos.csv” que se encuentra en el portal, cree modelos de clasificación usando Naive Bayes y Arboles de Decision para poder clasificar si un tipo de hongo, según sus características, es venenoso o no.

- La columna “class” es una variable categoría de dos niveles “no venenoso=e, venenoso=p”
- Se adjunta también un archivo llamado “Descripcion de atributos, Datos de Hongos.xlsx” que explica cada uno de los atributos y sus diferentes categorías

Con los datos de dicho archivo realice lo siguiente en Python:

- Visualice si existen datos faltantes o vacíos, de ser así descártelos, adjunte que atributos vienen incompletos (de haber alguno)
- evalúe que variables son categóricas y cuales son continuas, para aquellas categóricas cambie los tipos de atributos a “factor”
- Separe la data en dos data frames, uno para entrenar al modelo y otro para probarlo. Utilice un método aleatorio para realizar dicha separación.
- Cree los dos modelos de clasificación anteriormente descritos.
- Para el árbol de decisión realice lo siguiente:
 - Adjunto un plot del dibujo de dicho árbol
 - ¿Cuáles son las probabilidades de cada clase en las hojas del árbol?
 - ¿Qué atributos son más significativos?
 - ¿Se descarto algún atributo? De ser así, ¿Cuál o cuáles?
 - ¿Qué clases (valores de los atributos) producen un mayor impacto en si un hongo es venenoso o no? Por ejemplo “los hongos que tienen figuras convexa, plana y hundida tienden a tener una probabilidad más alta de ser venenosos”
- Para Bayes:
 - ¿Cuáles son las probabilidades de los resultados de la clasificación de si un hongo es venenoso o no?
 - Adjunte las tablas de probabilidad de cada atributo
- Cree objetos de los resultados de la predicción de cada modelo usando el data frame de pruebas
- Usando los resultados de la predicción de ambos modelos, cree y adjunte las matrices de confusión
 - ¿Qué modelo tiene más errores Tipo 1?
 - ¿Qué modelo tiene más errores Tipo 2?
 - ¿Qué modelo tiene una mejor precisión?
 - ¿Cuál de los dos modelos elegiría y por qué?

Parte IV – Power BI (15%)

Usando la tabla de Casos de COVID-19 de México que indica si una persona falleció y sus quebrantos de salud (tales como asma, insuficiencia renal, etc..) cree un reporte de Power BI que tendrá las por lo menos las siguientes visualizaciones:

- Pestaña Línea de tiempo
 - Defunciones (Fecha_def)
 - Muestre un gráfico de línea de tiempo del conteo de defunciones de COVID por Año-mes
 - Muestre una tabla que tenga dos columnas
 - Año-mes
 - Conteo de defunciones
 - Asegúrese de excluir pacientes que no fallecieron
 - Ingresos (Fecha_ingreso)
 - Muestre un gráfico de línea de tiempo del conteo de ingresos de COVID por Año-mes
 - Muestre una tabla que tenga dos columnas
 - Año-mes
 - Conteo de ingresos
- Pestaña Clustering
 - Muestre la gráfica de los clusters con el mejor K del inciso 1 corriendo su código de dicho inciso (únicamente muestre el mejor K, no varios)
 - Muestre una tabla con los promedios de las métricas por grupo para el K elegido