



Національний технічний університет України  
«КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»

Навчально-науковий  
фізико-технічний інститут  
Кафедра математичного  
моделювання та аналізу даних

# Інтелектуальний аналіз даних 30 6

## Робоча програма навчальної дисципліни (Силабус)

### Реквізити навчальної дисципліни

Рівень вищої освіти	Другий (магістерський)
Галузь знань	11 Математика і статистика
Спеціальність	113 Прикладна математика
Освітня програма	Математичні методи криптографічного захисту інформації
Статус дисципліни	Обов'язкова
Форма навчання	Очна (денна)
Рік підготовки, семестр	1 курс, осінній семестр
Обсяг дисципліни	Загальна кількість: 120 годин / 4 кредити Лекційних занять: 36 год Лабораторних занять: 18 год Самостійна робота студентів: 66 год
Семестровий контроль / контрольні заходи	Екзамен, модульна контрольна робота лабораторні роботи
Розклад занять	<a href="http://ipt.kpi.ua/navchalnij-protses">http://ipt.kpi.ua/navchalnij-protses</a>
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д.ф.-м.н., Сахненко Наталія Костянтинівна Лабораторні: д.ф.-м.н., Сахненко Наталія Костянтинівна
Розміщення курсу	Лекції: <a href="https://sites.google.com/view/nsakhnenko-data-analysis">https://sites.google.com/view/nsakhnenko-data-analysis</a> Практичні: <a href="https://github.com/natsakh/Data-Analysis">https://github.com/natsakh/Data-Analysis</a>

### Програма навчальної дисципліни

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

Мета - вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Analysis), дослідження масивів даних метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Analysis; розгляд практичних прикладів застосування Data Analysis; підготовка студентів до самостійної роботи з вирішення задач засобами Data Analysis.

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі результати навчання:

1) *Знання*: 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.

2) *Уміння*: обробка датасетов, рішення задач розпізнавання та генерації.

3) *Досвід*: вільно використовувати набутий апарат.

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі програмні компетентності та результати навчання за освітньою програмою (див. на сайті <https://osvita.kpi.ua/op>):

### **Загальні компетентності**

ЗК 1 – Здатність до самонавчання, пошуку, оброблення та інтелектуального аналізу інформації з різних джерел, вміння виявляти, ставити та вирішувати проблеми

### **Фахові компетентності**

ФК 1 – Здатність формалізувати та розв'язувати складні задачі й проблеми, які потребують оновлення й інтеграції знань, часто в умовах неповної, неточної чи недостатньої інформації та суперечливих вимог

ФК3 – Здатність формалізувати, будувати та використовувати у практичній діяльності моделі та методи інтелектуального аналізу даних.

### **Програмні результати навчання**

РН 1 – Використовувати та адаптувати математичні теорії та моделі для забезпечення теоретичного підґрунтя розв'язання наукових та практичних задач

РН 2 – Застосовувати існуючий математичний апарат, розробляти нові моделі, методи та алгоритми при вирішенні актуальних практичних задач широкого спектру

РН 8 – Застосовувати методи здобуття знань із даних, методи оцінки та інтерпретації знайдених закономірностей

РН 9 – Здійснювати математичне і комп'ютерне моделювання складних систем та процесів, обчислювальні експерименти з використанням сучасних методів інтелектуального аналізу даних та комп'ютерних технологій

## **2. Пререквізити та постреквізити дисципліни**

Для засвоєння матеріалу курсу студент повинен знати (на університетському рівні) курси теорії ймовірностей, математичної статистики, математичного аналізу, алгебри.

Отримані знання та практичні навички та засвоєні під час вивчення навчальної дисципліни «Інтелектуальний аналіз даних» теоретичні знання в подальшому можна використовувати під час вивчення багатьох інших навчальних дисциплін, зокрема «Моделювання складних систем».

## **3. Зміст навчальної дисципліни**

### **Розділ 1. Вступ.**

Тема 1.1. Навчання по розміченим та нерозміченим даним.

Тема 1.2. Лінійна регресія.

### **Розділ 2. Задачі класифікації.**

Тема 2.1. Оцінки ефективності моделей, логістична регресія, дерево ухвалення рішень, метод k-ближніх сусідів.

Тема 2.2. Метод опорних векторів, наївний баєсів класифікатор.

Тема 2.3. Ансамблеве навчання.

### **Розділ 3. Задачі кластеризації.**

Тема 3.1. Метод k-середніх, EM алгоритм, ієрархічна кластеризація.

### **Розділ 4. Методи зниження розмірності.**

Тема 4.1. Метод головних компонент.

Тема 4.2. Лінійний дискримінантний аналіз, tSNE

### **Розділ 5. Обробка природної мови.**

Тема 5.1. Передобробка текстових даних.

Тема 5.2. Класифікація та кластеризація текстових даних, побудова мовної моделі, тематичне моделювання.

#### **Розділ 6. Штучні нейронні мережі.**

Тема 6.1. Повнозв'язані нейромережі.

Тема 6.1. Згорткові нейромережі.

Тема 6.1. Рекурентні нейромережі.

#### **Розділ 7. Генеративні нейронні мережі.**

Тема 7.1. Варіаційний автокодувальник.

Тема 7.2. Генеративні змагальні мережі.

#### **Розділ 8. Механізм уваги в нейронних мережах.**

Тема 8.1. Механізм уваги.

Тема 8.2. Архітектура Transformer та мовні моделі, що базуються на цій архітектурі.

Тема 8.3. Генерація зображень за текстовим описом.

### **4. Навчальні матеріали та ресурси**

#### **Базова рекомендована література**

1. *A. Geron*. Hands on Machine Learning with Scikit-Learn, Keras, and Tensorflow. – O'Reilly, 2019. – pp. 684
2. *F. Chollet*. Deep Learning with Python. - Manning, 2021. – pp. 504
3. *I. Goodfellow, Y. Bengio, and A. Courville*. Deep Learning. – MIT Press, 2016. – pp. 651
4. *B. Bengfort, T. Ojeda, and R. Bilbro*, Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning. – O'Reilly, 2018. – pp. 283

#### **Допоміжна рекомендована література**

1. *Н.Б. Шаховська, Р.М. Камінський, О.Б. Вовк*. Системи штучного інтелекту. Львів, 2018. – 389 стор.

#### **Онлайн курси:**

<https://www.coursera.org/learn/machine-learning>

<https://www.coursera.org/learn/python-text-mining>

<https://www.coursera.org/specializations/deep-learning>

**Stanford University Courses:** CS229 Machine Learning, CS230 Deep Learning, CS231n Convolutional Neural Networks for Visual Recognition

**Berkley University:** CS249-158-SP20 Deep Unsupervised Learning

## Навчальний контент

### 5. Методика опанування навчальної дисципліни (освітнього компонента)

#### Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань
<b>I семестр</b>	
1	Вступ. Основні означення. Задача регресії. Метод градієнтного спуску.
2	Методи класифікації. Показники ефективності моделей. Логістична регресія. Дерево ухвалення рішень. Метод к-ближніх сусідів.
3	Метод опорних векторів. Наївний баєсів класифікатор.
4	Ансамблеве навчання.
5	Задачі кластеризації: метод к-середніх, EM алгоритм, ієрархічна кластеризація.
6	Методи зниження розмірності: метод головних компонент.
7	Методи зниження розмірності: лінійний дискримінантний аналіз, tSNE.
8	Методи обробки природної мови (NLP).
9	NLP: класифікація та кластеризація текстових даних, побудова мовної моделі, тематичне моделювання.
10	Штучні нейронні мережі. Повнозв'язані нейромережі.
11	Тренування штучних нейронних мереж.
12	Згорткові нейронні мережі та опрацювання зображень.
13	Часові ряди. Рекурентні нейронні мережі.
14	Штучні нейромережі для опрацювання природної мови.
15	Автокодуювальники. Варіаційний автокодуювальник.
16	Генеративні мережі. Змагальне навчання (GANs).
17	Механізми уваги в нейронних мережах. Архітектура Transformer та мовні моделі, що базуються на цій архітектурі.
18	Генерація зображень за текстовим описом, архітектура DALL E

#### Лабораторні заняття

№ з/п	Назва теми заняття та перелік основних питань
<b>I семестр</b>	
1	Python, NumPy, Pandas, візуалізація даних.
2	Бібліотека scikit-learn: методи класифікації.
3	Бібліотека scikit-learn: методи кластеризації.
4	Обробка природної мови: NLTK та scikit-learn бібліотеки.
5	Знайомство з Keras, високорівневою бібліотекою для глибокого навчання. Повнозв'язні нейромережі.
6	Згорткові нейромережі.
7	Рекурентні нейромережі.
8	Генеративні нейромережі.
9	Механізм уваги в нейронних мережах. Бібліотека Hugging Face

## **6. Самостійна робота студента**

Студент повинен завчасно готуватись до лекцій та лабораторних занять. Перед лекціями необхідно повторити теоретичний матеріал, наданий у попередніх лекціях. Перед лабораторними заняттями необхідно повторити відповідний теоретичний матеріал. Виконання лабораторних робіт є обов'язковим.

## **Політика та контроль**

### **7. Політика навчальної дисципліни (освітнього компонента)**

#### **Відвідування занять**

Матеріал занять, які були з тих чи інших причин пропущені, необхідно опанувати самостійно. У будь-якому випадку студентам рекомендується відвідувати усі види занять, оскільки на них викладається теоретичний матеріал та розвиваються навички, необхідні для виконання домашніх завдань, контрольних та розрахункових робіт. Система оцінювання орієнтована на отримання балів за активність студента, а також виконання завдань, які здатні розвинути практичні уміння та навички.

#### **Пропущені контрольні заходи**

Результат самостійної або модульної контрольної роботи для студента, який не з'явився на контрольний захід, є нульовим. Студент має можливість написати контрольний захід, але максимальний бал за нього буде дорівнювати 50% від загальної кількості балів. Повторне написання модульної контрольної роботи не допускається.

Пропущений екзамен не зараховується; у такому випадку студент отримує запис у відомості «не з'явився» та повинен скласти екзамен на додатковій сесії.

#### **Оголошення результатів контрольних заходів**

Результати виконання тематичних самостійних робіт оголошуються кожному студенту окремо у присутності або у дистанційній формі та супроводжуються оціночними листами, в яких студенти можуть побачити свою оцінку за певними критеріями, а також позначення основних помилок та коментарі до них.

Під час захисту студент зобов'язаний вміти розповісти про розв'язування вказаних викладачем задач та відповісти на теоретичні питання за темами задач. Результати захисту оголошуються кожному студенту окремо у присутності або в дистанційній формі та супроводжуються позитивними коментарями та зауваженнями стосовно помилок.

Результати модульної контрольної роботи вказуються на бланках для модульної контрольної роботи (завдання, які виконували студенти) з позначенням усіх помилок, коректної або некоректної відповіді, а також з коментарями, зауваженнями тощо.

Результати письмово екзамену вказуються на бланках для письмової екзаменаційної роботи (завдання, які виконували студенти) з позначенням усіх помилок, коректної або некоректної відповіді, а також з коментарями, зауваженнями тощо.

### **Академічна доброчесність**

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

### **Норми етичної поведінки**

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

### **Процедура оскарження результатів контрольних заходів**

Студенти мають можливість підняти будь-яке питання, яке стосується процедури проведення та/або оцінювання контрольних заходів, та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами.

Студенти мають право оскаржити результати контрольних заходів, але обов'язково аргументовано, пояснивши, з яким критерієм не погоджуються відповідно до оціночного листа та/або зауважень.

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

### I семестр

№ з/п	Контрольний захід	Макс бал	Ваговий бал	Кіл-ть	Всього
1.	Лабораторні роботи	10	1	4	40
2.	Модульна контрольна робота	20	1	1	20
3.	Екзамен	40	1	1	40
	Всього				100

Проміжна атестація студентів (далі – атестація) є календарним рубіжним контролем та проводиться двічі за семестр, на 8-му та 14-му навчальному тижнях кожного семестра. Для одержання першої атестації поточний рейтинг студента повинен бути щонайменше 5 балів, для одержання другої атестації – щонайменше 10 балів.

Семестрова атестація проводиться усно зі студентами, які були допущені за результатами роботи протягом семестру. Необхідними умовами допуску є:

- семестровий рейтинг  $RD \geq 30$
- виконання та захист лабораторних робіт;
- написання контрольної роботи.

Рейтингова оцінка складається з результатів роботи в семестрі (RD) та результатів усного екзамену. Під час екзамену, забороняється використання будь-яких додаткових довідкових матеріалів.

Студенти, які протягом семестру отримали менше ніж 30 балів, можуть з метою допуску до семестрової атестації (екзамену) виконувати додаткове завдання, яке містить десять практичних задач та оцінюється максимум в 20 балів. Якщо результати написання додаткового завдання є позитивними, студент отримує додаткові бали та допуск до екзамену.

### 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

### Робочу програму навчальної дисципліни (силабус):

Складено доцентом кафедри ММАД НН ФТІ, д.ф.-м.н. Сахненко Н.К.

Ухвалено кафедрою ММАД НН ФТІ (протокол №14 від 22.06.2022)

Затверджено Методичною комісією НН ФТІ (протокол №6 від 30.06.2022).