

<b>Назва курсу</b>	«Аналіз і обробка великих даних»
<b>Викладач (-і)</b>	Комар Мирослав Петрович
<b>Профайл викладача (-ів)</b>	<a href="http://www.tneu.edu.ua/fkit/department-iosu-fkit/">http://www.tneu.edu.ua/fkit/department-iosu-fkit/</a>
<b>Контактний тел.</b>	+380352-475050ext.12321
<b>E-mail:</b>	<a href="mailto:mko(@)tneu.edu.ua">mko(@)tneu.edu.ua</a>
<b>Сторінка курсу в moodle</b>	<a href="https://moodle.tneu.edu.ua">https://moodle.tneu.edu.ua</a>
<b>Консультації</b>	Очні консультації: понеділок: 9-35, ауд. 2014. Онлайн - консультації: у вібер групі курсу кожного дня з 16-00 до 18-00.

**1. Анотація до курсу.** Даний курс знайомить з предметною областю великих даних (big data), а також показує взаємозв'язки з наукою про дані (data science) та аналізом даних (data analytics) та забезпечує: оволодіння навичками оперування великими даними шляхом використання спеціалізованих програмно-апаратних засобів, зокрема хмарних сервісів, спеціалізованих систем зберігання даних, розподілених файлових систем тощо; розвиток уміння працювати з великими даними з урахуванням їх ключових характеристик: обсягу, різноманітності, мінливості й забезпечуючи відповідний рівень швидкості їх опрацювання; розвиток навичок ефективного використання математичного, алгоритмічного та програмного забезпечення для розв'язання основних задач предметної області великих даних. Ви зможете ефективно використовувати парадигми паралельного опрацювання даних, зокрема MapReduce та системи Apache Hadoop, Apache Spark, відповідні хмарні служби Amazon Web Services та IBM Bluemix; розгортати надійні та швидкі сховища для надвеликих обсягів даних; використовувати програмні бібліотеки та фреймворки з ефективними алгоритмами опрацювання надвеликих обсягів даних. Вміти аналізувати й ефективно застосовувати хмарні системи опрацювання великих даних.

## **2. Пререквізити.**

Раніше вивчені дисципліни, пов'язані з інтелектуальним аналізом даних, організацією баз даних та знань.

**Постреквізити.** Дисципліни, які будуть використовувати результати навчання даного курсу: підготовка дисертаційної роботи.

## **3. Мета та цілі курсу.**

**Метою курсу «Аналіз і обробка великих даних»** є отримання теоретичних знань та практичних умінь і навиків, достатніх для успішного виконання професійних обов'язків в області комп'ютерних наук та підготувати студентів для подальшого працевлаштування в області видобутку й інтелектуального аналізу великих даних різної природи в розподілених інформаційних системах.

Курс направлений на вивчення і набуття уміння в сфері інженерії даних та знань в технологічних процесах розроблення і підтримки комп'ютерних систем, які базуються на інтелектуальному аналізі великих даних.

В процесі навчання студенти набувають системних практичних навиків як в технологічній області, так і в області науково-дослідницької діяльності.

В результаті навчання студенти отримують сучасні знання, які необхідні для провадження інженерної та наукової діяльності в області видобутку знань у наборах великих даних, їх інтелектуальної обробки і створення прикладних інформаційних продуктів.

#### **Результати навчання:**

1. Використовувати інструментальні засоби інтеграції різнотипових даних у наборах різної розмірності.
2. Здатність до професійної роботи з даними: видобуток (у тому числі повторний), фільтрація, інтеграція, зберігання, перевірка актуальності, валідація і репрезентивність.
3. Видобувати знання шляхом інтеграції та аналізу великих даних, отриманих з різноманітних та різнорідних джерела інформації.
4. Проектувати сховища великих даних, для видобутку даних і знань.
5. Проектувати та розробляти прикладні інформаційні продукти для видобутку корисних знань і вироблення оптимальних стратегій прийняття рішень.
6. Аналізувати науково-технічну інформацію при проведенні науково-дослідних робіт.

#### **4 Загальна інформація про дисципліну**

<b>Ступінь вищої освіти</b>	<b>Третій (освітньо-науковий)</b>
<b>Спеціальність</b>	<b>121 Інженерія програмного забезпечення</b>
<b>Курс (рік навчання)</b>	<b>1</b>
<b>Семестр</b>	<b>1</b>
<b>Нормативна \ вибіркова</b>	<b>Вибіркова</b>
<b>Загальна кількість год. / кредитів</b>	<b>150/5</b>
<b>Аудиторні, год.</b>	<b>30</b>
<b>Самостійна робота, год.</b>	<b>120</b>

#### **5. Перелік тем**

1. Вступ в аналіз і обробку великих даних.
2. Життєвий цикл проекту по аналітиці великих даних.
3. Основні техніки (підходи) щодо роботи з великими даними.
4. Когнітивний аналіз даних.
5. Аналіз та візуалізація великих даних.
6. Основні технології та інструменти роботи з великими даними.
7. Наукові проблеми у сфері великих даних.

#### **6. Рекомендовані джерела інформації**

1. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб.: Питер, 2017. – 336 с.
2. Дэви С. Основы Data Science и Big Data. Python и наука о данных // С. Дэви, М. Арно, А. Мохамед. – СПб.: Питер, 2017. – 336 с.: ил.
3. Плас Дж.В. Python для сложных задач. Наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.
4. Свейгарт Э. Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих. – М.: Вильямс, 2017. – 573 с.

5. Гобарева Я.Л., Городецкая О.Ю., Золотарюк А.В. Бизнес-аналитика средствами Excel М.: Вузовский учебник, ИНФРА-М, 2013. – 336 с.
6. Дэвенпорт Том, Хо Ким Джин. О чем говорят цифры. Как понимать и использовать данные Манн, Иванов и Фербер, 2014.
7. Маккинни У. Python и анализ данных М.: ДМК Пресс, 2015. – 482 с.
8. Соммервилл И. Инженерия программного обеспечения, 6-е издание: Пер. с англ. – М.: Издательский дом "Вильямс", 2002. – 624 с.
9. Фрэнк Билл. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики М.: Альпина Паблишер, 2014. – 430 с.
10. Фрэнкс Билл. Укрощение больших данных М.: Манн, Иванов и Фербер, 2014. – 352 с.
11. Шаховська Н. Б., Болюбаш Ю. Я. Модель великих даних “сутність-характеристика”. Режим доступу: [http://ena.lp.edu.ua:8080/bitstream/ntb/29775/1/20\\_186-196.pdf](http://ena.lp.edu.ua:8080/bitstream/ntb/29775/1/20_186-196.pdf)

### 7. Система оцінювання та вимоги.

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Аналіз і обробка великих даних» визначається за шкалою оцінювання:

За шкалою ТНЕУ	За національною шкалою	За шкалою ECTS
90–100	Відмінно	A (відмінно)
85–89	Добре	B (дуже добре)
75-84		C (добре)
65-74	Задовільно	D (задовільно)
60-64		E (достатньо)
35-59	Незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)

### 8. Навчальні ресурси

№	Найменування
1.	<b>Обладнання:</b> Проектор, комп'ютери з доступом до мережі Інтернету.
2.	<b>Програмне забезпечення:</b> Прикладні інструментальні засоби – MATLAB, R, SciPy. Засоби візуалізації даних – Advizor Analyst, R (ggplot2), Weave, Visalix, Databionic ESOM Tools, KNIME, Lattice Miner. Фреймворк організації розподілених обчислень – Hadoop MapReduce, Hadoop, Spark, CLUTO. Фреймворки для web-аналітики – Google Analytics, AMI Enterprise Intelligence, Automation Anywhere, iWebScraping, Web Data Extraction Services. Хмарні технології Microsoft Azure – Storage – Distributed cloud computing (Hadoop), Amazon Elastic MapReduce (EMR), Google Cloud Platform Засоби візуалізації – ggplot (R), Weka. Прикладні пакети для вирішення практичних задач на основі глибокого машинного навчання – R та Caffè.

## **9. Політики курсу.**

**Академічна доброчесність. Дотримання академічної доброчесності студентами передбачає:**

- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);
- посилення на джерела інформації у разі використання ідей, розробок, тверджень, відомостей;
- дотримання норм законодавства про авторське право і суміжні права; - надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використані методики досліджень і джерела інформації.

**Порушенням академічної доброчесності вважається:**

**академічний плагіат** - оприлюднення (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства;

**самоплагіат** - оприлюднення (частково або повністю) власних раніше опублікованих наукових результатів як нових наукових результатів;

**фабрикація** - вигадкування даних чи фактів, що використовуються в освітньому процесі або наукових дослідженнях;

**фальсифікація** - свідомо зміна чи модифікація вже наявних даних, що стосуються освітнього процесу чи наукових досліджень;

**списування** - виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання, зокрема під час оцінювання результатів навчання.

**За порушення академічної доброчесності здобувачі освіти можуть бути притягнені до такої академічної відповідальності:**

- повторне проходження оцінювання (контрольна робота, іспит, залік тощо);
- повторне проходження відповідного освітнього компонента освітньої програми.

**Політика запізнення.** За несвоєчасно виконані завдання буде накладено штраф 10 відсотків від загальної кількості балів за це завдання. Примітка. Виключення можуть бути зроблені до невчасно зданих завдань з поважних причин.

**Політика щодо відвідування:** Відвідування занять є обов'язковим. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбутись в он-лайн формі за погодженням із керівником курсу.