РЕПУБЛИКА БЪЛГАРИЯ

**СТОПАНСКА АКАДЕМИЯ “Д. А. ЦЕНОВ” – СВИЩОВ**

|  |  |
| --- | --- |
| Факултет“СТОПАНСКА ОТЧЕТНОСТ” | Катедра  “СТАТИСТИКА И ПРИЛОЖНА МАТЕМАТИКА” |
|  |  |
| **Утвърждавам:** |  |
| Декан: …...…………….............................  (проф. д-р Атанас Атанасов) | Приета от ФС, Протокол № 1 от 14.09.2022 г. Приета от КС, Протокол № 1 от 07.09.2022 г. |
| **УЧЕБНА ПРОГРАМА**  **на** | |
| Учебна дисциплина“Анализ на големи данни ” | |
| **Код на дисциплината: ФСО-КСПМ-М-326**  **Брой кредити по учебен план: (6)** | |
| Образователно-квалификационна степен: МАГИСТЪР | Код на документа:УД/УПР-ФСО-КСПМ-М-326 |
| Форма на обучение: РЕДОВНА/ЗАДОЧНА/ДИСТАНЦИОННА Език: БЪЛГАРСКИ | Версия:v.01/2022 |

**І. ОРГАНИЗАЦИЯ НА ОБУЧЕНИЕТО**

**Часове учебна заетост (семестър с продължителност 14 седмици)**

***Таблица № 1***

|  |  |  |  |
| --- | --- | --- | --- |
| **Учебна заетост**  **/аудиторна и извънаудиторна/** | **Редовна форма на обучение** | **Задочна форма на обучение** | **Дистанционна форма на обучение** |
| **1. Аудиторна заетост (АЗ)** | **56** | **28** | **56** |
| 1.1. Лекции | 42 | 21 | 42 |
| 1.2. Семинарни занятия | 14 | 7 | 14 |
| **2. Извънаудиторна заетост (ИАЗ)** | **94** | **122** | **94** |
| 2.1.Самостоятелна работа | 47 | 61 | 47 |
| 2.2. Академични задания | 47 | 61 | 47 |
| 2.2.1. Курсови разработки и проекти | 23 | 27 | 23 |
| 2.2.2. Есета/доклади | 0 | 0 | 0 |
| 2.2.3. Казуси и делови игри | 10 | 14 | 10 |
| 2.2.4. Онлайн тестови и изпитни модули | 14 | 20 | 14 |
| **Всичко:** | **150** | **150** | **150** |

**Схема за формиране на крайната оценка по дисциплина**

***Таблица № 2***

|  |  |  |  |
| --- | --- | --- | --- |
| **Критерий** | **Тежест на критерия** | | |
| **Редовна форма на обучение** | **Задочна форма на обучение** | **Дистанционна форма на обучение** |
| **1. Изпълнение на ангажименти през семестъра, в т.ч.:** | **50%** | **50%** | **50%** |
| * 1. Посещение на учебни занятия   *(% от комплексната оценка)* | 5% | 5% | 5% |
| 1.2. Семестриални контролни  *(% от комплексната оценка)* | 25% | 25% | 25% |
| 1.3. Академични задания  *(% от комплексната оценка)* | 20% | 20% | 20% |
| **2. Семестриален изпит**  *(% от комплексната оценка)*  *Форма на провеждане:*  *Поливариантен тест с теоретични въпроси и практически задания* | **50%** | **50%** | **50%** |
| **Общо за дисциплината** | **100%** | **100%** | **100%** |

**II. Анотация**

**2.1. Цел на курса**

Дисциплината „Анализ на големи данни“ е включена в учебния план на магистърска програма „Бизнес анализ на данни със специализиран софтуер" (съвместна програма), която се реализира съвместно с Университета за национално и световно стопанство – София в рамките на проект BG05M2OP001-2.016-0004-C01 „Икономическото образование в България 2030“, финансиран по ОП „Наука и образование за интелигентен растеж“, съфинансиран от Европейския съюз чрез Европейските структурни и инвестиционни фондове.“

Целта на обучението по дисциплината е студентите да усвоят теоретичните аспекти на понятието „големи данни“, да придобият знания за източниците на големи данни и подходите за тяхната обработка, да формират практически умения за анализ на големи данни. Като база служат статистическите методи и модели. Като софтуерен инструмент се използва програмният продукт R.

**2.2. Предварителни изисквания**

Необходими са базисни знания по статистика, както и умения за работа със софтуерни продукти за обработка на данни.

**2.3. Използвани методи на преподаване**

***2.3.1 Редовна и задочна форма***

При обучението на студентите се използва дигитално учебно съдържание, за усвояването на което се прилагат различни традиционни и дигитални методи на обучение и оценяване на придобитите знания и умения (традиционни и видео лекции, казуси, симулации, ролеви игри, дебати, дискусии, демонстрации, мозъчни атаки, директни инструкции, кооперативно учене, независими и групови проекти, учене чрез преживяване, дигитални инструменти, интерактивни методи на преподаване). Учебните занятия се провеждат както във вид на традиционни лекции чрез използване на съвременни технически средства за презентация (интерактивни дъски и екрани, холограмни проектори, устройства за виртуална реалност и др.), така и във вид на синхронни лекции и семинарни занятия във виртуални класни стаи. Записите на занятията и дигиталните обучителни ресурси се съхраняват в облачно базирани библиотеки с учебни материали, което дава възможност за асинхронен достъп до тези ресурси в удобно за обучаемите време.

***2.3.2 Дистанционна форма***

При обучението на студентите се използва дигитално учебно съдържание, за усвояването на което се прилагат различни традиционни и дигитални методи на обучение и оценяване на придобитите знания и умения (традиционни и видео лекции, казуси, симулации, ролеви игри, дебати, дискусии, демонстрации, мозъчни атаки, директни инструкции, кооперативно учене, независими и групови проекти, учене чрез преживяване, дигитални инструменти, интерактивни методи на преподаване). Учебните занятия се провеждат във вид на синхронни лекции и семинарни занятия във виртуални класни стаи. Записите на занятията и дигиталните обучителни ресурси се съхраняват в облачно базирани библиотеки с учебни материали, което дава възможност за асинхронен достъп до тези ресурси в удобно за обучаемите време.

**2.4. Очаквани резултати**

След успешното завършване на курса студентите ще имат знания за източниците на големи данни и умения да ги визуализират и анализират, като използват за обработката им софтуерния продукт R.

**III. разпространение на дисциплината**

1. Софийски университет "Св. Климент Охридски", България - "Предварителна обработка на големи данни при провеждане на аналитично изследване"

2. Университет за национално и световно стопанство, София, България - "Съхранение и обработка на големи данни", "Големи данни с Python"

3. Икономически университет, Варна, България - "Големи данни", "Складове от данни и big data"

4. University of California San Diego - "Introduction to Big Data", "Big Data Modeling and Management Systems"

5. Hong Kong Polytechnic University - "Knowledge Management and Big Data in Business"

**IV. Учебно съдържание**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ТемИ** | **Редовна форма на обучение** | | **Задочна форма на обучение** | | **Дистанционна форма на обучение** | |
| **Л** | **У** | **Л** | **У** | **Л** | **У** |
| **${temi#1}Тема I. ТЕОРЕТИЧНИ ОСНОВИ НА АНАЛИЗА НА ГОЛЕМИ ДАННИ** | 5 | 0 | 2.5 | 0 | 5 | 0 |
| *1.1. Дефиниция на големи данни и основни понятия 1.2. Характеристики на големите данни 1.3. Видове големи данни 1.4. Жизнен цикъл на големите данни* |
| **${temi#2}Тема II. ИЗТОЧНИЦИ НА ГОЛЕМИ ДАННИ** | 5 | 0.2 | 2.5 | 0.1 | 5 | 0.2 |
| *2.1. Големи данни, генерирани от населението  2.2. Големи данни, генерирани от информационни системи 2.3. Големи данни, генерирани от сензорни устройства 2.4. Умна доверена статистика (Trusted smart statistics)* |
| **${temi#3}Тема III. МЕТОДОЛОГИЧЕСКИ АСПЕКТИ НА ГОЛЕМИТЕ ДАННИ В ОФИЦИАЛНАТА СТАТИСТИКА** | 5 | 0.2 | 2.5 | 0.1 | 5 | 0.2 |
| *3.1. Големите данни в контекста на статистическото изследване 3.2. Предизвикателства и подходи за тяхното решаване 3.3. Оценка на качеството на големите данни в статистическата практика* |
| **${temi#4}Тема IV. ТЕХНОЛОГИЧНИ РЕШЕНИЯ ПРИ ОБРАБОТКАТА НА ГОЛЕМИ ДАННИ** | 5 | 0.2 | 2.5 | 0.1 | 5 | 0.2 |
| *4.1. Технологии за обработка на големи данни 4.2. Извличане на данни от уебсайтове (Webscraping) 4.3. Софтуерни инструменти за обработка на големи данни 4.4. Европейска тестова ИТ инфраструктура за големите данни* |
| **${temi#5}Тема V. ВЪВЕДЕНИЕ В R** | 4 | 2.4 | 2 | 1.2 | 4 | 2.4 |
| *5.1. Обща характеристика на R и основни понятия 5.2. Типове данни в R 5.3. Елементи на програмиране в R* |
| **${temi#6}Тема VI. ПРЕДВАРИТЕЛНА ОБРАБОТКА НА ГОЛЕМИ ДАННИ** | 4 | 2.5 | 2 | 1.2 | 4 | 2.5 |
| *6.1. Екстрахиране 6.2. Трансформация 6.3. Зареждане* |
| **${temi#7}Тема VII. ВИЗУАЛИЗАЦИЯ НА ГОЛЕМИ ДАННИ** | 4 | 2.5 | 2 | 1.3 | 4 | 2.5 |
| *7.1. Същност и предназначение на визуализацията на големи данни 7.2. Видове визуализация на големи данни 7.3. Принципи за визуализация* |
| **${temi#8}Тема VIII. АНАЛИЗ НА ГОЛЕМИ ДАННИ ЧРЕЗ МАШИННО ОБУЧЕНИЕ** | 5 | 3 | 2.5 | 1.5 | 5 | 3 |
| *8.1. Същност на машинното обучение. Основни понятия 8.2. Подходи за анализ чрез машинно обучение 8.3. Анализ при количествена резултативна променлива (регресия) 8.4. Анализ при качествена резултативна променлива (класификация)* |
| **${temi#9}ТЕМА IX. ПРОИЗВОДСТВО НА ЕКСПЕРИМЕНТАЛНА СТАТИСТИКА ЧРЕЗ ДАННИ ОТ ИНТЕРНЕТ ИЗТОЧНИЦИ (ПРАКТИЧЕСКИ АСПЕКТИ)** | 5 | 3 | 2.5 | 1.5 | 5 | 3 |
| *9.1. Дефиниране на случаи на използване. Източници на данни за свободните работни места и характеристиките на предприятията 9.2. Етапи на производствения процес 9.3. ИТ инфраструктура 9.4. Анализ на данните с Python и методи за машинно обучение 9.5. Производство на експериментална статистика* |
| **${temi\_sum#1}Общо:** | **42** | **14** | **21** | **7** | **42** | **14** |

**V. ИЗПОЛЗВАНИ УЧЕБНО-ТЕХНИЧЕСКИ СРЕДСТВА**

|  |  |  |
| --- | --- | --- |
| **Наименование на**  **Учебно-техническото средство** | **Използвани учебно-технически средства по вид на занятията** | |
| **Лекции** | **Семинарни**  **занятия** |
| 1. Мултимедийни системи за презентиране | X | X |
| 2. Интернет | X | X |
| 3. Специализирани програмни продукти: |  |  |
| 3.1  R Studio | X | X |
| 3.2  Python | X | X |

**VI. ПреПОРЪЧИТЕЛНИ МАТЕРИАЛИ И НОРМАТИВНИ ИЗТОЧНИЦИ**

**6.1. Основна литература**

|  |
| --- |
| 1. Учебен курс в Платформата за дистанционно и електронно обучение на СА “Д. А. Ценов“ |
| 2. Статева, Г. (2016). Статистическите изследвания и „големите данни“: допълващи се източници или конкуренти. В: Годишен алманах Научни изследвания на докторанти, книга 10, с.169-190. Свищов: АИ Ценов. |
| 3. Богданов, Б., Статева, Г. (2016). Въздействието на големите данни (Big Data) върху официалната статистика: възможност или провокация. Статистика, № 4, с. 9-33. |
| 4. Йорданова, С., Стефанова, К. (2019). Предизвикателствата на големите данни - същност, характеристики, технологии.  Научни трудове на УНСС, том 1/2019, ИК – УНСС, София |
| 5. Бизев, Н. (2016). Какво представляват големите данни? Публични политики, кн. 1., с. 60-66. http://ejpp.eu/index.php/ejpp/article/viewFile/218/19 |
| 6. Irizarry, R. A. (2019). Introduction to data science: Data analysis and prediction algorithms with R. CRC Press. https://rafalab.github.io/dsbook/ |
| 7. Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data mining, inference and prediction. Second edition. Springer. https://hastie.su.domains/Papers/ESLII.pdf |
| 8. Oleksy, A. (2018). Data science with R: a step by step guide with visual illustrations and examples. Independently published by Andrew Oleksy |
| 9. Wiskman, H., Grolemund, G. (2017). R for data science: visualize, model, transform, tidy and import data. O’Reily Media. |
| 10. Boyd D., Crawford K. (2012). Critical questions for Big Data, 2012 https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878 |

**6.2. Допълнителна литература**

|  |
| --- |
| 1. Иванов, М. (2016). Съвременни методи за интелигентен анализ на данни. Научен електронен архив на НБУ, София. https://eprints.nbu.bg/id/eprint/3082/1/%D0%98%D0%BD%D1%82%D0%B5%D0%BB%D0%B8%D0%B3%D0%B5%D0%BD%D1%82%D0%B5%D0%BD%20%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7%20%D0%BD%D0%B0%20%D0%B4%D0%B0%D0%BD%D0%BD%D0%B8%20-%20%D0%BF%D1%83%D0%B1%D0%BB%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F.pdf |
| 2. Статева, Г. (2016). Оценка на качеството на „големите данни“ в официалната статистическа практика. В: Годишен алманах Научни изследвания на докторанти, книга 11, с. 448-458. Свищов: АИ Ценов. |
| 3. Popov, V., Emilova, P. (2020). Big data analytics in support of information security. In: Online International Scientific Practical Conference 1st Edition, December 11, 2020, Chișinău, Moldova |
| 4. Лобанов А. (2014). Большие данные: проблемы обработки. Вестник МГТУ МИ- РЭА, №3, с. 51-58. |
| 5. Попов, В. (2016). Анализ на Big Data - методи, технологии и инструменти. В: Предизвикателства пред информационните технологии в контекста на "Хоризонт 2020" - юб. науч. конфер., 7-8 окт. 2016 г., Сб. с докл. Свищов, АИ Ценов, с. 101-107. |
| 6. Иванов П. , Лопуховский А. (2014). Технологии Big Data и различные методы представления больших данных. Инженерный журнал: наука и инновации, вып. 9. http://engjournal.ru/catalog/it/asu/1229.html |
| 7. Chikhale К. (2014). Data Mining: Exploring Big Data Analytics, Hadoop and Mapreduce. International journal of engineering science and research technology, 3(8). |
| 8. Antenucci, D., Cafarella, M., Levenstein, M., Re, C, Shapiro, M. (2014). Using social media to measure labor market flows. NBER working paper series, workig paper 20010. https://www.nber.org/papers/w20010 |
| 9. Buyya, R., Calheiros, R., Dastjerdi, A. (2016). Big Data. Principles and Paradigms. http://dhoto.lecturer.pens.ac.id/lecture\_notes/internet\_of\_things/Big%20Data%20Principles%20and%20Paradigms.pdf |

**6.3. Нормативни документи**



**6.4. Интернет ресурси**

|  |
| --- |
| 1. https://cran.r-project.org/ |
| 2. https://www.rstudio.com/ |
| 3. https://r4ds.had.co.nz/ |
| 4. https://ec.europa.eu/eurostat/cros/content/essnet-big-data-1\_en |
| 5. https://ec.europa.eu/eurostat/cros/essnet-big-data-2\_en?fbclid=IwAR2DXSQ9-kUOOpjeP2a07RXVDIginrSacmWWXudrbFZnzsXR8LtfNjClJ00 |
| 6. https://rafalab.github.io/dsbook/ |

Съставил/и/:

|  |
| --- |
| …………………………………  ${authors#1}(доц. д-р Маргарита Шопова) |
| …………………………………  ${authors#2}(доц. д-р Красимира Славева) |

Ръководител катедра:

|  |  |
| --- | --- |
| …………………………………  (доц. д-р Пламен Петков) |  |