

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра системного проектування

Затверджено

На засіданні кафедри системного проектування
факультету електроніки та комп'ютерних
технологій
Львівського національного університету імені
Івана Франка
(протокол №1 від 28 серпня 2023 р.)

Завідувач кафедри  Роман ШУВАР

Силабус з навчальної дисципліни
«Вибрані розділи науки про дані / Selected Sections of Data Science»,
що викладається в межах ОП «Комп'ютерні науки»
другого (магістерського) рівня вищої освіти
для здобувачів з спеціальності
122 – Комп'ютерні науки

Львів 2023

Назва дисципліни	Вибрані розділи науки про дані / Selected Sections of Data Science
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра системного проектування
Галузь знань, шифр та назва спеціальності	12 Інформаційні технології, 122 Комп'ютерні науки
Викладачі дисципліни	Ляшкевич Василь Яремович, канд. техн. наук, доцент, доцент
Контактна інформація викладачів	vasyl.liashkevych@lnu.edu.ua https://electronics.lnu.edu.ua/employee/liashkevych-v-ya
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю): ауд. 317, корпус факультету електроніки та комп'ютерних технологій, м. Львів, вул. Драгоманова, 50. Також можливі онлайн-консультації через MS Teams або систему електронного навчання Moodle. Для погодження часу онлайн-консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://moodle.elct.lnu.edu.ua/enrol/index.php?id=94 https://electronics.lnu.edu.ua/course/aktual-ni-pytannia-nauky-pro-dani
Інформація про дисципліну	Дисципліна «Вибрані розділи науки про дані» є нормативною дисципліною з спеціальності 122 – Комп'ютерні науки для освітньої програми «Комп'ютерні науки», яка викладається в 1 семестрі в обсязі 8 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено таким чином, щоб надати учасникам необхідні знання, обов'язкові для того, щоб оволодіти базовими поняттями, пов'язаними з організацією та використанням даних, використанням технологій роботи з даними. Тому у дисципліні представлено як огляд базових понять та інструментів роботи з даними, так і засобів, які потрібні для вирішення типових завдань при використанні, налаштуванні середовищ та технологій роботи з даними, розробки програм та програмних інтерфейсів.
Мета та цілі дисципліни	<i>Метою</i> вивчення дисципліни «Вибрані розділи науки про дані» є оволодіння базовими поняттями, теоретичними знаннями про дані, можливості інформаційних систем, побудованих на основі опрацювання та аналізу даних, візуалізації даних, побудови конвеєрів для аналізу і перетворення даних з подальшим використанням в різних предметних областях людської діяльності з метою вирішення різного роду задач і бізнес проблем. <i>Ціллю</i> є вміння налаштування сервісів роботи з даними, проектування етапів роботи з даними та розробка програм для роботи з даними.

**Література для
вивчення дисципліни**

Основна література:

1. Christopher M. Bishop (2018) Pattern Recognition and Machine Learning, 738p.
2. Sarah Guido (2016) Introduction to Machine Learning with Python: A Guide for Data Scientists, 400p.
3. EMC Education Services (2015) Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, 432p.
4. Cole Nussbaumer Knaflic (2015) Storytelling with Data: A Data Visualization Guide for Business Professionals, 288p.
5. Peter Bruce (2017) Statistics for Data Scientists: 50 Essential Concepts, 298p
6. Data Mining: The Complete Guide. – Columbia Engineering, 2023. URL: <https://bootcamp.cvn.columbia.edu/blog/data-mining-guide/>
7. Paul Crickard. Data Engineering with Python - Birmingham: Packt Publishing, 2020. - 337 p. - ISBN 978-1-83921-418-9.
8. Wang L., Fu X. Data Mining with Computational Intelligence. –Springer, 2005. –280 p.
9. Wes McKinney. Python for Data Analysis - Sebastopol: O'Reilly Media, 2018. - 522 p. - ISBN 978-1-491-95766-0.
10. Joakim Sundnes. Introduction to Scientific Programming with Python - Lysaker: Simula SpringerBriefs, 2020, Volume 6. - ISBN: 978-3-030-50355-0. (eBook)
11. Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser. Data Structures & Algorithms in Python. Wiley: Courier Westford, 2013. - 748 p. (eBook)
12. Massimo di Pierro. Annotated Algorithms in Python - Chicago: Experts4Solutions, 2017. - 227 p. - ISBN: 978-0-9911604-0-2.

Додаткова література:

13. Allen B. Downey. Think Stats. Exploratory Data Analysis in Python - Needham: Green Tea Press, 2014. - 244 p.
14. Jake VanderPlas. Python Data Science Handbook - Sebastopol: O'Reilly Media, 2017. - 517 p. - ISBN: 978-1-491-91205-8.
15. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: concepts and techniques - Waltham: Elsevier, 2012. - 703 p.
16. Peter Bruce, Andrew Bruce, Peter Gedeck. Practical Statistics for Data Scientists. - Sebastopol: O'Reilly, 2020. - 329 p. - ISBN: 978-1-492-07294-2.
17. Brian Godsey. Think Like a Data Scientist. - Shelter Island: Manning Publications, 2017. - 299 p. - ISBN: 9781633430273.
18. Meher Krishna Patel. Pandas Guide. - May, 2020. - 62 p.
19. Aurelien Geron. Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow. - Sebastopol: O'Reilly, 2019. - 482 p. - ISBN: 978-1-492-03264-9.
20. Lewandowska, A.; Joachimiak-Lechman, K.; Kurczewski, P. A Dataset Quality Assessment—An Insight and Discussion on Selected Elements of Environmental Footprints Methodology. *Energies* **2021**, *14*, 5004. <https://doi.org/10.3390/en14165004>
21. Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. Data Quality Assessment / Communications of the ACM, Volume 45, Issue 4, April 2002 pp. 211–218. - <https://doi.org/10.1145/505248.506010>
22. J. Bicevskis, Z. Bicevska, A. Nikiforova and I. Oditis, "An Approach to Data Quality Evaluation," *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2018, pp. 196-201, doi: 10.1109/SNAMS.2018.8554915.
23. Mats Bergdahl, Manfred Ehling, Eva Elvers and others. Handbook on Data Quality Assessment Methods and Tools. - Wiesbaden, 2007. - 139 p.
24. The Ultimate Guide to Basic Data Cleaning: Atlan, 2014. - 66 p.
25. Dr. Ossama Embarak. Data Analysis and Visualization Using Python - Abu Dhabi: Apress Media LLC, 2018. - 374 p. - ISBN-13 (pbk): 978-1-4842-4108-0.
26. Dimensionality reduction [Режим доступу]: <http://bioconductor.org/books/3.15/OSCA.basic/dimensionality-reduction.html>
27. Data exploration with alluvial plots [Режим доступу]: https://www.datisticsblog.com/2018/10/intro_easyluvial/#features
28. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow: O'Reilly, 2017. - 718 p.
29. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
30. Andrew Ng. Machine Learning Yearning. - [Електронний ресурс]. - Режим доступу: https://nessie.ilab.sztaki.hu/~kornai/2020/AdvancedMachineLearning/Ng_MachineLearningYearning.pdf
31. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
32. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
33. Charu C. Aggarwal. Recommender Systems: Springer, 2016. - 518 p.
34. Kishan G. Mehrotra Chilukuri K. Mohan HuaMing Huang. Anomaly Detection Principles and Algorithms: Springer. - 2017. - 229 p. - DOI: <https://doi.org/10.1007/978-3-319-67526-8>
35. Machine Learning in Computer Vision / N. Sebe, Ira Cohen, Ashutosh Garg, Thomas S. Huang// Springer, 2005. - 249 p. - Режим доступу: [http://silverio.net.br/heitor/disciplinas/eeica/papers/Livros/\[Sebe\]20-%20Machine%20Learning%20in%20Computer%20Vision.pdf](http://silverio.net.br/heitor/disciplinas/eeica/papers/Livros/[Sebe]20-%20Machine%20Learning%20in%20Computer%20Vision.pdf)
36. Mark Richards. Software Architecture Patterns. - Sebastopol: O'Reilly Media, 2015. - 45 p. - ISBN: 978-1-491-92424-2.

	<p>37. Amazon. Lambda Architecture for Batch and Stream Processing. - AWS, 2018. - 12 p.</p> <p>38. Tomcy John, Pankaj Misra. Data Lake for Enterprises. - Packt Publishing, 2017. - 855p.</p> <p>39. Hernandez, M.; Epelde, G.; Beristain, A.; Álvarez, R.; Molina, C.; Larrea, X.; Alberdi, A.; Timoleon, M.; Bamidis, P.; Konstantinidis, E. Incorporation of Synthetic Data Generation Techniques within a Controlled Data Processing Workflow in the Health and Wellbeing Domain. <i>Electronics</i> 2022, <i>11</i>, 812. https://doi.org/10.3390/electronics11050812</p> <p>40. Viktor Mayer-Schonberger, Kenneth Cukier (2013) Big Data: A Revolution That Will Transform How We Live, Work and Think, 256 p.</p> <p>41. Alex Holmes. Hadoop in Practice: Manning Publications, 2012. - 537 p. - Режим доступу: https://ia600201.us.archive.org/7/items/HadoopInPractice/Hadoop%20in%20Practice.pdf</p> <p>42. Apache HBase Team. Apache HBase™ Reference Guide. - [Електронний ресурс]. - Режим доступу: https://hbase.apache.org/apache_hbase_reference_guide.pdf</p> <p>Google. Cloud Bigtable. - [Електронний ресурс]. - Режим доступу: https://cloud.google.com/bigtable</p>
--	--

Обсяг курсу	64 години аудиторних занять. З них 32 години лекцій, 32 годин лабораторних робіт, а також 176 години самостійної роботи.
Очікувані результати навчання	<p>Після завершення цього курсу студент буде:</p> <ul style="list-style-type: none"> - знати основні принципи розробки алгоритмів та програмного забезпечення розв'язування задач добування даних на ПК; - вміти досліджувати алгоритми добування даних, виявляти їх переваги та недоліки, обирати оптимальні алгоритми розв'язування задач, обробки даних та розробляти програми розв'язування задач; виконувати аналіз і опрацювання результатів розв'язування задач, використовувати методи оптимізації. <p>Після вивчення курсу здобувачі набудуть таких компетентностей і програмних результатів:</p> <p>ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>ЗК2. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК5. Здатність вчитися й оволодівати сучасними знаннями.</p> <p>ЗК7. Здатність генерувати нові ідеї (креативність).</p> <p>СК1. Усвідомлення теоретичних засад комп'ютерних наук.</p> <p>СК4. Здатність збирати та аналізувати дані (включно з великими) для забезпечення якості проектних рішень.</p> <p>СК8. Здатність розробляти та реалізовувати проекти зі створення програмного забезпечення, у т. ч. в непередбачуваних умовах, за нечітких вимог і необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.</p> <p>СК9. Здатність розробляти та адмініструвати бази даних і знань.</p> <p>СК13. Здатність застосовувати методи і підходи штучного інтелекту, інтелектуального аналізу та науки про дані та підходів оптимізації до розв'язання конкретних проблем комп'ютерних наук.</p> <p>РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.</p> <p>РН8. Розробляти математичні моделі та методи аналізу даних (включно з великими).</p> <p>РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).</p> <p>РН12. Проектувати та супроводжувати бази даних та знань.</p> <p>РН19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.</p> <p>РН20. Володіти методами та засобами штучного інтелекту, інженерії та</p>

	аналізу даних, розпізнавання образів і адаптивного опрацювання інформації, аналізу та обробки природної мови, моделювання та оптимізації. PH21. Створювати нові системи даних, високорівневі вбудовані системи, спеціалізовані комп'ютерні системи та інтелектуальні системи із застосуванням базових знань апаратного і програмного забезпечення мікроконтролерів і мікрокомп'ютерів.
--	---

Ключові слова	Науки про дані, data science, data, data analysis, регресія, структура даних.
Формат курсу	Очний
	Проведення лекцій, практичних робіт та консультації для кращого розуміння тем.
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Екзамен в кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань у галузі 12 – Інформаційні технології, зокрема з дисциплін «Алгоритми та структури даних», «Об'єктно-орієнтоване програмування», «Крос-платформне програмування».
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, практичні роботи, обговорення, дискусія.
Необхідне обладнання	Комп'ютерний клас, мультимедіа, платформа Moodle, середовище RStudio для програмування мовою R, Moodle, Git, GitHub Pages, Visual Studio Code, мова програмування Python, open source веб-інтерактивне обчислювальне середовище Jupyter, Community Edition PyCharm

<p>Критерії оцінювання (окремо для кожного виду навчальної діяльності)</p>	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою. Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • 7 лабораторних робіт: 32% семестрової оцінки; максимальна кількість балів 32 (5 лабораторних робіт по 4 бали і 2 роботи по 6 балів) • контрольні заміри (2 тестові модулі): 18% семестрової оцінки (максимально 2x9=18 балів); • іспит: 50% семестрової оцінки; максимальна кількість балів 50. <p>Загалом упродовж семестру 100 балів.</p> <p>Оцінки за лабораторні заняття розподіляються наступним чином: виконання практичних завдань – 60%, відповіді на запитання викладача по темі заняття – 40%.</p> <p>Контрольні заміри проводяться у формі тестових завдань.</p> <p>Академічна доброчесність: Очікується, що практичні та контрольні роботи студентів будуть їхніми оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і практичні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися усіх термінів визначених для виконання усіх видів робіт, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані на поточному тестуванні, самостійній роботі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до контрольних робіт</p>	<p>Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-сторінці курсу (система електронного навчання Moodle: https://moodle.elct.lnu.edu.ua/course/view.php?id=94).</p>
<p>Опитування</p>	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності	Література. Ресурси в Інтернеті.	Завдання (лаб. робота), 2 год	Термін виконання, тиж.
1	Базові відомості про дані. Поняття даних. Поняття інформації та знань. Ентропія. Набір даних та їх атрибутів. Вимірювання. Шкали. Типи наборів даних: записи, хімічні дані, графи. Формати зберігання даних. Основні положення про бази даних. Системи управління базами даних. Класифікація видів даних. Метадані.	лекція, лаб. робота	1-4	Вступне заняття. Академічна доброчесність. Добування даних та їх атрибутів, 4 год.	2
2	Методи добування даних. Поняття про процес добування даних, процес KDD (Knowledge Discovery in Databases). Математична база методів добування даних, статистичні методи. Причини виникнення та мотивація розвитку методів добування даних. Суть технології добування даних, мультидисциплінарність. Порівняння методів машинного навчання та методів добування даних. Проблеми підготовки даних. Огляд алгоритмів та методів попередньої обробки даних: очистка, перетворення та скорочення.	лекція, лаб. робота	1-4		3
3	Поняття залежностей по даних. Поняття коваріації. Матриця коваріації. Змінні, що корелюють. Лінійна кореляція. Кореляція Пірсона. Вивчення лінійних залежностей. Метод спроб і помилок. Оптимізація. Кореляція за багатьма змінними. Залишки регресії. Матриця кореляції. Поняття коваріації. Часові ряди. Що таке графік часових рядів. Моделювання даних часових рядів. Часовий ряд і стаціонарність. Модель наполегливості. Авторегресійна модель. Авторегресійне інтегроване ковзне середнє. Сезонність. Аналіз часових рядів. Серійна кореляція, автокореляція. Формування дата-сетів для передбачення аномалій на основі часових рядів.	лекція, лаб. робота	5-11	Побудова бізнес-моделі на основі даних, 4 год.	4
4	Вимірювання та балансування ознак даних. Статистичні дані. Розподіли та гістограми. Нормальний розподіл Гауса. Інші розподіли. Оцінка розподілу за вибіркою. Ядерна оцінка щільності. Властивість рівномірного розподілу. Бімодальний розподіл. Розкид випадкової величини. Центральна гранична теорема. Довірчі інтервали. Передбачувальні інтервали. Перевірка гіпотез. Популяція. Z-нормалізація. Моделювання методом Монте Карло. Вибір границі та діапазону даних для вирішення задач. Використання генетичних алгоритмів для балансування ознак даних.	лекція, лаб. робота	12-21		5
5	Метрики якості даних. Поняття оціночної метрики. Метрики оцінки вирішення задач регресії, класифікації, навчання без учителя, на	лекція, лаб. робота	18-23	Моделювання даних на основі методу Монте	6

	ін. Метрика точності, акуратності, f1-score. Розрахунок метрик. Задовільнення критеріїв та оптимізація метрик. Упередженість, якої можна уникнути. Типові помилки даних. Поняття аналізу помилок. Інтерпретація помилок. Інтерпретаційні моделі. Властивості інтерпретаційних моделей. Бізнес-метрики. Інтеграція бізнес-метрик у конвеєрі даних.			Карло, 4 год.	
6	Моделювання предметної області. Поняття предметної області. Визначення цілей, задач та границь предметної області. Ситуаційні підходи. Збір та підготовка даних. Тематична модель. Методи аналізу тематичних моделей. Візуалізація предметної області. Оцінка та покращення моделей. Формалізований та неформалізований опис предметної області. Текстовий опис предметної області. Використання великих мовних моделей для опису предметної області. Ієрархічний підхід для опису предметної області.	лекція, лаб. робота	3-4, 8, 13-14		7
7	Методи видобування знань. Методи видобування знаня. Виявлення прихованої інформації. Методи прийняття рішення. Кластеризація. Класифікація. Регресія. Асоціативні правила. Аналіз аномалій. Аналіз текстових ресурсів. Аналіз зображень. Аналіз соціальних мереж. Древа рішень. Ансамблі моделей. Зв'язкові правила. Методи глибинного навчання.	лекція, лаб. робота	9-10, 15, 17, 19	Формування метрик якості даних, 4 год.	8
8	Методи очищення даних та знань. Декомпозиція та агрегація даних. Трансформація даних. Виявлення та вилучення відсутніх даних. Виявлення аномальних даних. Нормалізація та Стандартизація даних. Кодування даних. Пошук та видалення копій. Валідація даних. Скорочення розмірності. Ітераційні дані. Таксономія. Поняття таксономії тематичної області. Діаграма Хассе. Експертні системи. Системи логічного висновку. Очищення правил-продукцій. Приклади та контр-приклади в базах знань.	лекція, лаб. робота	14, 17, 20, 24		8
9	Основні положення аналізу даних. Поняття процесу перетворення даних. Інтерпретація даних. Збір та очищення даних. Підготовка даних. Програмні конвеєри даних. Дослідницький аналіз даних. Моделювання даних. Валідація моделей та формування висновків. Способи подання даних. Візуалізація даних. Повторний цикл аналізу даних. Керування та застосування даних. Поняття штучних даних.	лекція, лаб. робота	1-4, 13, 16	Вибір засобів візуалізації бізнес-значень, 4 год.	10
10	Особливості візуального аналізу даних. Поняття простору даних. Поняття великої розрядності. Метод аналізу основних компонентів (PCA). Особливості роботи PCA. Поняття Eigenvector. Поняття генів. Дисперсія. Застосування PCA. Вибір системи оцінювання даних. Атрибути якості даних. Т-розподілене стохастичне вбудовування сусідів (t-SNE). Масштабування відстані та складність. tSNE проєкції. Застосування t-SNE. Поняття UMAP. Переваги UMAP. Практичні рекомендації	лекція, лаб. робота	4, 25-27		11

	щодо застосування PCA + tSNE/UMAP. Методи просторового кодування даних. Робота з даними у векторному просторі.				
11	Дані і штучний інтелект. Життєвий цикл науки про дані. Взаємозв'язок між даними та знаннями. Поняття семантичних мереж. Різновиди алгоритмів машинного навчання. Типи та види навчання. Штучний інтелект. Компоненти штучного інтелекту. Ера когнітивних обчислень. Вузкий штучний інтелект. Когнітивні обчислення. Просторово-часова аналітика. Генеративний інтелект. Супер інтелект. Великі мовні моделі.	лекція, лаб. робота	1-2, 15, 19, 28-30	Використання ситуаційного підходу для побудови бази-знань, 4 год.	12
12	Бази знань та онтології. Декомпозиція та агрегація даних. Трансформація даних. Поділ даних. Визначення факту, правила, поняття, події та взаємозв'язків між ними. Поняття словника. Принципи збереження знань. Сховище знань. Експертні системи. Керування знаннями. Поняття онтології. Семантична модель онтології. Графи знань. Формат представлення знань. Контекст. Виразність. Семантичний рівень. Коцептуальний рівень. Абстрактний рівень.	лекція, лаб. робота	29-32		13
13	Самонавчання та інтелектуальні системи. Поняття інтелектуалізованих та інтелектуальних систем. Поняття інтелекту. Роль машинного навчання для інтелектуальних систем. Складові та характеристики інтелектуальних систем. Збір та збереження знань. Моделі аналізу даних. Моделі генерації знань. Вибір моделі самонавчання. Навчання системи. Оцінка продуктивності. Адаптація та покращення поделей. Безперервний процес навчання інтелектуальних систем. Метрики та методи прийняття рішень. Сучасні засоби для побудови інтелектуальних систем. Архітектура інтелектуальних систем.	лекція, лаб. робота	32-35		14
14	Проектування систем, керованих даними. Визначення сфери інтересу. Визначення цілей. Вибір метрик. Вибір ресурсів та колекцій даних. Візуалізація даних. Визначення алгоритму прийняття рішень. Концепція дизайну систем, керованих даними. Архітектура даних. Вибір бази чи сховищ даних. Процеси керування. Синхронізація процесів керування даних. Цілі та гіпотези. Вибір стратегії проектування на основі даних. Використання хмарних технологій для проектування систем, керованих даними.	лекція, лаб. робота	36-38	Проектування системи, що керується даними на основі розробленої бізнес-моделі, 4 год.	15
15	Концепція великих даних. Труднощі та рішення. Модель Spark. Швидкодія Spark. Оцінка технічного стеку Spark. Поняття Spark SQL. Програмний інтерфейс. Модель даних. Поняття DataFrame. Операції з DataFrame. Запити. Оптимізація та виконання. Генерація коду. Розширення. Розширені функції аналітики. Spark MLib конвеєри. Дослідницька трансформація.	лекція, лаб. робота	39-41		16
16	Проектування інтелектуалізованих масштабованих розподілених систем.	лекція, лаб. робота	42-43		16

<p>Розподілені системи опрацювання даних. Поняття масштабування в розподілених системах опрацювання даних. Поняття ETL (Extract, Transform, Load). Розбиття даних за ключами. Розбиття даних за файлами. Перекіс даних. Зсуваюче об'єднання. Інші типи об'єднань. Різниця між ETL та ELT. Робочий потік даних в AWS GLU. Azure Data Factory. Сервіси бізнес даних. Технологічний стек. Поточкові дані з Apache Kafka. Побудова Kafka кластеру. Створення та поглинання з Python. Приклади.</p>				
--	--	--	--	--