

1.	Наслов на наставниот предмет	Подготовка и истражувачка анализа на податоци Wrangling and exploratory data analysis
2.	Код	СДП-3-4
3.	Студиска програма	Статистика за аналитика на податоци, Статистика за аналитика на податоци
4.	Организатор на студиската програма (единица, односно институт, катедра, оддел)	Факултет за информатички науки и компјутерско инженерство
5.	Степен (прв, втор, трет циклус)	втор циклус
6.	Академска година / семестар 5 / зимски /	7. Број на ЕКТС кредити 6
8.	Наставник	проф. д-р. Жанета Попеска, проф. д-р. Марија Михова
9.	Предуслови за запишување на предметот	
10.	Цели на предметната програма (компетенции): Подготовката (wrangling) на податоците троши приближно 50-80% од времето на аналитичарот, пред да биде можна било каква анализа. Курсот поминува низ процесот на подготовка со истражување на неколку фактори - време, грануларност, обем и структура - се за што треба да се размислува кога ќе се почне да се работи со податоци. Студентите ќе разберат какви податоци се достапни, да изберат кои податоци да ги користат и на кое ниво на деталност, да комбинираат повеќе извори на податоци, да одлучат како да ги извадат резултатите за да ја поедностават понатамошната анализа. Вториот дел од курсот ги опфаќа суштинските истражувачки техники за сумирање на податоци. И овие техники се применуваат пред да започне формалното моделирање и помага како информација при развојот на посложените статистички модели.	
11.	Содржина на предметната програма: Процес на подготовка на податоци за анализа: откривање, структурирање, чистење, збогатување, валидирање, објавување. Чекори во истражувачката анализа на податоци. Аналитички графици. Истражувачки графици. Основи на аналитичка графика и базичен систем за цртање. Графички уреди. Работа со бои. Хиерархиско кластерирање, K најблиски соседи, редукција на димензионалност. Практични вежби во R.	
12.	Методи на учење: Предавања поддржани со презентации преку слајдови, интерактивни предавања, вежби (користење на софтверски пакети), самостојна изработка и одбрана на проектна задача и семинарска работа	

13.	Вкупен расположив фонд на време	6 ECTS x 30 часа = 180 часа		
14.	Распределба на расположивото време	60 + + 45 + 45 + 30 = 180 часа		
15.	Форми на наставните активности	15.1.	Предавања- теоретска настава настава	60 часови
		15.2.	Вежби (лабораториски, аудиториски), семинари, тимска работа	часови
16.	Други форми на активности	16.1.	Проектни задачи	45 часови
		16.2.	Самостојни задачи	45 часови
		16.3.	Домашно учење	30 часови
17.	Начин на оценување			
	17.1.	Тестови	бодови	
	17.2.	Семинарска работа/ проект (презентација: писмена и усна)	бодови	
	17.3.	Активности и учење	бодови	
	17.4.	Завршен испит	бодови	
18.	Критериуми за оценување (бодови/ оценка)	до 50 бода		5 (пет) (F)
		од 51 до 60 бода		6 (шест) (E)
		од 61 до 70 бода		7 (седум) (D)
		од 71 до 80 бода		8 (осум) (C)
		од 81 до 90 бода		9 (девет) (B)
		од 91 до 100 бода		10 (десет) (A)
19.	Услов за потпис и полагање на завршен испит			
20.	Јазик на кој се изведува наставата	Македонски – Англиски		
21.	Метод на следење на квалитетот на наставата	механизам на интерна евалуација и анкети		
22.	Литература			
	22.1.	Задолжителна литература		

	Ред.бр.	Автор	Наслов	Издавач	Година
	1	Tye Rattenbury, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, Connor Carreras	Principles of Data Wrangling: Practical Techniques for Data Preparation	O'Reilly	2017
	2	Bradley C. Boehmke	Data Wrangling with R	Springer	2016
	3	John Wilder Tukey	Exploratory Data Analysis	Addison- Wesley	1977
	4	Roger Peng	Exploratory Data Analysis with R	Leanpub	2016
22.2.	Дополнителна литература				
	Ред. број	Автор	Наслов	Издавач	Година