

MI002	Grupiranje podataka i primjene	P	V	S	ECTS 5
		2	1	1	

**Cilj predmeta.** Studente upoznati s osnovnim činjenicama i rezultatima iz područja grupiranja podataka te mogućim primjenama.

**Potrebna predznanja.** Preddiplomski studij matematičkog, računarskog ili srodnog smjera.

### Sadržaj predmeta.

1. Uvod i motivacija. Definiranje problema i osnovna svojstva. Razni primjeri iz primjena.
2. Reprezentant skupa podataka. Osnovne kvazimetričke funkcije (LS i LAD kvazimetrička funkcija). Reprezentant konačnog skupa iz  $R^n$ . Mahalanobis kvazimetrička funkcija. Interpretacija periodičnih podataka na jediničnoj kružnici.
3. Tvrdo grupiranje podataka. Grupiranje podataka s jednim, dva i više obilježja. Primarni i dualni problem uz primjenu LS kvazimetričke funkcije. Kriterijske funkcije cilja i princip minimalnih udaljenosti.
4. K-means algoritam. Inkrementalni algoritmi.
5. Izbor particije s najprikladnijim brojem klastera – indeksi (Calinski-Harabasz, Davies-Bouldin, Silhouette Width Criterion). Uspoređivanje particija (Rand indeks, Hausdorffova udaljenost).
6. Analiza i prepoznavanje teksta.
7. Prepoznavanje geometrijskih objekata u ravnini. Pravac, kružnica i elipsa kao reprezentant skupa točaka u ravnini. Prepoznavanje više istovrsnih geometrijskih objekata u ravnini.
8. Meko grupiranje podataka. Fuzzy c-means algoritam. Gustafson-Kessel k-means algoritam.

### ISHODI UČENJA

R.b.	ISHODI UČENJA
1.	Prepoznati probleme iz odgovarajućih baza podataka na koje se mogu primijeniti dobivena znanja.
2.	Samostalno odabrati odgovarajuću metodu za problem grupiranja podataka.
3.	Analizirati složenost optimizacijskog problema grupiranja podataka i primijeniti osnovni k-means algoritam, kao i neke druge metode.
4.	Primijeniti ideju raznih geometrijskih objekata kao reprezentanata podataka.
5.	Provesti matematički dokaz utemeljenosti postupaka i formula kojima se služe u okviru ovog kolegija.

### POVEZIVANJE ISHODA UČENJA, ORGANIZACIJE NASTAVNOG PROCESA I PROCJENA ISHODA UČENJA

ORGANIZACIJA NASTAVNOG PROCESA	ECTS	ISHOD UČENJA **	AKTIVNOST STUDENATA*	METODA PROCJENE	BODOVI	
					min	max
Pohađanje predavanja i vježbi	1	1-5	Prisutnost na nastavi, rasprava, timski rad, samostalan rad na zadacima i kratke provjere znanja	Potpisne liste, praćenje aktivnosti na nastavi	0	4
Provjera znanja (kolokvij)	2	1- 5	Priprema za pismenu provjeru znanja	Provjera točnih odgovora (ocjenjivanje)	25	48
Završni ispit	2	1- 5	Ponavljanje gradiva	Usmeni ispit	25	48
UKUPNO	5				50	100

**Izvođenje nastave i vrednovanje znanja.** Predavanja i vježbe su ilustrirani gotovim programima. Vježbe su djelomično auditorne, a djelomično laboratorijske uz korištenje računala. Predavanja, vježbe i seminari su obvezni. Ispit se sastoji od pismenog i usmenog dijela, a polaže se nakon odslušanih predavanja. Prihvatljivi rezultati postignuti na kolokvijima, koje studenti pišu tijekom semestra, zamjenjuju pismeni dio ispita. Studenti mogu utjecati na ocjenu tako da tijekom semestra pišu domaće zadaće ili izrade seminarski rad. Domaće zadaće sadrže proširenje gradiva, a očekuje se samostalan i kreativan rad. Seminarski radovi shvaćaju se kao proširenje domaćih zadaća.

**Može li se predmet izvoditi na engleskom jeziku:** Da

**Osnovna literatura:**

1. R. Scitovski, K. Sabo, Grupiranje podataka i primjene, recenzirani nastavni materijal na web stranici
2. S. Theodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, Burlington, 2009, 4th edition.
3. J. Kogan, Introduction to Clustering Large and High-Dimensional Data, Cambridge University Press, 2007.

**Dopunska literatura:**

1. J. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal, Fuzzy models and algorithms for pattern recognition and image processing, Springer, 2005.
2. S. Butenko, W. A. Chaovalitwongse, P. M. Pardalos, Clustering Challenges in Biological Networks, World Scientific, 2009.
3. S. Dhillon, S. Mallela, R. Kumar, A divisive information theoretic feature clustering algorithm for text classification, Journal of Machine Learning Research, 3(2003) 1265–1287.
4. L. Kaufman, P. J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, John Wiley & Sons, Chichester, UK, 2005.
5. B. Mirkin, Data clustering for Data Mining, Chapman & Hall/CRC, 2005.
6. R. Scitovski, M. B. Alić, Grupiranje podataka, Odjel za matematiku, Sveučilište u Osijeku, 2016.
7. H. Späth, Cluster-Formation und- Analyse, R. Oldenburg Verlag, München, 1983.
8. P. N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Wesley, 2006.