

---

**EVALUACIÓN 2 - PARTE 1 - MODELOS DE REGRESIÓN**  
**INTELIGENCIA ARTIFICIAL**

Docente Jazna Meza Hidalgo  
Mayo 2024

## CONTEXTO DEL NEGOCIO

PeopleGrow Solutions se dedica a la consultoría de desarrollo organizacional y gestión de talento. El objetivo del negocio es optimizar las estrategias de compensación salarial para mejorar la retención del talento y la satisfacción laboral, considerando múltiples factores del perfil de los empleados.

El conjunto de datos proporciona información sobre:

1. ID:número de identificación del empleado.
2. Edad: edad del empleado.
3. Nivel Educativo: Universitaria, Técnica, Secundaria, Postgrado.
4. Experiencia en años: cantidad de años que lleva trabajando.
5. Departamento: departamento en donde se desempeña.
6. Horas entrenamiento mensual: cantidad de horas destinadas, mensualmente, a capacitación.
7. Satisfacción laboral: Alta, Muy Alta, Media, Baja.
8. Remoto: modalidad de trabajo.
9. Salario mensual: sueldo mensual, expresado en miles de peso.

El objetivo principal es desarrollar modelos de regresión que permitan predecir el monto de las **ventas semanales** para cada tienda, basándose en variables que reflejan tanto factores económicos como eventos estacionales y externos. Esta predicción permitirá a la cadena anticiparse a fluctuaciones en la demanda, optimizando así su cadena de suministro y maximizando la eficiencia operativa.

Dentro de los beneficios de realizar este trabajo se encuentran:

1. Mejorar la planificación de inventarios.
2. Ajustar estrategias de marketing y promociones basadas en la predicción de la demanda.
3. Aumentar la eficiencia operativa y la rentabilidad de las tiendas.

## DETALLES DE LA ENTREGA

1. Considerar la fase 2 de entendimiento de los datos para responder a la siguiente pregunta:
  - (a) ¿Existen diferencias salariales significativas entre empleados remotos y presenciales?
2. Construir modelos de regresión usando los tres algoritmos revisados en clases. Cada modelo debe ser generado considerando:
  - (a) Incluir todas las variables incluidas en el conjunto de datos.
  - (b) Incluir variables independientes elegidas por el equipo. Se debe fundamentar la elección de ellas.
3. Generar la siguiente tabla de resultados de todos los modelos solicitados

Table 1: Tabla Resultados

Modelo	$R^2$ train	$R^2$ test	MAE	MSLE	Overfitting
Modelo	9,999	9,999	9,99	99,99%	Si/No

4. Obtener conclusiones acerca de los resultados indicados en la tabla anterior. Deberá comentar respecto del resultado de  $R^2$  en test, MAE y MSLE y acerca de la existencia/inexistencia de overfitting.
5. Usando el mejor modelo realizar predicciones usando el siguiente esquema:
  - (a) Cargar, desde un archivo .json los valores de las variables independientes
  - (b) Realizar las predicciones de cada entrada contenida en el archivo cargado en el punto anterior.
  - (c) Generar un archivo .json que contenga las predicciones realizadas por el modelo

## FORMATO DE LA ENTREGA

Debe entregar el notebook de base entregado completando cada una de las secciones contenidas en él. Todos los comentarios y las tablas de resultados que se requieren deben estar incluidas en el notebook y en una celda de texto

Para el caso del archivo JSON que se pide con las entradas para realizar las predicciones, se debe incluir un enlace para cargar el archivo usando el comando *wget*.

## PLAZO DE ENTREGA

En la plataforma ADECCA, **JUEVES 15 DE MAYO** hasta las 23:55. Se aceptan entregas posteriores de acuerdo con los siguientes descuentos de la calificación final:

FECHA DE ENTREGA	DESCUENTO A APLICAR
15 mayo a las 23:56 horas hasta el 16 mayo a las 00:55 horas	Descuento = 20%
16 mayo a las 00:56 horas hasta el 16 mayo a las 01:55 horas	Descuento = 30%
Después de las 01:56 del 16 mayo	Descuento = 70%

## LISTA DE COTEJO

- Fase 2 - Entendimiento de los datos
  1. Responde, correctamente, a la pregunta
  2. Respalda, de forma correcta, la respuesta a la pregunta usando gráficos y/o tablas.
- Modelo usando todas las variables
  1. Justifica, correctamente, la técnica de transformación de la variable Nivel educativo.
  2. Aplica, correctamente, la técnica de transformación de la variable Nivel educativo.
  3. Justifica, correctamente, la técnica de transformación de la variable Departamento.
  4. Aplica, correctamente, la técnica de transformación de la variable Departamento.
  5. Justifica, correctamente, la técnica de transformación de la variable Departamento.
  6. Aplica, correctamente, la técnica de transformación de la variable Departamento.
  7. Crea y entrena, correctamente, modelo de regresión usando LinearRegression
  8. Crea y entrena, correctamente, modelo de regresión usando DecisionTreeRegressor
  9. Crea y entrena, correctamente, modelo de regresión usando RandomForest
  10. Genera tabla de resultados con todas las métricas solicitadas.
  11. Explicar los criterios usados para seleccionar el mejor modelo.
  12. Selecciona de forma automática y correctamente el mejor modelo.
  13. Obtener conclusión correcta (en el negocio) acerca de  $R^2$  test del mejor modelo seleccionado
  14. Obtener conclusión correcta (en el negocio) acerca de MAE del mejor modelo seleccionado
  15. Obtener conclusión correcta (en el negocio) acerca de MSLE del mejor modelo seleccionado
  16. Obtener conclusión correcta acerca de existencia/inexistencia de overfitting del mejor modelo seleccionado
  17. Carga un archivo JSON con los datos para poder predecir usando el mejor modelo
  18. Generar, correctamente, predicciones en archivo JSON usando el mejor modelo obtenido
- Modelo usando variables seleccionadas por el equipo
  1. Justifica, correctamente, la elección de las variables independientes.
  2. Crea y entrena, correctamente, modelo de regresión usando LinearRegression
  3. Crea y entrena, correctamente, modelo de regresión usando DecisionTreeRegressor

4. Crea y entrena, correctamente, modelo de regresión usando RandomForest
  5. Genera tabla de resultados con todas las métricas solicitadas.
  6. Explicar los criterios usados para seleccionar el mejor modelo.
  7. Selecciona, correctamente y de forma automática, el mejor modelo
  8. Obtener conclusión correcta (en el negocio) acerca de  $R^2$  test del mejor modelo seleccionado
  9. Obtener conclusión correcta (en el negocio) acerca de MAE del mejor modelo seleccionado
  10. Obtener conclusión correcta (en el negocio) acerca de MSLE del mejor modelo seleccionado
  11. Obtener conclusión correcta acerca de existencia/inexistencia de overfitting del mejor modelo seleccionado
  12. Cargar correctamente un archivo JSON con los datos con los cuales será probado el modelo
  13. Generar, correctamente, predicciones en archivo JSON usando el mejor modelo obtenido
- Consideraciones generales
    1. En el notebook se identifica a los integrantes del equipo
    2. Considera buenas prácticas en el desarrollo del notebook
    3. Conclusiones y comentarios carecen de faltas a la ortografía
    4. Conclusiones y comentarios se redactan en tercera persona
    5. Todos los gráficos presentan títulos y etiquetas en sus ejes

**NOTA** Cada ítem de la lista de cotejo se evalúa con presencia o ausencia.