Star Formation Rates with ML methods

(a sub-project in a much larger one....)

Michele Delli Veneri[(2)], Stefano Cavuoti[(1-2)] , Massimo Brescia[(2)],O. Razim[(1)], Giuseppe Riccio[(2)]
O. Torbaniuk[1], M. Paolillo[1,3] & **Giuseppe Longo**[(1,3,4)]

1- University of Naples Federico II
2- Astronomical Observatory of Capodimonte
3- INFN - Napoli UNit
*4. **Merle Kingsley distinguished Visitor** - Caltech*

Horizon2020
European Union Funding
for Research & Innovation

The larger project...

Use ML methods to obtain a **physical classification of galaxies**, i.e. a classification scheme based on physical properties such as: AGN/non-AGN, SFR, Mass, SMBH mass, B/D ratio, sersic indexes, galaxy components, etc.

- galaxy formation and evolution
- DM and DE studies
- AGN census and properties
- exploitation of non optically selected samples (e.g. EMU survey for SKA, e-Rosita, etc)
- etc....

AGN/non AGn:
Torbaniuk, Paolillo + core team

AGN/nonAGN.
Deep learning A. Razim + coreteam

Sersic indexes and other optically derived parameters:
SUNDIAL teams

Morphologies... SUNDIAL teams (Wilkinson and Biehl)

Radio selected samples:
Ray Norris + core team

etc...

Main problem to be solved

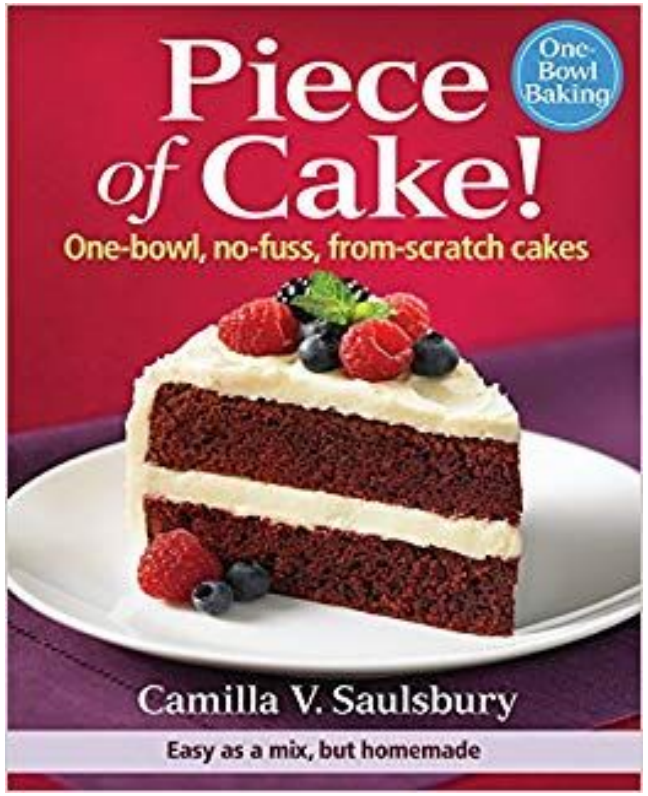
(cf. Umma Rebbapragada tutorial)

Astronomy LACKS of suitable, clean, reliable knowledge base on which to train supervised methods or to be used for labeling of unsupervised methods

Why SFRs?

at first sight a nice, simple regression problem for supervised learning...

hence having a long (too long) experience in photo-z evaluation should have proven useful....



1. 2017MNRAS.464.2577S, Sacrificing information for the greater good: how to select photometric bands for optimal accuracy, Stensbo-Smidt, K., et al.
2. 2019A&A...622A.137B , Star formation rates and stellar masses from machine learning, Bonjean, V. et al.
3. 2019MNRAS.486.1377D, \Star formation rates for photometric samples of galaxies using machine learning methods, Delli Veneri, M. bet al.
4. **it will explode ....**

Training data

Brinchmann Catalogue
Spectroscopic SFRs for a set of 10^6 galaxies in SDSS Dr7

→

Crossmatch
SDSS –DR7 - Brinchmann

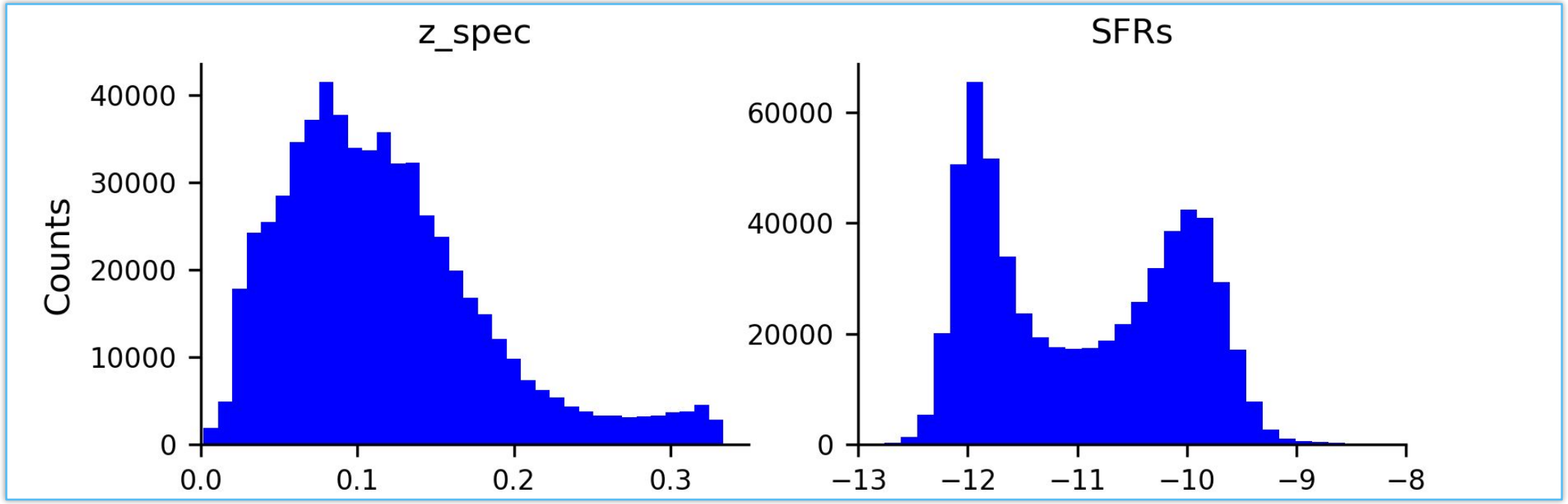
→

Data Pruning

→

603,680 galaxies, 55 features (magnitudes, colors, photometric redshift and spectroscopic redshift)

SFR as regression target



Crucial to the whole project… a new all relevant feature selection method

PHiLAB (*Parameter Handling investigation LABoratory*)

Able to solve the All-relevant feature selection!

Based on two concepts: «**shadow features**» and **Naïve-LASSO regularization** and exploiting Random Forest model as importance computing engine.

SHADOW FEATURES represent the noisy versions of the real ones and their calculated importance can be used to estimate the relevance of the real features.

A shadow feature for each real one is introduced by randomly shuffling its values among the N samples of the given dataset.

Kursa & Rudnicki 2010, Journal of Statistical Software, 36, 11

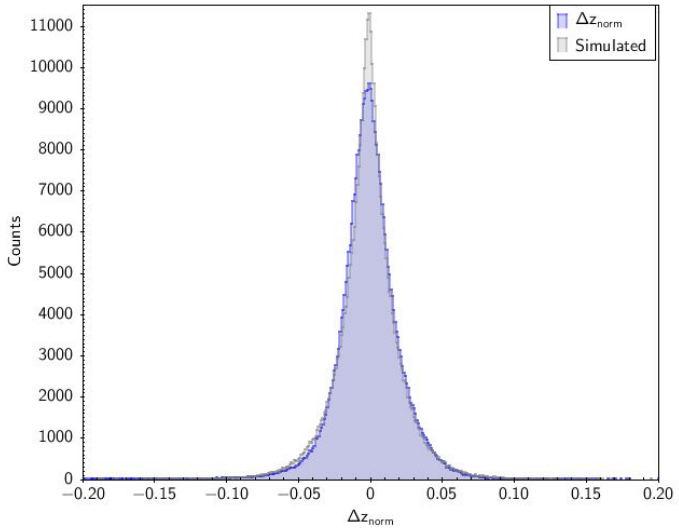
LASSO penalizes regression coefficients with an L_1-norm penalty, shrinking many of them to zero. Features with non-zero regression coefficients are "selected".

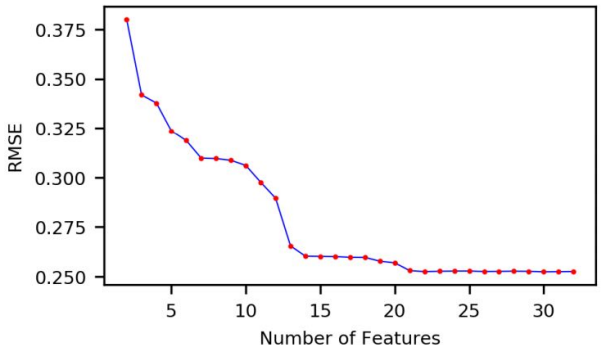
Regularization in Machine Learning is a process of introducing additional information to solve learning overfitting or to perform Feature Selection in a sparse Parameter Space. The regularization is a penalty term added to any loss function L.

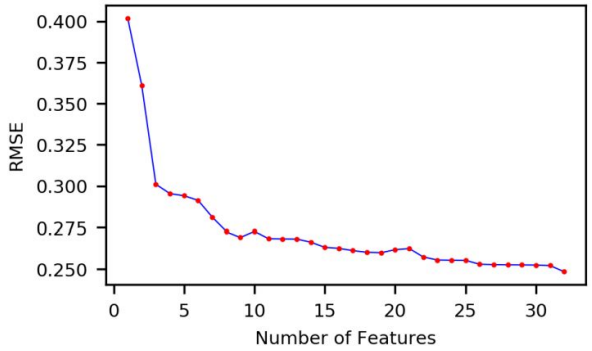
$$min_f \sum_{i=1}^{n} L(f(x)) + \boldsymbol{\lambda} \mathbf{L_{1-norm}}(\mathbf{w})$$

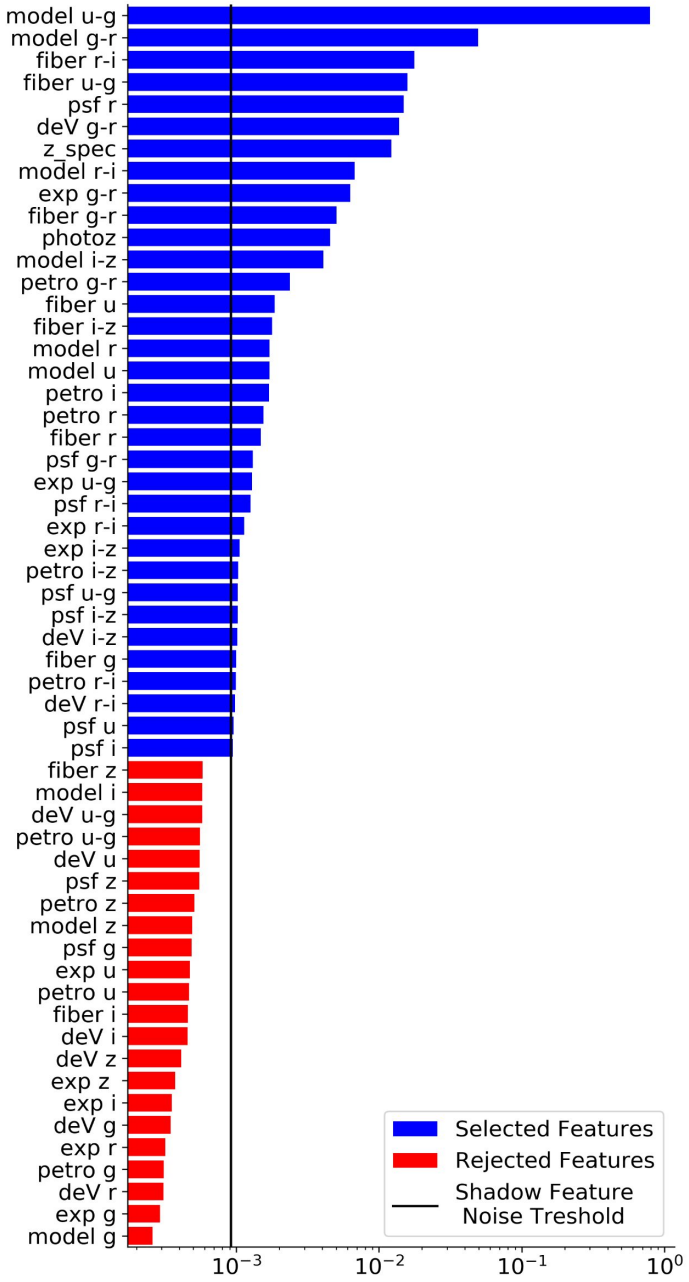
Hara & Maehara 2016, Proceedings of NIPS 2016, Barcelona, Spain

Feature selection









ID	n. of features	RMSE	Median	σ	η_{frac}
M+C	54	0.252	-0.021	0.252	2.07
ΦLABps	32	0.252	-0.021	0.252	2.03
REP	8	0.264	-0.020	0.264	1.86
SS4	8	0.274	0.013	0.274	1.85

ID	n. of features	RMSE	Median	σ	η_{frac}
PHI	32	0.252	-0.021	0.252	2.03
ZS	33	0.233	-0.017	0.233	2.24
ZP	33	0.252	-0.021	0.252	2.04

redshift used	RMSE	Median	σ	η_{frac}
$\sigma = 0.022$	0.249	-0.019	0.249	2.08
$\sigma = 0.015$	0.244	-0.019	0.244	2.11
$\sigma = 0.007$	0.238	-0.018	0.238	2.18
$\sigma = 0.005$	0.236	-0.018	0.236	2.21
z_{spec}	0.233	-0.017	0.233	2.24

Number of training objects	RMSE	Median	σ	η_{frac}
36,000	0.278	-0.022	0.278	1.99
100,000	0.265	-0.022	0.265	1.97
362,208	0.252	-0.021	0.252	2.03

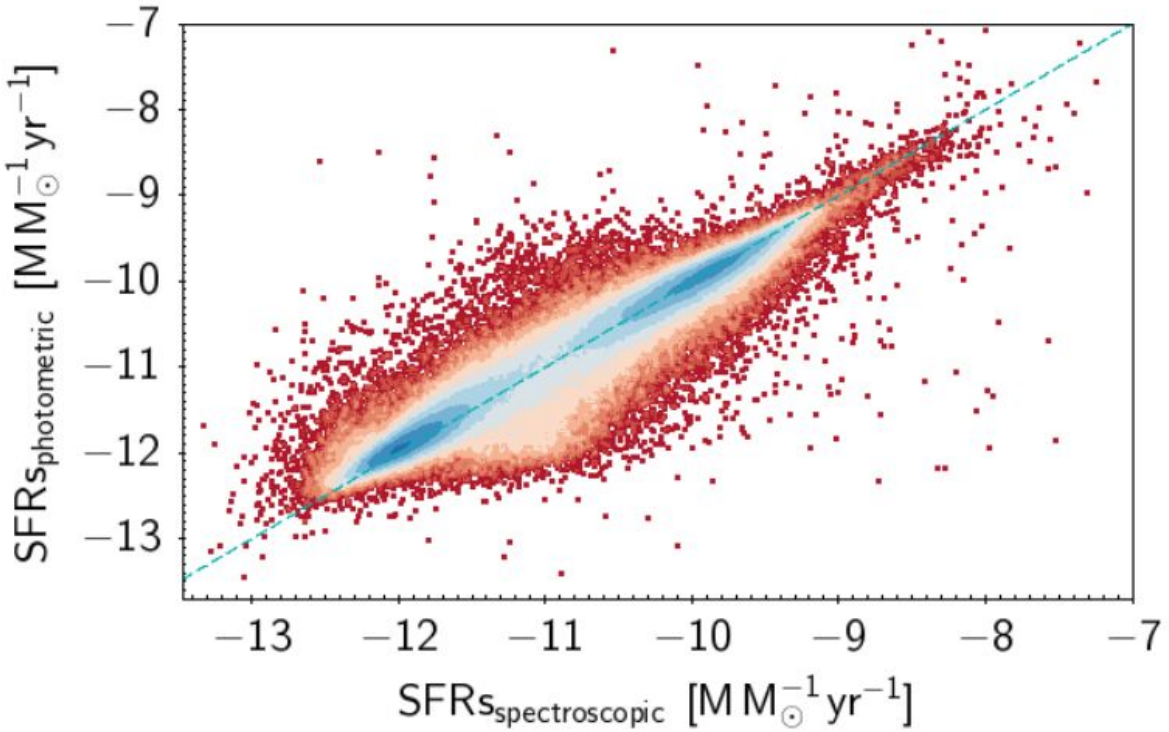
Random Forest

Number of training objects	RMSE	Median	σ	η_{frac}
36,000	0.337	-0.015	0.337	1.53
100,000	0.281	-0.017	0.281	1.62
362,208	0.248	-0.017	0.248	1.99

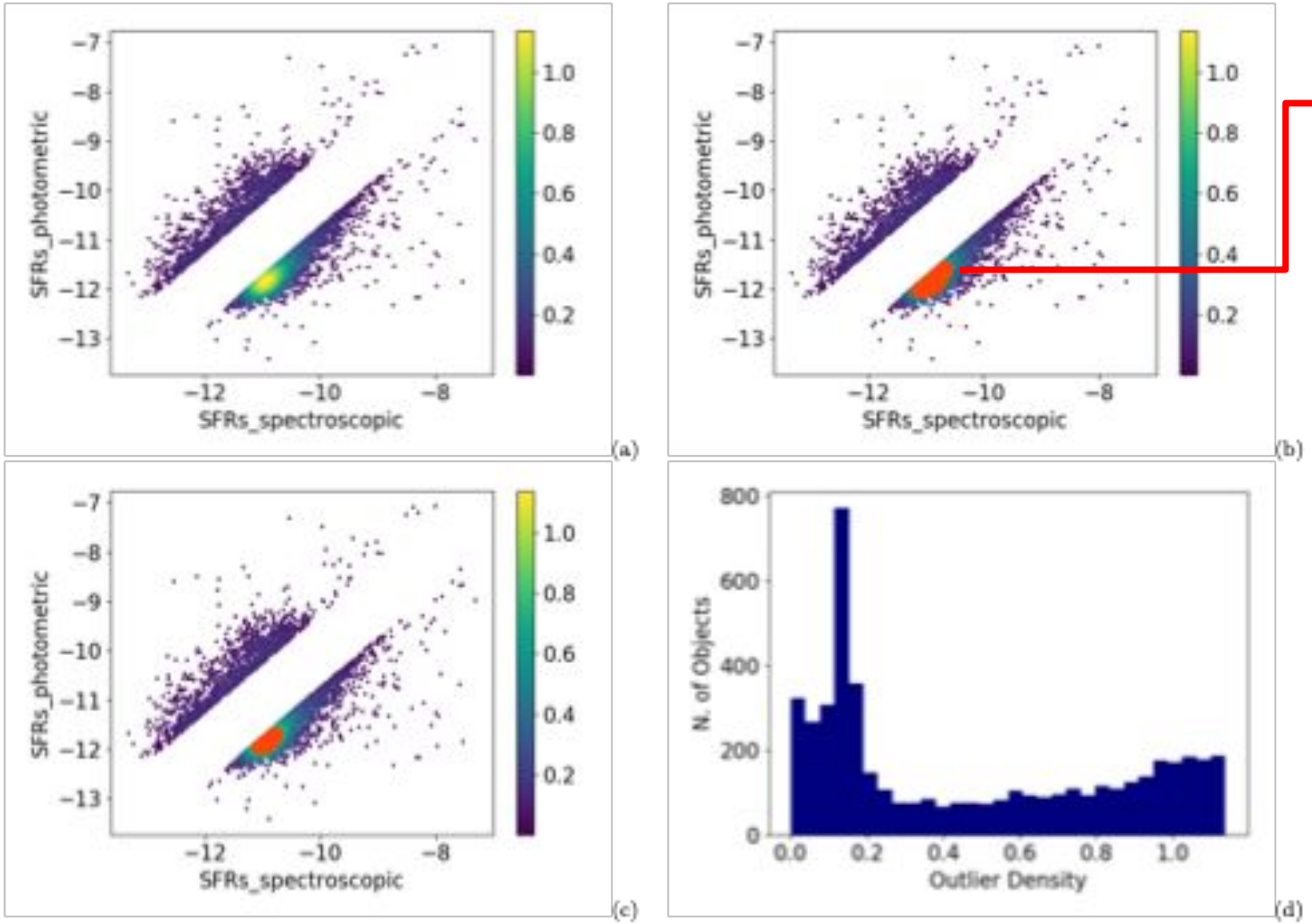
MLPQNA
(MLP with Quasi Newton Approximation)

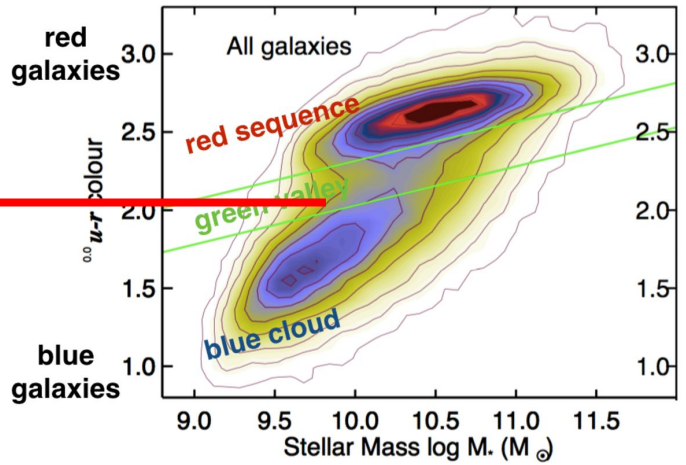
Problem not saturated
Need for more samples

Model	RMSE	Median	σ	η_{frac}
RF (paper 3)	0.252	-0.021	0.252	2.03%
MLPQNA (paper 3)	0.248	-0.017	0.248	1.99%
Stensbo-Smidt et al. 2016 (RF - paper 1)	0.274	0.013	0.274	1.85%



Understanding the outliers ....

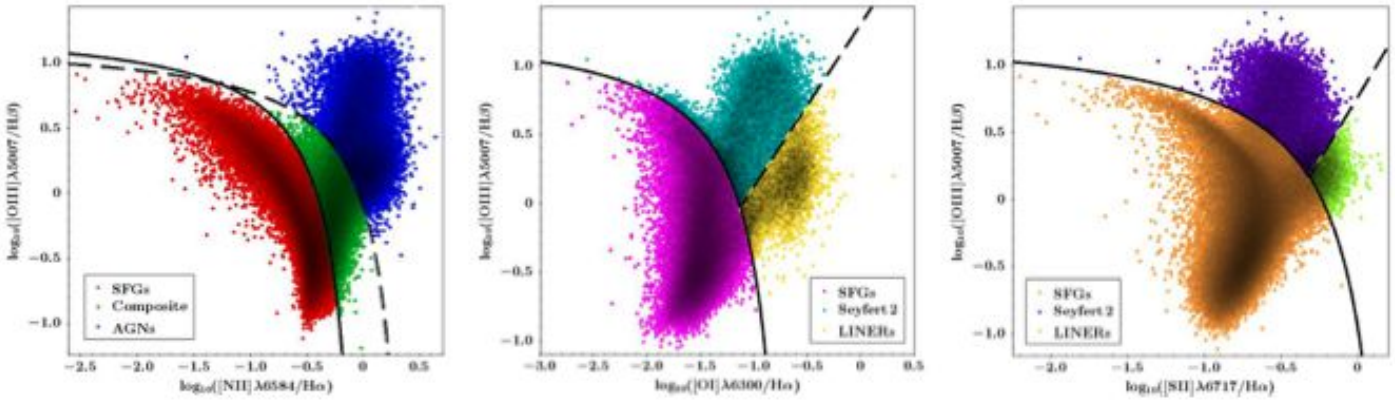




most objects in the overdensity region seemed to belong to the so called "green valley"

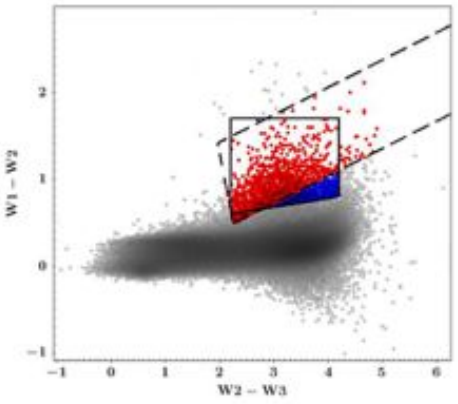
Figure 6. The scatter plot in the top left corner (a) shows the distribution of outliers in the $SFRs_{spectroscopic}$ VS $SFRs_{photometric}$ space with a superimposed density map, while the diagrams int the top right (b) and bottom left (c) corners show highlighted in orange all the objects with a density, respectively, six and eight times higher than the average point density. The histogram in the bottom right

Fig. 1: BPT diagrams for three set of lines: [NII] 6584 Å, [OI] 6300 Å and [SII] 6717 Å.

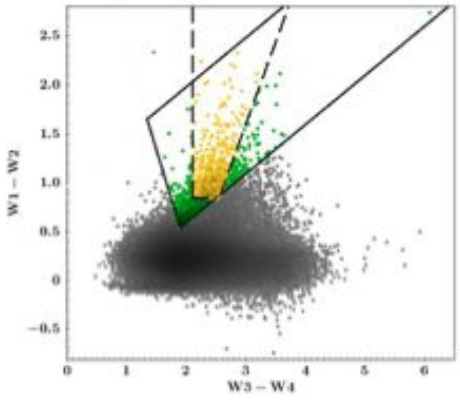


BPT1 [NII] line BPT2 [OI] line BPT3 [SII] line

Fig. 2: WISE 3-bands and 4-bands color-color diagrams.



WISE **3 bands** color-color diagram:
blue: AGNs selected by criteria from Jarrett et al. (2011)
red: AGNs selected by criteria from Mateos et al. (2012)



WISE **4 bands** color-color diagram:
green: AGNs selected by criteria from Mateos et al. (2012)
yellow: AGNs selected by criteria from Assef et al. (2010)

go back to the training set and explore biases in the KB.... (SDSS spectroscopic data)

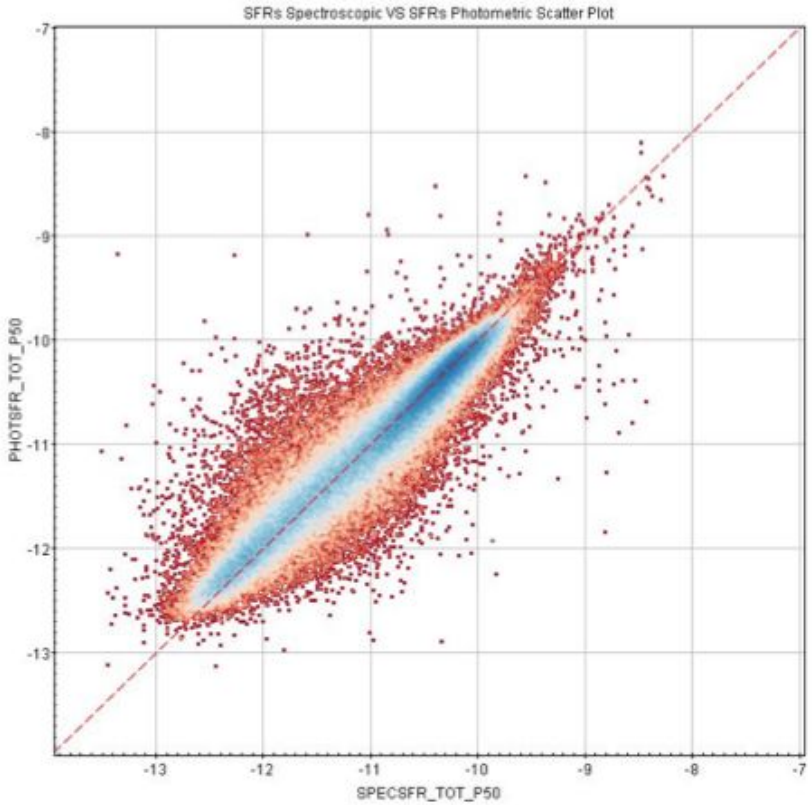
reclassify objects using BPT diagrams (AGN, normal, starburst)

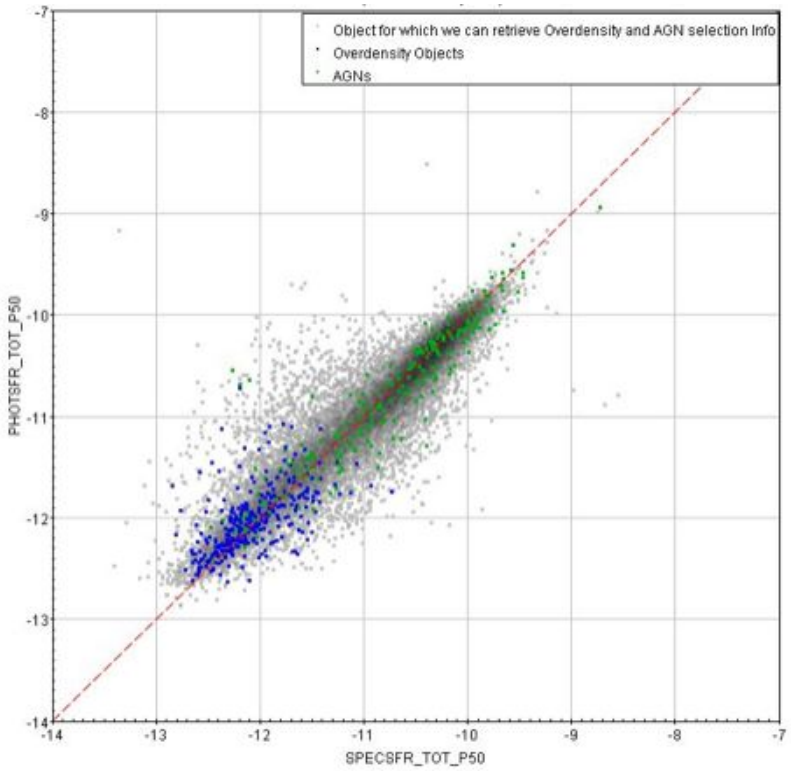
correct for obvious mistakes in line fluxes due to poor fitting

clean the KB

re-run FS and experiments

AND....





KB 1/3 of previous experiment

Run	RMSE	Median	η
Old	0.248	-0.017	1.99%
New	0.238	0.003	1.95%

- **No more green valley objects**
- **Confirmed correlation between SFR and the presence of AGN**
  ... needs further investigation
- **A catalogue of photometric SFR for 27 million of galaxies soon available on Vizier.**

Conclusions (just one in different flavours)

In the tutorial.... Umma put the finger in the wound....

In many (most) cases our results are strongly affected by biases in the knowledge base
GOOD TRAINING DATA ARE MUCH BETTER THAN A LOT OF TRAINING DATA

Careful analysis of the results is crucial and may lead to substantial improvemnts to the KB ...

Many iterations needed to get significant results

WE NEED GOOD (ANNOTATED OR WELL MEASURED)
TRAINING SAMPLES

Thank You for the Attention

Feature Selection with ΦLAB

What's behind the ***ΦLAB*** (*Parameter Handling investigation Laboratory)* project?....the property of **feature importance and relevance** in the context of a parameter space used to approach any prediction/classification task with machine learning methodology.

The **importance** of a feature is the relevance of its informative contribution to the solution of a learning problem.
The **relevance** of a feature can be formally defined as follows:

- Feature x is **strongly relevant** when removal of x from the parameter space always results in degradation of learning accuracy
- Feature x is **weakly relevant** if is not strongly relevant and there exists at least one subset S of features such that learning accuracy on S is worse than S U {x}
- Feature x is **irrelevant** if it is neither strongly nor weakly relevant.

feature selection problem taxonomy:
Minimal-optimal feature selection: selection of the smallest parameter space giving best accuracy. There are plenty of methods proposed in literature, either for prediction and classification problems (PCA, leave-one-out, forward selection, backward elimination, RF, PPS, Naive-Bayes, etc.).
All-relevant feature selection: the identification of the exact parameter space (all features) which are in some circumstances relevant for the problem solution. Basically, finding all relevant features, instead of only the non-redundant or unuseful ones, may help to understand the hidden mechanisms behind the problem. In more philosophical terms, it makes a predictive/classification model as a gray box, instead of merely as a black box!
There are very few methods proposed in literature to solve this type of feature selection.

Why all-relevant feature selection is challenging?

Random accuracy fluctuation: the impact of random fluctuation in the prediction/classification accuracy of a learning system. Such effect, common in all real problems, may condition and mask the true importance contribution of a weakly relevant feature.
The random fluctuation of accuracy usually does not affect the selection of strongly relevant features, thus not impacting on the minimal-optimal feature selection. But, since the core of all-relevant feature selection is the extraction of all weakly important features, it may dramatically affect such problem.

Obscuration of weakly relevance: the detection of weakly relevant features can be completely obscured by the strongly relevant ones. Therefore, in case of high-dimensionality problems, hand-made forward/backward/leave-one-out selection techniques may result impracticable for the all-relevant problem.

High-correlation compromise: in the frequent case of important features highly correlated, it is difficult to find the exact relevance contribution of single features. In such cases the most frequent compromise adopted by wrappers methods is to equally partition their importance by assigning them to the same relevance class. But this is always a simplification that could bring residual redundancy and mistakes in some cases.

Shadow features method is specialized to solve first issue, Naïve-LASSO the third issue, while both solve the second.

ΦLAB voting algorithm

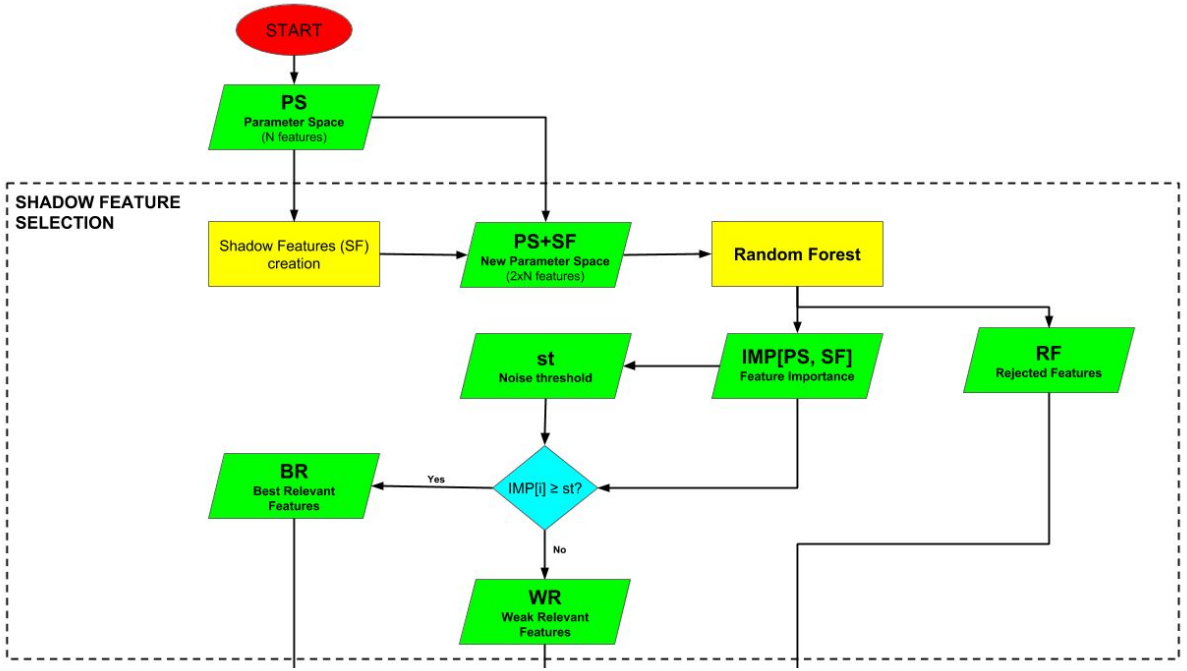
0. Let it be PS={$x_1 \ldots x_N$} the initial complete Parameter Space composed by N real features;

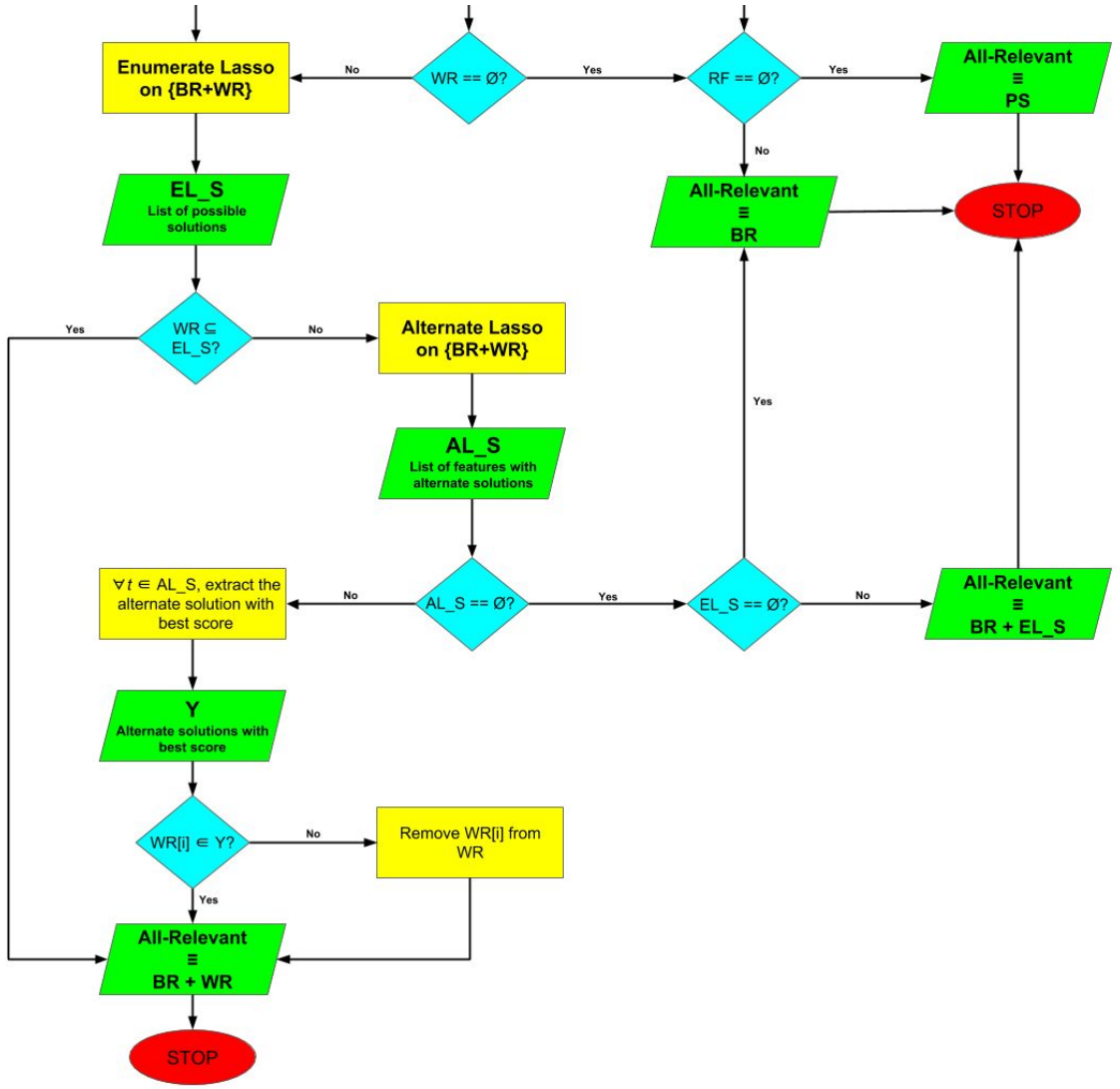
1. Apply the Shadow Feature Selection (**SFS method**) and produce the following items:
   - ➢ **SF**={$x\_s_1 \ldots x\_s_N$}, the list of shadow features, obtained by randomly shuffling the values of real features;
   - ➢ **IMP[PS, SF]** for each x ϵ PS & for each x_s ϵ SF, the importance list of all 2N features, original and shadows;
   - ➢ **st**: noise threshold, defined as the **max{IMP[SF],** for each **x_s ϵ SF}**;
   - ➢ **BR**={x ϵ PS t.c. IMP[x] ≥ st}, the set of best relevant real features;
   - ➢ **RF**={x ϵ PS, rejected by the Shadow Feature Selection}, the set of excluded real features, i.e. not relevant;
   - ➢ **WR**={x ϵ PS t.c. IMP[x] < st}, the set of weak relevant real features;

2. From the previous step, it resulted that PS ≡ {BR+WR+RF}. Now we consider the **PS_{red}= {BR+WR}**, by excluding the rejected features. In principle it may correspond to the original PS, in case of no rejections from the SFS;
   a) If RF==ø && WR==ø, the SFS method confirmed all real features as high relevant, therefore return **ALL-RELEVANT(PS)**, i.e. the full PS, as the optimized parameter space and **EXIT**.
   b) If RF≠ø && WR==ø, the SFS method rejected some features and confirmed others as high relevant, therefore return **ALL-RELEVANT(BR)** as the optimized parameter space and **EXIT**.
   c) If WR≠ø, regardless some rejections, SFS confirmed the presence of some weak relevant features that must be evaluated by LASSO methods, therefore goto 3;

ΦLAB voting algorithm

3. Given PS_{red}= {BR+WR}, the set of candidate features, apply **E-LASSO method**. It produces:
   - ➢ **EL_S**, a list of M subsets of features, considered as possible solutions, ordered by decreasing score;
   a) If WR ⊆ EL_S, then all weak relevant features are possible solutions, therefore return **ALL-RELEVANT(BR+WR)** as the optimized parameter space and **EXIT.**
   b) Else goto 4;
4. Given PS_{red}= {BR+WR}, the set of candidate features, apply **A-LASSO method**. It produces:
   - ➢ **AL_S**, a set of T features, each one with a list of features List(t) considered as alternate solutions with a certain score;
   a) if AL_S ==ø then no alternate solutions exist, therefore:
      i. If EL_S==ø then return **ALL-RELEVANT(BR)** as the optimized parameter space and **EXIT.**
      ii. Else if EL_S≠ø then return **ALL-RELEVANT(BR+EL_S)** as the optimized parameter space and **EXIT.**
   b) Else extract for each t ϵ T the alternate solution xas, t.c. *Score(xas) = min{Score(y), ∀ y ϵ List(t)};*
   c) goto 5.
5. For each x ϵ WR:
   a) If x is alternate solution of at least one feature t ϵ T, t.c. [t ϵ BR || t ϵ EL_S], then retain x within WR set;
   b) Else reject x (by removing x from WR);
6. Return **ALL-RELEVANT(BR+WR)** as the final optimized parameter space and **EXIT.**





ΦLAB voting algorithm

ΦLAB voting algorithm