



# Discusión de proyecto 2



# Discusión de proyectos

---

- Datos públicos para **clasificación de texto** (solo texto, no usar otras variables aunque tengan disponibles). Puede ser binario o multiclase y debe constar en texto y etiqueta (si tiene más variables, omitirlas).
- Presentar el tema oficial a más tardar el miércoles 6 de noviembre (mediante mediación virtual)
- Qué deben hacer? Utilizando python y lo visto en los laboratorios:
  - Preparación de datos para red neuronal (division de datos en entrenamiento/validación/test, limpieza y pre-procesamiento de texto)
  - Construir al menos 3 redes neuronales con distintas arquitecturas para clasificar las imagenes (**MLP/CNN y al menos una RNN o transformer**)
  - En cada modelo, al menos optimizar uno de los hiperparámetros (e.g intentar distintos tipos de optimizador o de valores de tasa de aprendizaje, número de neuronas/capas)
  - Utilizar embeddings o BoW
  - Medir el rendimiento en un set de datos separado (test)
  - Crear visualizaciones y tablas de resumen sobre los resultados que den una idea del rendimiento del modelo (ejemplos de texto bien/mal clasificado, matriz de confusion, curvas de aprendizaje, table resumen de métricas más relevantes)
  - Al menos una referencia de otro modelo realizado con esos datos y el rendimiento que se ha obtenido
- Presentación:
  - duración máxima 15min + 5min de preguntas (todas las personas del grupo deben presentar)
  - Orden de presentación de los grupos será aleatoria

# Discusión de proyectos

---

- Documento entre 5 y 10 páginas con las siguientes partes:
  - Introducción y motivación (cuál es el problema? Por qué es interesante? **Qué se ha hecho con esos datos y cuál ha sido el rendimiento?**)
  - Datos y Métodos (descripción de los datos que se van a usar y los modelos, visualizaciones de exploración de datos, etc)
  - Resultados (métricas de rendimiento, gráficos y comparación)
  - Conclusión (cuál es el mejor modelo? Cómo se pueden mejorar los resultados? Qué se aprendió del problema?)
  - Referencias (formato APA)
- El documento debe ir escrito en lenguaje formal y académico. Los gráficos, imágenes y tablas debidamente etiquetadas y numeradas con títulos descriptivos y referencias en el texto. Asegurarse de la legibilidad del documento. **De preferencia en el formato necesario para la revista Serengeti**
- En mediación virtual, deben entregar:
  - Diapositivas de la presentación en pdf
  - Documento en **pdf**
  - Código de python (.py o .ipynb)
  - NO enviar el set de datos
- Evaluación:
  - Presentación 40% del valor total
  - Documento 60% del valor total
  - Verificación de trabajo en equipo: Existirá un sistema en el que me indicarán el porcentaje de participación de cada estudiante, acordada por todos. Esta se usará para escalar la nota final de cada persona