



# One Click Mining

## Interactive Local Pattern Discovery through Implicit Preference and Performance Learning\*

Mario Boley, Michael Mampaey, Pavel Tokmakov, Bo Kang, Stefan Wrobel  
Fraunhofer IAIS / University of Bonn

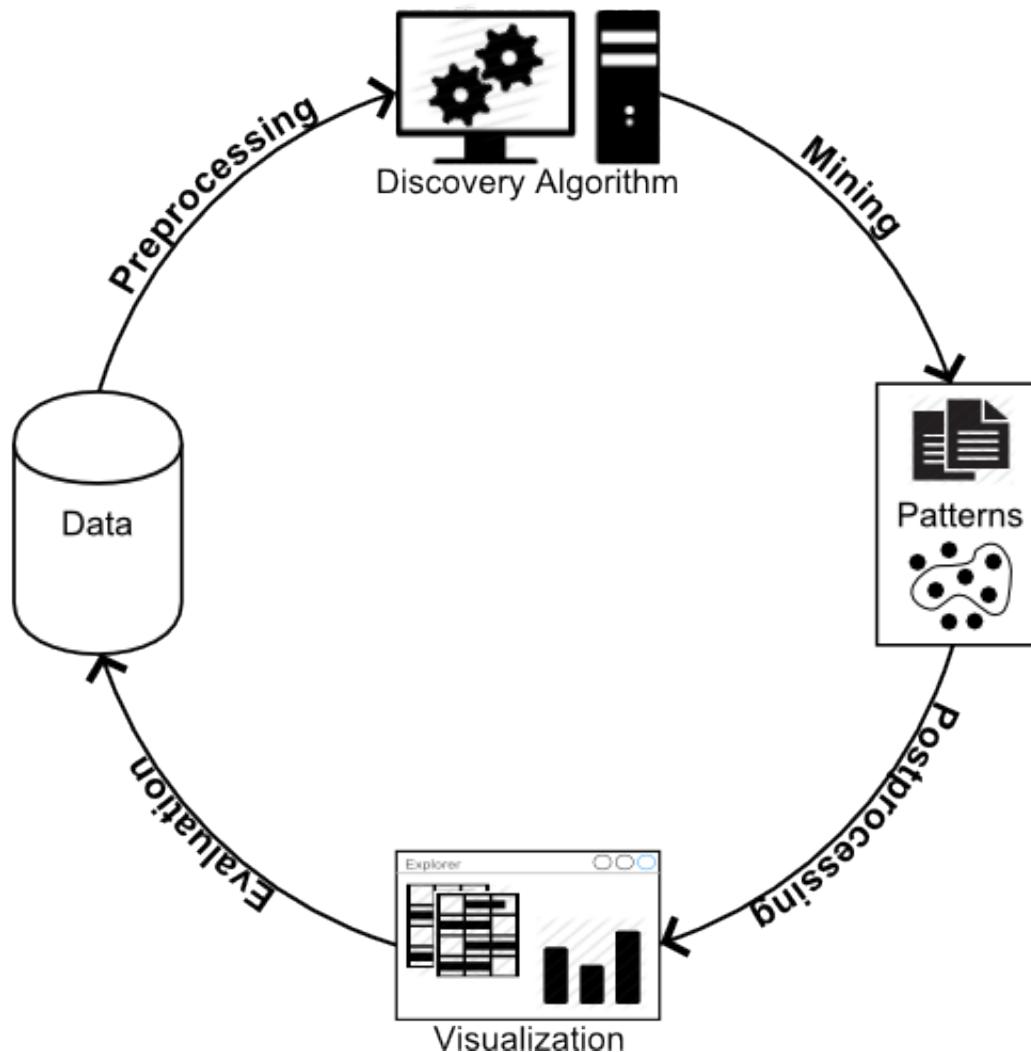


---

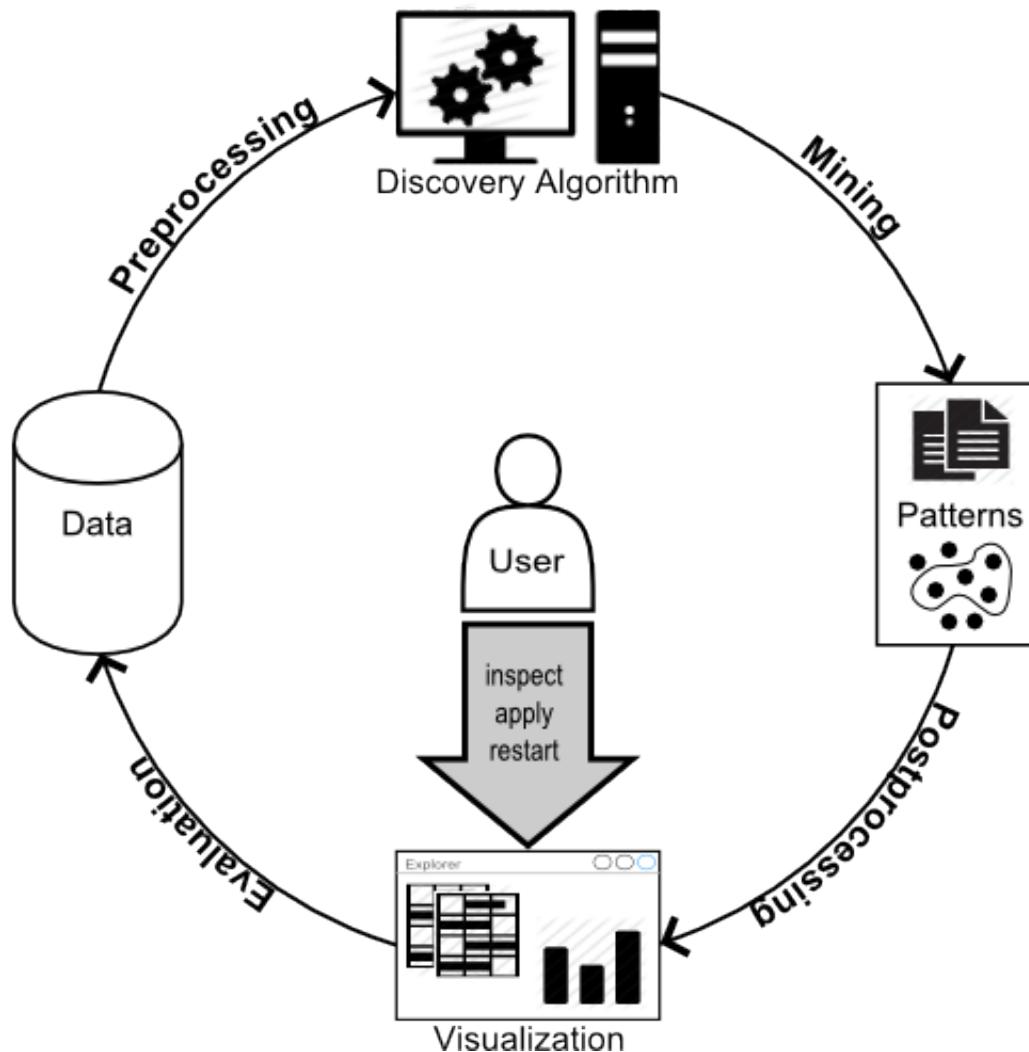
\*Slides based on Mario Boley, IDEA 2013 at KDD, Chicago

- **Build tools, refine them, make life easier**
- **Making coffee using coffee machine**
  - Beans, water as input
  - Press the run button
  - Get a nice cup of coffee as output
- **Subroutines are encapsulated for solving complex problems**

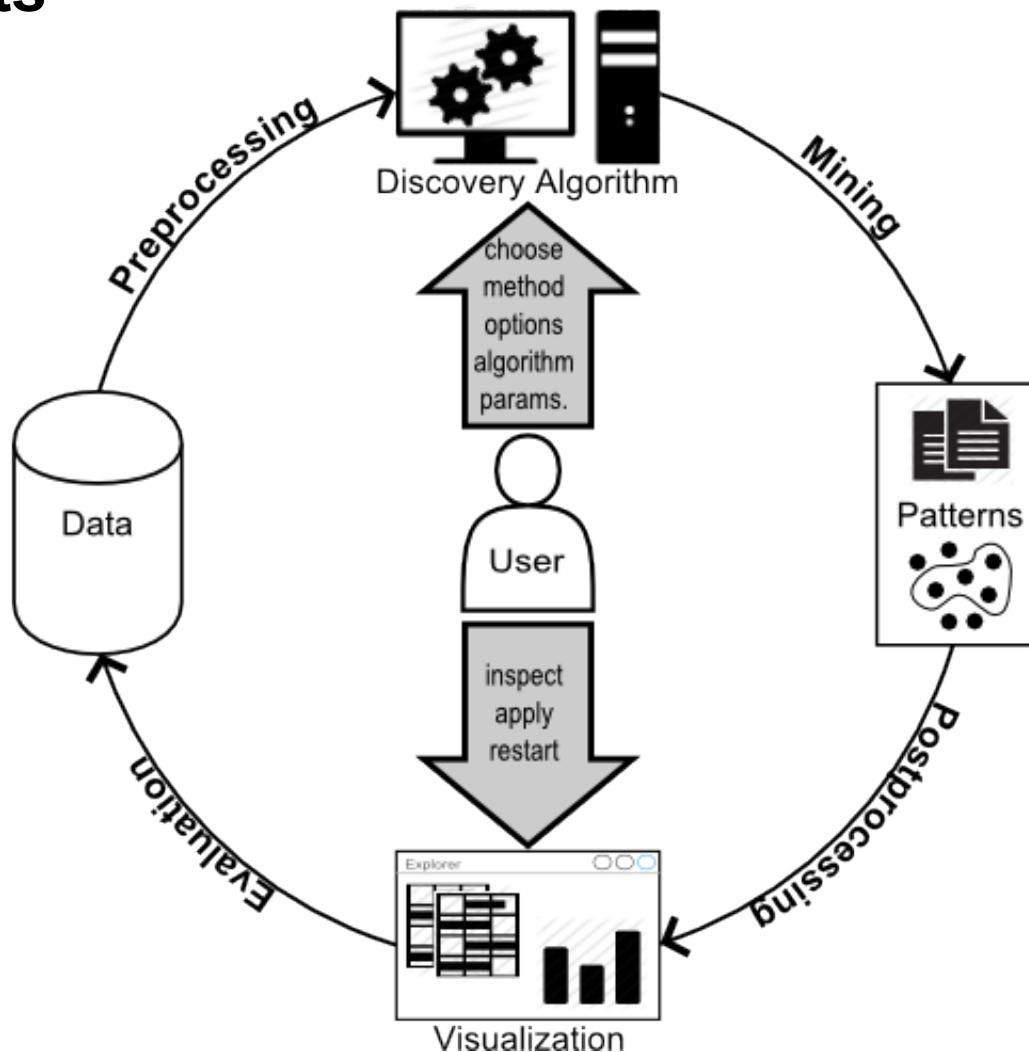
# Knowledge Discovery has iterative workflow



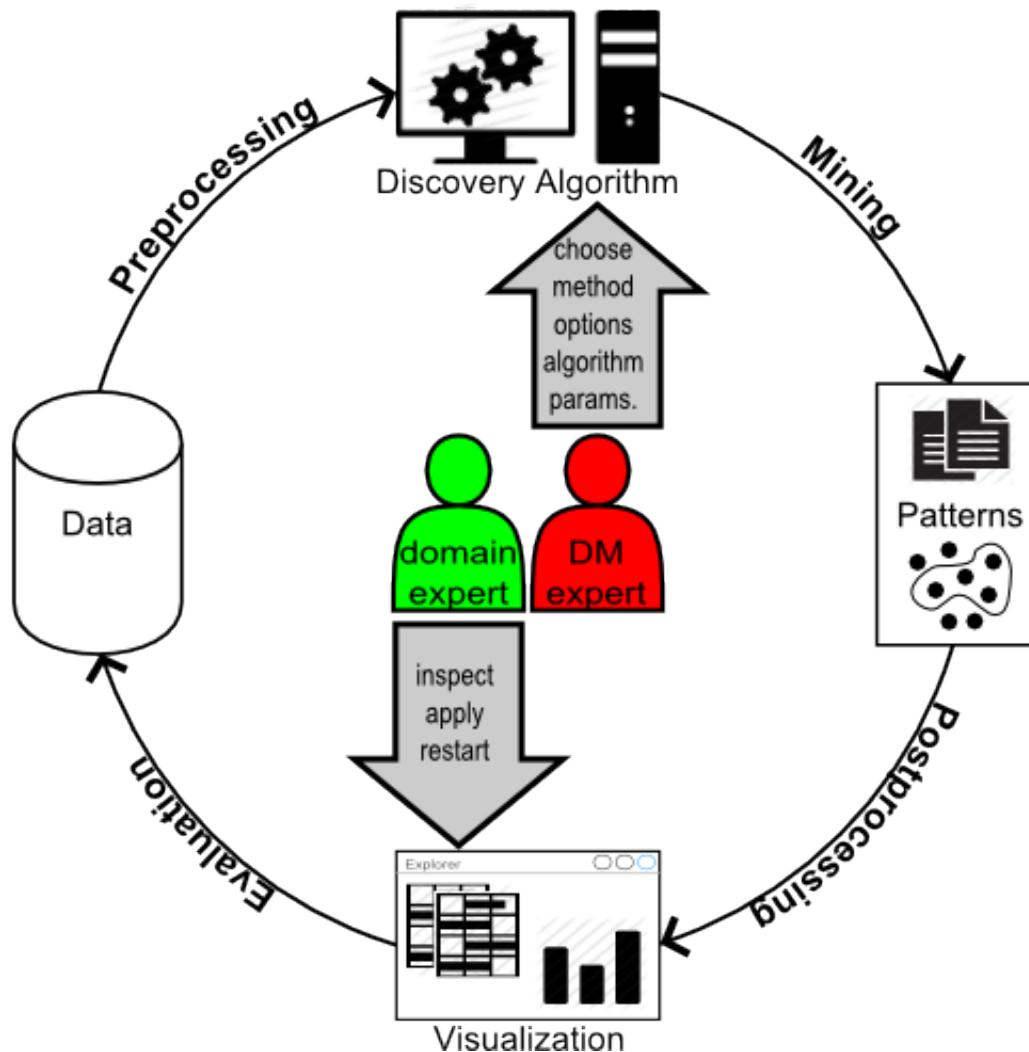
# User is required for evaluation



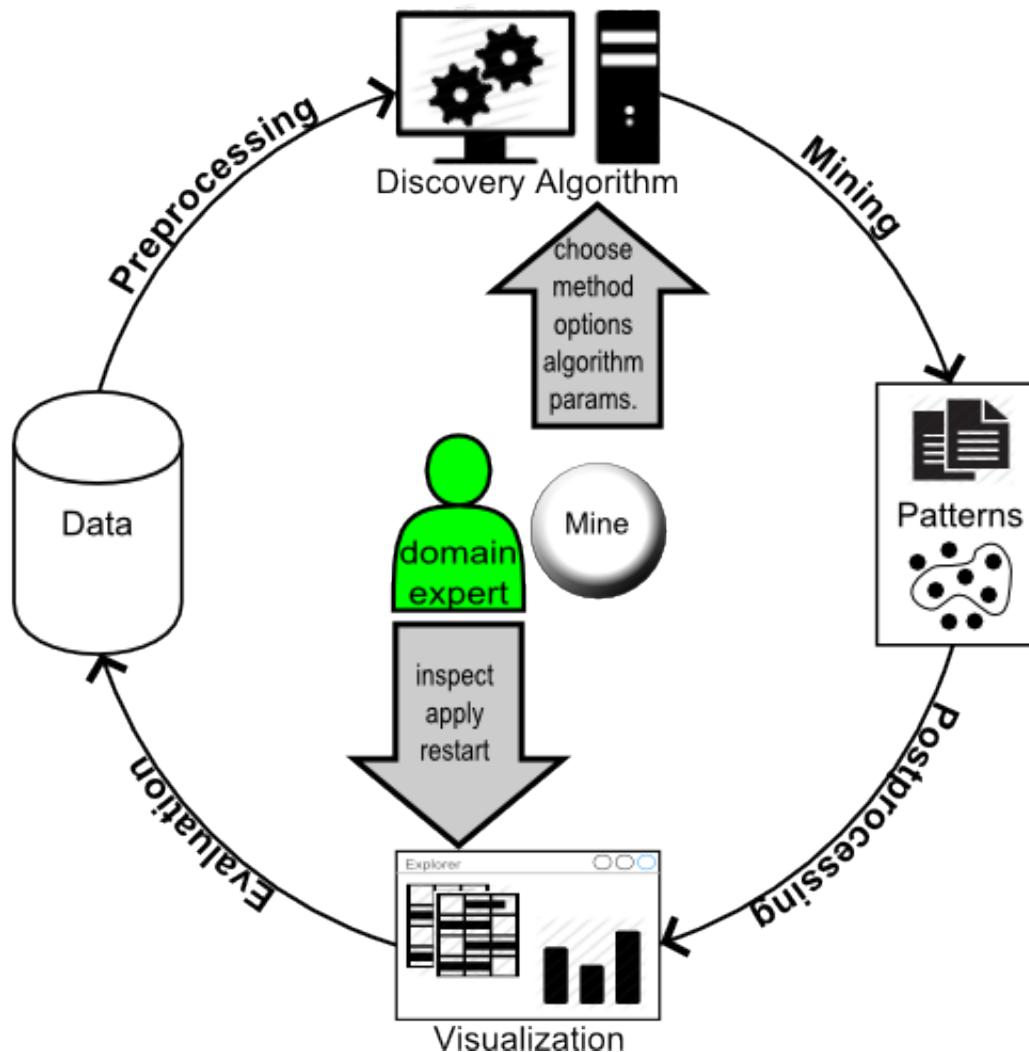
# In current tools she is also required to control technical components



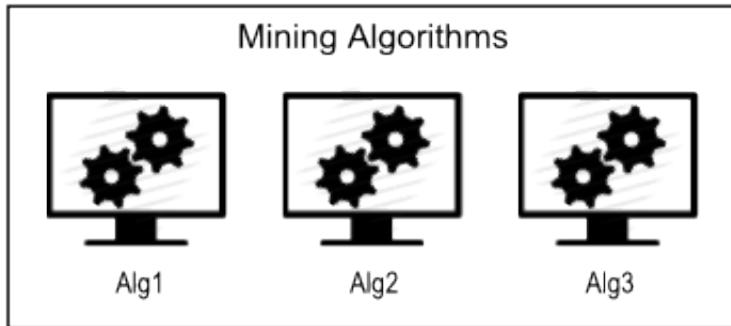
# Hence, user actually has two different roles



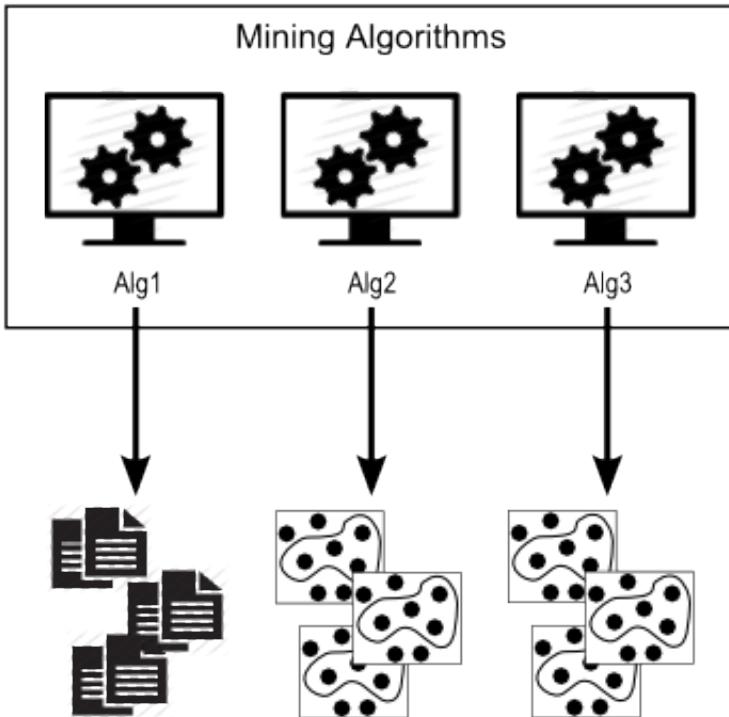
# One-click-mining aims to eliminate the technical role



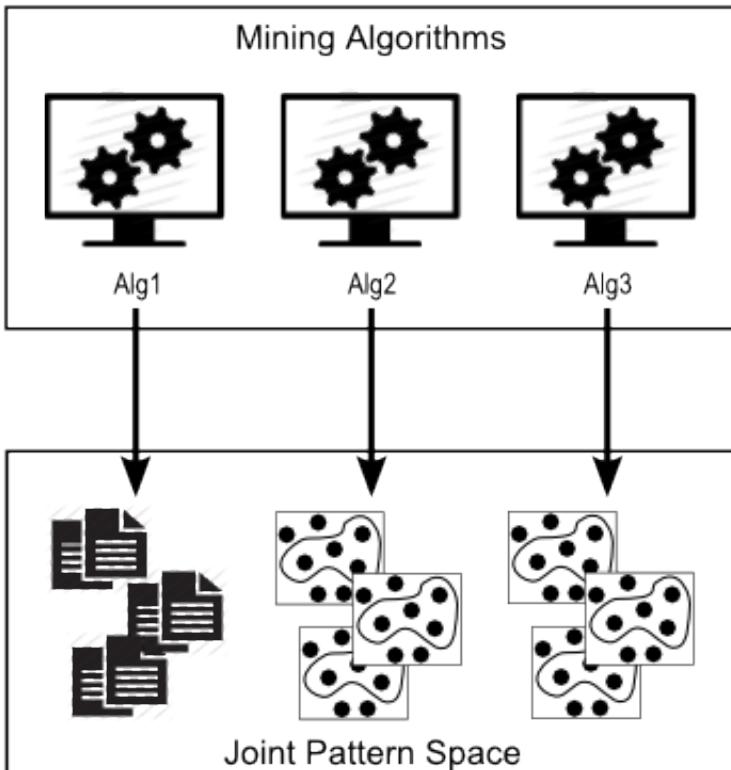
# Design of one-click-mining system starts with the algorithms to be included



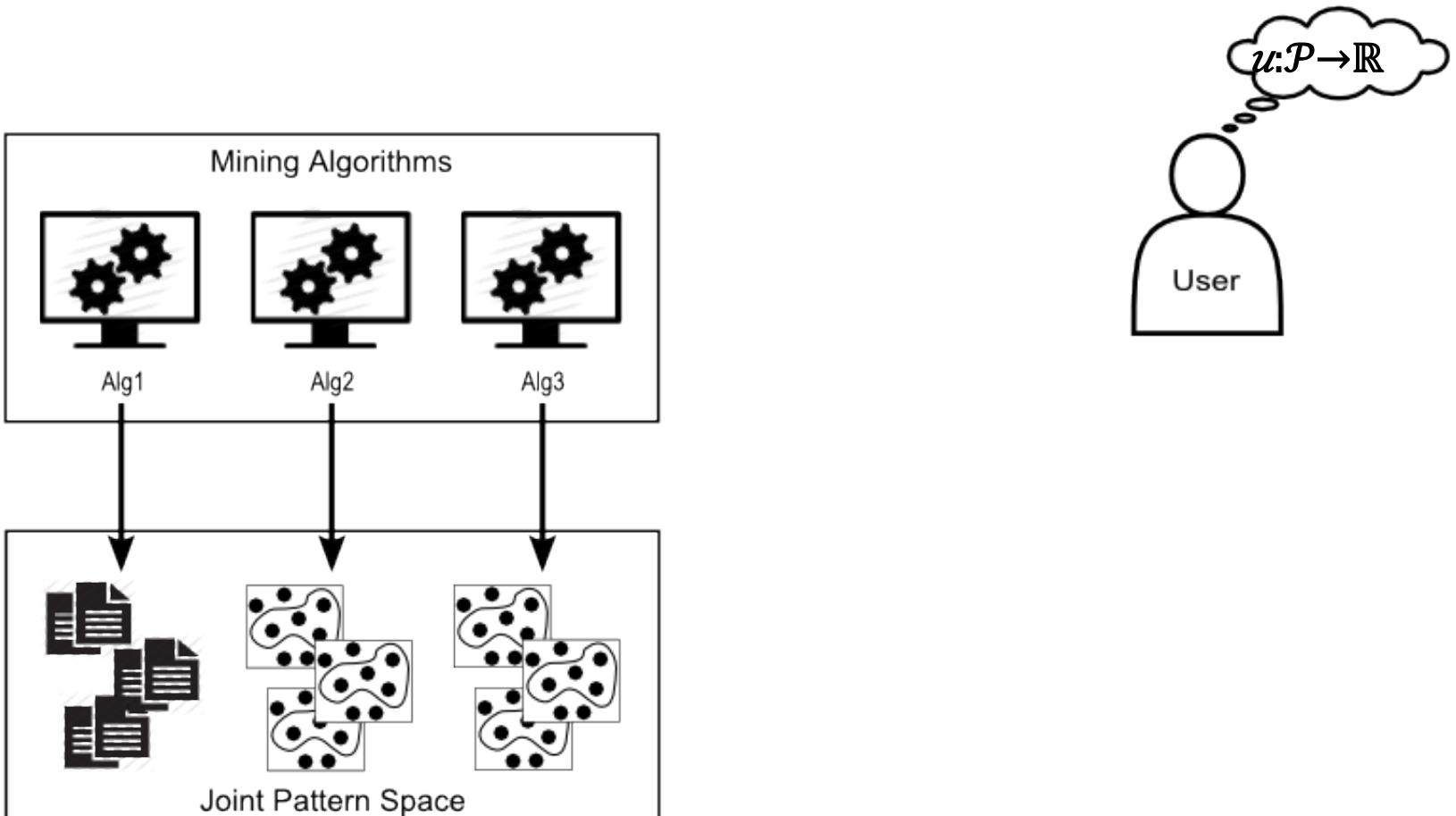
# Diverse set of algorithm from different method classes are possible



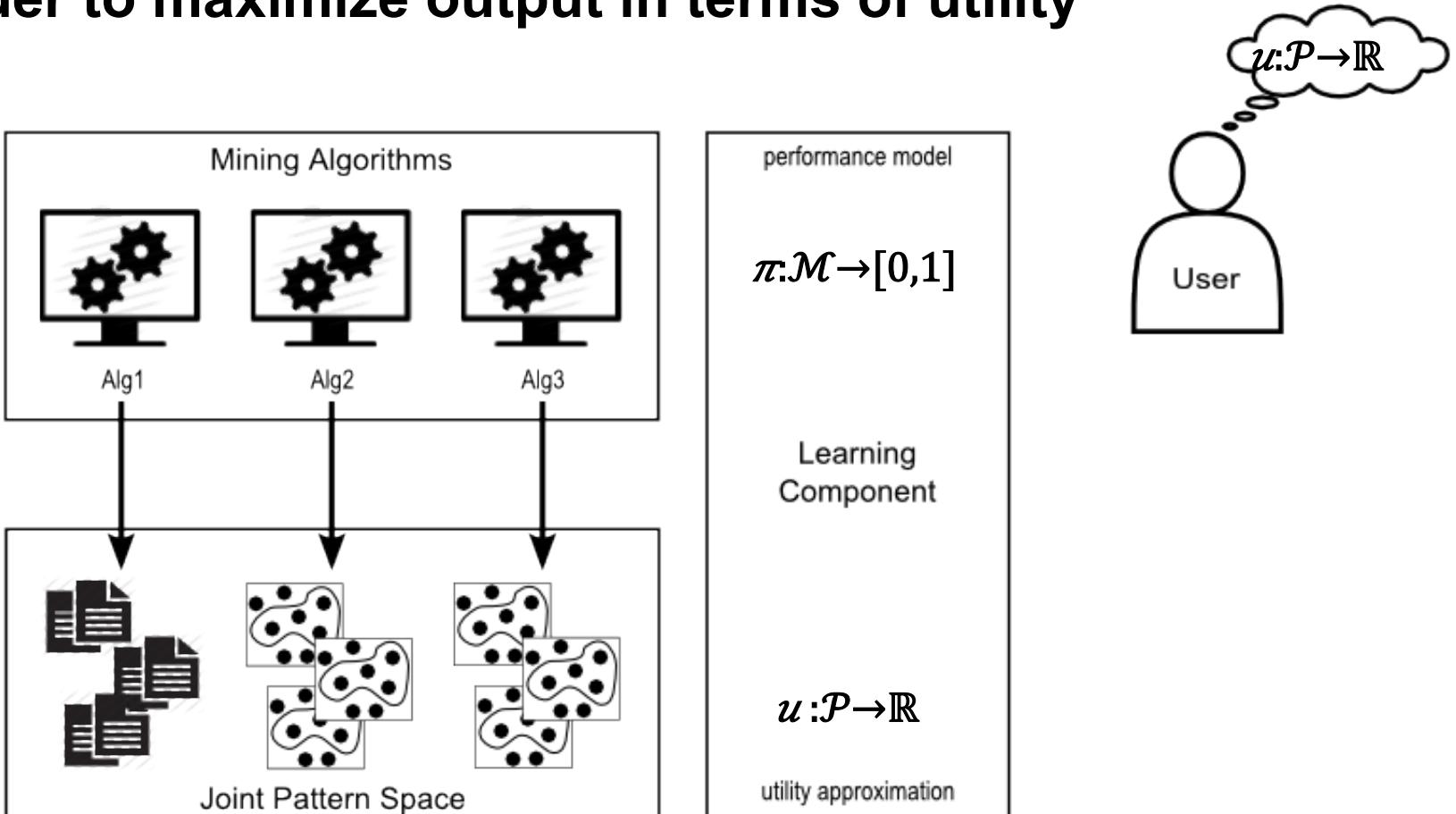
# Together they define space of patterns that can potentially be discovered by the system



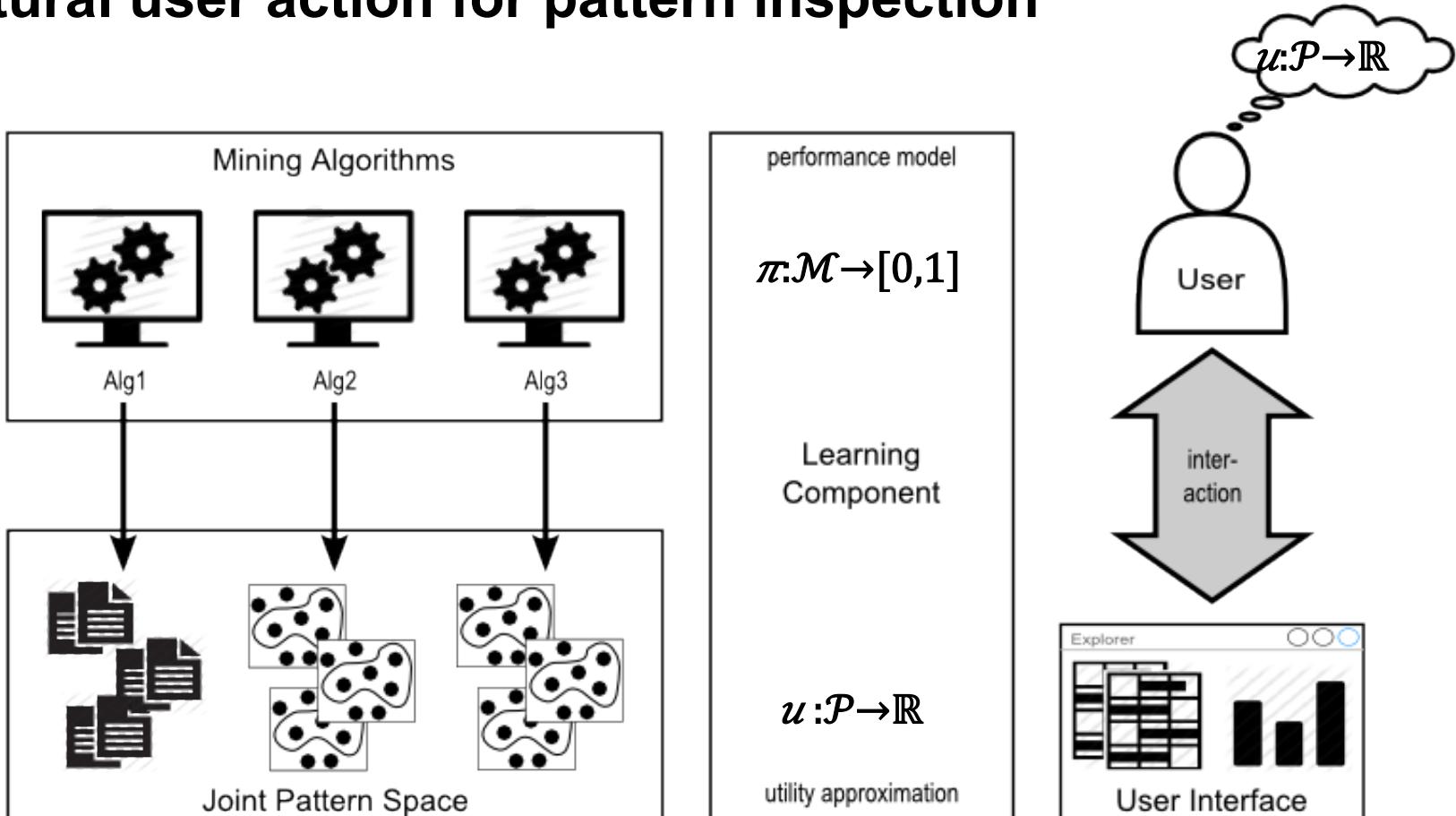
# User brings some latent utility notion



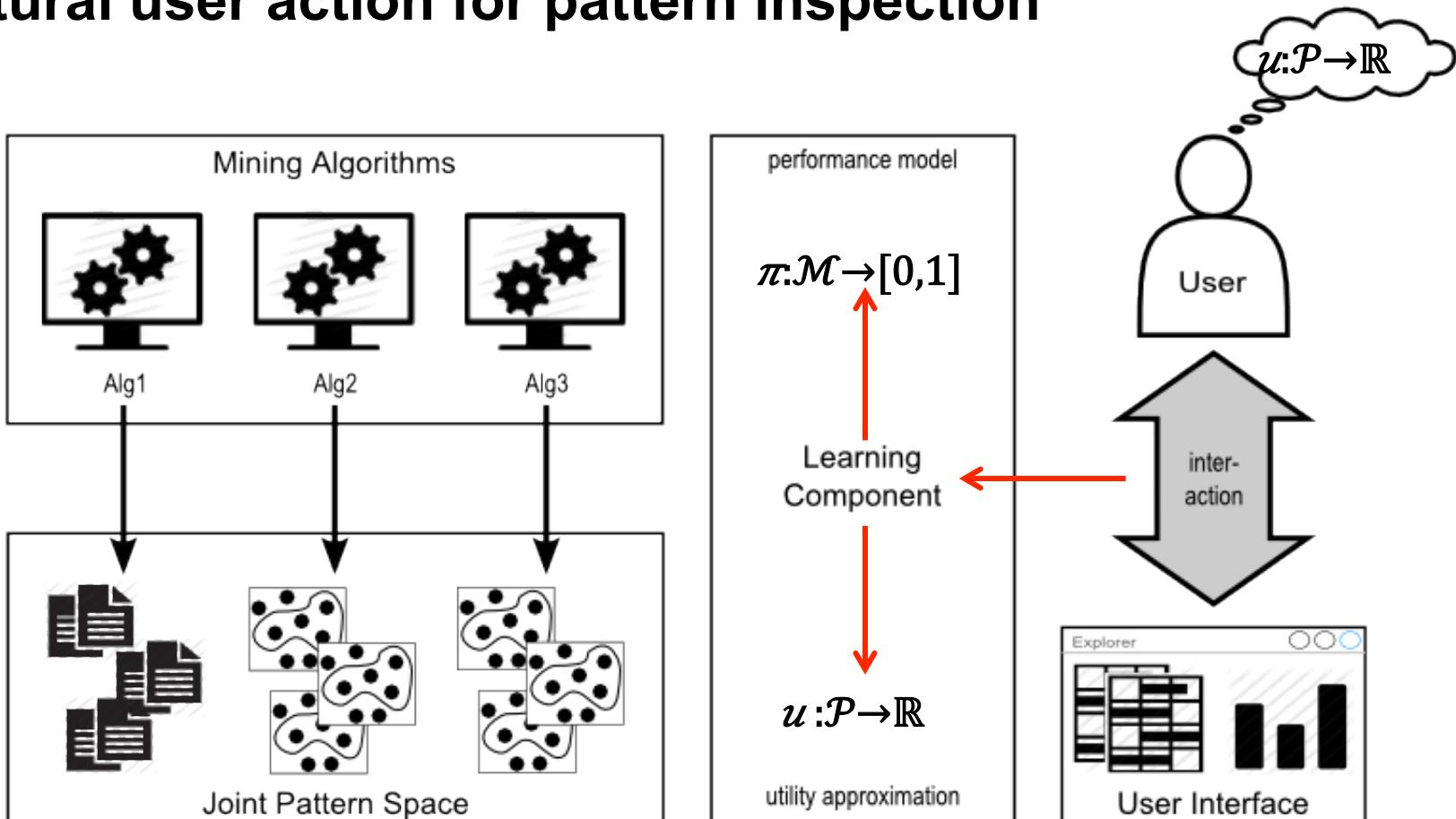
# Learning component controls mining algorithms in order to maximize output in terms of utility



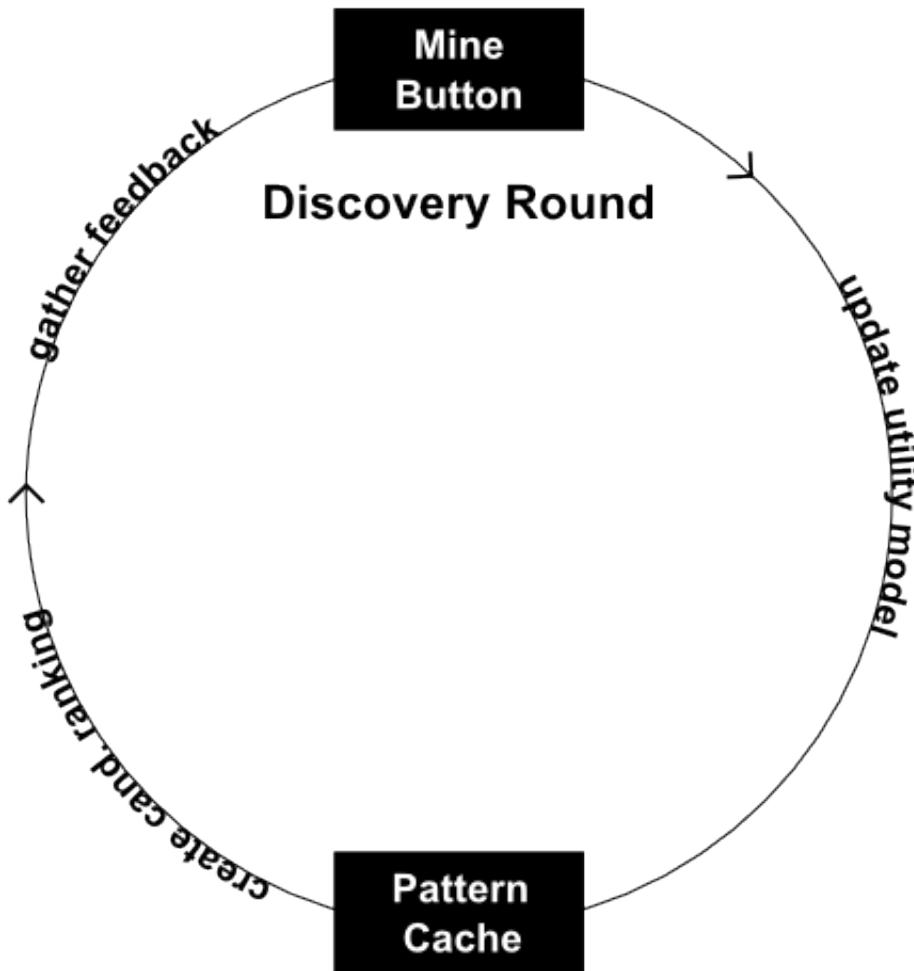
# Learning component updates models only based on natural user action for pattern inspection



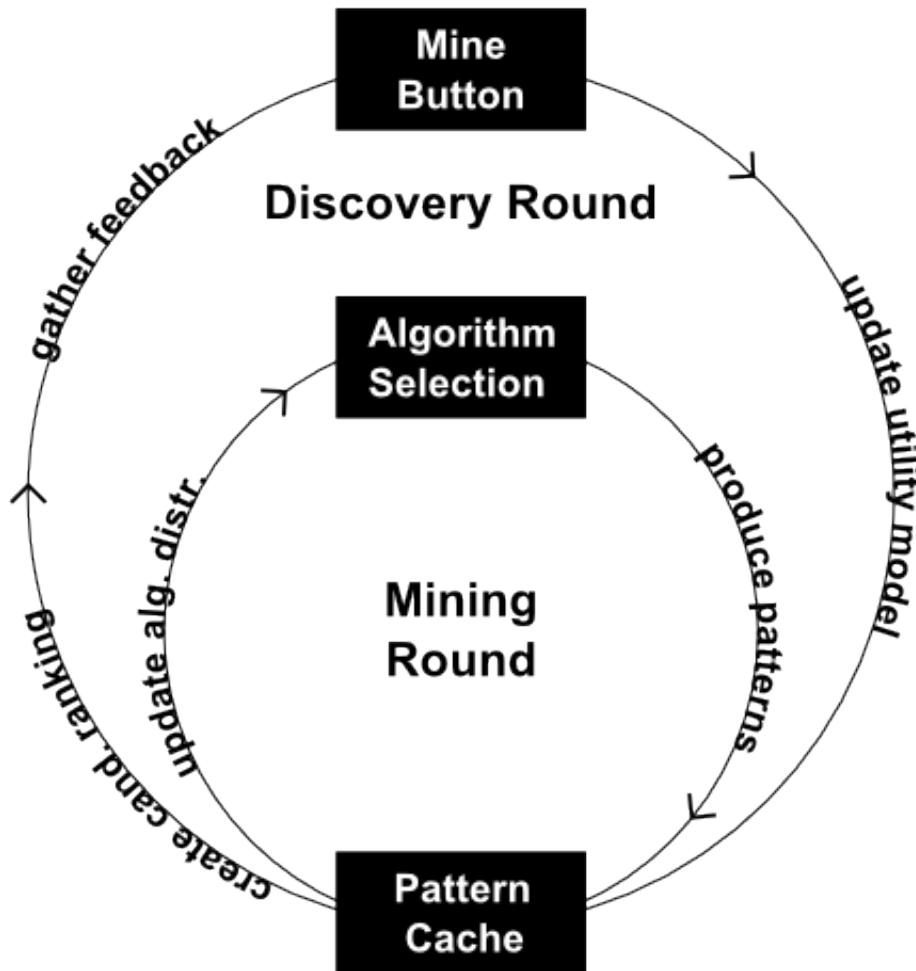
# Learning component updates models only based on natural user action for pattern inspection



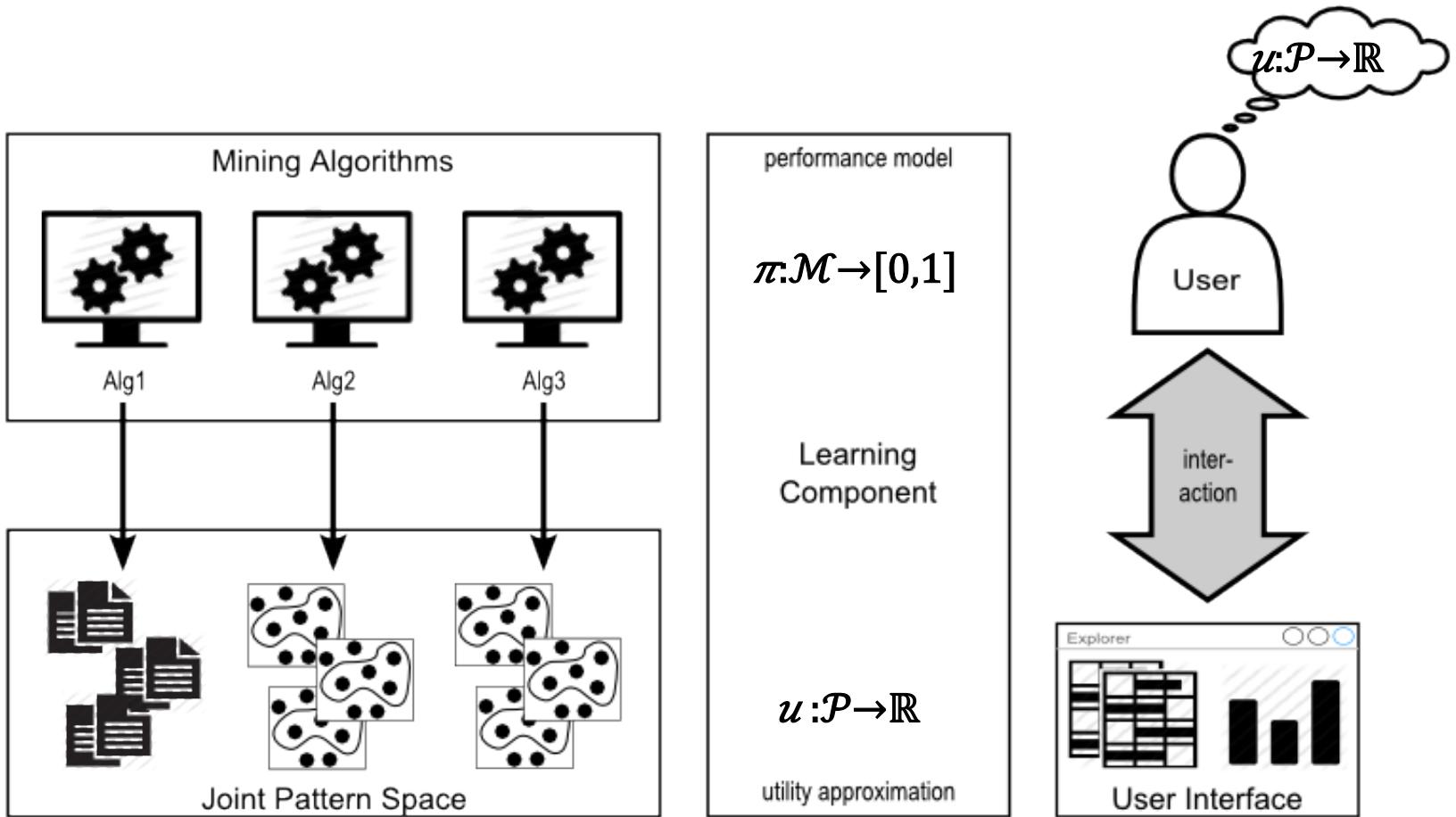
**One cycle of the kd process corresponds to the period between two *mine* clicks**



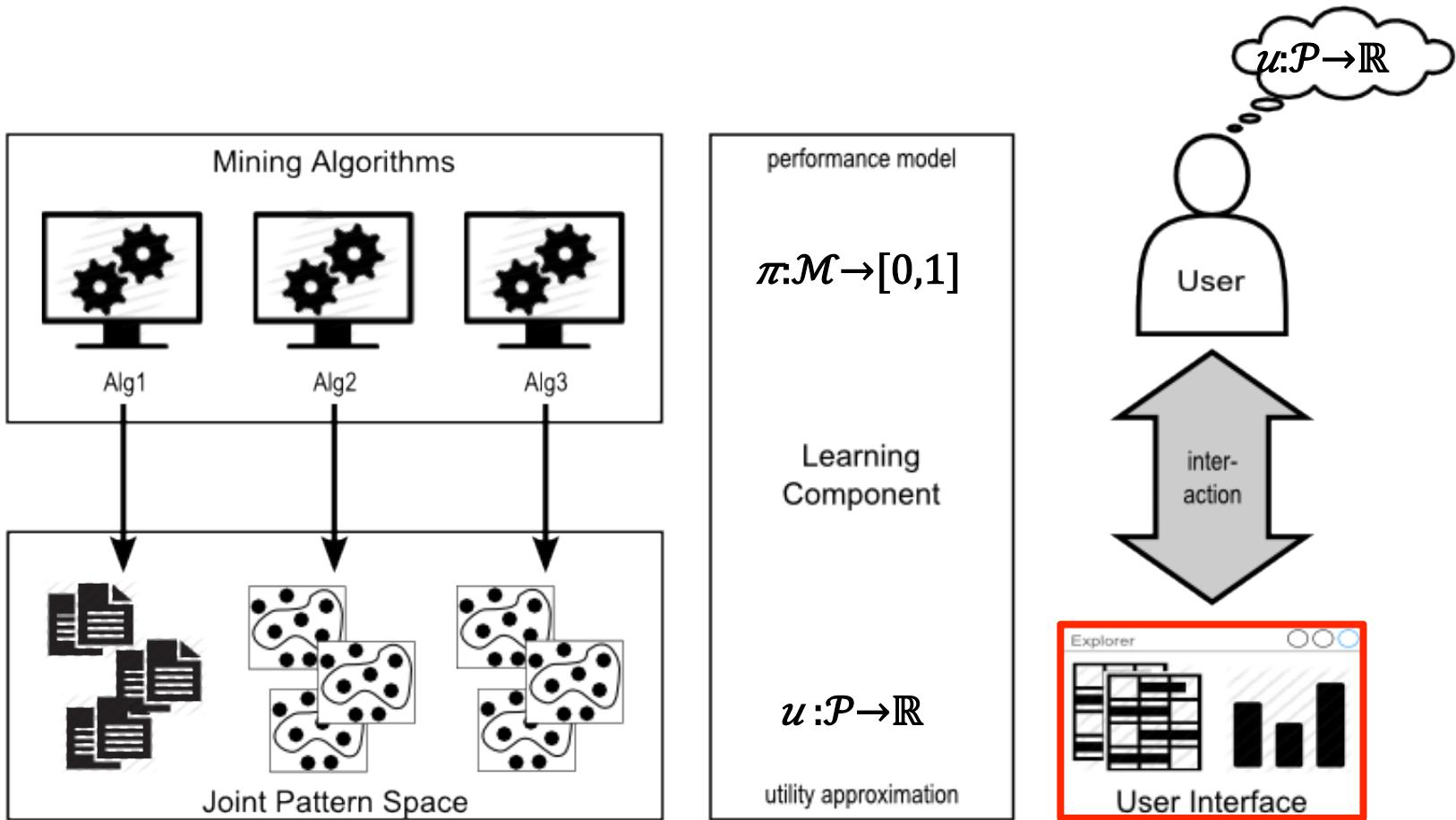
# During one user discovery round many mining rounds can be executed in the background



# Overview



# Interface



# Implicit feedback can be gathered by simple UI

Bonn Click Mining
[Login](#)
[Researcher view](#)

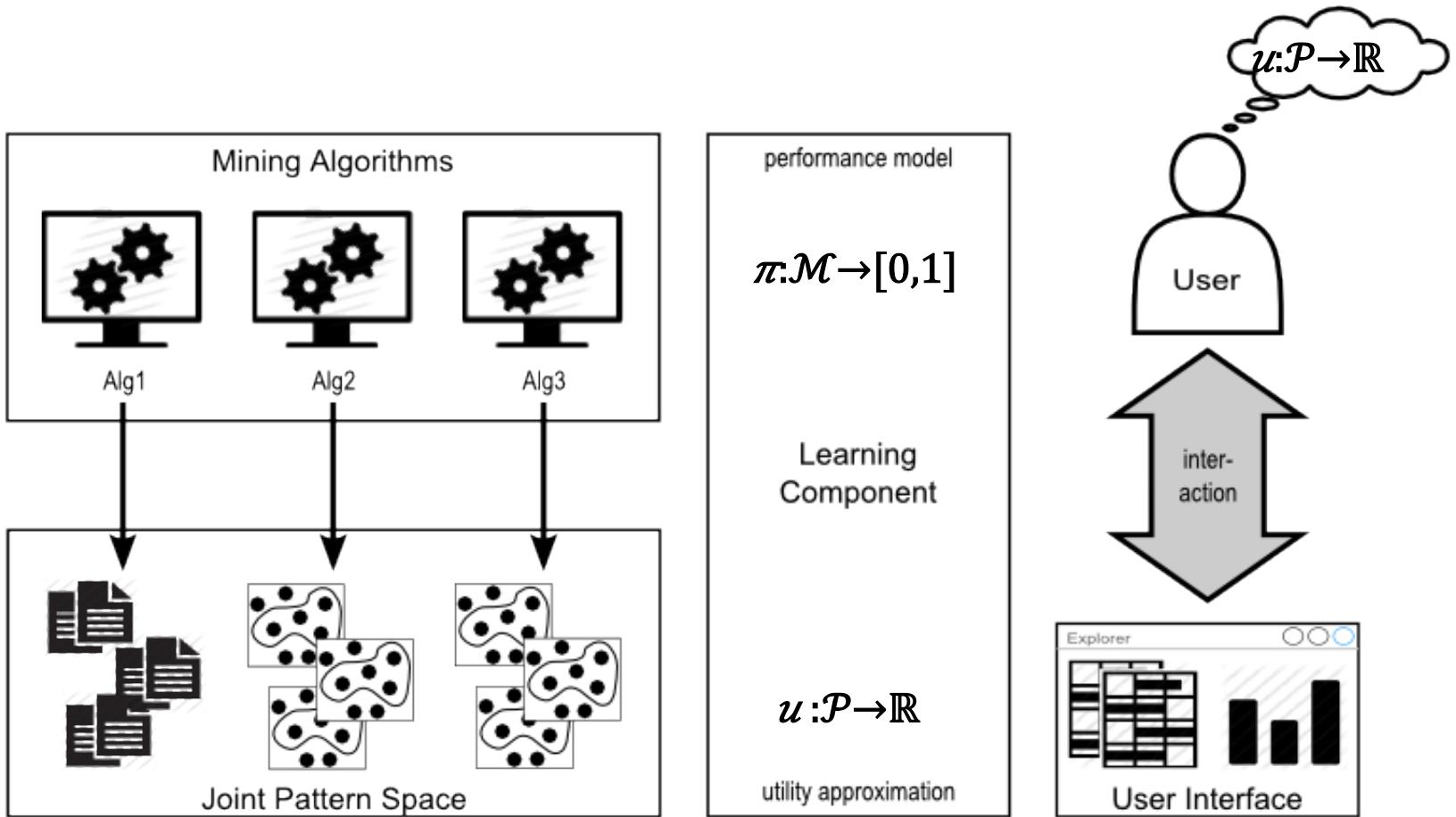
A One-Click Mining Prototype by [KDML Group](#), University of Bonn and Fraunhofer IAIS.

You are working on table: Socio-economics of Germany

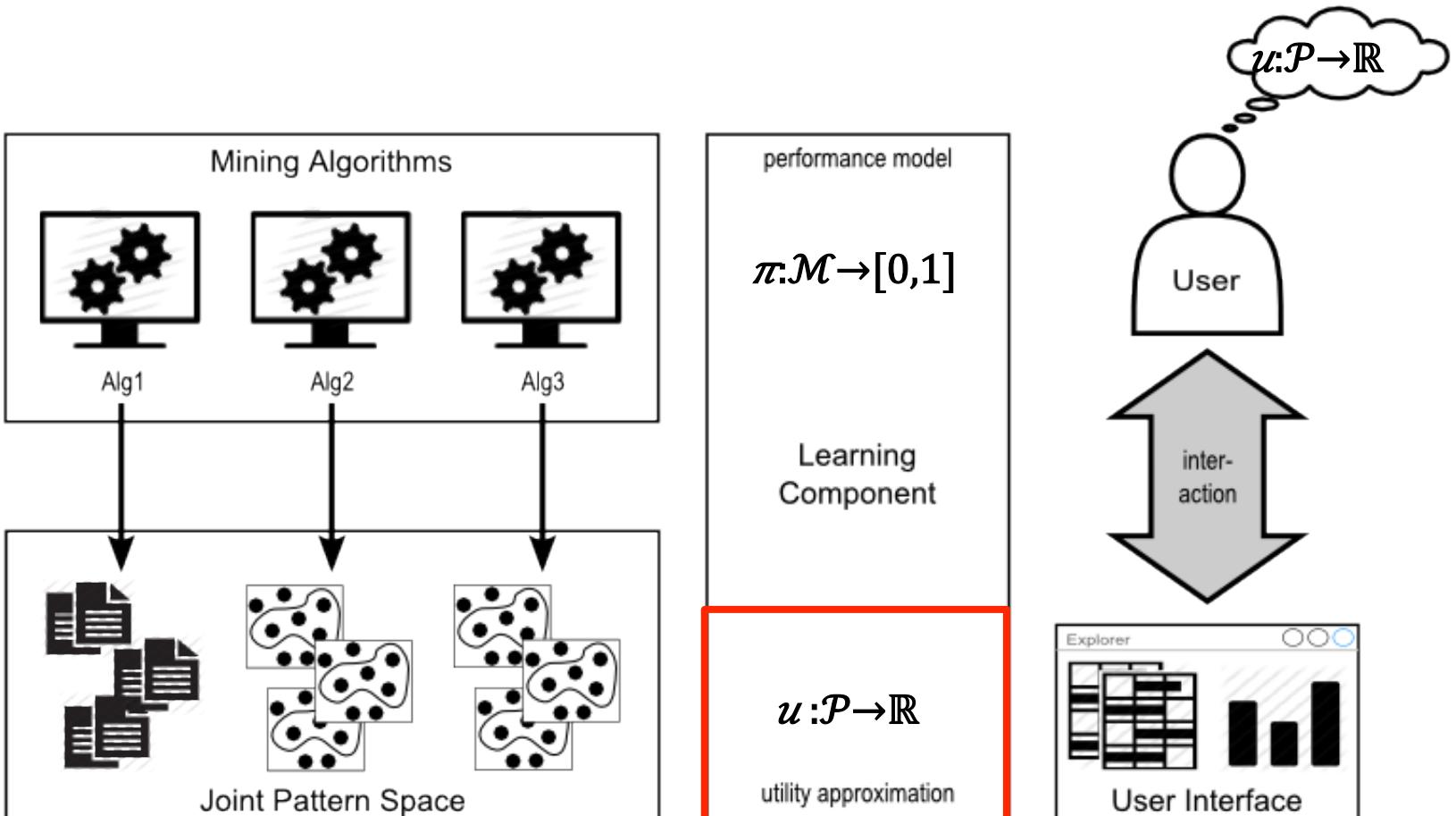
Area Code	Area Name	Type	State	Region	CDU Gain	SPD Gain	FDP Gain	GREEN Gain	LEFT Gain	Elect. Particptn. Gain	CDU	SPD	FDP	GREEN	LEFT	Elect. Particptn
9173	Bad Toelz-Wolfratshausen	Rural	Bavaria	South	-9.2	-5.9	5.8	2.4	2.2	-6.8	46.7	12	17.1	11.2	4.6	73.1
9188	Sternberg	Rural	Bavaria	South	-7.7	-6.2	6.3	2.2	1.6	-4.3	39.2	14.1	22.1	14.7	3.7	80
9175	Ebersberg	Rural	Bavaria	South	-7.9	-7.5	5.3	2.8	1.7	-5.9	42.4	14.9	16.9	13.1	4.2	77.7
9172	Berchtesgadener Land	Rural	Bavaria	South	-7.9	-6.9	5	4	2.1	-8.5	50.7	12.3	13.2	10.6	4.9	68.2
9177	Erding	Rural	Bavaria	South	-9.5	-7.9	5.3	4.6	1.9	-6.2	45.5	12.4	14.7	12	4.8	73.3
9184	Munich (district)	Rural	Bavaria	South	-5.5	-7.4	5	2.1	1.9	-4.3	39.8	16.7	19.6	12.7	4.5	79.1
9182	Miesbach	Rural	Bavaria	South	-6.7	-7	4.8	2.6	1.5	-6.5	48.1	12.2	17.6	10.2	3.9	73.8
9185	Neuburg-Schrobenhausen	Rural	Bavaria	South	-5.2	-8.9	5.7	2.4	2.7	-7.4	52.6	13.2	13.5	7.2	5.7	70.1
9186	Pfaffenhofen a.d.Ilm	Rural	Bavaria	South	-4.8	-10	4.8	2.7	2.4	-6.7	48.3	13.7	14.1	9.1	5.6	71.6
9572	Erlangen-Hoechstadt	Rural	Bavaria	South	-5.7	-9.9	5.8	2.6	3	-5.3	39.1	20.5	15.1	11.5	6	77.3
9773	Dillingen a.d.Donau	Rural	Bavaria	South	-8	-8.1	6.5	2.5	3	-8.7	47.3	13.2	16.2	7.5	6	68.4
8315	Breisgau-Hochschwarzwald	Rural	Baden-Wuerttemberg	South	-3.9	-10.8	6.6	2.7	2.9	-6.3	34.1	19.8	17.4	16.7	6.4	74.7
9779	Donau-Ries	Rural	Bavaria	South	-7.8	-8.7	6.2	2.6	3.3	-7.1	49.1	15.1	13.5	7.8	5.8	72.2
8337	Waldshut	Rural	Baden-Wuerttemberg	South	-4.7	-11.5	8.3	2.6	3.2	-7	37.8	18.6	18.9	12.1	6.3	68.4
9174	Dachau	Rural	Bavaria	South	-7.8	-6.8	5.3	2.2	1.9	-6	45.9	14.7	15.5	10.6	4.6	75.1
9373	Neumarkt i.d.OPf.	Rural	Bavaria	South	-8.8	-7.8	5.4	3.7	2.7	-6.5	49.3	13.4	12.9	8.9	5.7	73.7
9180	Garmisch-Partenkirchen	Rural	Bavaria	South	-8.6	-5.5	5.5	2.1	1.5	-6.8	51.6	10.2	17.1	8.7	3.8	72.1

!!!

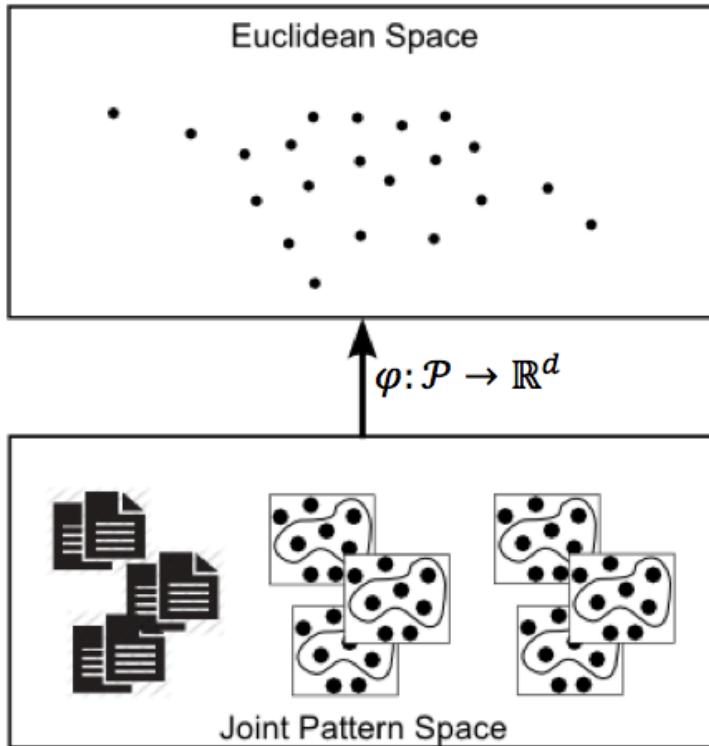
# Overview



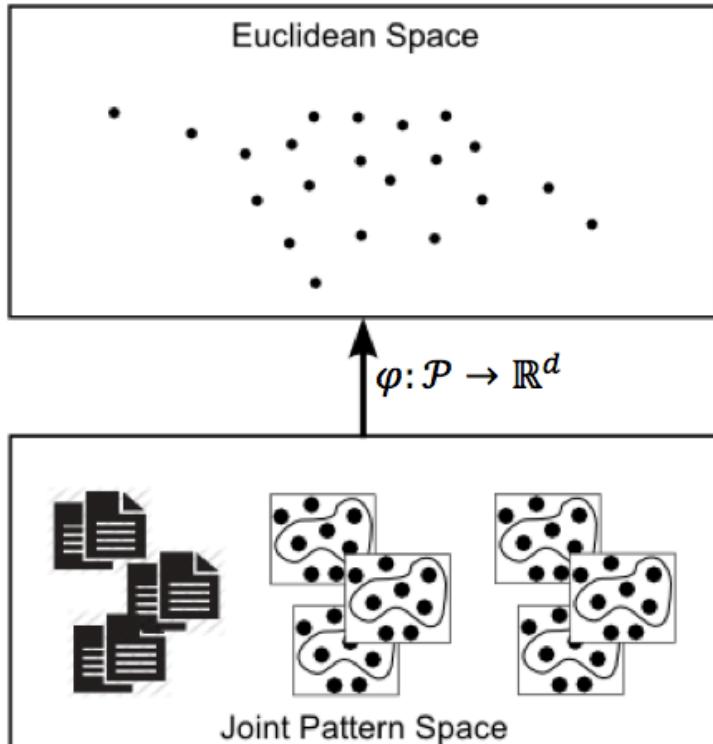
# Utility model



# Can be solved as *online learning* problem with appropriate features

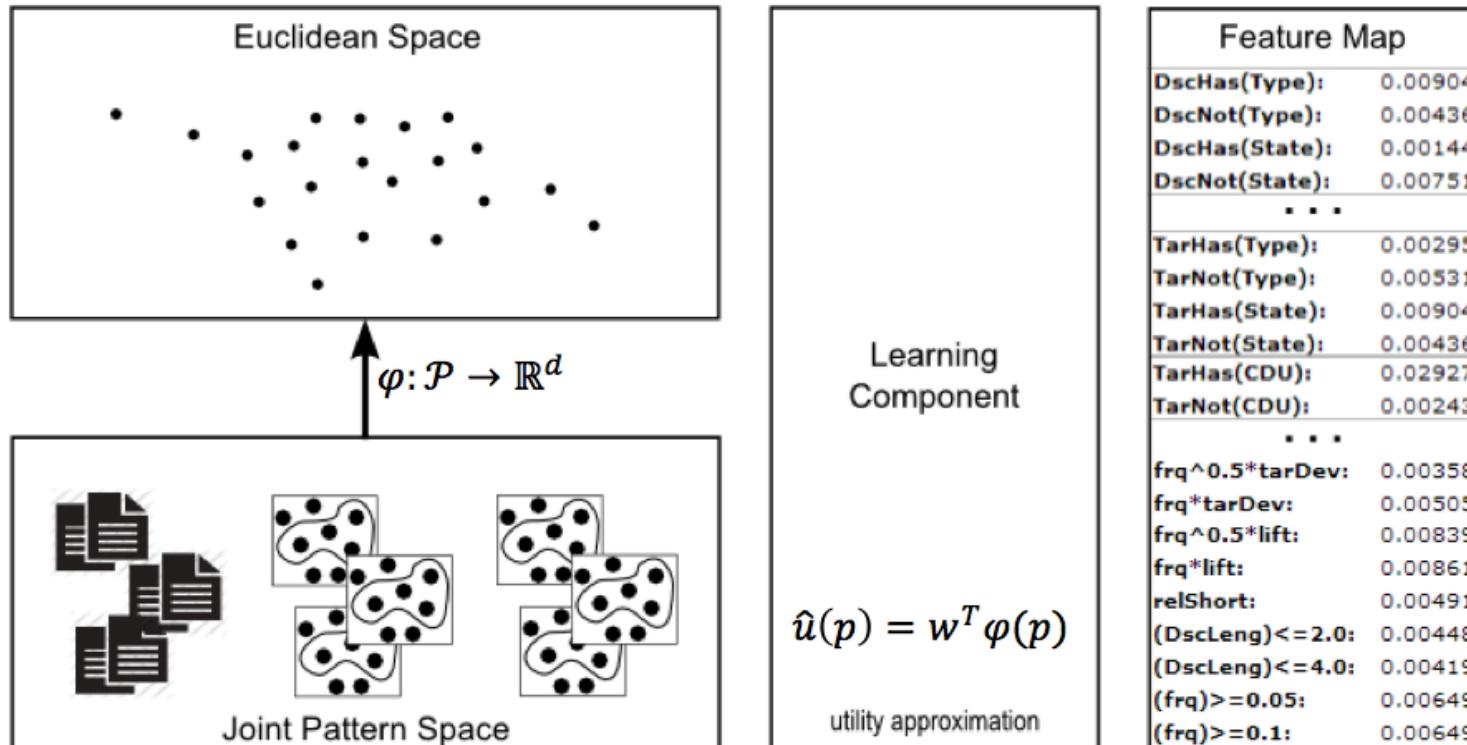


# Can be solved as *online learning* problem with appropriate features

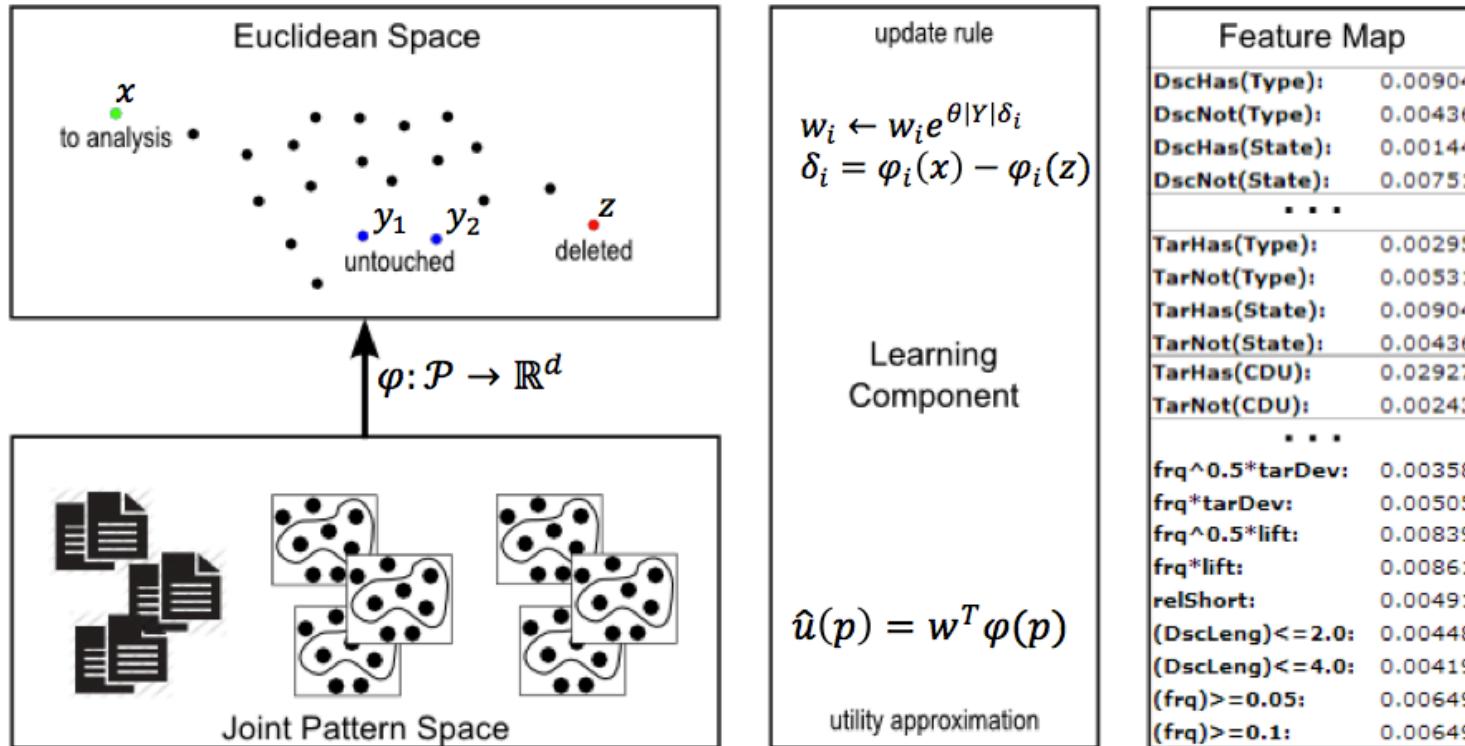


Feature Map	
DscHas(Type):	0.00904
DscNot(Type):	0.00436
DscHas(State):	0.00144
DscNot(State):	0.00751
***	
TarHas(Type):	0.00295
TarNot(Type):	0.00531
TarHas(State):	0.00904
TarNot(State):	0.00436
TarHas(CDU):	0.02927
TarNot(CDU):	0.00243
***	
frq <sup>0.5</sup> *tarDev:	0.00358
frq*tarDev:	0.00505
frq <sup>0.5</sup> *lift:	0.00839
frq*tarDev:	0.00861
relShort:	0.00491
(DscLeng)<=2.0:	0.00448
(DscLeng)<=4.0:	0.00419
(frq)>=0.05:	0.00649
(frq)>=0.1:	0.00649

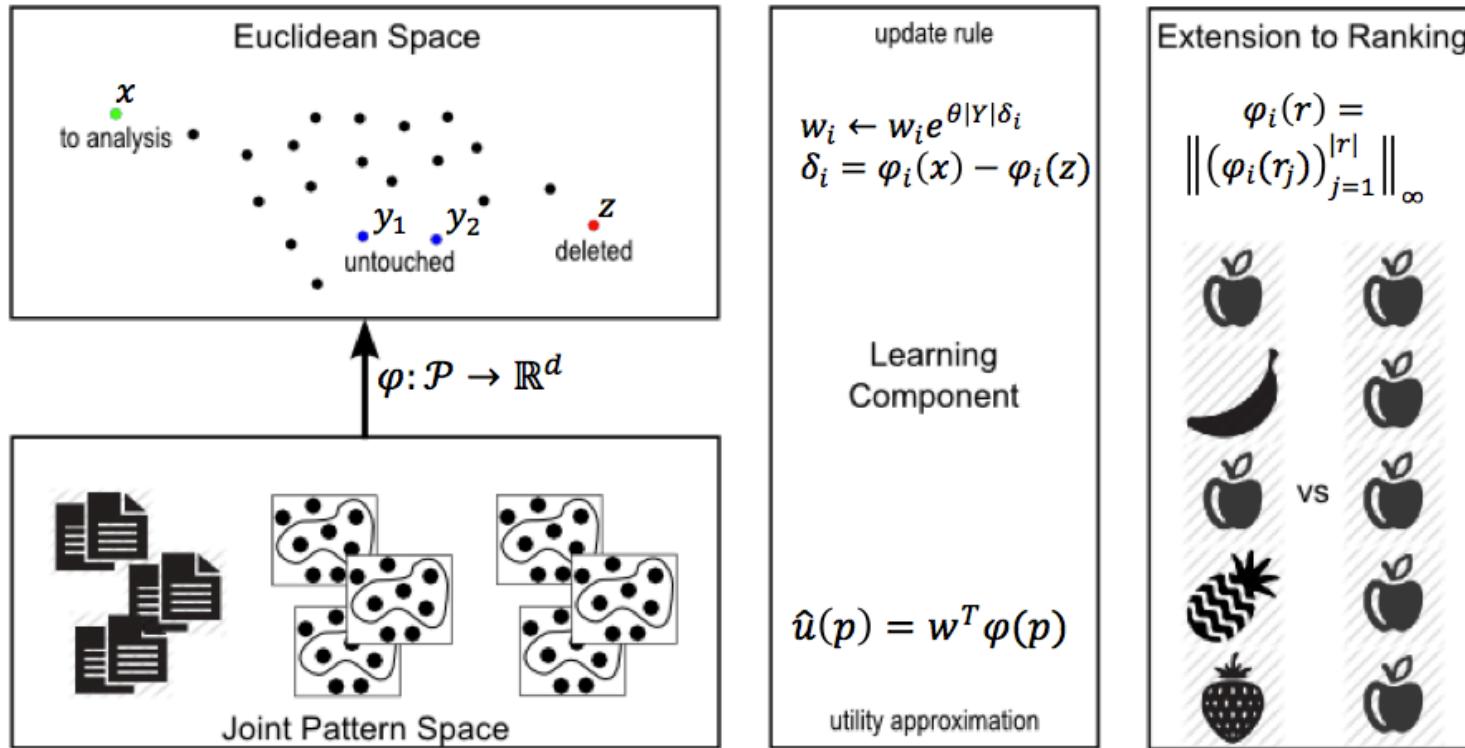
# Can be solved as *online learning* problem with appropriate features



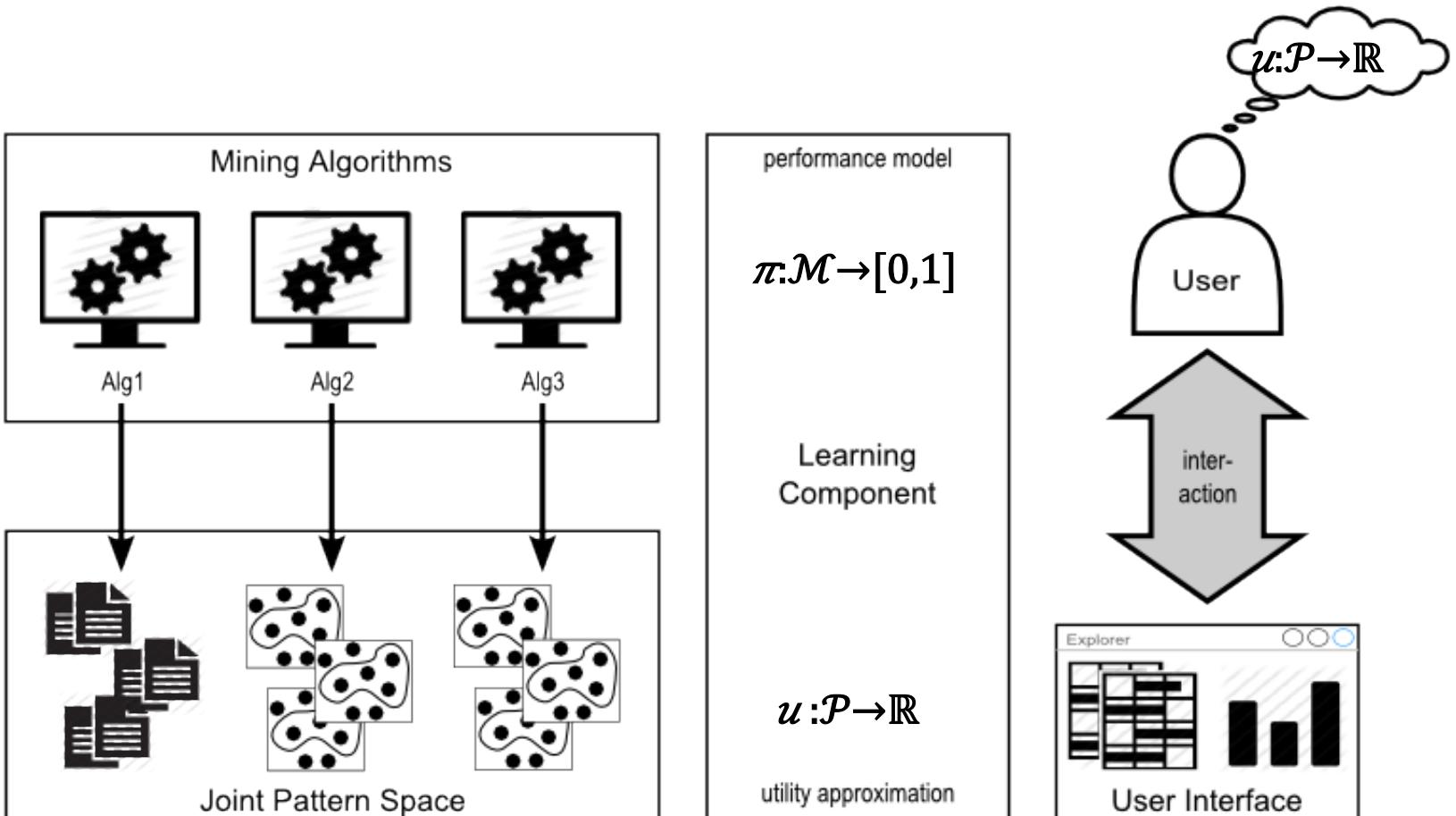
# Can be solved as *online learning* problem with appropriate features



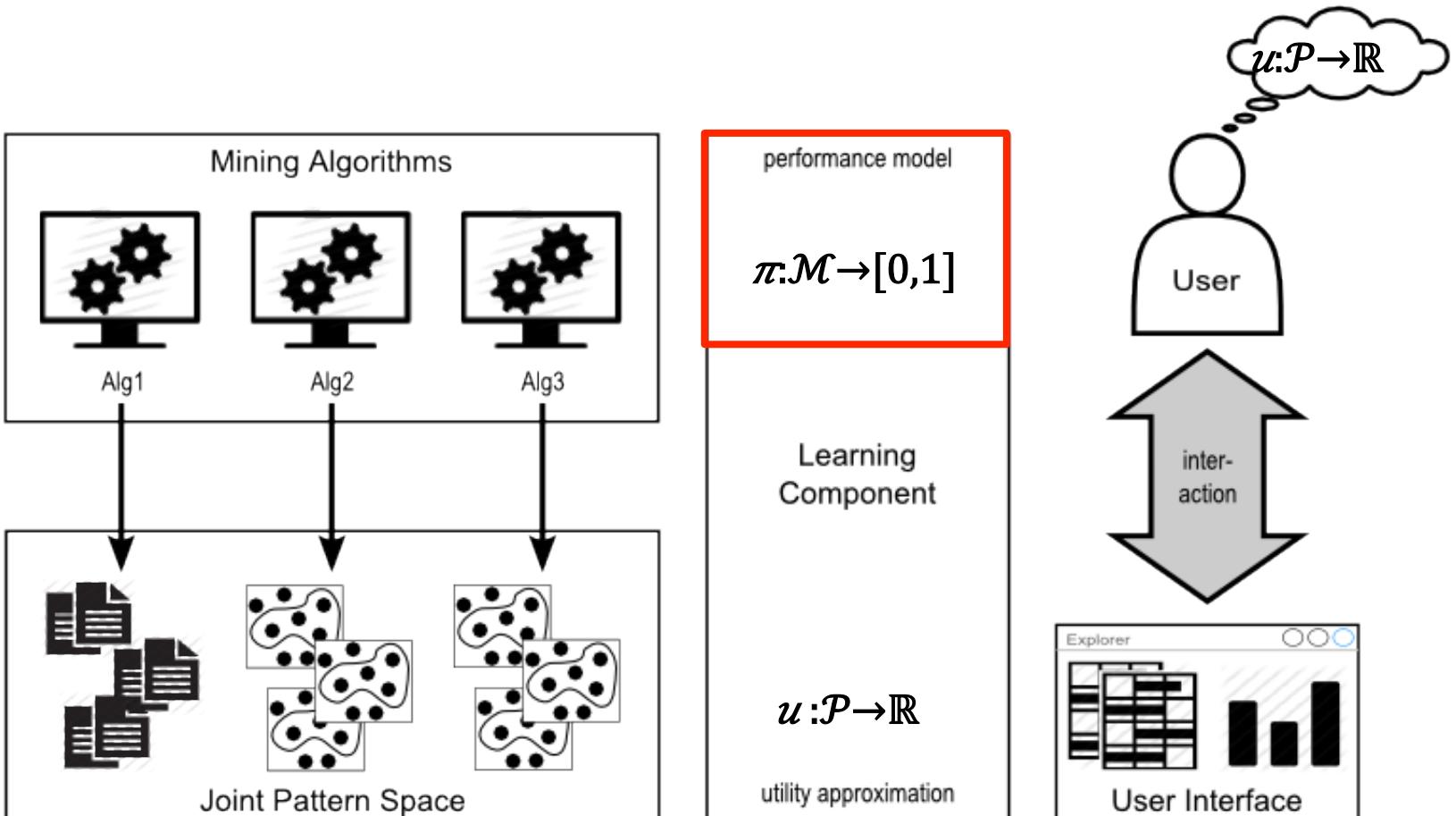
# Ranking score that balances *relevance* and *diversity* is crucial within user-supported search



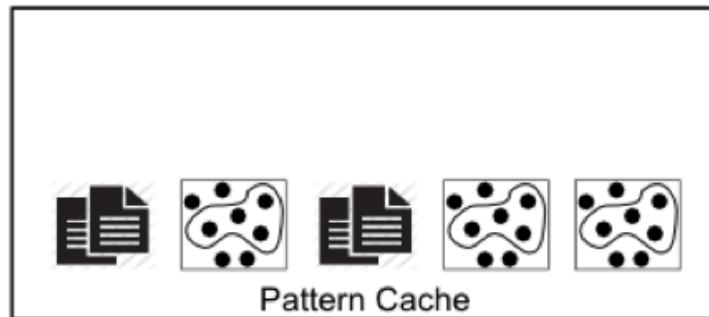
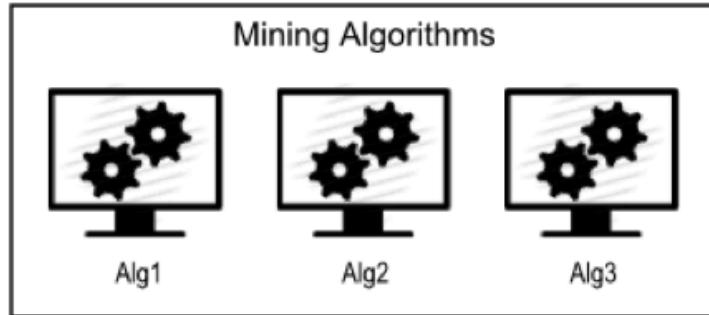
# Overview



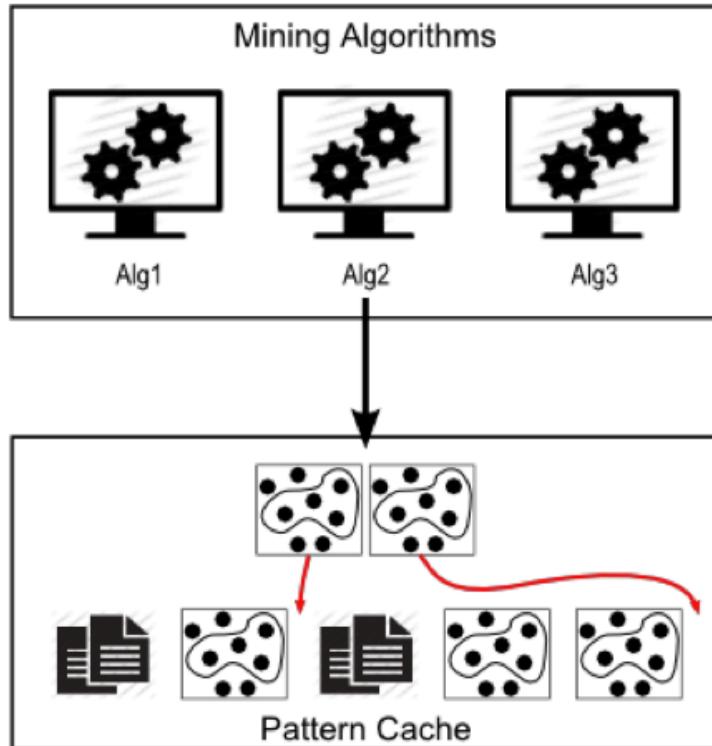
# Performance model



# Algorithm management can be modeled as *multi-armed bandit problem* with shifting rewards

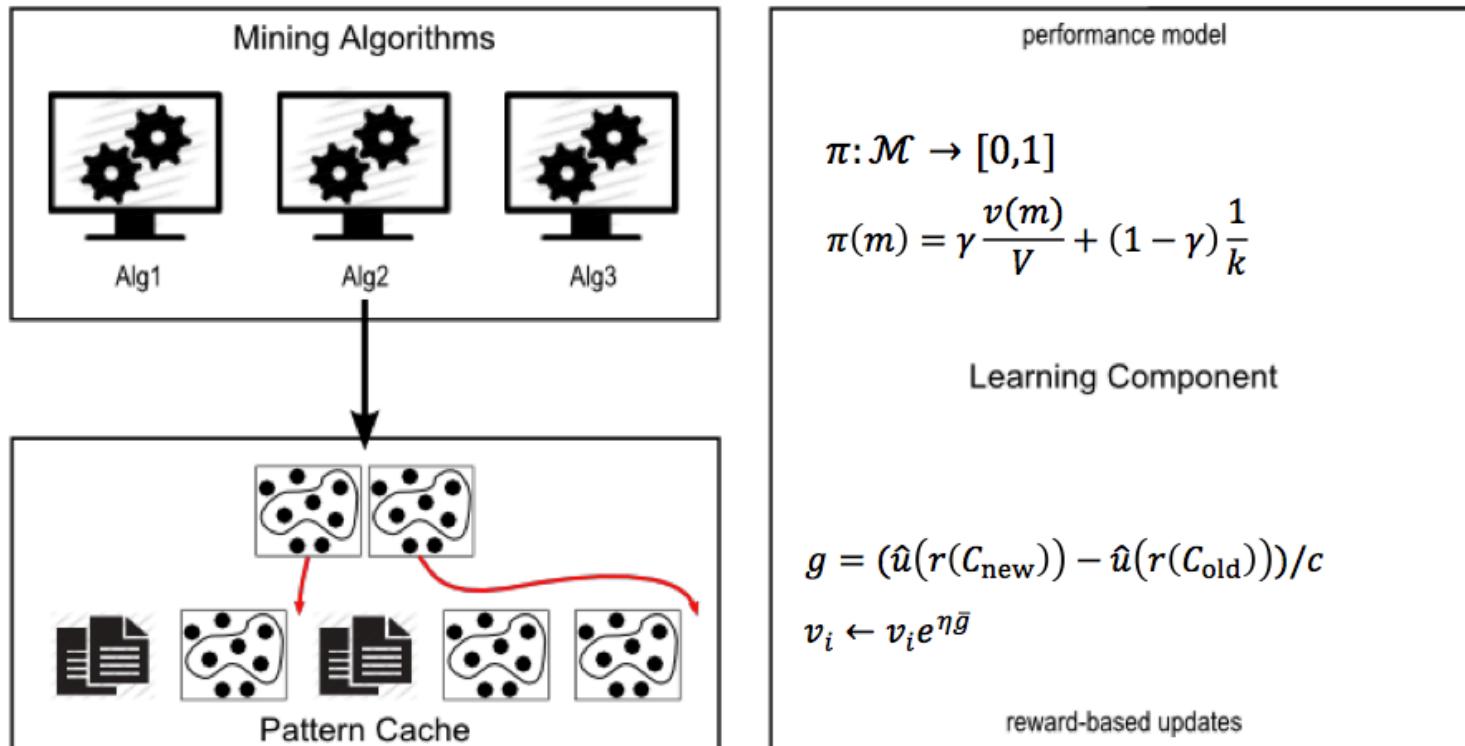


# Algorithm management can be modeled as *multi-armed bandit problem* with shifting rewards

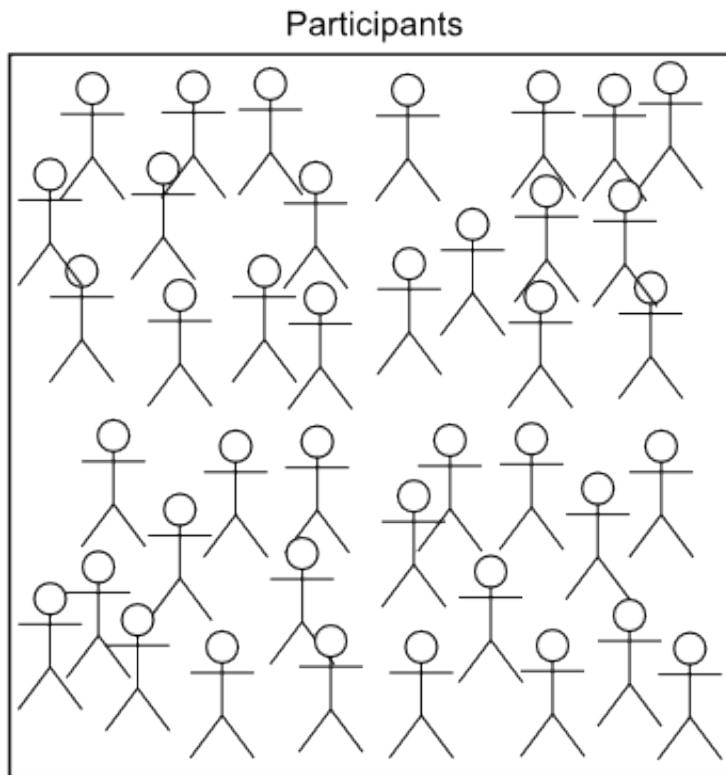


$$g = (\hat{u}(r(C_{\text{new}})) - \hat{u}(r(C_{\text{old}}))) / c$$

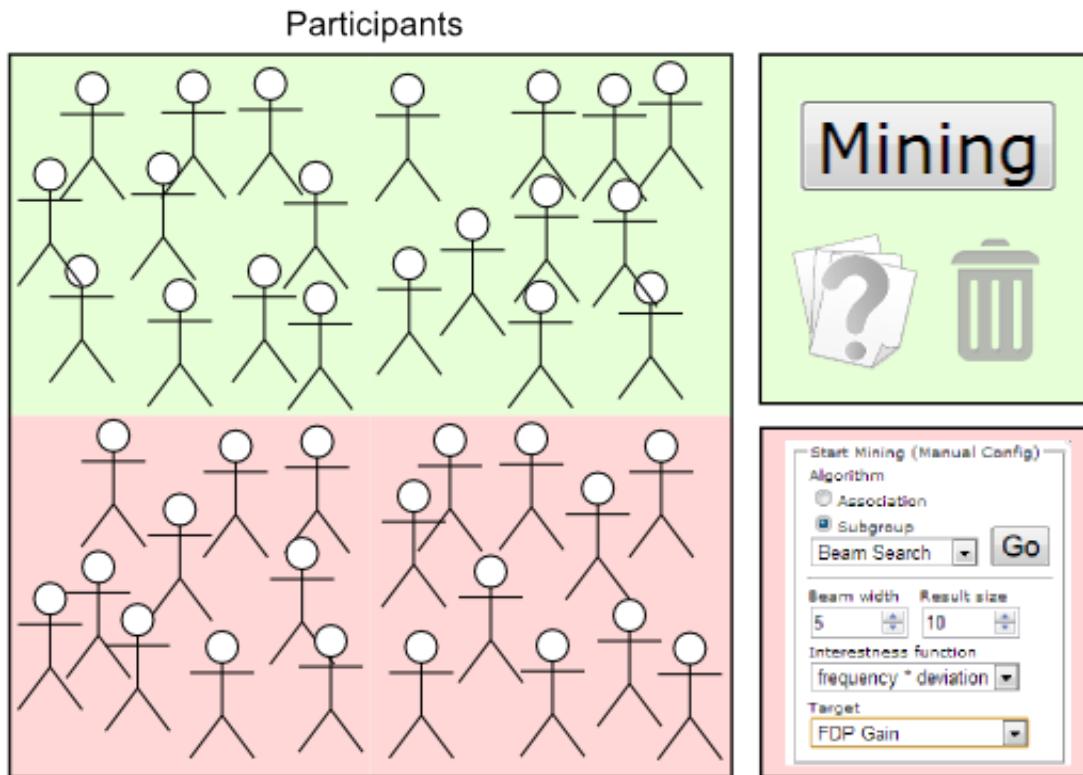
# Algorithm management can be modeled as *multi-armed bandit problem* with shifting rewards



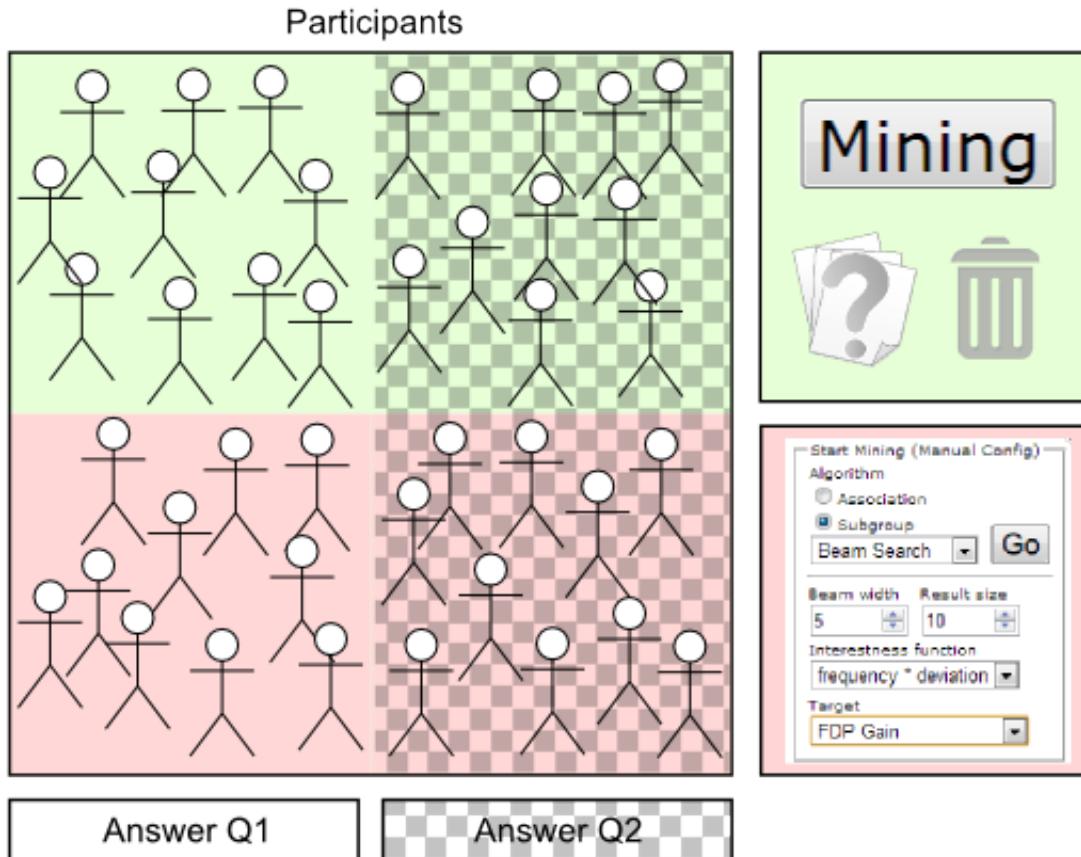
# Controlled experiment with real users is planned to evaluate effectiveness of approach



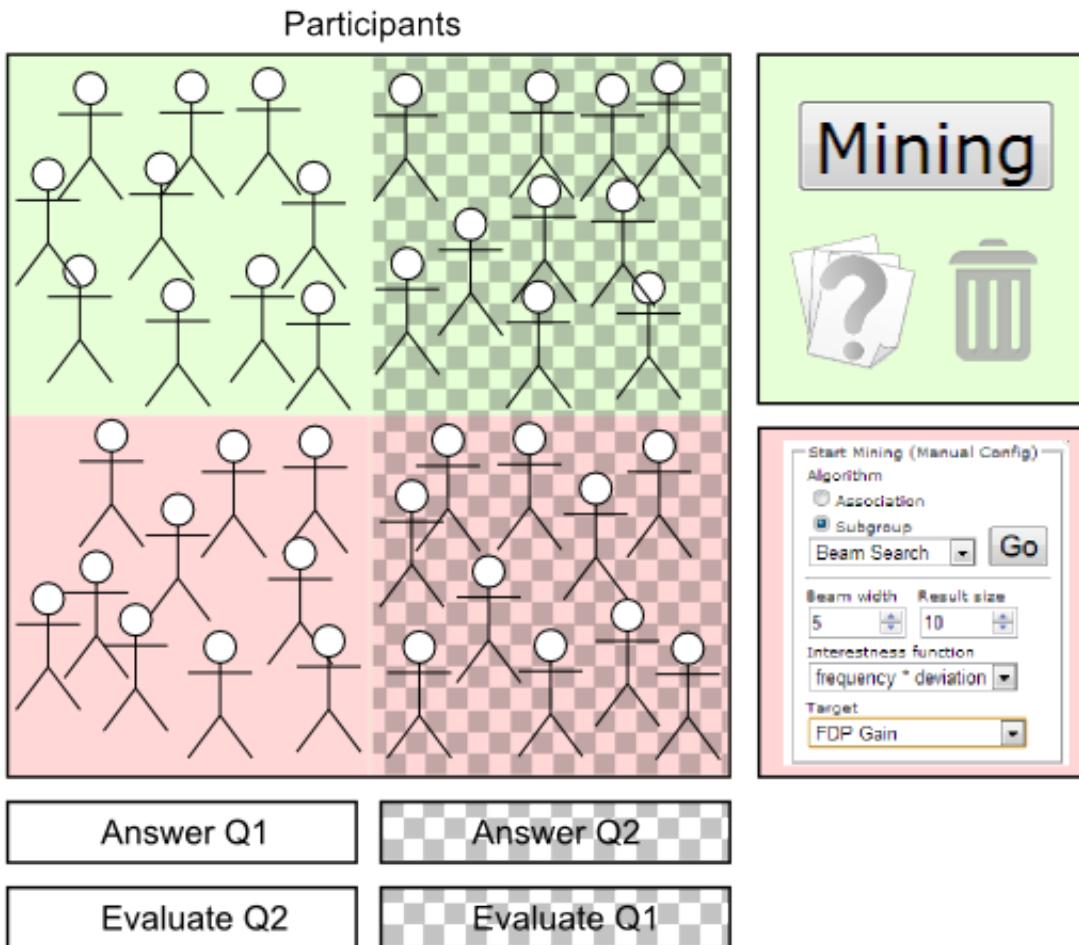
# We are interested in performance of the system compared to *manual variant*



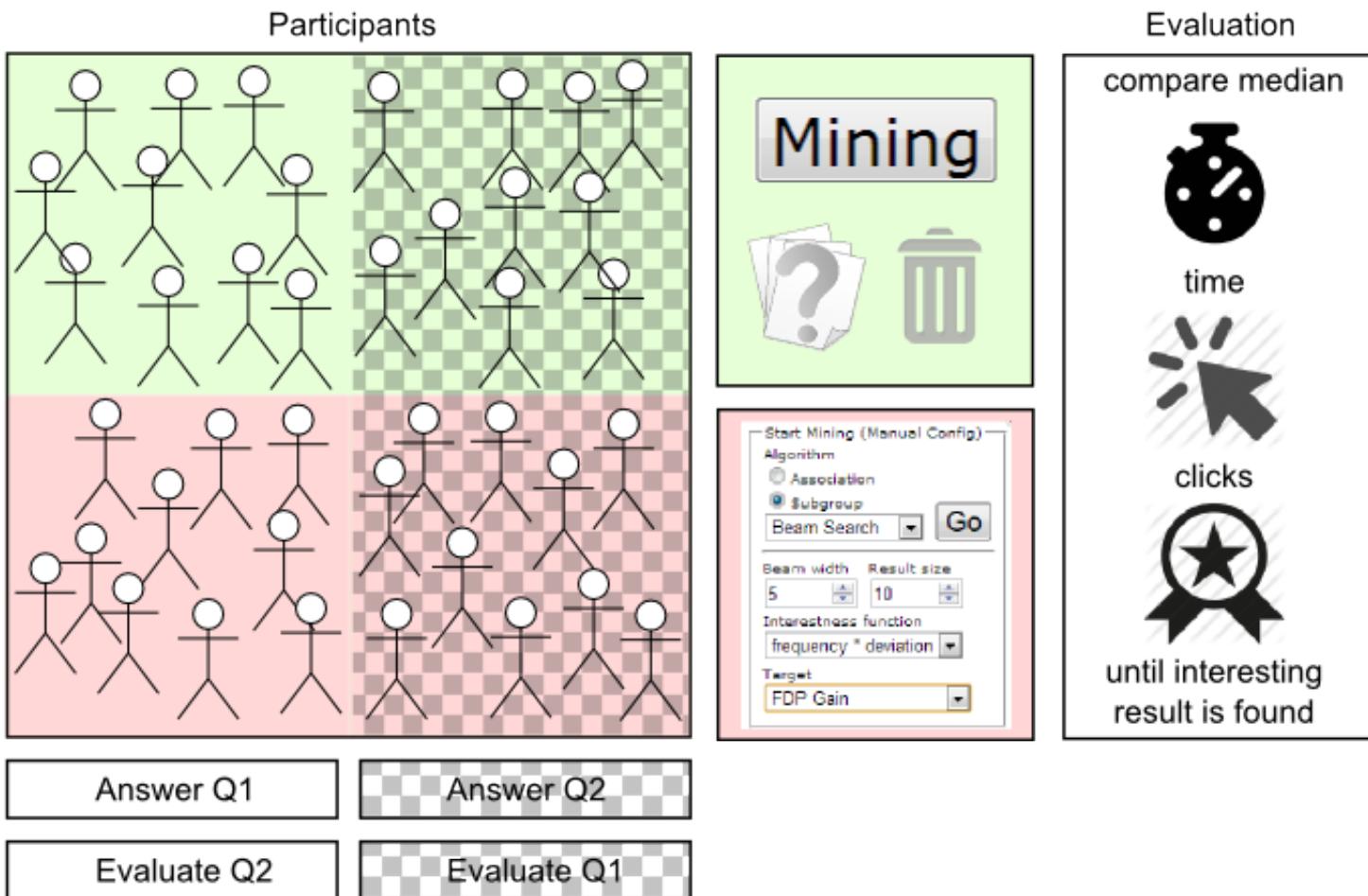
# Users are further split into groups according to different analysis questions they have to tackle



# After producing answers (patterns) to questions; quality of solutions is assessed by members of other group



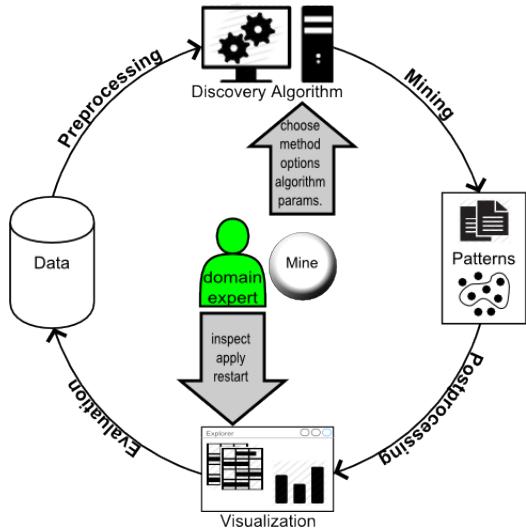
# Setup allows to compute expressive evaluation metrics



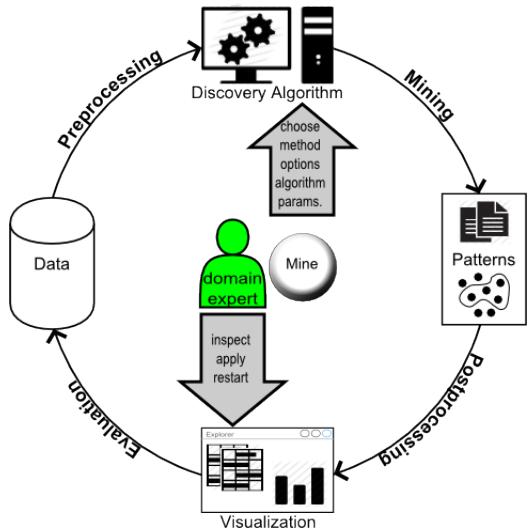
# Current Work

- **Evaluation**
- **Improve user interface**
- **Embedding representation**

# Summary



# Summary



Bonn Click Mining  
A One-Click Mining Prototype by KOML Group, University of Bonn and Fraunhofer IAIS.  
You are working on table: Socio-economics of Germany

Area Code	Area Name	Type	State	Region	CDU Gain	SPD Gain	FDP Gain	GREEN Gain	LEFT Gain	Elect. Particpnt. Gain	CDU	SPD	FDP	GREEN	LEFT	Elect. Particpnt.
9173	Bad Tölz-Wolfratshausen	Rural	Bavaria	South	-9.2	-5.9	5.8	2.4	2.2	-6.8	46.7	12	17.1	11.2	4.6	73.1
9181	Starnberg	Rural	Bavaria	South	-7.7	-6.2	6.3	2.2	1.6	-4.3	39.2	14.1	22.1	14.7	3.7	80
9175	Ebersberg	Rural	Bavaria	South	-7.9	-7.5	5.3	2.8	1.7	-5.9	42.4	14.9	16.9	13.1	4.2	77.7
9172	Berchtesgadener Land	Rural	Bavaria	South	-7.9	-6.9	5	4	2.1	-8.5	50.7	12.3	13.2	10.6	4.9	68.2
9177	Erding	Rural	Bavaria	South	-9.5	-7.9	5.3	4.6	1.9	-6.2	45.5	12.4	14.7	12	4.8	73.3
9184	Munich (district)	Rural	Bavaria	South	-5.5	-7.4	5	2.1	1.9	-4.3	39.8	16.7	19.6	12.7	4.5	79.1
9176	Eichstätt	Rural	Bavaria	South	-2.8	-10.8	4.4	2.4	2.6	-5.9	51.4	15.7	11.2	7.8	5.3	75.3
9182	Miesbach	Rural	Bavaria	South	-6.7	-7	4.8	2.6	1.5	-6.5	48.1	12.2	17.6	10.2	3.9	73.8
9185	Neuburg-Schrobenhausen	Rural	Bavaria	South	-5.2	-8.9	5.7	2.4	2.7	-7.4	52.6	13.2	13.5	7.2	5.7	70.1
9186	Pfaffenhofen a.d.Ilm	Rural	Bavaria	South	-4.8	-10	4.8	2.7	2.4	-6.7	48.3	13.7	14.1	9.1	5.6	71.6
9189	Traunstein	Rural	Bavaria	South	-9.2	-7.4	4.5	4.7	2.3	-7.9	47.7	12.7	12.8	12.1	5.1	70.3
9572	Erlangen-Höchstadt	Rural	Bavaria	South	-5.7	-9.9	5.8	2.6	3	-5.3	39.1	20.5	15.1	11.5	6	77.3
9773	Dillingen a.d.Donau	Rural	Bavaria	South	-8	-8.1	6.5	2.5	3	-8.7	47.3	13.2	16.2	7.5	6	68.4
8315	Breisgau-Hochschwarzwald	Rural	Baden-Württemberg	South	-3.9	-10.8	6.6	2.7	2.9	-6.3	34.1	19.8	17.4	16.7	6.4	74.7
8311	Freiburg im Breisgau	Urban	Baden-Württemberg	South	-2.2	-9.6	4.4	-0.8	4.5	-5.2	24.2	21.6	12.8	25.4	10	74.1
9779	Donau-Ries	Rural	Bavaria	South	-7.8	-8.7	6.2	2.6	3.3	-7.1	49.1	15.1	13.5	7.8	5.8	72.2
9679	Würzburg	Rural	Bavaria	South	-4.6	-8.6	4.5	3.2	2.3	-4.9	41.3	18.9	13.5	12.1	6	78.7

Region: East  
Unemploy.=high  
Children Pop.=low  
CDU Gain=high  
Frequency: 0.211165  
Deviation of LEFT: 0.162206  
Pattern Mean: 28.848276  
Global Mean: 12.504369

Subgroup  
Region=East  
Unemploy.=high  
Income=low  
Frequency: 0.211165  
Dev. of LEFT: 0.162206

Subgroup  
CDU=low  
LEFT Gain=high  
Frequency: 0.310680  
Dev. of SPD Gain: 0.225062

Subgroup  
SPD Gain=low  
LEFT=high  
Frequency: 0.342233  
Dev. of LEFT Gain: 0.463916

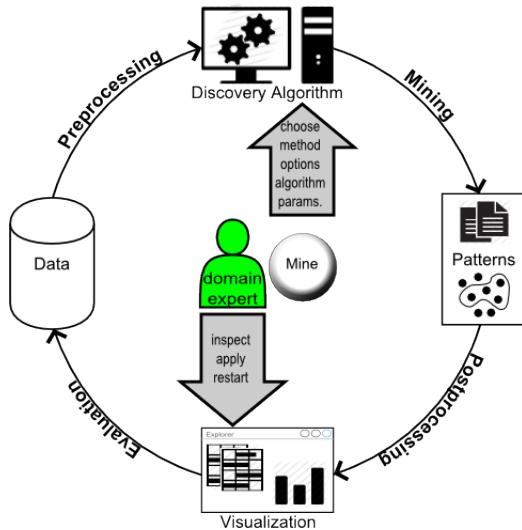
Subgroup  
Constr. workf.=low  
Housing workf.=high  
GDP growth=2009=low  
Frequency: 0.152913  
Dev. of GDP growth 2008: 0.120925

Subgroup  
Agricul. workf.=low  
sectoral workf.=high  
Finance workf.=high  
Frequency: 0.288835  
Dev. of Pop. Density: 0.001365

Subgroup  
Elect. Particpnt.=low  
GREEN=low  
Frequency: 0.322816  
Dev. of Elect. Particpnt. Gain: 0.406395

Subgroup  
CDU=low  
LEFT Gain=high  
Frequency: 0.310680  
Dev. of SPD Gain: 0.225062

# Summary



**Bonn Click Mining**  
A One-Click Mining Prototype by KOML Group, University of Bonn and Fraunhofer IAIS.  
You are working on table: Socio-economics of Germany

Area Code	Area Name	Type	State	Region	CDU Gain	SPD Gain	FDP Gain	GREEN Gain	LEFT Gain	Elect. Particpnt. Gain	CDU	SPD	FDP	GREEN	LEFT	Elect. Particpnt.
9173	Bad Tölz-Wolfratshausen	Rural	Bavaria	South	-9.2	-5.9	5.8	2.4	2.2	-6.8	46.7	12	17.1	11.2	4.6	73.1
9181	Starnberg	Rural	Bavaria	South	-7.7	-6.2	6.3	2.2	1.6	-4.3	39.2	14.1	22.1	14.7	3.7	80
9175	Ebersberg	Rural	Bavaria	South	-7.9	-7.5	5.3	2.8	1.7	-5.9	42.4	14.9	16.9	13.1	4.2	77.7
9172	Berchtesgadener Land	Rural	Bavaria	South	-7.9	-6.9	5	4	2.1	-8.5	50.7	12.3	13.2	10.6	4.9	68.2
9177	Erding	Rural	Bavaria	South	-9.5	-7.9	5.3	4.6	1.9	-6.2	45.5	12.4	14.7	12	4.8	73.3
9184	Munich (district)	Rural	Bavaria	South	-5.5	-7.4	5	2.1	1.9	-4.3	39.8	16.7	19.6	12.7	4.5	79.1
9176	Eichstätt	Rural	Bavaria	South	-2.8	-10.8	4.4	2.4	2.6	-5.9	51.4	15.7	11.2	7.8	5.3	75.3
9182	Miesbach	Rural	Bavaria	South	-6.7	-7	4.8	2.6	1.5	-6.5	48.1	12.2	17.6	10.2	3.9	73.8
9185	Neuburg-Schrobenhausen	Rural	Bavaria	South	-5.2	-8.9	5.7	2.4	2.7	-7.4	52.6	13.2	13.5	7.2	5.7	70.1
9186	Pfaffenhofen a.d.Ilm	Rural	Bavaria	South	-4.8	-10	4.8	2.7	2.4	-6.7	48.3	13.7	14.1	9.1	5.6	71.6
9189	Traunstein	Rural	Bavaria	South	-9.2	-7.4	4.5	4.7	2.3	-7.9	47.7	12.7	12.8	12.1	5.1	70.3
9572	Erlangen-Nürnberg	Rural	Bavaria	South	-5.7	-9.9	5.8	2.6	3	-5.3	39.1	20.5	15.1	11.5	6	77.3
9773	Dillingen a.d.Donau	Rural	Bavaria	South	-8	-8.1	6.5	2.5	3	-8.7	47.3	13.2	16.2	7.5	6	68.4
8315	Breisgau-Hochschwarzwald	Rural	Baden-Württemberg	South	-3.9	-10.8	6.6	2.7	2.9	-6.3	34.1	19.8	17.4	16.7	6.4	74.7
8311	Freiburg im Breisgau	Urban	Baden-Württemberg	South	-2.2	-9.6	4.4	-0.8	4.5	-5.2	24.2	21.6	12.8	25.4	10	74.1
9779	Donau-Ries	Rural	Bavaria	South	-7.8	-8.7	6.2	2.6	3.3	-7.1	49.1	15.1	13.5	7.8	5.8	72.2
9679	Würzburg	Rural	Bavaria	South	-4.6	-8.6	4.5	3.2	2.3	-4.9	41.3	18.9	13.5	12.1	6	78.7

**Subgroup**  
SPD Gain=low  
LEFT=high  
Frequency : 0.342233  
Dev. of LEFT Gain: 0.463916

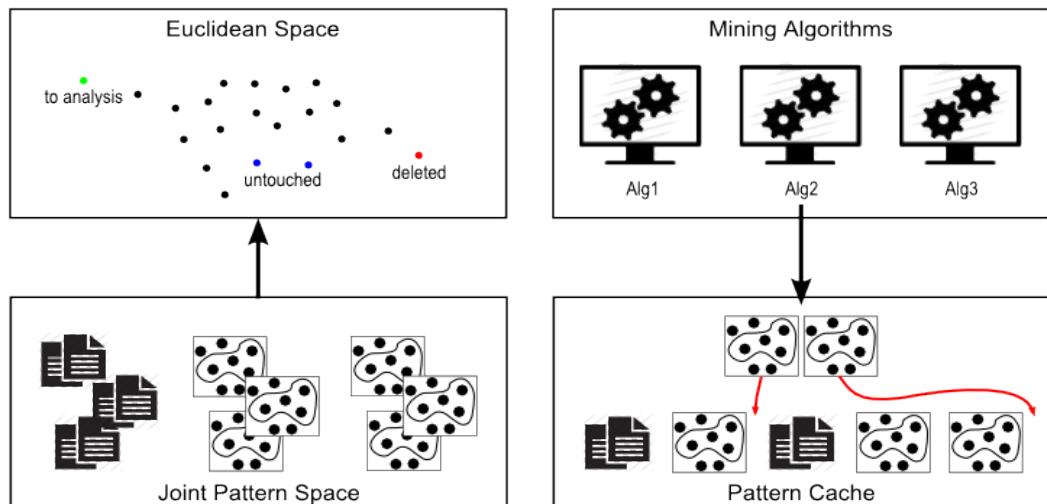
**Subgroup**  
Constr. workf.=low  
Housing workf.=high  
GDP growth 2009=low  
...  
Frequency : 0.152913  
Dev. of GDP growth 2008: 0.120925

**Subgroup**  
Agricul. workf.=low  
sector workf.=high  
Finance workf.=high  
Frequency : 0.288835  
Dev. of Pop. Density: 0.001365

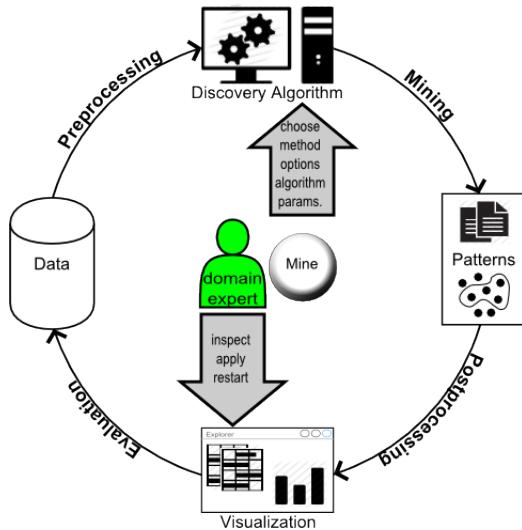
**Subgroup**  
Elect. Particpnt.=low  
GREEN=low  
Frequency : 0.322816  
Dev. of Elect. Particpnt. Gain: 0.406395

**Subgroup**  
CDU=low  
LEFT Gain=high  
Frequency : 0.310680  
Dev. of SPD Gain : 0.225062

**Explorer** icon with a checkmark and a trash can icon.



# Summary



**Bonn Click Mining**  
A One-Click Mining Prototype by KOML Group, University of Bonn and Fraunhofer IAIS.  
You are working on table: Socio-economics of Germany

Area Code	Area Name	Type	State	Region	CDU Gain	SPD Gain	FDP Gain	GREEN Gain	LEFT Gain	Elect. Particptn. Gain	CDU	SPD	FDP	GREEN	LEFT	Elect. Participants
9173	Bad Tölz-Wolfratshausen	Rural	Bavaria	South	-9.2	-5.9	5.8	2.4	2.2	-6.8	46.7	12	17.1	11.2	4.6	73.1
9181	Starnberg	Rural	Bavaria	South	-7.7	-6.2	6.3	2.2	1.6	-4.3	39.2	14.1	22.1	14.7	3.7	80
9175	Ebersberg	Rural	Bavaria	South	-7.9	-7.5	5.3	2.8	1.7	-5.9	42.4	14.9	16.9	13.1	4.2	77.7
9172	Berchtesgadener Land	Rural	Bavaria	South	-7.9	-6.9	5	4	2.1	-8.5	50.7	12.3	13.2	10.6	4.9	68.2
9177	Erding	Rural	Bavaria	South	-9.5	-7.9	5.3	4.6	1.9	-6.2	45.5	12.4	14.7	12	4.8	73.3
9184	Munich (district)	Rural	Bavaria	South	-5.5	-7.4	5	2.1	1.9	-4.3	39.8	16.7	19.6	12.7	4.5	79.1
9176	Eichstätt	Rural	Bavaria	South	-2.8	-10.8	4.4	2.4	2.6	-5.9	51.4	15.7	11.2	7.8	5.3	75.3
9182	Miesbach	Rural	Bavaria	South	-6.7	-7	4.8	2.6	1.5	-6.5	48.1	12.2	17.6	10.2	3.9	73.8
9185	Neuburg-Schrobenhausen	Rural	Bavaria	South	-5.2	-8.9	5.7	2.4	2.7	-7.4	52.6	13.2	13.5	7.2	5.7	70.1
9186	Pfaffenhofen a.d.Ilm	Rural	Bavaria	South	-4.8	-10	4.8	2.7	2.4	-6.7	48.3	13.7	14.1	9.1	5.6	71.6
9189	Traunstein	Rural	Bavaria	South	-9.2	-7.4	4.5	4.7	2.3	-7.9	47.7	12.7	12.8	12.1	5.1	70.3
9572	Erlangen-Nürnberg	Rural	Bavaria	South	-5.7	-9.9	5.8	2.6	3	-5.3	39.1	20.5	15.1	11.5	6	77.3
9773	Dillingen a.d.Donau	Rural	Bavaria	South	-8	-8.1	6.5	2.5	3	-8.7	47.3	13.2	16.2	7.5	6	68.4
8315	Breisgau-Hochschwarzwald	Rural	Baden-Württemberg	South	-3.9	-10.8	6.6	2.7	2.9	-6.3	34.1	19.8	17.4	16.7	6.4	74.7
8311	Freiburg im Breisgau	Urban	Baden-Württemberg	South	-2.2	-9.6	4.4	-0.8	4.5	-5.2	24.2	21.6	12.8	25.4	10	74.1
9779	Donau-Ries	Rural	Bavaria	South	-7.8	-8.7	6.2	2.6	3.3	-7.1	49.1	15.1	13.5	7.8	5.8	72.2
9679	Würzburg	Rural	Bavaria	South	-4.6	-8.6	4.5	3.2	2.3	-4.9	41.3	18.9	13.5	12.1	6	78.7

Frequency : 0.342233 Dev. of LEFT Gain: 0.463916

Subgroup

Constr. workf.=low Non-const. workf. GDP growth=2009=low ... Frequency : 0.152913 Dev. of GDP growth 2008: 0.120925

Subgroup

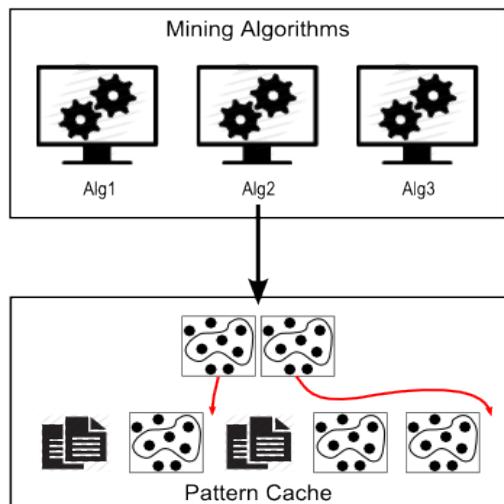
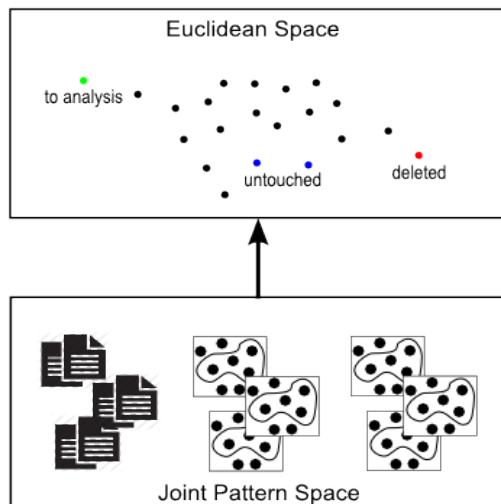
Agricul. workf.=low agri. area=high Finance workf.=high Frequency : 0.288835 Dev. of Pop. Density: 0.001365

Subgroup

Elect. Particptn.=low GREEN=low Frequency : 0.322816 Dev. of Elect. Particptn. Gain: 0.406395

Subgroup

CDU=low LEFT Gain=high Frequency : 0.310680 Dev. of SPD Gain: 0.225062



# We are hiring...



Post-doc position for 1y+ at mining and learning group of U. of Bonn, FhG

Working on one click mining; e.g., bringing in your analysis technique

Get in touch with me ([mario.boley@iais.fraunhofer.de](mailto:mario.boley@iais.fraunhofer.de))