Final Model Choice

- Symmetric Features
- Base model selection:
  - **Dense**: Deepnose (96) + selected Mordred (70)
  - **Sparse**: Morgan50+ Leffingwell98 (might choose Morgan + another 50 dim fingerprint)
- Slightly undertrained base models: reduce # of trees to create variability for the meta model to explore —> Recovers the rare meta model that performs better than bases
  - Random Forest/ XGboost as the meta model

	10 fold CV		Leaderboard	
Symmetric	R	RMSE	R	RMSE
dense	0.6549 ± 0.0083	0.1198	0.6334	0.1256
sparse	0.6387 ± 0.0043	0.1209	0.7126	0.1179
meta	0.6204	0.124	0.7234	0.1126

non-	10 fold CV		Leaderboard	
Symmetric	R	RMSE	R	RMSE
dense	0.6513 ± 0.0074	0.1212	0.724	0.12
sparse	0.6344 ± 0.0089	0.121	0.682	0.12
meta	0.6364	0.1207	0.7314	0.1135

Performance comparison, symmetric v.s. non-symmetric

Symmetric

Non-symmetric

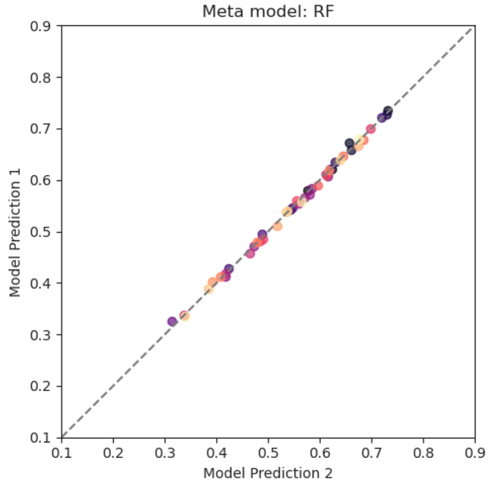
Leaderboard

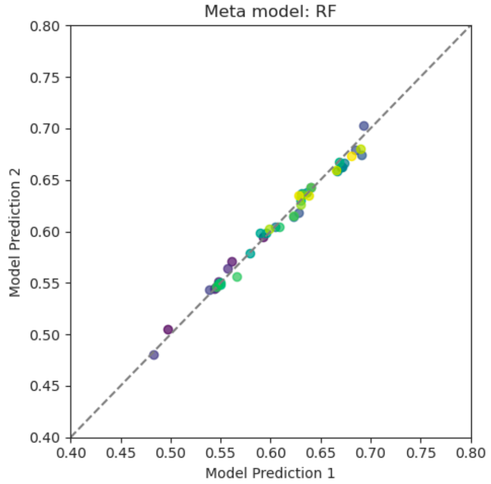
Test

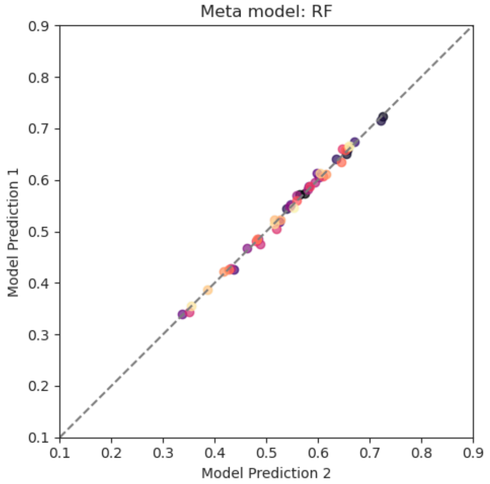
Leaderboard

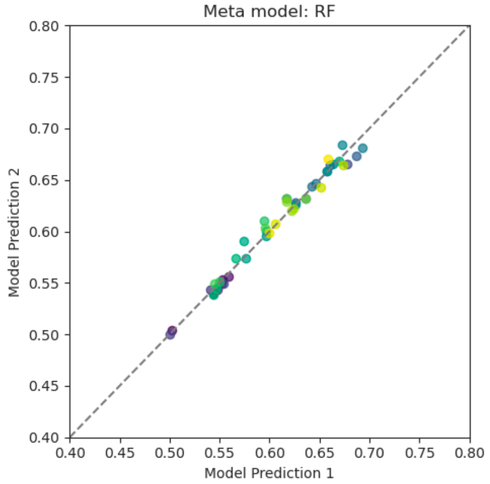
Test

2 set of seeds









Performance