

Trabajo sobre caso “Enfermedad Renal Crónica”

Este es el ejercicio de competencia del curso o denominado como la tarea académica 1 (TA1). El desempeño grupal de este trabajo incidirá en la nota final del curso, así como en el examen final.

En este ejercicio se pide que se “identifiquen” qué pacientes dentro de la base de datos **ERC_Test.csv** debe recomendárseles realizar una “prueba de detección” (*screen test*) para la enfermedad ERC (**Enfermedad Renal Crónica**). Los pacientes de la base de datos **ERC_Train.csv** son similares (*en cuanto estructura*) a los de la base de datos de prueba y deben ser la base para el aprendizaje de sus modelos.

La finalidad de la competencia consiste en determinar qué pacientes deben tomar la prueba de detección y subirlos a Kaggle para determinar la “rentabilidad” de sus datos.

Cada grupo deberá entregar como resultado de este trabajo cuatro (04) documentos:

- 1 archivo PDF describiendo detalladamente los pasos que se siguieron para obtener las respuestas cuyos resultados obtuvieron la más alta rentabilidad. Se deberán utilizar el formato de informe, contemplando, carátula y porcentaje de participación, índice, introducción, definición de la problemática, marco teórico y metodológico, resultados, conclusiones y recomendaciones. Asegurar incluir una fuerte argumentación estadísticas en sus decisiones.
- 1 archivo CSV o TXT que contenga para cada uno de los pacientes de los datos provistos por **ERC_Test.csv** un “Id” y una variable binaria llamada “CKD” que indica su recomendación para ese paciente. La primera fila de este archivo debe contener los nombres de las variables. Las demás filas deben contener los valores para esas dos columnas. Esto corresponde al archivo que se utilizará para subir a la plataforma en Kaggle.
- 2 archivos, uno por modelo, de los modelos Seleccionados (*los de mayor rentabilidad*) en Orange (*se pueden incluir archivos en R, Python y/o SQL Server si fueron necesarios*).

Soluciones que no puedan recrearse serán descalificadas, es decir, solo se aceptarán resultados que provengan de forma natural con la aplicación de un modelo (que supone una construcción lógica a priori). Es importante advertir que toda decisión de modelamiento debe ser documentada ordenadamente para tales fines.

Es preciso resaltar que este no es un caso para una escuela de medicina, sino para una empresa que se dedica a proveer servicios de salud. La tabla siguiente reporta como la plataforma de InClassKaggle.com evaluará las soluciones presentadas.

	Padece de ERC	No padece de ERC
Se recomienda la prueba	+ 1200	- 120
No se recomienda la prueba	- 360	0

En su documentación de respuestas de texto (PDF) asegúrense de discutir la lógica de esta matriz de pagos. ¿Les hace sentido?

Reglas Sobre La Competencia

- Todos los días cada grupo puede subir a InClassKaggle un máximo de 10 soluciones. InClassKaggle les brindará información sobre la calidad de la solución en el 30% de la base de datos.
- Las posiciones finales y una parte substancial de la posición relativa entre todos los grupos que compiten en este trabajo están determinadas por los resultados del otro 70% (datos no públicos) de los datos.
- Al final de la competencia, cada grupo deberá seleccionar 2 de las múltiples soluciones como las selecciones oficiales, de lo contrario se escogerán de forma automática en función solamente al 30% de los resultados públicos.
- Si las soluciones de cada grupo no pueden ser reproducidas exactamente con ayuda de los programas que se especifiquen en su trabajo de forma “a priori” serán descalificadas y no se tendrá en cuenta para la bonificación.
- En caso usen un modelo que no haya sido cubierto en el curso, deberán explicar la teoría del mismo. El uso de temas que hayan sido investigados por el grupo de forma independiente recibirá bonificación adicional si se justifican de manera correcta.