Rudarjenje podatkov za vsakogar

Blaž Zupan
Univerza v Ljubljani

Znanost o podatkih

Linearna algebra

Verjetnost

Statistika

Optimizacijske metode

Predstavitev podatkov in modelov

Strojno učenje

Računski pristopi

Vizualizacija

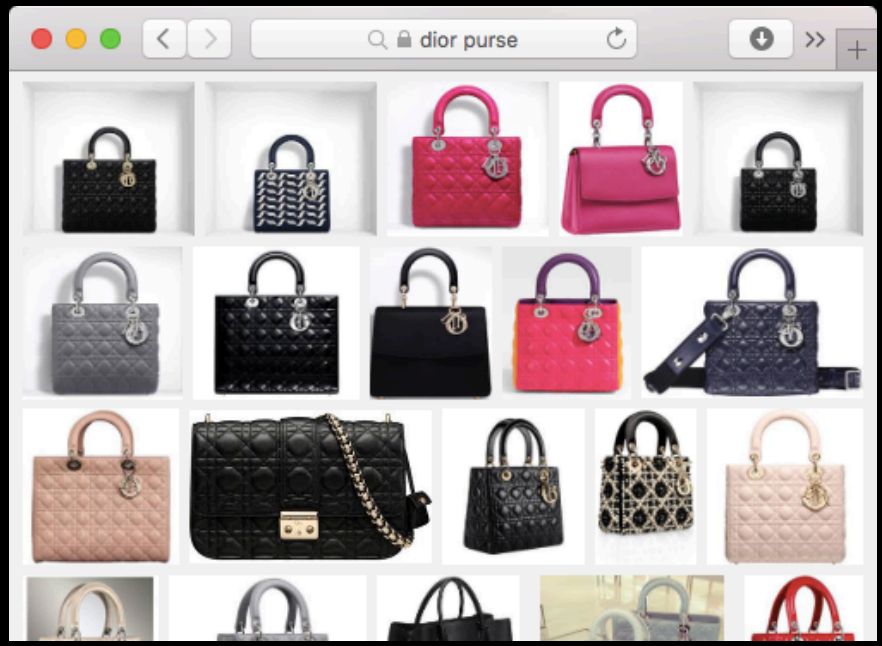
Programski jeziki in okolja

Silvana Žorž

```
def edges_from_osmdb(osmdb, vertex_namespace, slogs, profiledb=None):
    """generates (vertex1_label, vertex2_label, edgepayload) from osmdb"""

    street_id_counter = 0
    street_names = {}

    # for each edge in the osmdb
    for i, (id, parent_id, node1, node2, distance, geom, tags) in enumerate( (

        # Find rise/fall of edge, if profiledb is given
        rise=0
        fall=0
        if profiledb:
            profile = profiledb.get( id )
            if profile:
                rise, fall = get_rise_and_fall( profile )

        # insert end vertices of edge to graph
        vertex1_label = "%s-%s"%(vertex_namespace,node1)
        vertex2_label = "%s-%s"%(vertex_namespace,node2)

        # create ID for the way's street
        street_name = tags.get("name")
        if street_name is None:
            street_id_counter += 1
            street_id = street_id_counter
        else:
            if street_name not in street_names:
                street_id_counter += 1
                street_names[street_name] = street_id_counter
            street_id = street_names[street_name]

        # Create edges to be inserted into graph
        s1 = Street( id, distance, rise, fall )
        s2 = Street( id, distance, fall, rise, reverse_of_source=True )
        s1.way = street_id
        s2.way = street_id

        # See if the way's highway tag is penalized with a 'slog' value; if s(
        slog = slogs.get( tags.get("highway") )
        if slog:
            s1.slog = s2.slog = slog

        # Add the forward edge and the return edge if the edge is not oneway
    yield vertex1_label, vertex2_label, s1

        oneway = tags.get("oneway")
        if oneway != "true" and oneway != "yes":
        yield vertex2_label, vertex1_label, s2
```

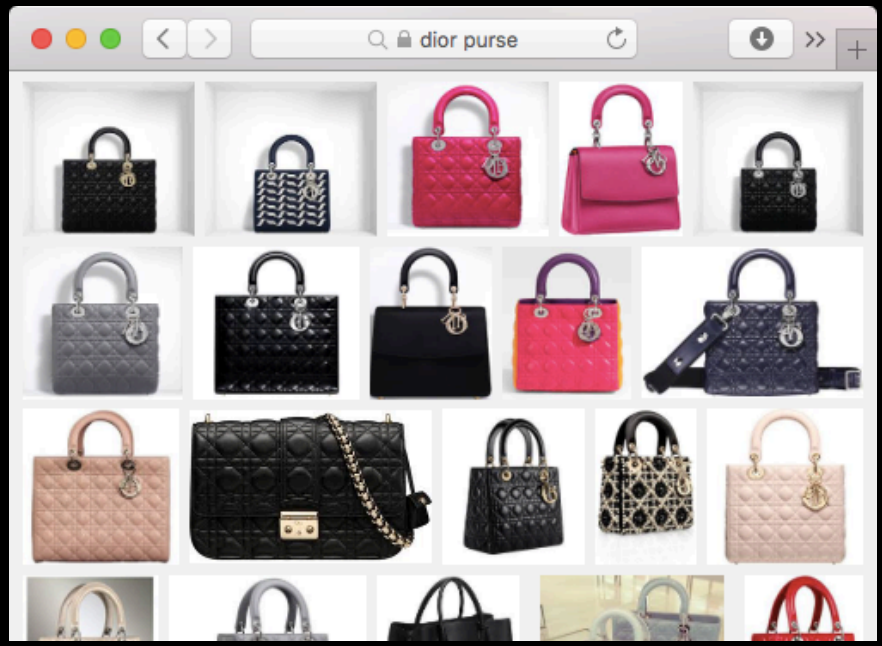

europeana.eu
europeana collections
MENU
THEMATIC COLLECTIONS
EUROPEANA MUSIC
THEMATIC COLLECTIONS
EUROPEANA ART
EXHIBITION
FACES OF EUROPE: ACADEMIC ART AND NEW DIRECTIONS


Dior

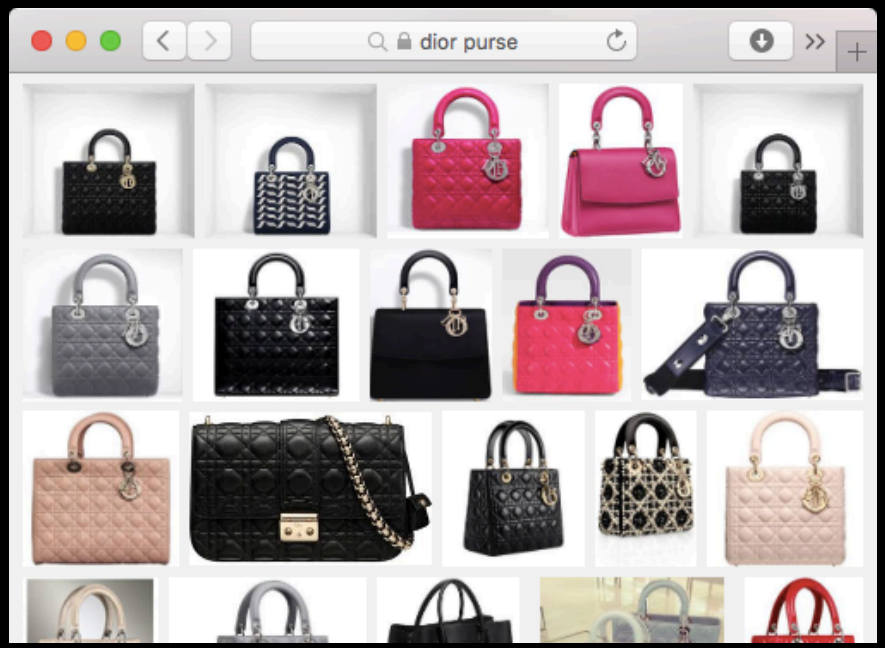

dior purse

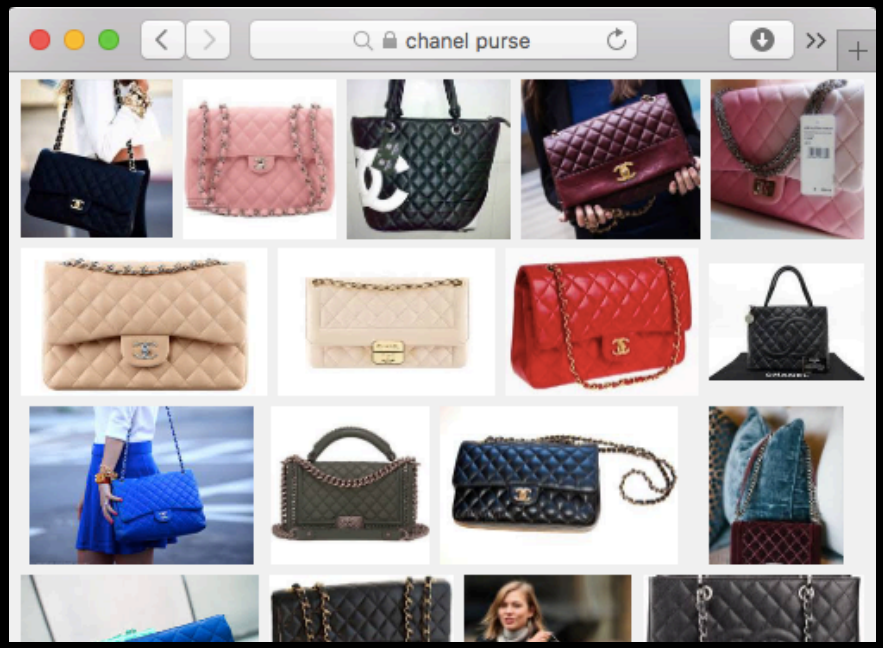

Dior


hermes purse
460 × 579 - br...

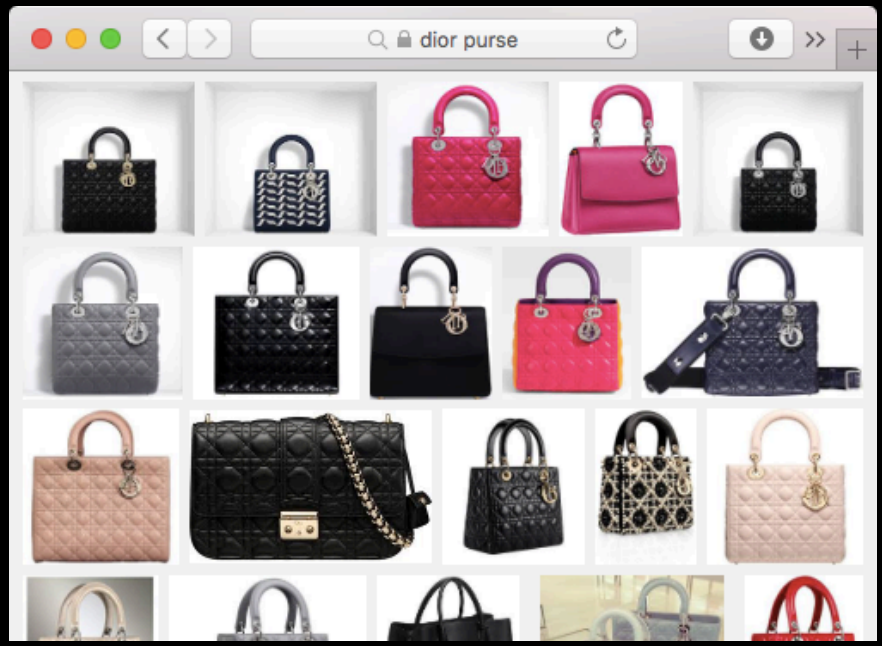

dior purse

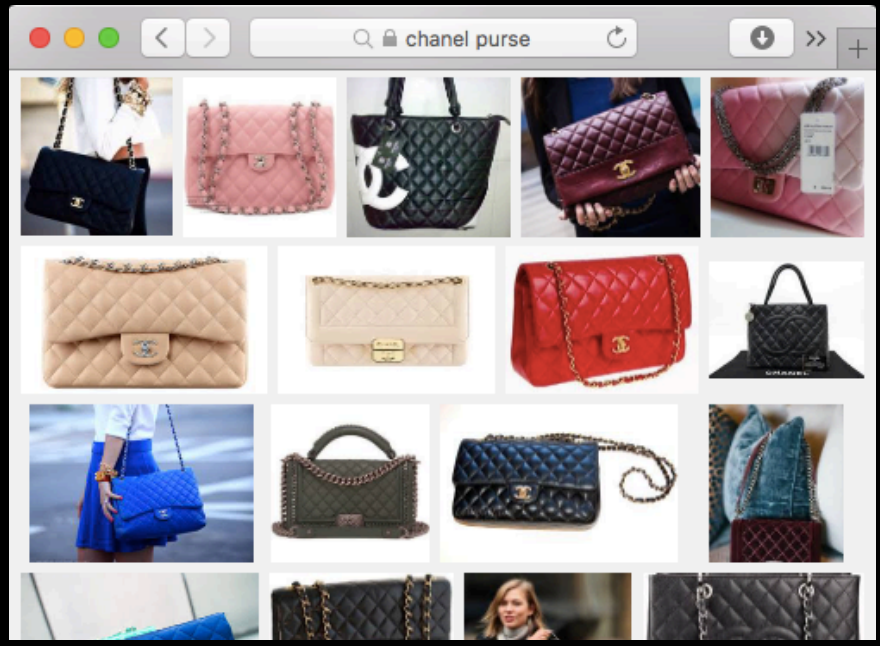

hermes purse
460 × 579 - br...

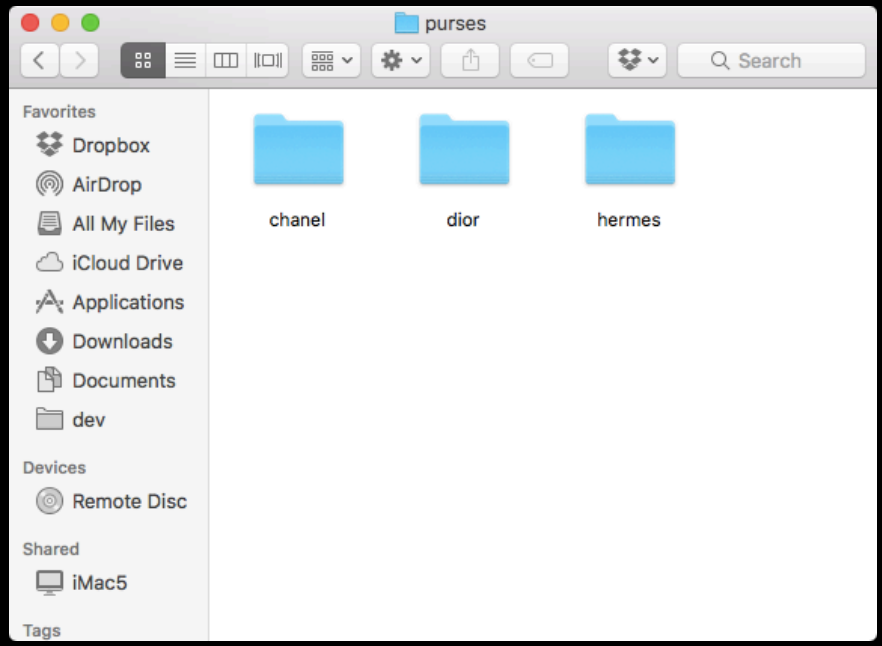

dior purse

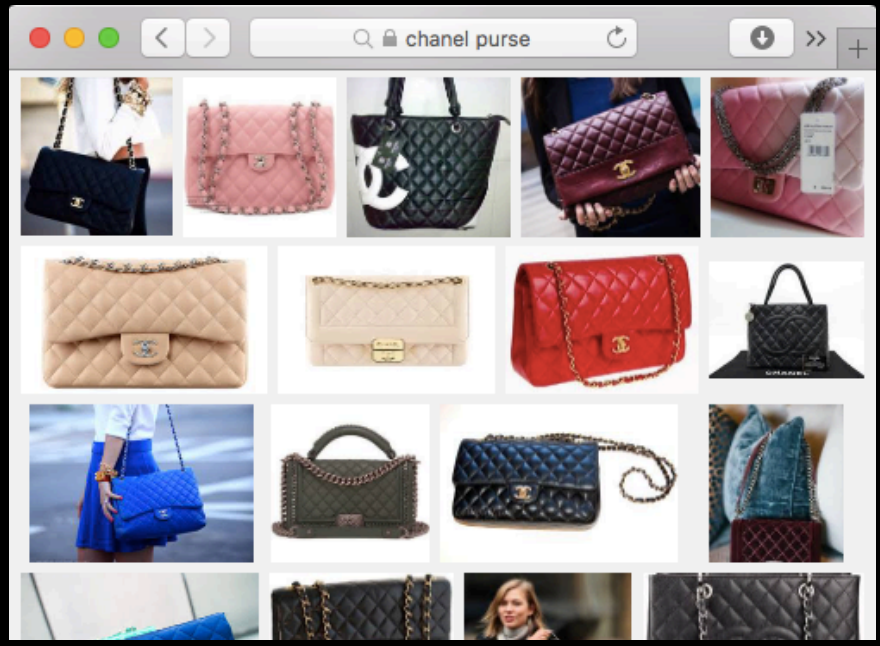

chanel purse

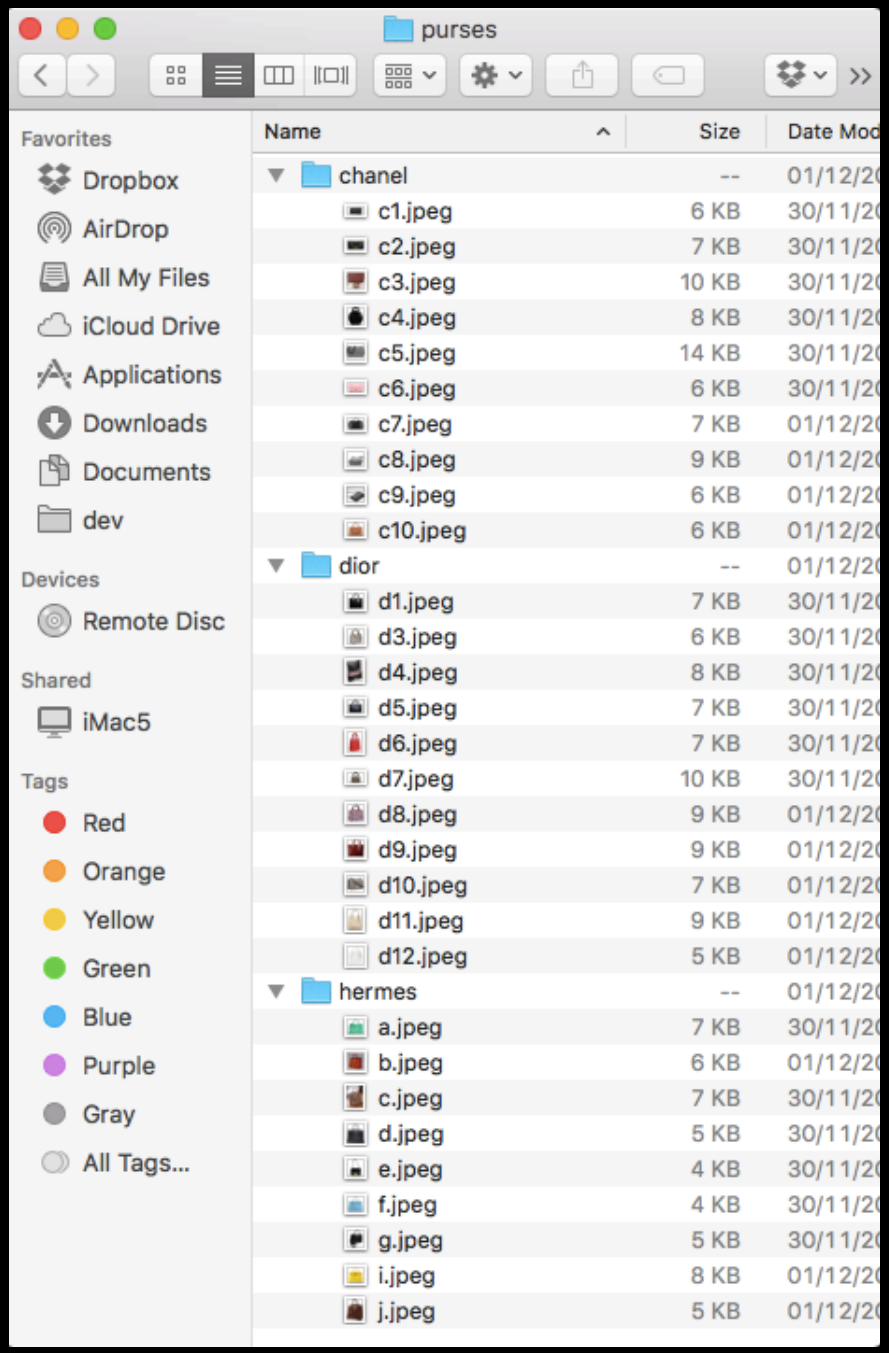
```
def edges_from_osmdb(osmdb, vertex_namespace, slogs, profiledb=None):
    """generates (vertex1_label, vertex2_label, edgepayload) from osmdb"""

    street_id_counter = 0
    street_names = {}

    # for each edge in the osmdb
    for i, (id, parent_id, node1, node2, distance, geom, tags) in enumerate( (

        # Find rise/fall of edge, if profiledb is given
        rise=0
        fall=0
        if profiledb:
            profile = profiledb.get( id )
            if profile:
                rise, fall = get_rise_and_fall( profile )

        # insert end vertices of edge to graph
        vertex1_label = "%s-%s"%(vertex_namespace,node1)
        vertex2_label = "%s-%s"%(vertex_namespace,node2)

        # create ID for the way's street
        street_name = tags.get("name")
        if street_name is None:
            street_id_counter += 1
            street_id = street_id_counter
        else:
            if street_name not in street_names:
                street_id_counter += 1
                street_names[street_name] = street_id_counter
            street_id = street_names[street_name]

        # Create edges to be inserted into graph
        s1 = Street( id, distance, rise, fall )
        s2 = Street( id, distance, fall, rise, reverse_of_source=True )
        s1.way = street_id
        s2.way = street_id

        # See if the way's highway tag is penalized with a 'slog' value; if s(
        slog = slogs.get( tags.get("highway") )
        if slog:
            s1.slog = s2.slog = slog

        # Add the forward edge and the return edge if the edge is not oneway
    yield vertex1_label, vertex2_label, s1

        oneway = tags.get("oneway")
        if oneway != "true" and oneway != "yes":
        yield vertex2_label, vertex1_label, s2
```

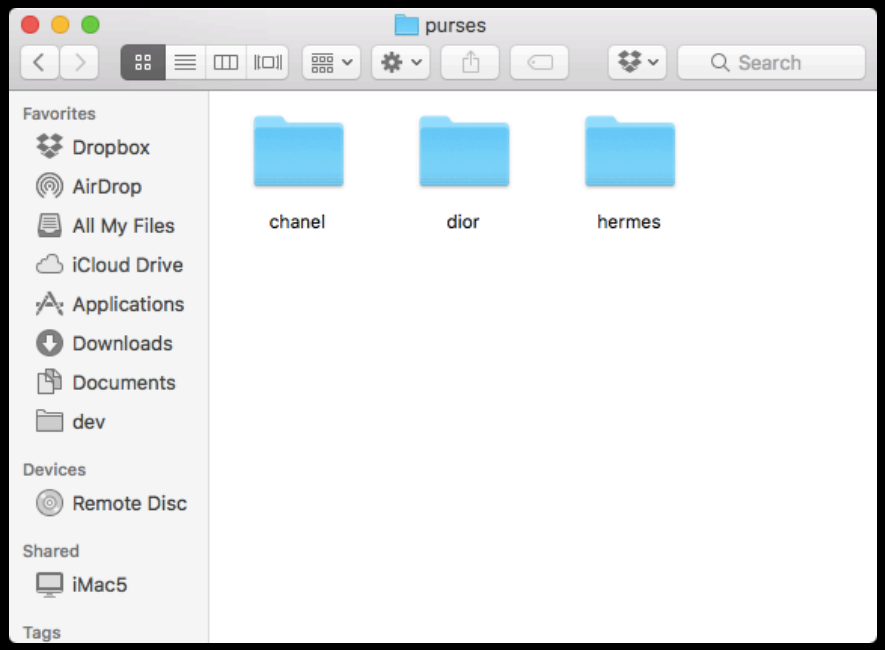


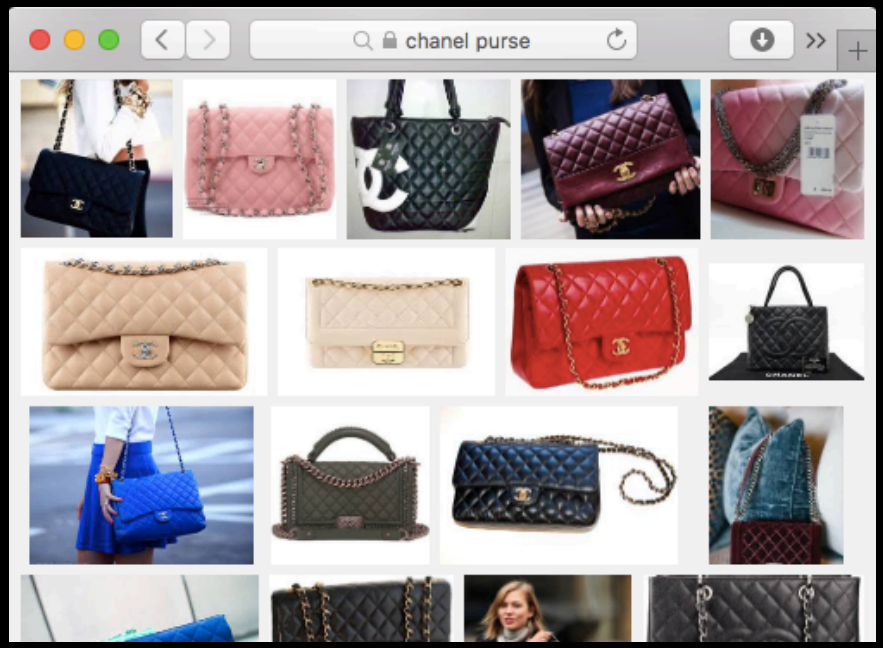



purses
Search
Favorites
Dropbox
AirDrop
All My Files
iCloud Drive
Applications
Downloads
Documents
dev
Devices
Remote Disc
Shared
iMac5
Tags
chanel
dior
hermes

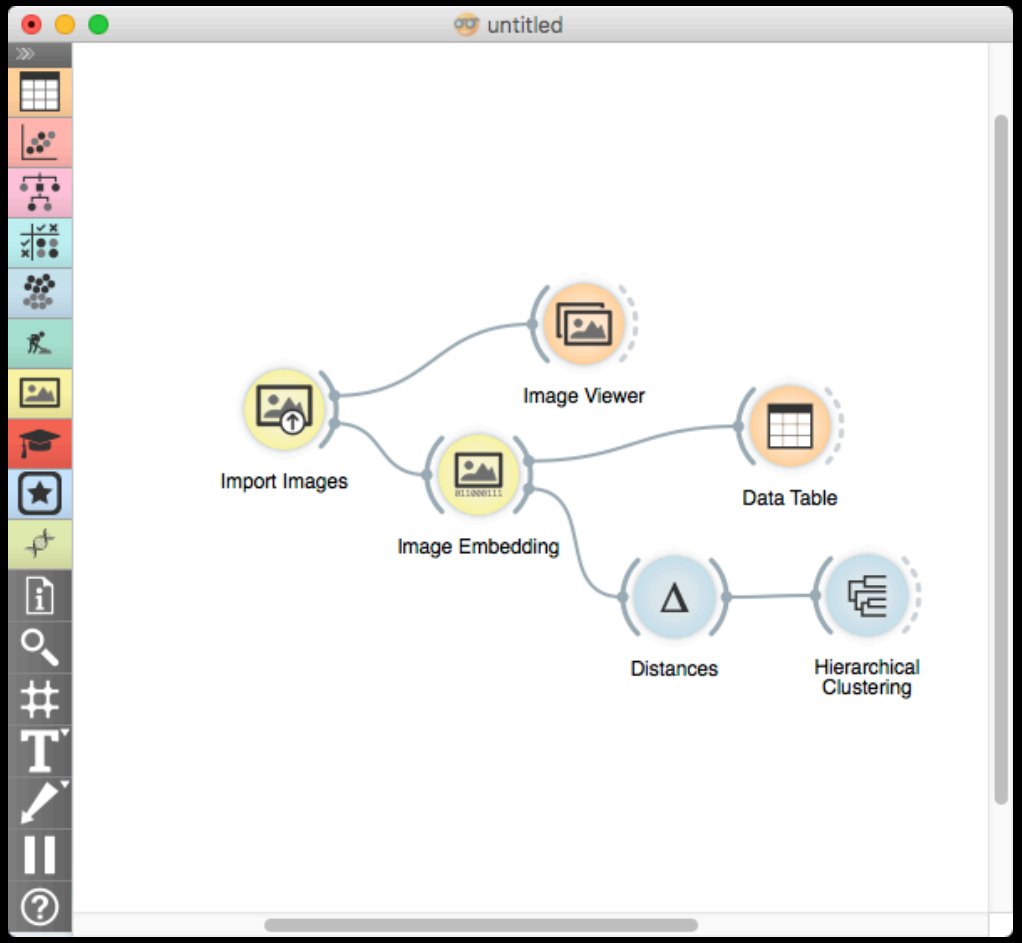

chanel purse


purses
Favorites
Dropbox
AirDrop
All My Files
iCloud Drive
Applications
Downloads
Documents
dev
Devices
Remote Disc
Shared
iMac5
Tags
Red
Orange
Yellow
Green
Blue
Purple
Gray
All Tags...
Name
Size
Date Mod
chanel -- 01/12/2
c1.jpeg 6 KB 30/11/2
c2.jpeg 7 KB 30/11/2
c3.jpeg 10 KB 30/11/2
c4.jpeg 8 KB 30/11/2
c5.jpeg 14 KB 30/11/2
c6.jpeg 6 KB 30/11/2
c7.jpeg 7 KB 01/12/2
c8.jpeg 9 KB 01/12/2
c9.jpeg 6 KB 01/12/2
c10.jpeg 6 KB 01/12/2
dior -- 01/12/2
d1.jpeg 7 KB 30/11/2
d3.jpeg 6 KB 30/11/2
d4.jpeg 8 KB 30/11/2
d5.jpeg 7 KB 30/11/2
d6.jpeg 7 KB 30/11/2
d7.jpeg 10 KB 30/11/2
d8.jpeg 9 KB 01/12/2
d9.jpeg 9 KB 01/12/2
d10.jpeg 7 KB 01/12/2
d11.jpeg 9 KB 01/12/2
d12.jpeg 5 KB 01/12/2
hermes -- 01/12/2
a.jpeg 7 KB 30/11/2
b.jpeg 6 KB 01/12/2
c.jpeg 7 KB 30/11/2
d.jpeg 5 KB 30/11/2
e.jpeg 4 KB 30/11/2
f.jpeg 4 KB 30/11/2
g.jpeg 5 KB 30/11/2
i.jpeg 8 KB 01/12/2
j.jpeg 5 KB 01/12/2


purses
Search
Favorites
Dropbox
AirDrop
All My Files
iCloud Drive
Applications
Downloads
Documents
dev
Devices
Remote Disc
Shared
iMac5
Tags
chanel
dior
hermes


chanel purse

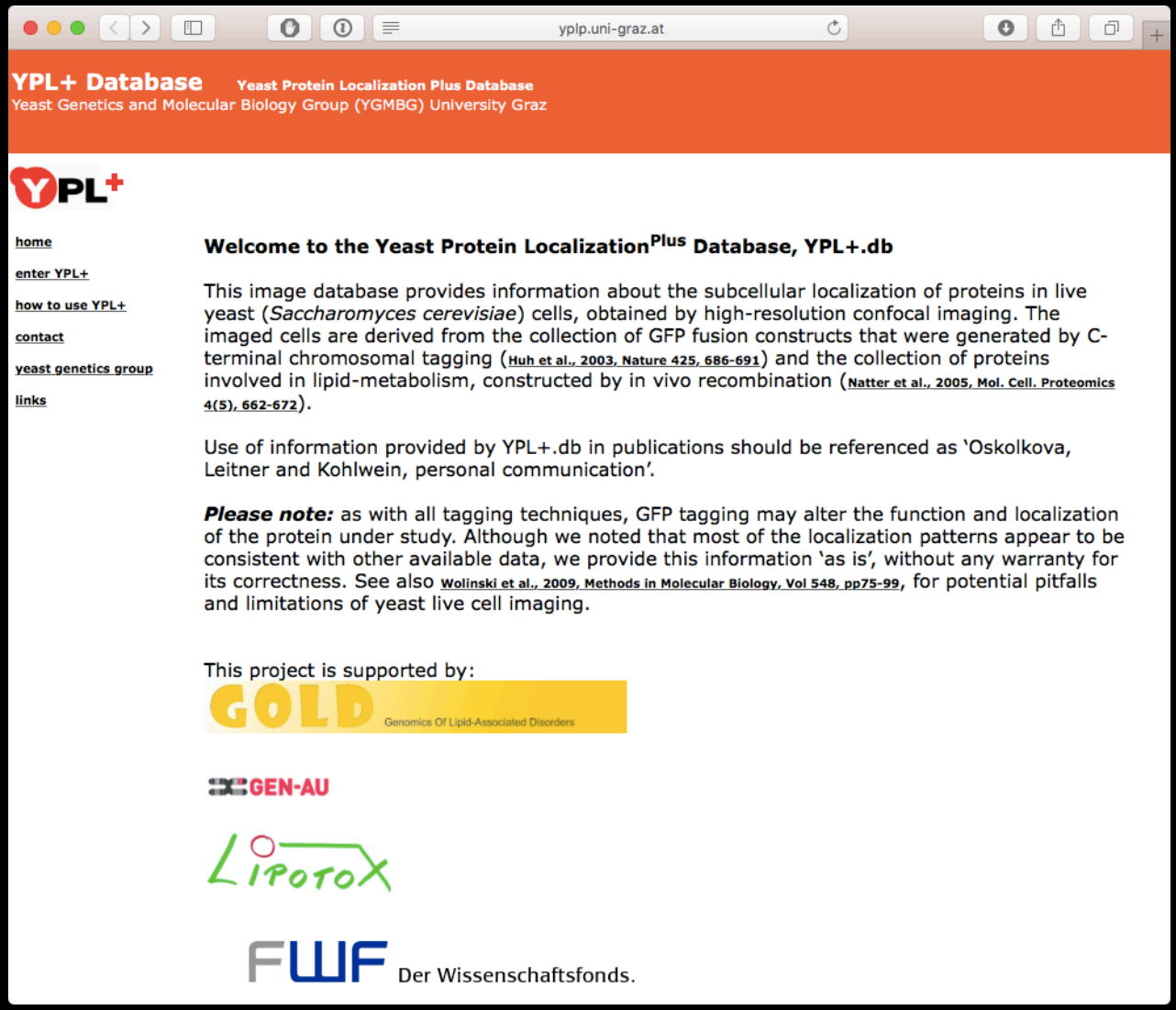
Znanost o torbicah



(tukaj je bil en daljši demo Orange-a)

Celična lokalizacija proteinov

Uroš Petrovič



http://yplp.uni-graz.at

Zarodne celice

Riccardo Bellazzi Mar 18 (1 day ago)

to me

Thanks Blaz. We have tested the image embedder with images of oocytes of two types, competent and not competent for development: embedding + plus random forests 96% accuracy 10 fold-cv, expert accuracy 92%, both on gold standard (e.g. lab evaluated). Unbelievable. What is the reference of the embedder you are using?

orange.biolab.si

orange

Screenshots Download Docs Blog

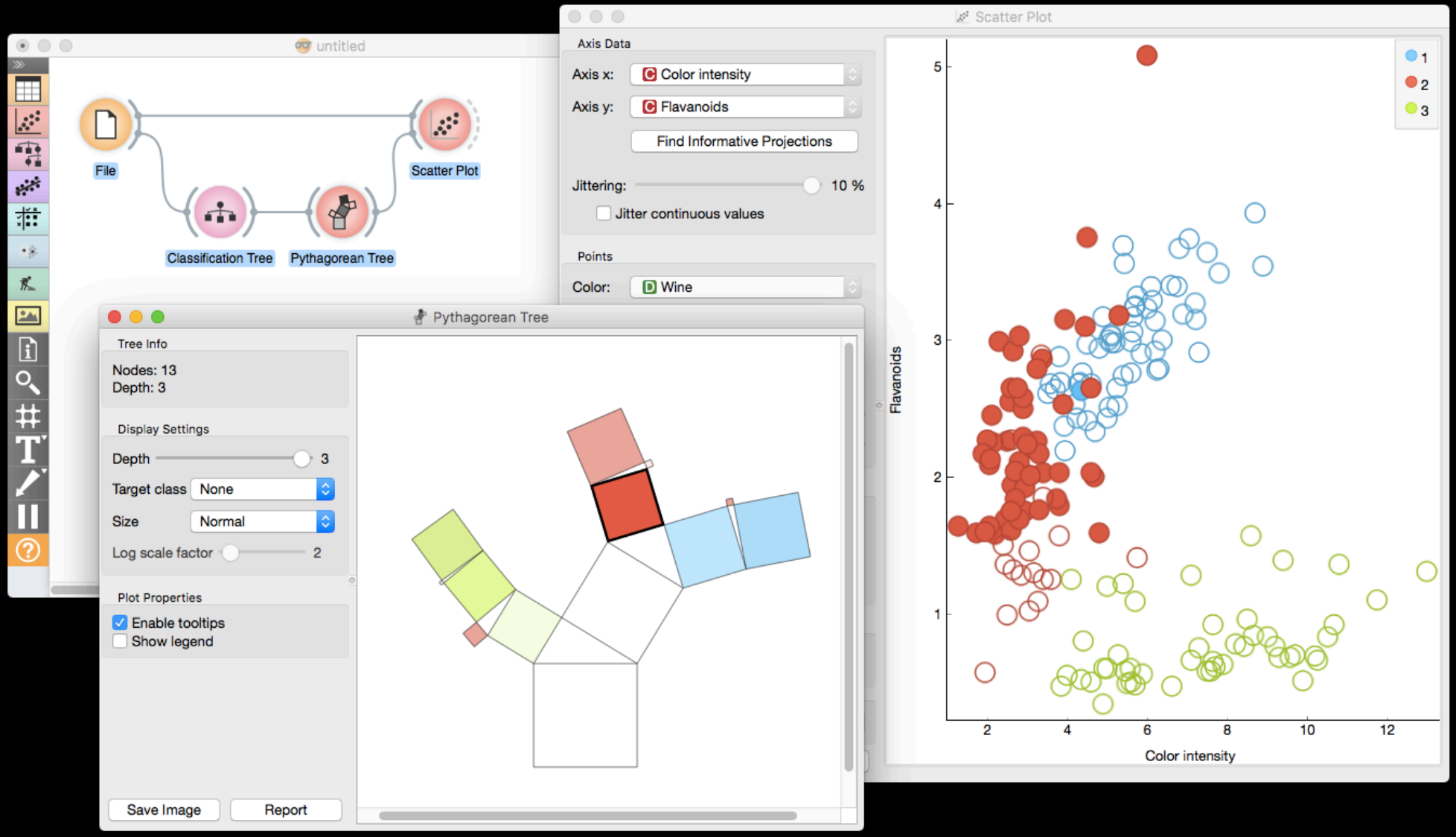
Data Mining Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

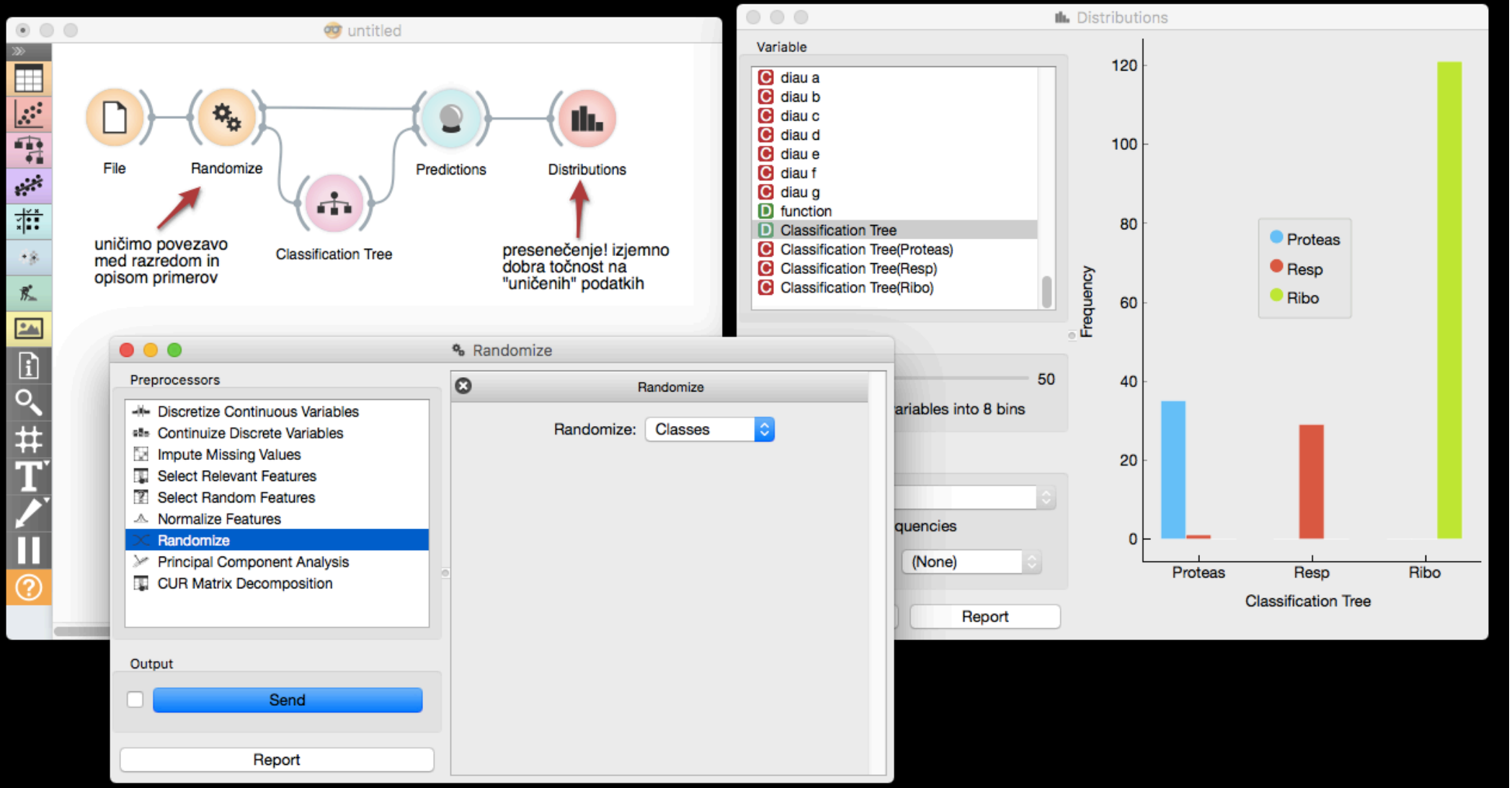
Download Orange

The old version, Orange 2.7, is still available.

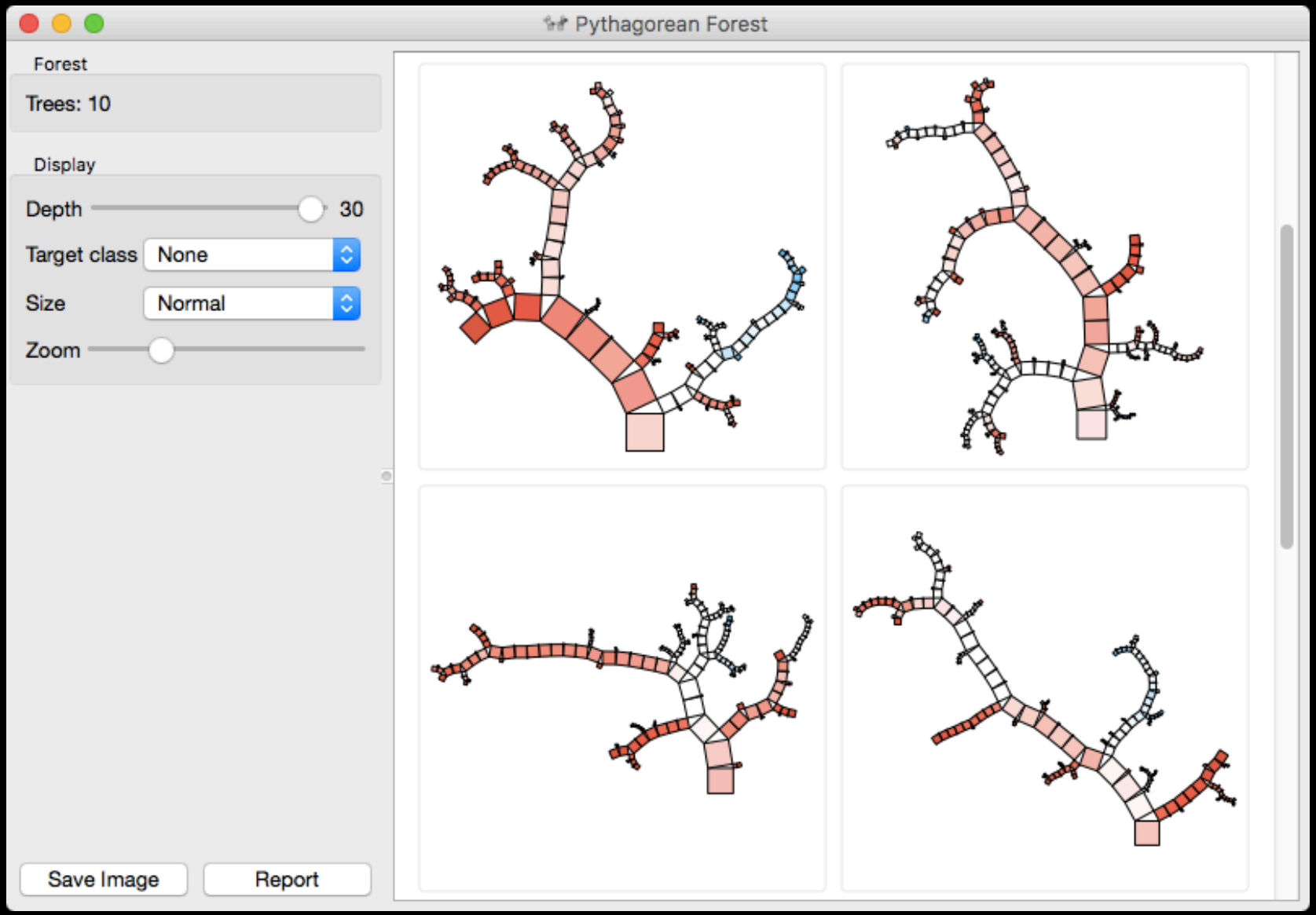
Interaktivnost



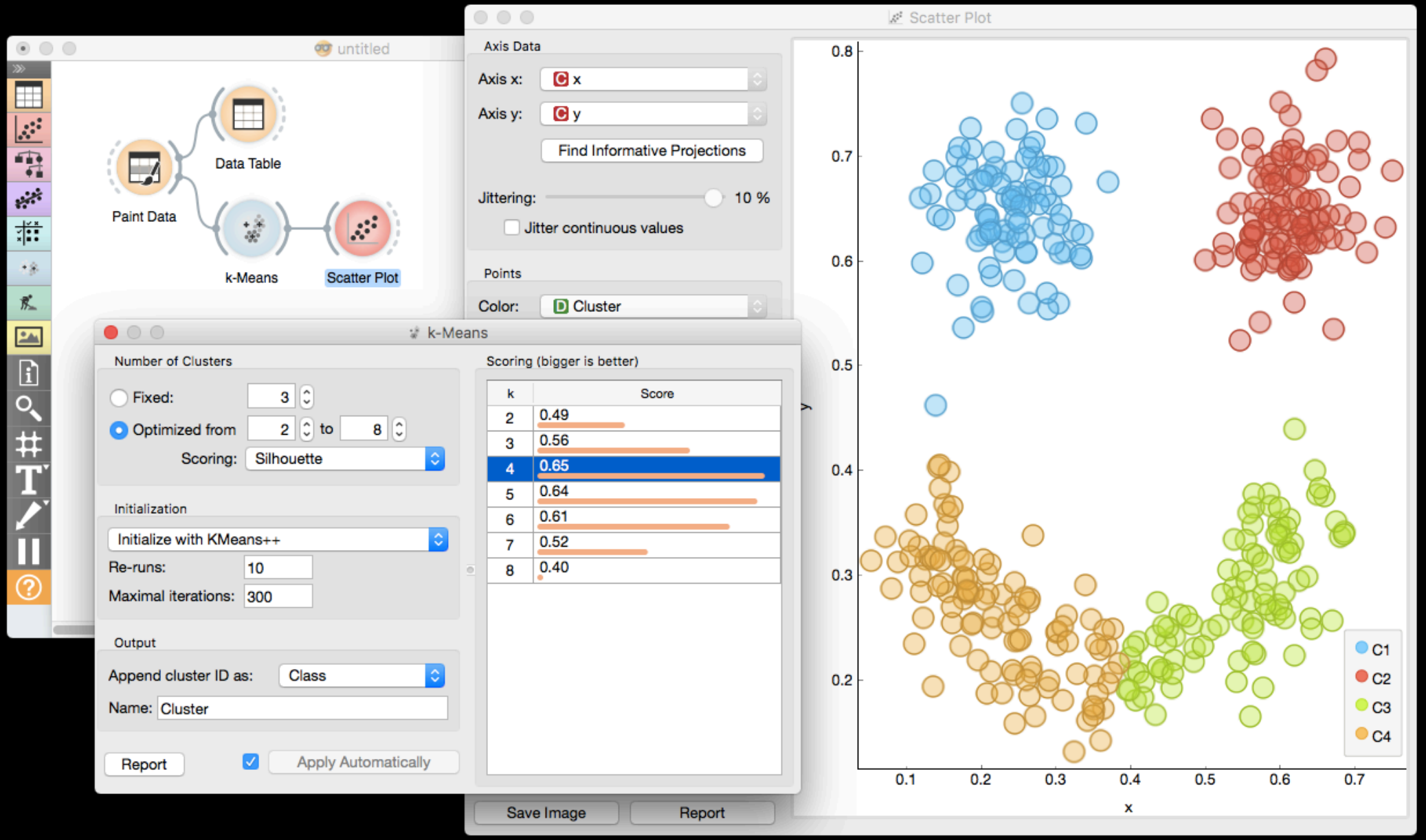
Eksperimentiranje



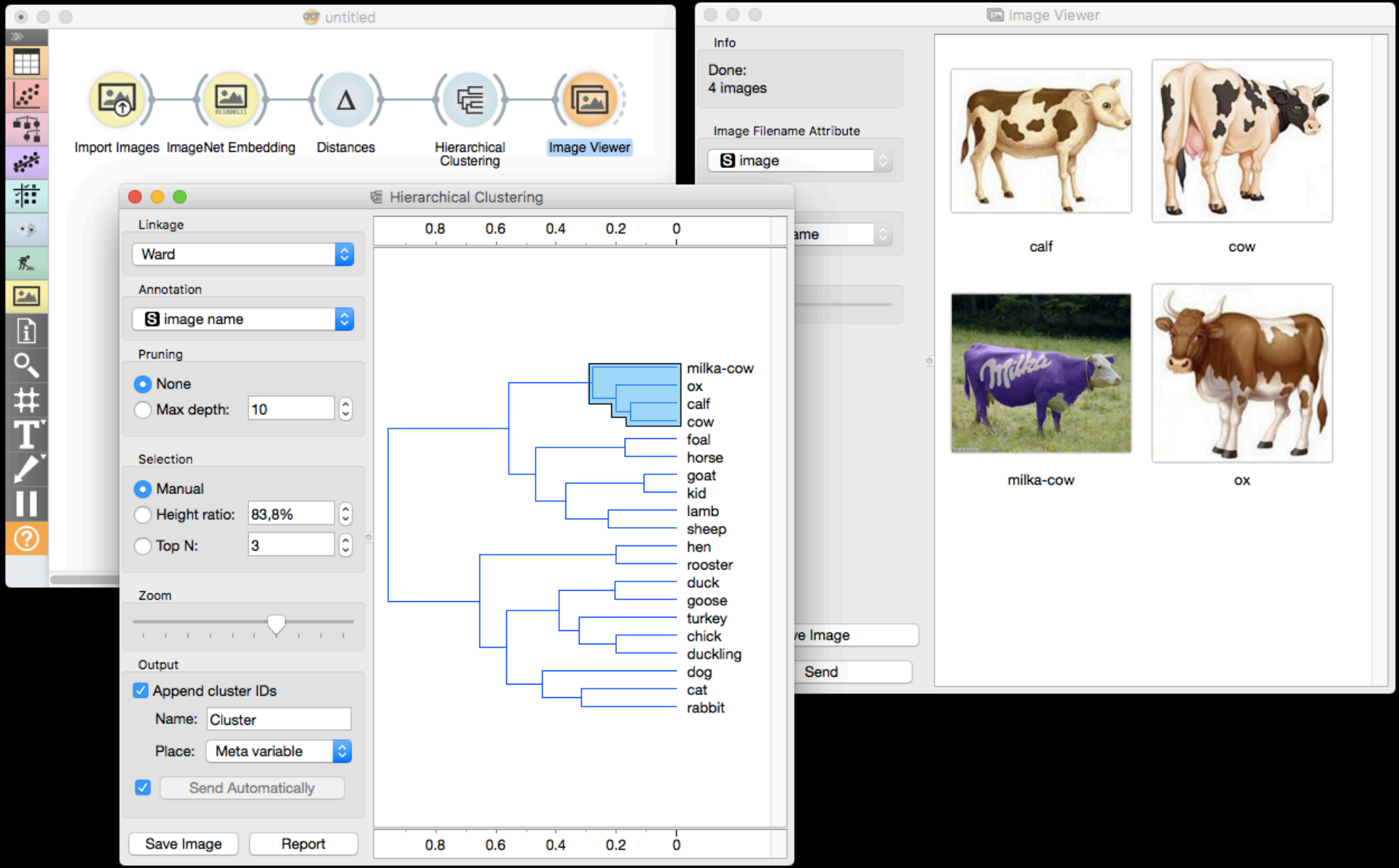
Všečne vizualizacije



Orodja za poučevanje



… in še kaj norega



Nada Lavrač

Donald Michie

Ivan Bratko

Programme of WebLab97

held in association with the Jožef Stefan Institute
at Grand Hotel Toplice, Bled, Slovenia, 2-3 September 1997

Tuesday, 2nd September, 1997

9:30 **Donald Michie**

Introduction to the meeting. WebLab as a shared testing shed. What question has the meeting been called to consider? Walk-through of the programme and the planning agenda.

10:00 **Matthew Webster**

IBM's Java mission. Description of corporate internet requirements. Commonalities and divergencies from scientific requirements.

10:30 **Questions and discussion**

11:00 **Coffee**

11:20 **Matthew Webster and Stephen Blackheath**

Java language, Java Development Kit, Java Beans, Common object models.

12:20 **Questions and discussion**

12:45 **Lunch**

14:15 **Java-less approaches:**

14:15 **Nada Lavrač**

ILPNET. Downloadable software. Cross-platform issues. Comparative testing. (http://www-ai.ijs.si/ilpnet.html)

14:30 **Blaž Zupan**

Issues in the development of an integrated Machine Learning environment, using a modular, object-oriented approach.

14:45 **Carl Rasmussen**

DELVE - Data for Evaluating Learning in Valid Experiments. Experiences with a Web archive for comparing the predictive performance of statistical learning algorithms. (http://www.cs.utoronto.ca/~delve)

15:15 **Specific questions and dicussions on the above three presentations**

15:30 **General discussion**

How far can such facilities be pushed, short of Java-based solutions?

16:00 **Tea**

16:20 **Matthew Webster, Stephen Blackheath**

Recap of Java language and library facilities of special relevance to the above.

16:50 **Adjourn for demos**

including Co-operative Agents video (**Hyacinth Nwana**), Test your own (a) rule-discovery, (b) rule-comprehension (**Rupert Parson**), BOT-WORLD (**Claude Sammut**).

18:15 **Jožef Stefan Evening Discourse**

Chair: **Prof. Ivan Bratko**
Lecturer: **Prof. J.R. Quinlan**
Title: **Boosting Inductive Learning Systems**

19:15 **Questions and Discussion**

19:40 **Depart for Vila Bled Restaurant**

20:00 **Banquet**

Welcome to foreign visitors from **Dr. Tomaž Kalin**, Deputy Director, Jožef Stefan Institute.

↦

1990 2000 2010 2020

ukazna vrstica

C++

Janez Demšar

1990 2000 2010 2020

Matjaž Kukar

Python

Cross-Validation

Validating the accuracy of classifiers on the training data, as we did above, serves demonstration purposes only. Any performance measure that assess accuracy should be estimated on the independent test set. Such is also a procedure called cross-validation, which averages performance estimates across several runs, each time considering a different training and test subsets as sampled from the original data set:

```
data = Orange.data.Table("voting")
bayes = Orange.classification.bayes.NaiveLearner()
res = Orange.evaluation.testing.cross_validation([bayes], data, folds=5)
print "Accuracy: %.2f" % Orange.evaluation.scoring.CA(res)[0]
print "AUC:      %.2f" % Orange.evaluation.scoring.AUC(res)[0]
```

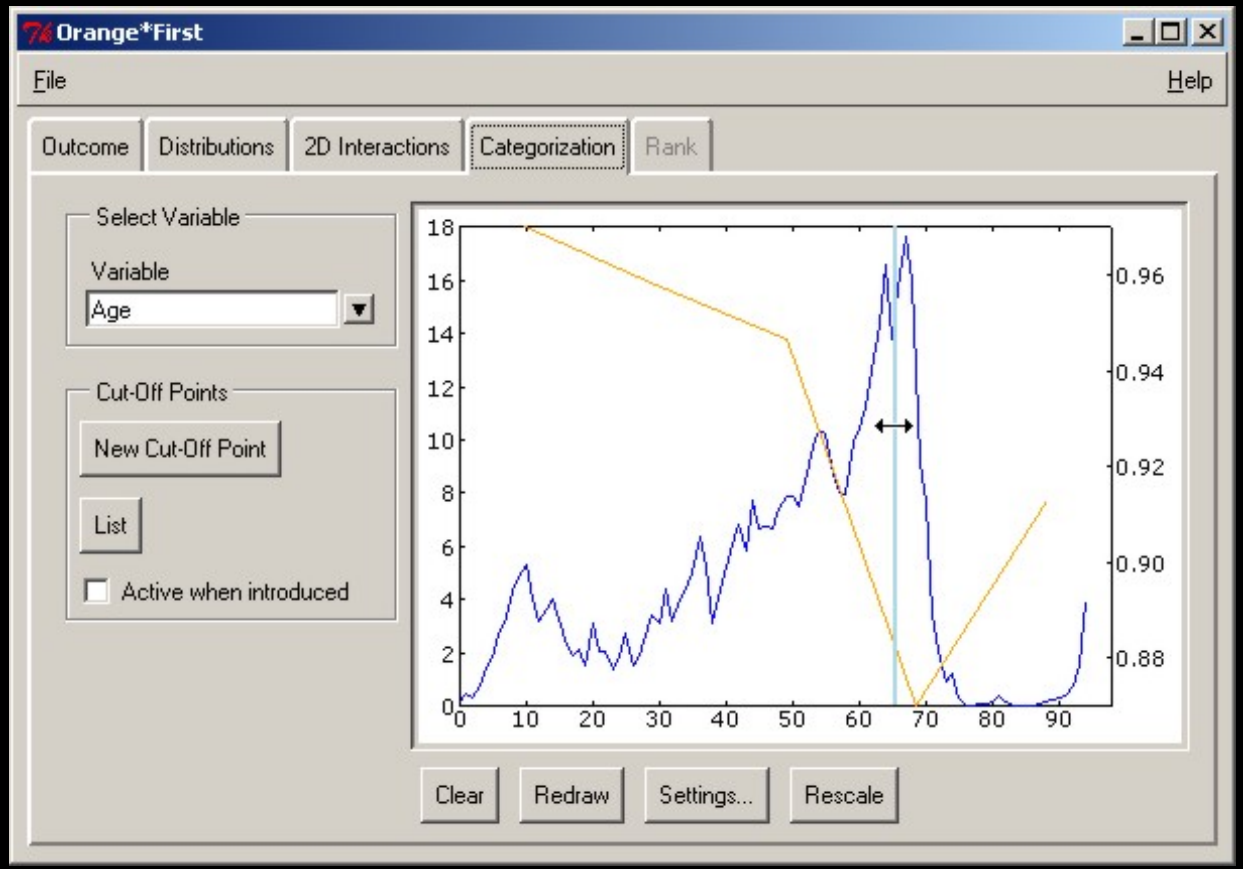
Cross-validation is expecting a list of learners. The performance estimators also return a list of scores, one for every learner. There was just one learner in the script above, hence the list of size one was used. The script estimates classification accuracy and area under ROC curve. The later score is very high, indicating a very good performance of naive Bayesian learner on senate voting data set:

```
Accuracy: 0.90
AUC:      0.97
```

1990 2000 2010 2020

interaktivnost

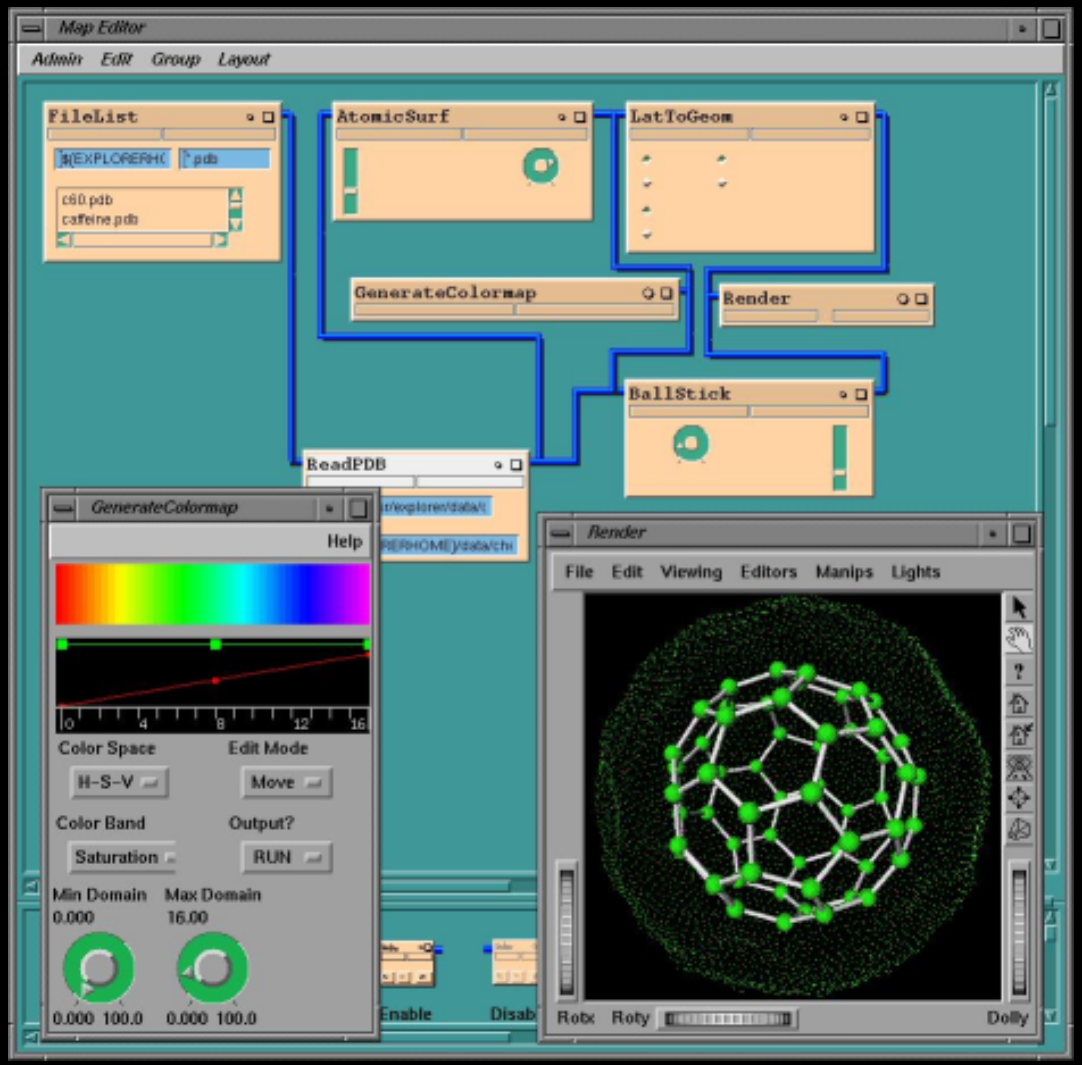
Noriaki Aoki



1990 2000 2010 2020

vizualno programiranje

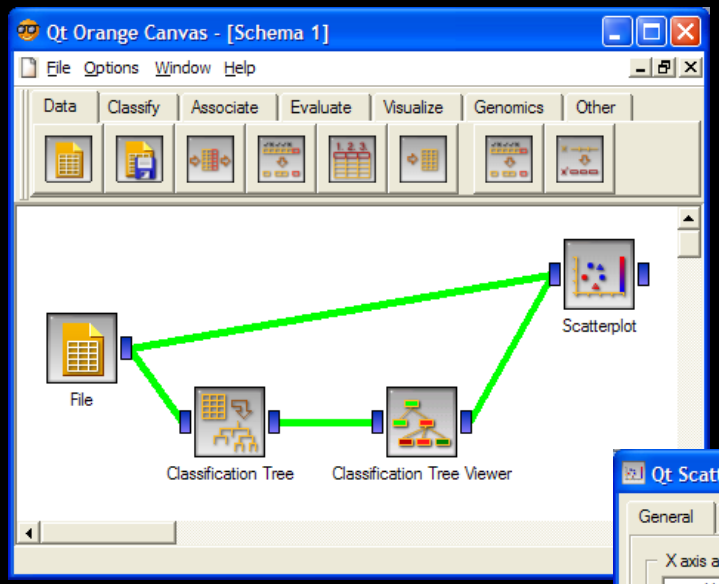
SGI Data Explorer

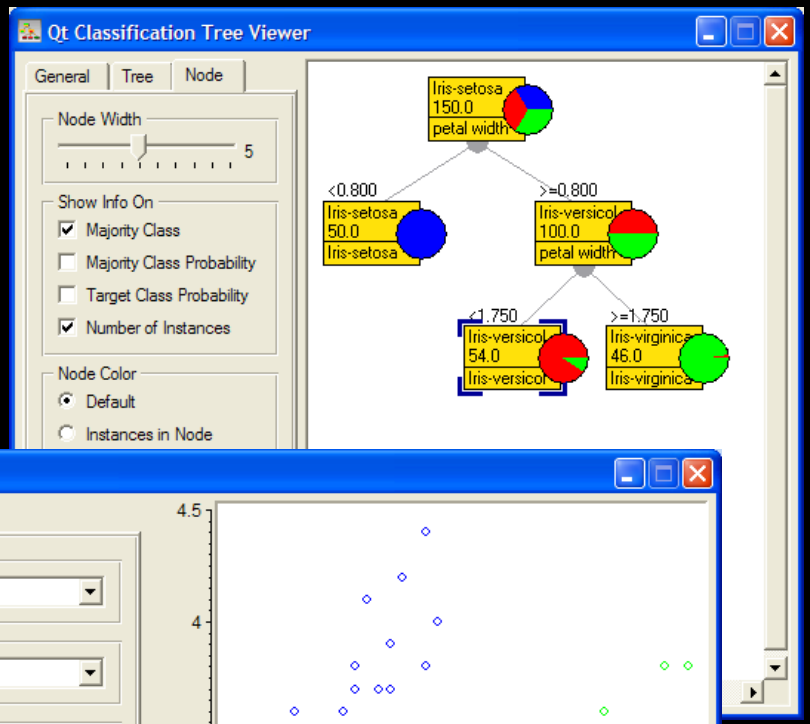


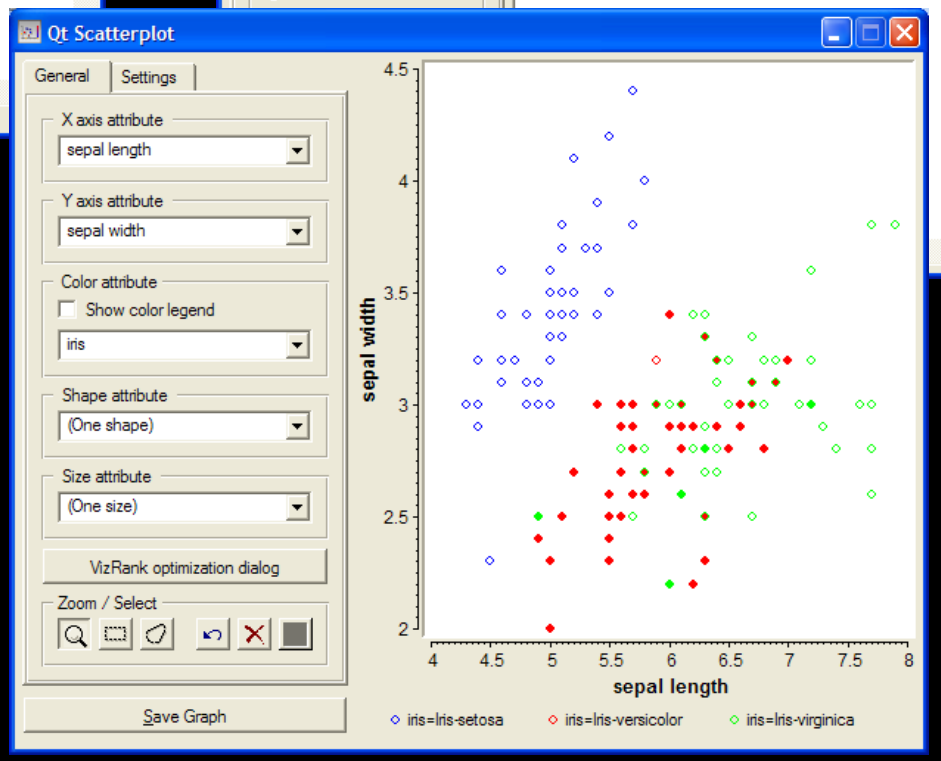
Olin Johnson

1990 2000 2010 2020

orange!

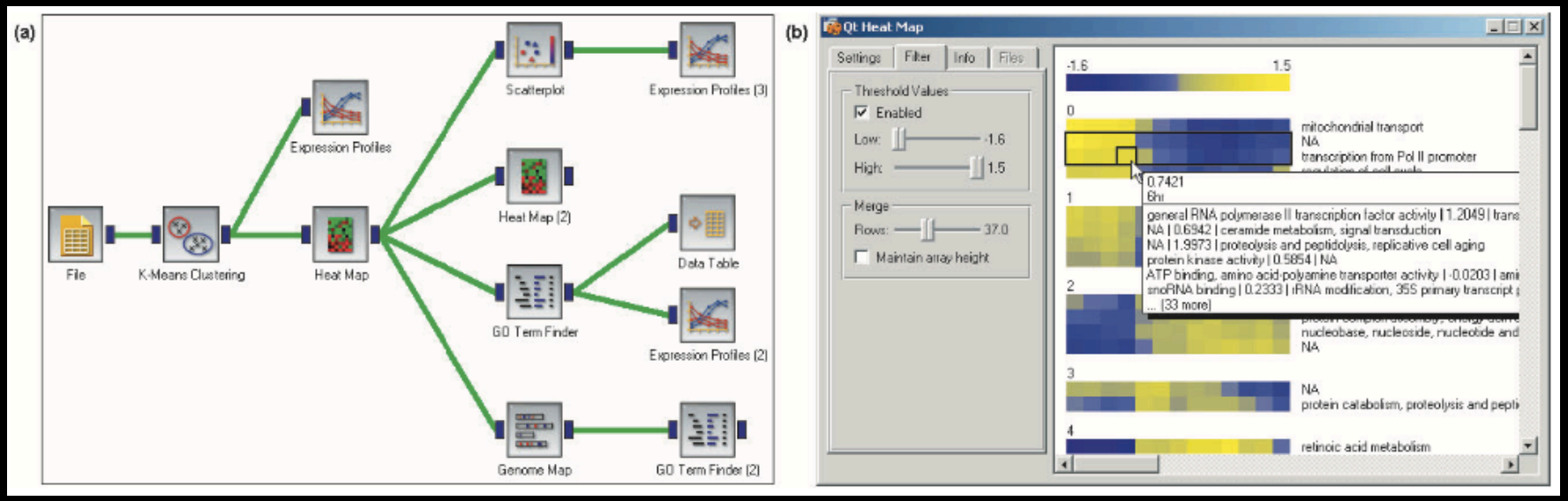






Gregor Leban

1990 2000 2010 2020



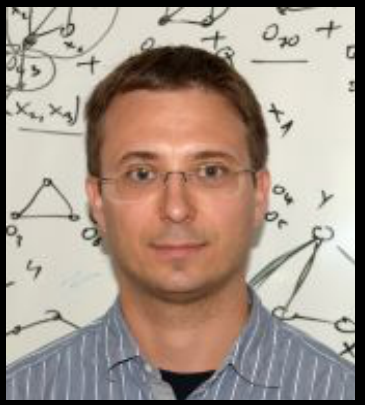
bioinformatika

Gad Shaulsky

Uroš Petrovič

Marko Toplak

Tomaž Curk



1990 2000 2010 2020

Miha Štajdohar

omrežja

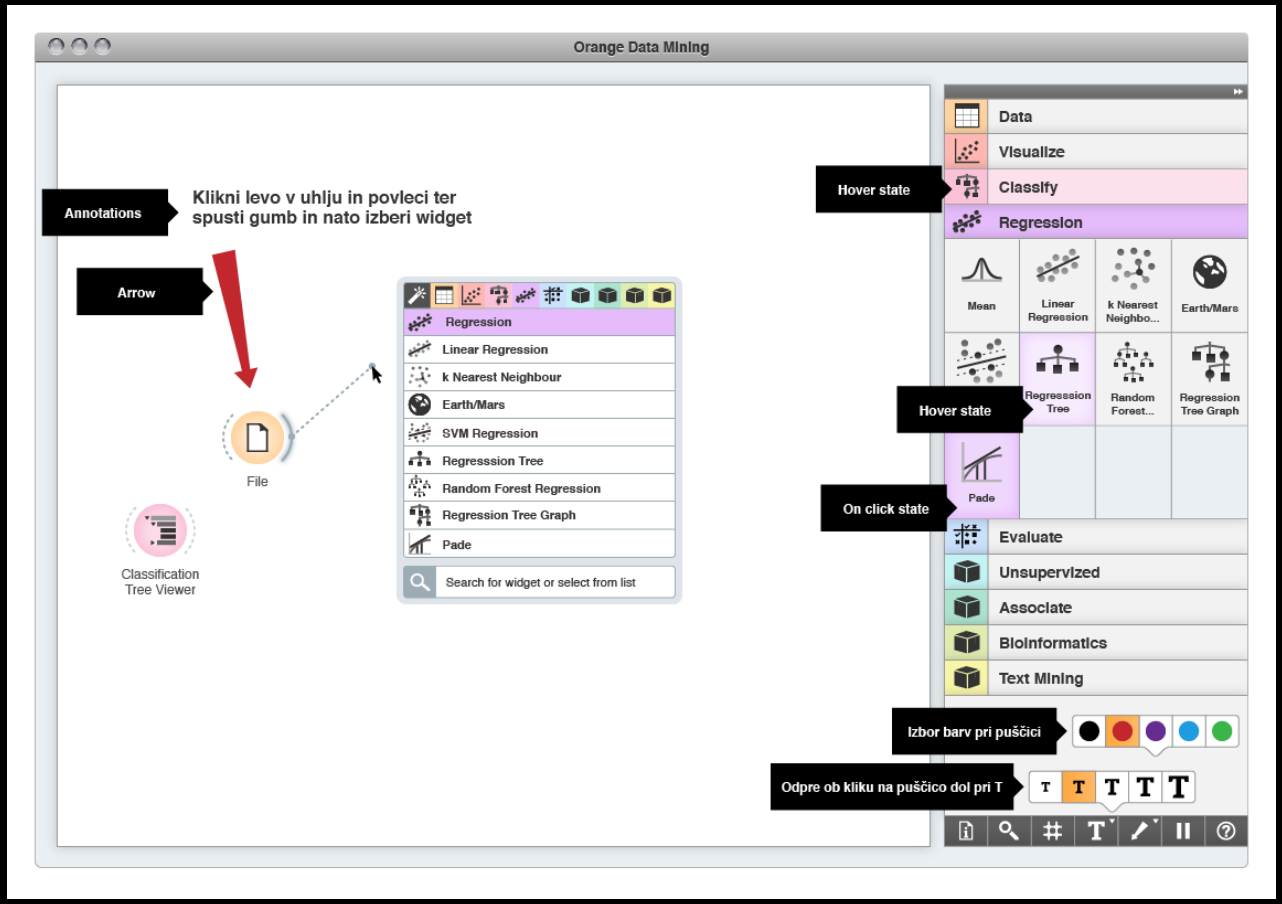
glavni programer

Aleš Erjavec

1990 2000 2010 2020

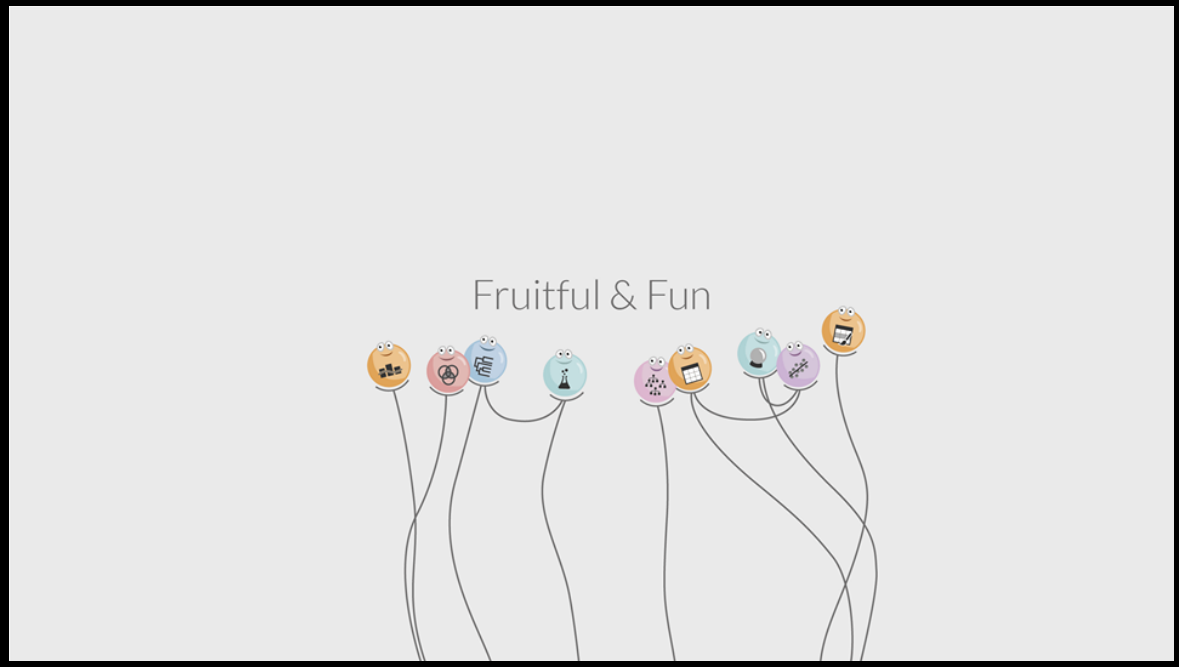
preobleka

Peter Čuhalev



1990 2000 2010 2020

celostna podoba



Agnieszka Rovšnik

1990 2000 2010 2020

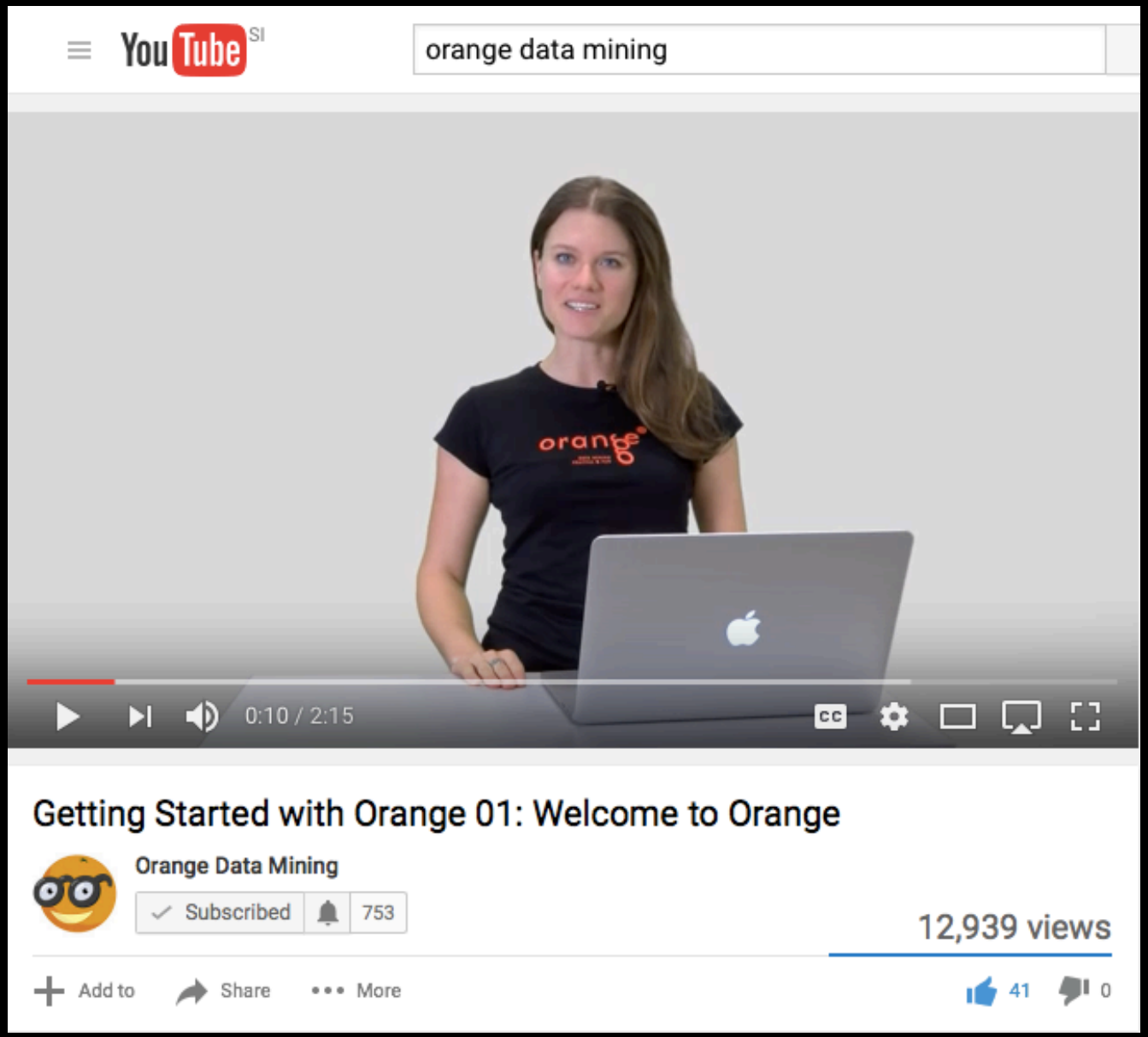
Lan Žagar

renesansa

Anže Starič

1990 2000 2010 2020

video kanal

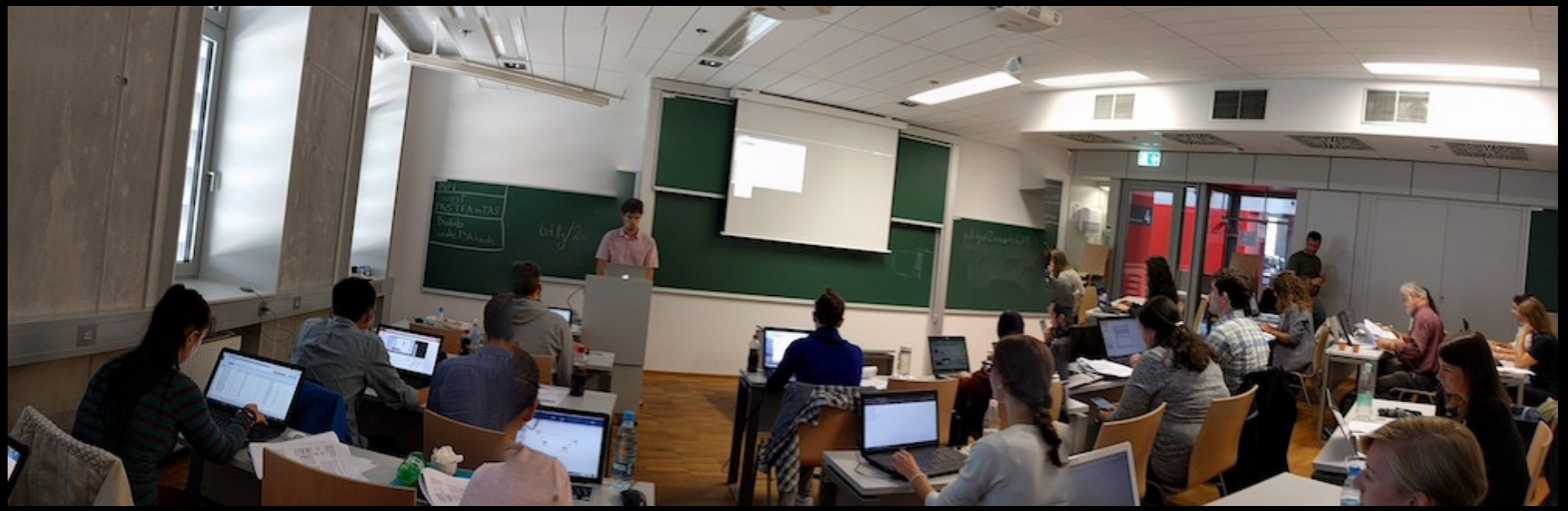


Ajda Pretnar

1990 2000 2010 2020

tečaji, delavnice

praktično vsi
sodelavci biolaba



1990 2000 2010 2020

500.000 programskih vrstic

čez 1000 obiskovalcev spletne strani dnevno (40% download)

več kot 40 aktivnih razvijalcev

tedenski blog

youtube kanal

twitter, facebook, stackoverflow

uporabniki vseh strok (humanisti, fiziki, ekonomisti)

omenjen na vseh “best of data mining” seznamih

najbrž najboljše orodje za uvajanje v znanost o podatkih

John Gorman @jp_gorman · Apr 4

Spent my morning reviewing Ireland's HospitalWaitingList data from #NTPF using Orange orange.biolab.si #SuperIntuitiveUX

Veeresh @veereshelango · Mar 25

@OrangeDataMiner I loved this tool. The tutorials are well structured and easy as pie 😀

Maia Majumder @maiamajumder · Mar 17

Replying to @maiamajumder

I personally started out on @OrangeDataMiner. I love it & still use it, because it allows you to visualize your data AND your narrative. 7/N

Pradeep K Paijwar @paijwar · Mar 15

RT @OrangeDataMiner: [⛔coding #AI #DataScience #datamining] Orange UI allows you to focus on exploratory #dataanalysis instead of coding…

2ndQuadrant Postgres @2ndQuad · Jan 23

Want informative visualizations of your data? Learn how #2UDA brings together #PostgreSQL & Orange Data Mining - youtu.be/_fCqxiXQVis

biolab

Andrej Čopar, strojna oprema, zlivanje podatkov

Niko Colnerič, text mining

Tomaž Curk, bioinformatika, infrastruktura

Janez Demšar, razvijalec jerda, picajzla pri widgetih

Aleš Erjavec, glavni razvijalec

Jernej Kernc, nove tehnologije, razvoj

Tomaž Hočevar, algoritmi, razvoj

Ajda Pretnar, PR, dokumentacija, digitalna humanistika

Anže Starič, projektni vodja

Martin Stražar, bioinformatika

Vesna Tanko, razvoj, testiranje

Marko Toplak, biomedicina, razvoj

Blaž Zupan, uživa

Lan Žagar, spremljanje razvoja

fri

Nejc Ilenič, globoko učenje, embedding

Pavlin Poličar, pitagorejska drevesa, klasifikacija

Jaka Kokošar, bioinformatika

Matjaž Pančur, storitve v oblaku

Fabio Ricciato, telekomunikacije

Veljko Pejović, telekomunikacije

finance

ARRS

NIH

Astra Zeneca, Lek, Elletra, Soleil

EU FP7 & Horizon 2020

SPS S4