



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?

Son 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?

Es estratificado.

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

Es un estadístico.

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

Que los datos recopilados sean confiables para trabajar con ellos.

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 que una persona distinta al accionista contestara la encuesta.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



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

Es un análisis que se utiliza cuando solo se tiene una sola variable, su principal función es estudiar a fondo los estadísticos de la variable.

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

La media es el promedio de los datos, representa cual debería ser el valor central de una muestra o población. La mediana es el dato que está en la mitad cuando estos se ordenan. La desviación estándar refleja que tan alejados o cercanos están los datos de la media o promedio.

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

Es el estudio exhaustivo de bases de datos para poder encontrar anomalías y patrones para poder predecir resultados futuros. La utilizan los bancos para entender a sus clientes y ver riesgos de mercado, empresas de comunicación, servicios de streaming, entre muchas más.

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

C (50%)

Visual Basic (50%)

Python (35%)

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

No he manejado ninguno notebook digital

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

No, no he trabajado, aunque me gustaría en el futuro



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS

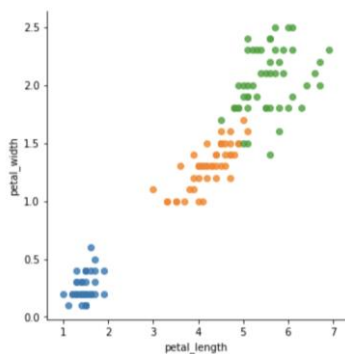


VIII. Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:

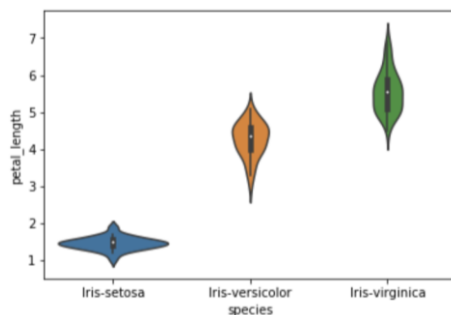
No, es posible que alguna vez lo hiciera, pero no sabía que era machine learning, aunque la probabilidad es mínima

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.
- 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

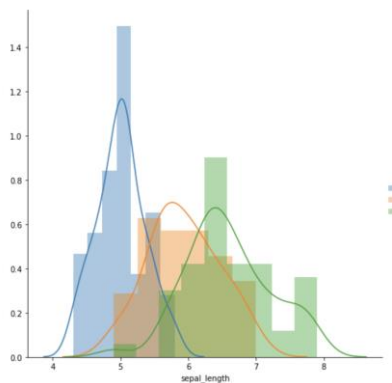


UANL

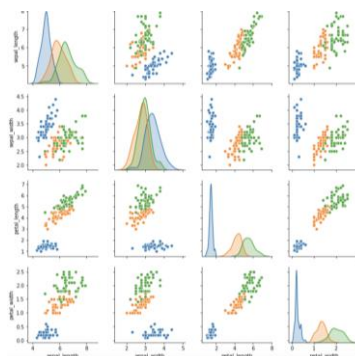
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

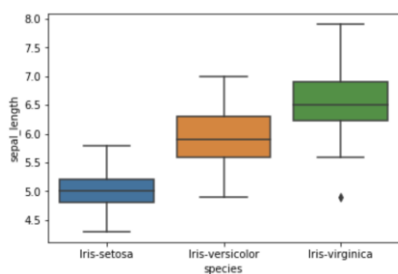
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



A



E



B

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos):
Solamente una conexión algo inestable de internet.
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)
No tengo ningún problema.
- XII. Que sistema operativo tienes:
Windows 10.
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
 - Github. (10%)
 - Jupyter Notebook (30%)



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- Google Collab (2%)
- Python 3 (60%)
- Terminal de tu computadora. (50%)