



Copyright © 2021 NeuralWorks. Confidential and proprietary

# Challenge Machine Learning Engineer

## Instrucciones

- Debes entregar tu solución en un repositorio GitHub
- En el repositorio deben estar todos los archivos utilizados para la resolución de tu desafío. - La solución debe estar implementada en un notebook .ipynb utilizando python 3, indicando claramente la pregunta que estás resolviendo. No serán revisados otros lenguajes como R o similar.
- Recuerda que no estamos en tu cabeza! Escribe los supuestos que estás asumiendo. - Para este desafío te recomendamos que describas claramente cómo mejorar cada parte de tu ejercicio en caso de que tenga opción de mejora.
- Debes enviar el link al repositorio vía mail a [rayen@neuralworks.cl](mailto:rayen@neuralworks.cl) contestando el correo en el que se te envió este enunciado.

## Problema

Te encuentras trabajando en un equipo conformado por varios Data Scientists. Uno de ellos, Juan, ha terminado sus experimentos y te ha pedido que habilites un servicio para que su modelo pueda ser consumido por varios actores fuera del equipo.

Juan escribió en un Jupyter Notebook una serie de modelos para predecir la probabilidad de atraso de los vuelos que aterrizan o despegan del aeropuerto de Santiago de Chile (SCL). Para eso utilizó un dataset de datos públicos y reales donde cada fila corresponde a un vuelo que aterrizó o despegó de SCL.

Tu desafío consiste en tomar el trabajo de Juan y exponerlo para que sea utilizado por el resto de la compañía:

- Escoge el modelo que a tu criterio tenga una mejor performance, argumentando tu decisión.
- Implementa mejoras sobre el modelo escogiendo la o las técnicas que prefieras.
- Habilita el modelo seleccionado como API REST para ser expuesto.
- Haz pruebas de estrés a la API con el modelo expuesto con al menos 50.000 requests durante 45 segundos. Para esto debes utilizar [esta herramienta](#) y presentar las métricas obtenidas.
- ¿Cómo podrías mejorar la eficiencia de las pruebas anteriores?