

DATA CHALLENGE 2022



Descripción del caso y de los datos

Competencia dirigida
a la comunidad de
ciencia de datos.


Apoya:



Inscripción: Del 10 de octubre al 4 de noviembre
Inicio de la competencia: 8 de noviembre
Bases e inscripción: Web de ANCAP / www.ancap.com.uy

The background of the slide is an abstract composition of various shades of green and yellow. It features a series of overlapping, semi-transparent geometric shapes, primarily triangles and quadrilaterals, which create a sense of depth and movement. The colors range from dark, forest green to bright, sunny yellow, with many intermediate tones. The overall effect is a complex, layered pattern that changes as the viewer's perspective shifts.

Descripción del caso

The background of the slide is an abstract composition of various shades of green and yellow. It features a series of overlapping, semi-transparent geometric shapes, primarily triangles and quadrilaterals, which create a sense of depth and movement. The colors range from dark forest green to bright, almost white-yellow, with many intermediate tones. The overall effect is a complex, layered pattern that resembles a modern digital art style or a close-up of a textured surface.

Descripción de los datos

Registros de variables

Variables Predictoras

- Variables de control de la operación (de entrada)
- Variables para posibilitar a reconciliación del balance de masa.
- Las variables pueden representar:
 - Presiones
 - Temperaturas medidas en distintas partes de la torre y en otros sitios adicionales
 - Caudales
- Se cuenta con aproximadamente 5 años de registro de datos, cada un minuto

Variables a predecir

- Contenido de “C7+”
- Contenido de precursores de benceno
- La medición de las 2 variables objetivo, se obtienen mediante análisis de laboratorio, el cual insume varias horas, con distintas frecuencias, desde 1 vez por día hasta 3 veces por semana.

Correlaciones conocidas

Variables predictoras que impactan en mayor grado:

- La composición de la carga (en particular el contenido de benceno),
- Temperatura del plato sensible (TI-21049),
- Temperatura de salida del reboiler 2105-C (TI-21050),
- Temperatura de cabeza (TIC-21047).

De impacto menor:

- Presión de operación (PIC-21034),
- Temperatura de la carga (TI-21046),
- Niveles (LIC-21022 y LI-21018).

Datos a disponibilizar

- Los datos se pondrán a disposición en archivos .csv

Variables de control:

- El formato de la información será fecha-hora (timestamp) del registro, y cada uno de los valores de las variables predictoras del proceso.
- Los valores son detallados cada un minuto.
- Se dispondrá además descripciones y unidades de las variables.

E

TimeStamp	tdc_phd.pi_21030	tdc_phd.pic_21034	tdc_phd.pi_21035	tdc_phd.ti_21046	tdc_phd.tic_21047	tdc_phd.ti_21048
2017-01-01 00:00:00.000	0.81798	0.83386	0.90135	103.9000	75.04008	46.70211
2017-01-01 00:01:00.000	0.81787	0.83426	0.90164	103.9122	75.02803	46.88399
2017-01-01 00:02:00.000	0.81777	0.83466	0.90194	103.9485	74.91962	46.88803
2017-01-01 00:03:00.000	0.81767	0.83505	0.90223	103.9243	74.96780	47.01734
2017-01-01 00:04:00.000	0.81756	0.83545	0.90252	103.9000	74.90757	46.37471
2017-01-01 00:05:00.000	0.81746	0.83585	0.90282	103.9243	75.08826	46.76272
2017-01-01 00:06:00.000	0.81735	0.83624	0.90311	103.9485	75.18163	46.99309

Información adicional para Balance de masa

Datos a disponibilizar

Variables a predecir para entrenamiento y testeo

- Resultados de los análisis de laboratorio para el valor de los Precusores de Benceno
 - Con la fecha-hora del análisis, valor y el nombre del archivo origen de la información del análisis de laboratorio. (análisis del período 11/2017 a 9/2021)
- Resultados de los análisis de laboratorio para el Contenido C7 Total
 - Con la fecha-hora del análisis y el nombre del archivo origen de la información del análisis de laboratorio. (análisis del período 11/2017 a 9/2021)

Ejemplo:

Fecha	Hora	Contenido.C7.Total	archivo
2021-05-03	07:00	1.23	CARGA_ISOM_CAB_2102E.xlsx
2021-05-07	07:00	0.90	CARGA_ISOM_CAB_2102E.xlsx
2021-05-10	07:00	1.23	CARGA_ISOM_CAB_2102E.xlsx
2021-05-11	07:00	1.05	CARGA_ISOM_CAB_2102E.xlsx

Datos a disponibilizar

Variables a predecir

- Timestamp de los resultados de las de los análisis de laboratorio para el valor de los Precusores de Benceno
 - Período 10/2021 a 7/2022
- Timestamp de los rresultados de los análisis de laboratorio para el Contenido C7 Total
 - Período 10/2021 a 7/2022

Deben ser entregados con los valores estimados.

Ejemplo:

Fecha	Hora
-------	------

Información a entregar

Notebook requisitos obligatorios:

- Que sea reproducible (dados los datos de entrada, genere las predicciones como salida) y que se vea claramente el uso de los datos proporcionados (en las secciones preparación de datos, creación del modelo, entrenamiento y generación de las salidas).

Notebook requisitos opcionales:

- Que muestre la exploración de propiedades de los datos generando insights, la existencia de gráficas mostrando la performance del modelo, etc. y la existencia de textos (comentarios) aclaratorios.

Archivo con valores a predecir de ambas variables completo.

Criterio de selección

- Se seleccionarán hasta cinco soluciones con más predicciones en el rango valor real $\pm 10\%$ y, entre ellas, la ganadora será la que minimice el error RMSE con los resultados obtenidos por laboratorio.
- En caso de que un ganador no cumpla con los requisitos de las bases, será descalificado por incumplimiento y podrá ser premiado el que sigue en orden en calificación



¿Dudas?

DATA CHALLENGE ANCAP 2022

Podes contactar al equipo organizador en datachallenge@ancap.com.uy