



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.

MATERIA/UNIDAD DE APRENDIZAJE: Minería de datos.

I. Lee el caso de ejemplo y responde las preguntas correspondientes:

La Coca Cola Company tiene 366,000 accionistas y efectúa una encuesta por medio de la selección aleatoria de 30 accionistas de cada uno de los 50 estados de Estados Unidos. Se registra el número de acciones de cada accionista de la muestra.

- a) ¿Los valores obtenidos son discretos o continuos?

Datos Discretos.

- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales

De razón.

- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?

Muestreo aleatorio estratificado.

- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?

Estadístico.

- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?

El promedio y desviación estándar de los accionistas tanto por estado como del conjunto completo.

- f) ¿Qué es lo que está incorrecto al evaluar la opinión del accionista enviado un cuestionario por correo que los accionistas podrían llenar y regresar por el mismo medio?

Puede haber un sesgo por el limitante de las mismas pregunta o contradicciones en las mismas respuestas.

II. Describe con tus palabras para que se utiliza el análisis uni-variante de datos:

Busca resumir (con estadísticos) o buscar patrones de un conjunto de datos, dejando de lado si estos se relacionan.

III. Describe qué es la media, la mediana y la desviación estándar:



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

La media es un promedio de un conjunto de datos- La mediana es el valor central de un conjunto de datos ordenados, ya sea de menor a mayor o viceversa. La desviación estándar muestra que tan dispersos están los datos con respecto al promedio de los mismos (media).

IV. Con tus palabras explica qué es minería de datos y en donde puede ser aplicada:

La minería de datos son técnicas de manipulación de conjuntos de datos para tratar de describir a estos, ya sea encontrando patrones o relaciones entre los mismos, o simplemente describiéndolos de forma generas. Lo anterior de puede aplicar a la hora de querer analizar el comportamiento de las masas (personas), estudiar resultados de estrategias de marketing o simplemente para evaluar la eficiencia de una empresa, materias, capacitaciones, etc., lo anterior sujeto a tener grandes cantidades de tatos para que sea significativo.

V. Que lenguajes de programación dominas o en qué porcentaje lo manejas:

Visual Basic (un 50% o 60%), C (15% o 20%) y R (10%)

VI. Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):

No

VII. ¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?

No trabajado directamente, pero he visto como mi padre trabaja, aunque el no lo vea, manipula muchos datos y hace reportes de estos.

VIII. Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:

Lo he visto en videos, pero nunca manipulado.

IX. Relaciona los conceptos con sus gráficas correspondientes:

- a) Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
- b) Forma estandarizada de mostrar la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo"). Puede informarle si sus datos son simétricos, qué tan estrechamente están agrupados y si sus datos están sesgados.
- c) Método para visualizar la distribución de datos numéricos de diferentes variables. Es similar al diagrama de caja, pero con un diagrama rotado en cada lado, que brinda más información sobre la estimación de densidad en el eje y.



UANL

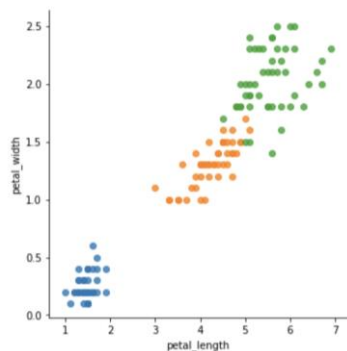
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

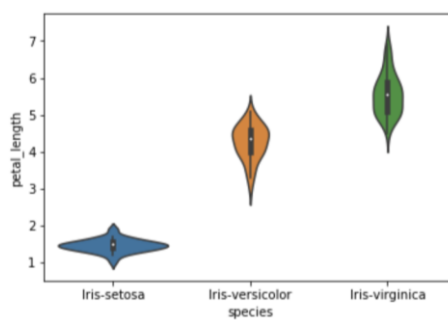
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



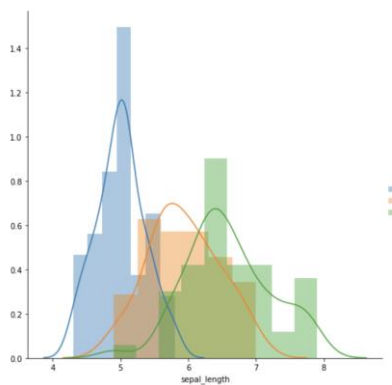
- d) Visualización de datos bidimensionales que utiliza puntos para representar los valores obtenidos para dos variables diferentes: una trazada a lo largo del eje "xy" otra trazada a lo largo del eje y.
- e) Produce una matriz de relaciones entre cada variable en sus datos para un examen instantáneo de nuestros datos. También puede ser un gran punto de partida para determinar los tipos de análisis de regresión a utilizar.



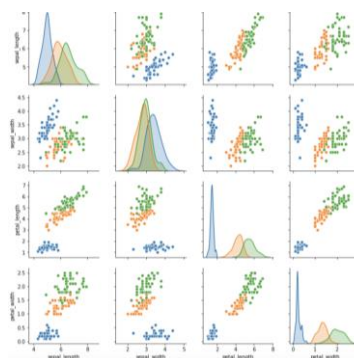
D



C



A



E

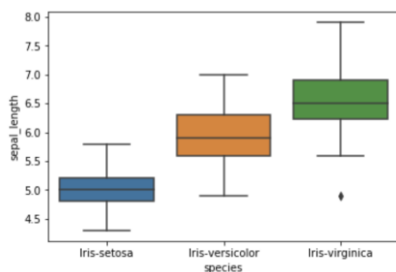


UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



B

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos):

Ningún problema

- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)

Solamente no aparecen algunas materias, por lo demás, todo correcto

- XII. Qué sistema operativo tienes:

Windows 10

- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.

- Github: **(0%, nuevo para mi)**
- Jupyter Notebook **(5%, lo vi en un curso de introducción al mismo)**
- Google Collab **(0%, nuevo para mi)**
- Python 3 Notebook **(5%, lo vi en un curso de introducción al mismo)**
- Terminal de tu computadora **(10%, no la uso mucho)**