МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

"ЗАТВЕРДЖУЮ"
Заступник керівника
(проректор з науково-педагогічної роботи)
_____ Афанасьєв М.В.

МЕТОДИ ТА МОДЕЛІ DATA SCIENCE

робоча програма навчальної дисципліни

Галузь знань	**05 Соціальні та поведінкові науки**
Спеціальність	**051 Економіка**
Освітній рівень	**другий (магістерський)**
Освітня програма	**Економічна кібернетика**

Вид дисципліни	**базова**
Мова викладання, навчання та оцінювання	**українська**

Завідувач кафедри економічної кібернетики Гур'янова Л.С.

Харків
ХНЕУ ім. С. Кузнеця
2019

ЗАТВЕРДЖЕНО
на засіданні кафедри економічної кібернетики
Протокол № 1 від 27.08.2019 р.

Розробник:
Івахненко О.В., к.е.н., доц. кафедри економічної кібернетики

Лист оновлення та перезатвердження робочої програми навчальної дисципліни

Навчальний рік	Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри

1. Вступ

Анотація навчальної дисципліни: Програма вивчення навчальної дисципліни «Математичні методи та моделі data science» складена відповідно до освітньо-професійної програми підготовки магістрів зі спеціальності 051 «Економіка».

Предметом вивчення навчальної дисципліни є методи та моделі Data science, як науки о даних, зокрема засоби побудови дескриптивної, предиктивної аналітики, візуалізації даних та машинного навчання.

Програма навчальної дисципліни складається з таких змістових модулів:

1. Основні ідеї Data science та вступ до аналітики даних
2. Алгоритми машиного навчання

Мета навчальної дисципліни: Метою викладання навчальної дисципліни «Математичні методи та моделі data science» є засвоєння основних ідей сучасного аналізу даних, вивчення сучасних методів та моделей аналізу даних і машинного навчання.

Основними завданнями вивчення дисципліни «Математичні методи та моделі data science» є

засвоєння основних ідей Data science та машинного навчання;

засвоєння основних принципів та підходів до обробки кількісних та якісних даних;

засвоєння прикладних аспектів роботи з великими даними, що описують стан та розвиток складних систем;

засвоєння основних методичних підходів та сучасного теоретичного та практичного базису з побудови моделей та алгоритмів обробки, візуалізації та аналізу даних, побудови рекомендаційних систем та машинного навчання.

Курс	**1м**	
Семестр	**1**	
Кількість кредитів ECTS	**5**	
Аудиторні навчальні заняття	**лекції**	**20**
	семінарські, практичні	**-**
	лабораторні	**20**
Самостійна робота		**110**
Форма підсумкового контролю	**іспит**	

Структурно-логічна схема вивчення навчальної дисципліни:

Попередні дисципліни	Наступні дисципліни
Економетрія Економетрія-2 Статистика Економічна кібернетика	Комплексний тренінг Переддипломна практика

2. Компетентності та результати навчання за дисципліною:

Компетентності	Результати навчання
СК4.Здатність пояснювати економічні та соціальні процеси і явища на основі теоретичних та прикладних моделей, аналізувати і змістовно інтерпретувати отримані результати.	РН 17. Виконувати міждисциплінарний аналіз соціально-економічних явищ і проблем в однієї або декількох професійних сферах з врахуванням ризиків та можливих соціально-економічних наслідків
СК6. Здатність застосовувати економіко-математичні методи та моделі для вирішення економічних задач.	Н 8. Застосовувати відповідні економіко-математичні методи та моделі для вирішення економічних задач.
СК7.Здатність застосовувати комп'ютерні технології та програмне забезпечення з обробки даних для вирішення економічних завдань, аналізу інформації та підготовки аналітичних звітів.	РН 19. Використовувати інформаційні та комунікаційні технології для вирішення соціально-економічних завдань, підготовки та представлення аналітичних звітів.

3. Програма навчальної дисципліни

Змістовий модуль 1. Основні ідеї Data science та вступ до аналітики даних

Тема 1. Вступ до Data Science

Переваги прийняття рішень заснованих на аналітиці даних. Категорії Data Science, Machine Learning, Data Mining, Business Intelligent, Artifical Intelligence та їх перетин. Основні завдання Data Science.

Еволюція аналізу даних. Описова (descriptive), діагностична (diagnostic), прогнозна (predictive), наказова (prescriptive) аналітики даних.

Основи статистичного (машинного навчання). Сутність навчальної та контрольної вибірок. Особливості оцінки якості моделей машинного навчання.

Лабораторна робота №1. Вступ до мови R.

Тема 2. Інструментальні засоби аналізу даних

Інструменти та методи Data Science. Прикладний інструментарій. Мови програмування, найбільш пристосовані для вирішення завдань аналітики даних. Можливості аналітики даних на R/Python.

Лабораторна робота №2. Основи опрацювання великих масивів даних у R

Тема 3. Передобробка та візуалізація даних

Аналіз структури та типів даних, обробка вхідних даних та їх чистка, використання неповних даних. Підготовка даних. Сирі дані та їх обробка. Побудова дескриптивної аналітики.

Основи візуалізації даних. Правила побудови вдалої візуалізації. Види візуалізацій та їх мікси. Особливості сприйняття інформації людиною. Використання кольорів та особливостей розміщення інформації для створення аналітичного звіту. Сучасний інструментарій для візуалізації даних.

Лабораторна робота №3. Передобробка та візуалізація даних

Тема 4. Перевірка гіпотез та закони розподілу

Формулювання гіпотези та її формальний запис. Квантілі законів розподілу та використання основних функцій для роботи з розподілами.

Лабораторна робота №4. Перевірка гіпотез та закони розподілу

Змістовий модуль 2. Алгоритми машиного навчання

Тема 5. Відбір та регуляризація лінійних моделей

Проблеми побудови регресійних моделей на великих обсягах даних. Методи пошагового відбору. Методи стиснення (ласо, гребенева регресія). Методи зниження розмірності. Інтерпретація даних в задачах великої розмірності.

Лабораторна робота №4. Перевірка гіпотез та закони розподілу

Тема 6. Кластерізація

Задача навчання без вчителя. Реалізація методів кластерного аналізу. Найпоширеніші алгоритми кластерізації у R (K-Means та GMM). Основи кластерізації та візуалізації результатів.

Лабораторна робота №6. Кластерізація

Тема 7. Методи та алгоритми класифікації

Навчання з вчителем. Особливості різних видів алгоритмів класифікації. Байесовський класифікатор. KNN. Оцінка якості класифікації. Види помилок. Проблема вибору моделі.

Лабораторна робота №7. Методи та алгоритми класифікації

Тема 8. Дерева класифікації та ансамблі дерев

Вирішення задачі навчання з вчителем методами дерев класифікації. Побудова дерев класифікації різними методами. Скупий алгоритм.

Побудова ансамблів дерев. Беггінг, бустінг та випадкові ліси (random forests). Вибір остаточного виду моделі.

Лабораторна робота №9 Дерева класифікації та ансамблі дерев

Тема 9. Машини опорних векторів

Побудова класифікаторів на опорних векторах. Машини опорних векторів (SVM). SVM з декількома класами. Взаємозв'язок з логістичною регресією.

Лабораторна робота №9 Машини опорних векторів

4. Порядок оцінювання результатів навчання

Система оцінювання сформованих компетентностей у студентів враховує види занять, які згідно з програмою навчальної дисципліни передбачають лекційні, лабораторні заняття, а також виконання самостійної роботи. Оцінювання сформованих компетентностей у студентів здійснюється за накопичувальною 100-бальною системою. Відповідно до Тимчасового положення "Про порядок оцінювання результатів навчання студентів за накопичувальною бально-рейтинговою системою" ХНЕУ ім. С. Кузнеця, контрольні заходи включають:

поточний контроль, що здійснюється протягом семестру під час проведення лекційних, лабораторних занять і оцінюється сумою набраних балів (максимальна сума – 100 балів);

модульний контроль, що проводиться у формі колоквіуму з ініціативи викладача з урахуванням поточного контролю за відповідний змістовий модуль і має на меті *інтегровану* оцінку результатів навчання студента після вивчення матеріалу з логічно завершеної частини дисципліни – змістового модуля.

Порядок проведення поточного оцінювання знань студентів. Оцінювання знань студента під час лабораторних занять та виконання індивідуальних завдань 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; вміння поєднувати теорію з практикою при розгляді виробничих ситуацій, розв'язанні задач, проведенні розрахунків у процесі виконання індивідуальних завдань та завдань, винесених на розгляд в аудиторії; логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки; арифметична правильність виконання індивідуального та комплексного розрахункового завдання; здатність проводити критичну та незалежну оцінку певних проблемних питань; вміння пояснювати альтернативні погляди та наявність власної точки зору, позиції на певне проблемне питання; застосування аналітичних підходів; якість і чіткість викладення міркувань; логіка, структуризація та обґрунтованість висновків щодо конкретної проблеми; самостійність виконання роботи; грамотність подачі матеріалу; використання методів порівняння, узагальнення понять та явищ; оформлення роботи.

Загальними критеріями, за якими здійснюється оцінювання позааудиторної самостійної роботи студентів, є: глибина і міцність знань, рівень мислення, вміння систематизувати знання за окремими темами, вміння робити обґрунтовані висновки, володіння категорійним апаратом, навички і прийоми виконання практичних завдань, вміння знаходити необхідну інформацію, здійснювати її систематизацію та обробку, самореалізація на практичних та лабораторних заняттях.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час поточного контролю, та балів, отриманих під час модульного контролю за накопичувальною системою. Сумарний результат у балах за семестр складає: "60 і більше балів – зараховано", "59 і менше балів – не зараховано" та заноситься у залікову "Відомість обліку успішності" навчальної дисципліни.

Розподіл балів за тижнями

(вказати засоби оцінювання згідно з технологічною картою)

Теми змістового модуля			Лекційні заняття	Лабораторні заняття	Індивідуальні завдання	Колоквіум	Усього
Змістовий модуль 1.	Тема 1	1-3 тижні	-	3	-	-	3
	Тема 2	4-5 тижні	4	3	-	-	7
	Тема 3	6-7 тижні	-	3	-	-	3
	Тема 4	8-9 тижні	-	3	-	8	11
Змістовий модуль 2.	Тема 5	10-11 тижні	-	3	-	-	3
	Тема 6	12 тиждень	-	3	-	-	3

	Тема 7	13 тиждень	-	3	-	-	3
	Тема 8	14 тиждень	-	3	-	-	3
	Тема9	15-16 тижні	-	6	10	8	24
Іспит			-	-	-	-	40
Усього			4	30	10	16	100

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82 – 89	B	добре	
74 – 81	C		
64 – 73	D	задовільно	
60 – 63	E		
35 – 59	FX	незадовільно	не зараховано
1 – 34	F		

5. Рекомендована література

5.1. Основна:

1) Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining /А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. - СПб.: БХВ-Петербург, 2004. - 336 с.

2) Джеймс Г., Уиттон Д., Хасти Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. / Г. Джеймс, Д. Уиттон, Т. Хасти, Р. Тибширани - Пер. с англ. С. Э. Мастицкого. – М.: ДМК Пресс, 2016. – 450 с.: ил.

3) Дюк В. Data Mining: учебный курс / Дюк В., Самойленко А. : СПб, 2001. – 368 с.

4) Зарядов И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика. – М.: Изд-во РУДН, 2010. – 207 с.

5) Зарядов И. С. Статистический пакет R: теория вероятностей и математическая статистика. – М.: Изд-во РУДН, 2010. – 141 с.

6) Кабаков Р. И. R в действии. Анализ и визуализация данных в программе R / пер. с англ. П. А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.

7) Магнус Я. Р. Эконометрика / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. - М.: Дело, 1997. - 248 с.

8) Мастицкий С. Э., Шитиков В. К. Статистический анализ и визуализация данных с помощью R. — 2014. — 401 с. Электронная книга, адрес доступа: http://r-analytics.blogspot.com

9) Шипунов А. Б. и др. Наглядная статистика. Используем R!. — 2014. — 296 с.

10) Статистический анализ данных в системе R. Учебное пособие / А. Г. Буховец, П. В. Москалев, В. П. Богатова, Т. Я. Бирючинская; Под ред. проф. Буховца А. Г. — Воронеж: ВГАУ, 2010. — 124 с.

5.2. Додаткова:

11) Бабешко Л. О. Основы эконометрического моделирования. - М.: КомКнига, 2006. - 432 с.

12) Дубров А. М. Многомерные статистические методы / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. - М.: Финансы и статистика, 1998. - 350 с.

13) Сошникова Л. А. Многомерный статистический анализ в экономике / Л. А. Сошникова, В. Н. Тамашевич, Г. Уебе, М. Шефер [Под ред.проф. В. Н. Тамашевича. - М.: ЮНИТИ-ДАНА, 1999. - 598с.

14) Плюта В. Сравнительный многомерный анализ в экономических исследованиях. - М.: Статистика, 1980. - 151 с.