

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Кафедра програмної інженерії та інформаційних технологій управління  
(назва)

**«ЗАТВЕРДЖУЮ»**

Голова науково-методичної комісії \_\_\_\_\_  
(назва комісії)

\_\_\_\_\_  
(підпис) (ініціали та прізвище)

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_\_ року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

**Машинне навчання BigData**

(назва навчальної дисципліни)

рівень вищої освіти другий (магістерський)  
перший (бакалаврський) / другий (магістерський)

галузь знань 12 Інформаційні технології  
(шифр і назва)

спеціальність 126 Інформаційні системи та технології  
(шифр і назва)

вид дисципліни професійна підготовка (вибіркова)  
(загальна підготовка / професійна підготовка)

форма навчання денна  
(денна / заочна)

Харків – 2018 рік

## ЛИСТ ЗАТВЕРДЖЕННЯ

Робоча програма з навчальної дисципліни

Машинне навчання BigData

(назва дисципліни)

Розробники:

доцент, к.т.н., доц.

(посада, науковий ступінь та вчене звання)

(підпис)

Шматко О.В.

(ініціали та прізвище)

(посада, науковий ступінь та вчене звання)

(підпис)

(ініціали та прізвище)

Робоча програма розглянута та затверджена на засіданні кафедри

програмної інженерії та інформаційних технологій управління

(назва кафедри)

Протокол від « 31 » \_\_\_\_\_ 2018 року № 1

Завідувач кафедри \_\_\_\_\_

(назва кафедри)

(підпис)

Годлевський М.Д.

(ініціали та прізвище)

## ЛИСТ ПЕРЕЗАТВЕРДЖЕННЯ РОБОЧОЇ НАВЧАЛЬНОЇ ПРОГРАМИ

Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри	Підпис голови НМК (для дисциплін загальної підготовки та дисциплін професійної підготовки за спеціальністю) або завідувача випускової кафедри (для дисциплін професійної підготовки зі спеціалізації, якщо РПНД розроблена не випусковою кафедрою)

## МЕТА, КОМПЕТЕНТНОСТІ, РЕЗУЛЬТАТИ НАВЧАННЯ ТА СТРУКТУРНО-ЛОГІЧНА СХЕМА ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**Мета** навчальної дисципліни “Машинне навчання BigData” є — підготовка здобувачів вищої освіти, яка необхідна для успішного освоєння сучасних методів машинного навчання. Отримані в результаті освоєння дисципліни знання необхідні при вирішенні практичних задач в сфері професійної діяльності, проектуванні та розробці інформаційних систем. Ознайомлення студентів з теоретичними основами і основними принципами машинного навчання – а саме, з класами моделей (лінійні, логічні), метриками якості та підходами до підготовки даних. Формування у студентів практичних навичок роботи з даними і вирішення прикладних задач аналізу даних.

### **Компетентності:**

#### *Загальні компетентності:*

- Здатність застосовувати знання у практичних ситуаціях.
- Знання та розуміння предметної області та розуміння професійної діяльності.
- Здатність вчитися і оволодівати сучасними знаннями.
- Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

#### *Спеціальні (фахові) компетентності:*

- здатність проведення збору, обробки і аналізу даних з використанням існуючих методів машинного навчання;
- здатність розробляти і реалізовувати в виді програмного модуля алгоритм вирішення поставленої теоретичної або прикладної задачі на основі математичної моделі.

### **Нормативний зміст підготовки здобувачів вищої освіти, сформульований у термінах результатів навчання**

<b>Класифікація компетентностей за НРК</b>	<b>Знання</b>	<b>Уміння</b>	<b>Комунікація</b>	<b>Автономія та відповідальність</b>
Вміння розкрити наукову суть проблем у професійній галузі Здатність вирішувати проблеми в професійній галузі на основі аналізу та синтезу	Знання наукових і методологічних основ створення та застосування інформаційних технологій та інформаційних систем для автоматизованої переробки інформації та управління	Уміння застосовувати теоретичні знання у практичних ситуаціях у науковій діяльності.	Спілкування в діалоговому режимі з широкою науковою спільнотою та громадськістю в певній галузі наукової та/або професійної діяльності	Ініціювання інноваційних комплексних проектів, лідерство та повна автономність під час їх реалізації Соціальна відповідальність за результати прийняття стратегічних рішень. Здатність саморозвиватися і

Класифікація компетентностей за НРК	Знання	Уміння	Комунікація	Автономія та відповідальність
				самовдосконалюватися протягом життя, відповідальність за навчання інших

### Структурно-логічна схема вивчення навчальної дисципліни

Попередні дисципліни:	Наступні дисципліни:
Бази даних та сховища даних	Преддипломна практика
	Дипломне проектування

### ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

(розподіл навчального часу за семестрами та видами навчальних занять)

Семестр	Загальний обсяг (годин) / кредитів ECTS	з них		За видами аудиторних занять (годин)			Індивідуальні завдання студентів (КП, КР, РГ, Р, РЕ)	Поточний контроль	Семестровий контроль	
		Аудиторні заняття (годин)	Самостійна робота (годин)	Лекції	Лабораторні заняття	Практичні заняття, семінари			Залік	Екзамен
1	2	3	4	5	6	7	8	9	10	11
<b>12</b>	<b>90/3</b>	<b>40</b>	<b>50</b>	<b>0</b>	<b>0</b>	<b>40</b>		<b>1</b>		<b>Е</b>

Співвідношення кількості годин аудиторних занять до загального обсягу складає 44% (%):

## СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

№ з/п.	Види навчальних занять (Л, ЛЗ, ПЗ, СР)	Кількість годин	Номер семестру (якщо дисципліна викладається у декількох семестрах). Назви змістових модулів. Найменування тем та питань кожного заняття. Завдання на самостійну роботу.	Рекомендована література (базова, допоміжна)
1	2	3	4	5
<b>Змістовий модуль 1. Вступ до машинного навчання</b>				
<b>1</b>	ПЗ	8	<b>Тема 1. Вступ до машинного навчання</b> Вступ до дисципліни. Історія аналізу даних. Постановки задач в машинному навчанні: класифікація, регресія, ранжування, кластеризація, латентні моделі. Приклади задач. Види даних: структуровані таблиці, тексти, зображення, звук. Ознаки.	<b>[1, 4, 6]</b>
	СР	10		
<b>2</b>	ПЗ	8	<b>Тема 2. Лінійні методи машинного навчання</b> Аналітичний і чисельний розв'язок задачі методу найменших квадратів. Градієнтний спуск, методи оцінки градієнту. Функції втрат. Регуляризація лінійної регресії	<b>[3-6]</b>
	СР	10		
<b>3</b>	ПЗ	8	<b>Тема 3. Особливі види регресії в машинному навчанні</b> Квантильна регресія (постановка задачі і приклади використання). Методи оцінювання узагальнюючої здатності, крос-валідація. Метрики якості регресії. Прогнозування часових рядів як задача регресії: авторегресія, тренди і сезонності. Оцінювання якості ковзним вікном	<b>[4-6]</b>
	СР	10		
<b>4</b>	ПЗ	8	<b>Тема 4. Лінійні методи класифікації</b> Апроксимація емпіричного ризику. Перцептрон. Метод опорних векторів, його двоїста задача (без ядер). Задача оцінювання ймовірностей, логістична регресія. Ідея калібрування ймовірностей. Оптимізація другого порядку (ідея і передумови для використання). Узагальнені лінійні моделі. Метрики якості в задачах класифікації.	<b>[2-4]</b>
	СР	10		
<b>5</b>	ПЗ	8	<b>Тема 5. Методи мультикласової класифікації</b> Multiclass- і multilabel-класифікація. Особливості багатокласових задач. Метрики якості. Методи вирішення multilabel-задач, що базуються на матричних розкладах	<b>[1-6]</b>
	СР	10		
Разом (годин)		90		

## САМОСТІЙНА РОБОТА

№ з/п	Назва видів самостійної роботи	Кількість годин
1	Опрацювання практичного матеріалу	10
2	Підготовка до практичних(лабораторних, семінарських) занять	15
3	Самостійне вивчення тем та питань, які не викладаються на практичних заняттях	10
4	Виконання індивідуального завдання:	10
5	Інші види самостійної роботи	5
	<b>Разом</b>	<b>50</b>

## МЕТОДИ НАВЧАННЯ

При викладанні навчальної дисципліни для активізації навчального процесу передбачено застосування сучасних навчальних технологій, таких, як: проблемні лекції; робота в малих групах; семінари-дискусії; кейс-метод; ділові ігри.

**Проблемні лекції** спрямовані на розвиток логічного мислення студентів. Коло питань теми лекції обмежується двома-трьома ключовими моментами, увага студентів концентрується на матеріалі, що не знайшов широкого відображення в підручниках, використовується досвід закордонних навчальних закладів з роздаванням студентам під час лекцій друкованого матеріалу та виділенням головних висновків з питань, що розглядаються. При викладанні лекційного матеріалу студентам пропонуються питання для самостійного розмірковування. При цьому лектор задає запитання, які спонукають студента шукати розв'язання проблемної ситуації. Така система примушує студентів сконцентруватися і почати активно мислити в пошуках правильної відповіді.

На початку проведення проблемної лекції необхідно чітко сформулювати проблему, яку необхідно вирішити студентам. При викладанні лекційного матеріалу слід уникати прямої відповіді на поставлені запитання, а висвітлювати лекційний матеріал таким чином, щоб отриману інформацію студент міг використовувати при розв'язанні проблеми.

**Міні-лекції** передбачають викладення навчального матеріалу за короткий проміжок часу й характеризуються значною ємністю, складністю логічних побудов, образів, доказів та узагальнень. Міні-лекції проводяться, як правило, як частина заняття-дослідження. На початку проведення міні-лекції за вказаними темами лектор акцентує увагу студентів на необхідності представити викладений лекційний матеріал у так званому структурно-логічному вигляді. На розгляд виносяться питання, які зафіксовані у плані лекцій, але викладаються вони стисло. Лекційне заняття, проведене у такий спосіб, пробуджує у студента активність та увагу при сприйнятті матеріалу, а також спрямовує його на використання системного підходу при відтворенні інформації, яку він одержав від викладача. Проблемні лекції та міні-лекції доцільно поєднувати з такою формою активізації навчального процесу, як робота в малих групах.

**Робота в малих групах** дає змогу структурувати лекційні або лабораторні заняття за формою і змістом, створює можливості для участі кожного студента в роботі за темою заняття, забезпечує формування особистісних якостей та досвіду соціального спілкування. Після висвітлення проблеми (при використанні проблемних лекцій) або стислого викладання матеріалу (при



використанні міні-лекцій) студентам пропонується об'єднуватися у групи по 5-6 осіб та презентувати наприкінці заняття своє бачення та сприйняття матеріалу.

**Презентації** – виступи перед аудиторією, що використовуються для представлення певних досягнень, результатів роботи групи, звіту про виконання індивідуальних завдань. Однією з позитивних рис презентації та її переваг при використанні в навчальному процесі є обмін досвідом, який здобули студенти при роботі у певній малій групі.

**Лабораторні заняття (з елементами семінарської дискусії)** дозволяють формувати у студентів навички особистого експериментального дослідження фізичних процесів що відбуваються під час роботи компонентів операційної системи, проводити аналіз умов її функціонування, а також розробляти нові елементи та системні компоненти відповідно до вимог, що пред'являються до них, узагальнювати отримані результати, формулювати висновки та думки, вести подальший обмін думками та поглядами з іншими учасниками щодо отриманих результатів досліджень з даної теми, а також розвивають творче мислення, допомагають формувати погляди і переконання, вчать об'єктивно оцінювати результати і пропозиції опонентів, критично підходити до власних результатів та поглядів.

**Ділові та рольові ігри** – форма активізації студентів, за якої вони задіяні в процесі інсценізації певної виробничої ситуації у ролі безпосередніх учасників подій.

**Кейс-метод** – метод аналізу конкретних ситуацій, який дає змогу наблизити процес навчання до реальної практичної діяльності спеціалістів і передбачає розгляд виробничих, управлінських та інших ситуацій, складних конфліктних випадків, проблемних ситуацій, інцидентів у процесі вивчення навчального матеріалу.

## МЕТОДИ КОНТРОЛЮ

Система оцінювання знань, вмінь та навичок студентів передбачає виставлення оцінок за усіма формами проведення занять. Перевірка та оцінювання знань студентів може проводитись у таких формах:

1. Оцінювання роботи студентів у процесі практичних занять.
2. Проведення проміжного контролю.
3. Проведення модульного контролю.

Загальна модульна оцінка складається з поточної оцінки, яку студент отримує під час практичних занять та оцінки за виконання модульної контрольної роботи.

Загальна оцінка з дисципліни визначається як середнє арифметичне модульних оцінок та оцінки яка отримана під час проведення заліку.

### **Порядок поточного оцінювання знань студентів**

Поточне оцінювання здійснюється під час проведення лабораторних занять і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- 1) активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни; відвідування занять;
- 2) виконання проміжного контролю;
- 3) виконання модульного контрольного завдання.

### **Контроль систематичного виконання самостійної роботи та активності на лабораторних заняттях**

Оцінювання проводиться за 5-бальною шкалою за такими критеріями:

- 1) розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- 2) ступінь засвоєння матеріалу дисципліни;
- 3) ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- 4) уміння поєднувати теорію з практикою при розгляді виробничих ситуацій, розв'язанні задач, проведенні розрахунків при виконанні завдань, винесених для самостійного опрацювання, та завдань, винесених на розгляд в аудиторії;
- 5) логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінка “відмінно” ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх п'яти зазначених критеріїв.

Відсутність тієї чи іншої складової знижує оцінку на відповідну кількість балів.

При оцінюванні практичних завдань увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

### **Проміжний модульний контроль**

Проміжний модульний контроль рівня знань передбачає виявлення опанування студентом матеріалу лекційного модуля та вміння застосовувати його для вирішення практичної ситуації і проводиться у вигляді контрольної роботи за темами 1-го або 2-го модулю.

### **Проведення модульного контролю**

Модульний контроль здійснюється та оцінюється за допомогою проведення контрольної роботи за всіма темами дисципліни.

**Підсумковий/семестровий контроль** проводиться у формі семестрового екзамену. Семестрові екзамени – форма оцінки підсумкового засвоєння студентами теоретичного та практичного матеріалу з окремої навчальної дисципліни, що проводиться як контрольний захід.

Підсумкова оцінка з дисципліни розраховується як середня з кількох складових, що враховує оцінки кожного виду контролю (дві оцінки за результатами поточного модульного контролю, оцінку за курсовий проект і оцінку за семестрову контрольну роботу).

## **РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ СТУДЕНТИ, ТА ШКАЛА ОЦІНЮВАННЯ ЗНАНЬ ТА УМІНЬ (НАЦІОНАЛЬНА ТА ECTS)**

**Таблиця 1 – Розподіл балів**

	Поточний контроль			Семестровий контроль	Всього за семестр
	КР	ПЗ	ІНДЗ		
Підсумкові бали	75			25	100
Макс. проміжні бали	12	5	26		
Кільк. од. обліку у семестрі	2	5	1		
Макс. проміжних балів, всього	24	25	26		100
Коеф. перерахунку	1				
Макс. кільк. підсумкових балів	24	25	26	25	100

**Таблиця 2 – Шкала оцінювання знань та умінь: національна та ЄКТС**

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
90 ... 100	A	відмінно
82 ... 89	B	добре
74 ... 81	C	
64 ... 73	D	задовільно
60 ... 63	E	
35 ... 59	FX	незадовільно з можливістю повторного складання
0 ... 34	F	незадовільно з обов'язковим повторним вивченням дисципліни

### **НАВЧАЛЬНО-МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

1. Комплект слайдів презентацій з матеріалами лекцій.
2. Методичні вказівки до практичних занять з дисципліни “Вступ до BigData” – Режим доступу: <https://trello.com/b/4w4jmR8S/introduction-to-bigdata>

### **РЕКОМЕНДОВАНА ЛІТЕРАТУРА**

#### *Базова*

1. Hastie T., Tibshirani R, Friedman J. The Elements of Statistical Learning (2nd edition). Springer, 2009.
2. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
3. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MIT Press, 2012.
4. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
5. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
6. Николенко С.И. Курс лекцій по машинному обучению — слайды. Електронний ресурс. Режим доступу: <http://logic.pdmi.ras.ru/sergei/index.php?page=mlaptu09>
7. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.

8. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.
9. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985.
10. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.