

**Machine learning python code that  
replicates this paper**

---

## The problem :

Predicting synthetic lethality pairwise interaction upon knowledge of protein domains of the pair

gene A



gene B



Are gene A and B, SL?

# **What do we need first?**

# DATA!!!

## Data

→ known SL pairs



→ known non-SL pairs



→ known all prot. domains



- All the current knowledge on yeast genetic interaction is in **BioGrid**

- All the current knowledge on yeast genetic interaction is in **BioGrid**
- For protein domains you can go to **Yeastmine**

- All the current knowledge on yeast genetic interaction is in **BioGrid**
- For protein domains you can go to **Yeastmine**
  - To retrieve data using Python you should create an account in yeastmine and copy the python code generate by your search.

## What do we need 2nd?

- if we have data

## What do we need 2nd?

- if we have data
- we have information available about our problem,

## What do we need 2nd?

- if we have data
- we have information available about our problem,
- that we can use to learn !!! (not us but an algorithm :D)

## What do we need 2nd?

- if we have data
- we have information available about our problem,
- that we can use to learn !!! (not us but an algorithm :D)
- The question , what do we have to supply to the algorithm in order to learn and do the job for us?

## Features of the paper

### Data

→ Known SL pairs



→ Known non-SL pairs



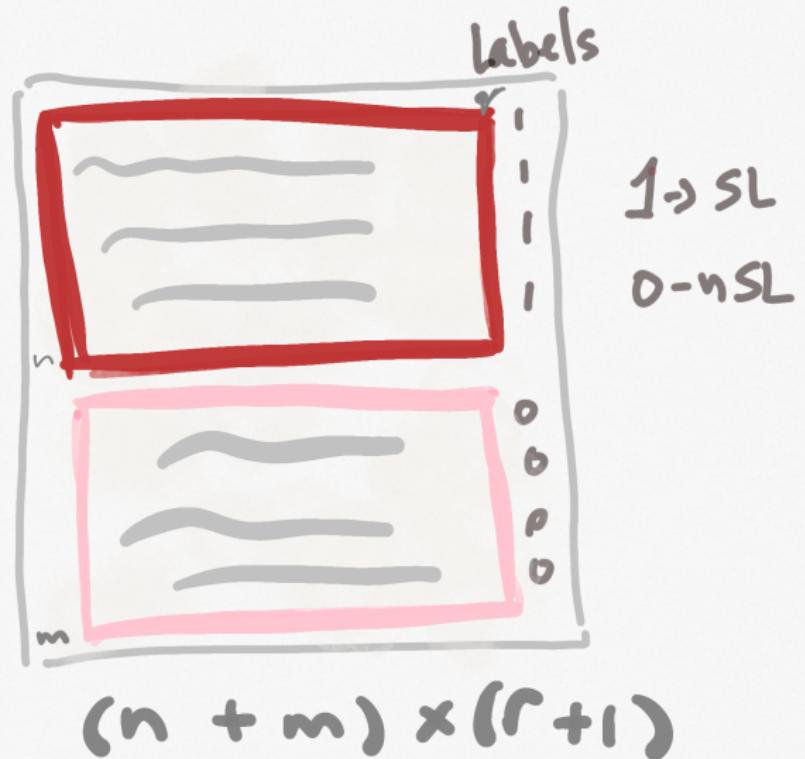
→ Known all prot. domains

## Feature matrix

$d_0$	$d_1$	$d_2$	$\dots$	$d_{k-1}$	$d_k$	$d_{k+1}$	$\dots$	$d_r$	
0	1	1	0	1	0	2	0	0	Pair 1
2	0	0	0	2	1	1	1	1	Pair 2
0	1	0	0	0	2	0	1	1	Pair 3
0	1	0	!	0	1	0	0	0	⋮
0	2	0	1	0	0	1	0	0	⋮
1	0	1	;	2	0	2	0	0	Pair n

$n \times r$  feature matrix





# Everything starts! :)

- Splitting the data for training and testing

# Everything starts! :)

- Splitting the data for training and testing
- Train a classifier with the training data

# Everything starts! :)

- Splitting the data for training and testing
- Train a classifier with the training data
- Make some predictions

# Everything starts! :)

- Splitting the data for training and testing
- Train a classifier with the training data
- Make some predictions
- Evaluate the predictions based on the testing data.

## One form of evaluating the methods

