

The background features four stylized molecular structures in the corners, composed of green and dark green spheres connected by rods. The top-left structure has a dark green sphere and a green sphere. The top-right structure has a white sphere, a dark green sphere, and a green sphere. The bottom-left structure has a green sphere, a dark green sphere, and a green sphere. The bottom-right structure has a green sphere and a dark green sphere.

QSAR based Metabolism and Hydration Free Energy Prediction

Team Panic Storm

Shaik Afroz
Parth B Sanghavi

TABLE OF CONTENTS

01

**Classification
Model**

02

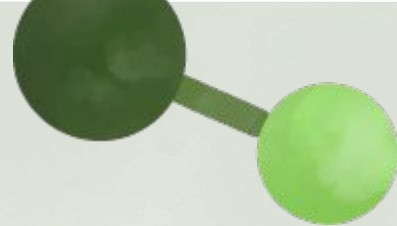
**Regression
Model**

03

**Molecule
Clustering**

04

Prediction



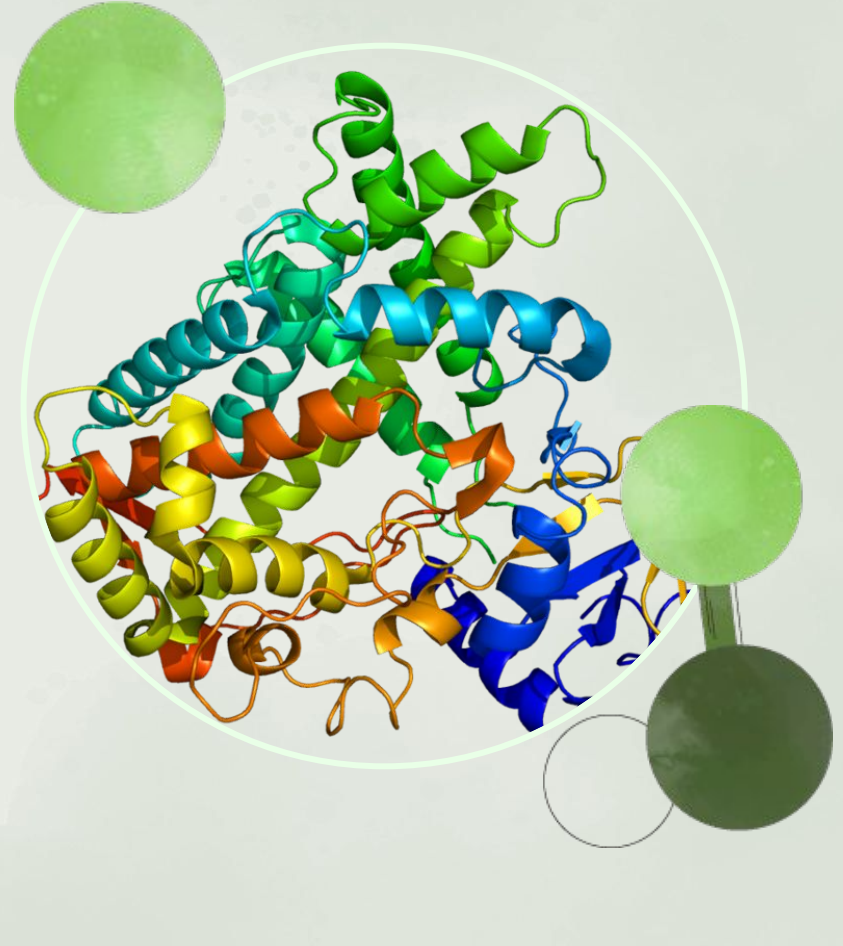
01



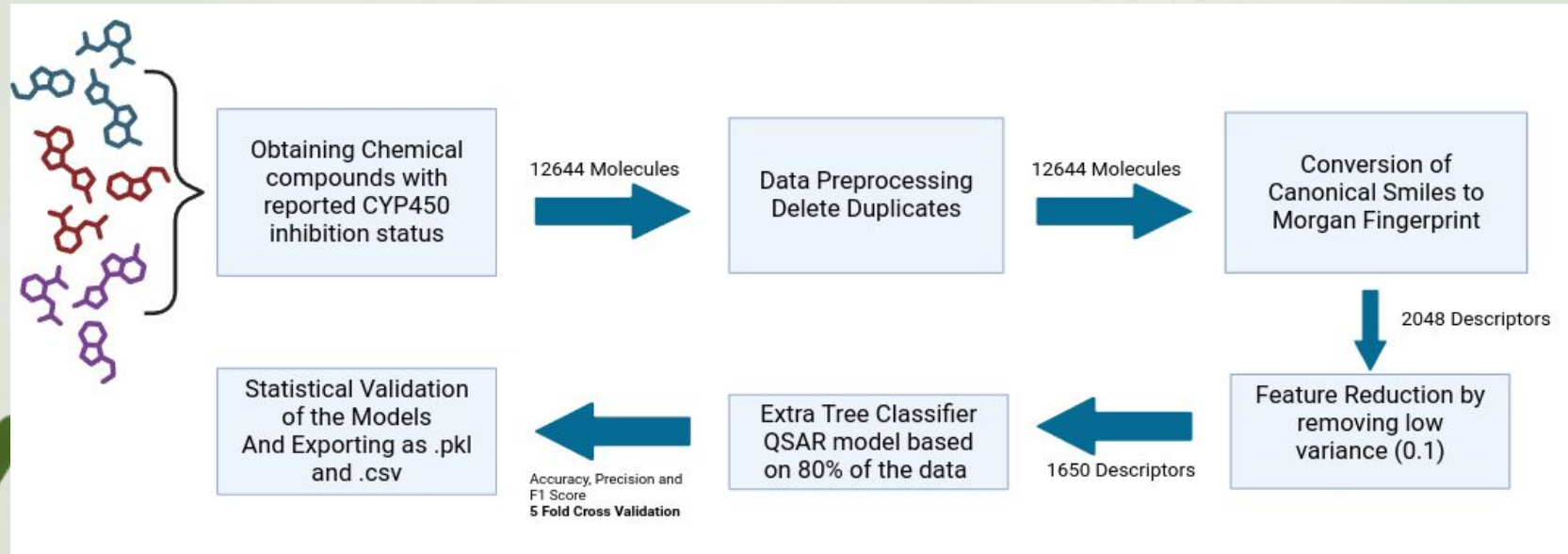
Classification Model- P450 inhibitor

CYP P450 2C19

- **CYP P450 2C19** is a key enzyme involved in the **metabolism** of several **drugs**, affecting their efficacy and safety.
- Understanding its inhibition potential is crucial for predicting and **avoiding harmful drug-drug interactions**.
- **Inhibition** of this enzyme can lead to **altered drug metabolism**, impacting **therapeutic outcomes** and **patient safety**.



Methodology



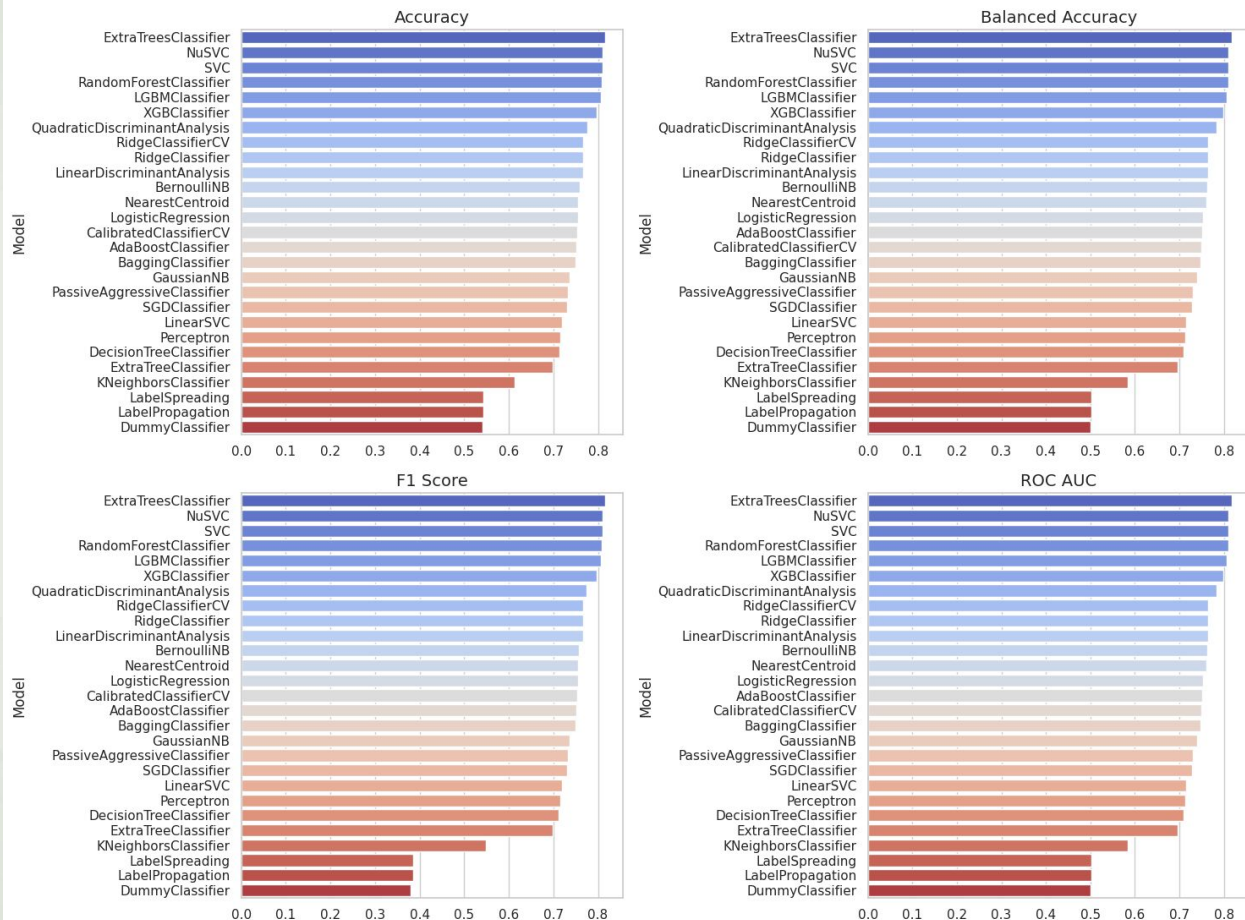
Results

The model that performed the best was Extra Tree Classifier

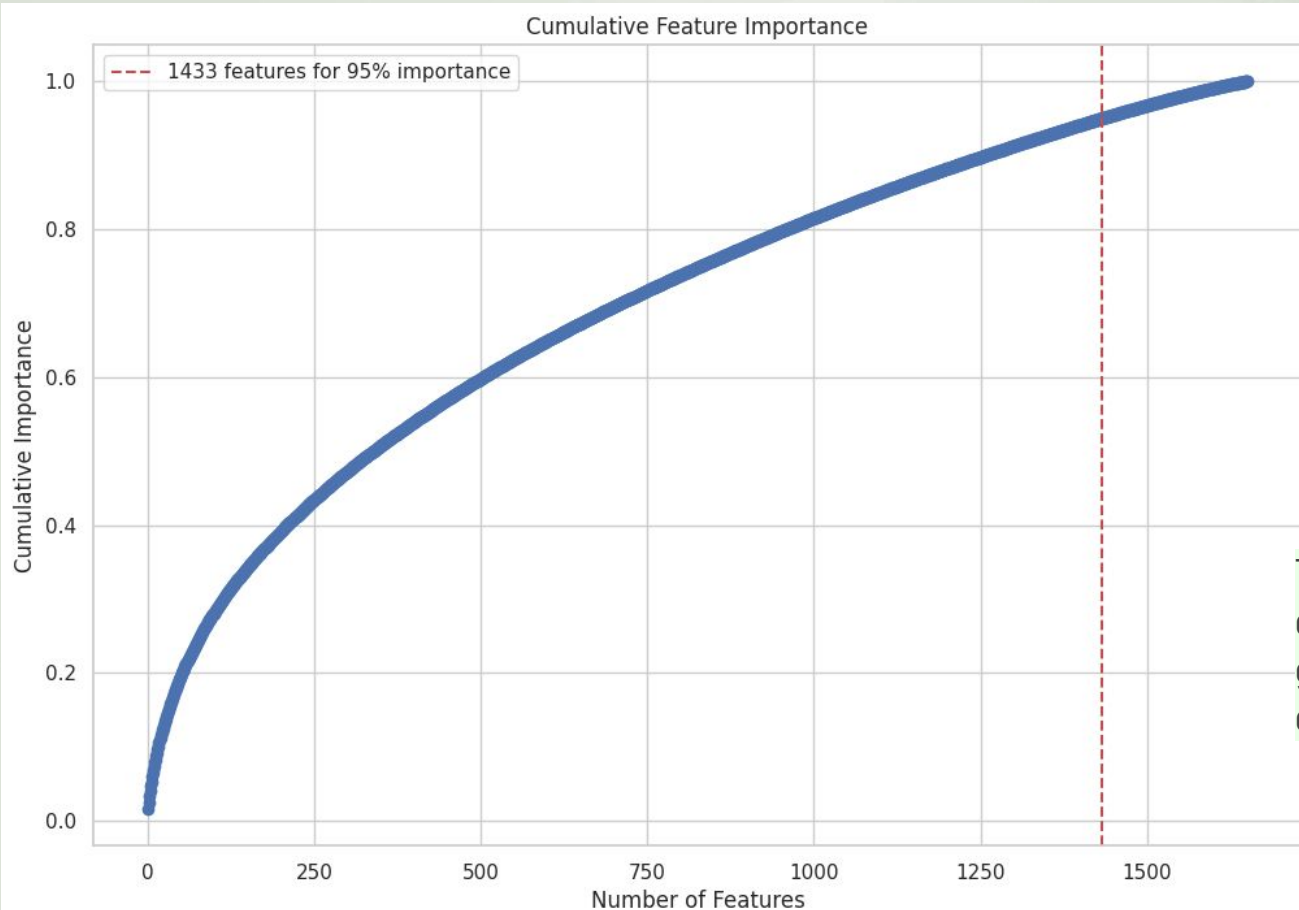
The performance metrics are as follows

Accuracy	0.81
ROC_AUC	0.82
F1 Score	0.82
5- Fold CV (Accuracy)	0.69
Std. Dev. of accuracy	0.04

Performance Metrics Comparison



Results

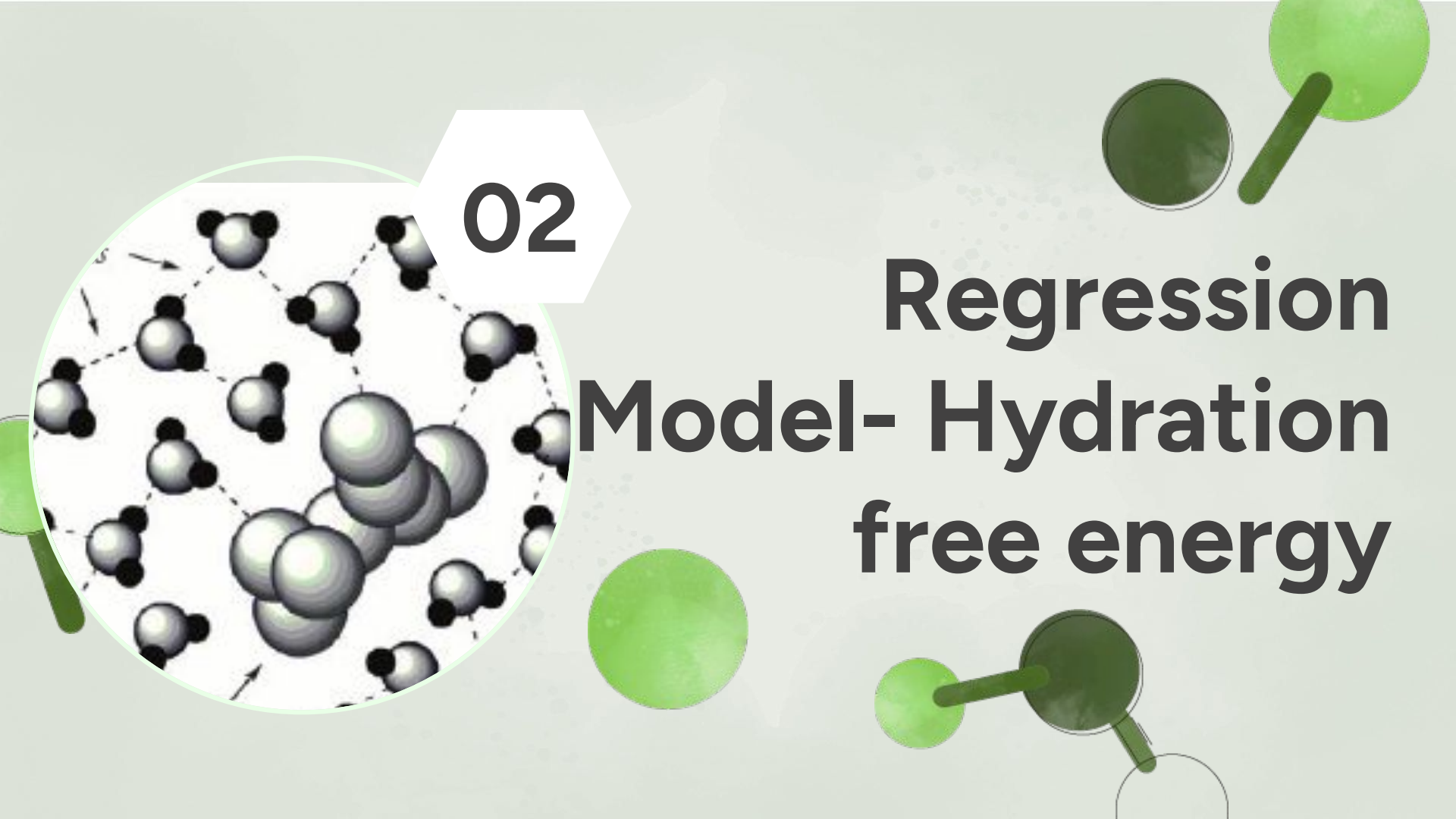


- True Positives (TP): 968
- False Positives (FP): 278
- True Negatives (TN): 1092
- False Negatives (FN): 195

These results successfully conclude that the model is good and can be used for classification tasks.

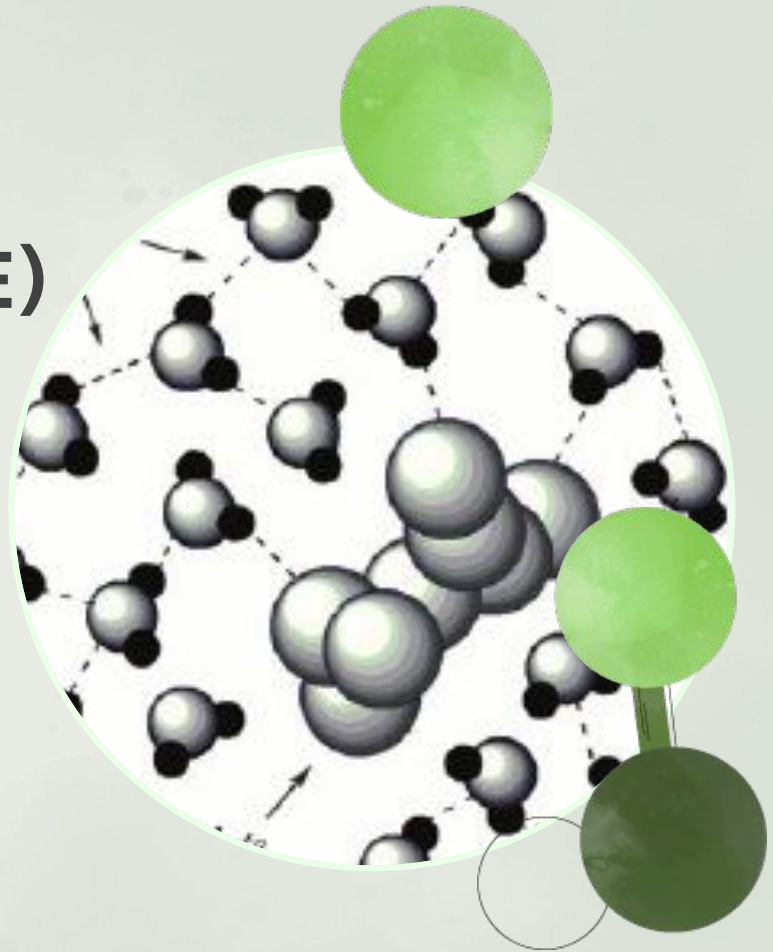
02

Regression Model- Hydration free energy

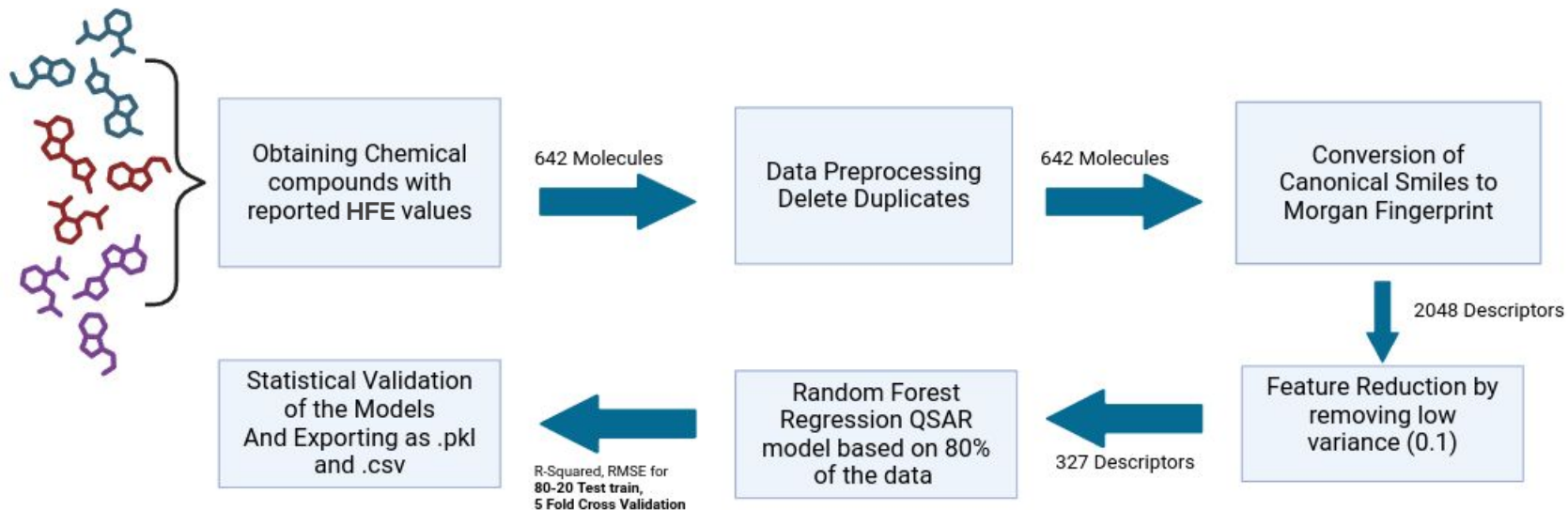


Hydration Free Energy (HFE)

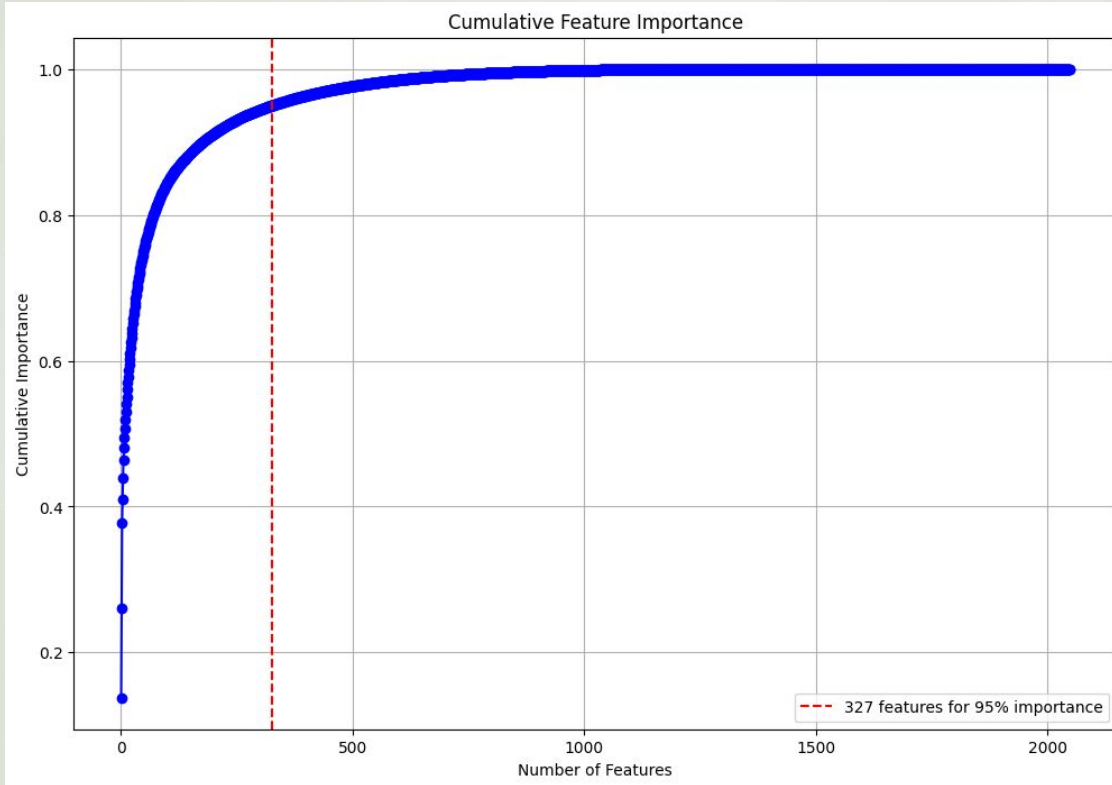
- Understanding **HFE** is crucial for **predicting** the **stability of molecules** in cellular environments.
- HFE influences various **biological processes**, including protein-ligand binding and **drug absorption**.
- **Accurate prediction** of HFE aids in **rational drug design** by optimizing molecular properties for **enhanced solubility** and **bioavailability**.



Methodology



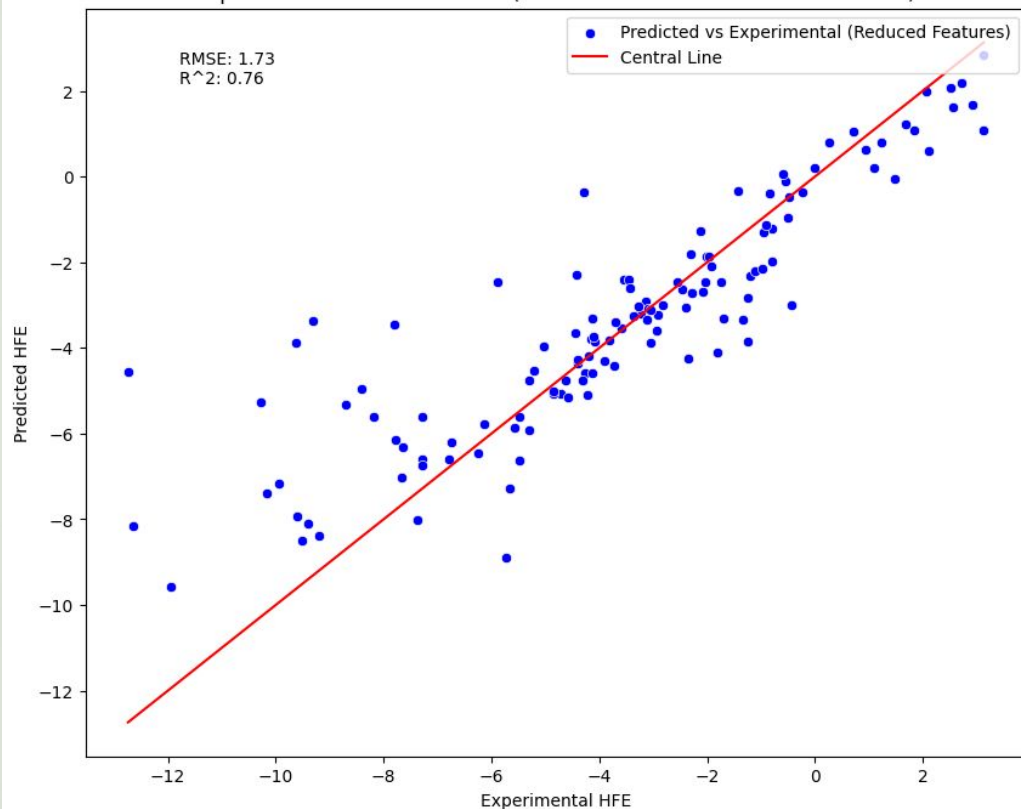
Results



Extracting features that explained 95% of variance revealed 327 features, indicating the model can handle large datasets and has high interpretability.

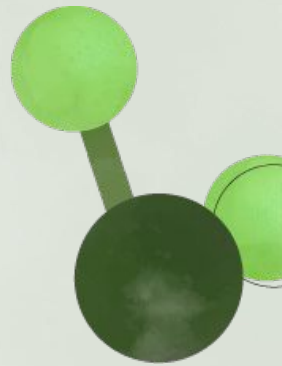
Results

Experimental vs Predicted HFE (Random Forest with Reduced Features)

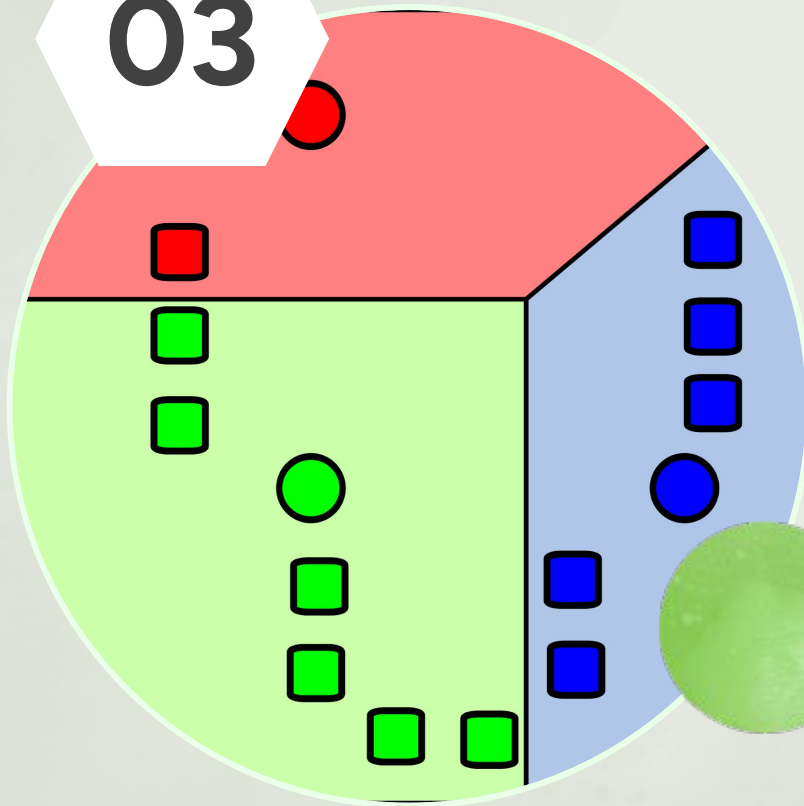


R-Squared of 0.76 can be improved but this can be utilized to perform high throughput virtual screening of large datasets like moose.

RMSE of 1.73 indicates that the predictions made are not deviating by a huge degree.



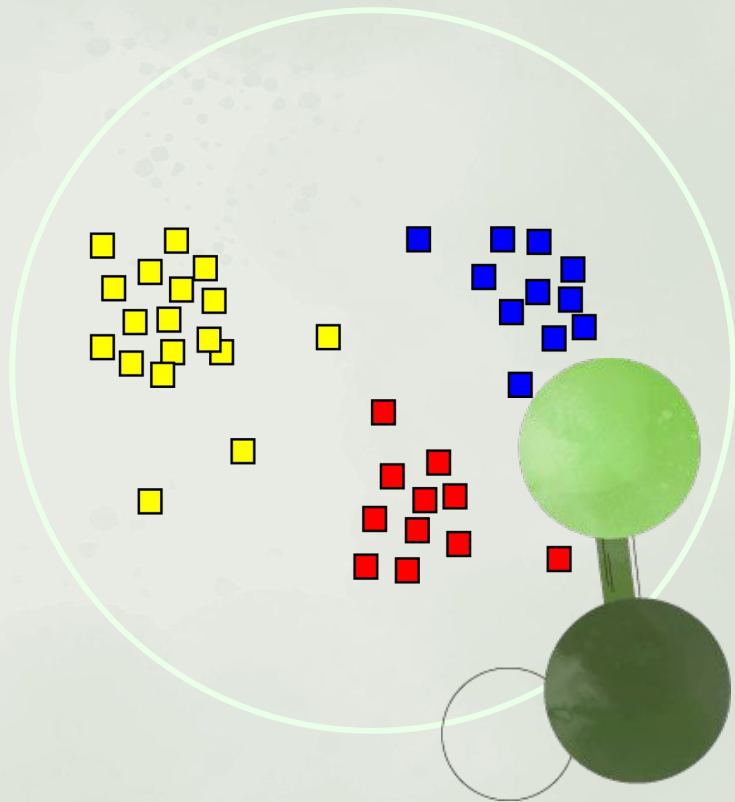
03



Clustering - Moses Dataset

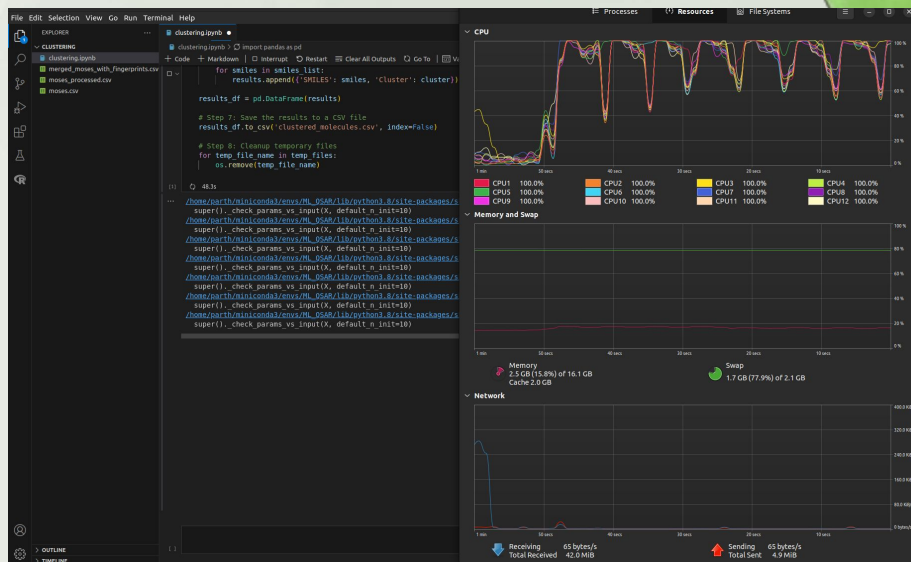
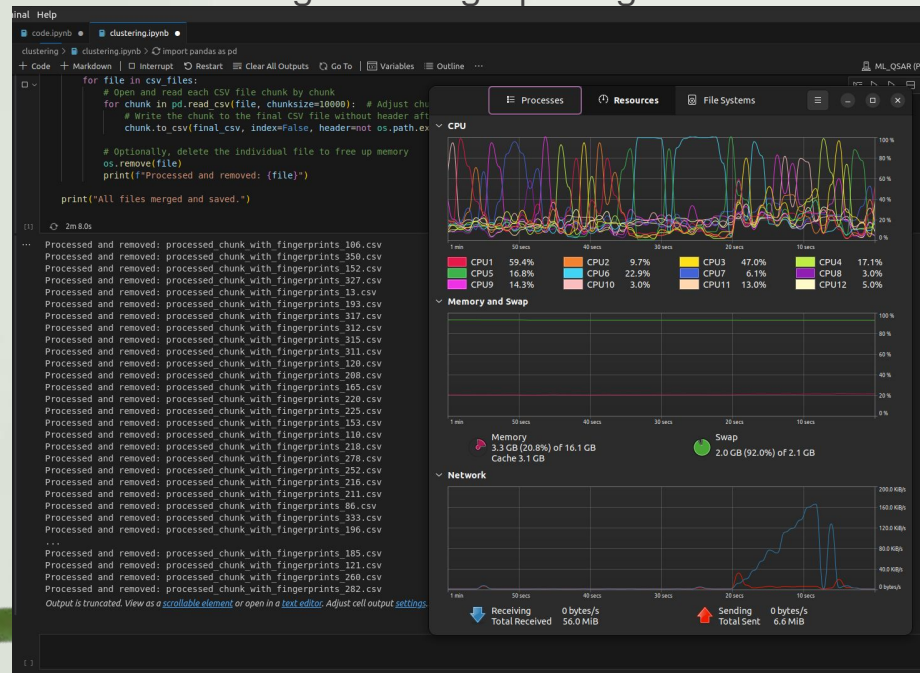
Memory efficient Clustering

- Random Sampling: Reduces memory footprint by processing a subset of the dataset at a time.
- Incremental PCA (IPCA): Enables batch processing of large datasets, reducing memory usage.
- Sparse Matrix Representation: Memory-efficient handling of large datasets with many zero values.
- Temporary Files and Chunk Processing: Minimizes memory usage by saving intermediate results to disk and processing data in chunks.
- Swap Memory: Provides additional virtual memory when physical RAM is full, enabling processing of large datasets with slower speed.



Resources used

Resource usage on Fingerprint generation

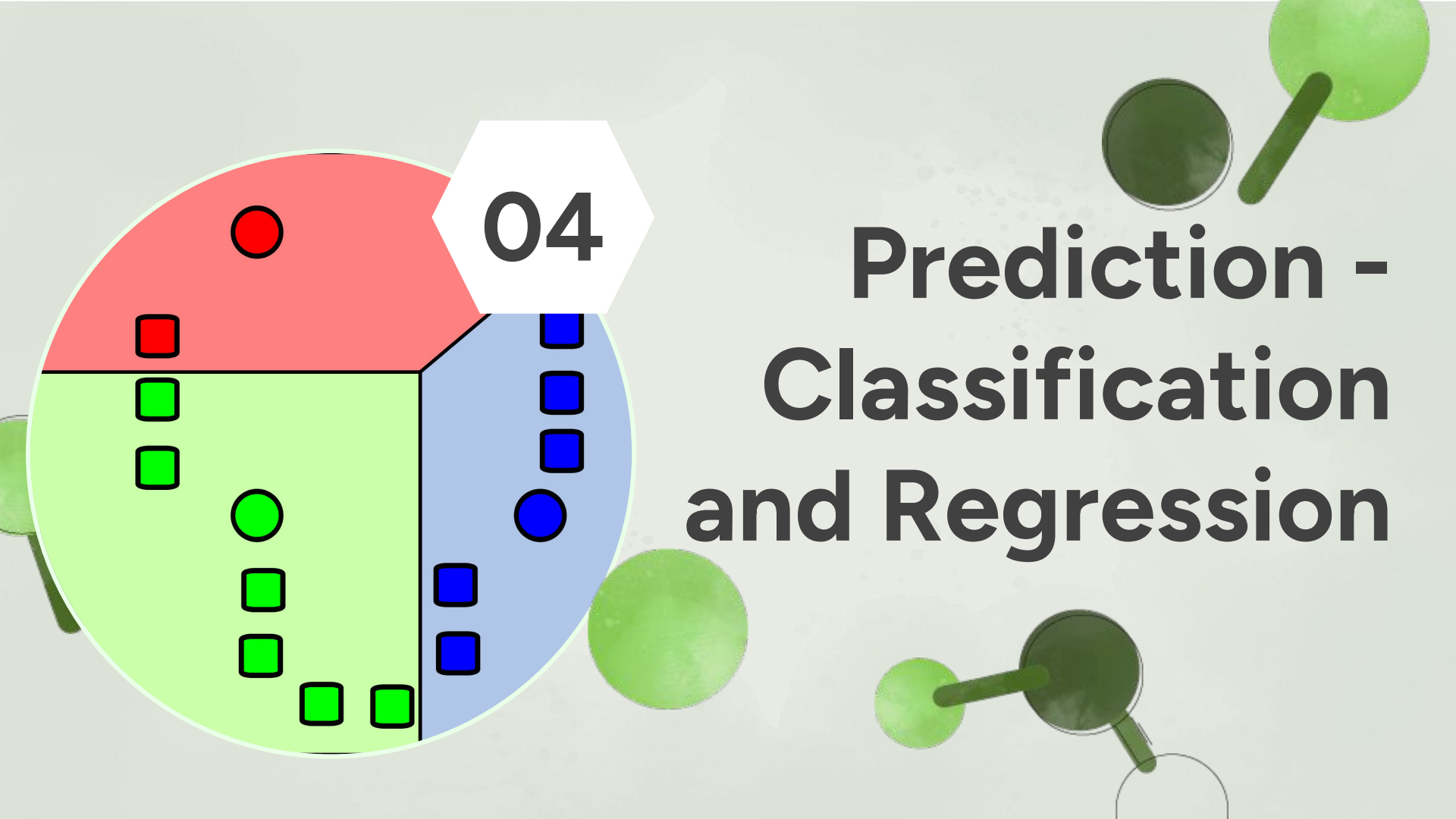


Resource usage on Clustering

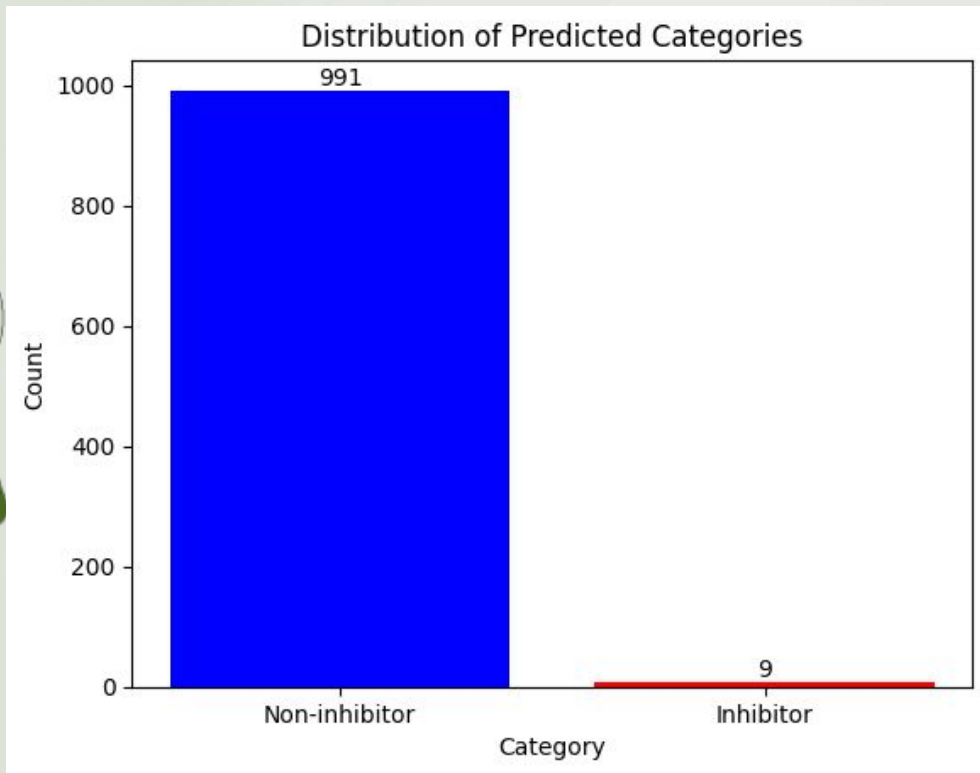
Low ram usage was maintained in both scenarios

04

Prediction - Classification and Regression



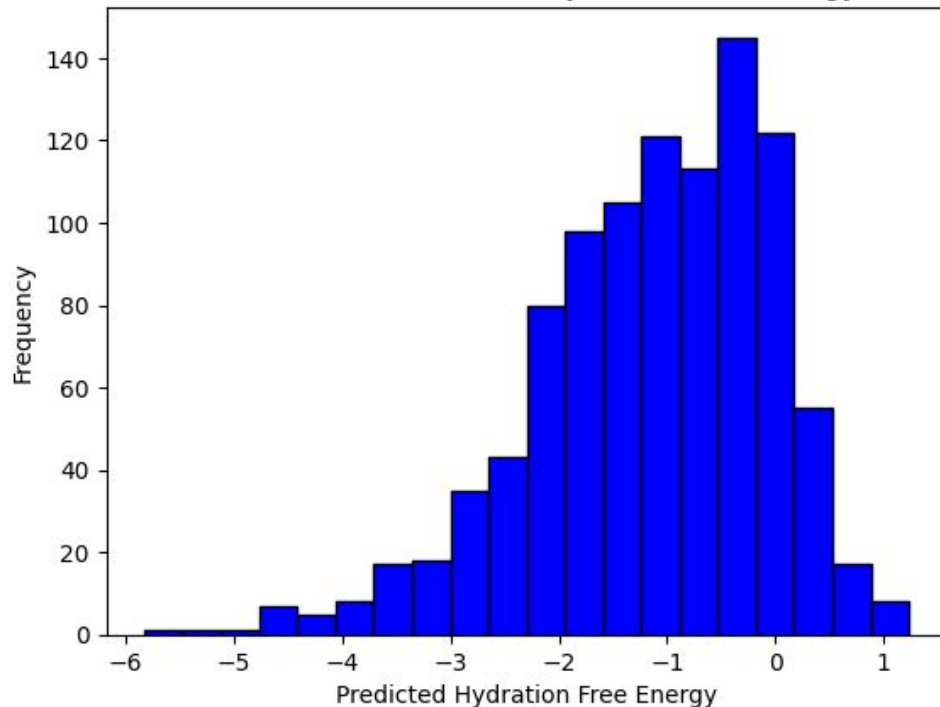
Classification Model



- Most were non inhibitors
- Most molecules could be metabolized easily

Regression Model

Distribution of Predicted Hydration Free Energy



- Most molecules had low solubility
- Emphasises the role of excipients in drug industry