

Strojno učenje & Presejanje velike količine podatkov za razvoj zdravil in odkrivanje novih funkcij

Sašo Džeroski

Jozef Stefan Institute, Ljubljana, Slovenia

Ekipa: Dragi Kocev,

Panče Panov, Tomaž

Stepišnik, Martin Breskvar

Interreg
ITALIA-SLOVENIJA



TRAIN



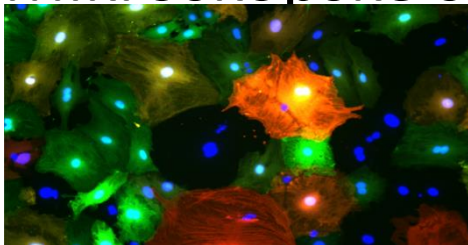
Progetto standard co-finanziato dal Fondo europeo di sviluppo regionale
Standardni projekt sofinancira Evropski sklad za regionalni razvoj



Presejanje velike količine podatkov

- Presejalni testi spojin
- Genomski presejalni testi (miRNA)

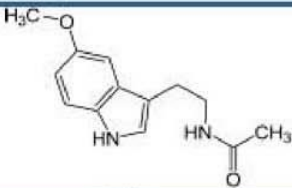
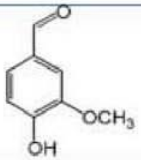
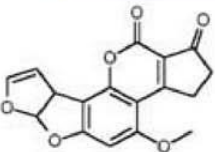
- Mikroskopske slike



- Uporaba strojnega učenja za pridobivanje modelov za virtualne presejalne teste in identifikacijo novih kandidatnih spojin

Učenje iz presejalnih testov spojin

Struktura učenja–
Odnosi aktivnosti

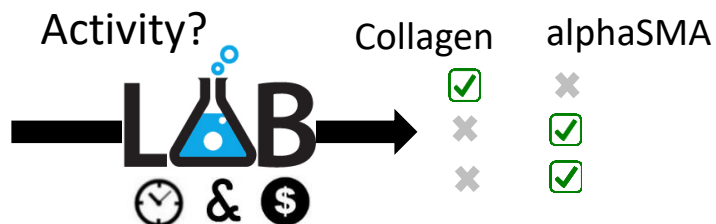
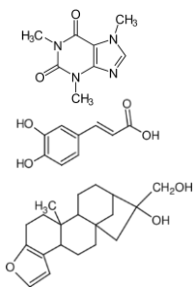
		input datatype specification	output datatype specification
		input: molecule datatype	output: real datatype
		compound	activity
data example			0.25
			0.28
			0.37



Učenje iz presejalnih testov spojin

Compound screen data

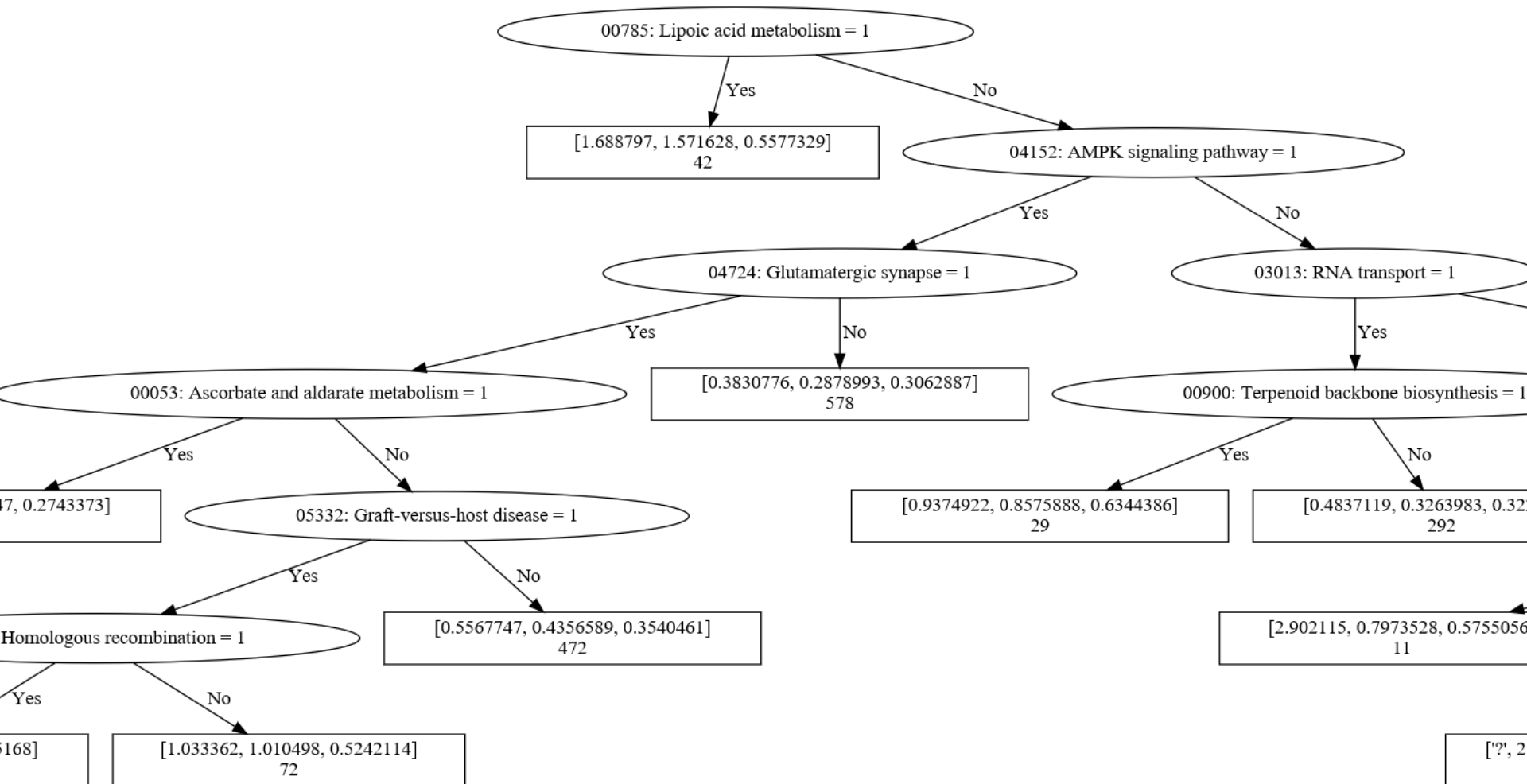
	Descriptive space				Target space	
Example 1	1	TRUE	0.49	0.69	0.68	3.91
Example 2	2	FALSE	0.08	0.07	0.56	7.59
Example 3	1	FALSE	0.08	0.07	0.10	7.57
Example 4	2	TRUE	0.49	0.69	0.08	8.86
...





Tehnologija ML

Več-tarčna regresijska drevesa





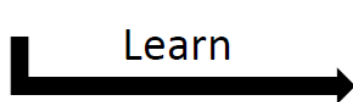
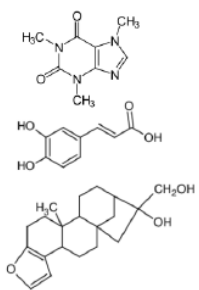
Virtualno presejanje

Labeled data

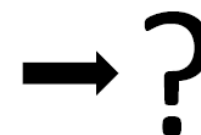
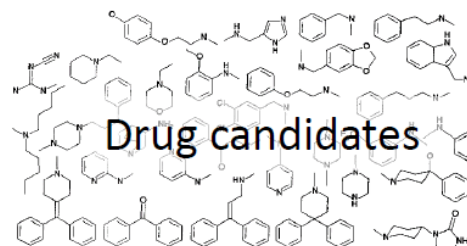
	Descriptive space				Target space	
Example 1	1	TRUE	0.49	0.69	0.68	3.91
Example 2	2	FALSE	0.08	0.07	0.56	7.59
Example 3	1	FALSE	0.08	0.07	0.10	7.57
Example 4	2	TRUE	0.49	0.69	0.08	8.86
...

Unlabeled data

Example N+1	1	TRUE	0.86	0.35	?	?
Example N+2	2	FALSE	0.09	0.05	?	?
Example N+3	4	FALSE	0.07	0.01	?	?
Example N+4	2	TRUE	0.91	0.78	?	?
Example N+5	2	TRUE	0.42	0.69	?	?
...



**VIRTUAL SCREENING
MODEL**



Predict



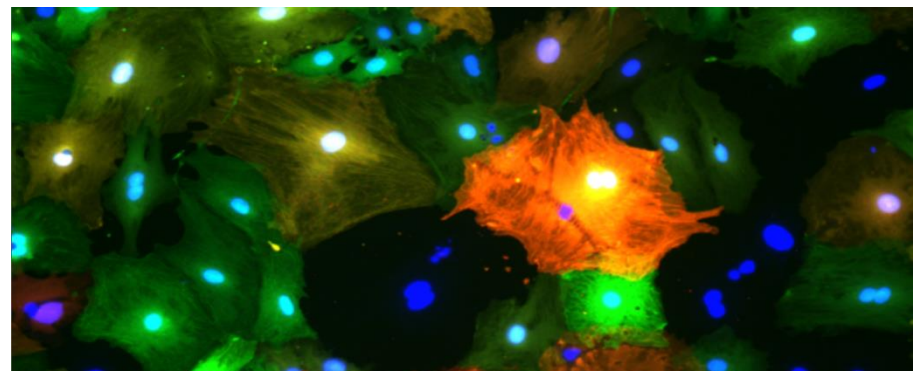
Analiziranje podatkov presejanja velike količine podatkov

- Spojine zaznamujejo njihovi prstni odtisi
- Generacija odprtokodnega kemoinformatske SW knjižnice RDkit
- prstni odtisi FCFP2 so bili uporabljeni (1024 značilnosti)
- Upoštevali smo tudi profile tarčnih proteinov
- Atributi
- Izvedeni testi so bili fotografirani pod mikroskopom
- Značilnosti so bile ekstrahirane iz slik
- Tarče



Zmanjšanje fibroze v miokardnem infarktu

- Presejanje velike količine podatkov z uporabo knjižnice 640, zdravila potrjena s strani FDA (ENZO)
- Identifikacija zdravil za zmanjšanje fibroze v miokardnem infarktu
- Test je uporabili mišje srčne fibroblaste, ki se diferencirajo v miofibroblaste v kulturi, ki izražajo alpha SMA-RFP in collagen-alpha1-EGFP
- Tarče: intenzivnost
 - alphaSMA
 - Collagen
- Atributi
 - Prstni odtisi



Testiranje predpostavk

- Nekaj domensko uporabljenih znanj / omejitev:
- Napovedane spojine filtrirajo za zdravila, odobrena s strani FDA, ki niso kortikosteroidi SMILE strings, ki se uporabljajo v Chemmine za identifikacijo snovi s strukturno podobnostjo nekomercialnim spojinam z visokimi predvidenimi vrednostmi
- Identificirane so tri povezane spojine, ki so opisane v literaturi, da imajo antifibrotični učinek
- Identificirane so štiri povezane spojine, za katere prej ni bilo opisano, da imajo antifibrotični učinek
- Testirano v mokrem laboratoriju in eden deluje zelo dobro



Izkušnje s projektom

- Analiza večih presejanj
 - spojin
 - miRNA
- Modele iz miRNA presejanj lahko uporabimo virtualno presejanje spojin
- Odlično sodelovanje z ICGEB
- Atraktivna VR predstavitev