

Enunciado Caso

El propósito de este proyecto es brindar a usted y a su equipo la oportunidad de enfrentarse a un problema real y resolverlo utilizando herramientas de ciencia de datos. En esta ocasión, se abordará un problema de clasificación que busca determinar si un(a) estudiante reprobará o aprobará un curso.

Para poder realizar esta labor, se les entregan los siguientes archivos de datos:

- **'logs_entrenamiento.csv'** : Este archivo tiene información respecto a la actividad que presentan los/as estudiantes dentro de la plataforma EOL para un ramo obligatorio de la FCFM.
- **'notas_por_subseccion.csv'** : Este archivo tiene las notas de evaluaciones parciales.

El trabajo que se les encomienda es poder cargar los datos, procesarlos de manera de tener una base de datos *"limpia"* con la cual puedan entrenar diferentes modelos de aprendizaje supervisado que le permitan responder la pregunta del negocio:

¿Cuáles estudiantes reprobarán el curso?

Para evaluar este caso se considerarán los siguientes puntos:

1. **Informe Avance (NIA)** : Se debe entregar un documento **.pdf**, en donde se detallen los avances realizados en el proyecto grupal. Lo anterior incluye análisis que se hayan realizado, iteraciones que haya tenido el proyecto y hallazgos relevantes de su trabajo. La estructura mínima para el informe es la siguiente:
 - Breve descripción del problema que se aborda.
 - Imputaciones y tratamiento de datos (análisis descriptivo y outliers).
 - Transformación y escalamiento de atributos.

Extensión máxima: 4 páginas sin contar portada o anexos.

Fecha de entrega: Lunes 6 de Noviembre

2. **Vídeo (NV)** : Esta entrega reemplaza una instancia de presentación. El objetivo es presentar el proyecto desarrollado de manera que todos puedan ver cuáles fueron los principales problemas y desafíos del proyecto, y cómo se resolvieron. Cada integrante del grupo debe presentar una parte de esta presentación. La estructura recomendada para la presentación es la siguiente:
 - Breve descripción de la empresa u organización (foco en el área del proyecto) y problema que se aborda.
 - Descripción (breve) de la base de datos que les fue entregada.
 - Acciones ejecutadas con la base de datos.
 - Evaluación y validación de los modelos desarrollados.
 - Comparación de los resultados obtenidos con la situación actual de la empresa u organización.

Extensión máxima del vídeo 7 minutos

Fecha de entrega: Domingo 3 de Diciembre

3. **Informe Final (NIF)** : Se debe entregar un documento **.pdf**, en donde se detalle lo realizado, justificando las decisiones tomadas en cada una de las etapas, así como también las conclusiones alcanzadas por el grupo con respecto al trabajo realizado y cómo éste agrega valor. La estructura es libre, pero debe contener las siguientes formalidades:
- Resumen Ejecutivo/Abstract
 - Introducción
 - Desarrollo
 - Conclusiones

El desarrollo/cuerpo del informe final debe contener:

- Descripción de la base de datos que les fue entregada
- Acciones ejecutadas con la base de datos
- Evaluación y validación de los modelos desarrollados
- Relevancia de su trabajo: en base a los resultados, señalar implicancias para el cliente y el usuario.

Se recomienda escribir el informe final en Latex usando el [template ACL de dos columnas](#)

Extensión máxima: 6 páginas sin contar portada o anexos.

Fecha de entrega: Domingo 3 de Diciembre

4. **Evaluación de desempeño (D)**: Este proyecto forma parte de una investigación en curso, por lo que ya se cuenta con un modelo que responde a esta problemática, por lo cual la solución que usted entregue debe ser mejor a esta.

Al mismo tiempo, usted se encuentra en una situación en la que existen otros consultores buscando solucionar el problema de mejor manera. Es por ello que los resultados del grupo serán evaluadas siguiendo el siguiente criterio:

- Métrica objetivo: Se utilizará el [F1-score](#).
- Rendimiento del modelo: Su modelo debe tener un resultado mejor a **0.88**.
- Para la evaluación de resultados se les habilita el siguiente [link](#) en el cual deben subir sus resultados. Para poder hacer un “submit” en la plataforma deben aplicar su modelo entrenado sobre los datos del archivo “**logs_validación.csv**”.
- La entrega de los resultados funcionará con 5 submit máximos por día hasta la fecha de finalización de la competencia.
- La nota se calculará aplicando una escala lineal entre el puntaje *baseline* y el mejor puntaje de la competencia. Por ejemplo, si el mejor puntaje del curso fuera 0.98 y su grupo obtuvo un puntaje de 0.93 (justo en la mitad), obtendrá un 4 en este ítem.

Fecha de entrega: Domingo 26 de Noviembre.

5. **Código (C)**: Junto a cada informe deben entregar un Jupyter Notebook (archivo **.ipynb**) donde quede claro el proceso que se realizó sobre los datos. Este documento será sujeto de dos macro-correcciones:
- Revisión funcional : Se ejecutará el código completo, y el código debe **ejecutarse completamente**, entregar los **mismos resultados mostrados en el informe y no presentar ningún error**.
 - Revisión integral : Se comparará su código con el de sus pares, en caso de haber sospecha de copia de código entre grupos se investigará la situación y se sancionará fuertemente en caso de comprobar dicho comportamiento.

En caso de que su trabajo no cumpla con este punto se les aplicará un descuento sobre la nota grupal (NG) de entre 0.5 y 2 puntos.

Cálculo de nota

La nota grupal del caso se calculará de la siguiente manera:

$$\text{Nota Grupal (NG)} = (0.1 * NIA + 0.3 * NIF + 0.3 * NV + 0.3 * D) - C$$

Además, se considera una coevaluación grupal. Para dicho proceso se habilitará un documento en el momento adecuado, a continuación se detalla el funcionamiento de la coevaluación:

- Cada estudiante debe asignar 100 puntos entre las y los integrantes del grupo (incluido quien evalúa). Una mayor asignación de puntos se relaciona con un mayor aporte al proyecto.
- Al final del proceso de coevaluación, cada estudiante tendrá un puntaje total de aporte al desarrollo del trabajo (CT), correspondiente a la suma de los puntajes que les fueron asignados por el resto del grupo.
- La nota del proyecto se les calculará de la siguiente manera:

$$\text{Nota Personal (NP)} = 1 + CT * ((NG - 1)/100)$$

Finalmente, la nota del caso viene dada por:

$$\text{Nota Caso (NC)} = 0.25 * NG + 0.75 * NP$$

Un ejemplo de autoevaluación es el que se indica:

<i>Integrante</i>	<i>CT</i>	<i>NP</i>
1	100	5.5
2	125	6.9
3	55	3.1
4	120	6.6

Se tiene un grupo de 4 integrantes, con Nota Grupal (NG) igual a 5.5. La tabla muestra el puntaje total recibido por estudiante (CT) y la nota personal (NP). Se observa que la suma de los CT es igual a 400. Esto debe ser así, pues cada evaluador/a debe asignar un total de 100 puntos entre los 4 integrantes del grupo. Finalmente, si el grupo decide que cada integrante mantendrá la nota grupal, cada evaluador/a debe asignar la misma cantidad de puntos a cada integrante del grupo. Es decir, cada evaluador/a deberá asignar 25 puntos a cada integrante del grupo.

Consideraciones

Recuerde que la autoevaluación grupal mide el grado de contribución personal al trabajo, y no el resultado de éste. Un integrante no podrá asignarse a sí mismo más del 50% de los puntos. Por ejemplo, si el grupo tiene 4 integrantes hay 100 puntos para repartir, el o la estudiante no podrá asignarse más de 50 puntos.

Situaciones excepcionales serán analizadas con el cuerpo docente. Finalmente, es importante indicar que la autoevaluación de cada estudiante es tratada de forma confidencial, y serán revisadas sólo por el profesor y profesor auxiliar del curso. De esta manera, la información final que poseerá cada

estudiante corresponderá a la nota final obtenida en esta evaluación, que estará relacionado al total de puntaje recibido.

Formulario de autoevaluación

Para realizar la autoevaluación se dispondrá de un formulario especial que deberá completar cada estudiante. Este formulario estará disponible en la sección “Tareas” de u-cursos en el módulo de entrega de la autoevaluación. Recuerden que el total de puntos que deben distribuir es 100.