



Студијски програм: Основне академске студије информатике					
Назив предмета: УВОД У НАУКУ О ПОДАЦИМА					
Статус предмета: Обавезни на модулима: Рачунарске науке и Информационо-комуникационе технологије, изборни на модулу Софтверско инжењерство					
Број ЕСПБ: 6					
Услов: Уписан одговарајући семестар					
Циљ предмета стицање конкурентних и модерних знања о фундаменталним процесима, техникама и методама које чине науку о подацима и могућностима њихове практичне примене.					
Исход предмета Савладане методе и технике оспособиће студента за: Припрему података за анализу (data wrangling) – вештину довођења података у облик погодан за визуелизацију и моделовање; Ефикасно управљање подацима; Истраживачку анализу података, постављање статистичких хипотеза и извођење закључака; Предвиђање помоћу модела базираних на подацима; Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје.					
Садржај предмета <i>Теоријска настава</i> Увод. Дефиниције. Интердисциплинарност. Подаци. Процеси: спецификација и разумевање проблема, 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. Типови 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. PCA - 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. Остали приступи кластеризацији. Кластеризација категоријских података. FRM (Frequency Recency Monetary) анализа. Комуникацију резултатима кроз визуелизацију података и ефективне сумарне извештаје. Преглед блиских тема. Data Mining у текстуалним документима. Sentiment Analysis. Association rules. Обрада и препознавање слика. Временске серије. Internet of Things. Big Data. 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.					
Литература 1. Wickham, Hadley, and Garrett Grolemond, R за статистичку обраду података, 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.					
Број часова	активне наставе	Теоријска настава:	3	Практична настава:	2
Методе извођења наставе Проблемски-оријентисана настава, практична настава и вежбе уз софтверску подршку, самостални рад студената, домаћи задаци и консултације.					
Оцена знања (максимални број поена 100)					
Предиспитне обавезе	70	поена	Завршни испит	30	поена
колоквијуми	20 + 20		писмени испит	20	
семинар	30		усмени испит	10	