

PEC 2: Desarrollo del trabajo - Fase 2

Álvaro Román Gómez

5/5/25

Table of contents

1 Avance del proyecto	1
1.1 Grado del cumplimiento de los objetivos	1
2 Relación de las actividades realizadas	2
2.1 Actividades previstas en el plan de trabajo	2
2.2 Actividades no previstas	2
3 Listado de resultados obtenidos	2

1 Avance del proyecto

1.1 Grado del cumplimiento de los objetivos

De los objetivos que estaban previstos para cumplir en esta fase se han alcanzado los siguientes:

- Hiperparametrización del mejor modelo de todos los candidatos de la fase 1: XGBoost.
- Refinamiento del preprocesado de datos y selección de variables utilizando PCA y un algoritmo genético.
- Entrenamiento de una red neuronal convolucional a partir de la representación de las moléculas como grafos.
- “Transfer-learning” de una red neuronal roberta entrenada previamente con un conjunto de 40.000 moléculas de la base de datos ZINC.

2 Relación de las actividades realizadas

2.1 Actividades previstas en el plan de trabajo

- Una vez evaluados los modelos, se ha realizado la hiperparametrización del modelo XGBoost con los fingerprints MACCS. Para ello se ha utilizado la librería optuna.
- Además del filtrado de variables basado en la correlación de las mismas y su varianza, se ha realizado una selección de descriptores moleculares a través de PCA y un algoritmo genético.
- Se ha entrenado una red neuronal convolucional a partir de la representación de las moléculas como grafos. Se ha utilizado para ello la librería deepchem. La hiperparametrización del modelo se ha realizado con optuna.
- Se ha realizado un “transfer-learning” de una red neuronal roberta entrenada previamente con un conjunto de 40.000 moléculas de la base de datos ZINC representadas a través de su SMILES canónico. Se ha utilizado para ello la librería simpletransformers y el modelo ha sido obtenido de la base de datos HuggingFace. Para esta tarea se ha empleado Google Colab.

2.2 Actividades no previstas

- Se ha realizado una revisión de los datos originales debido a los resultados preliminares de la hiperparametrización del modelo XGBoost.

3 Listado de resultados obtenidos

Los resultados obtenidos que se adjutan son los siguientes:

- 05_models_training_and_selection.pdf
- 06_hyperparameter_optimization.pdf
- 07_neural_network_gcn.pdf
- 08_neural_network_roberta.pdf