

# Taller de Aprendizaje Automático

Instituto de Ingeniería Eléctrica - FING - UdelaR

## Proyecto 2 – 2024

### *Freesound Audio Tagging*

## Descripción

El proyecto consiste en diseñar e implementar un modelo de Audio Tagging para determinar automáticamente las fuentes de audio presentes en una grabación. Se trabaja en grupos de tres estudiantes.

## Objetivos de aprendizaje

Para la resolución del proyecto debería ser capaz de realizar lo siguiente.

1. Explorar y visualizar los audios y sus espectrogramas para identificar aspectos relevantes que permitan caracterizar el problema.
2. Realizar una lectura eficiente de los audios mediante la utilización de estructuras que permitan el manejo de grandes volúmenes de datos.
3. Implementar, entrenar y refinar modelos de AudioTagging.
4. Indicar por qué es razonable abordar este problema de clasificación de audio utilizando *redes neuronales convolucionales* aplicadas sobre los espectrogramas.
5. Comparar al menos dos arquitecturas diferentes de *redes neuronales convolucionales* aplicadas sobre los espectrogramas para resolver el problema.
6. Implementar alguna variante de la técnica *transferencia de aprendizaje*.
7. Proponer e implementar al menos una forma de utilizar los datos *ruidosos* para el entrenamiento del modelo.
8. Realizar *aumento de datos* durante el entrenamiento y evaluar su impacto en el desempeño del modelo.
9. Manejar adecuadamente conjuntos de entrenamiento, validación y prueba.
10. Identificar el sobreajuste y sub-ajuste de un modelo a los datos.
11. Aplicar técnicas de regularización y de selección de hiperparámetros.
12. Para algunas instancias y clases del conjunto de validación, interpretar las valoraciones realizadas por el clasificador.

13. Diseñar y registrar experimentos usando una herramienta como *Comet*.
14. Evaluar cuantitativamente el desempeño de los modelos. Discutir y analizar resultados.
15. Subir los resultados con el conjunto de datos de test al *InClass* de la plataforma *Kaggle* asociado al proyecto.

## Informe y entrega

El documento no debe exceder las **8 páginas**. El informe debe presentar muy brevemente el problema abordado y describir en detalle la solución desarrollada. Se deben justificar las decisiones tomadas en el diseño de la solución, considerando aspectos como la transferencia de aprendizaje, el aumento de datos, la comparación entre arquitecturas (al menos dos); y cualquier otra estrategia que el equipo haya considerado relevante. Se valorará especialmente la discusión de los resultados y las conclusiones.

Deben incluir en la entrega (junto al informe) el modelo final con un código que permita ejecutarlo sobre los datos de test. Además, deben entregar un registro de los principales experimentos realizados (link a *Comet*) junto al código usado para entrenar su mejor modelo.

Recuerden hacer referencia en el texto a las figuras que incluyan, y que cada figura cuente con una nota de descripción. Tengan la precaución de que las etiquetas y números de las figuras sean legibles. Puede ser conveniente resumir los resultados de desempeño usando tablas. Incluyan una sección de referencias bibliográficas, y citen cuando crean necesario. Las páginas que contengan solamente carátula o referencias no serán contabilizadas para el límite de 8.

## Defensa

Cada grupo de estudiantes tendrá **10 minutos** para hacer una presentación oral de su trabajo (seremos estrictos en cuanto al tiempo) seguido de una sesión de preguntas de unos **20 minutos**. En la defensa el grupo deberá presentar su solución al problema planteado, así como el proceso que condujo a ella, los principales experimentos realizados haciendo énfasis en el análisis de resultados. Se valorará el análisis de la solución vinculado con los conceptos y técnicas vistas en el curso. No es necesario describir el problema en la presentación.

## Fechas

- Entrega de informe: martes 9 de julio, 13:00 hs.
- Defensas: jueves 11 de julio, de 8:00 hs. a 15:00 hs.