

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної
роботи
Пантелеймонов А.В.

“ _____ ” _____ 2018 р.

Робоча програма навчальної дисципліни

“Введення в Data Science”

(шифр і назва навчальної дисципліни)

рівень вищої освіти _____ перший (бакалаврський) _____

галузь знань _____ 05 «Соціальні та поведінкові науки» _____

спеціальність _____ 051 «Економіка» _____

освітня програма _____ «Бізнес-аналітика та міжнародна статистика» _____

спеціалізація _____ _____

вид дисципліни _____ обов'язкова _____

факультет _____ економічний _____

2018 / 2019 навчальний рік

Програму рекомендовано до затвердження Вченою радою факультету

“22” червня 2018 року, протокол № 7

РОЗРОБНИКИ ПРОГРАМИ:

Чала Т. Г., доцент кафедри статистики, обліку та аудиту, канд. екон. наук, доцент
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Програму схвалено на засіданні кафедри статистики, обліку та аудиту

Протокол від “21” червня 2018 року № 9

Завідувач кафедри статистики, обліку та аудиту

_____ (Соболев В.М.)
(підпис) (прізвище та ініціали)

Програму погоджено методичною комісією економічного факультету

назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна

Протокол від “21” червня 2018 року № 9

Голова методичної комісії економічного факультету

_____ (Євтушенко В.А.)
(підпис) (прізвище та ініціали)

ВСТУП

Програма навчальної дисципліни “**Введення в Data Science**” складена відповідно до освітньо-професійної (освітньо-наукової) програми підготовки першого рівня вищої освіти - **бакалавра**

(назва рівня вищої освіти, освітньо-кваліфікаційного рівня)

спеціальності (напряму) 051 «Економіка»

спеціалізації

1. Опис навчальної дисципліни

1.1. *Метою* викладання навчальної дисципліни є надання теоретичних знань і практичних навиків збирання, узагальнення, оброблення, аналізу, зберігання та архівування різною за природою інформації.

1.2. Основними *завданнями* вивчення дисципліни є:

- організація процесу дослідження даних на основі наукових принципів організації статистичних спостережень;
- використання сучасних інформаційних технологій збирання, систематизації, оброблення та аналізу статистичної інформації.

1.3. Кількість кредитів – 5

1.4. Загальна кількість годин – 150

1.5. Характеристика навчальної дисципліни

Нормативна	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	–
Семестр	
2-й	–
Лекції	
32 год.	–
Практичні, семінарські заняття	
32 год.	–
Лабораторні заняття	
–	–

Самостійна робота	
86 год.	–
Індивідуальні завдання	
14 год.	

1.6. Заплановані результати навчання.

Згідно з вимогами освітньо-наукової програми студенти повинні досягти таких **результатів навчання**:

знання:

- усвідомлювати сутність методики організації збору даних;
- усвідомлювати сутність великих даних та ідентифікувати їх джерела;
- розуміти принципи використання великих даних в офіційній статистиці та підприємницькій діяльності;
- визначати сфери застосування великих даних;
- розуміти порядок систематизації, узагальнення та групування даних;
- демонструвати розуміння методів візуалізації даних;

уміння:

- організувати процес дослідження даних;
- визначати основні напрями аналізу веб-даних.
- усвідомлювати основні принципи культури інновацій в галузі аналітики
- використовувати сучасні інформаційні технології збирання, систематизації, оброблення та аналізу статистичної інформації;
- застосовувати графічно-візуальні методи узагальнення та аналізу даних;

2. Тематичний план навчальної дисципліни

Розділ 1. Теоретичні засади Data Science

Тема 1. Ознайомлення з Data Science (наукою про дані)

Визначення “Data Science”. Історія науки про дані. Вивчення великих даних та ідентифікація їх джерела. Взаємозв’язок науки про дані з іншими науками. Предмет дослідження науки про дані. Методологія науки про дані як сукупність загальних правил (принципів) і спеціальних прийомів і методів отримання знань на основі даних.

Тема 2. Історія формування науки про дані

Лекція Ч.-Ф. Джеффа Ву «Статистика = Наука про дані?». Організація видання «Data Science Journal». Організація видання «International Journal on Data Science and Analytics». Заснування Європейської асоціації з питань науки про дані (European Association for Data Science (EuADS)).

Тема 3. Сутність даних та інформації

Поняття про статистичні дані та інформацію. Види даних та аналітичні функції статистичних групувань. Структурні, типологічні, аналітичні групування. Ряд розподілу як основа групувань. Види рядів розподілу. Принципи формування груп. Інтервали групувань. Способи перегрупування даних (вторинні групування).

Статистична таблиця, її підмет і присудок. Характеристика простих, групових та комбінаційних таблиць, що вибираються в залежності від цілі їх побудови. Правила складання таблиць.

Розділ 1. Основні напрямки досліджень в Data science

Тема 4. Бази даних та інтеграція інформації

Роль міжурядових організацій у формуванні статистичних баз даних. Основні статистичні бази даних міжурядових організацій. Міжнародна статистична база даних інтелектуальної власності. База даних ООН з міжнародної торгівлі товарами та послугами (UN Comtrade Database).

Тема 5. Виявлення знань в соціальних та інформаційних мережах

Аналіз тексту та інших неструктурованих джерел даних. Різноманітність аналізованих даних. Різноманітність технологічних платформ обробки інформації

Тема 6. Математичне мислення для data scientist

Мозок створений для складних розрахунків. Налагодження «розумового насоса». Використання сфокусованого мислення. Ілюзія компетентності. Удосконалення пам'яті.

Тема 7. Візуалізація інформації

Поняття про візуалізацію інформації. Аналітичне значення статистичних графіків. Значення графіків в узагальнені статистичної інформації. Основні елементи статистичного графіку: графічний образ, поле графіку, масштабні орієнтири, експлікація графіку.

Класифікація статистичних графіків за способом побудови, за формою графічних образів, за характером завдань, що вирішуються. Розподіл графіків за способом побудови: діаграми та статистичні карти. Діаграми порівняння, динаміки та структурні діаграми. Основні види графіків, що застосовуються в економічно-статистичному аналізі: лінійні, радіальні, стовпчикові, секторні, смугові, квадратні, кругові, фігурні. Фонові та крапкові картограми, картодіаграми.

Тема 8. Типи вчених в області даних

Навички вчених в області даних. Поєднання навичок і самооцінки. Різноманітність вчених в області даних. Дата бізнесмен (Data Businessperson). Дата творець (Data Creative). Дата розробник (Data Developer). Дата дослідник (Data Researcher). «Т-образні навички» Data Scientists.

3. Структура навчальної дисципліни

Назви Розділів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
		л	п	лаб	інд	с.р.
1	2	3	4	5	6	7
Розділ 1. Теоретичні засади Data Science						
Тема 1. Ознайомлення з Data Science (наукою про дані)	22	6	6	-	-	10
Тема 2. Історія формування науки про дані	18	4	4	-	-	10
Тема 3. Сутність даних та інформації	18	4	4	-	-	10
Разом за розділом 1	58	14	14		-	30
Розділ 2. Основні напрямки досліджень в Data science						
Тема 4. Бази даних та інтеграція інформації	22	4	4	-	-	14
Тема 5. Виявлення знань в соціальних та інформаційних мережах	14	4	4	-	-	6
Тема 6. Математичне мислення для data scientist	14	4	4	-	-	6
Тема 7. Візуалізація інформації	10	2	2	-	-	6
Тема 8. Типи вчених в області даних	18	4	4	-	-	10
Разом за розділом 2	78	18	18	-	-	42
Усього годин	136	32	32	-	-	72
ІНДЗ – розрахункове завдання	14		-	-	-	14
Усього годин	150	32	32	-	-	86

4. Теми практичних занять

№ з/п	Зміст практичного заняття	Кількість годин
1	Ознайомлення з Data Science (наукою про дані)	6
2	Історія формування науки про дані	4
3	Сутність даних та інформації	4
4	Бази даних та інтеграція інформації	4
5	Виявлення знань в соціальних та інформаційних мережах	4
6	Математичне мислення для data scientist	4
7	Візуалізація інформації	2
8	Типи вчених в області даних	4

5. Завдання для самостійної роботи

Самостійна робота студента включає: опрацювання навчального матеріалу, виконання індивідуальних завдань і науково-дослідної роботи.

Зміст самостійної роботи студента над кожною темою дисципліни визначається кількістю годин на підготовку до:

- 2 година на підготовку до кожної аудиторної лекції;
- 2-3 години на підготовку до кожного практичного заняття;
- 20 годин на підготовку до заліку (по 2-3 години на кожен тему).

№ з/п	Види, зміст самостійної роботи (назва теми)	Кількість годин
1	Тема 1. Ознайомлення з Data Science (наукою про дані)	10
2	Тема 2. Історія формування науки про дані	10
3	Тема 3. Сутність даних та інформації	10
4	Тема 4. Бази даних та інтеграція інформації	14
5	Тема 5. Виявлення знань в соціальних та інформаційних мережах	6
6	Тема 6. Математичне мислення для data scientist	6
7	Тема 7. Візуалізація інформації	6
8	Тема 8. Типи вчених в області даних	10
	Всього	72
	ИНДЗ – розрахункове завдання	14
	Разом	86

Методи навчання

Основними відмінностями активних та інтерактивних методів навчання від традиційних вважаються не лише методика й техніка викладання, але й висока ефективність навчального процесу, яка проявляється через:

- високу мотивацію студентів;
- закріплення теоретичних знань на практиці;
- підвищення самосвідомості студентів;
- вироблення здатності ухвалювати самостійні рішення;
- вироблення здатності приймати колективні рішення;
- вироблення здатності до соціальної інтеграції;
- придбання навичок вирішення конфліктів;
- розвиток здатності до компромісів.

При викладанні дисципліни “Введення в Data Science” для активізації навчального процесу передбачено використання таких сучасних навчальних технологій як проблемні лекції, міні-лекції, робота в малих групах та ін.

Проблемні лекції спрямовані на розвиток логічного мислення студентів. Коло питань теми лекції обмежується кількома ключовими моментами, й увага студентів має бути сконцентрована на матеріалі, що не знайшов відображення в підручниках. При проведенні проблемної лекції має бути розданий друкований матеріал і виділені головні висновки з питань, що розглядаються.

При викладанні лекційного матеріалу студентам пропонуються питання для самостійного розмірковування. При цьому лектор ставить питання, які спонукають студента шукати вирішення проблемної ситуації. Така ситуація змушує студентів сконцентруватися, й почати активно мислити в пошуках правильної відповіді.

На початку проведення проблемної лекції необхідно чітко сформулювати проблему, яку необхідно вирішити студентам. При викладанні лекційного матеріалу слід уникати прямої відповіді на поставлені питання, а висвітлювати матеріал, таким чином, щоб отриману інформацію студент міг використати при вирішенні проблеми.

Міні-лекції передбачають викладення навчального матеріалу за короткий проміжок часу й характеризуються значною ємністю, складністю логічних побудов, образів, доведень та узагальнень.

На початку проведення міні-лекції за обраними темами лектор акцентує увагу студентів на необхідності представити викладений лекційний матеріал у так званому структурно-логічному вигляді. На розгляд виносять питання, які зафіксовані в плані лекцій, але викладають їх стисло. Лекційне заняття, проведене в такий спосіб, пробуджує в студента активність та увагу до сприйняття матеріалу, а також спрямовує його на використання системного підходу при відтворенні інформації, яку він отримав під час лекції.

Проблемні лекції та міні-лекції доцільно поєднувати з такою формою активізації навчального процесу, як робота в малих групах.

Робота в малих групах дає змогу структурувати лекційні або практичні заняття за формою та змістом, створює можливості для участі кожного студента в роботі за темою заняття, забезпечує формування досвіду соціального спілкування.

Після висвітлення проблеми (проблемна лекція) або стислого викладення матеріалу (міні-лекція) студентам пропонується об'єднатися в групи по 5-7 осіб та представити наприкінці заняття своє бачення та сприйняття матеріалу, тобто виступити перед аудиторією з результатами роботи групи.

Дискусії передбачають обмін думками та поглядами учасників щодо певної теми, а також розвивають мислення, допомагають формувати погляди й переконання, виробляють уміння формулювати думки й висловлювати їх, учать оцінювати пропозиції інших людей, критично підходити до власних поглядів.

Кейс-метод – метод аналізу конкретних ситуацій, який дає змогу наблизити процес навчання до реальної практичної діяльності спеціалістів, і передбачає розгляд виробничих, управлінських та інших ситуацій, інцидентів у процесі вивчення навчального матеріалу.

7. Методи контролю

Оцінювання знань, умінь та навичок студентів включає ті види занять, які згідно з програмою навчальної дисципліни “Введення в Data Science” передбачають лекційні, практичні заняття, самостійну роботу та виконання індивідуального розрахункового завдання.

Перевірка та оцінювання знань студентів проводиться в наступних формах:

- оцінювання роботи і знань студентів під час практичних занять;
- складання проміжного контролю знань за розділами (тестування);
- оцінювання виконання та захист індивідуального розрахункового завдання;
- складання заліку.

Поточне оцінювання знань студентів здійснюється під час проведення практичних, і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни;
- відвідування занять;
- виконання індивідуального науково-дослідного завдання;
- складання проміжного контролю за розділами.

Контроль систематичного виконання самостійної роботи та активності на практичних заняттях проводиться за такими критеріями:

- розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- ступінь засвоєння фактичного матеріалу навчальної дисципліни;
- ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- уміння поєднувати теорію із практикою при розгляді практичних ситуацій, розв'язанні задач, проведенні розрахунків при виконанні індивідуальних завдань, та завдань, винесених на розгляд в аудиторії;
- оволодіння методами економіко-статистичної обробки даних із використанням комп'ютерних технологій;
- логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, уміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінювання знань студента під час виконання завдань для самостійної роботи проводиться за 4- бальною шкалою.

Оцінка “відмінно” ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх зазначених критеріїв. Відсутність тієї чи іншої складової знижує оцінку.

При оцінюванні практичних занять увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

Проміжний контроль рівня знань передбачає виявлення опанування студентом лекційного матеріалу та вміння застосування його для вирішення практичної ситуації й проводиться у вигляді тестування. При цьому тестове завдання може містити як запитання, що стосуються суто теоретичного матеріалу, так і запитання, спрямовані на вирішення невеличкого практичного завдання.

8. Схема нарахування балів

Структура засобів контролю та розподіл балів із дисципліни «Введення в Data Science» наведена в таблиці.

Структура засобів контролю з навчальної дисципліни

Назви Розділів і тем		Види та засоби контролю	Розподіл балів
Розділ 1. Теоретичні засади Data Science	Тема 1. Ознайомлення з Data Science (наукою про дані)	<i>Тестування</i>	6
	Тема 2. Історія формування науки про дані	<i>Тестування</i>	4
	Тема 3. Сутність даних та інформації	<i>Тестування</i>	4
Розділ 2. Основні напрямки досліджень в Data science	Тема 4. Бази даних та інтеграція інформації	<i>Тестування</i>	4
	Тема 5. Виявлення знань в соціальних та інформаційних мережах	<i>Тестування</i>	4
	Тема 6. Математичне мислення для data scientist	<i>Тестування</i>	6
	Тема 7. Візуалізація інформації	<i>Тестування</i>	6
	Тема 8. Типи вчених в області даних	<i>Тестування</i>	6
	Індивідуальне науково-дослідне завдання	Написання та захист ІНДЗ	20
	Підсумковий контроль знань	залік	40
	Разом за семестр		100

Проміжний тестовий контроль проводиться два рази на семестр (в рамках вивчення Розділу). Загальна тривалість тестів за розділами – по 1,5 години. Поточне тестування з кожного Розділу складається з 20 тестів. Одна правильна відповідь на кожен із тестів дорівнює 1 бал. Тестове завдання містить запитання одиничного і множинного вибору різного рівня складності. Для оцінювання рівня відповідей студентів на тестові завдання використовуються наступні критерії оцінювання.

Шкала перерахунку оцінок результатів поточного тестування з Розділів

Оцінка за національною шкалою	Кількість набраних балів
1	2
ВІДМІННО	18 – 20
ДОБРЕ	12 – 17
ЗАДОВІЛЬНО	6 – 11
НЕЗАДОВІЛЬНО	0 – 5

Тести можуть бути застосовані як із метою контролю, так і для закріплення теоретичних знань і практичних навичок.

Тести для проміжного контролю обираються із загального переліку тестів за відповідними темами.

Підсумковий контроль за курсом – у формі заліку.

До складання заліку допускають студентів, що мають задовільну кількість балів із складених тестів з основних навчальних елементів розділів курсу, написання та захисту індивідуального науково-дослідного завдання та інших завдань передбачених програмою дисципліни.

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90–100	відмінно	зараховано
70–89	добре	
50–69	задовільно	
1–49	незадовільно	не зараховано

9. Рекомендована література

Основна література

1. Чала Т. Г. Статистика: навч.-метод. посібн. / О. С. Корепанов, Ю.О. Лазебник, Т.В. Пономарьова, В.М. Соболев, Т. Г. Чала. – Х. : Харківський національний університет ім. В. Н. Каразіна, 2014. – 151 с.
2. Roger D. Peng & Elizabeth Matsui The Art of Data Science. A Guide for Anyone Who Works with Data, Skybrude Consulting, LLC, 2015, 155 p.
3. Mike Barlow. Real-Time Big Data Analytics: Emerging Architecture. O'Reilly Media, 2013.
4. Jeffrey Stanton. Introduction to Data Science, Syracuse University, By Robert De Graaf, 2013.

Допоміжна література

1. Єріна А. М. Статистика : підручник / А. М. Єріна. – К. : Знання, 2009. – 484 с.
2. Ковалевский Г. В. Статистика : учебник / Г. В. Ковалевский. – Х. : ХНАГХ, 2012. – 445 с.
3. Кулинич О. І. Теорія статистики : підручник / О. І. Кулинич, Р. О. Кулинич. – 5-те вид. перероб. і доп. – К. : Знання, 2010. – 239 с.
5. Чекотовський Е. В. Графічний метод у статистиці на основі програми EXCEL : навч. посіб. / Чекотовський Е. В. – К. : Знання, 2000. – 518 с.

10. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Нормативно-правова база [Електронний ресурс] / Кабінет Міністрів України; Урядовий портал. – Офіц. веб-сайт. – Режим доступу : <http://www.kmu.gov.ua/control/npd/list>.
2. Нормативно-правові акти [Електронний ресурс] // Урядовий портал. Єдиний веб-портал виконавчої влади України. – Режим доступу : http://www.me.gov.ua/control/uk/publish/category/main?cat_id=32854.
3. Статистична інформація [Електронний ресурс] / Державна служба статистики України ; за ред. О. Г. Осауленка // Офіційний сайт державної служби статистики України. – Режим доступу : <http://www.ukrstat.gov.ua>.
4. Economicus.ru – економічний портал. Галерея економістів [Електронний ресурс] / А. Абрамков, Е. Лукьяненко, О. Куликова и др. ; сост. и ред. А. Скоробогатов, А. Дмитриев, М. Сторчевой. – Режим доступу : http://gallery.economicus.ru/cgi-ise/gallery/g_homen.pl.
5. Загальноакадемічний портал наукової періодики. Національна академія наук України [Електронний ресурс] / Л. Костенко, Є. Копанева // Національна бібліотека України імені В.І. Вернадського. – Режим доступу : <http://www.nbuv.gov.ua/portal/natural/NTI/search.html>.