

Proyecto final

En el proyecto final diseñarán un modelo de inteligencia artificial para aplicaciones en ciencias de la salud. Cada equipo tiene la libertad de seleccionar la problemática que quieran resolver, para ello seleccionarán un paper que aborde y detalle el problema que van a investigar.

Su proyecto puede basarse en visión computacional, procesamiento de lenguaje natural, datos tabulares o un modelo multimodal que convine varios tipos de datos. En caso de trabajar con imágenes o datos tabulares, está el requerimiento que entre la familia de modelos para su solución deben incluir una arquitectura diferente a los modelos ya estudiados en el semestre (XGBoost, LightGBM, Catboost y YOLO).

El proyecto será evaluado en base a los siguientes rubros:

1. Investigación de un paper
2. Construcción de un modelo de aprendizaje automático
3. Presentación
4. Coevaluación

Investigación de un paper

Por equipo, seleccionarán un paper que aborde la problemática que les interesa investigar. Deben detallar cómo las y los investigadores de la publicación abordaron el problema, indicando cuáles fueron sus metodologías y resultados. Adicional a explicar la solución del paper, también deben indicar cuáles son las limitaciones y áreas de oportunidad que identificaron del paper.

Construcción de un modelo de aprendizaje automático

En su diseño y construcción del modelo se evaluarán los siguientes aspectos:

I. Recolección y procesamiento de datos

- a. Detallen cuáles fueron las técnicas de recolección empleadas para construir su set de entrenamiento y validación. Su set de datos debe incluir por lo menos una base de datos pertinente al problema que están abordando, en su repositorio indiquen de dónde la obtuvieron. Efectúen un análisis exploratorio de datos en el que grafiquen e inspeccionen los datos recolectados. Identifiquen cuáles son las limitaciones y sesgos de los datos que están utilizando. Determinen si su distribución y cantidad

son apropiadas para entrenar su modelo. En caso de que la distribución de los datos recolectados no sea equitativa, mencionen las técnicas de aumentación de datos que usaron. Mencionen y justifiquen el procesamiento de datos que efectuaron para disminuir la cantidad de ruido y errores presentes.

II. Selección y entrenamiento de modelos

- a. Investiguen el estado del arte respecto a modelos de inteligencia artificial para la problemática que están analizando. Especifiquen y justifiquen los modelos que seleccionaron para el proyecto. Identifiquen cuáles son las limitaciones de los modelos seleccionados. Detallen su plan de entrenamiento indicando los hiper parámetros, la función de costo y el método para ajustar parámetros. Mencionen las técnicas para evitar el sobre y sub ajuste que incorporaron en su modelo o técnica de entrenamiento.

III. Evaluación y métricas de desempeño

- a. Identifiquen cuáles son las métricas más apropiadas para medir el desempeño de su modelo. Construyan matrices de confusión para evaluar los clasificadores que diseñaron.

IV. Despliegue y reproducibilidad del modelo

- a. Construyan un repositorio de github como control de versiones del proyecto que incluya instrucciones de cómo correr su código y los enlaces a las bases de datos que utilizaron. El repositorio debe incluir un requirement.txt que incluya todos los paquetes que se requieren instalar para correr su código. El .ipynb debe ser desarrollado en Google colab o en un servidor de Kaggle. Incluyan un README que contenga la documentación apropiada para reproducir sus experimentos. Sigán las indicaciones de <https://help.github.com/en/github/creating-cloning-and-archiving-repositories/about-readmes> para generar un README file apropiado.

Presentación

Cada equipo debe grabar una presentación de 30 minutos exponiendo el proyecto que desarrollaron. En la presentación indiquen:

- La problemática que estudiaron
- Análisis del paper seleccionado
- Los datos que recolectaron y el procesamiento que les efectuaron
- La demostración de su modelo corriendo

El video lo deben subir a youtube en alta calidad. También habrá una presentación vía Zoom en la que respondan una serie de preguntas de su proyecto.

Entregables

En canvas van a subir los siguientes archivos y enlaces:

- Repositorio github de su proyecto
- Archivo pptx de su presentación
- Enlace al video de youtube de su presentación

La ponderación de cada uno de los rubros del proyecto es la siguiente:

- 15% investigación
- 40% construcción del modelo de aprendizaje automático
- 25% presentación
- 20% auto y coevaluación

Fecha de entrega: 28 de Noviembre