High Level Design(HLD)

Prediction of LC50 value
using (QSAR models)

written by-Aditya Rokade

contents:-

Abstract

Now we all are doing progress in various sectors/fields. Day by day finding a new technology in every fields ,scientist are finding a so many chemical which is used in various industry for different use cases.but some chemical are very harmful to nature .some are toxic in nature .
In toxilogy in major part is median lethal concentration(LC50),by checking the concentration of LC50 we can determine the toxicity of that perticular chemical by using the Quantitative structure–activity relationship models(QSAR models)

1 Introduction

1.1 why this high level design document?
The purpose of high level design document is to ass necessary detail to the current project description to represent a sutable model for coding this document is also indented to help detect contradictions prior to coding

The HLD will
- present all the design aspects and define them in detail
- Desribe the user interface being implemented
- Describe the hardware and software interfaces
- describe the performance requirements
- list and describe the non functional attributes like
  1. Security
  2. Reliability
  3. Maintainability
  4. Portablity
  5. Reusability

1.2 Scope
The HLD documentation present the structure of the system.such as the database architecture,application architecture. The HLD uses the nontechnical to midily-technical words.

1.3 definations

Terms	Description
QSAR models	Quantitative structure–activity relationship models
database	Collection of all the data
IDE	Integrated Development Envoirement
AWS	amazon web services

2. General Description

2.1 Product perspective

Quantitative structure–activity relationship model is a machine learning model which will help us to calculate the concentration of LC50 in various chemical.

2.2 Problem statement

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.

2.3 technical requirements

To calculate the concentration of LC50 we need the values of 6molecular descriptors
1) CIC0
2) SM1_Dz(Z)
3) GATS1i
4) NdsCH
5) NdssC
6) MLOGP

to calculate these values we need the some sensors/technical reading

2.4 Data requirements

   data requirement is totally depend on our problem statements

   1.to train the ml module we need the sufficient data in which 6 molucular descriptors and
   their respective LC50 value so we can train our module based on that data
      besause the ml module is worked on relantionship between data

   2 . data file in numeric format besauce this is regresion problem

   3. some data to test the model

2.5 Tools Used

python programming langauge and frameworks such as pandas numpy,scikit-learn,flask
   AWS

1.VScode used as IDE

2. For data visualization matplotlib,plotly,seaborn are used

3.AWS is used to deploy the whole project

4.cassandra &sqlite databases are used for data operation

5.front end development is done using HTML/CSS/Javascript

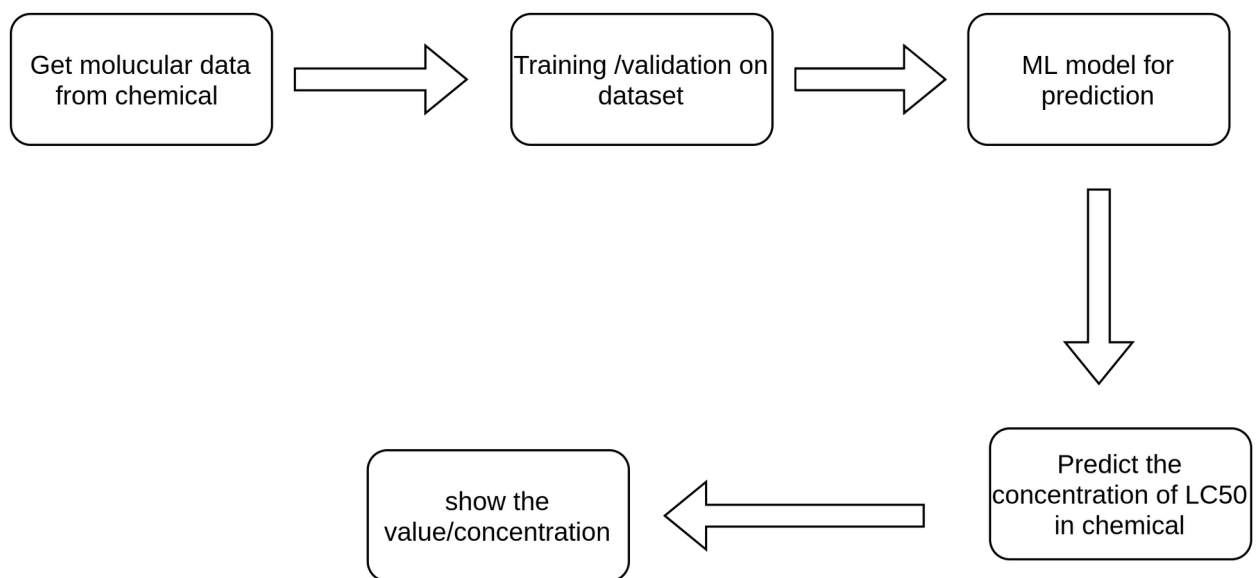6.python Django are used as backend development

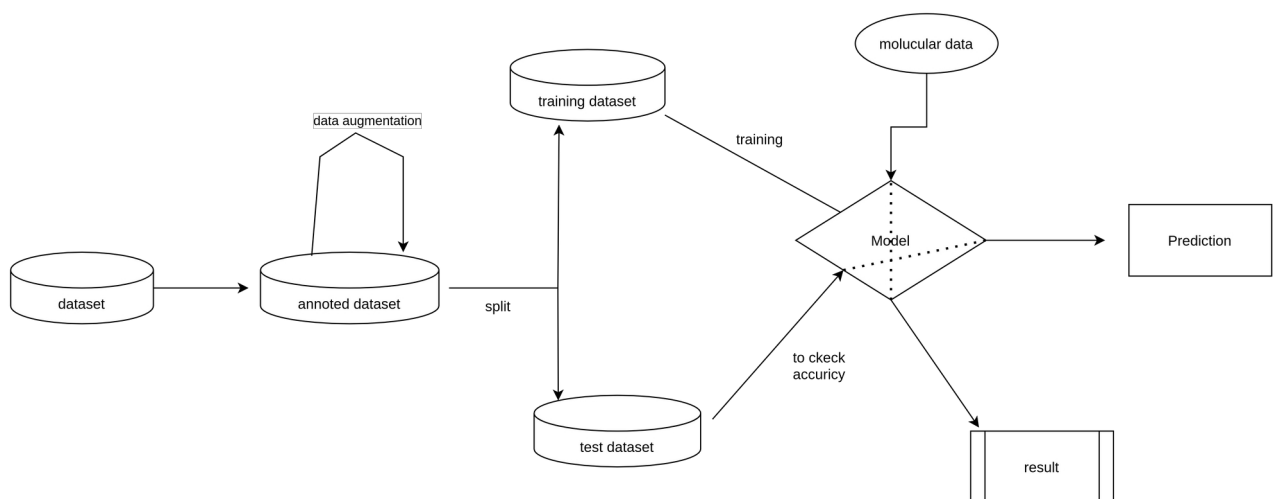7. Github is used as version contraol system

3. design details

3.1 Process flow
To calculate the concentration to LC50 ,we will use machine learning modeule
process flow is shown in below diagram

Proposed methodology

model training and Evaluation

3.2 Event log
  The system should log every event so that the user will know what process is running internaly

1. system should be able to log each and every system flow.
2.Developer can choose looging method.we can choose database logging /file logging

3.3 Error Handling
The system should be handled the error which is occoured during running .
We can use try and except statements to handled the error.

3.4 Performence

The our solution is used to calculate the concentration of LC50 ,to check the toxicity of chemical based on the LC50 value we can determine the Toxicity of perticular chemical

3.5 Reusability
The code is written and the components used should have the ability to be reused with no problems

4 Deployment