FUNCTIONAL and TECHNICAL REQUIREMENTS DOCUMENT

Title:

Drug Discovery

Develop By: Kashaf Naz

TABLE OF CONTENTS

1. General information

We have to download and pre-process biological activity data from the ChEMBL database that you can use to perform Computational Drug Discovery. The dataset is comprised of compounds (molecules) that have been biologically tested for their activity towards target organism/protein of interest

1.1. Purpose

The purpose of this project is to build a relationship between structural and biological activity of a molecule. exploratory data analysis will help to find differences of the active and inactive sets of compounds. We will build a regression model for predicting the pIC50 values.

1.2. Scope

This drug designing project will build a Bioinformatics tool that will allow users the ability to predict whether a compound of interest has favorable biological activity against the target protein or not.

2. Acronyms and/or definitions

- Python (ID Pycharm)
- PADEL-Descriptor software (molecular descriptors)
- Sciket learn (ML model)
- Streamlit (for deployment)

3. Project Overview

use the SMILES notation (representing the unique chemical structure of compounds) to compute molecular descriptors. The descriptors that we will be computing are the Lipinski's descriptors (molecular weight, LogP, number of hydrogen bond donors and number of hydrogen bond acceptors). we will perform exploratory data analysis by making simple box plots and scatter plots to discern differences of the active and inactive sets of compounds.
prepare the dataset (X and Y dataframes) that will be used in the next part for Model Building.

Functional Requirements and user Impact

- Data Collection and Preprocessing
- Exploratory Data Analysis
- Dataset Preparation
- Model Building
- Compare Models
- Deploy Model as Web App) | Streamlit

Data Resource:

In this project we will use the data from different databases like ChEMBL