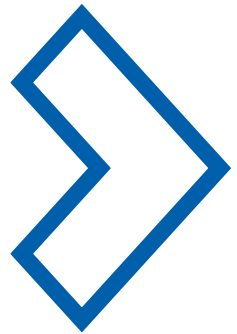


Fundamentos del Procesamiento de Lenguaje Natural

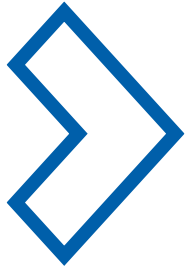


Entrega E3 – Primera Iteración de Resolución del Problema



Resumen de la Entrega

- Corresponde al **15% de la nota final** de la asignatura.
- Debéis **subir la entrega a ALUD** antes del **28/11 a las 23:59**.
- Para el código, podéis subir el **notebook** a **ALUD**, o mejor aún, crear un **repositorio de GitHub** para el proyecto.
- En cualquier caso, el código tiene que estar **adecuadamente documentado**. Es decir, no puede contener sólo código (y comentarios): debéis **describir y documentar las cosas en profundidad** (lo que hacéis, por qué lo hacéis, etc.)



Instrucciones de la Entrega

- El propósito principal de esta entrega es **empezar con el desarrollo** de la resolución de **vuestro problema...**
- Se llevará acabo la **primera iteración o aproximación** a la resolución de los problemas planteados en la E1, y con el *dataset* y las *features* extraídas en la E2.



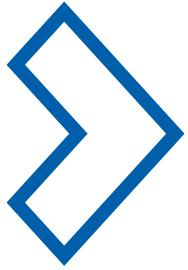
Instrucciones de la Entrega

- Para ello, se deberán usar técnicas tanto de **Shallow ML** (o ML tradicional), como algunos de los modelos de **CNNs o Redes Recurrentes** que hemos visto en clase.
- Por tanto, esta entrega debería contener los siguientes apartados...



Instrucciones de la Entrega

- Código y documentación de los **experimentos** que se han llevado a cabo **para cada una de las tareas** a abordar...
- Tabla comparativa de los **resultados obtenidos** con los distintos enfoques, usando las **métricas adecuadas y específicas** para cada caso. Es decir, considerando, e.g., el *problema* a resolver (clasificación, traducción, *summarization*, predicción complementaria de ST, etc.), la *distribución* de datos disponibles, etc.
- En caso de haber **optimizado hiperparámetros**, justificar cómo.



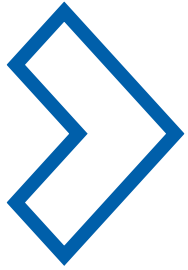
Instrucciones de la Entrega

- En caso de usar **modelos** (de clasificación/predicción/lo que sea) **diferentes** a los que se han visto en clase, justificar su uso, **referenciar su origen** y **explicar su funcionamiento**.
- En caso de **combinar datos** de **diversas tipologías**, **explicar cómo**.



Instrucciones de la Entrega

- Se valorará positivamente:
 - El uso de **atención**.
 - Combinar **distintos tipos de features** (e.g., no sólo texto).
 - Pre-entrenar los modelos con **otros datasets parecidos** para analizar si eso **aumenta/empeora el rendimiento** de vuestros modelos.
- Y otras **ideas, metodologías, o aproximaciones** al problema que **vayan más allá** de lo visto en clase 😊



Instrucciones de la Entrega

- **Mínimos exigibles indispensables:**
 - Dos técnicas de *Shallow Learning* utilizando técnicas de **representación dispersa/sparse**.
 - Dos técnicas de Deep Learning comparando **diferentes tipos de embeddings y fine-tuneandolos** dependiendo del caso. Ejemplos:
 - Word2Vec congelado vs Word2Vec fine-tuneado vs Word2Vec “from scratch”
 - Embedding fine-tuneado durante el entrenamiento vs Embedding inicializado
 - Comparar **al menos dos formas de embeddings** de cada tipo:
 - **Tradicionales:** e.g., Bag-of-Words, TF-IDF, etc.
 - **Semánticos No-Contextuales:** e.g., Glove, FastText, Word2Vec, etc.
 - **Contextuales:** e.g., ELMo, BERT, Modelos pre-entrenados de Hugging-Face, etc.



Instrucciones de la Entrega

- **Importante:** Esto es un mínimo absoluto de lo que se espera de esta entrega...
- Además, debido a la variedad de los proyectos, muchos de ellos exigirán que **adaptéis** o llevéis a cabo **pasos no mencionados** anteriormente que son específicos a vuestro problema.
- **¡No os limitéis a seguir estas instrucciones al pie de la letra!**



Instrucciones de la Entrega

- **Importante:** Como hemos comentado, esta entrega se plantea como una **primera iteración** de la resolución del problema. Por lo tanto, **no tenéis que limitaros** únicamente al uso de Shallow ML, CNNs/RNNs, etc.
- Si queréis, podéis empezar a trabajar con **Transformers** 😊
- Además, si tenéis intención de realizar **varias tareas**, podéis, por supuesto, desarrollar algunas de ellas **al completo** y entregar una **primera aproximación** de las otras.
- Eso sí, **todas las tareas tienen que aparecer en esta entrega.**

 ¿Preguntas?