



АКАДЕМИЈА СТРУКОВНИХ СТУДИЈА ЗАПАДНА СРБИЈА

Седиште Ужице, Трг Светог Саве 34

О Д С Е К Ужице

План рада

Назив предмета	Наука о подацима (2+3)				
Студијски програм/и (модул)	МАСТЕР Информационе технологије и системи				
Година студија	I	Семестар	II	ЕСПБ	6
Статус предмета	Обавезни		Услов	Нема	

Подаци о наставницима и сарадницима на предмету	
Име и презиме, академско звање, термин консултација, електронска адреса	<i>Предавања:</i> др Милован Миливојевић, проф. стр. студ. Консултације: четвртак, од 10 до 12 часова, кабинет 203 Електронска адреса: mmilivojevic031@gmail.com <i>Вежбе:</i> Жарко Богићевић

Циљеви предмета
СТИЦАЊЕ конкурентних и модерних знања о фундаменталним процесима, техникама и методама које чине науку о подацима и могућностима њихове практичне примене.

Садржај и структура предмета
<i>Теоријска настава</i> Увод. Интердисциплинарност. Подаци. Small Data. Big Data. Data Science. Процеси. Припрема података за анализу (data wrangling); истраживачка анализа (exploratory data analysis) и евалуација; моделирање засновано на подацима; анализа резултата; комуникација резултатима. Примена у различитим областима. Визуелизација података. Граматика графике. Технике динамичке и интерактивне визуелизације података. Припрема података. Квалитет података. Статистичко размишљање и закључивање. Примери на одабраној софтверској платформи. Одабрани примери расподела: Binomial, Geometric, Poisson, Exponential, Gaussian, Student's, Snedecor's F distribution, Beta, Weibull... Тестирање хипотеза о параметрима основних скупова и облику расподеле; тестови нормалности, анализа варијансе, непараметарски тестови. Редукција димензионалности и факторска анализа података. РСА - анализа главних компоненти. Моделирање података. Регресије. Вишеструка регресија (MLR). Stepwise регресије. Логистичка регресија. Моделирање помоћу вештачких неуронских мрежа. Врсте растојања. Концепти сличности. Концепти класификације и кластеризације података. Кластеризација. K-Means. K-Medoids. Хијерархијска кластеризација. FRM (Frequency Recency Monetary) анализа. Association rules. Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје. Преглед блиских тема. Временске серије. Sentiment Analysis. Етички проблеми. Будућност науке података. <i>Практична настава:</i>

Примена програмског језика R у науци о подацима. Рад на вежбама ће подразумевати примену стеченог знања на решавање конкретних актуелних проблема у различитим областима

План и распоред извођења наставе	
Наставна недеља	НАЗИВ НАСТАВНЕ ЈЕДИНИЦЕ
Предавања	
1	Увод. Интердисциплинарност. Подаци. Small Data. Big Data. Data Science. Процеси. Припрема података за анализу (data wrangling); истраживачка анализа (exploratory data analysis) и евалуација; моделирање засновано на подацима; анализа резултата; комуникација резултатима. Примена у различитим областима.
2	Визуелизација података. Граматика графике. Технике динамичке и интерактивне визуелизације података.
3	Припрема података. Квалитет података. Статистичко размишљање и закључивање.
4	Примери на одабраној софтверској платформи. Одабрани примери расподела: Binomial, Geometric, Poisson, Exponential, Gaussian, Student's, Snedecor's F distribution, Beta, Weibull...
5	Тестирање хипотеза о параметрима основних скупова и облику расподеле; тестови нормалности, анализа варијансе, непараметарски тестови.
6	Редукција димензионалности и факторска анализа података. PCA - анализа главних компоненти.
7	Моделирање података. Метрике квалитета моделирања. Регресије. Вишеструка регресија (MLR).
8	Stepwise регресије. Логистичка регресија.
9	Моделирање помоћу вештачких неуронских мрежа. Део А
10	Моделирање помоћу вештачких неуронских мрежа. Део Б
11	Врсте растојања. Концепти сличности. Концепти класификације и кластеризације података.
12	Кластеризација. Врсте кластеризације. Метрике квалитета кластеризације. Одабрани примери: K-Means. K-Medoids.
13	Хијерархијска кластеризација.
14	FRM (Frequency Recency Monetary) анализа. Association rules. Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје.
15	Преглед блиских тема. Временске серије. Sentiment Analysis. Етички проблеми. Будућност науке података.
Вежбе	
1	Кључни процеси и Науци о подацима. Упознавање са R програмском платформом. Инсталација. Интерфејс. Основне технике. Елементи синтаксе. Објекти у R-у.
2	Типови података. Визуелизација података у R-у. Примери. Граматика графике. Примери
3	Технике динамичке и интерактивне визуелизације података у R-у. Одабрани примери.
4	Статистика у R-у. Примери. Закони расподеле у R-у Примери.
5	Тестирање статистичких хипотеза у R-у.
6	Колоквијум А
7	Редукција димензионалности. Примена PCA (Principal Component Analysis) у R-у. Одабрани примери.
8	Моделирање регресијама у R-у. Одабрани примери. Проста линеарна регресија. MLR. Stepwise регресија. Логистичка регресија. Одабрани примери у R-у.
9	Моделирање помоћу вештачких неуронских мрежа (ANN - Artificial neural network) у R-у. Одабрани примери
10	Колоквијум Б
11	Концепти сличности. Примери у R-у. K-Means. K-Medoids кластеризација. Одабрани примери у K-Means. K-Medoids.-у.

12	Хијерархијска кластеризација. Одабрани примери у R-у.
13	FRM (Frequency Recency Monetary) анализа. Одабрани примери у R-у.
14	Помоћ студентима у изради Семинарских радова – део А.
15	Помоћ студентима у изради Семинарских радова – део Б.

Начин оцењивања – структура и број поена на предиспитним обавезама и испиту

Оцена знања (максималан број поена)			
Предиспитне обавезе	поена	Испит	поена
активност у току предавања	5	практични део на рачунару	20
практична настава	15	усмени испит	10
колоквијум	20		
семинарски рад	30		

Литература

1. Wickham, Hadley, and Garrett Golemund, R za statističku obradu podataka, Mikro knjiga, 2017.
2. Hastie, Trevor, Robert Tibshirani, and J. H Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer, 2009.
3. Pathak, Manas A., Beginning Data Science with R. Springer, 2014.
4. Schutt, Rachel, and Cathy O'Neil. Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.", 2013.