



Аналіз даних Інтернет медіа та соціальних мереж Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 інформаційні технології</i>
Спеціальність	<i>122 Комп'ютерні науки</i>
Освітня програма	<i>Системи і методи штучного інтелекту</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>I курс, весняний семестр</i>
Обсяг дисципліни	<i>загальна кількість: (4, кр.) 120 год , лекційних занять: 36 год практичних занять: 18 год, самостійна робота студента: 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>залік, МКР</i>
Розклад занять	<i>https://schedule.kpi.ua/ 2 год лекційних та 1 год практичних занять на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: доктор технічних наук, професор, Ланде Дмитро Володимирович d.lande@kpi.ua</i>
Розміщення курсу	<i>Сервісу Zoom / Google Meet (за узгодженням зі студентами)</i>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Курс ставить на меті формування у студентів глибинного розуміння сучасних теоретичних засад і методів інтелектуального аналізу даних із Інтернет медіа та соціальних мереж: алгоритмів, моделей, задач збору і попередньої обробки інформації, класифікації, кластерного аналізу, пошуку, глибинного аналізу даних (Data Mining), теорії складних мереж (Complex Networks), а також навиків, необхідних для комп'ютерного моделювання та аналізу складних систем і мереж, що утворюють сучасне інформаційне середовище. Для цього передбачається системне ознайомлення студентів із науковими напрямками та методами, які використовуються при розв'язку задач аналізу даних із Інтернет медіа та соціальних мереж в рамках фахової діяльності. Аналіз даних - це комплексний науковий напрям, що знаходиться на перетині таких наук, як дискретна математика, теорія штучного інтелекту, комп'ютерна лінгвістика, теорія графів, теорія алгоритмів тощо. Вивчення курсу є необхідним етапом освіти фахівця з інформаційних технологій, що займається прикладними проблемами і закладає базу для подальшої спеціалізації.

Дисципліна знайомить з перспективними напрямками і тенденціями розвитку інформаційних технологій в галузі інтелектуального аналізу даних із Інтернет медіа та соціальних мереж і базується на сучасних наукових досягненнях в цій сфері за останні роки.

Особлива увага приділяється розгляду засібів агрегації великих обсягів неструктурованих даних із Інтернет медіа та соціальних мереж, їх збиранню, пошуку, обробки та візуалізації. В рамках цього курсу, крім іншого, розглядається екосистема компонентів Elastic Stack, а саме, Elasticsearch – це інформаційно-пошукова система, ядро Elastic Stack, яка дозволяє здійснювати обробку неструктурованих даних, інформаційний пошук, аналіз даних; утиліта Kibana – це вікно в Elastic Stack, засіб маніпуляції, аналізу і візуалізації інформації.

Також курс містить відомості щодо інших технологічних платформ оброблення даних із Інтернет медіа та соціальних мереж, серед яких система Manticore Search, призначена для організації розподіленого пошуку в великих масивах слабкоструктурованих даних. Окремо розглядаються засоби аналізу великих мереж, графових СКБД. Розглядаються можливості двох основних систем – програм аналізу та візуалізації графів Gephi і графова система керування базами даних Neo4j. Графова СУБД Neo4j забезпечує збереження і обробку мережевих даних великих обсягів, містить декларативну мову запитів до графів Cypher.

Вивчення курсу базується на ознайомленні з останніми практичними результатами та теоретичними науковими дослідженнями в галузі інтелектуального аналізу даних із Інтернет медіа та соціальних мереж. З огляду на специфіку курсу його засвоєння передбачає знання англійської мови на рівні достатньому для вільного читання наукових текстів. Значний обсяг самостійної роботи спрямований на розвиток у студентів навичок ефективного пошуку науково-технічної інформації, її систематизації та викладу її у концентрованому вигляді в презентації та/або літературного огляду.

Здобувачі на лекціях беруть участь в бліц-опитуваннях та виконують значний обсяг самостійної роботи, що включає самостійний пошук, систематизацію, узагальнення свіжих наукових робіт, підготовки звітів-оглядів за окремими темами. Заохочується презентації власних наукових здобутків дотичних до тематики курсу, участі в конференціях, школах, презентації лекцій на тематику курсу для колег та студентів молодших курсів тощо.

Під час навчання використовуються:

- Технічні засоби подання інформації (мультимедійні комплекси, інтерактивні дошки тощо);
- Google-диск з комплектом методичного забезпечення в електронному вигляді;
- Засоби дистанційного навчання (електронна пошта, Google meet, zoom тощо)

Навчання здійснюється на основі студентоцентрованого підходу та стратегії взаємодії викладача та здобувача з метою засвоєння здобувачами матеріалу та розвитку у них відповідних компетентностей.

Програмні результати навчання

Згідно з вимогами освітньо-професійної програми студенти після засвоєння навчальної дисципліни мають продемонструвати такі результати навчання:

PH 8 Розробляти математичні моделі та методи аналізу даних (включно з великими)

PH 9 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими)

PH 13 Оцінювати та забезпечувати якість інформаційних та комп'ютерних систем різного призначення

PH 26 Розробляти адекватні методи навчання та самонавчання, включаючи методи глибокого навчання (Deep Learning) та використовувати їх для налаштування нейронних мереж для вирішення конкретних задач прогнозування, керування, класифікації та інтелектуального аналізу даних

Набуті знання та практичні навички сформують у здобувачів:

Загальні компетентності:

ЗК 07 Здатність генерувати нові ідеї й нестандартні підходи до їх реалізації (креативність)

Спеціальні (фахові, предметні) компетентності:

СК 04 Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень

СК 06 Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук

СК 08 Здатність розробляти і реалізовувати проєкти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проєктом

СК 14 Здатність вибирати адекватні методи навчання, включаючи методи глибокого навчання (Deep Learning) і самонавчання та використовувати їх для налаштування нейронних мереж для вирішення конкретних задач прогнозування, керування, класифікації та інтелектуального аналізу даних)

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Дисципліна базується на знаннях теоретичних і технологічних засадах дискретної математики, зокрема, теорії графів, алгебри. інформаційних технологій, алгоритмів та структур даних, баз даних та інформаційних систем, моделей захисту інформації, а також іноземної мови професійного спрямування.

3. Зміст навчальної дисципліни

Розділ 1. Теоретичні основи аналізу даних із Інтернет медіа та соціальних мереж

Тема 1. Введення в дисципліну «Аналіз даних із Інтернет медіа та соціальних мереж»

Тема 2. Збір даних, консолідація, класифікація, візуалізація.

Тема 3. Класифікація. Кластерний аналіз.

Тема 4. Основи глибокого навчання (Deep Learning)

Тема 5. Концепція комплексних мереж (Complex Networks)

Розділ 2. Застосування методів аналізу даних із Інтернет медіа та соціальних мереж

Тема 6. Технологічні засади аналітики даних із Інтернет медіа та соціальних мереж

Тема 7. Технології роботи з Big Data. Концепція Map Reduce. Інструменти Apache Hadoop

Тема 8. Технології роботи з Big Data. Інструменти пошуку неструктурованих даних Elasticsearch.

Тема 9. Технології роботи з Big Data. Інструмент візуалізації і аналізу даних (Kibana)

Тема 10. Технології роботи з Big Data. Технології роботи з Big Data. Інструменти мережевого аналізу (Gephi , Neo4j).

4. Навчальні матеріали та ресурси

Основна:

1. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник.
/ Д.В. Ланде, І.Ю. Субач, А.Я. Гладун. / - Київ 2021. - 168 с.
<https://ela.kpi.ua/handle/123456789/46129>

2. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник / Д.В. Ланде, І.Ю. Субач, Ю.Є. Бояринова. – Київ: ІСЗЗІ КПІ ім. Ігоря Сікорського, 2018. – 300 с. <https://ela.kpi.ua/handle/123456789/45721>
3. Information Operations Recognition. From Nonlinear Analysis to Decision-Making A. Dodonov, D. Lande, V. Tsyganok, O. Andriichuk, S. Kadenko, A. Graivoronskaya. - LAP Lambert Academic Publishing, 2019. - 292 p. ISBN-13: 978-620-0-27697-1, ISBN-10: 6200276978, EAN: 9786200276971. url: https://www.researchgate.net/profile/D-Lande/publication/323399024_Information_Operations_Recognition/links/5a94133fa6fdccceff063d60/Information-Operations-Recognition.pdf
4. Karl Seguin. "The Little MongoDB Book". –34 p. URL: <http://github.com/karlseguin/the-little-mongodb-book>
5. Ланде Д.В., Субач І.Ю. Візуалізація та аналіз мережевих структур : навчальний посібник. - Київ : КПІ ім. Ігоря Сікорського, Вид-во "Політехніка", 2021. - 80 с. <https://ela.kpi.ua/handle/123456789/45722>

Додаткова:

6. Elasticsearch: The Definitive Guide / Clinton Gormley and Zachary Tong. - O'Reilly Media, Inc., 2015. – 719 p. (за запитом викладачу)
7. Elasticsearch Blueprints. A practical project-based guide to generating compelling search solutions using the dynamic and powerful features of Elasticsearch / Vineeth Mohan. – Packt Publishing, 2015. – 192 p. (за запитом викладачу)
8. Learning Kibana 5.0. Exploit the visualization capabilities of Kibana and build powerful interactive dashboards / Bahaaldine Azarmi. – Packt Publishing, 2017. – 275 p. (за запитом викладачу)
9. Fabio Nelli. Python Data Analytics: Data Analysis and Science Using Pandas, matplotlib, and the Python Programming Language. – Apress, 2015. – 350 с. (за запитом викладачу)
10. Holzschuher, Florian and Peinl, Rene (2013). "Performance of Graph Query Languages: Comparison of Cypher, Gremlin and Native Access in Neo4J" in EDBT '13. Proceedings of the Joint EDBT/ICDT 2013 Workshops: 195-204, Genoa, Italy: ACM. DOI:10.1145/2457317.2457351 (за запитом викладачу)

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

В рамках дисципліни заплановано наступні види навчальних занять:

- лекції;
- практичні заняття;
- самостійна робота.

Теми дисципліни взаємозв'язані, матеріал вивчається в логічній послідовності. На лекціях розкриваються найбільш суттєві теоретичні питання, які дозволяють забезпечити аспірантам можливість глибокого самостійного вивчення всього програмного матеріалу. Теми та порядок виконання практичних завдань сформовано в логічній послідовності і повністю узгоджуються з лекційним матеріалом. Теоретичні і практичні знання поглиблюються шляхом самостійної роботи з використанням рекомендованої літератури та глобальної мережі Internet.

На заняттях використовуються звичайна дошка та/або інтерактивна дошка, віртуальна дошка (в умовах карантинних обмежень), а також презентації лекцій з використанням мультимедіапроектора або дистанційно.

Контроль засвоєння навчального матеріалу здійснюється індивідуальним опитуванням, перевіркою домашніх завдань та тестів, контрольною роботою та семестровим іспитом.

№ з/п	Тема	Основні завдання	
		Контрольний захід	Термін виконання
Розділ 1. Теоретичні основи аналізу даних із Інтернет медіа та соціальних мереж			
1.	Тема 1. Введення в дисципліну «Аналіз даних із Інтернет медіа та соціальних мереж» (1 лекція) Поняття інтелектуального аналізу даних. Поняття «данні», «інформація» і «знання», співставлення и порівняння цих понять. Визначення і методи глибинного аналізу даних (Data Mining). Література: [1], [2], [3]	Бліц-опитування на лекціях	1-й тиждень
2.	Тема 2. Збір даних, консолідація, класифікація, візуалізація. (1 лекція) Витяг даних з різних джерел, забезпечення їх якості, перетворення в єдиний формат і завантаження в сховище даних. Перетворення чисельних даних в деякий візуальний образ, з метою спрощення сприйняття великих масивів інформації. Розподіл об'єктів дослідження по заздалегідь відомим класам на підставі подібності ознак. Література: [1], [2], [3], [5], [6]	Бліц-опитування на лекціях. Практичне заняття	2-й тиждень
3.	Тема 3. Класифікація. Кластерний аналіз. (2 лекції) Співвідношення класифікації та кластеризації. Класифікація як машинне навчання з учителем (Supervised Machine Learning). Типовий процес класифікації. Математичний апарат класифікації. Моделі класифікації: класифікаційні правила, дерева рішень, математичні (логічні) формули. Алгоритми класифікації. Оцінка якості класифікації. Кластеризація як машинне навчання без вчителя (Unsupervised Machine Learning). Математична формалізація кластерного аналізу. Алгоритми кластеризації. Література: [1], [2], [5]	Бліц-опитування на лекціях. Практичне заняття	4-й тиждень
4.	Тема 4. Основи глибинного навчання (Deep Learning) (2 лекції) Нейронні мережі як метод класифікації. Карти Кохонена як нейронні мережі. Визначення машинного навчання (machine learning, ML). Машинне навчання як засіб інтелектуального аналізу великих масивів даних. Обмеження машинного навчання. Формалізація задачі машинного навчання. Вибір готового	Бліц-опитування на лекціях. Практичне заняття	6-й тиждень

	<p>програмного інструменту на базі штучних нейронних мереж для вирішення задачі машинного навчання. Застосування готового програмного інструменту для вирішення задачі машинного навчання.</p> <p>Література: [1]</p>		
5.	<p>Тема 5. Концепція комплексних мереж (Complex Networks) (2 лекції)</p> <p>Основні поняття концепції комплексних мереж. Комплексні мережі і великі дані. Основні параметри комплексних мереж. Розподіл ступенів вузлів складних мереж, безмасштабні мережі.</p> <p>Ранжирування вузлів мережі – параметр PageRank.</p> <p>Моделювання комплексних мережах.</p> <p>Алгоритми децентралізованого пошуку в комплексних мережах.</p> <p>Література: [1-3], [5]</p>	<p>Бліц-опитування на лекціях.</p> <p>Практичне заняття</p>	8-й тиждень
Розділ 2. Застосування методів аналізу даних із Інтернет медіа та соціальних мереж			
6	<p>Тема 6. Технологічні засади аналітики даних (2 лекції)</p> <p>Поняття, сфери застосування аналітики даних. Машинне навчання в аналітичному обробленні надвеликих масивів даних. Програмне забезпечення для реалізації алгоритмів машинного навчання для великих даних.</p> <p>Технології NoSQL. Особливості розробки інформаційних систем на базі NoSQL-рішень. Особливості та Технологічні рішення на прикладі СУБД MongoDB, CouchDB та Redis.</p> <p>Порівняння та оцінювання сучасних рішень на базі концепції NoSQL.</p> <p>Література: [3], [6], [8]</p>	<p>Бліц-опитування на лекції.</p> <p>Практичне заняття</p>	10-й тиждень
7	<p>Тема 7. Технології роботи з Big Data. Концепція Map Reduce. Інструменти Apache Hadoop (2 лекції)</p> <p>Apache Hadoop, MapReduce – парадигма розподілених обчислень для великих даних. Технологія Apache Hadoop для організації розподіленого оброблення великих об'ємів даних.</p> <p>HDFS - файлова система для зберігання файлів надвеликих розмірів.</p> <p>YARN - модуль, що забезпечує керування ресурсами кластерів та планування завдань.</p> <p>Основи концепції MapReduce для великих</p>	<p>Бліц-опитування на лекціях.</p> <p>Практичне заняття</p>	12-й тиждень

	<p>даних. Масштабованість в MapReduce. Відмовостійкість в MapReduce Універсальність MapReduce. Література: [1-3]</p>		
8	<p>Тема 8. Технології роботи з Big Data. Інструменти пошуку неструктурованих даних Elasticsearch (2 лекції) Визначення термінів. Elastic Stack як екосистема компонентів, що служать для пошуку і обробки даних. Основні компоненти Elastic Stack - Kibana, Logstash, Beats, X-Pack і Elasticsearch. Інформаційно-пошукова система Elasticsearch – ядро Elastic Stack. Технологія Apache Lucene. Відмінності Elasticsearch від традиційних рішень для реляційних баз даних. Особливості Elasticsearch. REST API. Література: [3], [6], [4]</p>	<p>Бліц-опитування на лекціях. Практичне заняття</p>	14-й тиждень
9	<p>Тема 9. Технології роботи з Big Data. Інструмент візуалізації і аналізу даних (Kibana) (2 лекції) Утиліта Kibana як засіб візуалізації і призначеним для користувача інтерфейсом для Elastic Stack. Варіанти візуалізацій в Kibana: гістограма, карта, лінійні графіки, часові ряди. Інструменти для управління і розробки. Створення і тестування запитів REST (Representational State Transfer) API. Встановлення і аналіз можливостей утиліти Kibana. Візуалізації даних із компонент Elastic Stack. Інтерфейс доступу до створеної бази даних Elasticsearch засобами утиліти Kibana. Варіанти візуалізацій в Kibana: гістограма, карта, лінійні графіки, часові ряди. Література: [3], [6], [7]</p>	<p>Бліц-опитування на лекціях. Практичне заняття</p>	16-й тиждень
10	<p>Тема 10. Технології роботи з Big Data. Інструменти мережевого аналізу (Gephi , Neo4j) (2 лекції) Пакет програмного забезпечення для мережевого аналізу та візуалізації Gephi. Графічні формати даних: GML, GraphML, GraphViz, JSON, CSV. Neo4j - графова система управління базами даних. Графова СУБД Neo4j. Термінологія Neo4j і графових баз даних в цілому. Збереження даних в Neo4j. Основні транзакційні можливості. Cypher – декларативна мова запитів до графів. Основні області застосування Neo4j. Література: [3], [6]</p>	<p>Бліц-опитування на лекціях. Практичне заняття</p>	18-й тиждень

Практичні заняття

Основні завдання практичних занять:

Навчитись розв'язувати задачі в галузі інтелектуального аналізу даних великих обсягів даних (Big Data) із Інтернет медіа та соціальних мереж. Знати й вміти застосовувати відповідні програмно-інструментальні засоби, у тому числі, потужну пошукову систему Elasticsearch, систему обробки графових баз даних Neo4j, сучасні NoSQL системи керування базами даних.

6. Самостійна робота студента

Самостійна робота студентів має на меті розвиток творчих здібностей та активізація їх розумової діяльності, формування потреби безперервного самостійного поповнення знань та розвиток здатностей презентувати результати навчання та досліджень. Завданням самостійної роботи аспірантів є навчити аспірантів самостійно працювати з літературою, творчо сприймати навчальний матеріал і осмислювати його та формування навичок до щоденної роботи з метою одержання та узагальнення знань, умінь і навичок.

На самостійну роботу відводяться наступні види завдань:

- обробка і осмислення інформації, отриманої безпосередньо на заняттях;
- робота з відповідними підручниками, літературою, та особистим конспектом лекцій;
- підготовка реферату за узгодженою темою;
- підготовка до складання семестрового контролю.

Самостійна робота заохочується високим рейтинговим балом. Виконується за тематикою, яка вибирається здобувачем самостійно після обговорення з викладачем і є дотичною або охоплює тему дисертаційної роботи здобувача. Здобувач обирає 1-3 теми із курсу, які хотів би дослідити глибше. Робота за кожним із модулем закінчується коротким, змістовним звітом (від 5 до 10 сторінок). Звіт може бути у вигляді: А) Огляду найсвіжшої наукової літератури на обрану тему. Пріоритет надається оригінальним науковим статтям, оглядам, збіркам, монографіям відомих видавництв (Elsevier, Academic press, John Wiley & Sons тощо) із високим індексом цитування; В) Презентації обсягом від половини до однієї академічної години на обрану тему; С) Результат моделювання фізичного процесу; D) Презентації обсягом від половини до однієї академічної години результатів власних досліджень, дотичних до тематики курсу; E) інші форми СР (звіти з участі в конференціях, школах, лекції на тематику курсу тощо)

Критерії оцінювання СРС: максимальна кількість балів за звіт – 28 балів:

- Максимальна оцінка – лаконічно викладено 95% інформації, що стосується тематики, інформація релевантна, подана лаконічно, послідовно і структуровано, не калькована, наведено ілюстрації, посилання, формулювання та терміни, терміни роз'яснено. Презентація зроблена послідовно, структуровано, не переобтяжена деталями, якісно представлена.
- Зменшення від 1 до 3 балів – викладено не більше ніж 65% інформації, що стосується тематики, інформація релевантна, проте переобтяжена деталями, подана в основному послідовно і структуровано, не калькована, наведено ілюстрації, посилання, формулювання та терміни в основному точні, терміни роз'яснено. Презентація зроблена в основному послідовно, структуровано, можливо переобтяжена деталями, представлена задовільно.
- Зменшення від 4 до 7 балів – в огляді викладено не більше ніж 35% інформації, що стосується тематики, інформація переобтяжена деталями, подана не послідовно і не структуровано, часто калькована, мало ілюстрована, посилання відсутні або неповні, формулювання та терміни не точні, терміни не роз'яснено. Презентація зроблена не

послідовно і не структуровано, переобтяжена деталями, представлена погано або затягнута

Валідність оцінок забезпечується:

- чіткими критеріями оцінювання

Заохочується також самостійне опанування сертифікованого курсу за тематикою дисципліни або дотичною до неї (за наявності сертифікату).

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Програмні результати навчання, політика навчальної дисципліни, методика її опанування, контрольні заходи та терміни виконання оголошуються аспірантами на першому занятті.

Відвідування лекцій, а також відсутність на них, не оцінюється. Однак, здобувачам рекомендується відвідувати заняття, оскільки викладений на них теоретичний матеріал та надані інструкції дозволять ефективніше зорієнтуватися у темі, вибраній для самостійної роботи, спланувати її виконання та спосіб дослідження.

Заохочувальні та штрафні бали наведені в таблиці.

Заохочувальні бали		Штрафні бали	
Критерій	Ваговий бал	Критерій	Ваговий бал
Звіт з участі у міжнародних, всеукраїнських та/або інших заходах та/або конкурсах за тематикою близькою до тематики курсу	До 10 балів, залежно від рівня представництва і якості звіту	Порушення термінів виконання звітування за самостійну роботу	-5 балів
Виступ із лекцією перед студентами молодших курсів на обрану тему за тематикою близькою до тематики курсу	До 10 балів сумарно		
Самостійне опанування сертифікованого курсу за тематикою дисципліни або дотичною до неї.	Заміщення складників дисципліни		

Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>).

Норми етичної поведінки

Норми етичної поведінки здобувачів вищої освіти і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». (<https://kpi.ua/code>).

Процедура оскарження результатів контрольних заходів

Здобувачі вищої освіти мають можливість підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними

процедурами (згідно «Положення про систему забезпечення якості вищої освіти у Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського», «Положення про організацію навчального процесу»).

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: виступи за регламентом.

Календарний контроль: атестація проводиться двічі на семестр як моніторинг поточного стану виконання вимог програми.

Критерій		Перша атестація	Друга атестація	
Термін атестації		8-й тиждень	14-й тиждень	
Умови отримання атестації	Поточний рейтинг	≥ 15 балів	≥ 30 балів	
	Поточний контрольний захід	Модульна контрольна робота	-	+
	Практичні заняття	ПЗ№1	+	-
		ПЗ№2	+	-
		ПЗ№3	+	-
		ПЗ№4	+	-
		ПЗ№5	-	+
		ПЗ№6	-	+
		ПЗ№7	-	+
		ПЗ№8	-	+
ПЗ№9		-	+	

Академічна доброчесність

Семестровий контроль: залік

Умови допуску до семестрового контролю:

Обов'язкові:

- Активність на практичних заняттях
- Виступи за регламентом з доповіддю.
- Поточний рейтинг $RD \geq 42$ балів

Необов'язкові:

- Активність на лекційних заняттях.
- Позитивний результат першої атестації та другої атестації.
- Виконання семестрового індивідуального завдання.

Система рейтингових балів:

Розрахунок шкали рейтингу:

з/п	Контрольний захід	Ваговий бал	Кіл-ть	Всього
1.	Семестрове індивідуальне завдання	25	1	25
2.	Робота на практичних заняттях	5	9	45
3.	Модульна контрольна робота (МКР)	30	1	30
	Всього			100

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

Питання до заліку:

1. Визначення та термінологія із області аналізу соціальних мереж (SNA).
2. Визначити основні властивості систем || обробки великих даних: консолідація, візуалізація, класифікація.
3. Великі дані, як сучасний феномен (пояснити скорочення VVV).
4. Застосування даних із соціальних мереж.
5. Еталонна архітектура для великих даних (Big Data Reference Architecture)
6. Структуровані, неструктуровані та напівструктуровані типи даних.
7. Метадані.
8. Характеристики великих даних із соціальних мереж.
9. Концепція комплексних мереж.
10. Алгоритм HITS. Ранжування вузлів мережі
11. Основні параметри вузлів мережі.
12. Карти Кохонена як нейронні мережі з навчанням без вчителя.
13. Штучні нейронні мережі. Перцептрон Розенблата.
14. Алгоритм PageRank. Ранжування вузлів.
15. Модель малого світу Ваттса-Строгатца.
16. Класифікація джерел великих даних.
17. Походження даних.
18. Процес здобуття даних
19. Ідентифікація даних
20. Збір і фільтрація даних
21. Екстрагування даних
22. Перевірка й очищення даних
23. Агрегація та подання даних
24. Візуалізація даних
25. Використання результатів аналізу
26. Поняття аналітики даних для великих даних
27. Використання сховища даних .
28. Вітрини даних для надвеликих масивів даних.
29. Недоліки традиційних баз даних для зберігання надвеликих масивів даних. NoSQL-бази даних.
30. Система керування базою даних NoSQL. Реплікація масивів даних.
31. Особливості розробки інформаційних систем на базі NoSQL-рішень.
32. Особливості NoSQL СУБД (MongoDB, CouchDB та Redis).
33. Життєвий цикл аналітики великих даних
34. Процедура ETL: витяг, перетворення й завантаження великих даних.

35. Поняття кластера в сфері обробки надвеликих масивів даних.
36. Шардинг для організації сховищ великих даних.
37. Реплікація для організації обробки великих даних.
38. Основні поняття обробки великих даних
39. Методи класифікації. Складність алгоритмів класифікації.
40. Метод опорних векторів як реалізація машинного навчання. Основні ідеї.
41. Метод ієрархічного агрегування (НАС).
42. Метод матричного латентного семантичного індексування (LSI).
43. Розподілена обробка великих даних
44. Особливості платформ для організації розподіленої обробки великих даних
45. Визначення поняття глибинного навчання (Deep Learning)
46. Пакетна обробка великих даних
47. Транзакційна обробка великих даних
48. Обробка великих даних в пакетному режимі
49. Переваги та недоліки застосування моделі Map Reduce.
50. Технологія Apache Hadoop для організації розподіленого оброблення великих об'ємів даних.
51. Задачі Map і Reduce в обробці великих даних
52. Огляд технологій зберігання великих даних.
53. Бази даних та моделі даних.
54. Підготовка вихідних даних для аналізу: первинна обробка й візуалізація наявних даних.
55. Визначення термінів машинного навчання.
56. Типи задач машинного навчання: навчання з учителем, навчання без учителя та навчання з підкріпленням.
57. Взаємозв'язок машинного навчання з іншими областями аналітики великих даних.
58. Штучні нейронні мережі.
59. Мова Python у сфері роботи з даними. Основні елементи, можливості, особливості.
60. Можливості інформаційно-пошукової системи Elasticsearch.
61. Екосистема компонентів Elastic Stack.
62. Особливості технології Apache Lucene.
63. Відмінності Elasticsearch від традиційних рішень для реляційних баз даних.
64. Аналітичні засоби Kibana.
65. Варіанти візуалізацій в Kibana.
66. Властивості і можливості запитів REST API.
67. Основні можливості системи аналізу і візуалізації графів Gephi.
68. Особливості інтерфейсу користувача системи Gephi.
69. Основні графічні формати даних: GML, GraphML, GraphViz, JSON, CSV.
70. Графова СУБД Neo4j. Основні можливості і особливості.
71. Основні елементи мови запитів до графів Cypher.

Робочу програму навчальної дисципліни (силабус):

Складено: доктор технічних наук, професор, Ланде Дмитро Володимирович

Ухвалено кафедрою ШІ (протокол № 1 від 5.07.2022)

Погоджено Методичною комісією НН ІПСА (протокол № 8 від 17.06.2022)