阿布都赛米·阿布都外力 学号：2020182631 考试号：180150129

**实验名称：2D-QSAR**

**实验目的：**

1. 掌握数据集中训练集与测试集的拆分方法。

2. 掌握分子描述符的选择和建模方法的确定。

3. 掌握模型结果的分析。

4. 掌握未知化合物活性预测。

**实验原理：**

使用Discovery Studio 软件进行2D-QSAR模型的构建、外部数据集检测、未知活性化合物预测。

本实验所用软件环境：

DS Version：19.1.0.18287

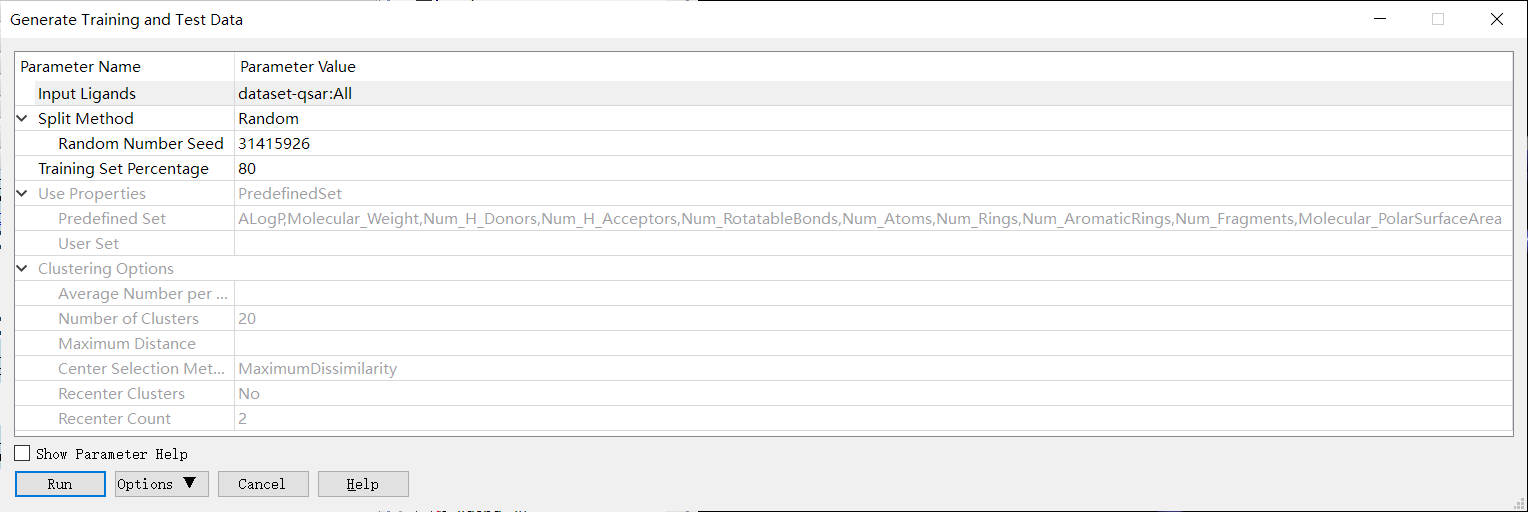
PP Version：19.1.0.1963

DS Client Version：19.1.0.18287

OS Distribution：Windows

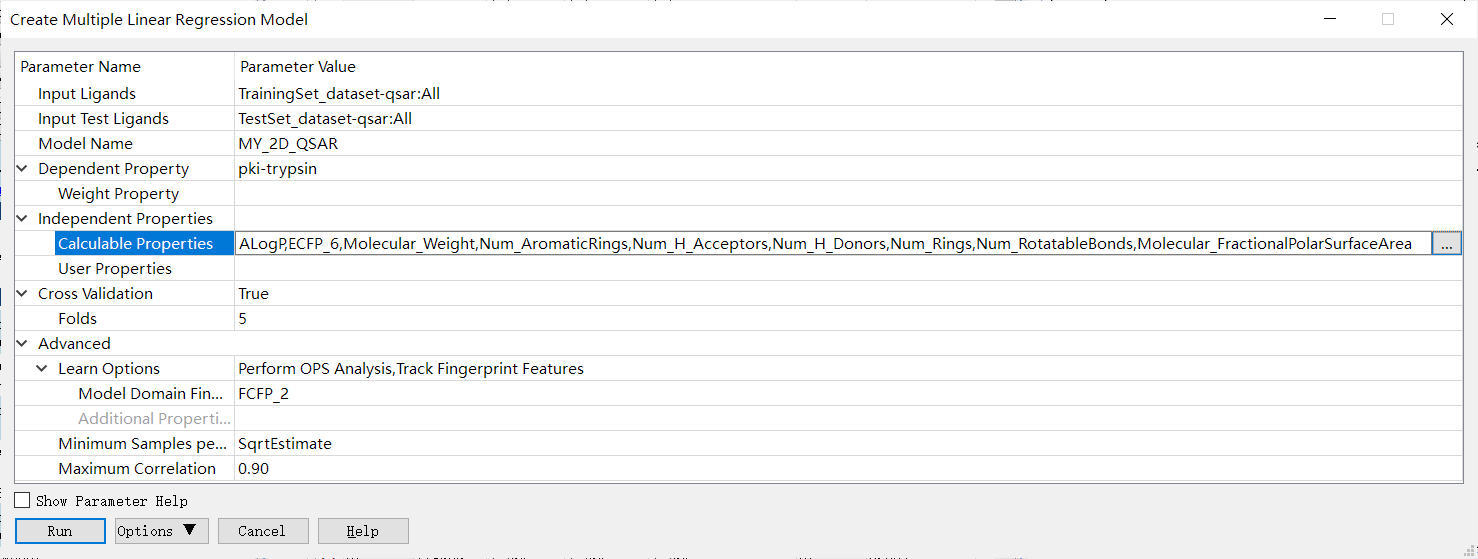
OS Version：10.0.19044

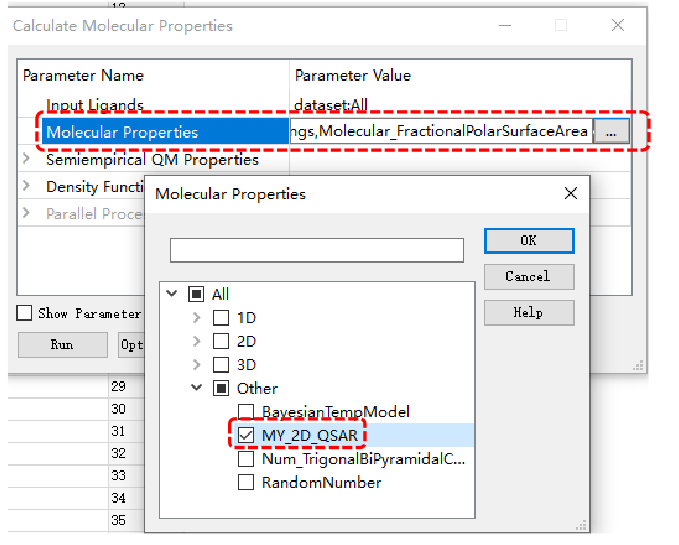
**实验步骤：**

1. 已知活性数据收集：本实验使用指导老师提供的dataset-qsar.sdf数据集。

2. 数据集准备（训练集与测试集拆分等）：点击Discovery Studio软件上的Small Molecules🡪Create QSAR Model🡪Generate Training and Test Data进行训练集与测试集拆分。设置参数如下：

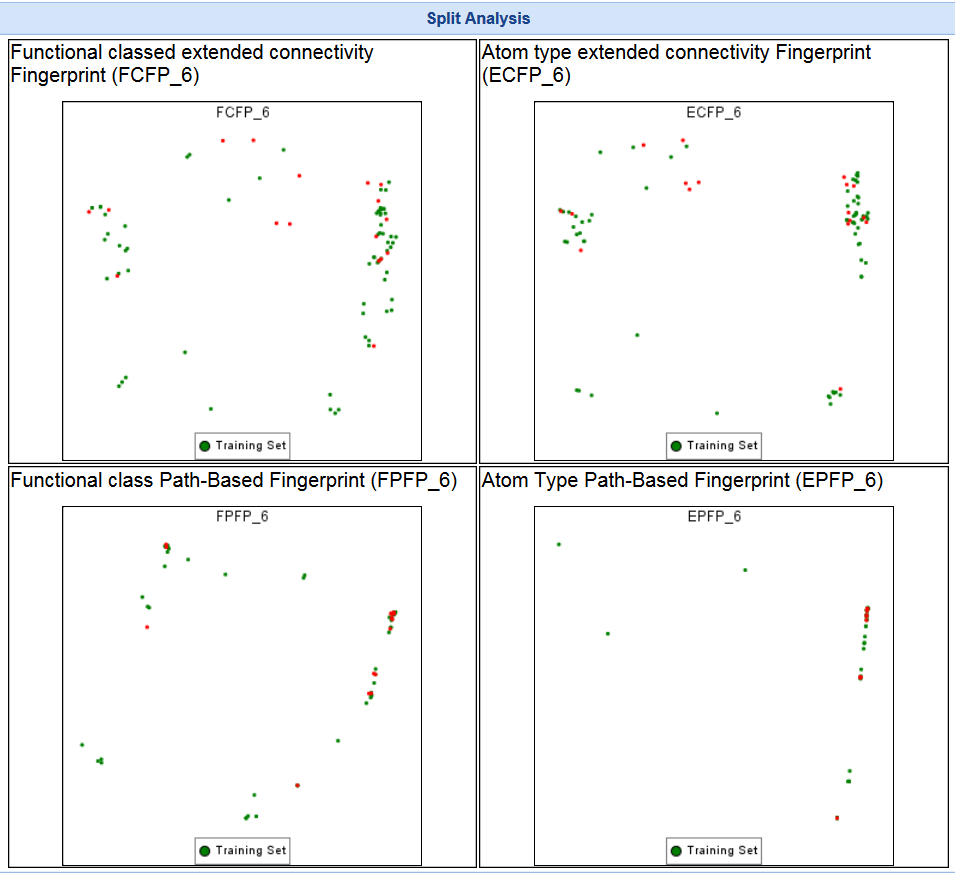
3. 分子描述属性计算（传统分子描述符、分子指纹等）：Discovery Studio会在模型的构建中自动计算。在构建模型时，只需在Calculable Properties中挑选要计算的描述符。

 4. 模型的参数设置与构建：点击Discovery Studio软件上的Small Molecules🡪 Create QSAR Model 🡪 Create Multiple Linear Regression Model进行多元线性回归模型的构建。设置参数如下：

 5. 未知活性化合物预测：点击Discovery Studio软件上的Small Molecules🡪Calculate Molecular Properties🡪 Calculate Molecular Properties🡪进行未知活性化合物预测。设置参数如下：

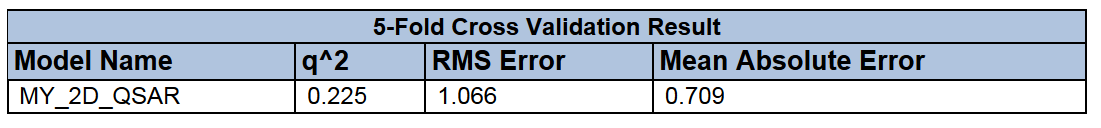
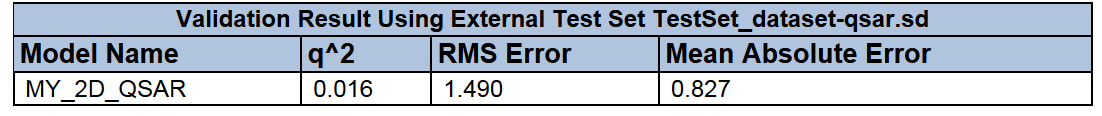
**实验结果：**

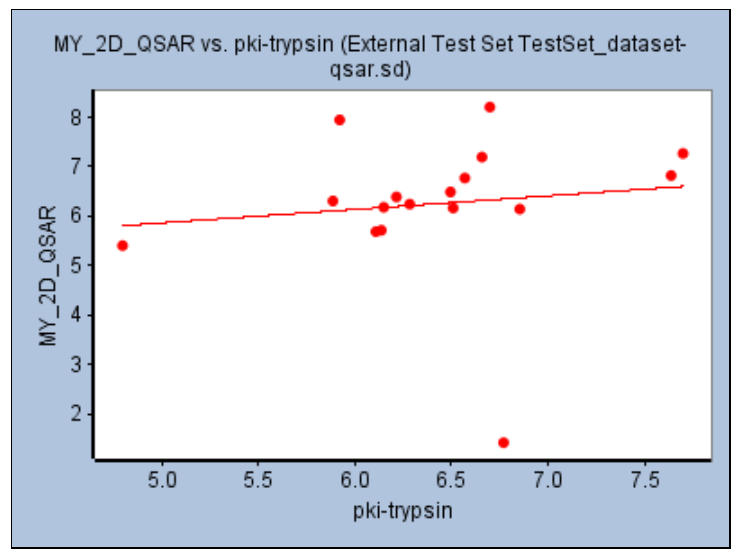
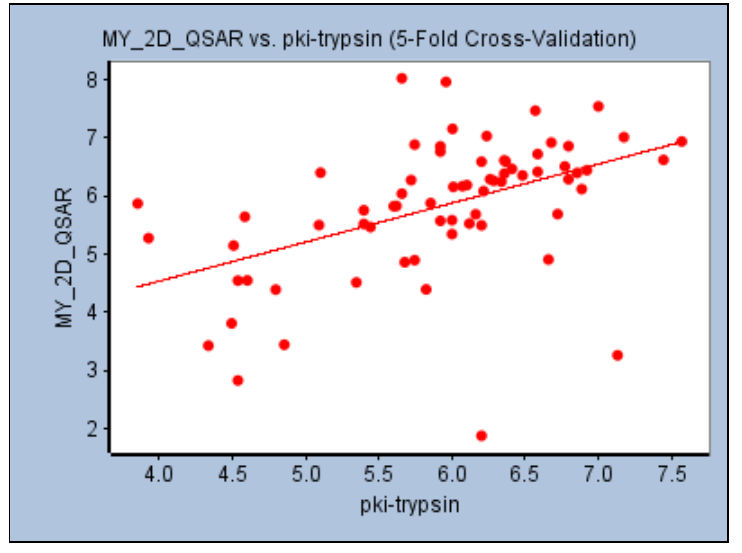
1. 数据集准备的结果：

 Status：Success Elapsed Time: 00:00:08Summary：Data split: 69 in training set, 17 in test set.

2. 模型的参数设置与构建的结果：

Status：Success Elapsed Time: 00:00:07

 Summary: Multiple Linear Regression Model MY\_2D\_QSAR has been created.



Y=0.943x+0.340 R-Square=0.943

Y=0.273x+4.494 R-Square=0.016

**讨论：**

模型的 q^2值在内部验证中，在外部验证中均小于0.5。所以此模型不可靠。