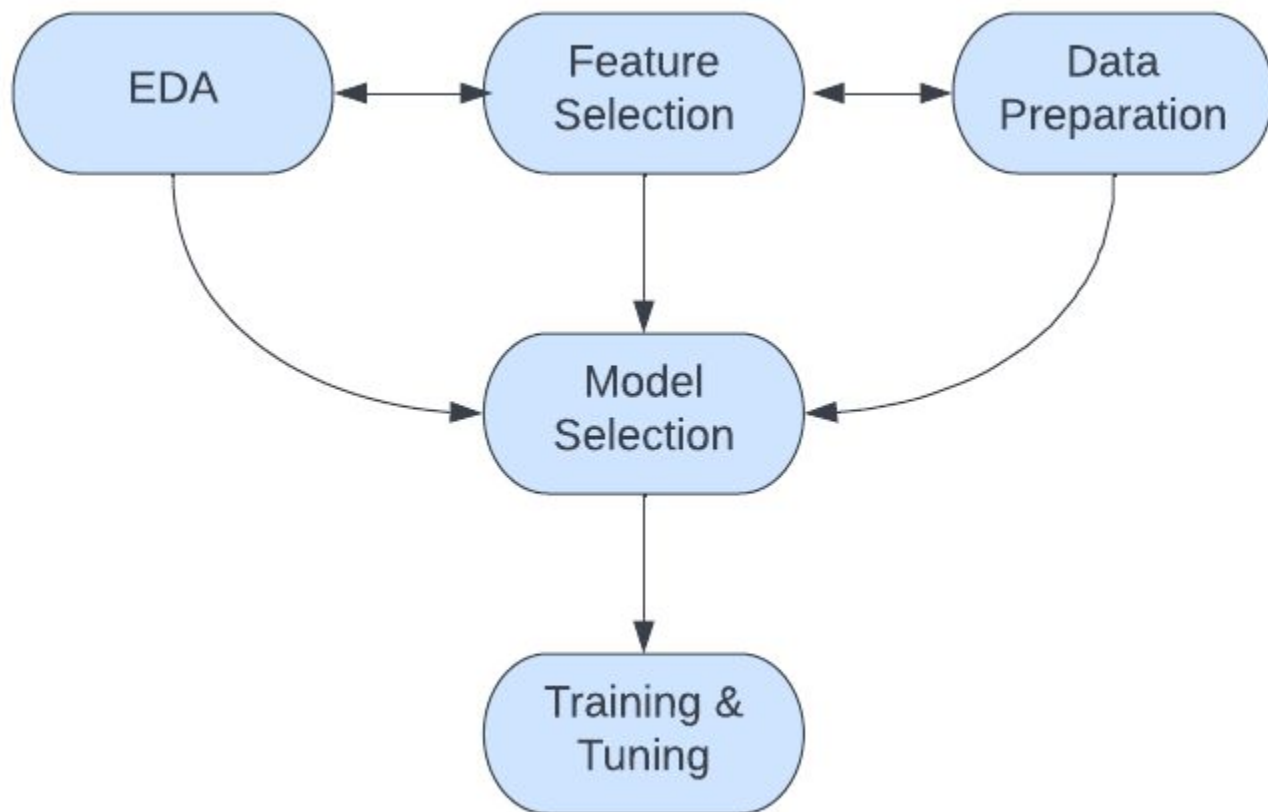




Activity prediction for chemical compounds

Group 9 - Mattia Evangelisti, Roeland Hooijmans, Gustav Normelli





Feature Selection & Data preparation

Feature Selection

- Arbitrary selection of functions from RDkit to get features to set
- **Feature selections techniques**

Data Preparation

- Problem with **imbalanced data set**
- Applied techniques



Model Selection

- Consideration of **different standard models** without tuning
- **Training** of selected models
- **Tuning** of selected models
- Model selected: **ExtraTreesClassifier**

Results

- The model has been tested multiple times for more reliable results
- Final mean AUC = 83.89%

