



Кластерски приступ у развоју модула Информационог система за развој каријерног вођења

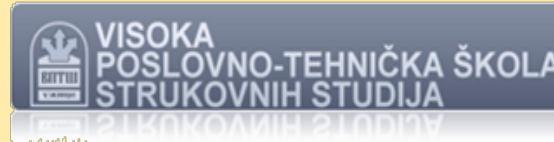
М. Миливојевић

milovan.milivojevic@vpts.edu.rs

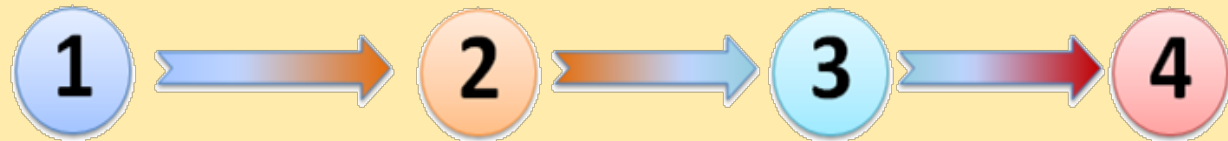
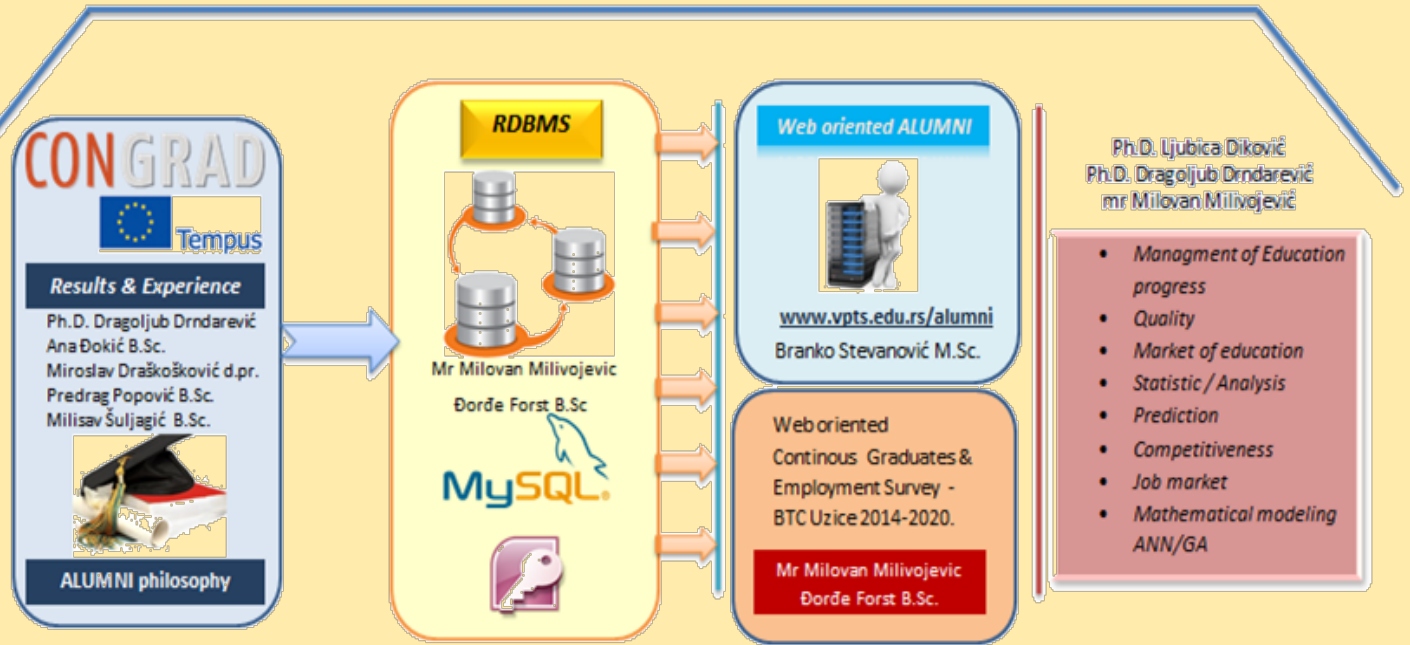


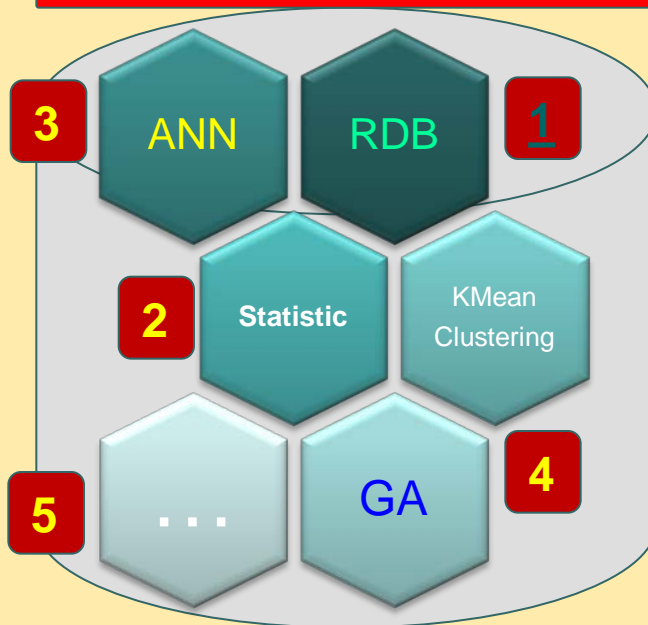
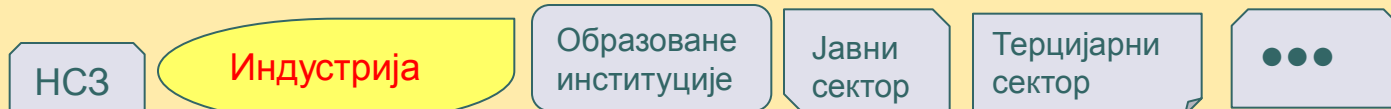
Висока
пословно-
техничка школа
Ужице
2015.





**BUSINESS and
TECHNICAL
COLLEGE – Uzice- 2013**





- RDB-Relation Data Bases
- Statistic (Pareto, Descriptive Statistic, Multiple Linear Regression, Statistic tests, Probability...)
- ANN-Artificial Neural network
- kMean Clustering / Genetic Algorithm
- ...

Информациони систем

Каријерног центра
ВПТШ, Ужице

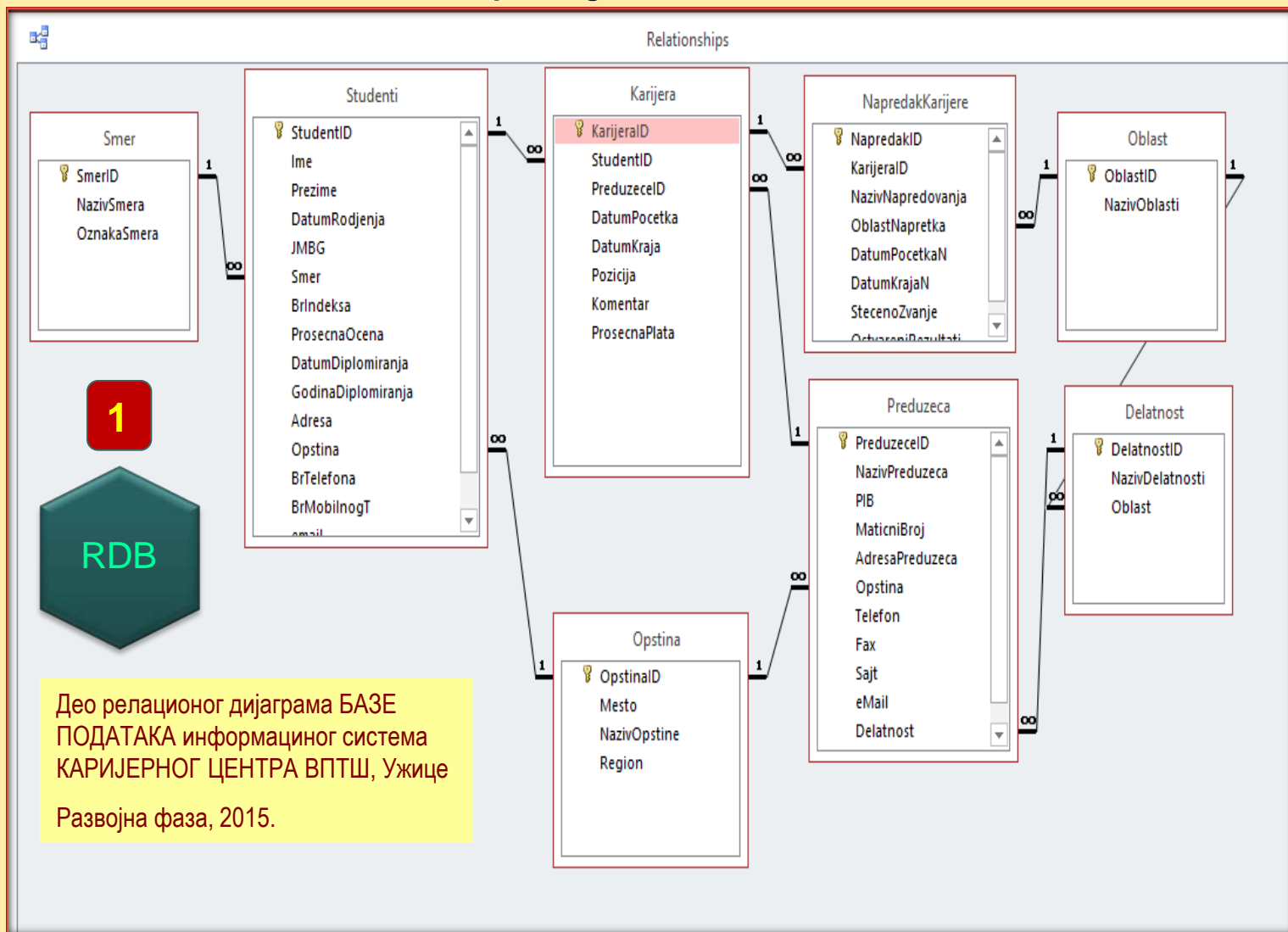
модуларни
концепт

поглед у
будућност

Каријерни центар ВПТШ, Ужице



Database Relationship Diagram



1

RDB





Application Main Window

Studenti
- □ ×

VISOKA
POSLOVNO-TEHNIČKA ŠKOLA
STRUKOVNIH STUDIJA

StudentID

Ime

Prezime

Datum rođenja

JMBG

Adresa

Opština

Broj fiksnog telefona

Broj mobilnog telefona

email

Slika

Smer

Broj indeksa

Prosečna ocena

DatumDiplomiranja

GodinaDiplomiranja

Лични подаци

Karijera

KarijeraID

PreduzeceID

Datum početka RO

Datum kraja RO

Pozicija

Komentar

Prosečna plata

ID	Naziv napredovanja	Oblast napretka	Datum početka	Datum kraja	Steceno zvanje	Ostvareni rezultati
1	Officer for student afa	IT	2.4.2009.		OSA	
####						

Изглед апликације за праћење пословне каријере студената основних и специјалистичких студија ВПТШ Ужице...

Подаци о напредовању на послу

Record: 1 of 1

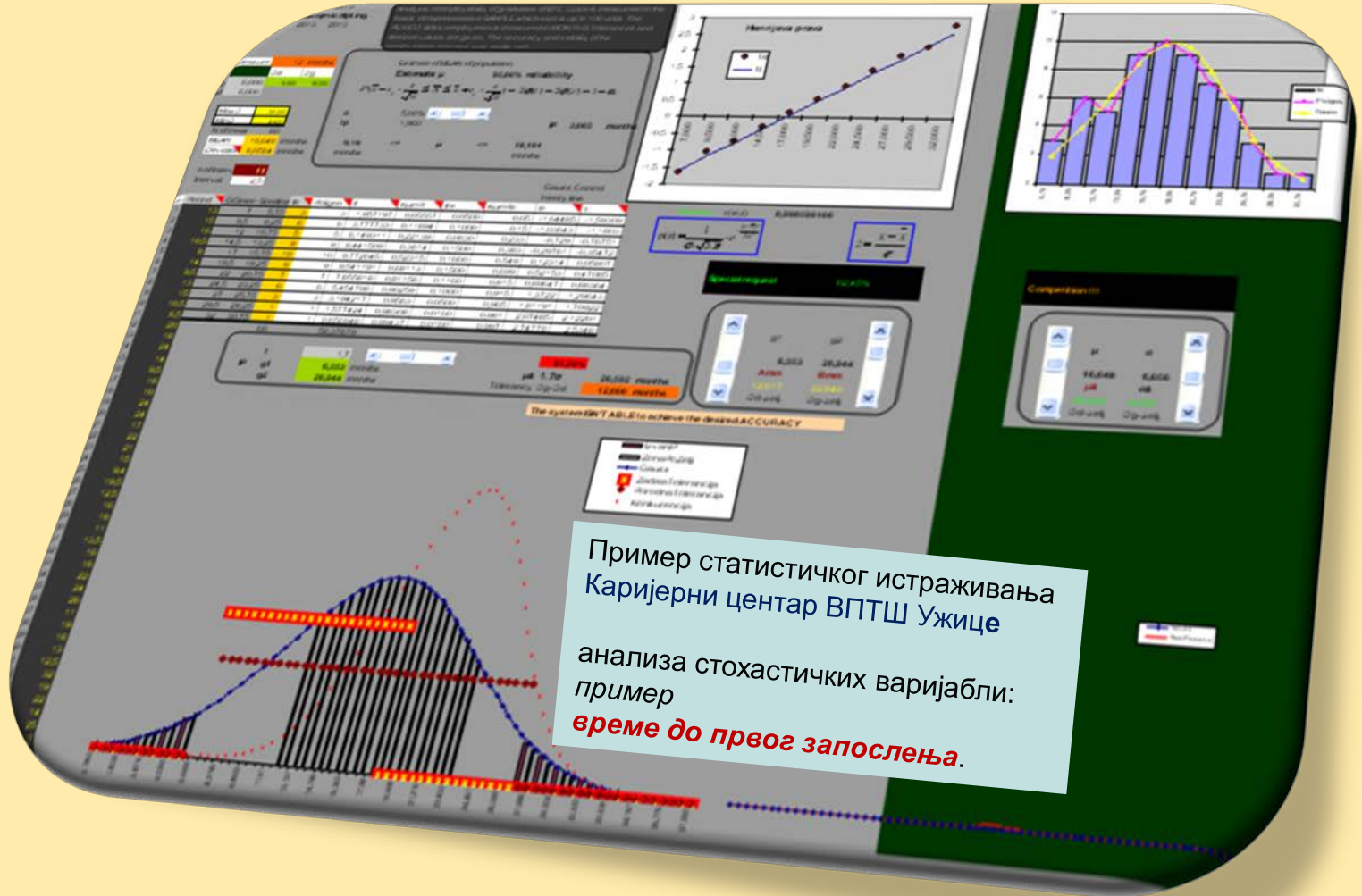
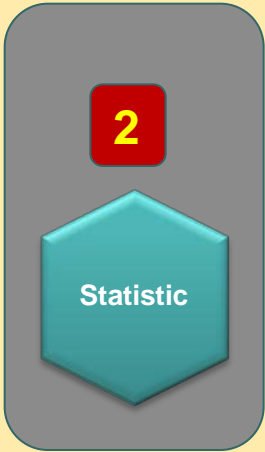
Record: 1 of 35

Радно искуство

Изглед апликације за праћење пословне каријере студената основних и специјалистичких студија ВПТШ Ужице...

Подаци о напредовању на послу

за запослене чланове ALUMNI заједнице ВПТШ / једна од апликација информационог система Каријерног центра ВПТШ Ужице

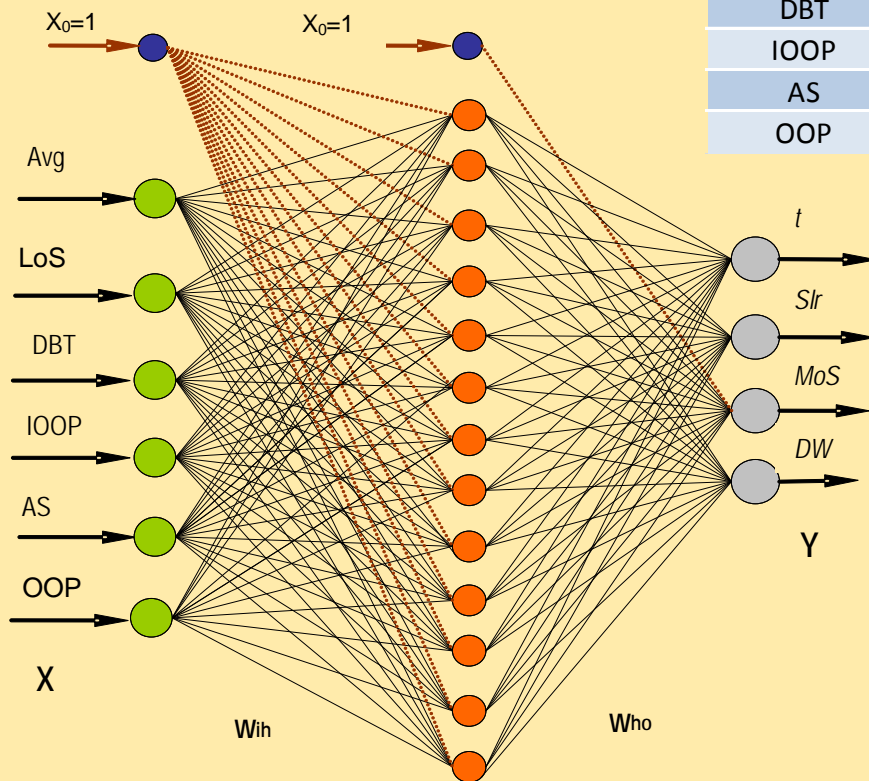


Пример статистичког истраживања
 Каријерни центар ВПТШ Ужице
 анализа стохастичких варијабли:
 пример
време до првог запослења.





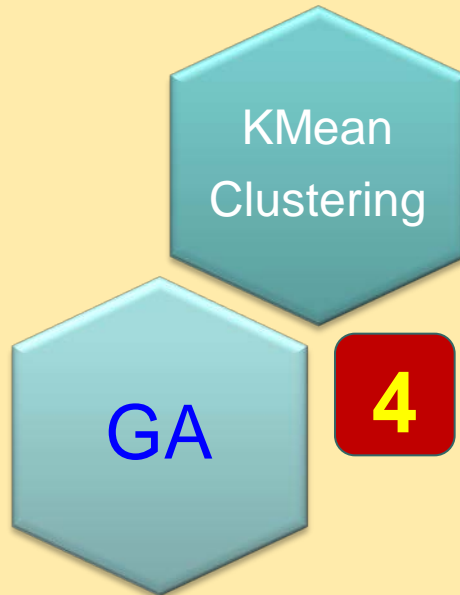
Acronym	Meaning
Avg	Average grade during the studies
LoS	Length of studies
DBT	DataBase Tools
IOOP	Intudaction to Object Oriented Programming
AS	Applicative Software
OOP	Object Oriented Programming



ПРИМЕР примене ANN
Каријерни центар ВПТШ, Ужице

Acronym	Meaning
<i>t</i>	Time before first employment
<i>Slr</i>	Average salary
<i>MoS</i>	Measure of Satisfaction
<i>DW</i>	Daily work in hours

Пример примене техника AI (Artificial Intelligence) у моделирању утицаја предиктора: просечне оцене, дужине студирања и оцена из стручних информатичких предмета на ВПТШ на запослење и квалитет добијеног посла израженог преко: *t*-време протекло од дипломирања до првог посла, *Slr*-просечна плата, *MoS*-мера задовољства послом, *DW*-дужина радног дана...



SOFTWARE FOR k-MEANS CLUSTERING

Кластери потенцијалних
послова за дипломиране
струковне инжењере,
економисте и менаџере
основних и специјалистичких
студија

Модул ИС: Каријерни центар,
ВПТШ, Ужице
развојна фаза



Кључне одреднице:

❖ *k-Means Clustering* је једна од најчешће коришћених техника у домену **Data Mining**-а.

❖ *Software* итеративно распоређује **m** објеката (**послова**) описаних са **n** димезија (**атрибута**) у **k** задатих кластера (**група послова**).

❖ Реалокација објеката између кластера, током итерација, је базирана на **концепту сличности**.

❖ Сличност је изражена кроз појам **дистанце**: Стандардна Еуклидова дистанца, Квадрирана Еуклидова дистанца, Менхетн дистанца, Косинусна дистанца као и мера сличности изражена преко **корелације** (*Pearson*-ов коефицијент просте линеарне корелације).

❖ Потрага за глобалним минимумом је базирана на минимизацији **функције трошка**.

❖ Функција трошка је дефинисана преко минимума унутар кластерског растојања (**Within Cluster Sumed Squered**).

❖ *kMean Clustering* омогућава редукацију података

❖ *kMean Clustering* омогућава истраживање унутрашње структуре података и интеракција међу варијаблима као и боље предикције у кластерским подзонама података

Кључне речи:

kMean кластеровање, **Сличност**, **Центроиди** кластера, **Within Clusters Sumed Squered**, **Функција трошка**





Кластеровање

- ❖ Примарни циљ је **класификовање послова** у групе на бази особина (карактеристика) које ови послови садрже
- ❖ У процесу кластеровања трага се за максимумом **интерне хомогености међу пословима** и **екстерном хетерогености међу групама послова**.
- ❖ Ова методологија припада домену Data Mining-а и карактерише се као ненадгледана техника машинског учења.

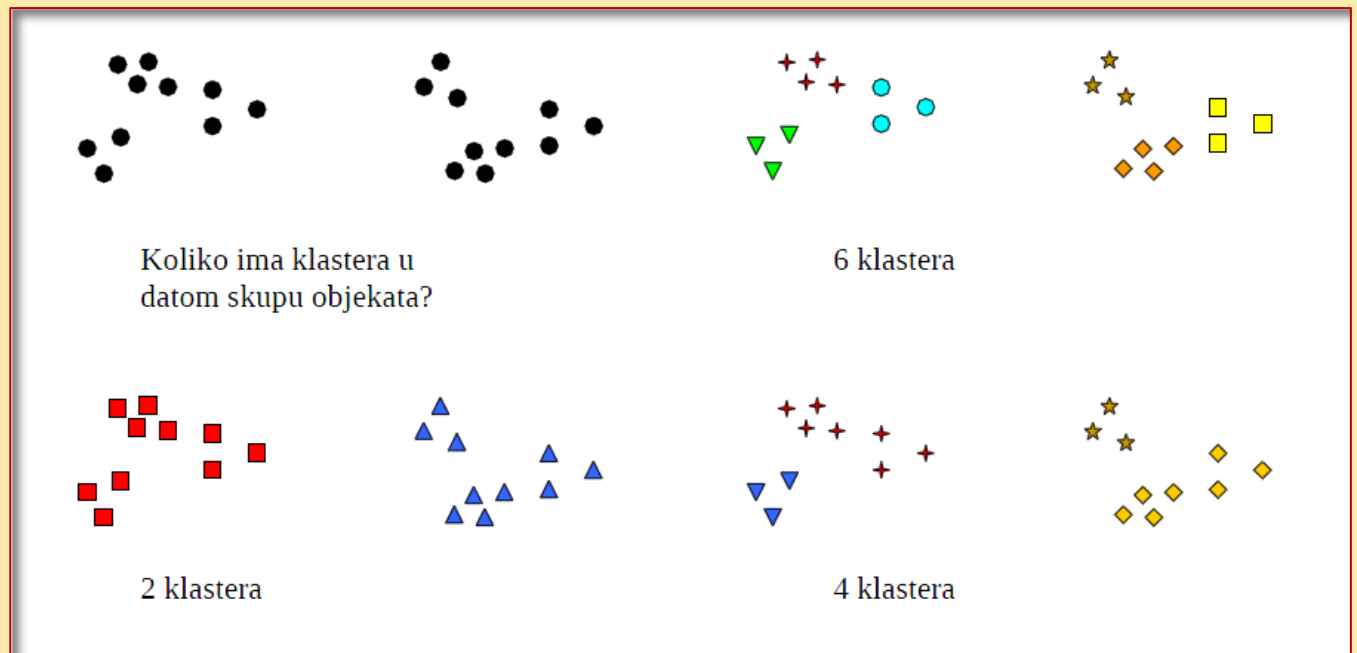


Кластеровање

- ❖ Кластерска анализа, на жалост, нема статистичку подлогу
- ❖ Треба је првенствено схватити као технику претраге
- ❖ Решење није једнозначно и зависи од бројних одлука које корисник (истраживач) мора донети
- ❖ За правилно коришћење потребно је пуно знања, као и пажљив приступ јер је реч о веома комплексној области.
- ❖ Интерпретација добијених резултата је од огромног значаја.
- ❖ Избор карактеристика којим се описује сваки од послова је од огромне важности.
- ❖ Решења нису јединствена (што је приказано на следећем сајду).
- ❖ Свако решење може бити добро. То је питање критеријума и намене.

Каква кластеризација је добра?

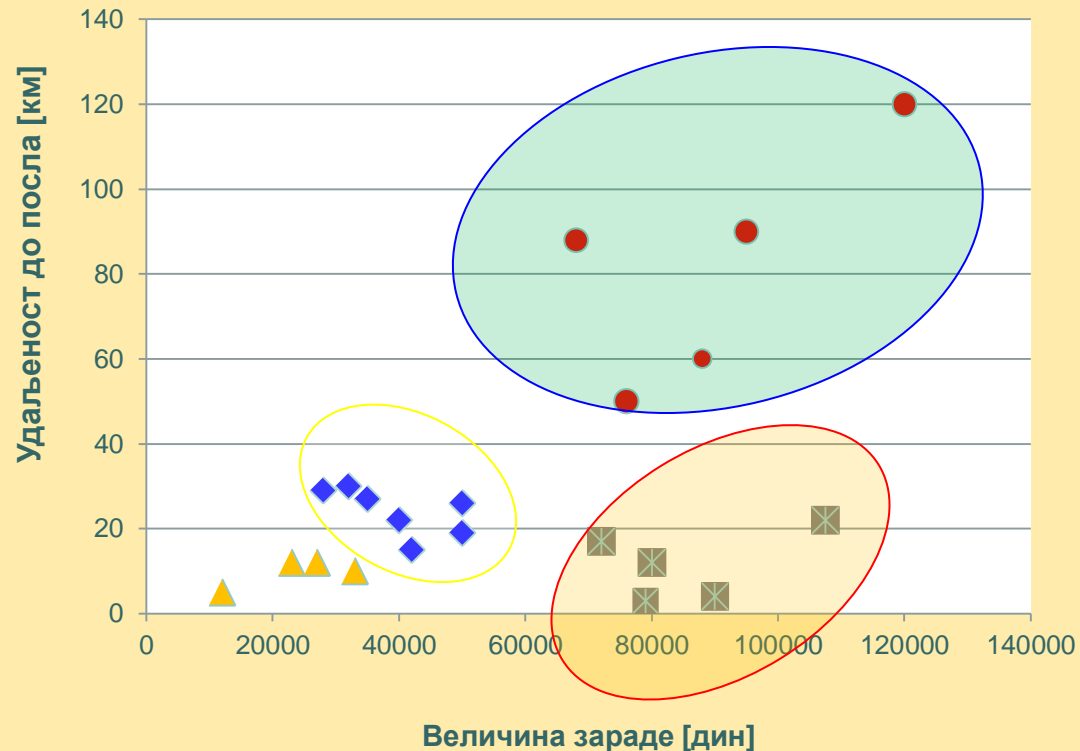
Свако решење може бити добро. Питање је критеријума и намене !



Каква кластеризација је добра?



Кластеровање послова?



Карактеристике посла

Зарада
Удаљеност

АЛИ!!!

- Дужина радног времена
- Опасност по здравље
- Стрес
- Ниво одговорности
- Стручност
- Тимски рад
- Смене
- Знање страних језика
- Информатичко знање
- Физичка снага
- Социјална интелигенција
- Године
- Пол
- Брачно стање
- Деца
- ...
- ...





Постоји велики број техника кластеровања , које се генерално могу поделити у три групе:

- k-Means кластеровање
- Хијерархијско кластеровање са две подгрупе
 - Агломерацијско кластеровање
 - Кластеровање дељењем
- Самоорганизујуће мапе



Историја

- k-Means методу први су независно користили и описали *James MacQueen* и *E.W. Forgy*, 1967. године.
- *Hartigan* и *Wong* су презентирали много ефикаснији алгоритам (имплементиран у Fortran-u) и објавили га 1975/79.
- Једно од познатијих побољшања k-Means алгоритма дали су Lloyd и Forgy.
- k-Means кластеровање се доста користи у пракси и има јаку софтверску подршку (SPSS, R, Free Software ...)





ТЕОРИЈСКА ПОДЛОГА

Циљ kMean кластеровања је налажење најбоље поделе n ентитета у k група, тако да укупна дистанца између чланова групе (скупа) и одговарајућег центроида, који репрезентује групу, буде минимална. Формално циљ је партиционисање n ентитета у k скупова $S_i, i = 1, 2, \dots, k$ тако да се минимизује **функција трошка** \mathfrak{J} која је дефинисана међукластерском сумом квадрата **WCSS (Within Cluster Sum of Squeres)**:

$$\mathfrak{J} = \sum_{\eta=1}^k \sum_{\chi=1}^n \|L_{\chi}^{\eta} - C_{\eta}\|^2$$



Algorithm 1 Basic K-means Algorithm.

- 1: Select K points as the initial centroids.
- 2: **repeat**
- 3: Form K clusters by assigning all points to the closest centroid.
- 4: Recompute the centroid of each cluster.
- 5: **until** The centroids don't change

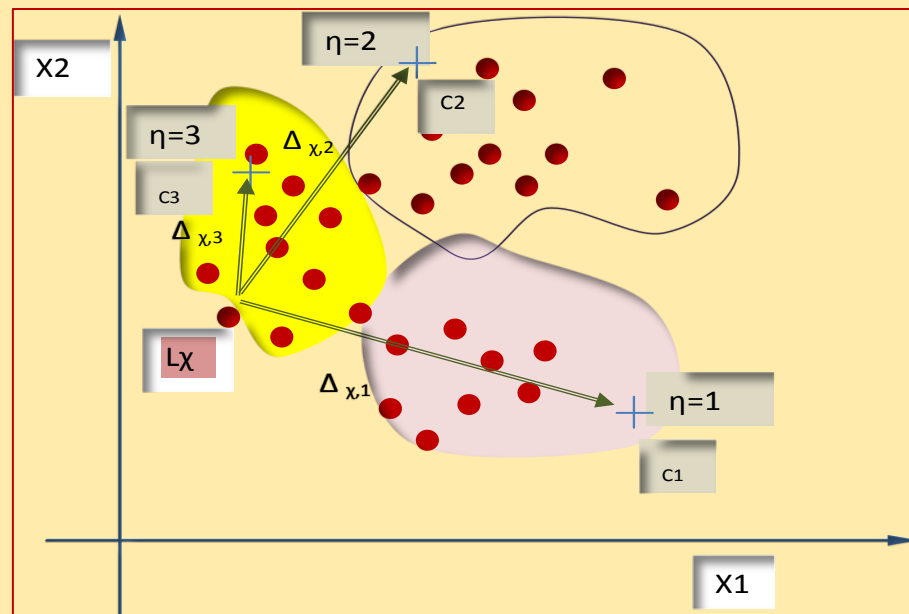
K-Means Clustering Algorithm



k-Mean кластеровање /кораци: **K0/** **Припрема података**

- ❖ *Нетипичне тачке*
- ❖ *Нормализација* $x_{j,i} = (x_{j,i} - \mu_i) / \sigma_i$
- ❖ *Weighting...* (пондерисање) / примена генетских алгоритама /GA

k-Mean clustering: **корак: K1/ Семена**





k-Mean clustering: кораџи: **K2/ Концепт сличности**

❖ Корелациона мера (за метричке податке)

$$r_{L\chi-C\eta} = \frac{n \cdot \sum_{j=1}^d x_j^{(\chi)} \cdot x_j^{(\eta)} - \sum_{j=1}^d x_j^{(\chi)} \cdot \sum_{j=1}^d x_j^{(\eta)}}{\sqrt{n \cdot \sum_{j=1}^d (x_j^{(\chi)})^2 - (\sum_{j=1}^d x_j^{(\chi)})^2} \cdot \sqrt{n \cdot \sum_{j=1}^d (x_j^{(\eta)})^2 - (\sum_{j=1}^d x_j^{(\eta)})^2}}$$

❖ Мера дистанџе (за метричке податке)

1. SEuclid – Squared Euclidean Distance

$$\Delta^2(\mathbf{x}_{L\chi}, \mathbf{x}_{S\eta}) = \left(\sum_{i=1}^d |x_{i,\eta} - x_{i,\chi}|^2 \right); \chi = 1, \dots, n; \eta = 1, \dots, k$$

2. Standard Euclidean distance

$$\Delta(\mathbf{x}_{L\chi}, \mathbf{x}_{S\eta}) = \sqrt{\left(\sum_{i=1}^d |x_{i,\eta} - x_{i,\chi}|^2 \right)}; \chi = 1, \dots, n; \eta = 1, \dots, k$$

3. Block-City-Block ili Manhattan distance

$$\dots \Delta(\mathbf{x}_{L\chi}, \mathbf{x}_{S\eta}) = \left(\sum_{i=1}^d |x_{i,\eta} - x_{i,\chi}| \right); \chi = 1, \dots, n; \eta = 1, \dots, k$$





k-Mean clustering steps: **K2/ Концепт сличности**

- ❖ *Мере дистанце (за метричке податке)*

4. *Cosine distance*

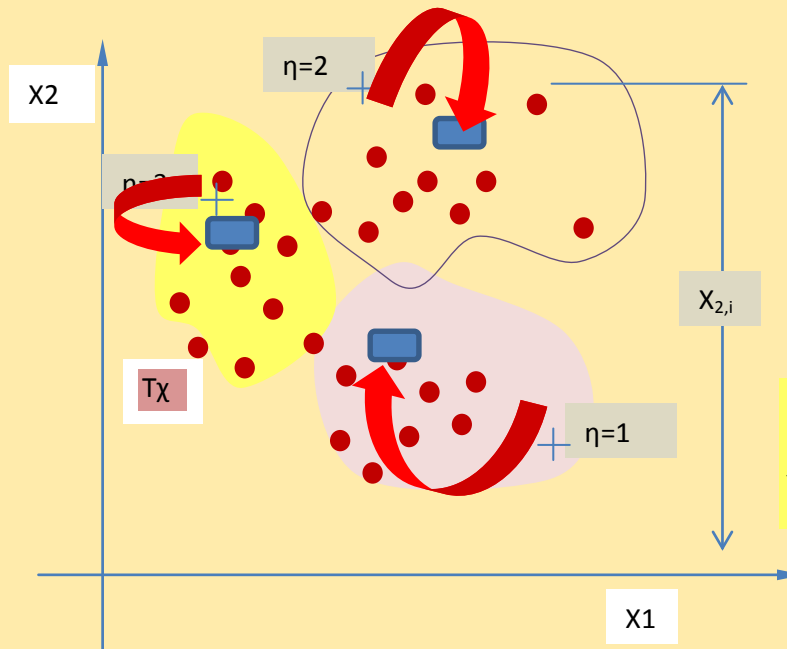
$$\Delta_{L_\chi, S_\eta} = \cos \varphi(L_\chi, S_\eta) = \frac{\overrightarrow{L_\chi} \cdot \overrightarrow{S_\eta}}{|\overrightarrow{L_\chi}| \cdot |\overrightarrow{S_\eta}|}, \chi = 1, \dots, n; \eta = 1, \dots, k$$

5. *Mahalanobis distance*

- ❖ *Мере асоцијације (за неметричке податке)*



k-Mean кластеровање: clustering кораци: K3/K4/K5 Репозиционирање центра



$$x_j^{(T\eta)} = \bar{x}_j^{(T\eta)} = \frac{\sum_{i=1}^{b_\eta} x_{j,i}}{b_\eta}, \quad j = 1, \dots, d$$

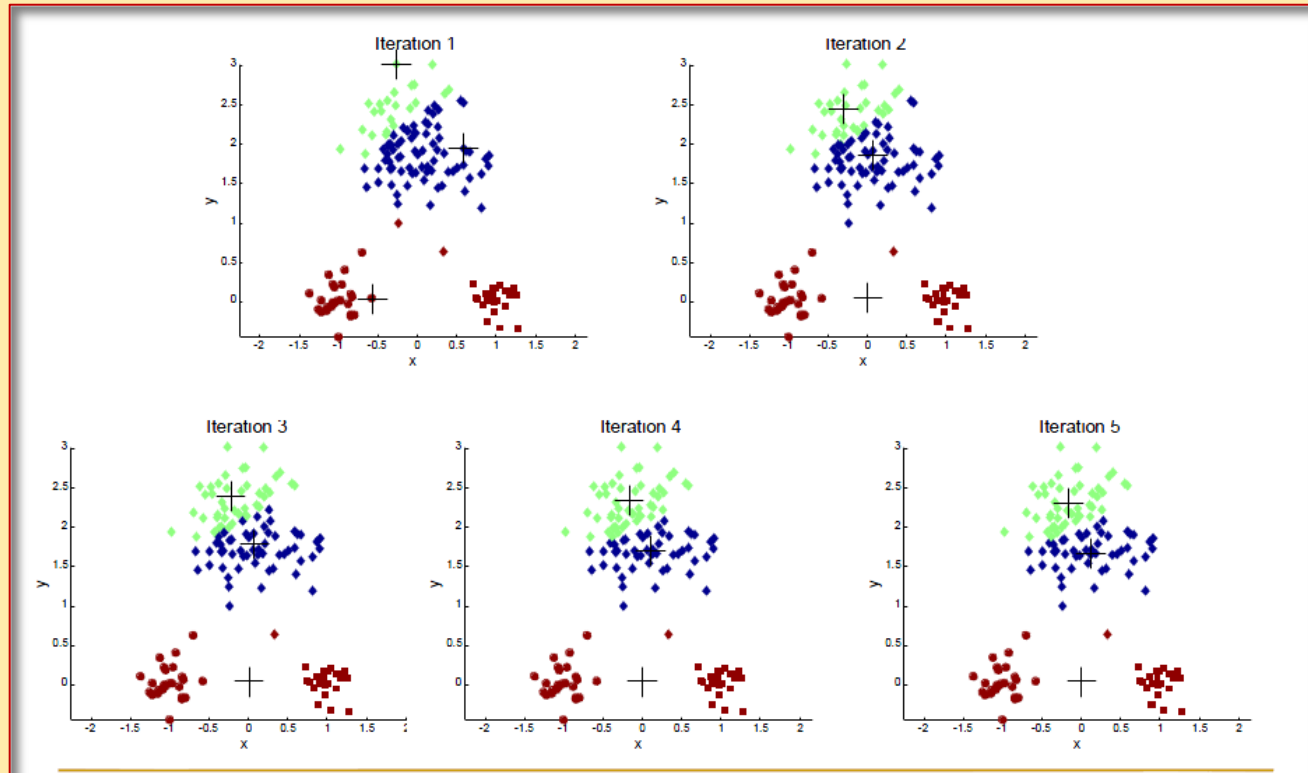
СТАТИЧКИ МОМЕНТИ ИНЕРЦИЈЕ

$$\sqrt{\sum_{j=1}^d |x_{(T\eta)j}^{(\tau)} - x_{(T\eta)j}^{(\tau+1)}|^2} \leq \varepsilon, \quad \eta = 1, \dots, k$$



Проблеми!!! Решење: евалуација

kMean алгоритам је хеуристички алгоритам, који не гарантује постизање глобалног оптимума. Резултат може зависити од почетне иницијализације кластера.



SOFTWARE-ска имплементација



Висока
пословно-
техничка школа
Ужице
2015.

kMean Clustering by M. Milivojevic & Dj. Forst & M. Stevanovic /// 2014

Uneti broj klastera: 5 Zoom

Odebrati distancu: Euclid

Normalizovani podaci Data

Centri klastera:
 Random
 Manual

Vrednost epsilon: Centri KLASTERA

0.01

Broj ciklusa: 1000 Broj iteracija: 25

kMean clustering

Broj test tačaka: 0 Test tačke

Object	X1	X2	Y
24	0.120803121	-0.273621514	137
116	1.188441512	-0.273621514	54
188	-0.235076343	2.668730014	3
148	-0.94683527	-0.808594519	86
100	-0.413016075	-0.808594519	86
110	1.01050178	-0.006135012	38
102	-0.768895538	-0.273621514	8
228	0.120803121	-0.808594519	30
160	0.832562048	2.936216517	17
113	-0.94683527	-0.541108017	158
131	2.96783883	1.598784004	4
18	0.120803121	-0.273621514	43
179	-0.590955807	-0.273621514	94
161	-0.94683527	-0.808594519	45
219	-0.94683527	-0.808594519	74

zscore

Microsoft Office Web Compon

EN English (United States) Help

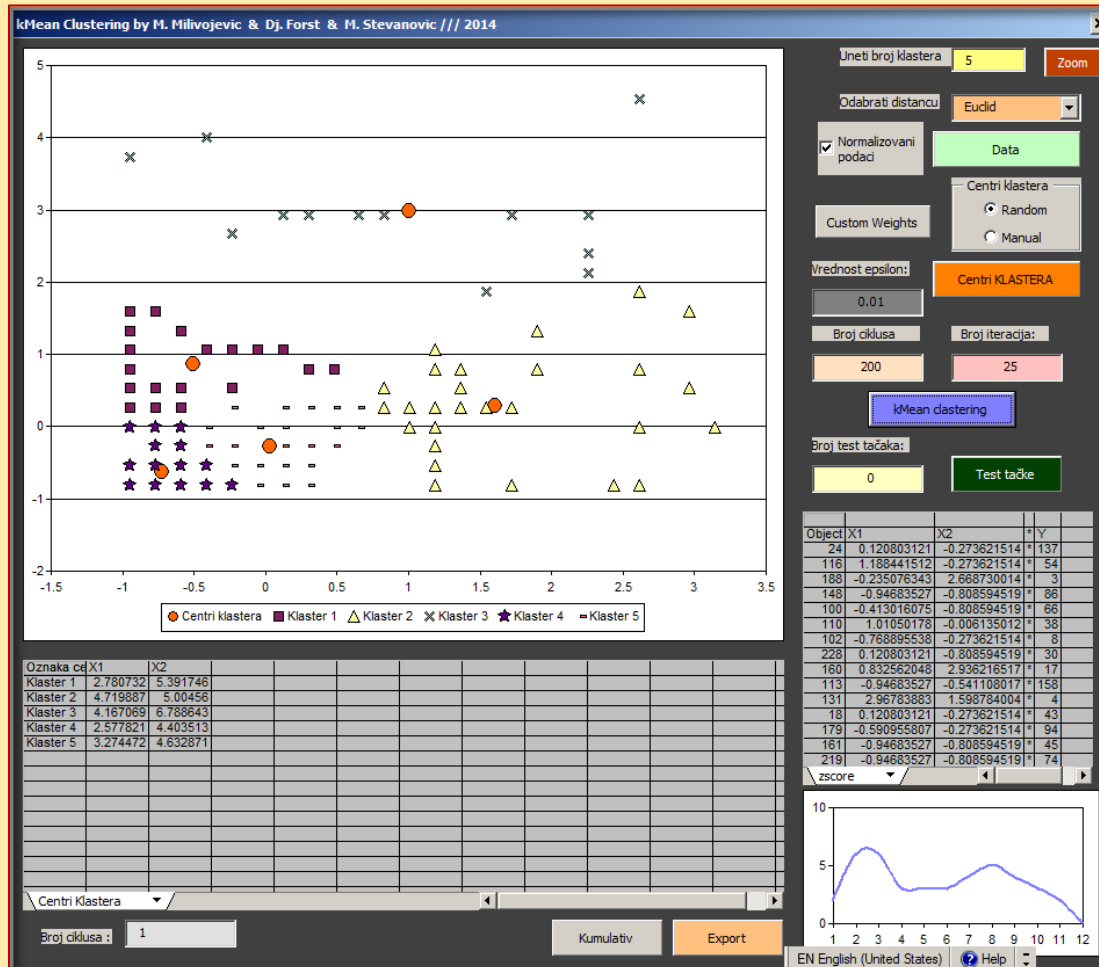
Sheet1 Broj ciklusa: Kumulativ Export

Software k-Means Clustering - interface

517200- TEMPUS-1-2011-1-BE TEMPUS-SMGR



Висока
пословно-
техничка школа
Ужице
2015.



Software k-Means Clustering - – 2D Demo

Прикупљање података



Висока
пословно-
техничка школа
Ужице
2015.



АНКЕТА

потенцијални послови
карактеристике послова

Висока пословно-техничка школа
Ужице
2015.

Поштовани,

Молимо вас да процените ваше потребе по питању кадрова за наредне 3 године и да што је могуће објективније опишете карактеристике послова на којима би кардови били упуслени.

Назив пословног субјекта		
Адреса		
Лице за контакт		
e-mail		
Телефон		

Карактеристике посла	Посао 1	Посао 2	Посао 3	...	Посао n
Категорија					
Занимање					
Број извршилаца					
Временски план запошљавања					
Очекивана висина зараде					
Удаљеност до посла					
Дужина радног времена					
Опасност по здравље					
Стрес					
Ниво одговорности					
Стручност					
Тимски рад					
Смене					
Знање страних језика					
Информатичко знање					
Физичка снага					
Социјална интелигенција					
Године					
Пол					
Брачно стање					
Деца					

517200- TEMPUS-1-2011-1-BE TEMPUS-SMGR



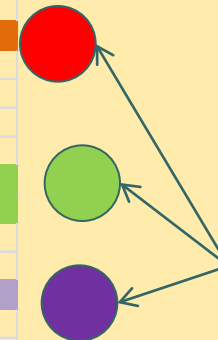
Кодирање података
ФАЗА А



Висока пословно-техничка школа
Ужице
2015.

PosaoN	Kategorija	Zanimanje	Izvrsilaca	Dinamika	Zarada	Udaljenost	dnevnoH	OpasnostZ	Stres	Odgovornost	Strucnost	Tim	Smene
1	1	17	3	5	71893	191	6	8	11	55	71	76	2
2	9	9	6	8	85435	113	8	16	14	69	93	97	3
3	7	6	8	35	49508	7	3	6	14	68	86	65	1
4	6	11	8	30	81243	97	9	12	2	53	82	57	2
5	3	2	1	33	46558	60	2	14	20	77	90	77	2
6	6	5	10	23	125554	42	8	17	5	71	85	55	2
7	2	13	8	14	78877	7	2	15	25	98	98	93	1
8	2	14	6	16	101719	72	4	3	18	51	77	97	3
9	5	18	10	32	82194	90	8	14	19	88	56	72	2
10	7	15	5	10	32436	81	2	4	24	85	53	89	2
11	4	15	3	23	64871	190	6	3	27	79	71	52	3
12	9	2	4	28	106644	126	7	3	17	57	98	70	1
13	7	8	3	29	52581	134	7	1	28	50	90	83	3
14	8	16	1	19	82223	89	2	6	15	65	83	53	1
15	8	8	8	21	81744	54	7	12	6	76	71	83	1
16	2	10	8	26	75716	81	2	15	20	72	80	60	2

Tim	Smene	Jezik	Informatika	Fizika	EI	Godine	Pol	Brak	deca	*	Firma
76	2	72	82	89	94	32	1	2	2	*	Futura
97	3	88	96	73	80	29	1	1	1	*	Maxi
65	1	52	56	69	58	36	2	2	1	*	Polo
57	2	73	75	71	86	37	2	1	1	*	AAG
77	2	70	87	89	76	38	2	1	3	*	Dolina
55	2	72	84	61	59	33	1	2	3	*	Dolina
93	1	84	63	74	63	34	1	1	2	*	Dolina
97	3	71	72	97	92	39	1	2	3	*	Futura
72	2	67	55	51	79	35	1	1	3	*	Futura
89	2	69	84	55	97	27	1	2	3	*	Maxi
52	3	70	93	82	83	32	2	2	2	*	Saturn
70	1	88	85	97	82	24	2	1	1	*	Media
83	3	95	86	93	62	36	2	2	2	*	Media
53	1	64	67	65	66	33	1	1	2	*	Media
83	1	88	60	92	55	38	2	1	2	*	Saturn
69	2	96	90	60	66	35	2	1	2	*	Saturn
74	2	72	68	92	79	32	1	1	2	*	Saturn
57	2	95	97	54	62	31	2	2	1	*	Futura
71	3	95	99	53	82	26	2	2	3	*	Futura



Тачке у m
димензионалном
простору
потенцијалних
послова

Простор за
кMean
кластеровање

517200- TEMPUS-1-2011-1-BE TEMPUS-SMGR



Висока
пословно-
техничка школа
Ужице
2015.

kMean Clustering by M. Milivojevic & Dj. Forst & M. Stevanovic /// 2014

Uneti broj klastera: 5

Odabrati distancu: Euclid

Normalizovani podaci

Centri klastera: Random

Vrednost epsilon: 0.01

Broj ciklusa: 300

Broj iteracija: 25

Broj test tačaka: 0

Centri Klastera

Oznaka	cd	Kategorija	Zanimanje	Izvrslaca	DinamikaF	Zarada	Udajenost	dnevnoH	Opasnost	Stres	Odgovorn	Strucnost
Klaster 1	3	517241	13.55172	5.137931	16.7931	57041.38	99.06897	5.724138	7	19.58621	66.24138	66.06897
Klaster 2	8	066667	9.86667	5.86667	9.2	97957.33	107.4667	4.466667	14.8	13.73333	65.66667	77.93333
Klaster 3	6	511429	6.214296	26.21429	59065.21	44	3.926571	7	13.07143	76.64296	72.35714	
Klaster 4	5	307692	12.23077	6.461538	17.61538	74761.38	104.1154	5.884615	11.88462	11.57692	74.34615	85.30769
Klaster 5	4	291667	8.916667	4.125	24.79167	67702.87	131.7917	4.916667	10.25	22.33333	81.79167	75.08333

Broj ciklusa: 1

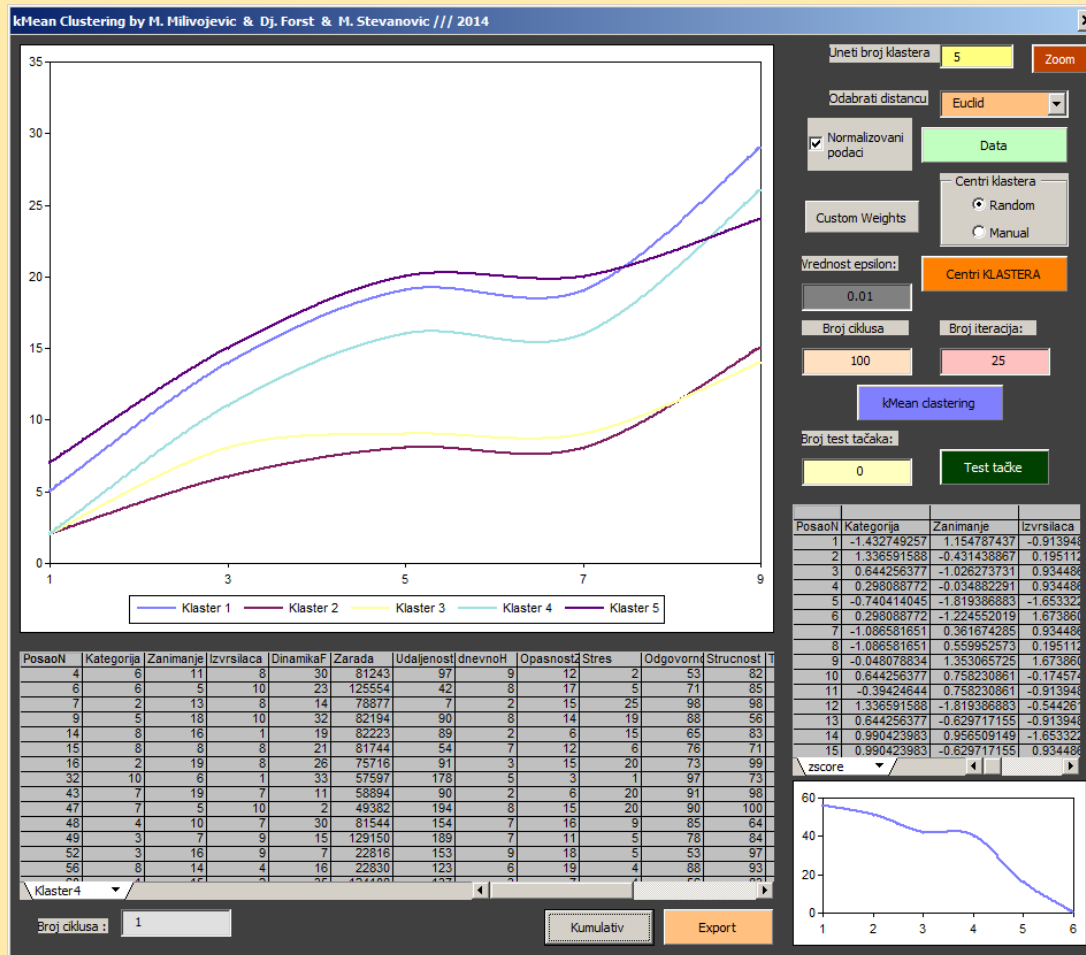
Kumulativ Export

Software k-Means Clustering – nD interface

Резултати и дискусија



Висока
пословно-
техничка школа
Ужице
2015.



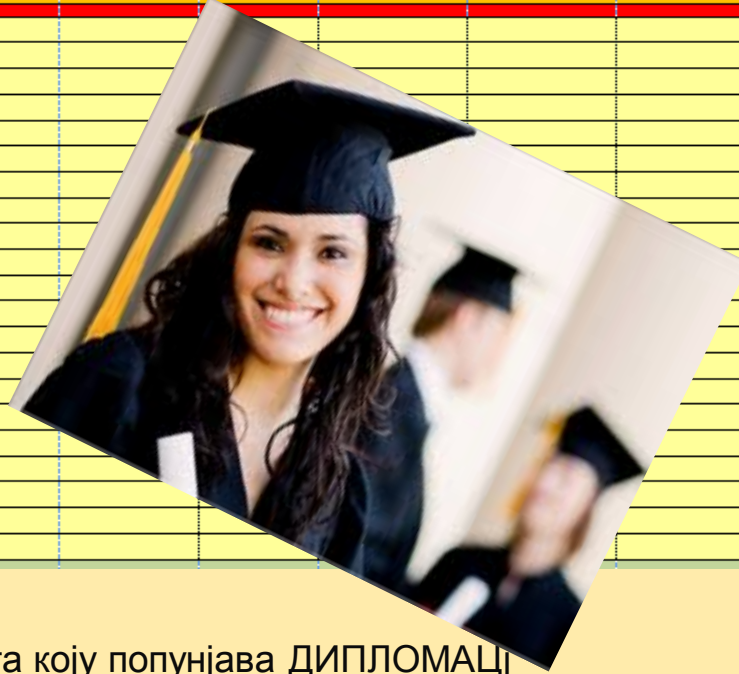
Потрага за послом / ДИПЛОМАЦ



АНКЕТА
потенцијални послови
карактеристике послова

Висока пословно-техничка школа
Ужице
2015.

Карактеристике посла	Посао 1	Посао 2	Посао 3	...	Посао n
Категорија					
Занимање					
Број извршилаца					
Временски план запослења					
Очекивана висина зараде					
Удаљеност до посла					
Дужина радног времена					
Опасност по здравље					
Стрес					
Ниво одговорности					
Стручност					
Тимски рад					
Смене					
Знање страних језика					
Информатичко знање					
Физичка снага					
Социјална интелигенција					
Године					
Пол					
Брачно стање					
Деца					



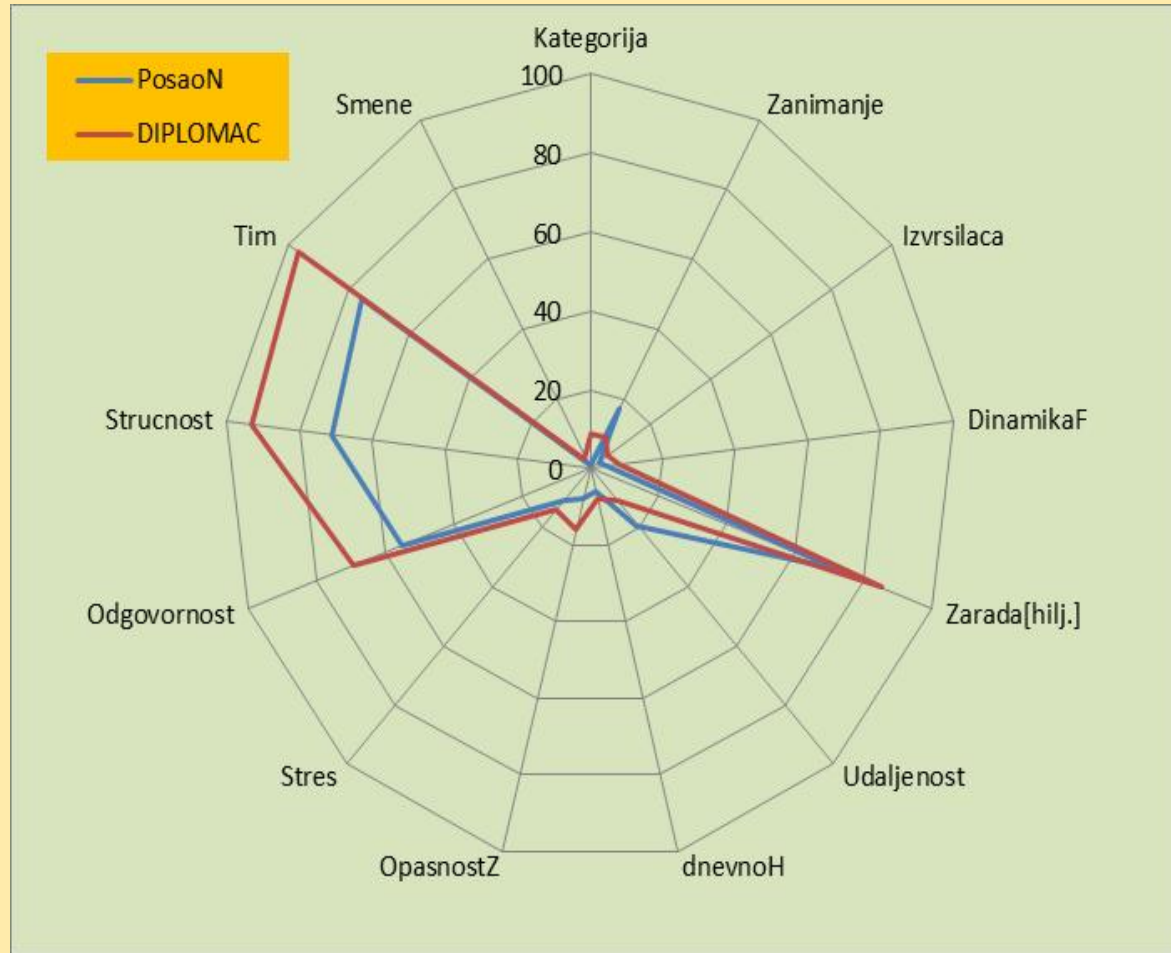
Анкета коју попуњава ДИПЛОМАЦ



Циљ



Висока
 пословно-
 техничка школа
 Ужице
 2015.





Хвала на пажњи!

мр Милован Миливојевић, дипл. инж.



Висока пословно-техничка школа Ужице, Србија

Technical and Business College, Uzice, Serbia



**Висока
пословно-
техничка школа
Ужице
2015.**