

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І  
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Кафедра інформаційних систем

**“ЗАТВЕРДЖУЮ”**

Декан факультету  
інформаційних технологій  
\_\_\_\_\_ доц. Глазунова О.Г.  
“ \_\_\_\_ ” \_\_\_\_\_ 2019 р.

**РОЗГЛЯНУТО І СХВАЛЕНО**

на засіданні кафедри  
інформаційних систем  
Протокол № 9 від “15” квітня 2019 р.  
Завідувач кафедри  
\_\_\_\_\_ Швиденко М.З.

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
«АНАЛІТИКА ВЕЛИКИХ ДАНИХ»**

галузь знань \_\_\_\_\_ 05 «Соціальні та поведінкові науки»  
спеціальність \_\_\_\_\_ 051 «Економіка»  
освітня програма \_\_\_\_\_ «Економічна кібернетика»  
факультет \_\_\_\_\_ інформаційних технологій  
розробники: \_\_\_\_\_ доц., к.е.н. Харченко В.В.

(посада, науковий ступінь, вчене звання)

Київ – 2019 р.

# 1. Опис навчальної дисципліни «Аналітика великих даних» (назва)

<b>Галузь знань, напрям підготовки, спеціальність, освітній ступінь</b>		
Освітній ступінь	<i>Магістр</i>	
Галузь знань	<i>05 «Соціальні та поведінкові науки»</i>	
Спеціальність	<i>051 «Економіка»</i>	
Освітня програма	<i>«Економічна кібернетика»</i>	
<b>Характеристика навчальної дисципліни</b>		
Вид	<i>Обов'язкова</i>	
Загальна кількість годин	<i>90 год.</i>	
Кількість кредитів ECTS	<i>3 ECTS</i>	
Кількість змістових модулів	<i>2</i>	
Курсовий проект (робота) (за наявності)	<i>–</i>	
Форма контролю	<i>Іспит</i>	
<b>Показники навчальної дисципліни для денної та заочної форм навчання</b>		
	денна форма навчання	заочна форма навчання
Рік підготовки (курс)	<i>1</i>	-
Семестр	<i>2</i>	-
Лекційні заняття	<i>15 год.</i>	-
Практичні, семінарські заняття	-	-
Лабораторні заняття	<i>30 год.</i>	-
Самостійна робота	<i>45 год.</i>	-
Індивідуальні завдання	-	-
Кількість тижневих аудиторних годин для денної форми навчання	<i>3 год.</i>	

## 2. Мета та завдання навчальної дисципліни

**Мета** сформувати у студентів фундаментальні знання з теорії та практики в області розробки і використання систем обробки та аналізу великих масивів даних.

**Завдання** вивчення теоретико-методичних засад та основних технологій щодо вирішення завдань обробки великих за обсягом, швидко змінюваних та погано структурованих даних, що об'єднуються терміном «великі дані».

У результаті вивчення навчальної дисципліни студент повинен

**знати:** основні поняття аналітики великих даних; основні технології, що застосовуються для зберігання і пошуку в великих даних.

**вміти:** застосовувати методи аналізу великих даних, вміти реалізовувати програми для аналітики великих даних.

### **3. Програма та структура навчальної дисципліни для повного терміну денної форми навчання:**

#### **Змістовий модуль 1. Введення в аналітику великих даних**

##### **Тема 1 Введення в аналітику великих даних**

Основні поняття та визначення. Історія розвитку. Джерела великих даних. Застосування в економіці, бізнесі, сільському господарстві, промисловості. Приклади використання. Великі дані в наукових сферах. Особливості застосування. Вимоги до професії аналітика великих даних.

##### **Тема 2 Життєвий цикл проекту по аналітиці великих даних**

Основні етапи життєвого циклу. Збір, консолідація і очищення даних. Побудова моделей, роль машинного навчання.

##### **Тема 3 Основні техніки (підходи) щодо роботи з великими даними**

Збір та консолідація даних, «аналітична пісочниця» (analytic sandbox) «озеро даних» (data lake), пакетна аналітика (batch oriented), аналітика реального часу (real time oriented), гібридна аналітика (hybrid), робота з СУБД.

##### **Тема 4 Введення в когнітивний аналіз даних**

Когнітивна система типу IBM Watson. Функції та можливості системи IBM Watson.

#### **Змістовий модуль 2. Технології та інструменти роботи з великими даними**

##### **Тема 5 Аналіз та візуалізація великих даних**

Візуалізація великих даних («big data visualization»), візуалізація текстів, візуалізація кластерів, візуалізація асоціацій, ландшафтна візуалізація, візуалізація гіпотез, візуалізація дерев рішень, багатовимірна візуалізація. Сіткова візуалізація. Класифікація. Gephi.

##### **Тема 6 Основні технології та інструменти роботи з великими даними**

Підхід Map/Reduce та його програмна реалізація, Apache Hadoop, HDFS, HBase, YARN, Hive, Pig, Storm як система потокової обробки, мова програмування Python, R, Apache Spark. IBM Bluemix. Microsoft HDInsight.

##### **Тема 7 Наукові проблеми у сфері великих даних. Перспективи розвитку великих даних**

Оцінка «великих даних»: проблеми та вирішення

## Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин													
	денна форма							Заочна форма						
	тижні	усього	у тому числі					усього	у тому числі					
			л	п	ла б	ін д	с.р.		л	п	ла б	ін д	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
<b>Змістовий модуль 1. Введення в аналітику великих даних</b>														
Тема 1. Введення в аналітику великих даних	1,2	11	2		4		5							
Тема 2. Життєвий цикл проекту по аналітиці великих даних	3,4	11	2		4		5							
Тема 3. Основні техніки роботи з великими даними	5,6	11	2		4		5							
Тема 4. Введення в когнітивний аналіз даних	7,8	13	2		6		5							
Разом за змістовим модулем 1	46		8		18		20							
<b>Змістовий модуль 2. Технології та інструменти роботи з великими даними</b>														
Тема 5. Аналіз та візуалізація великих даних	9,10	16	2		4		10							
Тема 6. Основні технології та інструменти роботи з великими даними	11,13	16	2		4		10							
Тема 7. Наукові проблеми у сфері великих даних. Перспективи розвитку великих даних	14-15	12	3		4		5							
Разом за змістовим модулем 2	44		7		12		25							
Усього годин	<b>90</b>		<b>15</b>		<b>30</b>		<b>45</b>							
Курсовий проект (робота)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Усього годин	<b>90</b>		<b>15</b>		<b>30</b>		<b>45</b>							

#### 4. Теми семінарських занять

Не передбачені навчальним планом.

#### 5. Теми практичних занять

Не передбачені навчальним планом.

## 6. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Готові рішення аналізу даних (Rapid Miner, Weka), мови Python та R, стек бібліотек аналізу даних.	4
2	Візуалізація даних. Gephi.	4
3	Хмарний сервіс Big Data IBM Bluemix, Azure HD Insights.	4
4	Налаштування кластеру Hadoop for Analytics.	6
5	Файлова система HDFS, Object Storage.	4
6	Модель Map/Reduce, Apache Spark.	4
7	NewSQL платформа SAP HANA, Oracle Exalytics.	4
8	Всього	30

## 7. Контрольні питання, комплекти тестів для визначення рівня засвоєння знань студентами

1. Основні характеристики великих даних
2. Роль великих даних в сільському господарстві
3. Консолідація даних
4. Візуалізація даних, Gephi.
5. Основні конструкції мови R, консолідація даних, візуалізація
6. HDFS – основи організації
7. Архітектура Hadoop
8. Виконання Map/Reduce
9. Виконання програм в Hadoop
10. Основи YARN
11. Аналітика потокових даних в платформі Storm
12. Архітектура Apache Spark
13. Організація даних в Apache Spark
14. Обробка даних в GraphX
15. Алгоритми класифікації
16. Алгоритми кластеризації
17. Нейронні мережі як реалізація алгоритмів машинного навчання
18. Інтелектуальні алгоритми
19. Застосування технологій великих даних для задач управління в реальному часі

## 8. Методи навчання.

Засвоєння матеріалу забезпечується на лекціях, лабораторних заняттях та самостійній роботі у комп'ютерних класах, обладнаних локальними мережами, Інтернет та програмним забезпеченням. Лекції супроводжуються використанням презентацій та мультимедійного обладнання для полегшення засвоєння матеріалу.

## 9. **Форми контролю.**

Контроль знань у слухачів магістерського курсу «Аналітика великих даних» передбачає такі контрольні заходи:

- самоконтроль – є первинною формою контролю знань, який обов'язково забезпечується дистанційним курсом шляхом надання студентам переліку запитань (питань та відповідей на них);
- поточний контроль – здійснюється через систему оцінки безпосередньо викладачем лабораторно-практичних практичних занять та виконаних завдань для самостійної роботи;
- модульний контроль – здійснюється дистанційно в автоматизованому режимі або очному режимі, основною формою якого є тестування;
- підсумковий контроль – це іспит, який складається очно в період призначений деканатом або за індивідуальним графіком, який затверджується навчальним планом. Основною формою підсумкового контролю є тестування.

**10. Розподіл балів, які отримують студенти.** Оцінювання студента відбувається згідно положенням «Про екзамени та заліки у НУБіП України» від 27.02.2019 р. протокол № 7 з табл. 1.

**Таблиця 1. Співвідношення між національними оцінками і рейтингом здобувача вищої освіти**

<b>Оцінка національна</b>	<b>Рейтинг здобувача вищої освіти, бали</b>
<b>Відмінно</b>	<b>90 – 100</b>
<b>Добре</b>	<b>74 – 89</b>
<b>Задовільно</b>	<b>60 – 73</b>
<b>Незадовільно</b>	<b>0 – 59</b>

Для визначення рейтингу студента (слухача) із засвоєння дисципліни  $R_{\text{дис}}$  (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи  $R_{\text{нр}}$  (до 70 балів):  
 $R_{\text{дис}} = R_{\text{нр}} + R_{\text{ат}}$ .

## 11. Методичне забезпечення

Електронна система дистанційного навчання MOODLE, адреса moodle.nauu.kiev.ua

## 12. Рекомендована література

### – основна:

1. Абдикеев Н.М. Когнитивная бизнес-аналитика Учебник. – М.: ИНФРА-М, 2014. – 511 с.
2. Гобарева Я.Л., Городецкая О.Ю., Золотарюк А.В. Бизнес-аналитика средствами Excel М.: Вузовский учебник, ИНФРА-М, 2013. – 336 с.
3. Дэвенпорт Том, Хо Ким Джин. О чем говорят цифры. Как понимать и использовать данные Манн, Иванов и Фербер, 2014.
4. Кулешова О.В. Microsoft Excel 2010. Уровень 2. Расширенные возможности М.: Центр компьютерного обучения «Специалист» при МГТУ им. Н.Э. Баумана, 2012. – 91 с.
5. Маккинни У. Python и анализ данных М.: ДМК Пресс, 2015. – 482 с.
6. Осетрова И.С., Осипов Н.А. Microsoft Excel 2010 для аналитиков Учебное пособие. – СПб.: НИУ ИТМО, – 2013. – 65 с.
7. Ситник В.Ф. та ін. Системи підтримки прийняття рішень – К.: Техніка, 1995-162 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)
12. National Research Council. 2013. Frontiers in Massive Data Analysis. Washington, D.C.: The National Academies Press
13. Big Data analytics: Future architectures, Skills and roadmaps for the CIO – 2011. – IDC/SAS

### – допоміжна:

1. Big Data Visualization: Turning Big Data into Big Insights. The Rise of Visualization-based Data Discovery Tools. White Paper. Intel IT Center. March 2013
2. Big Data: The Next Frontier for Innovation, Competition, and Productivity – 2011. – McKinsey Global Institute
3. Martin Hilbert. Big Data for Development: From Information- to Knowledge Societies", – 2013. – SSRN Scholarly Paper No. ID 2205145). Rochester, NY: Social Science Research Network
4. DJ Patil. Building Data Science Teams. O'Reilly. 2011. ISBN: 978-1-449-31623-5 <http://cdn.oreilly.com/radar/2011/09/Building-Data-Science-Teams.pdf>

### 13. Інформаційні ресурси

1. IBM Analytics <http://www.ibm.com/analytics/us/en/technology/hadoop/hadoop-trials.html>
2. IBM Cloud [https://www.ibm.com/cloud-computing/bluemix/?lnk=hp\\_trials\\_uauk](https://www.ibm.com/cloud-computing/bluemix/?lnk=hp_trials_uauk)
3. IBM Bluemix Promo Code - 6 Month Trial  
<https://ibm.onthehub.com/WebStore/OfferingDetails.aspx?o=bb3528b7-2b63-e611-9420-b8ca3a5db7a1>
4. Hadoop: Built for big data, insights, and innovation  
<http://www.ibm.com/analytics/us/en/technology/hadoop/>
5. IBM BigInsights <http://www.ibm.com/analytics/us/en/technology/biginsights/>
6. Виктор Маер-Шенбергер, Кеннет Кукьер. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим. – М.: «Манн, Иванов и Фербер», 2013, 240 с. ISBN 978-5-91657-936-9 [http://www.mann-ivanov-ferber.ru/books/paperbook/big\\_data/](http://www.mann-ivanov-ferber.ru/books/paperbook/big_data/)
7. Weka Machine learning software to solve data mining problems  
[https://sourceforge.net/projects/weka/?source=typ\\_redirect](https://sourceforge.net/projects/weka/?source=typ_redirect)
8. Books Ngram Viewer <https://books.google.com/ngrams>
9. Революция Big Data: Как извлечь необходимую информацию из «Больших Данных»? <http://statsoft.ru/products/Enterprise/big-data.php>
10. Бесплатные программы для статистического анализа данных  
<http://boris.bikbov.ru/2013/12/01/besplatnyie-programmyi-dlya-statisticheskogo-analiza-dannyih/>
11. <https://www.r-bloggers.com/>
12. Мова програмування R [Електронний ресурс]: <https://cran.r-project.org>
13. <http://r-analytics.blogspot.ru/p/rstudio.html#.WDifOrnzuwI>
14. Середовище для розробки програм на R – R Studio [Електронний ресурс]:  
<http://www.r-studio.com>
15. [http://www.tadviser.ru/index.php/Статья:Большие\\_данные\\_\(Big\\_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data))