



Modelos Datos II

1. IDENTIFICACIÓN DE ASPECTOS BÁSICOS

Denominación del Programa de Formación: Análisis y Desarrollo de Software

Instructor: Edicson Quiroz

Fase: Evaluación

Actividad de Proyecto: Desarrollar las tareas de configuración y puesta en marcha del software.

Competencia: Desarrollar la solución de software de acuerdo con el diseño y metodologías de desarrollo.

Resultados de Aprendizaje a Alcanzar: Codificar la solución que cumpla con el diseño establecido.

Actividad Evaluativa: Modelos de Datos II

Utilizando Python y el editor de Visual Studio Code a través de la línea de comandos, u otro editor que permita realizar la ejecución, utiliza el archivo anexo de acuerdo al siguiente listado:

DOCUMENTO	ARCHIVO PDF
1001361171	Dataset_Salarios
1131104572	Dataset_Vacunas
1025884757	Dataset_Salarios
1092455030	Dataset_Vacunas
1025883346	Dataset_Salarios
1017279074	Dataset_Vacunas
1128424088	Dataset_Salarios
1128424098	Dataset_Vacunas
1034987348	Dataset_Salarios
1023625636	Dataset_Vacunas
1005271815	Dataset_Salarios
1001580656	Dataset_Vacunas
1025884397	Dataset_Salarios
1001138698	Dataset_Vacunas
1038867065	Dataset_Salarios
1015067967	Dataset_Vacunas
1036926873	Dataset_Salarios
1037388309	Dataset_Vacunas
1021803262	Dataset_Salarios
1035971586	Dataset_Vacunas
1127925664	Dataset_Salarios



1. Realiza inicialmente actividades de extracción, limpieza y transformación de datos para calcular:

Dataset_Salarios:

Frecuencias (absoluta, acumulada, relativa) para la experiencia en años, e imprime la tabla de frecuencias con su gráfico al igual que para el salario 2023 y el Sexo.

- a. Utiliza las medidas de tendencia central calculando la media, mediana, moda, varianza y desviación estándar para las anteriores columnas e imprime los resultados.

Dataset_Vacunas:

Frecuencias (absoluta, acumulada, relativa) para la Edad, e imprime la tabla de frecuencias con su gráfico al igual que para la estatura y el Sexo.

- b. Utiliza las medidas de tendencia central calculando la media, mediana, moda, varianza y desviación estándar para las anteriores columnas e imprime los resultados.

2. Considerando que el análisis de la regresión lineal se utiliza para predecir el valor de una variable según el valor de otra. Donde la variable que desea predecir se denomina variable dependiente y la variable que se está utilizando para predecir el valor de la otra variable se denomina variable independiente, aplica regresión lineal realizando las siguientes operaciones:

Dataset_Salarios:

- a. Lectura y análisis de los datos
- b. Limpieza y transformación de los datos:
 - Eliminar filas con valores nulos
 - Eliminar las columnas que no utilizaremos (dejar solo las de SALARIO 2022, SALARIO 2023 y EXPERIENCIA AÑOS)
 - Reemplazar los salarios con valores ceros o menores a un salario mínimo por 1170000.
 - Cambiar el nombre de las columnas
 - Crear columna promedio y riesgo
- c. Entrenar el modelo de regresión lineal:
 - Seleccionar variables relevantes
 - Dividir el conjunto de datos en entrenamiento y prueba
 - Crear el modelo de regresión lineal



Entrenar el modelo

Realizar predicciones en el conjunto de prueba

d. Evaluar el modelo

Dataset_Vacunas:

e. Lectura y análisis de los datos

f. Limpieza y transformación de los datos:

- Eliminar filas con valores nulos
- Eliminar las columnas que no utilizaremos (dejar solo las de DOSIS COVID 1, DOSIS COVID 2 y Presión Arterial Alta)
- Reemplazar las dosis con valores ceros o con valor de 1, por 0,5
- Cambiar el nombre de las columnas
- Crear columna promedio y riesgo

g. Entrenar el modelo de regresión lineal:

- Seleccionar variables relevantes
- Dividir el conjunto de datos en entrenamiento y prueba
- Crear el modelo de regresión lineal

Entrenar el modelo

Realizar predicciones en el conjunto de prueba

h. Evaluar el modelo

NOTA: Genera un archivo .txt (block de notas) o .docx (Word) con los scripts. Puedes enviar los archivos anteriores en un archivo comprimido en un zip (no .rar).

“Cree en ti mismo y en lo que eres”