iNeuron

Project Title	Prediction of LC50 value using Quantitative structure–activity relationship models (QSAR models)
Technologies	Machine Learning
Domain	Chemical
Project Difficulties level	Intermediate

Problem Statement:

Thousands of chemical substances for which no ecological toxicity data are available can benefit from QSAR modelling to help prioritise testing. One of the data set encompassing in vivo test data on fish for hundreds of chemical substances using the ECOTOX database of the US Environmental Protection Agency, you can check that dataset through this link: ECOTOX Database and additional data from ECHA. We can utilise this to develop QSAR models that could forecast two sorts of end points: acute LC50 (median lethal concentration) and points of departure akin to the NOEC (no observed effect concentration) for any period (the "LC50" and "NOEC" models, respectively). Study factors, such as species and exposure route, were incorporated as features in these models to allow for the simultaneous use of many data types. To maximise generalizability to other species, a novel way of substituting taxonomic categories for species dummy variables was introduced.

The goal here is to build an end-to-end automated Machine Learning model that predicts the LC50 value, the concentration of a compound that causes 50% lethality of fish in a test batch over a duration of 96 hours, using 6 given molecular descriptors.

Dataset:

You can get Dataset through this link:Dataset

Project Evaluation metrics:

Click here to enter text.

Code:

- You are supposed to write a code in a modular fashion
- Safe: It can be used without causing harm.
- Testable: It can be tested at the code level.
- Maintainable: It can be maintained, even as your codebase grows.
- Portable: It works the same in every environment (operating system)
- You have to maintain your code on GitHub.
- You have to keep your GitHub repo public so that anyone can check your code.
- Proper readme file you have to maintain for any project development.
- You should include basic workflow and execution of the entire project in the readme file on GitHub
- Follow the coding standards: https://www.python.org/dev/peps/pep-0008/

Database:

- You are supposed to use a given dataset for this project which is a Cassandra database.
- https://astra.dev/ineuron

Cloud:

- You can use any cloud platform for this entire solution hosting like AWS, Azure or GCP

API Details or User Interface:

- You have to expose your complete solution as an API or try to create a user interface for your model testing. Anything will be fine for us.

Logging:

- Logging is a must for every action performed by your code use the python logging library for this.

Ops Pipeline:

- If possible, you can try to use AI ops pipeline for project delivery Ex. DVC, MLflow , Sagemaker , Azure machine learning studio, Jenkins, Circle CI, Azure DevOps , TFX, Travis CI

Deployment:

Click here to enter text.

- You can host your model in the cloud platform, edge devices, or maybe local, but with a proper justification of your system design.

Solutions Design:

- You have to submit complete solution design strategies in HLD and LLD document

System Architecture:

- You have to submit a system architecture design in your wireframe document and architecture document.

Latency for model response:

- You have to measure the response time of your model for a particular input of a dataset.

Optimization of solutions:

- Try to optimize your solution on code level, architecture level and mention all of these things in your final submission.
- Mention your test cases for your project.

Click here to enter text.
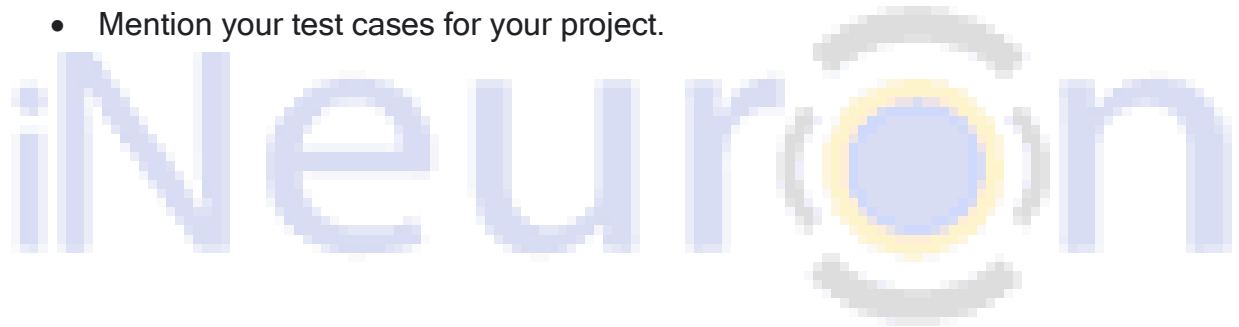
Submission requirements:

High-level Document:

You have to create a high-level document design for your project. You can reference the HLD form below the link.

Sample link:

HLD Document Link

Low-level document:

You have to create a Low-level document design for your project; you can refer to the LLD from the below link.

Sample link

LLD Document Link

Architecture: You have to create an Architecture document design for your project; you can refer to the Architecture from the below link.

Sample link

Architecture sample link

Wireframe: You have to create a Wireframe document design for your project; refer to the Wireframe from the below link.

Demo link

Wireframe Document Link

Project code:

You have to submit your code GitHub repo in your dashboard when the final submission of your project.

Demo link

Project code sample link :

Click here to enter text.

Detail project report:

You have to create a detailed project report and submit that document as per the given sample.

Demo link

DPR sample link

Project demo video:

You have to record a project demo video for at least 5 Minutes and submit that link as per the given demo.

Demo link

Project sample link :

The project LinkedIn a post:

You have to post your project detail on LinkedIn and submit that post link in your dashboard in your respective field.

Demo link

Linkedin post sample link :

Click here to enter text.