Drug Target Identification

Sponsor

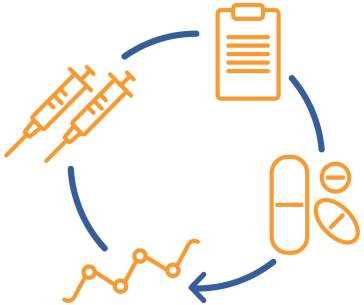
Tomas Kawalec (Executive Sponsor)
Sanjay Jaiswal (Internal MD Project Sponsor)

Team Members

Anurag Gunti
Ayush Dadhich
Kausar Perveen
Yin Yang

Problem Statement

Drug discovery and development pipeline is resource-intensive and time-consuming, which make them a major obstacle for rapid drug development.

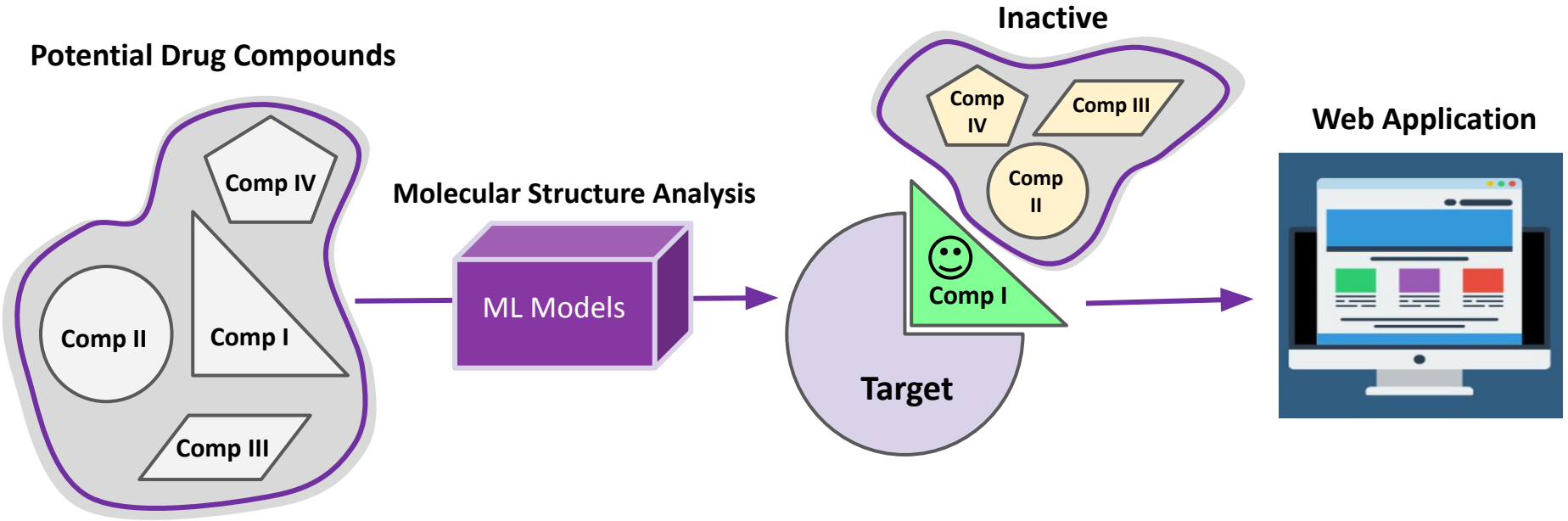


Current challenge is to develop discovery pipelines that can identify promising drug compounds early.

A reliable method that can identify the potential drug compound with respect to its successful clinical translatability is needed.

Objectives

- Create a prototype for identifying potential drug compounds.
- How we did it?



Pipeline

Literature Review

Targets Search

Google Scholar

PubMed.gov

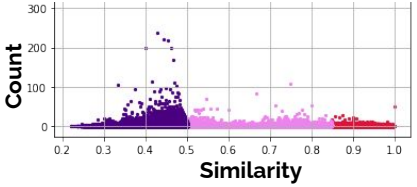
nature

Data Preparation

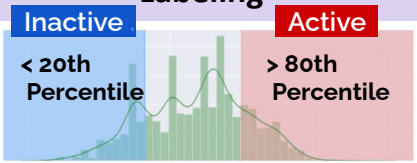
Data Acquisition

ChEMBL

Data Cleaning



Labeling



Data

PaDEL-Descriptor

Modeling

Binary Classifiers

Multi-labeled Classifiers

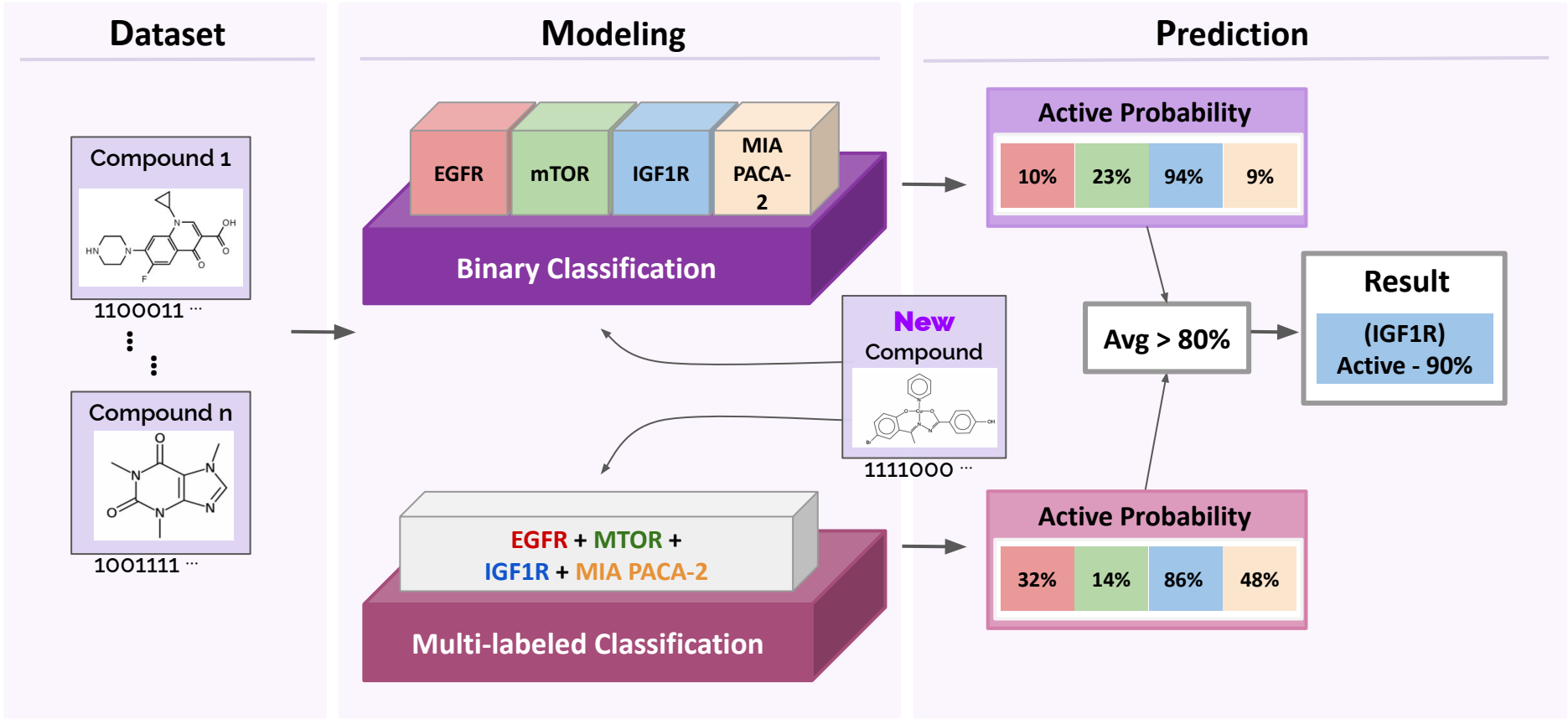
Deployment

Web Application

django

aws

Modeling


Dataset
Compound 1
1100011 ···
Compound n
1001111 ···
Modeling
EGFR
mTOR
IGF1R
MIA PACA-2
Binary Classification
New Compound
1111000 ···
EGFR + MTOR +
IGF1R + MIA PACA-2
Multi-labeled Classification
Prediction
Active Probability
10%
23%
94%
9%
Avg > 80%
Result
(IGF1R)
Active - 90%
Active Probability
32%
14%
86%
48%

Web Application Demo

Appendix

Preprocessing

Cleaning

- Filtered data by
  - standard_type = ‘IC50’
  - standard_unit = [‘nM’, ’uM’, ’pM’]
  - target_organism = ‘Homo Sapiens’
- Dropped data points with missing standard_value or canonical_smiles.
- Converted the standard values to have the same standard unit (nM).

Labeling

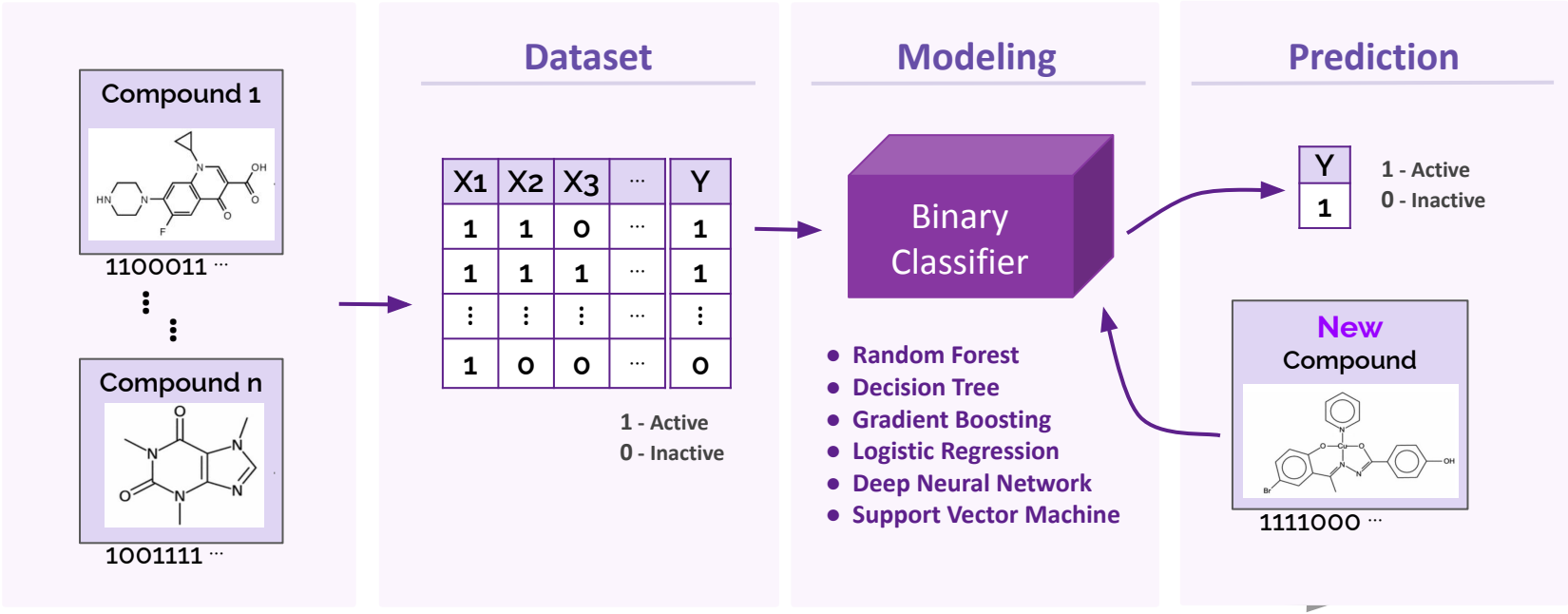
- Standardized IC50 to pIC50 to avoid skewed distribution.
- Checked the standard values at different percentiles to determine the cutoff values for labeling.

Transformation

- Used PaDEL descriptor to transform the SMILE notations into rule-based Pubchem Fingerprints.

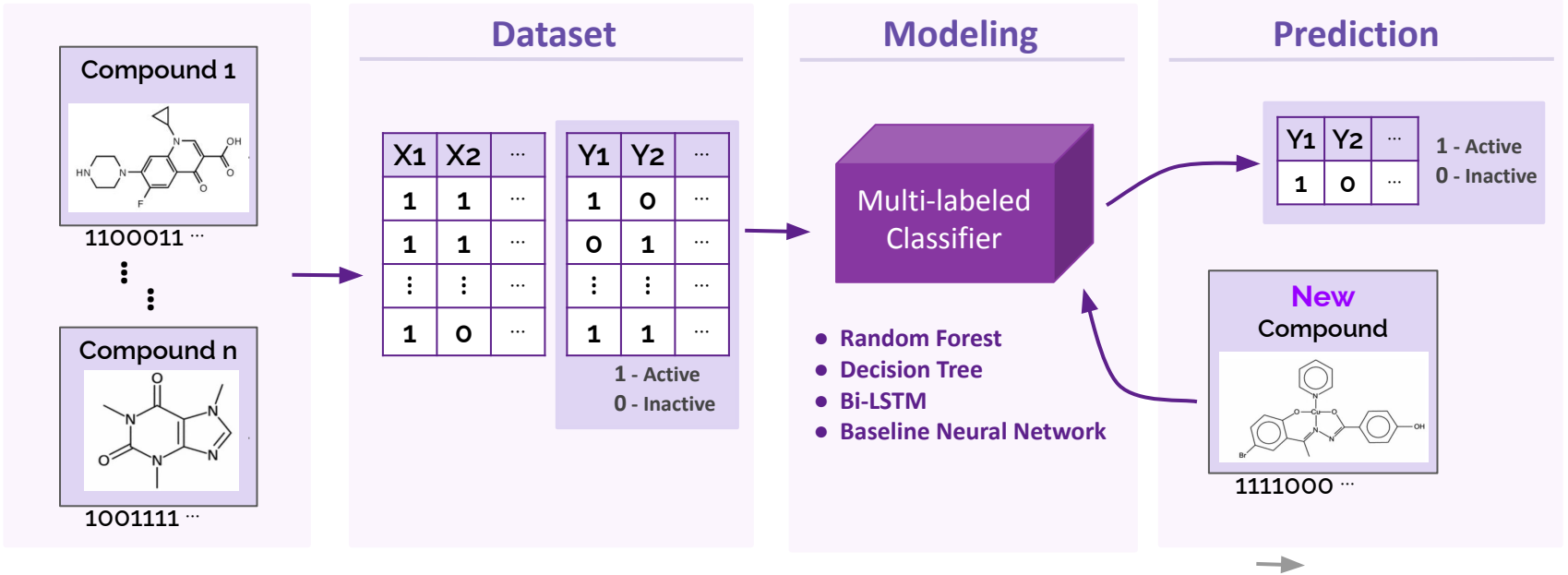
Modeling - Binary Classifiers

- Build models on each selected target to identify the bioactivity of candidate compounds.

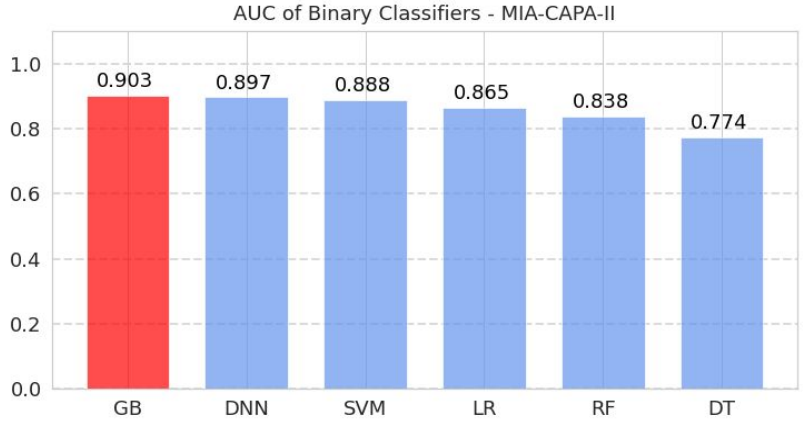


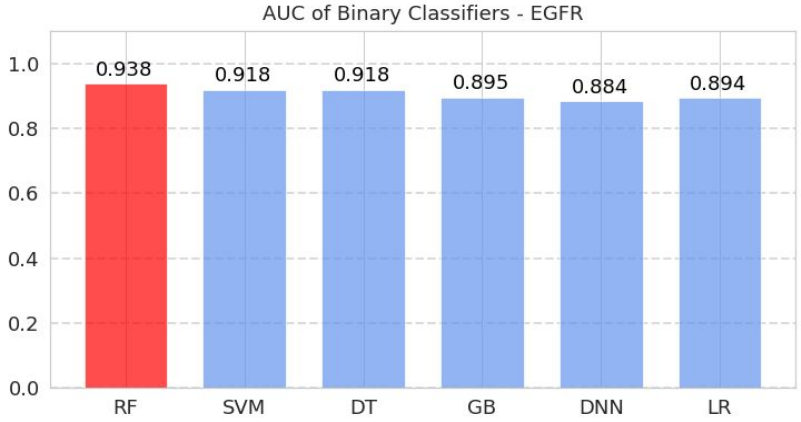
Modeling - Multi-labeled Classifiers

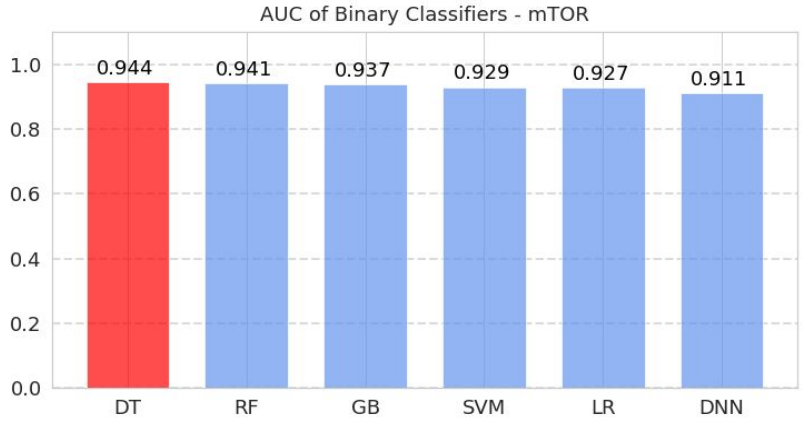
- Build models on all selected target to identify the bioactivity of candidate compounds.

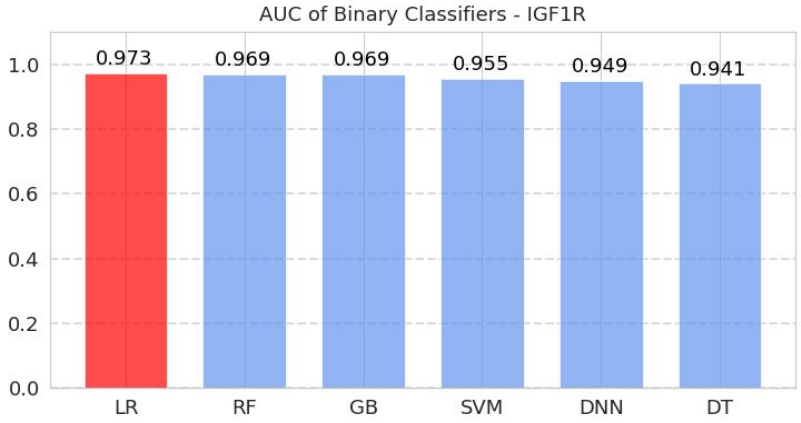


Modeling - Results









Modeling - Results