



Taller práctico

Apertura: miércoles, 20 de agosto de 2025, 00:00 **Cierre:** lunes, 15 de septiembre de 2025, 18:00



Taller Final: Predicción y Clasificación en la Industria Azucarera

IMPORTANTE: Este taller se debe realizar en grupos de mínimo 2 personas y máximo 4 personas.

Por favor abstenerse de compartir los datos del taller en repositorios públicos. Ustedes pueden compartir o publicar el código, pero no los datos.

Contexto y Objetivos

El **Ingenio La Providencia** cuenta con un conjunto de datos históricos que recoge información relevante sobre la producción de caña de azúcar. Este taller tiene como objetivo que ustedes, como estudiantes, apliquen los conocimientos adquiridos en el curso para desarrollar modelos que permitan desarrollar las siguientes tareas:

- Regresión (HISTORICO_SUERTES.xlsx):** Predecir dos variables clave:
 - Toneladas de caña por hectárea (TCH):** Indicador de la productividad de la tierra.
 - Porcentaje de sacarosa (%Sac.Caña):** Medida de la calidad de la caña y la cantidad de azúcar extraíble.
- Clasificación (BD_IPSA_1940.xlsx):** A partir de las variables continuas (TCH y %Sac.Caña), crear categorías que permitan clasificar los registros en niveles de desempeño:
 - Para **%Sac.Caña:** Niveles de sacarosa **alto, medio y bajo.**

- Para **TCH**: Niveles **alto, medio y bajo**.

La correcta predicción y categorización de estas variables le permitirá al ingenio tomar decisiones más informadas y eficientes en términos de planificación y manejo de cultivos.

Instrucciones Generales

Parte 1: Modelo de Regresión

Análisis Exploratorio de Datos (EDA):

- Realizar una exploración inicial del conjunto de datos: identificar variables relevantes, detectar valores faltantes y posibles outliers.
- Visualizar la distribución de las variables de interés (TCH y %Sac.Caña).

Desarrollo de Modelos:

- Ajustar modelos de **regresión lineal** para predecir TCH y %Sac.Caña.
- Analizar la significancia de las variables, interpretar los coeficientes y evaluar los supuestos del modelo (linealidad, homocedasticidad, normalidad de errores, etc.).
- Diagnosticar posibles problemas como la multicolinealidad o la presencia de observaciones atípicas, y proponer soluciones (por ejemplo, técnicas de regularización).

Validación y Evaluación:

- Utilizar estrategias de validación (holdout y/o validación cruzada) para evaluar el desempeño del modelo.
- Calcular al menos dos métricas de evaluación apropiadas (por ejemplo, R^2 , RMSE, MAE) y discutir los resultados.

Parte 2: Modelo de Clasificación

Creación de Categorías:

- Definir umbrales para transformar las variables continuas TCH y %Sac.Caña en categorías: **alto, medio y bajo**. Justificar la metodología empleada para definir estos cortes (por ejemplo, percentiles, criterios de negocio, etc.).

Desarrollo de Modelos de Clasificación:

- Ajustar un modelo de **regresión logística** y/o aplicar el algoritmo **K-Nearest Neighbors (KNN)** para clasificar las observaciones según los niveles definidos.
- Empleando regularización, analizar la importancia de las variables predictoras en el desempeño de la clasificación.

Validación y Evaluación:

- Emplear una estrategia de validación (por ejemplo, validación cruzada) y justificar la elección.
- Reportar y analizar métricas de evaluación para clasificación (accuracy, precision, recall, F1-score, kappa.).

Parte 3: Criterios de Evaluación

1. Claridad y Rigor en el Análisis:

- Presentación ordenada del EDA.
- Justificación de las decisiones tomadas durante la transformación y preprocesamiento de datos.

2. Desarrollo y Ajuste de los Modelos:

- Correcta implementación de los modelos de regresión y clasificación.
- Interpretación adecuada de los resultados, incluyendo la significancia de las variables y los supuestos del modelo.

3. Validación y Métricas:

- Selección y aplicación correcta de las estrategias de validación.
- Uso adecuado de las métricas de evaluación y análisis crítico de los resultados obtenidos.

4. Comunicación de Resultados:

- Informe escrito claro y conciso que explique el proceso, resultados y conclusiones.
- Visualizaciones y gráficos que respalden el análisis realizado.

5. Innovación y Propuesta de Mejora:

- Propuestas para mejorar el modelo (por ejemplo, la inclusión de técnicas de regularización o selección de variables) y discusión sobre posibles limitaciones.

Entregables

- **Código Fuente:** Archivo(s) con la implementación del análisis y modelado (Jupyter Notebook).
- **Informe Final:** Documento (de no más de 5 páginas) en el que se detalle el proceso completo del análisis, resultados, interpretaciones y conclusiones. No anexar el código al documento.

del análisis, resultados, interpretaciones y conclusiones. No anexas el código al documento.

Recomendaciones Finales

- Asegúrense de documentar bien el código y de comentar las secciones relevantes para facilitar la comprensión del proceso.
- Revisen y apliquen los conceptos vistos en clase, especialmente los relacionados con la validación, las métricas de evaluación y las técnicas de regularización.
- Ser críticos en el análisis de resultados y propongan mejoras o futuras líneas de trabajo basadas en los hallazgos.



19 de junio de 2025, 14:53

[BD_IPSA_1940.xlsx](#)



19 de junio de 2025, 14:53

[HISTORICO_SUERTES.xlsx](#)



Agregar entrega

Estado de la entrega

Estado de la entrega	Todavía no se han realizado envíos
Estado de la calificación	Sin calificar
Tiempo restante	25 días 2 horas restante
Última modificación	-
Comentarios de la entrega	▶ <u>Comentarios (0)</u>