

Empirical means to validate skills models and assess the fit of a student model

Behzad Beheshti

Supervisor : Michel C. Desmarais

Génie Informatique Et Génie Logiciel
École Polytechnique de Montréal

25 février 2016

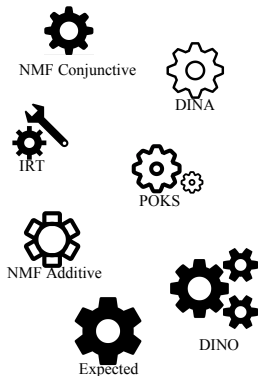
Problem Specification

- Student skills assessment models



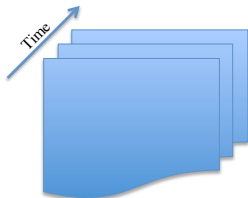
Problem Specification

- Student skills assessment models
- How to decide which are the most representative of the underlying ground truth ?



Problem Specification

- Student skills assessment models
- How to decide which are the most representative of the underlying ground truth ?
- Static Vs. Dynamic



Dynamic

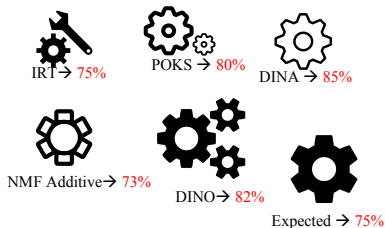
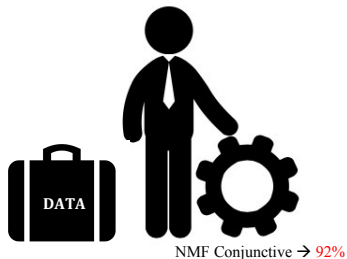


Static



Problem Specification

- Student skills assessment models
- How to decide which are the most representative of the underlying ground truth ?
- Static Vs. Dynamic
- Model selection and goodness of fit
- A general answer : best performer



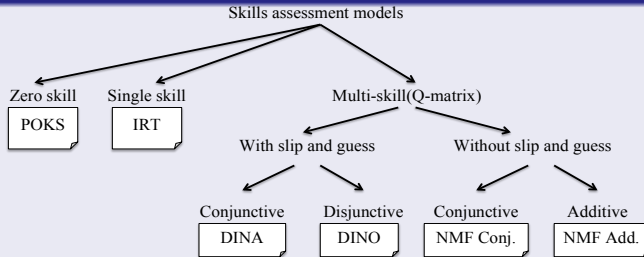
Problem Specification

- Student skills assessment models
- How to decide which are the most representative of the underlying ground truth ?
- Static Vs. Dynamic
- Model selection and goodness of fit
- A general answer : best performer
- Our contribution
 - To make a comprehensive comparison of educational data model performances
 - To propose a new approach to assessing model fit

Problem Specification

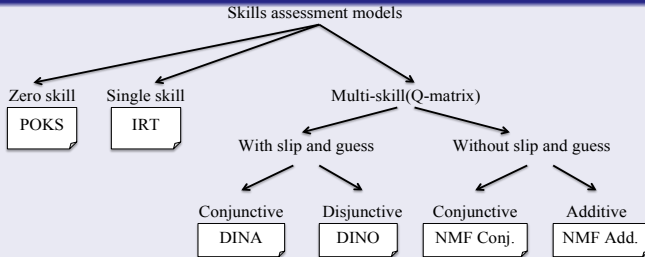
- Student skills assessment models
- How to decide which are the most representative of the underlying ground truth ?
- Static Vs. Dynamic
- Model selection and goodness of fit
- A general answer : best performer
- Our contribution
 - To make a comprehensive comparison of educational data model performances
 - To propose a new approach to assessing model fit
- The proposed approach :
 - Assessing the fit of the model to the underlying ground truth using a methodology based on **synthetic data**

Student skills assessment models



- Number of Skills

Student skills assessment models



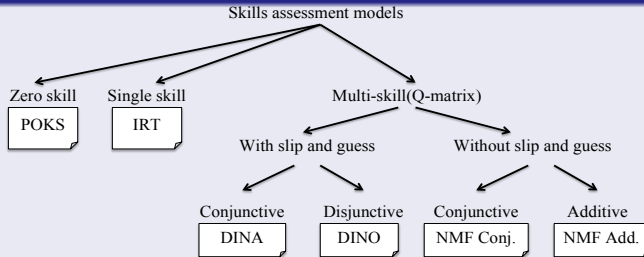
- Number of Skills
- Q-matrix

s_1 : fraction multiplication
 s_2 : fraction addition
 s_3 : fraction reduction

$$\begin{array}{lcl}
 i_1 & \frac{4}{12} + \frac{3}{5} = \frac{8}{5} \\
 i_2 & \frac{4}{12} = \frac{4 \times 3}{12} = \frac{12}{12} = 1 \\
 i_3 & 1 + \frac{3}{5} = \frac{8}{5} \\
 i_4 & 2 \times \frac{1}{2} = 1
 \end{array}$$

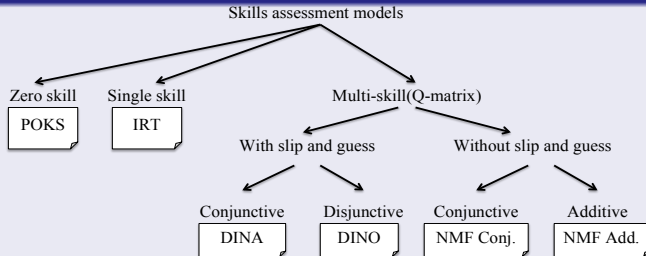
		Skills		
Items	i_1	s_1	s_2	s_3
	i_1	1	1	1
	i_2	1	0	1
	i_3	0	1	1
	i_4	1	0	1

Student skills assessment models



- Number of Skills
- Q-matrix
- Slip and Guess

Student skills assessment models

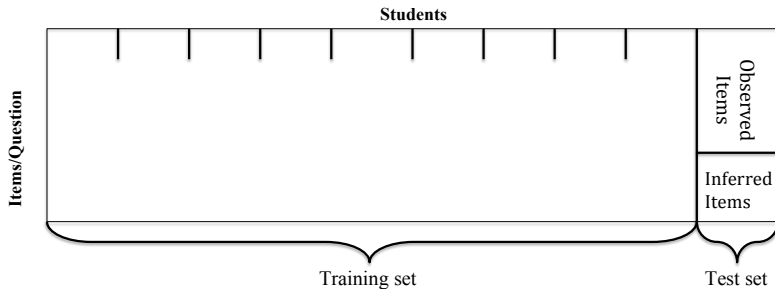


- Number of Skills
- **Q-matrices (types)**
- Slip and Guess

- 1 Conjunctive
- 2 Additive
- 3 Disjunctive

		Skills		
		s_1	s_2	s_3
Items	i_1	1	1	1
	i_2	1	0	1
	i_3	0	1	1
	i_4	1	0	1

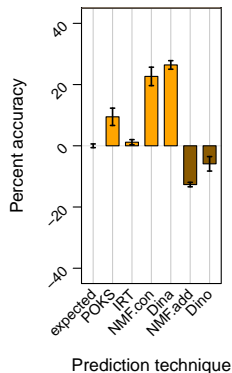
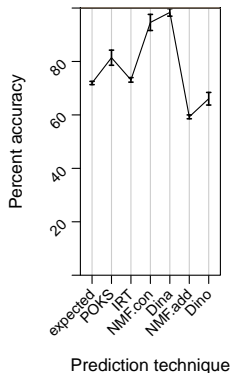
- **Performance of a model over a data set**



- **Performance of a model** over a data set
- Model parameters

		Parameters estimated from		
Skills Model		Training set		Observed items
Contributed skills	Multiple	NMF Conj.	• Q-matrix	• Students skills mastery matrix
		NMF Add.		
		DINA		
		DINO		
	Single	• Slip	• Item difficulty • Item discrimination	• Student Ability
		• Guess		
		Expected	• Item Odds	• Student Odds
	Zero	POKS	• Initial Odds • Odds ratio • Partial order structure	

- **Performance of a model** over a data set
- Model parameters
- Performance vector
- Performance signature



Model	Performace
<i>Expected</i>	0.72%
<i>POKS</i>	0.80%
<i>IRT</i>	0.74%
<i>NMF.Conj</i>	0.94%
<i>Dina</i>	0.99%
<i>NMF.Add</i>	0.60%
<i>Dino</i>	0.65%

- **Performance of a model** over a data set
- Model parameters
- Performance vector
- Performance signature
- Performance prototype

The *performance vector* associated with the synthetic data of a model class.

- **Performance of a model** over a data set
- Model parameters
- Performance vector
- Performance signature
- Performance prototype
- Target performance vector

The *performance vector* of the real data set to classify.

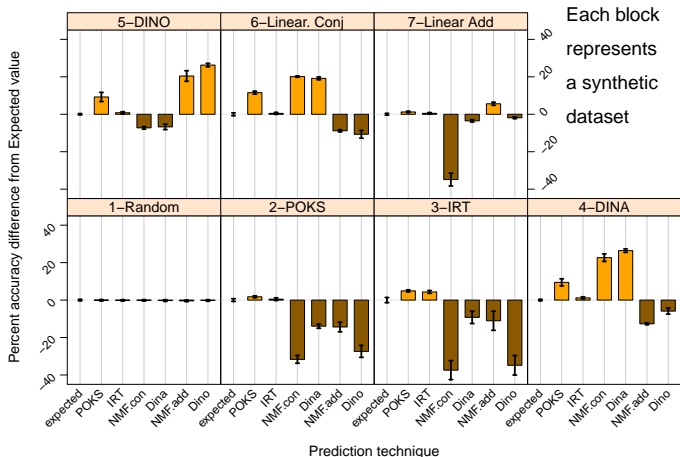
Research questions

- ① ✓ What is the *performance vector* of student skills assessment models over real and over synthetic data created using the same models ?
 - Experiment 1 : Predictive performance of models over real and synthetic data sets

Datasets

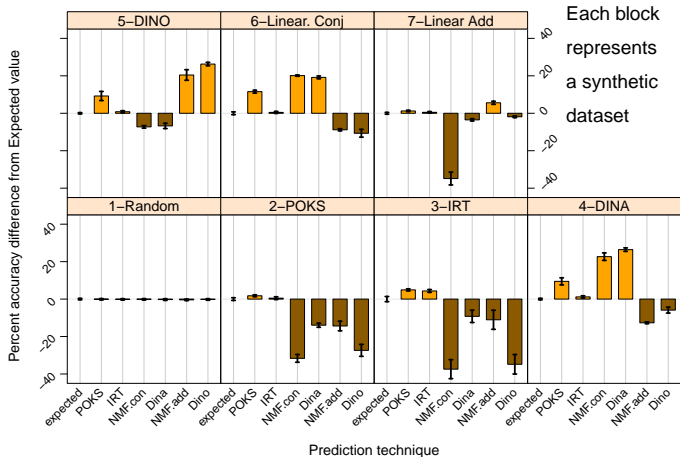
Data set	Number of			Mean Score	Q-matrix
	Skills	Items	Students		
Synthetic					
1.Random	7	30	700	0.75	Q₀₁
2.POKS	7	20	500	0.50	Q₀₂
3.IRT-2PL	5	20	600	0.50	Q₀₃
4.DINA	7	28	500	0.31	Q₅
5.DINO	7	28	500	0.69	Q₆
Linear (Matrix factorization)					
6. Conj.	8	20	500	0.24	Q₁
7. Comp.	8	20	500	0.57	Q₁
Real					
8.Fraction	8	20	536	0.53	Q₁
9.Vomlel	6	20	149	0.61	Q₄
10.ECPE	3	28	2922	0.71	Q₃
Fraction subsets and variants of Q₁					
11. 1	5	15	536	0.53	Q₁₀
12. 2/1	3	11	536	0.51	Q₁₁
13. 2/2	5	11	536	0.51	Q₁₂
14. 2/3	3	11	536	0.51	Q₁₃

Predictive performance of models over synthetic datasets



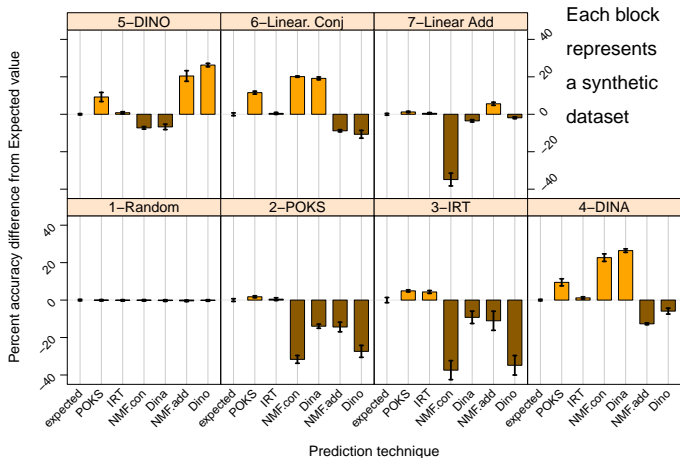
The highest performance is for the generative model behind the dataset

Predictive performance of models over synthetic datasets



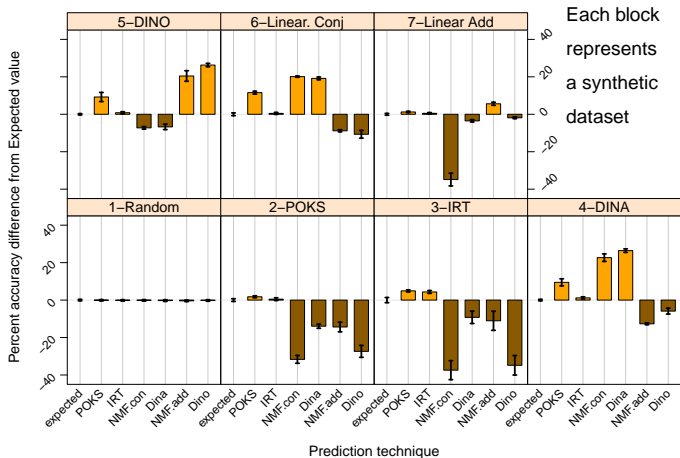
Data sets have unique pattern of performance vector across models

Predictive performance of models over synthetic datasets



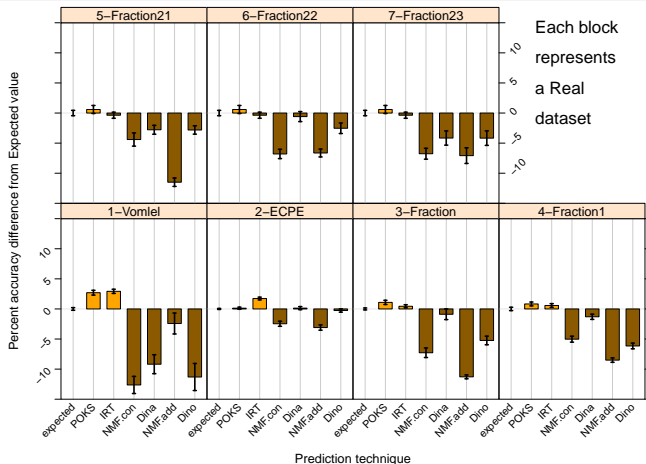
The random data set has a flat performance across techniques

Predictive performance of models over synthetic datasets



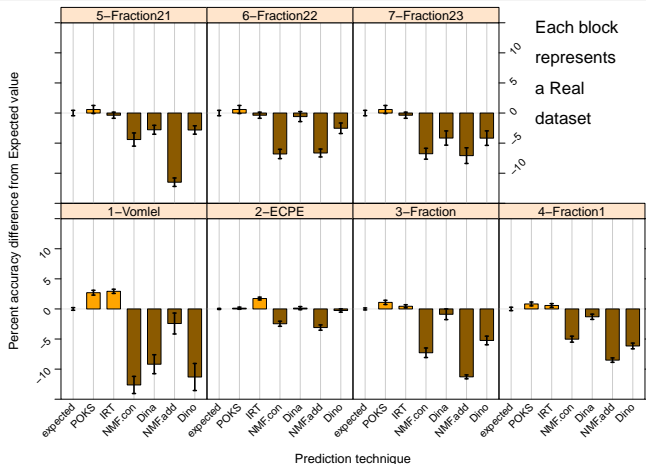
The capacity of recognizing a data set's true model relies on this uniqueness characteristic

Predictive performance of models over real datasets



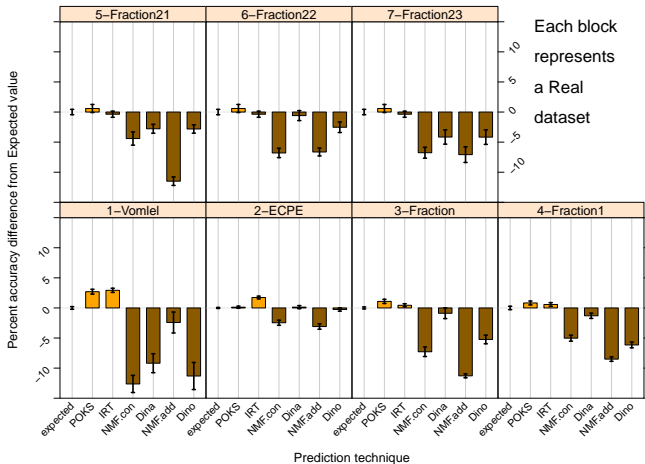
In most cases, the best performer is close to the baseline

Predictive performance of models over real datasets



The pattern of the Fraction performance data set repeats over its subsets

Predictive performance of models over real datasets



None of the real data sets show the large the amplitude and the differences found in the synthetic data sets models

Vector space of accuracy performances

- Performance vectors of datasets in columns(Data points in the performance space)

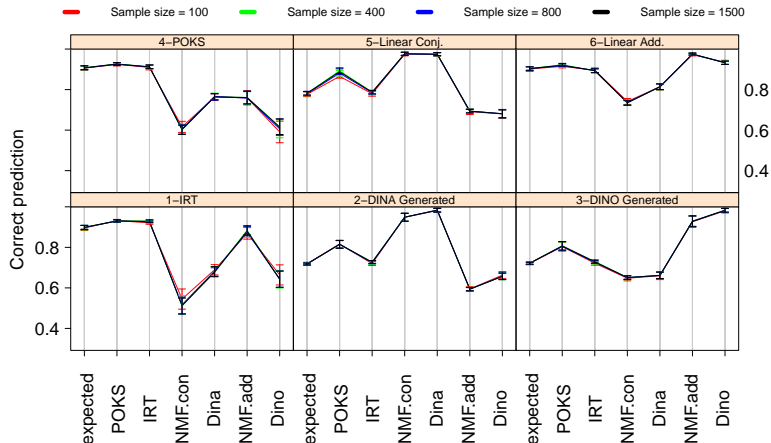
Model	Synthetic data set						
	<i>Random</i>	POKS	IRT	DINA	DINO	Linear .Conj	Linear .Comp
<i>Expected</i>	0.75	0.91	0.90	0.72	0.72	0.78	0.93
POKS	0.75	0.94	0.94	0.81	0.81	0.90	0.94
IRT	0.75	0.91	0.95	0.73	0.73	0.79	0.89
DINA	0.75	0.77	0.81	1.00	0.65	0.98	0.89
DINO	0.75	0.63	0.56	0.66	1.00	0.68	0.91
NMF.Conj	0.75	0.59	0.53	0.95	0.65	0.97	0.58
NMF.Comp	0.75	0.76	0.79	0.59	0.93	0.70	0.98

The diagonal generally displays the best performance

Research questions

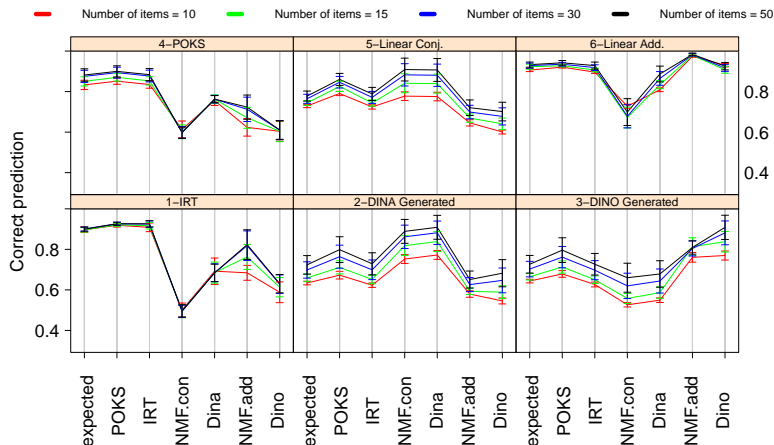
- ① ✓ What is the *performance vector* of student skills assessment models over real and over synthetic data created using the same models ?
 - Experiment 1 : Predictive performance of models over real and synthetic data sets
- ② **Is the performance vector unique to each synthetic data type (data from the same ground truth model) ?**
 - Experiment 2 : Sensitivity of the Model performance over different data generation parameters

Variation of sample size over synthetic data sets



Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1

Variation of number of items over synthetic data sets



Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1

Research questions

- ① ✓ What is the *performance vector* of student skills assessment models over real and over synthetic data created using the same models ?
 - Experiment 1 : Predictive performance of models over real and synthetic data sets
- ② ✓ Is the *performance vector* unique to each synthetic data type (data from the same ground truth model) ?
 - Experiment 2 : Sensitivity of the Model performance over different data generation parameters
- ③ **Can the performance vector be used to define a method to reliably identify the ground truth behind the synthetic data ?**
 - Experiment 3 : Model selection based on performance vector classification

Degree of similarity between six synthetic datasets based on the correlation

Synthetic Datasets

Synthetic Datasets		POKS	IRT	NMF Conj.	DINA	NMF Add.	DINO
	POKS	0.96					
	IRT	0.86	0.96				
	NMF Conj.	0.22	-0.20	0.96			
	DINA	0.02	-0.40	0.94	0.96		
	NMF Add.	0.44	0.75	-0.62	-0.73	0.93	
	DINO	-0.15	0.20	-0.70	-0.69	0.63	0.95

Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1

Degree of similarity between six synthetic datasets and the ground truth based on the correlation

Real Datasets

				Fraction subsets			
				1	21	22	23
Synthetic Datasets	Vomlel	ECPE	Fraction	0.43	0.24	0.61	0.57
	Random	0.58	0.73	0.61	0.88	0.60	0.77
	IRT	0.90	0.42	0.72	0.30	0.56	0.06
	DINA	-0.38	-0.09	0.23	-0.31	0.10	-0.08
	DINO	0.34	0.15	-0.18	0.95	0.70	0.83
	POKS	0.75	0.40	0.83	0.55	0.66	0.33
	NMF Conj.	-0.05	0.54	0.51	-0.19	-0.03	0.13
NMF Add.	0.39	0.06	-0.04	0.28			

Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1 Item1
Item1 Item1 Item1 Item1

Research questions

- ① ✓ What is the *performance vector* of student skills assessment models over real and over synthetic data created using the same models ?
 - Experiment 1 : Predictive performance of models over real and synthetic data sets
- ② ✓ Is the *performance vector* unique to each synthetic data type (data from the same ground truth model) ?
 - Experiment 2 : Sensitivity of the Model performance over different data generation parameters
- ③ ✓ Can the *performance vector* be used to define a method to reliably identify the ground truth behind the synthetic data ?
 - Experiment 3 : Model selection based on performance vector classification
- ④ **How does the method compare with the standard practice of using the model with the best performance ?**
 - Experiment 4 : Signature vs. best performer classification

Research questions

- ① ✓ What is the *performance vector* of student skills assessment models over real and over synthetic data created using the same models?
 - Experiment 1 : Predictive performance of models over real and synthetic data sets
- ② ✓ Is the *performance vector* unique to each synthetic data type (data from the same ground truth model)?
 - Experiment 2 : Sensitivity of the Model performance over different data generation parameters
- ③ ✓ Can the *performance vector* be used to define a method to reliably identify the ground truth behind the synthetic data?
 - Experiment 3 : Model selection based on performance vector classification
- ④ ✓ How does the method compare with the standard practice of using the model with the best performance?
 - Experiment 4 : Signature vs. best performer classification

Thank you

