Genobolitics

Abdurrahman Aboudakila
Ammar Raşid

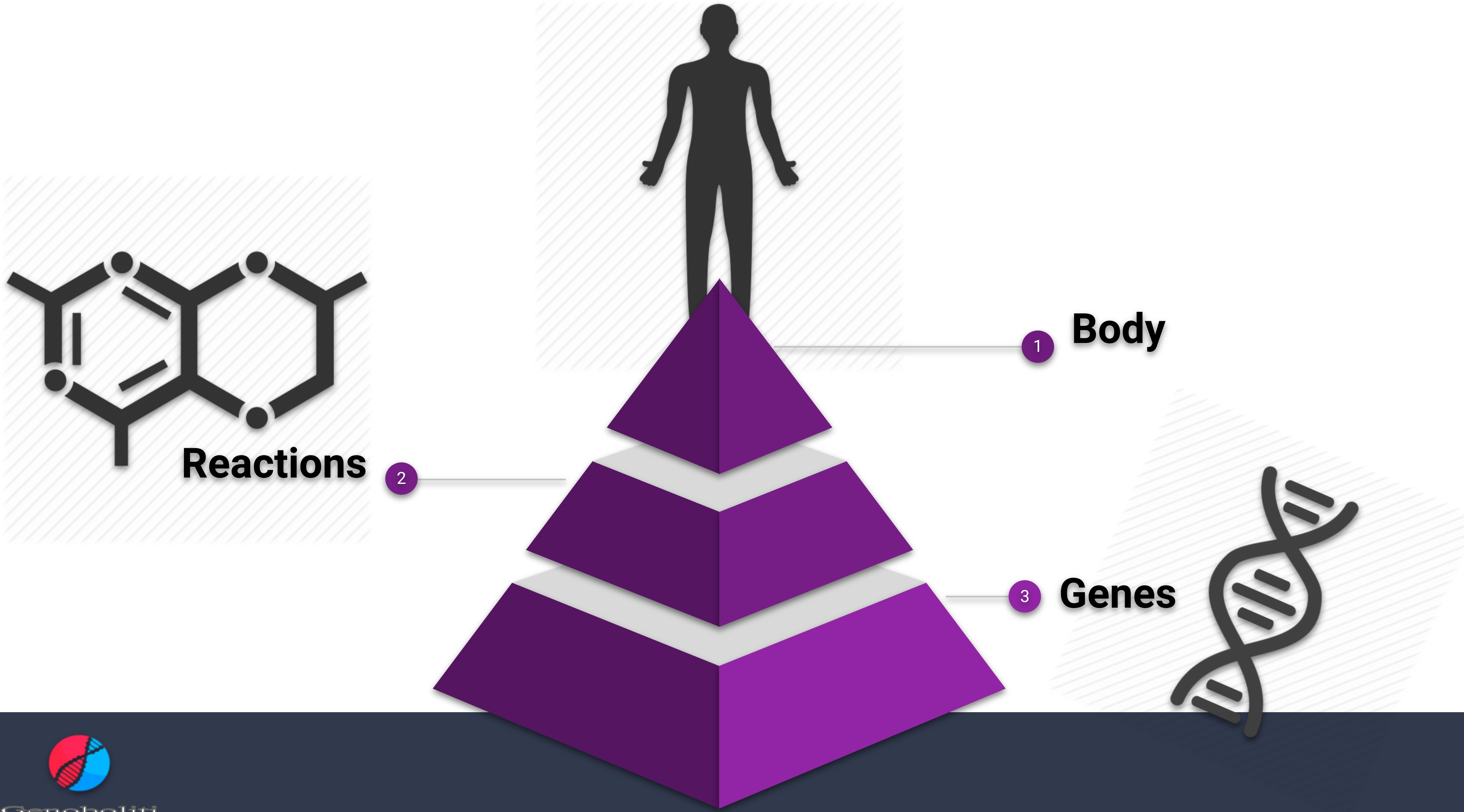
Outline

1. Introduction

2. Problem definition

3. Novelty and Methods

4. Results and Discussion



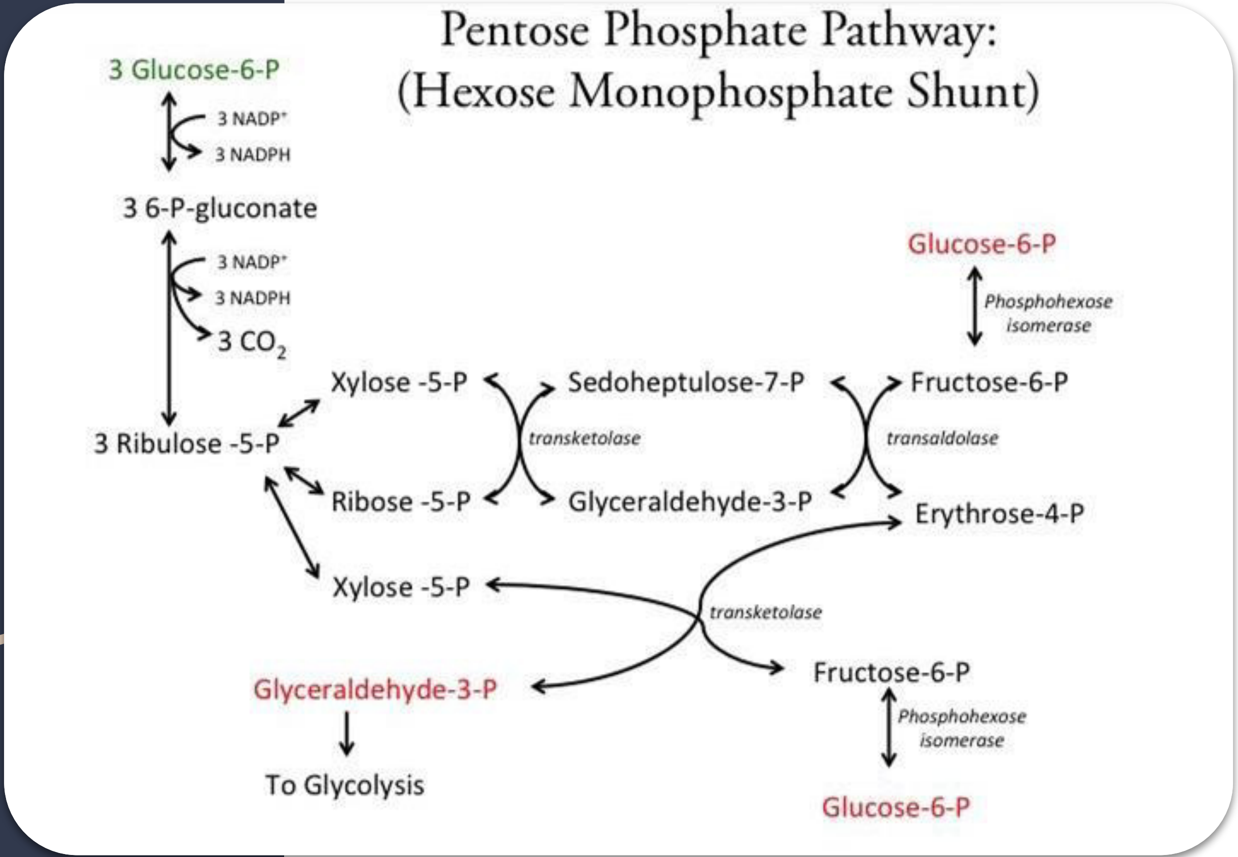
1 Body

2 Reactions

3 Genes

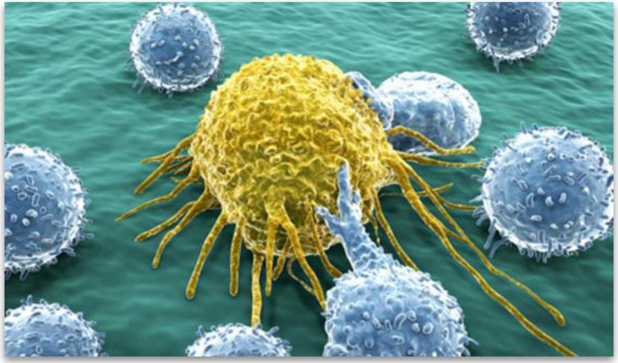
Genoboliti
cs

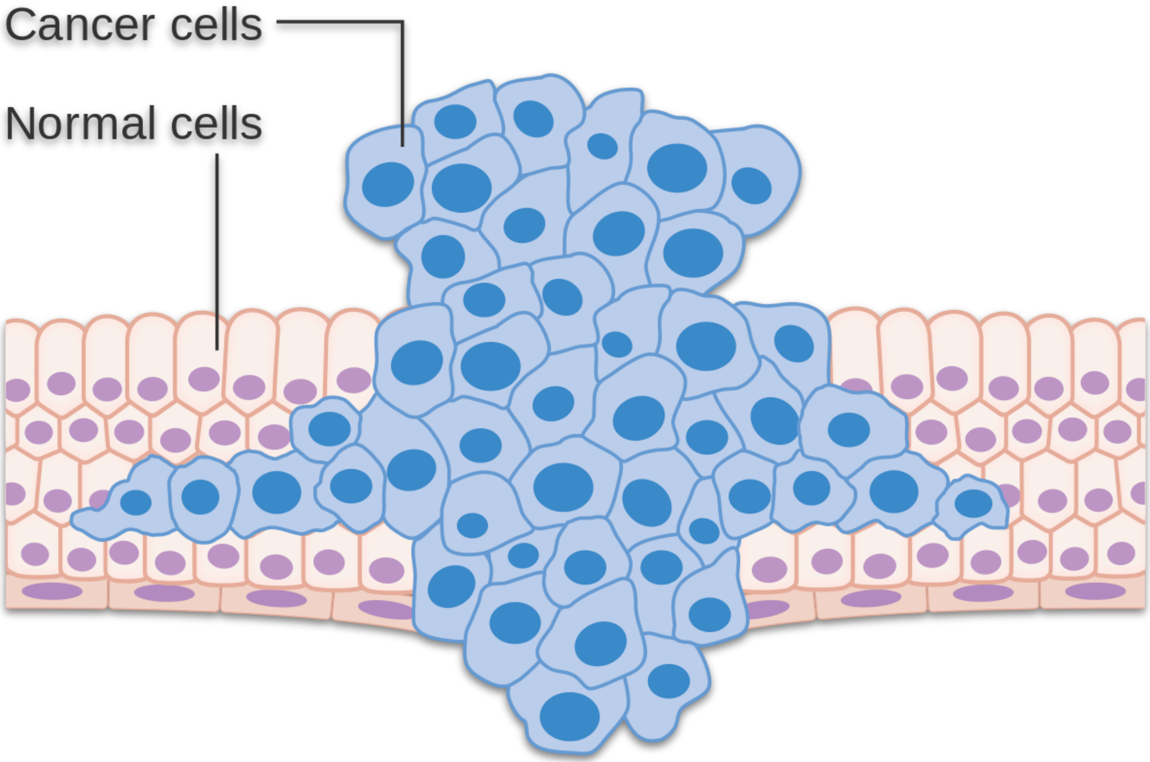
Pathways



Pentose Phosphate Pathway:
(Hexose Monophosphate Shunt)

Cancer

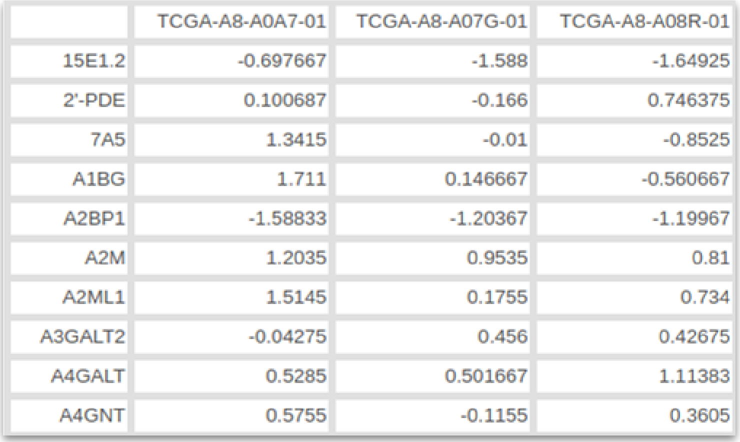




Problem Definition

Inputs

Gene Expression Arrays of healthy and patient samples of various diseases



	TCGA-A8-A0A7-01	TCGA-A8-A07G-01	TCGA-A8-A08R-01
15E1.2	-0.697667	-1.588	-1.64925
2'-PDE	0.100687	-0.166	0.746375
7A5	1.3415	-0.01	-0.8525
A1BG	1.711	0.146667	-0.560667
A2BP1	-1.58833	-1.20367	-1.19967
A2M	1.2035	0.9535	0.81
A2ML1	1.5145	0.1755	0.734
A3GALT2	-0.04275	0.456	0.42675
A4GALT	0.5285	0.501667	1.11383
A4GNT	0.5755	-0.1155	0.3605

Goal

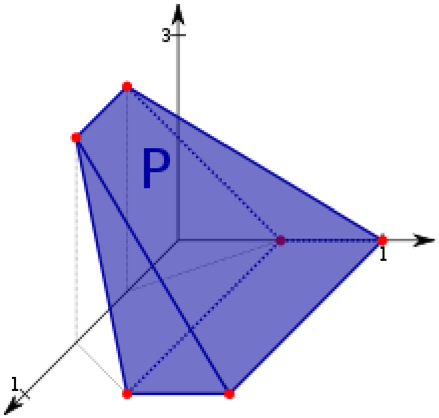
- Train a disease classifier
- Detect significant pathways in diseases
- Construct disease-ontology

Methods

- Dynamic Linear Programming
  - Flux Variability Analysis (FVA)
- Computing Diff values (Features)
- Machine Learning

Dynamic Linear Programming

1. Objective Coefficients = Fold Changes
2. Solve the objective function = C
3. Constrain ObjFun with C
4. Repeat for every reaction
5. Max+Min Reactions FCs



$$Objective\ Function = \sum_{m\,\in\,M} \sum_{R\,\in\,m_R} m_{fc} * R[m] \,/\, m_{TS}$$

Compute Diff Values

1. U_{ref} , L_{ref} = avg(U_i), avg(L_i) for i in healthy samples.

2. Diff = ((U -U_{ref}) + (L - L_{ref}))/2

Compute Diff Values

1. U_{ref} , L_{ref} = avg(U_i), avg(L_i) for *i* in healthy samples.

2. Diff = $((U - U_{ref}) + (L - L_{ref}))/2$

Disease Classification

1. Genobolitics Pipeline
   a. Nested Cross-Validation
   b. 10-Folds
   c. 10-Trials

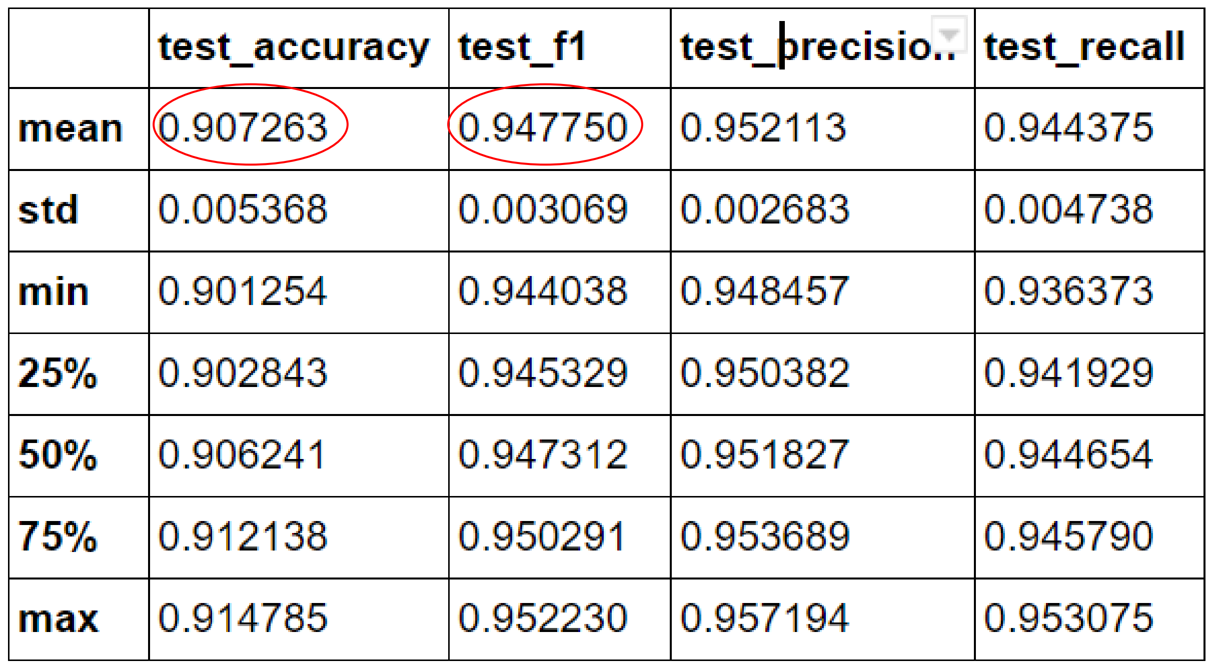
Dimensionality Reduction	PCA 21 components
Classifier	SVC C = 4.641

Breast-Cancer Case Study

UCSC Xena Agilent G4502A

64 Healthy

534 PT & Metastatic

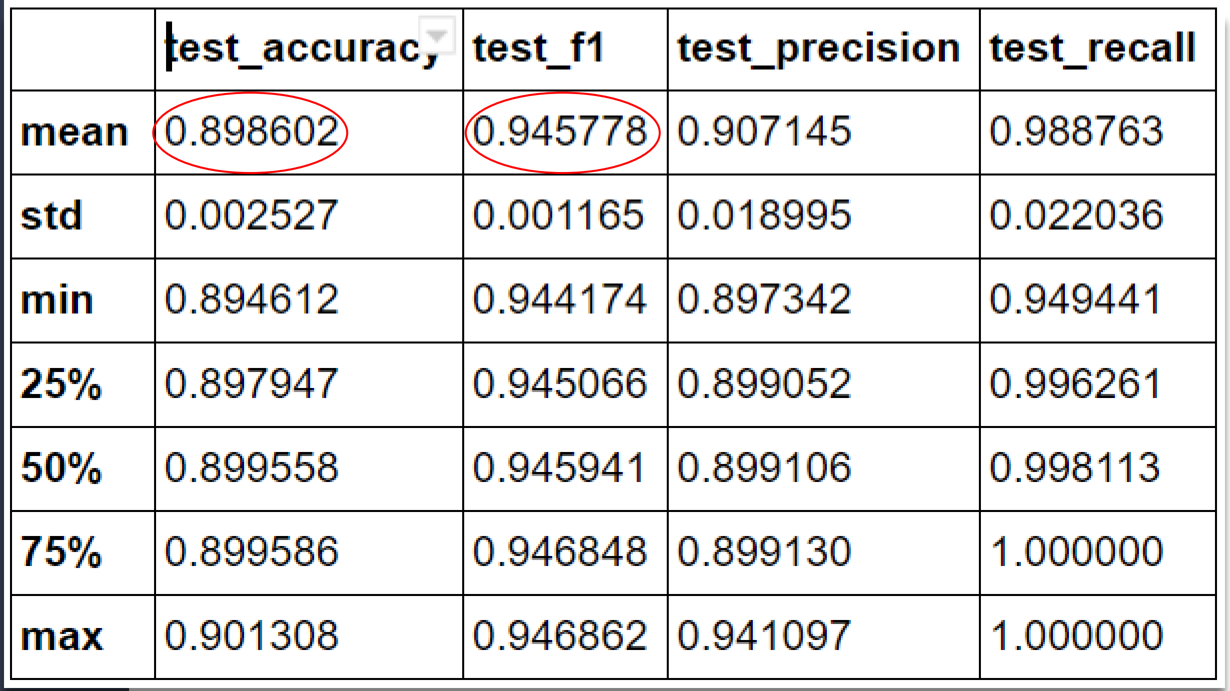


	test_accuracy	test_f1	test_precisio..	test_recall
mean	0.907263	0.947750	0.952113	0.944375
std	0.005368	0.003069	0.002683	0.004738
min	0.901254	0.944038	0.948457	0.936373
25%	0.902843	0.945329	0.950382	0.941929
50%	0.906241	0.947312	0.951827	0.944654
75%	0.912138	0.950291	0.953689	0.945790
max	0.914785	0.952230	0.957194	0.953075

Using Diff Values

Breast-Cancer Case Study

Genoboliti
cs



	test_accuracy	test_f1	test_precision	test_recall
mean	0.898602	0.945778	0.907145	0.988763
std	0.002527	0.001165	0.018995	0.022036
min	0.894612	0.944174	0.897342	0.949441
25%	0.897947	0.945066	0.899052	0.996261
50%	0.899558	0.945941	0.899106	0.998113
75%	0.899586	0.946848	0.899130	1.000000
max	0.901308	0.946862	0.941097	1.000000

Using No Diff Values

Breast-Cancer Case Study (cont.)

Most significant pathways (based on Raw-values):

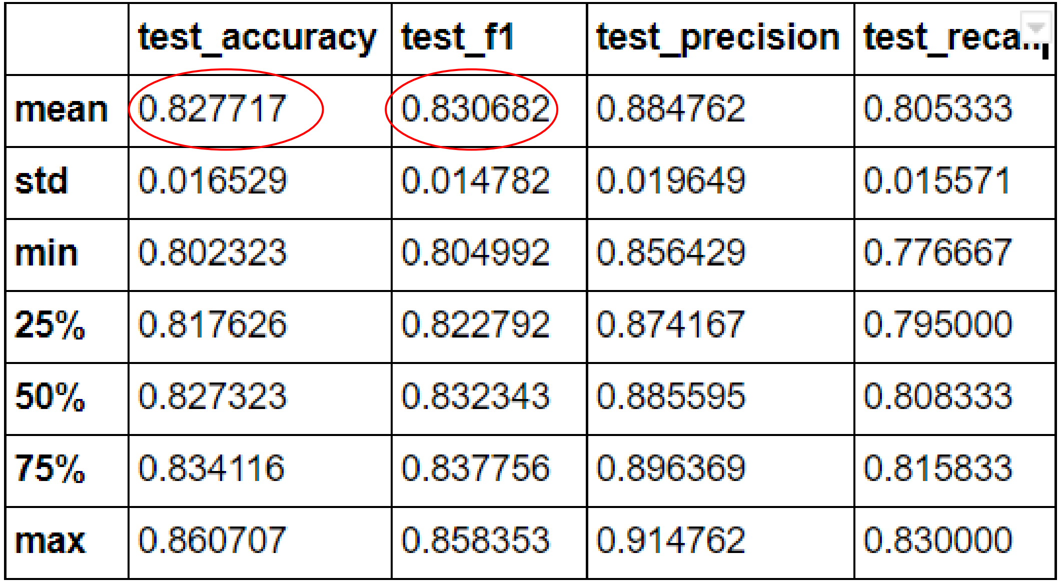
'Nucleotide interconversion',
'Purine synthesis',
'Pyrimidine synthesis',
'Vitamin A metabolism'

Genoboliti
cs

Lung-Cancer Case Study (GDS3257)

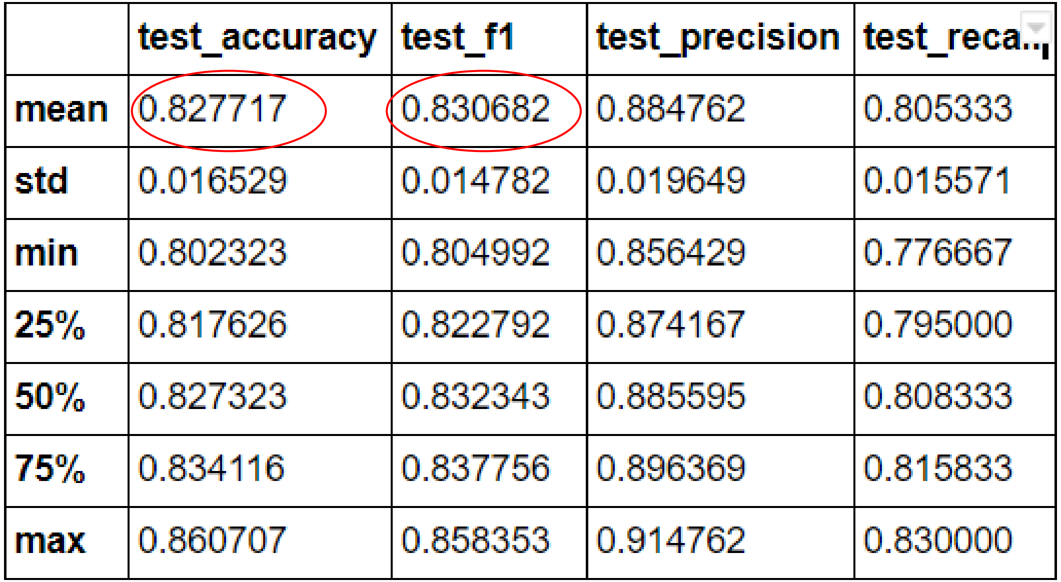
A lung cancer dataset of **107** samples, consisting of **49** healthy, **58** unhealthy samples was used.

Same procedure used in previous study was applied here.



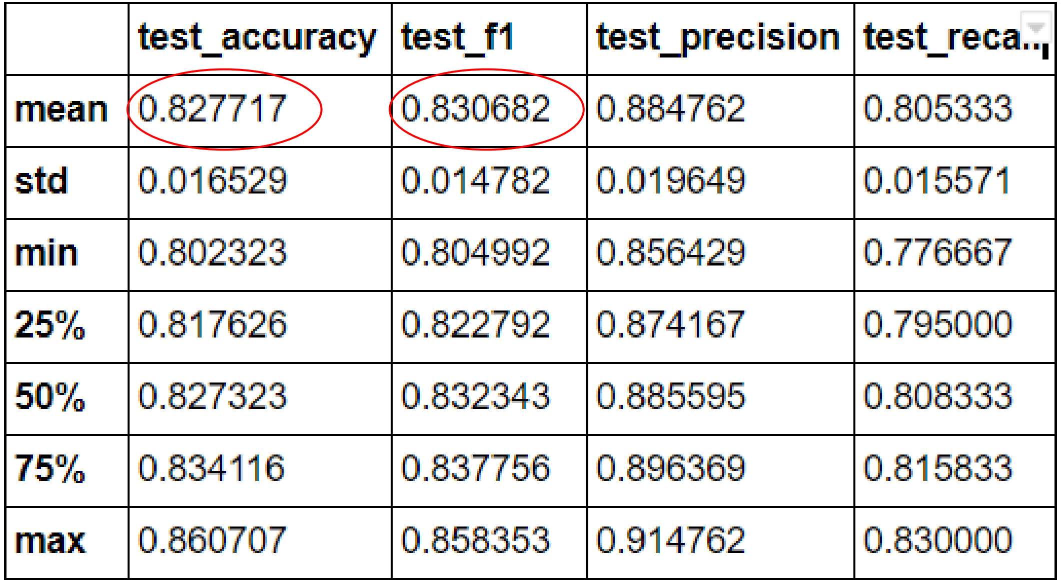
	test_accuracy	test_f1	test_precision	test_reca...
mean	0.827717	0.830682	0.884762	0.805333
std	0.016529	0.014782	0.019649	0.015571
min	0.802323	0.804992	0.856429	0.776667
25%	0.817626	0.822792	0.874167	0.795000
50%	0.827323	0.832343	0.885595	0.808333
75%	0.834116	0.837756	0.896369	0.815833
max	0.860707	0.858353	0.914762	0.830000

Lung-Cancer Case Study (GDS3257)



	test_accuracy	test_f1	test_precision	test_reca..
mean	0.827717	0.830682	0.884762	0.805333
std	0.016529	0.014782	0.019649	0.015571
min	0.802323	0.804992	0.856429	0.776667
25%	0.817626	0.822792	0.874167	0.795000
50%	0.827323	0.832343	0.885595	0.808333
75%	0.834116	0.837756	0.896369	0.815833
max	0.860707	0.858353	0.914762	0.830000

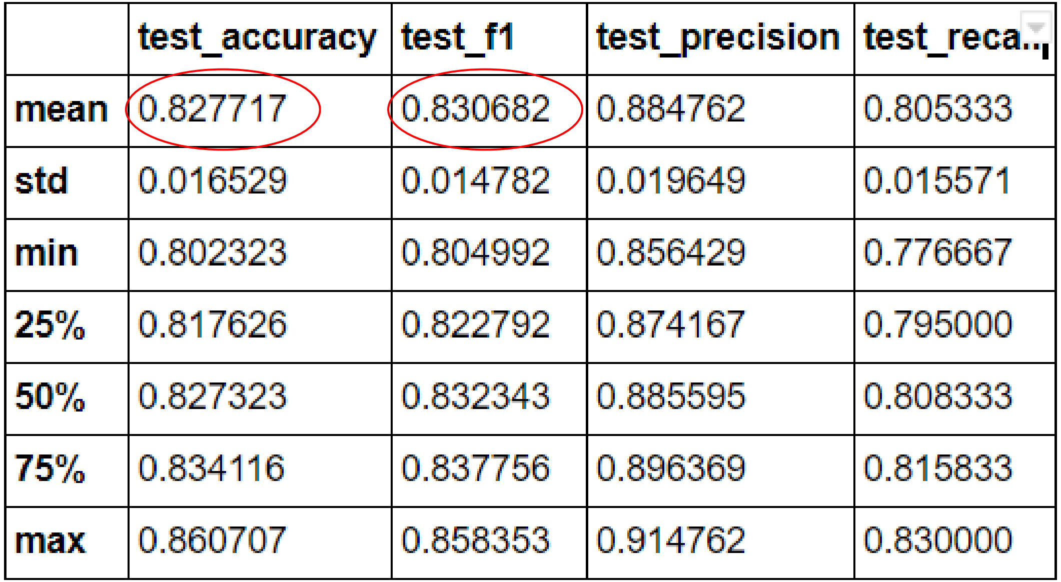
Lung-Cancer Case Study (GDS3257)



	test_accuracy	test_f1	test_precision	test_reca..
mean	0.827717	0.830682	0.884762	0.805333
std	0.016529	0.014782	0.019649	0.015571
min	0.802323	0.804992	0.856429	0.776667
25%	0.817626	0.822792	0.874167	0.795000
50%	0.827323	0.832343	0.885595	0.808333
75%	0.834116	0.837756	0.896369	0.815833
max	0.860707	0.858353	0.914762	0.830000

f1 score, precision and recall reported across 10-trials of nested cross-validation (Diff. scores)

Lung-Cancer Case Study (GDS3257)



	test_accuracy	test_f1	test_precision	test_reca...
mean	0.827717	0.830682	0.884762	0.805333
std	0.016529	0.014782	0.019649	0.015571
min	0.802323	0.804992	0.856429	0.776667
25%	0.817626	0.822792	0.874167	0.795000
50%	0.827323	0.832343	0.885595	0.808333
75%	0.834116	0.837756	0.896369	0.815833
max	0.860707	0.858353	0.914762	0.830000

f1 score, precision and recall reported across 10-trials of nested cross-validation (Diff. scores)

Lung-Cancer Case Study (GDS3257)

Most significant pathways:

'Blood group synthesis', 'Cholesterol metabolism', 'Eicosanoid metabolism', 'Fatty acid oxidation',

'Folate metabolism',
'Inositol phosphate metabolism',

'Nucleotide interconversion',
'Phosphatidylinositol phosphate metabolism',

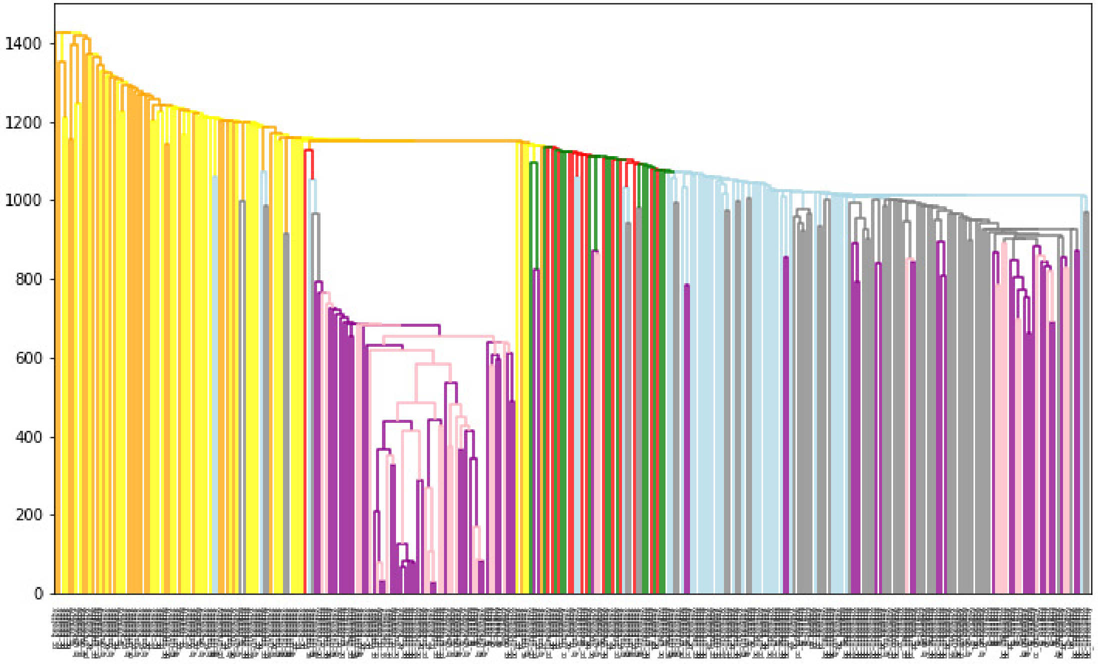
'Pyrimidine synthesis', 'Steroid metabolism',

'Thiamine metabolism', 'Urea cycle'

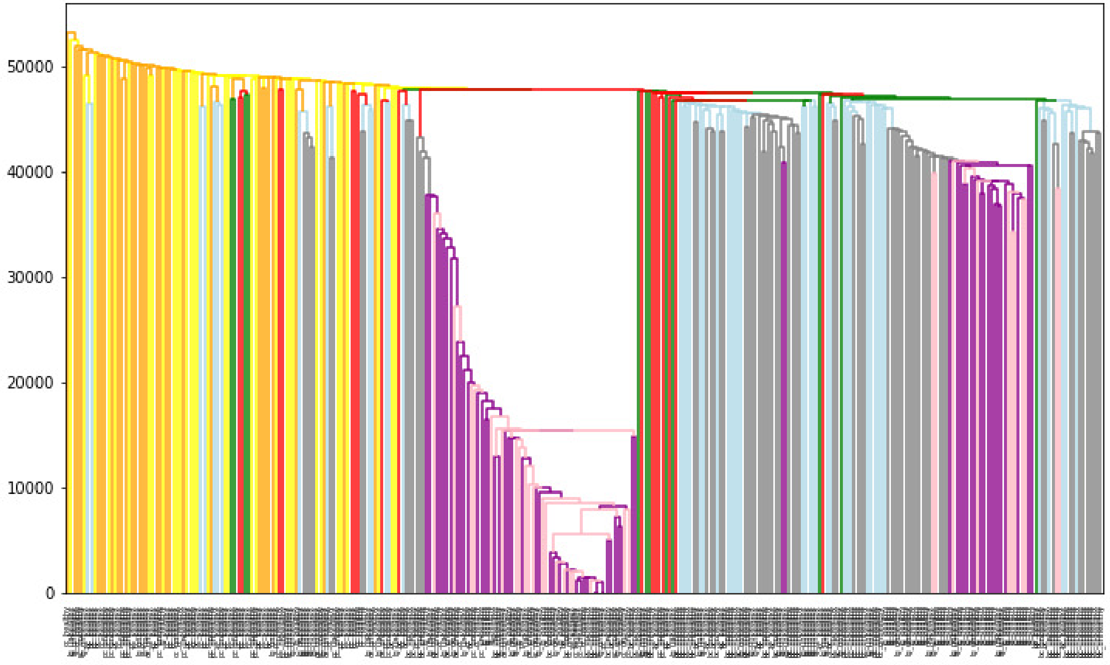
Clustering

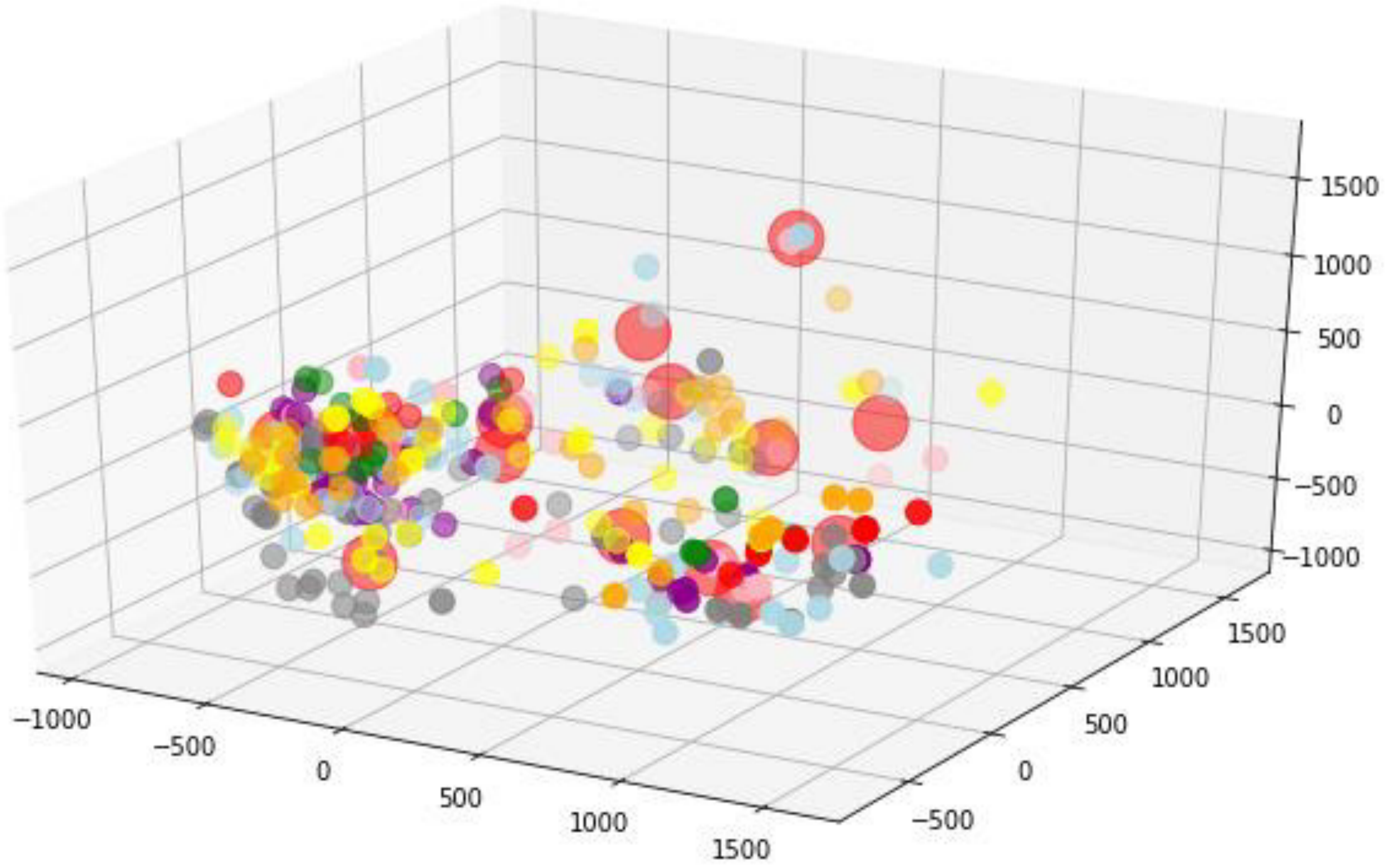
BC healthy	Light Pink
BC Patient	Dark magenta
CC Healthy	Green
CC Patient	Red
LG Healthy	Light Blue
LG Patient	Grey
PC Healthy	Yellow
PC Patient	Orange

Hierarchical Clustering (Diff)

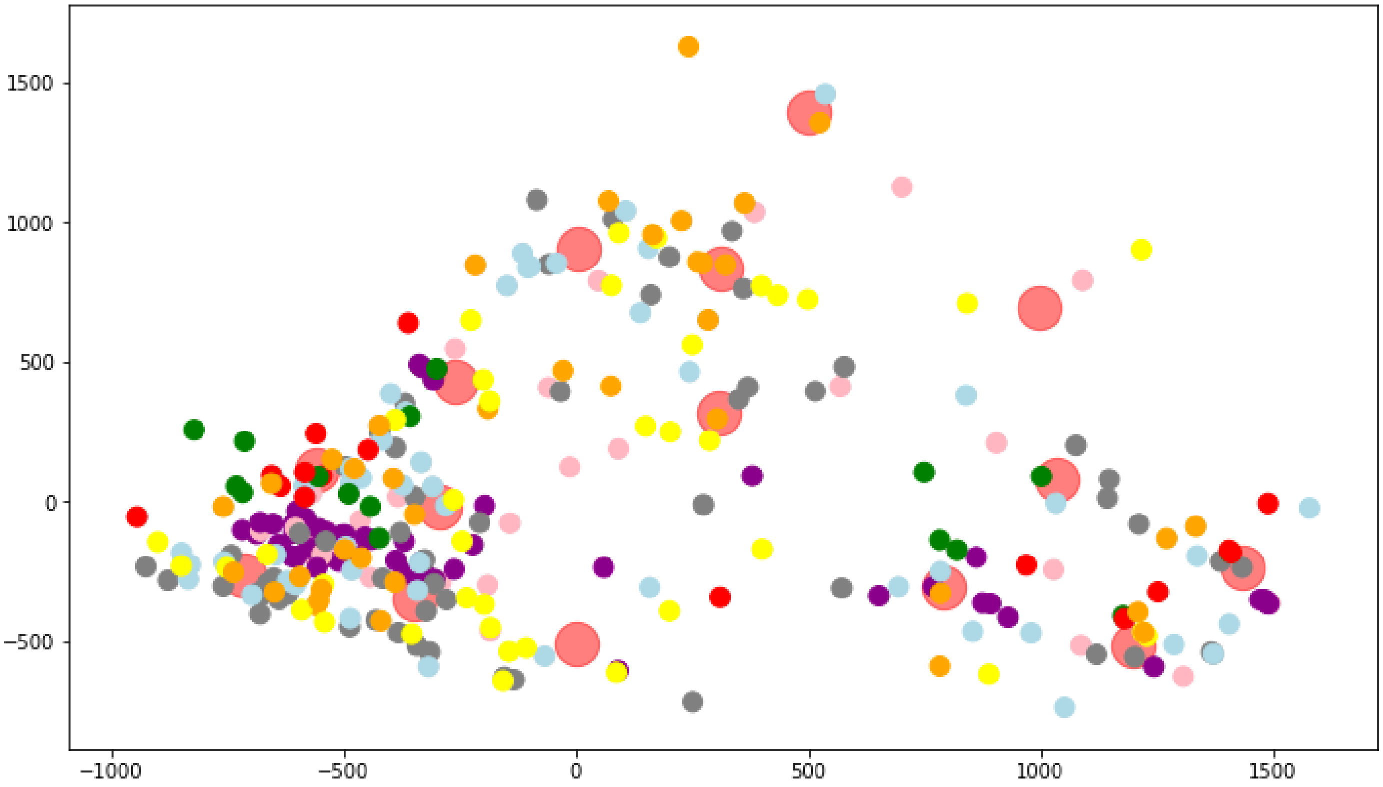


Hierarchical Clustering (Raw)



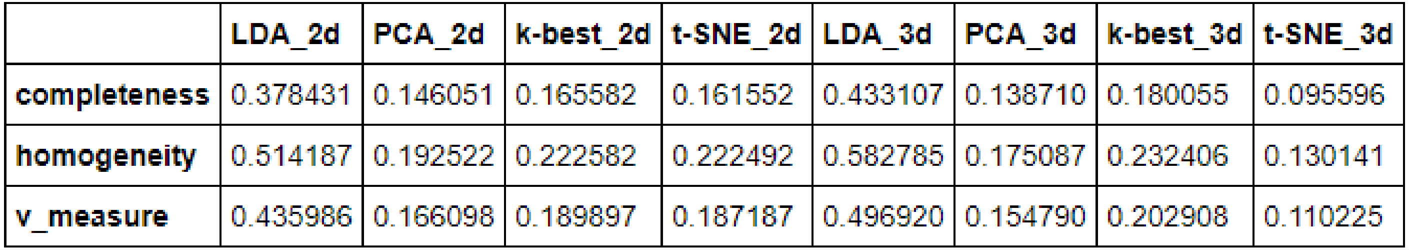


Clustering (PCA diff. score)

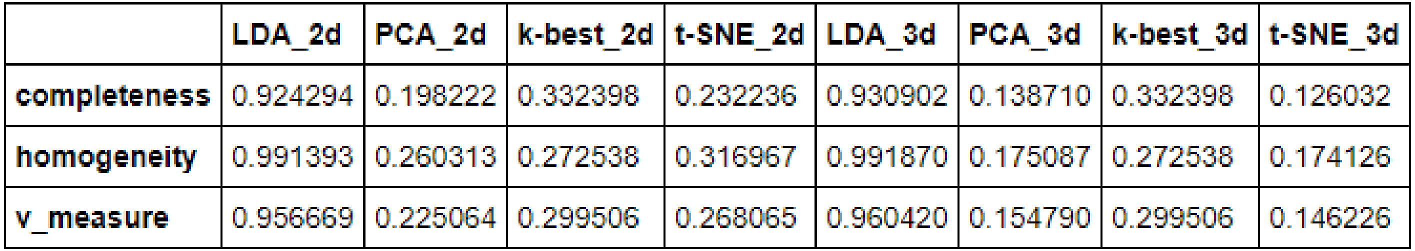


Clustering (PCA diff. score)

Clustering Metrics Results



	LDA_2d	PCA_2d	k-best_2d	t-SNE_2d	LDA_3d	PCA_3d	k-best_3d	t-SNE_3d
completeness	0.378431	0.146051	0.165582	0.161552	0.433107	0.138710	0.180055	0.095596
homogeneity	0.514187	0.192522	0.222582	0.222492	0.582785	0.175087	0.232406	0.130141
v_measure	0.435986	0.166098	0.189897	0.187187	0.496920	0.154790	0.202908	0.110225



	LDA_2d	PCA_2d	k-best_2d	t-SNE_2d	LDA_3d	PCA_3d	k-best_3d	t-SNE_3d
completeness	0.924294	0.198222	0.332398	0.232236	0.930902	0.138710	0.332398	0.126032
homogeneity	0.991393	0.260313	0.272538	0.316967	0.991870	0.175087	0.272538	0.174126
v_measure	0.956669	0.225064	0.299506	0.268065	0.960420	0.154790	0.299506	0.146226

Questions