



Co-funded by the
Erasmus+ Programme
of the European Union

Project Title	pArtneRship foR AddressiNG mEgatrends in ICT
Project Acronym	ARRANGE-ICT
Start of the project	1.9.2018
End of the project	30.11.2020
Name of Coordinator	Technical University of Sofia
Contract Number	2018-1-BG01-KA203-048023



ARRANGE-ICT
pArtneRship foR AddressiNG mEgatrends in ICT

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Заглавие на проекта	ARRANGE-ICT - PARTNERSHIP FOR ADDRESSING MEGATRENDS IN ICT
Акроним на проекта	ARRANGE-ICT
Инструмент	„ЕРАЗЪМ+“

Непосредствен резултат	Непосредствен резултат 5 – Осъвременени учебни програми и учебни планове, с които ефективно да се отговори на най-
Водещ автор (организация)	Джордж Палис, Мариос Дикаякос (Кипърски университет)
Сътрудник(-ци) (организация)	Димостенис Стефанидис (Кипърски университет)
Краен срок за подаване	
Дата на реално подаване	
Версия	1.0

Разпространение

<input checked="" type="checkbox"/>	PU: Публично (*онлайн платформа)
<input type="checkbox"/>	PP: Ограничено до другите участници в програмата (включително Комисията)
<input type="checkbox"/>	RE: Ограничено до група, посочена от консорциума (включително Комисията)
<input type="checkbox"/>	CO: Поверителна информация, само за членовете на консорциума (включително Комисията)

Версии и история на публикуваните раздели

Версия	Дата	Автор	Бележки
0.1	01.06.2020 г.	Кипърски университет	Първоначална версия

Отказ от отговорност

Настоящият документ съдържа материали и информация, които са собственост на Консорциума ARRANGE-ICT, поверителни са и не могат да бъдат копирани, възпроизведени или изменени изцяло или частично за каквато и да е цел без предварителното писмено съгласие на Консорциума ARRANGE-ICT.

Въпреки че материалите и информацията, съдържащи се в настоящия документ, се считат за прецизни и точни, нито координаторът на проекта, нито който и да е партньор на Консорциума ARRANGE-ICT, нито което и да е лице, действащо от името на някой от партньорите на Консорциума ARRANGE-ICT, не дават никакви гаранции или декларации, било то изрични или подразбиращи се, по отношение на използването на материалите, информацията, метода или процеса, оповестени в настоящия документ, включително продаваемостта и годността за определена цел, или че тази употреба не нарушава или не пречи на частни права.

Освен това нито координаторът на проекта, нито който и да е партньор на Консорциума ARRANGE-ICT, нито което и да е било лице, действащо от името на някой от партньорите на Консорциума ARRANGE-ICT, не носи отговорност за преки, косвени или произтичащи загуби, щети, отправени претенции или натрупани разходи, произтичащи от или във връзка с каквато и да е информация, материал, съвет, неточност или пропуск, съдържащи се в настоящия документ.

Съдържание

1. Въведение	4
2. Определяне на уменията.....	8
3. Изучаване на успешни учебни програми, които имат отношение	10
4. Разработване на осъвременената учебна програма по науката за данните	11
5. Прилагане на осъвременената учебна програма по Наука за данните в Кипърския университет.....	15
6. Кратки програми	31
7. Заключение	35

1. Въведение

ARRANGE-ICT работи върху установяването и записването на големите тенденции на развитие, прокарвани от новаторите на пазара, за да се помогне на висшите учебни заведения да следват бързия темп на нововъзникващите браншове. Основният приоритет на ARRANGE-ICT е да се повиши качеството и приложимостта на знанията и уменията на студентите, тъй като проектът има за цел да подкрепи ВУЗ-овете. Този приоритет е актуален и навременен, тъй като растящата дигитализация и произтичащите от нея големи налични масиви от данни, съчетани с ускоряващия се напредък в изкуствения интелект и машинното самообучение, водят до значителни трансформации в икономическите и социалните дейности, които революционизират заетостта и търсените умения, за поддържане на икономическата дейност^{1,23}

Първата ключова тенденция е стратегическата значимост на **цифровите данни** в широк спектър от икономически и научни дейности^{4,5} Тъй като огромни количества данни стават достъпни и могат да бъдат анализирани, все повече аспекти на икономиката, обществото и ежедневието стават зависими от данните. Наличието на данни и зависимостта от тях правят работата в почти всички области все по-често движена от данни и създават повече възможности и нужди за решаване на проблеми и вземане на решения, основани на данни. Тази тенденция разкрива злоещо *количествено несъответствие в уменията* на работната сила, тъй като на все повече работници и служители възлагат задачи да участват и да си сътрудничат в дейности, свързани с цикъла за решаване на проблеми на PPDAC (Разбиране и определяне на проблема; Планиране на събирането и регистриране на данни; Събиране, управление и отсяване на невалидните данни; Анализ на данни и изготвяне на хипотези; и Извличане на заключения и тълкуване на данните)⁶. Участието в тези дейности изисква разбиране и опит с разнообразни усъвършенствани умения и инструменти в областта на информационните технологии, статистиката и прилагането им в различни промишлени и научни области. Тъй като изобилието

¹ „L „Avenir du travail“. Jacques Attali. 2007. Fayard.

² “The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies.” Erik Brynjolfsson and Andrew McAfee. 2014. Norton.

³ “The Age of Analytics: Competing in a Data-driven World.” McKinsey & Company, December 2016.

⁴ “The New Digital Age.” Eric Schmidt and Jared Cohen. 2013. Knopf.

⁵ “The world’s most valuable resource is no longer oil, but data.” Economist, May 6, 2017.

⁶ “The Art of Statistics. Learning from Data.” David Spiegelhalter. 2019. Penguin.

от данни и разпространението на мощни инструменти за анализ на данни обаче увеличава нашата зависимост от подходите, основани на данни, тяхното възприемане често върви ръка за ръка с дълбоки последици, които пораждаат непредвидени, но сериозни етични опасения или вредни последици⁷. Този риск показва необходимостта от включване *на качествени умения* в решаването на проблеми и вземането на решения за подходяща контекстуализация и внимателно тълкуване на количествените резултати^{8,9}

Автоматизацията е друга основна тенденция, произтичаща от напредъка в дигитализацията и изкуствения интелект. Автоматизацията носи значителни ползи в производителността, но също така ще доведе до мащабна замяна на хората с работи и интелигентен софтуер в много задачи, както физически, така и когнитивни. Следователно, и за разлика от предишните индустриални революции, при които новите технологии бяха свързани с дългосрочно създаване на работни места и растеж на заплатите, в няколко проучвания се прогнозира, че четвъртата индустриална революция ще има голям отрицателен ефект върху работните места, което ще създаде значителни социални проблеми и ще засили икономическото неравенство вътре в държавите и между тях. Прогнозите сочат, че до 2030 г. работодателите ще се нуждаят от 20 – 25% по-малко служители, като 30 – 50% от работните дейности ще стане технически възможно да бъдат автоматизирани¹⁰. Безработицата, предизвикана от автоматизацията, се очаква да засегне както служителите, които се занимават с интелектуален труд, така и работниците, заети с физическа работа. Във вътрешен план, това ще окаже допълнителен натиск върху работниците, които ще станат свидетели на намаляване на възможностите им за [колективно] договаряне, на трудовите им права, на заплатите и сигурността на работните им места. В международен аспект, възможността за производство в заводи почти без работници в страни с високи разходи за труд може да намали конкурентното предимство на страни с по-ниски разходи за труд да привличат чуждестранни инвестиции, да създават работни места и да търсят път към икономически

⁷ "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy." Cathy O'Neil. 2016. Crown.

⁸ "Big Data needs Thick Data." Tricia Wang, May 13, 2013. <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>

⁹ "Spreadsheets can't predict behaviour." Tassos Stassopoulos. Trinetra, July 20, 2020. <https://www.trinetra-im.com/post/spreadsheets-cant-predict-behaviour>

¹⁰ "Technology and Jobs. Coming to an Office Near You." Economist, Jan. 18, 2014.

растеж, воден от износа¹¹. Социалните и икономическите трансформации, предизвикани от автоматизацията, на свой ред променят търсенето на умения. С постепенното автоматизиране на производствените и други задачи, които не изискват кой знае какви умения, необходимостта от рутинни когнитивни и занаятчийски умения намалява, а нараства търсенето на умения за обработка на информация и други когнитивни и междуличностни умения от високо равнище. Следователно, и в допълнение към овладяването на умения, специфични за професията и в областта на ИТ, работниците и служителите в 21-ви век също трябва да имат набор от различни „общи“ умения, включително междуличностна комуникация, самоуправление и способността да се учат, с помощта на които да издържат на несигурността на бързо променящия се пазар на труда¹².

Друга важна тенденция, която възниква в края на 20^{-ти} и началото на 21^{-ви} век е много страни по света да се опитват да насърчават дейности в сферата на **иновациите, ориентирани към предприемачеството**. Тези усилия обикновено се насърчават от стремежа да се приеме културата на Силициевата долина и да се възпроизведе нейното икономическо въздействие. Този стремеж изглежда разумен – многобройни икономически проучвания показаха, че технологичните промени повишават нивата на доходите, съществува силна връзка между високите нива на създаване на интелектуална собственост и растежа на БВП и има положително въздействие на иновациите върху производителността и финансовите резултати на предприятията¹³. Действително, с ускоряването на темпа на научно-техническия прогрес и с разширяването на тяхното въздействие върху икономическите и социалните дейности, способността на хората бързо да се адаптират и да се възползват от новите условия и разработки, като се ангажират с иновативно и творческо предприемачество, се превръща в стратегически приоритет за нациите, които желаят да насърчават и дават възможност за проявление на въображението и

¹¹ "Universities as Knowledge Platforms in a Period of Disruption." Marios D. Dikaiakos. Invited talk Conference on The Role of the University in the Crisis of Forced Displacement: Ethics, Innovation, and Immersive Learning, Boston University, September 26, 2019. <https://medium.com/swlh/universities-as-knowledge-platforms-in-a-period-of-disruption-c68062edf7f1>

¹² "OECD Skills Outlook 2019: Thriving in a Digital World." OECD. OECD Publishing, Paris, 2019. <https://doi.org/10.1787/df80bc12-en>

¹³ "Knowledge, Networks and Nations: Global scientific collaboration in the 21st century." Royal Society, March 2011.

таланта на своите хора, като им помагат да развиват творческите си способности, способностите си да генерират нови идеи, и да дадат началото на нови отрасли¹⁴. Въпреки това, приближаването до уникалните условия, довели до създаването и успехите на предприемаческата екосистема на Силициевата долина, не може лесно да бъде възпроизведено на различно място и по различно време. Обучението по иновативно мислене и за придобиване на предприемачески умения обаче може да помогне на хора с различно образование да набележат и да потърсят възможности и да поемат рискове в различен контекст, което да доведе до нови бизнес начинания или социални предприятия с глобално или местно въздействие.

В този широк контекст науката за данните бързо се превръща в област от основно значение за стратегията на съвременните организации и се превърна в мегатенденция в ИКТ. Днес науката за данните се превърна в свещения граал на уменията, нужни за пригодност за заетост. Важността и възможностите, присъщи на науката за данните, са ясни (вж. <http://cra.org/data-science/>). Както се посочва в статия, публикувана наскоро в SACM¹⁵, *“висшите учебни заведения в САЩ признават, че науката за данните е критично умение изследванията през 21-ви век и работна сила на 21-ви век. Във висшето образование учебните програми по науката за данните имат две аудитории: Нови специалисти в областта на науката за данните, както и учени и специалисти, които се нуждаят от умения в областта на науката за данните, за да допринесат за други области на ИКТ“*. През 2011 г. институтът Маккинзи нарича големите информационни масиви – следващата територия за покоряване. От тогава до сега *„Науката за данните се е развила, за да представлява неразделна част от цифровата трансформация и технологичните иновации за следващия свят. Това означава, че не само, че търсенето на специалисти в областта на науката за данните се увеличава с всеки изминал ден, а и, че получават високи заплати поради голямото търсене на пазара. Всеки отрасъл от сферата на бизнеса, финансите, държавното управление, здравеопазването,*

¹⁴ “Thank you for Being Late.” Thomas Friedman. 2016. Farrar, Straus and Giroux.

¹⁵ “Realizing the potential of data science.” Francine Berman, Rob Rutenbar, Brent Hailpern, Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Margaret Martonosi, Padma Raghavan, Victoria Stodden, and Alexander S. Szalay. 2018. Commun. ACM 61, 4 (April 2018), 67–72. DOI:<https://doi.org/10.1145/3188721>

социалните мрежи и технологиите търси хора с умения в областта на науката за данните“ (вж. <https://www.analyticsinsight.net/top-10-universities-in-usa-offering-ph-d-in-data-science/>).

В този отчетен резултат сме разработили съвременна учебна програма в областта на науката за данните. Очаква се тази учебна програма да бъде изпълнявана от Кипърския университет. Методологията, която следвахме, за да разработим осъвременената учебна програма в областта на науката за данните, може да бъде обобщена, както следва:

- Определяне на уменията, които студентите трябва да развият в следдипломната програма по наука за данните;
- Изучаване на успешни учебни програми в много престижни академични институции по целия свят (напр. Технологичния институт на Джорджия и Нюйоркския университет в САЩ, Амстердамския университета и Университета Бокони в Европа).
- Разработване на осъвременената учебна програма по науката за данните;
- Прилагане на осъвременената учебна програма по наука за данните в Кипърския университет.

Разделите по-долу представят подробно следваната методология.

2. Определяне на уменията

Науката за данните по своята същност е интердисциплинарна. Работата с данни изисква познаване и разбиране на различни умения и концепции, свързани с областите на статистиката, информатиката и мениджмънта. Който завърши Наука за данните трябва да бъде подготвен да работи с данни на всички етапи на едно проучване и от него ще се очаква да работи в екипна среда.

По-конкретно, необходимостта от програма по Наука за данните между различни катедри е породена именно от тази нужда учените и мениджърите от различни специалности да мислят в различни дисциплини:

- Специалистите по бизнес анализи трябва да разбират основните принципи на инструментите, които използват (необходимост от гледна точка на статистиката), и често се налага да се занимават с практически въпроси по изпълнение на

проекти, особено когато се занимават с нарастващия обем данни, които съвременните организации генерират (необходимост от умения в областта на информатиката).

- Информатиците специалисти по наука за данните трябва в крайна сметка да се грижат за утилитарното - икономическо или друго - въздействие от работата си и да могат ясно да комуникират, колко е ценна на всички заинтересовани страни (необходимост от гледната точка на бизнес анализа). В същото време тези специалисти не са защитени от това да злоупотребяват с инструментите на своя занаят, ако не разбират достатъчно научната основа и математиката на областта (необходимост от статистическа гледна точка).
- Статистиците специалисти по наука за данните често са натоварени с решаването на сложни и многостранни теоретични задачи, които често налагат прилагане на инженерен подход за стекове с данни за анализ, с които могат да се обработват данни с много висока пропускателна способност (необходимост от гледна точка на информатиката), като те също трябва често да осигуряват резултати в рамките на организационните приоритети, при които се цени създадената бизнес стойност и често полезността и други практически съображения се поставят над прецизността или методологична чистота (необходимост от гледна точка на бизнес анализа).

Програмата за следдипломна квалификация по Наука на данните се стреми да:

- Предложи на студентите възможност да придобият задълбочени знания, практически опит и изследователски опит в една или повече области на науката за данните.
- Подготви завършилите да могат да продължат кариерата си на отговорни позиции в академичните среди или в промишлеността, където ефективно ще стимулират развитието и прилагането на нови методи и идеи.
- Предложи на студентите образование в най-широкия смисъл на думата, да събуди у тях желанието за непрекъснато учене, което от своя страна води до зрялост и развива способността за независимо и критично мислене.
- Помогнете на своите бъдещи възпитаници да придобият дълбоко разбиране на наука за данните, както като наука, така и по отношение на по-общите ѝ приложения и ефекти върху обществото.

- Подготви своите бъдещи възпитаници така, че независимо от кариерния път, който избират, наученото да им даде добра основа, за да бъдат в крак с невероятно бързото научно и технологично развитие в областта на компютърните науки.
- Осигури умения в научната и техническата комуникация (както писмена, така и устна), събиране и трупане на практическа информация, работа в сътрудничество, творчество и иновативно мислене, които могат да се използват в различни професии.

3. Изучаване на успешни учебни програми, които имат отношение

Имаше бум на програмите по Наука за данните в академичните среди, след като ръководителите на университетите се втурнаха да отговорят на търсенето. На уебсайта <http://datascience.community/colleges> понастоящем са изброени 620 програми по Наука за данните, Анализи на данни и свързаните с тях области в над 200 университета по целия свят. По-голямата част от тях са магистърски програми и сертификационни програми, предлагани както традиционно, така и онлайн. Затова не е изненадващо, че в много престижни академични институции по целия свят (напр. Технологичния институт на Джорджия и Нюйоркския университет в САЩ, Амстердамския университета и Университета Бокони в Европа) са въведени интердисциплинарни програми по Наука за данните. Тези програми запазват цялостна структура, която съответства в голяма степен на предложената тук структура, което е в светлината на факта, че основната логика на необходимостта от обединяване на статистическите данни, компютърните науки и бизнес перспективите е обща за всички такива програми.

Успешната програма за следдипломна квалификация по Наука на данните трябва да предостави следните предимства:

- Цялостна мултидисциплинарна програма; Студентите ще бъдат добре запознати с различни инструменти, гледни точки и подходи, за да могат да установят най-подходящите методи и модели, които да се използват за решаване на всеки конкретен случай. В сравнение с чистите магистърски програми по една от трите дисциплини, завършилите магистърската програма по „Наука за данните“ ще получат основа и

функционални знания и в трите дисциплини (компютърни науки, математика и статистика, бизнес и публична администрация) и това ще ги постави в изгодна позиция за заетост в областта. В същото време, ще могат да продължат да се развиват в предпочитаната от тях дисциплина чрез избора на една от трите предлагани специализации (Компютърни науки/ Статистика/ Бизнес анализи).

- Студентите ще получат възможност да използват различни видове данни, за да направят сложни прогнози и изчисления в големи мащаби. В светлината на непрекъснатото увеличаване на данните по целия свят, въпросът за това как можем да използваме данните, за да получим ценна информация, е по-важен от всякога. Как можем да извлечем съответната информация от огромни количества данни? По какъв начин компютрите могат да се научат от опита, за да вземат интелигентни решения? Тези въпроси са от ключово значение за специализираната магистърска програмата по Наука за данните.
- Студентите ще имат възможност да чуят други ключови аспекти от значение за обработката и извличането на финансова полза от данните, като поверителност на данните, етика и иновативни бизнес модели.
- Студентите ще имат достъп от първа ръка до авангардни изследвания и вълнуващи възможности за стаж в областта и за кариера в различни сфери, като финанси, здравеопазване, застраховане, фармация, телекомуникации, туризъм, организиране на пътувания, транспорт и др.

4. Разработване на осъвременената учебна програма по науката за данните

Специализираният учебен план по Наука за данните, който се предлага в сътрудничество с **катедра „Компютърни науки“**, **катедра „Математика и статистика“**, както и с **катедра „Бизнес и публична администрация“**, осигурява висококачествено образование. Важно е да се отбележи, че занятията от съществуващите програми по Наука за данните се провеждат в помещенията на катедрите по компютърни науки, информатика, математика и статистика, както и управление. Много от най-успешните студенти, особено бакалаврите,

представяват широки коалиции от университета, често финансирани от интердисциплинарни институти, а не от определена катедра.

Осъвременената учебна програма по Наука за данните ще бъде магистърска програма от 1,5 години, изцяло на английски език. Научните изследвания в областта на науката за данните изискват солидни умения в областта на статистиката, управлението и съхранението на огромни количества данни, бизнес анализи, както и способността да се разработват ефективни алгоритми за анализ на данни. Тези техники се използват в сложни приложения в бизнеса, инженерството и науката. Част от програмата е проектът **„Капстоун“ в областта на Науката за данните**, в който студентите се занимават със специфични и практически проблеми на интердисциплинарните приложения. В този курс студентите се занимават с всички задачи – от моделирането на процесите до внедряването и валидирането на техниките на науката на данните. Ключов аспект на този курс ще бъде ангажираността с промишлеността и участието в публичния сектор.

Програмата е разработена да предложи 3 специализации (Информатика/ Статистика/ Бизнес анализи). Първите два семестъра ще съдържат основните курсове, а през третия студентите ще избират 3 курса по желание. Проектът „Капстоун“ ще започне през летния семестър (след края на втория семестър).

Това, което свързва тези програми, и осъвременената учебна програма е набор от общи основни курсове, както и накрая среща на всички специалности по време на окончателния проект „Капстоун“. Това общо ядро от курсове е набор от бакалавърски като ниво курсове, чиято цел е да се изравнят познанията на студентите от различни специалности и да се изгради солидна основа, на която се изграждат отделните специализации.

В зависимост от специалностите, които са следвали студентите, е естествено, част от студентите да считат някои от курсовете за по-голямо предизвикателство, отколкото други. Това може да се очаква от самото разработване на програмата и дори е вярно за мултидисциплинарните програми, предлагани в международен план, като споменатите по-горе.

Всеки от общите основни курсове е разработен около възможностите на студент с добра концентрация и различна от основната сфера на курса бакалавърска степен. Факултетът, отговорен за подготовката на общите основни курсове, е събрал допълнителна библиография, както и набор от ресурси за студенти, които могат да сметат своите курсове по-голямо предизвикателство, отколкото е било намерението в началото, и ще имат два допълнителни часа на седмица приемно време през семестъра, освен редовните три часа, които се спазват при всички други курсове.

Целева аудитория: Нови специалисти в областта на науката за данните, както и учени и професионалисти, които се нуждаят от умения в областта на науката за данните, за да имат принос в други области на ИКТ. Основната цел на тази програма е да окаже подкрепа на местната промишленост и да гарантира, че персоналът има съответните умения да прилага науката за данните на работното си място. В този контекст планираме да предложим стъпаловиден подход към образованието и обучението. Това ще включва индивидуални модули (наречени *кратки програми*), които могат да се вземат отделно и да се получават сертификати.

Език: Програмата ще бъде на английски език.

Кредити: 90

Образователни предпоставки. Успелите кандидати за магистърската програма по Наука за данните ще идват от много различни бакалавърски специалности, включително Статистика, Компютърни науки, Математика, инженерни специалности, Икономика, Бизнес, Биология, Физика и Психология. Независимо от академичната степен на кандидата, ще изискваме конкретно и съществено познаване на определени математически компетенции, както и да са преминали известно обучение по програмиране и да имат основни познания по компютърни науки. За да бъдат разгледани кандидатурите им за програмата, от студентите се изисква да са преминали обучение по следните дисциплини (или техен еквивалент):

- Основни математически понятия (напр. висша математика, линейна алгебра и т.н.)

- Въведение в компютърните науки (или еквивалентен курс за програмиране): Нямаме определени изисквания по отношение на конкретни езици, но като цяло очакваме академичен и/или професионален опит с езици за програмиране като Python, Scalla, R и др.
- Един от следните курсове: Вероятности, Статистика, инженерен курс, Иконометрика.

Планирани резултати от обучението:

Когато завършват, от студентите се очаква да:

- Имат задълбочени знания и изследователски опит в една или повече области на науката за данните чрез дисциплините, посочени в приложение 2.
- Владееят мощни инструменти, за справяне с широк спектър от теми.
- Имат статистически умения на напреднало ниво.
- Познават основните понятия в други природните и/или социалните науки, свързани с откривателството, водено от данни.
- Демонстрират задълбочено разбиране на широк спектър от дисциплини в областта на науката за данни, и да бъдат до голяма степен запознати с преобладаващи изследователски направления и авангардни проблеми.
- Проявяват гъвкавост и новаторско мислене при решаването и управлението на отворени въпроси в различни контексти, като съществено предимство за кариера в областта на научните изследвания, промишлеността, търговията, образованието и публичния сектор.
- Имат развити умения, които могат да се използват в различни професии, като: Устна и писмена научна комуникация, почти свободно използване на научен английски език, използване на информационни/комуникационни технологии, организация и планиране на групова работа.
- Имат придобит опит от независима работа, в идеалния случай в контекста на изследователски проекти, провеждани в час.
- Установяват и оценяват нуждите на една организация при задача, свързана с науката за данните.
- Събират и управляват данни, за разработване на решения на задачи, свързани с науката за данните.

- Тълкуват резултатите от анализа, с използване на техниките на науката за данните.
- Съобщават ефективно информацията, свързана с науката за данните, в различни формати, на подходяща аудитория.
- Ценят и опазват етичното използване на данните във всички аспекти на тяхната професия.
- Превръщат констатациите от ресурсите-данни в практически бизнес стратегии.

5. Прилагане на осъвременената учебна програма по Наука за данните в Кипърския университет.

В следващите абзаци е представен планът на курса с кратко описание на курсовете от учебната програма по Наука на данните, които се планира да бъдат изнесени от Кипърския университет (очаква се да започне септември 2021 г.).

Членовете на академичния състав, които са разработили програмата по Наука за данните в Кипърския университет са: Сергиос Агапиу (асистент, катедра „Математика и статистика“), Анастасия Бакшевани (доцент, катедра „Математика и статистика“), Мариос Дикаякос (професор, катедра „Компютърни науки“), Константинос Фокианос (професор, катедра „Математика и статистика“), Панос Маркопулос (асистент, катедра „Бизнес и администрация“), Христос Николайдес (преподавател, катедра „Бизнес и администрация“), Джордж Палис (доцент, катедра „Компютърни науки“).

Програмата е представена за акредитация от Кипърската агенция за осигуряване на качеството и акредитация във висшето образование (CYQAA), която е компетентният орган, отговорен за осигуряване на качеството на висшето образование в Кипър и за подкрепата на процесите, предвидени в съответното законодателство, за непрекъснатото подобряване и модернизиране на висшите учебни заведения и техните програми за обучение.

СЕМЕСТРИАЛЕН УЧЕБЕН ПЛАН НА КУРСА

EPL: КУРСОВЕ ПО КОМПЮТЪРНИ НАУКИ

MAS: КУРСОВЕ ПО СТАТИСТИКА

BUS: КУРСОВЕ ПО МЕНИДЖМЪНТ/БИЗНЕС

Година 1. Зимен семестър

Заглавие на курса	Кредити
ОБЩО КРЕДИТИ	28
EPL: Въведение в науката за данните и анализа	8

Този курс се разглежда как технологиите за анализ на данни могат да бъдат използвани за подобряване на вземането на решения. Целта е да се проучат основните принципи и техники на науката за данни и ще разгледаме реални примери и случаи, за да поставим техниките от науката на данните в контекст, за развиване на аналитично мислене с данни, и да се илюстрира, че правилното прилагане е наука, но също толкова е и изкуство. В допълнение, този курс ще работи практически с програмния език Python и свързаните с него библиотеки за анализ на данни.

MAS: Вероятности и статистика за Науката за данните	8
--	----------

Това е теоретичен курс, обхващащ основните теми на вероятностите и статистиката в контекста на науката за данните с присъщите си предизвикателства. Той ще започне с преглед на основната вероятност, като ще обхване теми, като случайни променливи, техните функции за разпределение, очаквани стойности, условия за определени събития и независимост. Студентите ще бъдат запознати с определени семейства на вероятностните разпределения и ще се научат как да правят приблизителна оценка или прогноза на определени величини, които представляват интерес,

въз основа на наблюдения. Ще бъдат изследвани редица свойства на оценките, включително достатъчност, безпристрастност и последователност, които дават възможност за оценка на тяхното качество. Студентите ще се научат също как да въвеждат различни видове хипотези, как да конструират проверки на техните хипотези, както и как да сравняват тестовете и как да изграждат доверителни интервали за оценките си.

MAS: Симулации и анализ на данни

8

Студентите ще бъдат запознати с езика за програмиране R, който е специално разработен за анализ на данни, и днес се използва широко в повечето организации, които извършват анализ на данни. Студентите ще се научат как да проучват наборите от данни в R, като използват основни инструменти за визуализация и обобщени статистически данни, как да извършват различни видове регресии и анализи, и как да правят статистически изводи за дисперсията на практика, например как да се проверяват някои хипотези относно данните или как да изчисляват доверителни интервали за величините, представляващи интерес. Студентите също ще се научат как да използват R, за да извършват симулации, изключително полезен инструмент, който може да изпълни широк спектър от аналитични задачи. Обхванатите симулационни техники ще включват метода Монте Карло, моделиране на значимост и отхвърляне. И накрая, студентите ще се научат как да правят оценка на точността на изчислената статистическа информация с помощта на методи за повторно моделиране. Курсът използва практически подход, като почти половината от работата се извършва в лаборатория.

Една избираема дисциплина (вж. списъка на избираемите дисциплини по-долу) 4

Бележка: Два избираеми курса (общо 8 кредита) могат да бъдат заменени от един курса с 8 кредита от специализациите (Компютърни науки, Статистика, Бизнес анализи).

Година 1. Летен семестър

Заглавие на курса	Кредити
ОБЩО КРЕДИТИ	28
MAS: Статистически методи и анализ	8

Студентите ще придобият знания за провеждане на статистически анализ на различни набори от данни, използвайки широк спектър от съвременни компютризирани методи. Ще се научат как да избират от кои инструменти имат нужда за анализ на различни видове набори от данни, как да прилагат тези инструменти във всеки отделен случай и как да използват диагностика за оценка на качеството на техните резултати. Ще научат за статистическите модели, тяхната сложност и относителните им ползи в зависимост от наличните данни. Някои от инструментите, които студентите ще научат, включват линейна проста и множествена регресия, метода на най-близките съседи, методите на свиване (хребет, ласо), методи за намаляване на димензиите (основни компоненти), логистична регресия, линеен дискриминантен анализ, методи, базирани на дърво на решенията и алгоритми за избор на модел с

критерий или чрез техники за ресемплиране. Фокусът на курса ще бъде по-малко на теория и повече за предоставяне на студентите на възможно най-много информация и запознаването им с възможно най-много методи. Курсът ще използва съществено езика за статистическо програмиране R и неговите библиотеки.

EPL: Управление на големи масиви от данни

8

Големите масиви от данни изискват съхранение, организация и обработка на данните в такива мащаби и с такава ефективност, които далеч надхвърлят възможностите на конвенционалните информационни технологии. В този курс ще изучаваме последните тенденции в управлението на големи масиви от данни: Ще научим за алгоритмите, техниките и инструментите, необходими за поддръжка на обработката на големи масиви от данни. В допълнение, в този курс ще бъдат разглеждани реални приложения, които изискват анализ на големи масиви от данни и как могат да бъдат използвани в платформи за големи масиви от данни. Курсът ще се състои от лекции, основани както на материали от учебници, така и на научни статии. Ще включва и задачи по програмиране, които ще предоставят на студентите практически опит в изграждането на приложения с интензивно използване на данни, чрез използване на съществуващи платформи за големи масиви от данни.

BUS: Приложения за бизнес анализи

8

Този курс предоставя знания и умения за прилагане на бизнес анализи за вземане на управленски решения в корпоративна среда. Темите включват описателни анализи (техники за категоризиране, характеризиране, консолидиране и класифициране на данни за превръщането им в полезна информация за целите на разбирането и анализа на бизнес резултатите), прогнозни анализи (техники за

откриване на скрити модели в големи количества данни за сегментиране и групиране на данни в кохерентни набори, за прогнозиране на поведение и тенденции), предписателни анализи (техники за установяване на най-добрите алтернативи за постигане на максимума или свеждане до минимум на бизнес целите). Студентите ще се научат как да използват ефективно данните, за да стимулират вземането на бързи, точни и печеливши решения, базирани на анализи. Рамката за използване на взаимосвързани данни, аналитични модели и инструменти за подпомагане на вземането на решения ще се прилага в рамките на наша специална основна структура за бизнес анализи и ще бъдат давани примери от различни функционални области на предприятието.

Една избираема дисциплина (вж. списъка на избираемите дисциплини по-долу) 4

Бележка: Два избираеми курса (общо 8 кредита) могат да бъдат заменени от един курс с 8 кредита от специализациите (Компютърни науки, Статистика, Бизнес анализи).

Година 1. Летен семестър

Заглавие на курса	Кредити
ОБЩО КРЕДИТИ	5
Проект „Капстоун“ в науката за данните (фаза I)	5

Целта на проекта „Капстоун“ е студентите да прилагат теоретичните знания, придобити по време на програмата по Наука за данните,

към проект, включващ актуални данни в реалистична обстановка. По време на проекта студентите участват в целия процес на решаване на реален проект по науката за данните: от събиране и обработка на действителни данни, до прилагане на подходящи и аналитични методи към проблема. Както самите проблеми от задачите по проекти, така и наборите от данни са от области от реалния свят, подобни на тези, с които студентите обикновено могат да се сблъскат в промишлеността, държавното управление, сектора на неправителствените организации (НПО) или в академичните изследвания. В рамките на този курс ще насърчаваме възможностите за намиране на работа и командироване в местната индустрия.

Година 2. Зимен семестър

Заглавие на курса	Кредити
ОБЩО КРЕДИТИ	29
Специализация по Компютърни науки/ Статистика/ Бизнес анализи	8
Специализация по Компютърни науки/ Статистика/ Бизнес анализи	8
Специализация по Компютърни науки/ Статистика/ Бизнес анализи	8
Проект „Капстоун“ в науката за данните (фаза II)	5

Целта на проекта „Капстоун“ е студентите да прилагат теоретичните знания, придобити по време на програмата по Наука за данните, към проект, включващ актуални данни в реалистична

обстановка. По време на проекта студентите участват в целия процес на решаване на реален проект по науката за данните: от събиране и обработка на действителни данни, до прилагане на подходящи и аналитични методи към проблема. Както самите проблеми от задачите по проекти, така и наборите от данни са от области от реалния свят, подобни на тези, с които студентите обикновено могат да се сблъскат в промишлеността, държавното управление, сектора на неправителствените организации (НПО) или в академичните изследвания. В рамките на този курс ще насърчаваме възможностите за намиране на работа и командироване в местната индустрия.

Специализация по компютърни науки

EPL: Управление на данните (8 кредита)

Този курс най-общо въвежда системите от бази данни, включително модела на релационните данни, езици за заявки, дизайн на база данни, индексирание и файлови структури, обработка на заявки и оптимизация, паралелни данни и възстановяване на данни, управление на трансакции и дизайн на бази данни. Студентите придобиват практически опит в работата със системи от бази данни и в изграждането на приложения за бази данни, достъпни от интернет.

EPL: Обработка на естествен език (8 кредита)

Как следва да се разбират и анализират човешките езици? Този курс ще разгледа съвременните изчислителни подходи, базирани на репрезентативно самообучение за разбиране, обработка и използване на човешки език. Те включват методи за дълбоко самообучение на база невронни мрежи и векторно-пространствени модели на смисъла на думите и взети заедно ще получите инструменти за изграждане на най-съвременни модели за трудни задачи за разбиране на език, като превод.

EPL: Извличане на информация (8 кредита)

Целта на този курс е да се проучат основните принципи на компютърните науки, които стоят зад Google и други търсачки. За тази цел ще предоставим въведение в Извличането на информация, което се описва като наука за търсене на информация в документи, търсене на самите документи, търсене на метаданни, които описват документи или търсене в бази данни, независимо дали са релационни самостоятелни или свързани с хипервръзки, като световната мрежа (уеб) например.

EPL: Дълбоко машинно самообучение (8 кредита)

В този курс учениците ще научат основите на дълбокото самообучение, ще разберат как да изградят невронни мрежи и ще научат как да ръководят успешни проекти за машинно самообучение. Студентите ще научат за конволюционните невронни мрежи, повтарящите се невронни мрежи (RNN), дълга краткосрочна памет (LSTM), алгоритъмът за оптимизация „Adam“, Dropout, BatchNorm, инициализация по Ксавие или Хе и др.

EPL: Уеб анализи и интелигентен анализ на текст (8 кредита)

Курсът по уеб анализи и интелигентен анализ на текст обхваща областите на уеб анализа, интелигентния анализ на текст и практическите приложения от тях. Част от курса по уеб анализа изследва показателите на уеб сайтовете, тяхното съдържание, поведението на потребителите и отчитането. Инструментът Google Analytics се използва за събиране на данни от уеб сайтове и извършване на анализа. Модулът за интелигентен анализ на текст обхваща анализа на текст, включително извличане на съдържание, съпоставяне на низове, клъстериране, класификация и системи за препоръки. Модулът по интелигентен анализ на текст разглежда как ботовете обработват и индексират съдържанието на уеб сайтове, как става търсенето и как се класират резултатите. Подробно ще бъдат проучени областите на приложение, интелигентния анализ на социалните мрежи и показателите при игрите.

EPL: Компютърни услуги „в облак“ (8 кредита)

Компютърните услуги “в облак” са широкомащабна разпределена изчислителна парадигма, която се превърна в движеща сила за информационните технологии през последните няколко години. Експоненциалният ръст на размера на данните от научните прибори/симулации и социалните медии предизвика по-широкото използване на компютърните услуги „в облак“. Този курс обхваща теми и технологии, свързани с компютърните услуги „в облак“ и техните практически реализации. Той се фокусира върху изучаването на нови въпроси, свързани с технологията за компютърни услуги „в облак“. Целите са: Разбиране на различни основни концепции, свързани с облачните компютърни технологии; разбиране на архитектурата и концепцията на различните модели „облаци“: IaaS, PaaS, SaaS; разбиране на инструментите и техниките за анализ на големи масиви от данни; разбиране на основния принцип на виртуализацията в облака, съхранението в облак, управлението на данни и визуализацията на данни; разбиране на различни платформи и инструменти за програмиране в облак. Да бъдат запознати с програмирането в облака чрез езика за програмиране на Google "Go"; Да имат подробни познания за четене и писане в облачно хранилище; да са запознати с разработването и внедряването на приложения, с използване на облачни платформи; да създават приложения чрез използване на облачни платформи като Google app Engine и Amazon Web Services (AWS); да се научат да разработват мащабируеми приложения, чрез използване на функциите на AWS; да научат основните концепции на моделите за програмиране на MapReduce за анализ на големи масиви данни в облака.

EPL: Визуализация на данни (8 кредита)

Визуализацията на данните е важен визуален метод за ефективна комуникация и анализ на големи масиви от данни. Чрез визуализациите на данните можем да правим изводи от данни, които понякога не са непосредствено очевидни и да взаимодействаме с данните по съвсем различен начин. Този курс ще предостави на студентите информативно въведение в методите, инструментите и процесите, свързани с визуализацията на големи масиви от данни. Темите ще включват: Въведение във визуализацията; визуализация на информацията;

научна визуализация; инструменти за визуализация; подходи при дизайна на визуализация; визуализация за комуникация. Ще използваме различни инструменти, да може студентите спокойно да изпробват различен софтуер и уверено да опитват нови софтуерни пакети, за да намерят тези, които най-добре отговарят на Вашите нужди. Това включва R, Tableau и D3.js. До края на семестъра се стремим студентите да се чувстват комфортно да проектират и разработват визуални истории с данни.

BUS: Информационни мрежи (8 кредита)

Този курс се фокусира върху това как са свързани социалните, икономическите, технологичните и естествените системи и как проучването на *мрежите* хвърля светлина върху тези връзки. Темите включват: как се моделира образуването на социални и икономически мрежи; разбиране и измерване на определени модели мрежи в реалния свят; установяване, количествено определяне и моделиране как мненията, актуалните за малко явления, политическите движения и болестите се разпространяват чрез взаимосвързани системи и да се измери тяхната стабилност и уязвимост. За да отговорим на тези въпроси ще съберем модели и техники от икономиката, социологията, математиката, физиката, статистиката и компютърните науки.

Специализация по Статистика

MAS: Бейсовска статистика (8 кредита)

Субективна вероятност, правило на Бейс, априорно и постериорно разпределение, неинформативни априори, точкова оценка и надеждни интервали, проверка на хипотези, въведение в Бейсовския анализ на решенията, въведение в емпиричния Бейсовски анализ, въведение във верижните методи на Марков и метода Монте Карло.

MAS: Правене на извадки от проучване (8 кредита)

Разработване на проучването, стандартно и други видове отклонения, проста случайна извадка, стратифицирана извадка, системна извадка, клъстерна

извадка, анализ на съотношения, регресии, определяне на оптималния размер на извадката, отклонение при извадката, съвременни техники за правене на извадки от проучвания.

MAS: Мултивариантен анализ (8 кредита)

Случайни вектори, мерки от центъра и вариация в многовариантни моменти. Многовариантно нормално разпределение. Тестове за нормалност. Оценка на средния вектор и анализа на дисперсията, независимост, многовариантност – ковариантна матрица. Разпределение на Wishart и Hotelling. Статистически изводи. Проверка обединяване – пресичане. Доверителни интервали. Многовариантен анализ на дисперсията и мултивариантен регресионен анализ. Метод на най-малките квадрати и разпределение на Wilks. Анализ на ковариатите. Основни компоненти, факторен анализ, дискриминантен анализ, клъстерен анализ.

MAS: Изчислителна статистика (8 кредита)

Линейна алгебра: Множествена регресия, разлагане на Cholesky, диагностика и колинеарност, основни компоненти и проблеми със собствените стойности. Нелинейни статистически методи: Оценка на максималната вероятност, метод на Newton-Raphson и свързаните с тях методи, многовариантни данни и метода на Newton-Raphson, техники за оптимизация (безусловни и под ограничения) Алгоритъм на очакване-максимизация . Цифрова интеграция и сближаване: Метод на Newton-Coates, сплайн интерполация, интеграция по метода Монте Карло, общи методи за апроксимация. Оценка на плътността на вероятността: Хистограма, линейно и нелинейно изглаждане, параметрична крива. Буутстрап.

MAS: Анализ на динамични редове (8 кредита)

Стохастични процеси, слаба и силна стационарност. Тенденции и сезонно поведение на динамичните редове. Функция за автоматично корелиране на извадката и функция за частично автоматично корелиране на извадката. Прогнози. Параметрични семейства на стохастични процеси. Модели ARMA, ARIMA и SARIMA. Свойства, приблизителни оценки и примери. ARCH и GARCH процеси, свойства и примери.

Специализация по Бизнес анализи

BUS: Управление на бизнес процесите с информационни системи и анализи (8 кредита)

Този курс предоставя на студентите основните инструменти за анализ и подобряване на бизнес процесите в организациите, с акцент върху сектора на услугите. Това се постига чрез обединяване на ключови идеи от областта на информационните системи, бизнес анализите и разработването и управлението на бизнес процеси. Курсът въвежда основните видове информационни системи, включително корпоративни системи (ERP, SCM, CRM), както и основните принципи за подкрепа на бизнес стратегията с информационните системи. Студентите ще се научат как да използват информационни системи в подкрепа на бизнес процесите на своята организация и как да използват бизнес анализи и техники за моделиране на бизнес процеси, за да вземат информирани ключови решения по време на реинженеринга на бизнес процесите. Студентите ще бъдат запознати с различни системи за бизнес анализи в области като маркетинг, търговия на дребно, управление на веригата за доставки, електронна търговия и др. и ще се научат как да измерват резултатите от бизнес процесите чрез подходящи показатели и рамки (напр. методологията на балансираната система от показатели).

BUS: Количествено и качествено вземане на решения (8 кредита)

Този курс изследва вземането на решения и формулирането на политики в организациите. Включва определяне на цели и процес на планиране, рационални модели на вземане на решения, ефективно съчетаване на качествени и количествени данни (напр. триангулация, комплементарност и т.н.) по отношение на набелязаната цел, оценка на алтернативи, прогнозиране на резултатите, анализ на разходите и ползите, дървета за вземане на решения, оценка на несигурността и риска, както и процедури за оценка на резултатите.

BUS: Финансови понятия (8 кредита)

Запознаване с понятията, методите и проблемите на счетоводството и финансовия анализ. Включва счетоводни принципи, въпроси, свързани с отчитането и оповестяването, анализ на финансови отчети, времева стойност на парите, прогнозиране и анализ на паричните потоци, капиталово бюджетиране и оценка на проекти, оценка на облигации и собствен капитал, цена на капитала и капиталова структура.

BUS: Уеб анализи в бизнеса (8 кредита)

Уеб анализи, интелигентен анализ на текст, извличане на информация от уеб и практически приложения. Част от курса по уеб анализа изследва показателите на уеб сайтовете, тяхното съдържание, поведението на потребителите и отчитането. Инструментът Google Analytics се използва за събиране на данни от уеб сайтове и извършване на анализа. Модулът за интелигентен анализ на текст обхваща анализа на текст, включително извличане на съдържание, съпоставяне на низове, клъстериране, класификация и системи за препоръки. Модулът по интелигентен анализ на текст разглежда как ботовете обработват и индексират съдържанието на уеб сайтове, как става търсенето и как се класират резултатите. Подробно ще бъдат проучени областите на приложение, интелигентния анализ на социалните мрежи и показателите при игрите.

BUS: Извличане на данни за бизнес анализи (8 кредита)

Предприятията, организациите и физическите лица създават, събират и използват огромно количество структурирани и неструктурирани данни с цел да превърнат информацията в знания, да подобрят качеството и ефективността на процеса на вземане на решения и да се позиционират по-добре на силно конкурентния пазар. Извличането на данни е процес на намиране, извличане, визуализиране и отчитане на полезна информация и идеи както от малки, така и от големи масиви от данни с помощта на сложни методи за анализ на данни. Той е част от бизнес анализа, който се отнася до процеса на използване на различни форми на аналитични техники за постигане на желаните бизнес резултати чрез изискване на бизнес целесъобразност, практична информация,

управление на производителността и управление на стойността. Студентите в този курс ще изучават основните принципи и техники на извличането на данни. Те ще се научат как да прилагат съвременни модели и софтуерни приложения за извличане на данни. И накрая, студентите ще се научат как да правят оценка на цялостния бизнес процес на една организация или проект с цел да се разбере (а) бизнес контекстът, където следва да бъдат установена и уловена скрита вътрешна и външна стойност, и (б) точно какво прави избраният метод за извличане на данни.

BUS: Информационни мрежи (8 кредита)

Този курс се фокусира върху това как са свързани социалните, икономическите, технологичните и естествените системи и как проучването на *мрежите* хвърля светлина върху тези връзки. Темите включват: как се моделира образуването на социални и икономически мрежи; разбиране и измерване на определени модели мрежи в реалния свят; установяване, количествено определяне и моделиране как мненията, актуалните за малко явления, политическите движения и болестите се разпространяват чрез взаимосвързани системи и да се измери тяхната стабилност и уязвимост. За да отговорим на тези въпроси ще съберем модели и техники от икономиката, социологията, математиката, физиката, статистиката и компютърните науки.

BUS: Управление на проекти с използване на аналитични инструменти (8 кредита)

Този курс разглежда процеса на управление на проекти с акцент върху техниките за бизнес анализ за преодоляване на клопките и пречките, които често се появяват по време на типичен проект. Предназначен е за бизнес лидери, отговорни за изпълнението на проекти, както и за начинаещи и средно напреднали ръководители на проекти. Включва теми за планиране, остойностяване и бюджетиране, методологии за набиране на персонал и организиране, за управление на проекти и използване на данни за вземане на информирани решения от страна на ръководителя на проекта през целия жизнен цикъл на проекта.

EPL: Визуализация на данни (8 кредита)

Визуализацията на данните е важен визуален метод за ефективна комуникация и анализ на големи масиви от данни. Чрез визуализациите на данните можем да правим изводи от данни, които понякога не са непосредствено очевидни и да взаимодействаме с данните по съвсем различен начин. Този курс ще предостави на студентите информативно въведение в методите, инструментите и процесите, свързани с визуализацията на големи масиви от данни. Темите ще включват: Въведение във визуализацията; визуализация на информацията; научна визуализация; инструменти за визуализация; подходи при дизайна на визуализация; визуализация за комуникация. Ще използваме различни инструменти, да може студентите спокойно да изпробват различен софтуер и уверено да опитват нови софтуерни пакети, за да намерят тези, които най-добре отговарят на Вашите нужди. Това включва R, Tableau и D3.js. До края на семестъра се стремим студентите да се чувстват комфортно да проектират и разработват визуални истории с данни.

Избираеми дисциплини

S4E: Предприемачество с анализ на данни (4 кредита)

Съществуващите бизнес модели се променят и се създават нови чрез анализ на данни и бизнес информация. Целта на този курс е да се въведат нови бизнес модели, произтичащи от анализа на данни, и да се предостави на студентите възможността да развият нови бизнес процеси, основани на данни в по-традиционните организации. За тази цел студентите ще се научат да балансират различни въпроси от областта на анализа, бизнеса, предприемачеството, правото, етиката и т.н. Курсът широко използва казуси и дава много примери за съществуващи предприятия и начина, по който техният бизнес модел се трансформира чрез използване на бизнес анализи, в областта на набирането на клиенти, препоръчваните системи (напр. Amazon/Netflix), здравеопазване (например 23andme), насочена/ персонализирана реклама (напр. Facebook), логистика (напр. Maersk), и т.н.

LAW: Етика на науката за данните (4 кредита)

Курсът е предназначен за изграждане на етичното въображение и умения на студентите за събиране, съхранение, споделяне и анализ на данни, получени от хора, включително данни, използвани в алгоритми. Курсът предоставя исторически контекст за разбиране на принципите на информираното съгласие, дискриминацията и поверителността на данните. Използвайки казуси, студентите ще проучат текущите приложения на количествените разсъждения в организациите, алгоритмичната прозрачност и непреднамерената дискриминацията чрез данни, които съдържат предразсъдъци, коренящи се в раса, пол, класа и други характеристики.

6. Кратки програми

Индивидуални модули (наречени *кратки програми*) могат да се вземат отделно и да се получават сертификати. Определени са следните програми:

Анализ на данни с използване на езика R (8 кредита)

Участниците ще могат: Да разберат основните понятия в областта на науката за данните и техните приложения в реалния свят; да работят ефективно с езика R; да се запознаят се с инструментите за управление на данните.

MAS: Вероятности и статистика за Науката за данните

Това е теоретичен курс, обхващащ основните теми на вероятностите и статистиката в контекста на науката за данните с присъщите си предизвикателства. Той ще започне с преглед на основната вероятност, като ще обхване теми, като случайни променливи, техните функции за разпределение, очаквани стойности, условия за определени събития и независимост. Студентите ще бъдат запознати с определени семейства на вероятностните разпределения и ще се научат как да правят приблизителна оценка или прогноза на определени величини, които

представляват интерес, въз основа на наблюдения. Ще бъдат изследвани редица свойства на оценките, включително достатъчност, безпристрастност и последователност, които дават възможност за оценка на тяхното качество. Студентите ще се научат също как да въвеждат различни видове хипотези, как да конструират проверки на техните хипотези, както и как да сравняват тестовете и как да изграждат доверителни интервали за оценките си.

EPL: Управление на големи масиви от данни (8 кредита)

Големите масиви от данни изискват съхранение, организация и обработка на данните в такива мащаби и с такава ефективност, които далеч надхвърлят възможностите на конвенционалните информационни технологии. В този курс ще изучаваме последните тенденции в управлението на големи масиви от данни: Ще научим за алгоритмите, техниките и инструментите, необходими за поддръжка на обработката на големи масиви от данни. В допълнение, в този курс ще бъдат разглеждани реални приложения, които изискват анализ на големи масиви от данни и как могат да бъдат използвани в платформи за големи масиви от данни. Курсът ще се състои от лекции, основани както на материали от учебници, така и на научни статии. Ще включва и задачи по програмиране, които ще предоставят на студентите практически опит в изграждането на приложения с интензивно използване на данни, чрез използване на съществуващи платформи за големи масиви от данни.

Анализ на данни за бизнеса

Участниците ще могат: Да подхождат към бизнес проблемите чрез анализ на данни; да мислят внимателно и систематично за това дали и как данните и бизнес анализите могат да подобрят резултатите на бизнеса; да разработват идеи за бизнес анализи; да анализират данни, използвайки софтуер за бизнес анализ; и да генерират бизнес информация.

Курсове:

EPL: Визуализация на данни (8 кредита)

Визуализацията на данните е важен визуален метод за ефективна комуникация и анализ на големи масиви от данни. Чрез визуализациите на данните можем да правим изводи от данни, които понякога не са непосредствено очевидни и да взаимодействаме с данните по съвсем различен начин. Този курс ще предостави на студентите информативно въведение в методите, инструментите и процесите, свързани с визуализацията на големи масиви от данни. Темите ще включват: Въведение във визуализацията; визуализация на информацията; научна визуализация; инструменти за визуализация; подходи при дизайна на визуализация; визуализация за комуникация. Ще използваме различни инструменти, да може студентите спокойно да изпробват различен софтуер и уверено да опитват нови софтуерни пакети, за да намерят тези, които най-добре отговарят на Вашите нужди. Това включва R, Tableau и D3.js. До края на семестъра се стремим студентите да се чувстват комфортно да проектират и разработват визуални истории с данни.

BUS: Извличане на данни за бизнес анализи (8 кредита)

Предприятията, организациите и физическите лица създават, събират и използват огромно количество структурирани и неструктурирани данни с цел да превърнат информацията в знания, да подобрят качеството и ефективността на процеса на вземане на решения и да се позиционират по-добре на силно конкурентния пазар. Извличането на данни е процес на намиране, извличане, визуализиране и отчитане на полезна информация и идеи както от малки, така и от големи масиви от данни с помощта на сложни методи за анализ на данни. Той е част от бизнес анализа, който се отнася до процеса на използване на различни форми на аналитични техники за постигане на желаните бизнес резултати чрез изискване на бизнес целесъобразност, практична информация, управление на производителността и управление на стойността. Студентите в този курс ще изучават основните принципи и техники на извличането на данни. Те ще се научат как да прилагат съвременни модели и софтуерни приложения за извличане на данни. И накрая, студентите ще се научат как да правят оценка на цялостния бизнес процес на една организация или проект с цел да се разбере (а) бизнес контекстът, където следва да

бъдат установена и уловена скрита вътрешна и външна стойност, и (б) точно какво прави избраният метод за извличане на данни.

Анализ на основни данни

Участниците ще могат: Да подготвят и анализират данни; да научат различните етапи на проучвателния анализ на данни и как да създадат интерактивни табла за управление; да се запознаят се с основните принципи на Python за анализатори на данни.

EPL: Въведение в науката за данните и анализа (8 кредита)

Този курс се разглежда как технологиите за анализ на данни могат да бъдат използвани за подобряване на вземането на решения. Целта е да се проучат основните принципи и техники на науката за данни и ще разгледаме реални примери и случаи, за да поставим техниките от науката на данните в контекст, за развиване на аналитично мислене с данни, и да се илюстрира, че правилното прилагане е наука, но също толкова е и изкуство. В допълнение, този курс ще работи практически с програмния език Python и свързаните с него библиотеки за анализ на данни.

EPL: Визуализация на данни (8 кредита)

Визуализацията на данните е важен визуален метод за ефективна комуникация и анализ на големи масиви от данни. Чрез визуализациите на данните можем да правим изводи от данни, които понякога не са непосредствено очевидни и да взаимодействаме с данните по съвсем различен начин. Този курс ще предостави на студентите информативно въведение в методите, инструментите и процесите, свързани с визуализацията на големи масиви от данни. Темите ще включват: Въведение във визуализацията; визуализация на информацията; научна визуализация; инструменти за визуализация; подходи при дизайна на визуализация; визуализация за комуникация. Ще използваме различни инструменти, да може студентите спокойно да изпробват различен софтуер и уверено да опитват нови софтуерни пакети, за да намерят тези,

които най-добре отговарят на Вашите нужди. Това включва R, Tableau и D3.js. До края на семестъра се стремим студентите да се чувстват комфортно да проектират и разработват визуални истории с данни.

7. Заключение

Този отчетен резултат може да послужи като начало за разработване на нови програми и за преминаване на съществуващи програми към областта на Науката за данните. В обобщение, основните моменти от това проучване включват следното:

- Науката за данните е бързо развиваща се дисциплина.
- Занятията от съществуващите програми по Наука за данните се провеждат в помещенията на катедрите по компютърни науки, информатика, математика и статистика, както и управление. Много от най-успешните студенти, особено бакалаврите, представляват широки коалиции от университета, често финансирани от интердисциплинарни институти, а не от определена катедра.
- Редизайнът на учебната програма, чрез включване на елементи от математиката и мисленето от информатиката и статистиката на всички нива ще осигури богата и ефективна поредица от курсове за подготовка на завършилите за кариера в областта на Науката за данните.