

**СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**  
**«ПРЕДИКАТИВНА АНАЛІТИКА ВЕЛИКИХ ДАНИХ»**



<b>Ступінь освіти</b>	Третій (освітньо-науковий)
<b>Галузь знань</b>	12 Інформаційні технології
<b>Тривалість викладання</b>	4 семестр
<b>Заняття:</b>	7 чверть
Лекції	3 години на тиждень
Практичні	2 години на тиждень
<b>Мова викладання</b>	українська

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/course/view.php?id=5304>

Кафедра, що викладає: Інформаційних технологій та комп'ютерної інженерії

**Інформація про викладача:**



<b>Викладач:</b>	Каштан В.Ю., доцент, к.т.н.
<b>Персональна сторінка:</b>	<a href="https://it.nmu.org.ua/ua/HR_staff/prepods/kashtan.php">https://it.nmu.org.ua/ua/HR_staff/prepods/kashtan.php</a>
<b>E-mail:</b>	Kashtan.V.Yu@nmu.one

## 1. Анотація до курсу

Навчальна дисципліна «**Предикативна аналітика великих даних**» відіграє важливу роль в ІТ-галузі, так як в даний час обробка великих обсягів даних є актуальним завданням для більшості підприємств та організацій. Розуміння методів та алгоритмів обробки даних, включаючи граничні, туманні, хмарні та паралельні обчислення, є важливим для розробки комп'ютерних систем та кіберфізичних систем, таких як інтернет-речей, мережі датчиків, автономні транспортні засоби та інші. Такі системи збирають великі обсяги даних, які потребують обробки та аналізу, щоб забезпечити ефективну та надійну роботу системи. Отже, вивчення предикативної аналітики великих даних допоможе аспірантам розвинути навички, необхідні для роботи в цих сферах.

Цей курс охоплює вивчення таких тем як: методи предикативної аналітики великих даних, граничних обчислень; інструменти та технології для предикативної аналітики великих даних; хмарні та паралельні обчислення для великих даних.

Після закінчення курсу «Обчислювальна техніка та програмування» аспіранти зможуть виконувати такі завдання:

- збирати дані з різних джерел та підготовлювати їх для подальшого аналізу;
- обробляти великі обсяги даних за допомогою спеціальних інструментів;
- розробляти та використовувати алгоритми машинного навчання для розв'язання різних задач;
- виконувати прогнозування на основі аналізу даних;
- візуалізувати результати аналізу даних та представляти їх у зручному для сприйняття вигляді;
- застосовувати знання з обробки великих даних для розв'язання різних завдань у ІТ галузі.

## 2. Мета та завдання навчальної дисципліни

**Мета дисципліни** – формування у здобувачів вищої освіти умінь та компетентностей щодо дослідження алгоритмів та методів обробки великих масивів даних, хмарних, паралельних обчислень для застосування у наукової діяльності.

### **Завдання курсу:**

- опанування теоретичних та практичних основ методології великих даних;
- аналізування алгоритмів для обробки великих даних;
- опанування сучасних фреймворків для обробки великих даних;
- опанування методів паралельних та хмарних обчислень та їх використання для обробки Big Data;
- реалізування алгоритмів обробки великих масивів даних для дослідження характеристик і поведінки складних об'єктів і комп'ютерних систем;
- опанування підходів для розв'язання завдань та прикладних задач збору, обробки, аналізу та візуалізації великих обсягів даних на сучасних мовах програмування.

## 3. Результати навчання

1. Здатність вирішувати загальні та складні спеціалізовані завдання та практичні проблеми для ефективної роботи з сучасними комп'ютерними засобами.

2. Здатність інтегрувати знання та вирішувати складні питання, формулювати судження щодо недостатньої або обмеженої інформації.

3. Знати принципи паралельних та розподілених обчислень та їх використання для ефективної обробки великих обсягів даних.

4. Вміти використовувати сучасні інструменти та технології для роботи з великими обсягами даних

5. Здійснювати ефективну комунікацію з колегами, кінцевими користувачами, керівництвом, фахівцями та експертами різного рівня інших галузей знань, аргументовано висловлювати свої думки щодо поточних чи майбутніх завдань у професійній сфері

6. Обґрунтовувати вибір програмних продуктів, їх взаємодію та потенційний вплив на вирішення організаційних проблем, здійснювати їх впровадження та використання з дотриманням професійних і етичних стандартів.

## ЛЕКЦІЇ

Лекція 1	<b>Тема 1. Вступ до предикативної аналітики великих даних та її застосування в ІТ-галузі.</b> Визначення термінів "предикативна аналітика" та "великі дані". Огляд технологічного стеку для роботи з великими даними. Приклади застосування предикативної аналітики в ІТ галузі.
Лекція 2	<b>Тема 2. Методи предикативної аналітики великих даних.</b> Визначення поняття "предикативна аналітика великих даних. Огляд основних методів предикативної аналітики великих даних. Застосування предикативної аналітики великих даних у сфері ІТ.
Лекція 3	<b>Тема 3. Методи граничних обчислень.</b> Огляд методів роботи з обмеженими ресурсами в системах з обробки великих даних. Технології оптимізації обчислень та зменшення обсягу даних, що обробляються. Використання граничних обчислень для відображення різних сценаріїв роботи. Вибір та оптимізація моделей машинного навчання. Оцінка точності моделей та уникнення перенавчання.
Лекція 4	<b>Тема 4. Інструменти та технології для предикативної аналітики великих даних.</b> Огляд популярних інструментів та платформ для роботи з великими даними. Використання сучасних мов програмування для предикативної аналітики. Робота з базами даних та системами управління даними для предикативної аналітики. Розробка та використання власних інструментів для предикативної аналітики.
Лекція 5	<b>Тема 5. Хмарні та паралельні обчислення для великих даних.</b> Моделі та архітектури хмарних обчислень. Огляд фреймворку Hadoop та його компонентів. Огляд Apache Storm та Apache Flink. Поточкові обчислення в режимі реального часу на платформі Apache Storm. Масштабованість та паралельне виконання обчислень на платформі Apache Spark. Алгоритми та практичні приклади паралельних обчислень в фреймворках Hadoop MapReduce Порівняння Hadoop, Kafka та Storm з іншими інструментами обробки великих даних. Використання Hadoop, Kafka та Storm для реалізації задач в області Big Data.

## ПРАКТИЧНІ ЗАНЯТТЯ

Практична робота 1	Дослідження методів Data Science та Big Data для комп'ютерних систем.
Практична робота 2	Використання методів машинного навчання для оптимізації роботи кіберфізичних систем.
Практична робота 3	Робота з інформаційними технологіями обробки великих даних у хмарних середовищах.
Практична робота 4	Дослідження алгоритмів обробки великих даних в фреймворку Hadoop.
Практична робота 5	Дослідження алгоритмів обробки великих даних в фреймворку MapReduce.

### 5. Технічне обладнання та/або програмне забезпечення

1. Персональний комп'ютер або ноутбук зі сталим доступом до мережі Інтернет
2. Активованій акаунт університетської пошти (student.i.p.@nmu.one) на Офіс365.
3. Активний обліковий запис у системі дистанційної освіти Moodle.
4. Програмне забезпечення:
  - Платформа Windows 10
  - HWINFO64;
  - LibreOffice 6.4;
  - MS Visual Studio Community 2019;
  - MS Visual Basic;
  - Python;
  - Інтернет-браузер;
  - Дистанційні платформи Moodle, MS Teams.

### 6. Система оцінювання та вимоги

**6.1. Навчальні досягнення** здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно / Excellent
74 – 89	добре / Good
60 – 73	задовільно / Satisfactory
0 – 59	незадовільно / Fail

Загальні критерії досягнення результатів навчання відповідають описам 8-го кваліфікаційного рівня НРК.

6.2. Здобувачі освітньо-наукового рівня PhD можуть отримати підсумкову оцінку з навчальної дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів з поточного тестування та виконання і захисту практичних робіт складатиме не менше 60 балів.

Максимальне оцінювання:

Теоретична частина	Практична частина		Разом
	При своєчасному складанні	При несвоєчасному складанні	
40	60	50	100

В рамках курсу передбачено виконання 5 практичних робіт. Під час захисту роботи аспірант відповідає на запитання стосовно ходу роботи, пояснює послідовність дій, демонструє результати роботи.

За результатами виконання роботи студенти складають звіт встановленого зразка, який завантажується до системи Moodle у відповідну категорію.

Звіт обов'язково має містити такі структурні компоненти:

- титульний лист;
- номер варіанту, текст завдання;
- скріншоти етапів виконання завдання, посилання на відповідні ресурси, коди програм тощо;
- звіт має бути завантажено у систему впродовж 3 днів після захисту роботи на занятті.

**Важливо!!!** Всі умови до лабораторних робіт з детальними поясненнями до них представлено на сторінці Moodle. Всі бали за лабораторні роботи фіксуються у журналі оцінок Moodle.

**6.3. Критерії оцінювання підсумкової роботи.** У випадку, якщо здобувач освітньо-наукового рівня PhD за поточною успішністю отримав менше 60 балів та/або прагне поліпшити оцінку проводиться **підсумкове оцінювання (диференційований залік)** під час сесії.

**Диференційований залік** проводиться у вигляді комплексної контрольної роботи, яка включає запитання з теоретичної та практичної частини курсу. Білет складається з **20 тестових завдань** з чотирма варіантами відповідей, одна правильна відповідь оцінюється в 3 бали (**разом 60 балів**) та **4 тестових завдань** з практичної частини, кожне з запитань оцінюється максимум у 10 балів (**разом 40 балів**), причому:

- 10 балів – відповідність еталону;
- 8 балів – відповідність еталону з незначними помилками;
- 5 балів – часткова відповідність еталону, питання розкриті не в повній мірі;
- 2 бали – невідповідність еталону, але відповідність темі запитання;
- 1 бал – фрагментарні результати у відповідності до теми запитання;
- 0 балів – відповідь не наведена або не відноситься до теми запитання.

Отримані бали за відкриті та закриті тести додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за підсумковою роботою здобувач вищої освіти може набрати 100 балів.

## 7. Політика курсу

**7.1. Політика щодо академічної доброчесності.** Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка" (<https://bit.ly/3ExtVKY>).

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

**7.2. Комунікаційна політика.** Здобувачі вищої освіти повинні мати активовану корпоративну університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

**7.3. Політика щодо перескладання.** Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання підсумкового оцінювання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

**7.4 Політика щодо оскарження оцінювання.** Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

**7.5. Відвідування занять.** Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

**7.6. Опитування.** Наприкінці вивчення курсу та перед початком сесії здобувача вищої освіти буде запропоновано анонімно заповнити електронні анкети (Microsoft Forms Office 365), які буде розіслано на ваші університетські поштові скриньки. Заповнення анкет є важливою складовою вашої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати ваші пропозиції стосовно покращення змісту навчальної дисципліни «Предикативна аналітика великих даних».

## 8. Рекомендовані джерела інформації

1. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly, 2019, 856 p.
2. Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016, 654 p.
3. Tyler Akidau, Slava Chernyak, Reuven Lax. Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing, O'Reilly Media, 2018, 352 p.
4. Sebastian Raschka, Vahid Mirjalili. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, Packt, 2019, 770 p.
5. Valliappa Lakshmanan. Data Science on the Google Cloud Platform: Implementing End-to-End Real-Time Data Pipelines: From Ingest to Machine Learning, O'Reilly Media, 2018, 332 p.
6. Каштан В.Ю. Дешифрування автодоріг на цифрових космічних знімках на основі нейронних мереж / В.Ю. Каштан, В.В. Гнатушенко, О.Г. Баглай// XVI міжнародна конференція «Проблеми використання інформаційних технологій в освіті, науці та промисловості» / Збірник наукових праць № 6. – м. Дніпро. – 15 грудня 2021, С.71 – 75.
7. Kashtan V.Yu. Deep Learning Technology for Automatic Burned Area Extraction Using Satellite High Spatial Resolution Images / V. Yu. Kashtan, V. V.Hnatushenko // // Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2022. Advances in Intelligent Systems and Computing, vol 1246. Pp. 55-76. Springer, Cham.
8. Kashtan, V., Hnatushenko, V. (2023). Deep Learning Technology for Automatic Burned Area Extraction Using Satellite High Spatial Resolution Images. In: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. IDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 149, 664-685p. Springer, Cham. [https://doi.org/10.1007/978-3-031-16203-9\\_37](https://doi.org/10.1007/978-3-031-16203-9_37)
9. Data Mining and Predictive Analytics. Dean Abbott. Електронний ресурс. – Режим доступу: <http://abbottanalytics.blogspot.com/2014/07/predictive-analytics-and-business.html?m=1>.
10. Predictive Analytics. Електронний ресурс. – Режим доступу: <https://www.gartner.com/itglossary/predictive-analytics>.
11. Kashtan V.Yu., Hnatushenko V.V. Machine learning for automatic extraction of water bodies using Sentinel-2 imagery. The scientific journal «Radio Electronics, Computer Science, Control». – Zaporizhzhia, 2024. – №1 (68). – P.118-127. DOI: <https://doi.org/10.15588/1607-3274-2024-1-11>.
12. Radionov Y. D., Kashtan V. Yu., Hnatushenko V. V., Kazymyrenko O. V. Aircraft detection with deep neural networks and contour-based methods. The scientific journal «Radio Electronics, Computer Science, Control». – Zaporizhzhia, 2024. №1 4(71). – P.121-129.

## **9 Інформаційні ресурси**

Репозиторій Національного технічного університету «Дніпровська політехніка»:  
<http://ir.nmu.org.ua/>