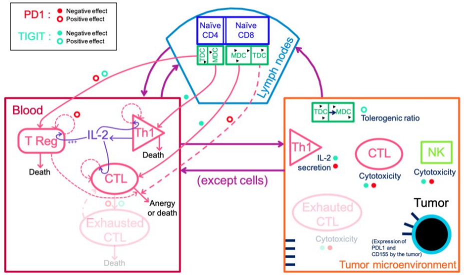
Key Objectives: ML Analysis

- Using the 144 QSP Parameters, is it possible to build a Machine Learning Model to predict the Categorical response(Classification)?
- For the top models what are the most influential parameters for each Task?

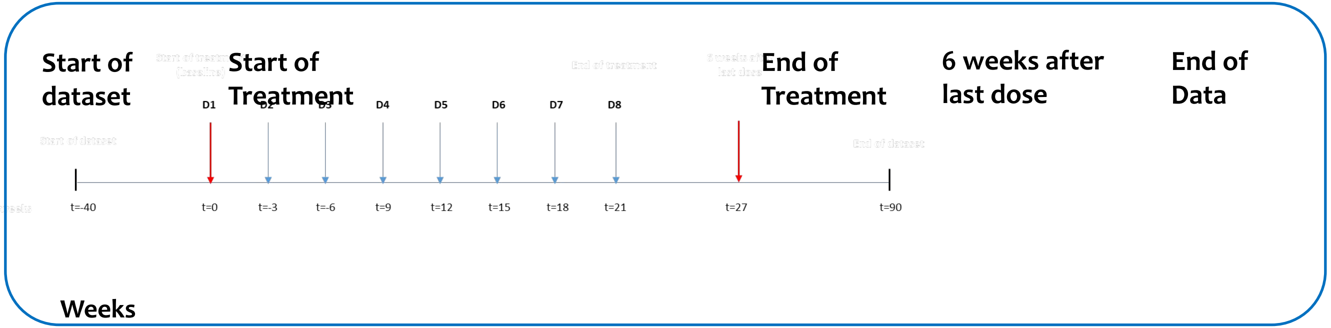


Lemaire et al., ASCPT 2019, Clinical Pharmacology & Therapeutics v105

*"**Biology is likely far too complex and messy to ever be encapsulated as a simple set of neat mathematical equations**. But just as mathematics turned out to be the right description language for physics, **biology may turn out to be the perfect type of regime for the application of AI**."*

- Demis Hassabis, CEO: DeepMind/Isomorphic Labs

Introduction: Dosing Description



Untreated case and 3 treatment arms:

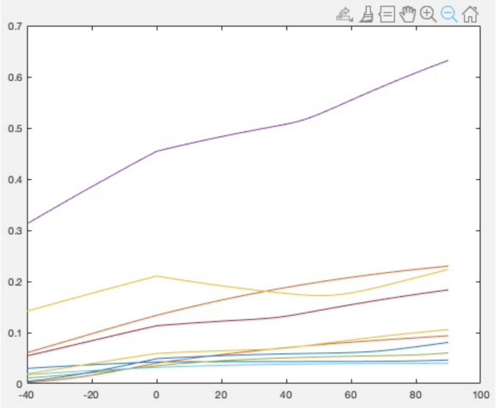
- Untreated
- Atezolizumab

1. T: Tumor size (pmol tumor cells)
2. Ag: Tumor antigens in the tumor space (pmol tumor antigens per liter of tumor space)
3. R: T Regulatory cells in peripheral blood (pmol cells)
4. H: T helper 1 cells in peripheral blood (pmol cells)
5. C: Cytotoxic T cells in peripheral blood (pmol cells)
6. RT: T Regs in the tumor space (pmol cells)
7. HT: T helper 1 cells in the tumor space (pmol cells)
8. CT: Cytotoxic T cells int the tumor space (pmol cells)
9. Cyde: Total cytotoxic death (total pmol tumor cells killed per liter of tumor space per day)

Output Variables

1. T: Tumor size (pmol tumor cells)
2. Ag: Tumor antigens in the tumor space (pmol tumor antigens per liter of tumor space)
3. R: T Regulatory cells in peripheral blood (pmol cells)
4. H: T helper 1 cells in peripheral blood (pmol cells)
5. C: Cytotoxic T cells in peripheral blood (pmol cells)
6. RT: T Regs in the tumor space (pmol cells)
7. HT: T helper 1 cells in the tumor space (pmol cells)
8. CT: Cytotoxic T cells int the tumor space (pmol cells)
9. Cyde: Total cytotoxic death (total pmol tumor cells killed per liter of tumor space per day)

Plots the tumor curves of 10 VPs drawn randomly from the VPop after treatment with atezolizumab:



Tumor Response Predictions

ODE parameter input (nVP x Parameter Total)

	param1	param2	param3	param4	param5	param6	param7	param8	param9	param10	...	param135	param136	param137	param138	parar
0	0.062634	0.009535	0.195818	0.541676	2.145568	0.589604	1.845076	0.788936	0.190571	0.109339	...	0.490933	0.045937	0.174120	0.219884	2.93
1	0.050782	0.007997	0.165107	0.059236	7.199288	0.280072	2.954273	0.358183	0.343335	0.290389	...	0.389982	0.116961	0.363144	0.212258	6.45
2	0.047898	0.003772	0.482840	0.538632	4.403517	0.489704	2.828143	0.866641	0.116924	0.321305	...	0.255452	0.151598	0.371907	0.352394	6.75
3	0.056124	0.007052	0.270573	0.607347	2.434118	0.331034	3.818806	0.567811	0.395329	0.452369	...	0.402877	0.172348	0.133558	0.495851	2.58
4	0.054890	0.002686	0.221682	0.341635	8.520480	0.377646	4.050971	0.414549	0.313547	0.779070	...	0.515292	0.152060	0.376253	0.483987	5.91
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
8342	0.079013	0.007331	0.538638	0.258569	5.877707	0.559215	3.475271	0.816070	0.307728	0.186831	...	0.554212	0.132839	0.196642	0.493638	2.17
8343	0.069440	0.008369	0.701176	0.512504	6.605340	0.284484	4.989012	0.978155	0.269484	0.358895	...	0.263358	0.138811	0.366587	0.329723	2.13
8344	0.068121	0.005464	0.680377	0.337729	4.962087	0.440508	2.086325	0.193523	0.224164	0.381748	...	0.563846	0.126276	0.318534	0.280019	3.96
8345	0.053753	0.005717	0.187533	0.540387	3.124140	0.533369	3.519511	0.899892	0.209199	0.207606	...	0.550188	0.123984	0.375216	0.371862	4.43
8346	0.037470	0.004705	0.471046	0.631253	1.194096	0.234200	2.200571	0.802492	0.298582	0.236690	...	0.479289	0.053088	0.383718	0.339195	2.56

8347 rows × 144 columns

nVP ≡ Number of Virtual Patients

Tumor VP Variability

	Untreated	Atezo	Tigit	Combination
0	0.012076	0.012076	0.012076	0.012076
1	0.003457	0.003457	0.003457	0.003457
2	0.107160	0.107160	0.107160	0.107160
3	0.040739	0.040739	0.040739	0.040739
4	0.042649	0.042649	0.042649	0.042649
...	...	...	...	...
8342	0.007925	0.007925	0.007925	0.007925
8343	0.025677	0.025677	0.025677	0.025677
8344	0.005865	0.005865	0.005865	0.005865
8345	0.029339	0.029339	0.029339	0.029339
8346	0.003000	0.003000	0.003000	0.003000

time=0 weeks

	Untreated	Atezo	Tigit	Combination
0	0.045169	0.041972	0.042706	0.039189
1	0.038780	0.034350	0.036882	0.028880
2	0.294082	0.277875	0.292828	0.267143
3	0.089850	0.078985	0.085866	0.052574
4	0.152771	0.063288	0.150522	0.060114
...	...	...	...	...
8342	0.109248	0.111286	0.095647	0.094065
8343	0.071422	0.060776	0.070177	0.058944
8344	0.058440	0.056153	0.037594	0.035912
8345	0.130368	0.127425	0.128854	0.124473
8346	0.013047	0.011721	0.009945	0.008213

time=27 weeks

Tumor size at week 0 and 27

$$\frac{T|t=27 - T|t=0}{T|t=0}$$

Response(Continuous)

	Untreated	Atezo	Tigit	Combination
0	0.362832	0.267531	0.289822	0.185283
1	0.675372	0.486541	0.593874	0.254085
2	0.354729	0.281445	0.349051	0.232545
3	0.229928	0.083760	0.174484	-0.269673
4	0.496833	-0.366377	0.475444	-0.397141
...	...	...	...	...
8342	0.756451	0.786647	0.540587	0.513981
8343	0.274611	0.092528	0.252350	0.060703
8344	0.813503	0.744150	0.184829	0.132998
8345	0.472169	0.439028	0.454819	0.406486
8346	0.521889	0.368776	0.165900	-0.034238

Response(Continuous)

	Untreated	Atezo	Tigit	Combination
0	0.362832	0.267531	0.289822	0.185283
1	0.675372	0.486541	0.593874	0.254085
2	0.354729	0.281445	0.349051	0.232545
3	0.229928	0.083760	0.174484	-0.269673
4	0.496833	-0.366377	0.475444	-0.397141
...	...	...	...	...
8342	0.756451	0.786647	0.540587	0.513981
8343	0.274611	0.092528	0.252350	0.060703
8344	0.813503	0.744150	0.184829	0.132998
8345	0.472169	0.439028	0.454819	0.406486
8346	0.521889	0.368776	0.165900	-0.034238

Conditions logic

Complete Responser(Condition 0) $\equiv \dfrac{T|t=27 - T|t=0}{T|t=0} = -1$

Partial Responser(Condition 1) $\equiv -1 < \dfrac{T|t=27 - T|t=0}{T|t=0} <= -0.3$

Stable Disease(Condition 2) $\equiv -0.3 < \dfrac{T|t=27 - T|t=0}{T|t=0} <= 0.2$

Progressive Disease(Condition 3) $\equiv 0.2 < \dfrac{T|t=27 - T|t=0}{T|t=0}$

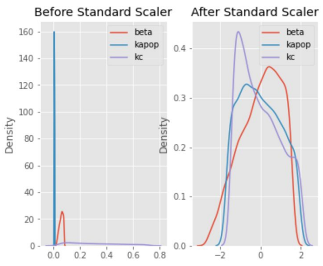
Response Type(Categorical)
(nVP x Treatment Total)

	Untreated	Atezo	Tigit	Combination
0	3	3	3	2
1	3	3	3	3
2	3	3	3	3
3	3	2	2	2
4	3	1	3	1
...	...	...	...	...
8342	3	3	3	3
8343	3	2	3	2
8344	3	3	2	2
8345	3	3	3	3
8346	3	3	2	2

Feature Scaling: Normalization

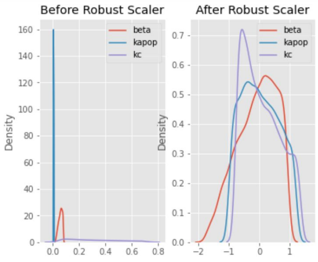
Standard Scaler

$$\frac{x_i - mean(x)}{stdev(x)}$$



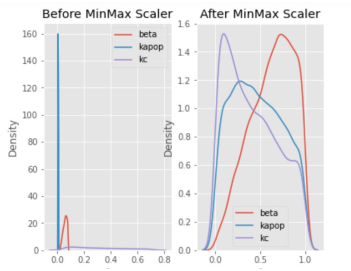
Robust Scaler

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$



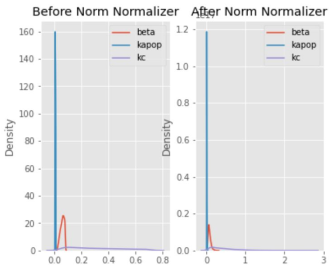
MinMax Scaler

$$\frac{x_i - min(x)}{max(x) - min(x)}$$



Norm Normalizer

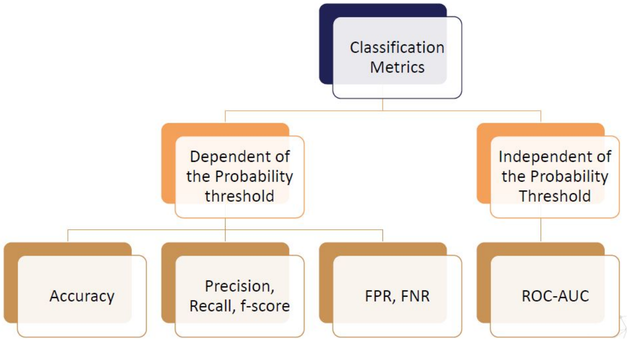
$$\frac{x_i}{\sqrt{x_i^2 + y_i^2 + z_i^2}}$$



Hyperparameter Tuning Description

- **Manual Search:** Time Consuming/does not scale.
- **Grid Search:** Examines all combinations of hyperparameters. Is often called the brute force approach. Computationally Expensive, so it cannot explore entire hyperparameter space.
- **Random Search:** Examines hyperparameters by randomly picking a set and training the models. Less computationally expensive, but it does not learn from its past and does not guaranteed an optimal solution.
- **Bayesian Optimization:** It considers the set of possibilities via a Probability density function. Sequential Model Based Global Optimization using Hypeopt and Optuna was used. Average metric cross validation was minimized.
    I. Random Forest
    II. Tree-structured Parzen Estimator: Utilizes acquisition functions i.e expected improvement/bayes theorem
    III. Gaussian Process

Model Performance Evaluations for Classification



		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

$TP \equiv$ True Positive | $FN \equiv$ False Negative

$FP \equiv$ False Positive | $TN \equiv$ True Negative

$$P = TP + FN \equiv \text{Positives}$$

$$N = FP + TN \equiv \text{Negatives}$$

$$\text{Accuracy}= \frac{TP + TN}{P + N}$$

$$\text{Precision}= \frac{TP}{TP + FP}$$

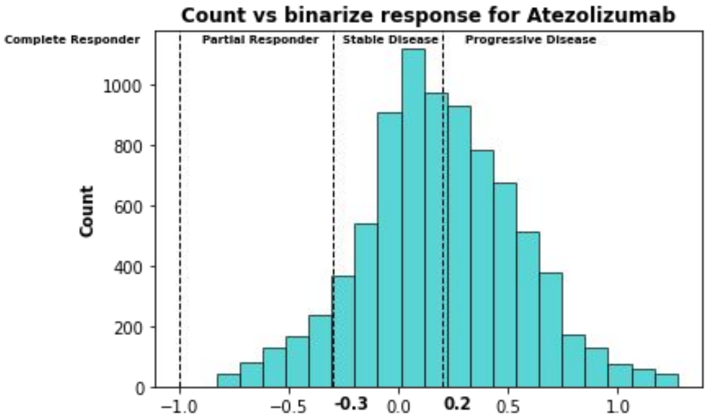
$$\text{True Positive Rate (TPR)}= \frac{TP}{P} \equiv \text{Recall}$$

$$\text{F1}Score\text{(F1)}= \frac{2\,.\,Precision\,.\,Recall}{(Precision + Recall)}$$

Other Metrics:

$$\text{True Negative Rate (TNR)}= \frac{TN}{N} \equiv \text{Specificity}$$

Response Distribution Atezolizumab



Conditions logic

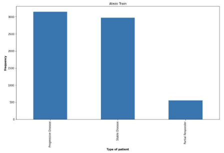
Complete Responser(Condition 0) $\equiv \frac{T|t{=}27 - T|t{=}0}{T|t{=}0} = -1$

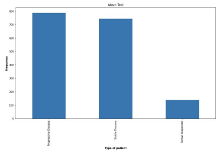
Partial Responser(Condition 1) $\equiv -1 < \frac{T|t{=}27 - T|t{=}0}{T|t{=}0} <= -0.3$

Stable Disease(Condition 2) $\equiv -0.3 < \frac{T|t{=}27 - T|t{=}0}{T|t{=}0} <= 0.2$

Progressive Disease(Condition 3) $\equiv 0.2 < \frac{T|t{=}27 - T|t{=}0}{T|t{=}0}$

Atezo Class Types Count/Frequencies





Key Definitions

XGBoost (Machine learning)

- Non-parametric tree-based method to predict variable of interest
- Supports use with missing data
- Amenable for binary classification, regression, and survival analyses

SHAP analysis (ML explainability)

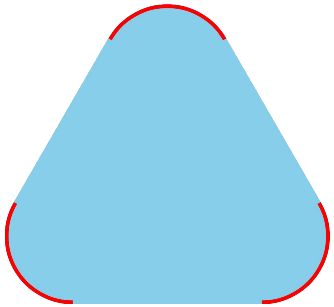
- Game theoretic approach to explain how ML models generate predictions based on inputs
- SHAP values represent the adjusted impact of an input on model prediction

Pareto Front

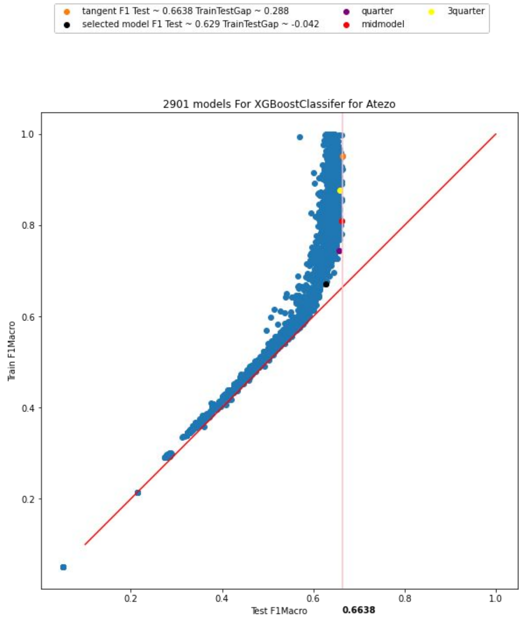
- Set of all Efficient Solutions in a multi-objective optimization set up
- Make trade-offs within the set

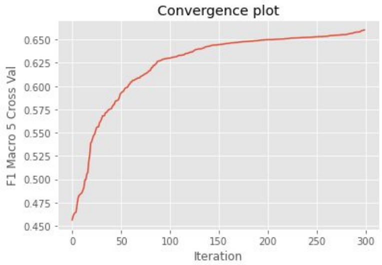
Convex Hull/Convex set

- Convex hull of a Shape is the smallest convex set that contains it.
- SHAP values represent the adjusted impact of an input on model prediction



Atezo; Model Selected: XGBoost with Random OverSampling





Optimal Models on Pareto Front