

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
КАФЕДРА ІНТЕЛЕКТУАЛЬНИХ ПРОГРАМНИХ СИСТЕМ**

«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

_____ Кашпур О.Ф.

«___» _____ 2019 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
СИСТЕМИ МАШИННОГО НАВЧАННЯ
для студентів**

| | |
|------------------|---|
| галузь знань | 12 Інформаційні технології |
| спеціальність | 121 Інженерія програмного забезпечення |
| освітній рівень | магістр |
| освітня програма | Програмне забезпечення систем |
| спеціалізація | Інтелектуальні програмні системи |
| вид дисципліни | вибіркова |

| | |
|--|-------------------|
| Форма навчання | денна |
| Навчальний рік | 2021/2022 |
| Семестр | 3 |
| Кількість кредитів ECTS | 4 |
| Мова викладання, навчання та оцінювання | українська |
| Форма заключного контролю | залік |

Викладачі: **д. ф.-м. н., професор Провотар О.І.** (лекції, лабораторні заняття).

Пролонговано: на 20__/20__ н. р. _____ (_____) «__» 20__ р.

на 20__/20__ н. р. _____ (_____) «__» 20__ р.

Розробники: Провотар Олександр Іванович, д. ф.-м. н., професор, завідувач кафедри інтелектуальних програмних систем; Сенько Іван Олександрович, асистент кафедри математики Державної наукової установи «Київський академічний університет».

ЗАТВЕРДЖЕНО

Завідувач кафедри інтелектуальних програмних систем

_____ О.І. Провотар

Протокол № __ від «__» _____ 2019 р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «__» _____ 2019 року №__

Голова науково-методичної комісії _____ Л.Л. Омельчук

«__» _____ 2019 року

Затверджено вченою радою факультету комп'ютерних наук та кібернетики

Протокол від «__» _____ 2019 року №__

Голова вченої ради факультету _____ А.В. Анісімов

1. Мета дисципліни – ознайомлення з основними видами задач машинного навчання, моделями, що описують ці задачі для різних типів доступних даних, алгоритмами розв'язування цих задач для різних моделей, способами попередньої підготовки даних та способи оцінювання якості розв'язків задач машинного навчання.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. **Знати:** базові відомості про розробку програмного забезпечення з використанням сучасних мов програмування, деякі глави математичного аналізу, дискретної математики, лінійної алгебри, програмування та теорії ймовірностей.
2. **Вміти:** розв'язувати та системи лінійних алгебраїчних рівнянь, знаходити фундаментальну систему розв'язків, обчислювати суми послідовностей та рядів аналітичними методами, обчислювати основні теоретичні та емпіричні характеристики випадкових розподілів.
3. **Володіти навичками:** елементами інтегрального числення та теорії ймовірностей.

3. Анотація навчальної дисципліни. Навчальна дисципліна «Системи машинного навчання» є складовою освітньо-наукової програми підготовки фахівців за другим (магістерським) рівнем вищої освіти у галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення в рамках освітньо-наукової програми «Програмне забезпечення систем».

Дана дисципліна належить до переліку дисциплін вільного вибору студента, спеціалізація «Інтелектуальні програмні системи». Викладається у **3 семестрі в обсязі – 120 год. (3 кредити ECTS)**, зокрема: лекції – 18 год., лабораторні заняття – 20 год., самостійна робота – 80 год., консультації – 2 год. У курсі передбачено 2 змістовні частини та 2 контрольні роботи. Завершується дисципліна – **заліком**.

В результаті вивчення навчальної дисципліни студент повинен:

знати: відомості про основні задачі машинного навчання, моделі для опису даних, типові методи розв'язання та умови застосовності цих методів.

вміти: вибирати алгоритми для розв'язання задач машинного навчання в залежності від природи та структури даних, проводити попередню обробку даних, оцінювати якість отриманих розв'язків.

Дисципліна «Системи машинного навчання» є логічним продовженням, доповненням та розширенням дисциплін «Прикладне застосування нейронних мереж» другого (магістерського) рівня вищої освіти у галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення, в рамках освітньо-наукової програми «Програмне забезпечення систем».

4. Завдання (навчальні цілі). Основними завданнями дисципліни «Системи машинного навчання» є набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень в області машинного навчання відповідно до освітньої кваліфікації магістр з інженерії програмного забезпечення за спеціалізацію «Інтелектуальні програмні системи». Зокрема, розвивати:

- Здатність до абстрактного мислення, аналізу та синтезу (ЗК01).
- Здатність проведення теоретичних та прикладних досліджень на відповідному рівні (ЗК03).
- Здатність удосконалювати свої навички на основі аналізу попереднього досвіду (ЗК06).

- Здатність аналізувати предметні області, формувати, аналізувати та моделювати вимоги до програмного забезпечення (СК01).
- Здатність проектувати програмне забезпечення, включаючи проведення моделювання його архітектури, поведінки та процесів функціонування окремих підсистем і модулів (СК03).
- Вміння планувати і проводити наукові дослідження, готувати результати наукових робіт з інженерії програмного забезпечення до оприлюднення (СК09).
- Здатність розробляти та аналізувати математичні моделі інтелектуальних процесів пошуку, аналізу, обробки та збереження інформації (СК12.2).
- Здатність розробляти, аналізувати та реалізовувати алгоритми для вирішення задач автоматизованого пошуку, аналізу, обробки та збереження інформації (СК13.2).
- Здатність використовувати сучасні програмні системи та інформаційні технології для проектування та розробки інтелектуальних програмних систем (СК14.2).
- Здатність проектувати та розробляти спеціалізовані, автономні, розподілені інтелектуальні програмні системи автоматизованого пошуку, аналізу, обробки та збереження інформації (СК15.2).

5. Результати навчання за дисципліною.

| Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність) | | Форми (та/або методи і технології) викладання і навчання | Методи оцінювання та пороговий критерій оцінювання (за необхідності) | Відсоток у підсумковій оцінці з дисципліни |
|--|--|--|--|--|
| Код | Результат навчання | | | |
| РН1.1 | Знати основні поняття і визначення машинного навчання | Лекції. | Контрольна робота 1 (тест), 60% правильних відповідей, захист лабораторної роботи. | 15% |
| РН1.2 | Знати основні алгоритми розв'язання задач машинного навчання, умови їх застосовності. | Лекції, лабораторні заняття. | Контрольна робота 1, 2 (тест), 60% правильних відповідей. | 20% |
| РН1.3 | Знати способи підготовки даних для різних алгоритмів машинного навчання, способи оцінки якості отриманих розв'язків. | Лекції, лабораторні заняття. | Контрольна робота 2 (тест), 60% правильних відповідей, захист лабораторної роботи. | 15% |
| РН2.1 | Вміти застосовувати на практиці інструментальні програмні засоби для розв'язання задач | Лабораторні заняття, самостійна робота. | Контрольна робота 1, 2 (тест), 60% правильних відповідей, захист лабораторної | 24% |

| | | | | |
|-------|---|--|-----------------------------|-----|
| | машинного навчання та обробки даних. | | роботи. | |
| РН2.2 | Вміти проектувати та розробляти моделі машинного навчання | Лабораторні заняття, самостійна робота. | Захист лабораторної роботи. | 10% |
| РН3.1 | Консультуватися з викладачем стосовно питань що виникають у ході опанування теоретичного матеріалу. | Лекції, лабораторні заняття, консультації. | Захист лабораторної роботи | 4% |
| РН4.1 | Організувати свою самостійну роботу для досягнення результату | Самостійна робота. | Захист лабораторної роботи. | 6% |
| РН4.2 | Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість | Самостійна робота. | Захист лабораторної роботи. | 6% |

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання.

| Результати навчання дисципліни | РН1.1 | РН1.2 | РН1.3 | РН2.1 | РН2.2 | РН3.1 | РН4.1 | РН4.2 |
|--|-------------------------------|-------|-------|-------|-------|-------|-------|-------|
| | Програмні результати навчання | | | | | | | |
| ПРН-1. Знати і системно застосовувати методи аналізу та моделювання прикладної області, виявлення інформаційних потреб і збору вихідних даних для проектування програмного забезпечення. | + | + | + | + | + | | + | + |
| ПРН03. Знати і застосовувати базові концепції і методології моделювання інформаційних процесів. | + | | | + | + | | + | + |
| ПРН06. Аналізувати, оцінювати і обирати методи, сучасні програмно-апаратні інструментальні та обчислювальні засоби, технології, алгоритмічні та програмні рішення для ефективного виконання конкретних виробничих задач з програмної інженерії. | | | | + | | | | + |
| ПРН07. Обґрунтовано обирати парадигми і мови програмування для вирішення прикладних завдань; застосовувати на практиці системні та спеціалізовані засоби, компонентні технології (платформи) та інтегровані середовища розробки програмного забезпечення. | | + | | + | | + | + | + |
| ПРН12. Формулювати, експериментально підтверджувати, обґрунтовувати і застосовувати на практиці в процесі розробки програмного забезпечення конкурентоспроможні ідеї, методи, технології вирішення професійних, науково-технічних завдань | + | + | | + | | | + | + |

| | | | | | | | | |
|---|---|---|---|---|---|--|---|---|
| в умовах невизначеності. | | | | | | | | |
| ПРН14. Пояснити, аналізувати, цілеспрямовано шукати і обирати необхідні для вирішення фахових наукових і прикладних задач інформаційно-довідкові та науково-технічні ресурси і джерела знань з урахуванням сучасних досягнень науки і техніки. | | | | | | | + | + |
| ПРН15.2. Знати основні методи структурного, лексичного, семантичного та концептуального аналізу предметних областей. | + | + | | | | | | + |
| ПРН17.2. Знати основні математичні моделі інтелектуальних процесів пошуку, аналізу, обробки та збереження інформації. | | | + | | | | | + |
| ПРН18.2. Знати основні алгоритми для вирішення задач автоматизованого пошуку, аналізу, обробки та збереження інформації. | | + | + | | | | | + |
| ПРН19.2. Вміти проводити структурний, лексичний, семантичний та концептуальний аналіз предметних областей. | | | + | | | | | + |
| ПРН20.2. Вміти розробляти, аналізувати та реалізовувати алгоритми для вирішення задач автоматизованого пошуку, аналізу, обробки та збереження інформації. | | | | + | | | | + |
| ПРН21.2 Вміти використовувати сучасні програмні системи та інформаційні технології для проектування та розробки інтелектуальних програмних систем. | | | | | + | | | + |
| ПРН22.2. Вміти проектувати та розробляти спеціалізовані інтелектуальні програмні системи автоматизованого пошуку, аналізу, обробки та збереження інформації. | | | | + | + | | | |

7. Схема формування оцінки.

7.1 Форми оцінювання студентів.

Семестрове оцінювання:

1. Контрольна робота 1: РН1.1, РН1.2, РН2.1 – **20/12 балів.**
2. Контрольна робота 2: РН1.2, РН1.3, РН2.1 – **20/12 балів.**
3. Лабораторна робота 1: РН1.1, РН1.3, РН2.1, РН2.2, РН3.1, РН4.1, РН4.1 – **20/12 балів.**
4. Лабораторна робота 2: РН1.1, РН1.3, РН2.1, РН2.2, РН3.1, РН4.1, РН4.1 – **20/12 балів.**
5. Лабораторна робота 3: РН1.1, РН1.3, РН2.1, РН2.2, РН3.1, РН4.1, РН4.1 – **20/12 балів.**

Підсумкова оцінювання (у формі заліку):

- Залікові бали визначаються як сума оцінок/балів за всіма успішно оціненими результатами навчання передбачених даною програмою.
- Оцінки нижче від мінімального порогового рівня не додаються.
- Мінімальний пороговий рівень для кожного окремого компонента становить 50% від максимально можливої кількості балів.
- Мінімальний пороговий рівень для сумарної оцінки за всіма компонентами становить 60% від максимально можливої кількості балів.

7.2 Організація оцінювання:

Терміни проведення форм оцінювання:

1. Контрольна робота 1: до 7 тижня семестру.
2. Контрольна робота 2: до 13 тижня семестру.
3. Лабораторна робота 1: до 4 тижня семестру.
4. Лабораторна робота 2: до 7 тижня семестру.
5. Лабораторна робота 3: до 13 тижня семестру.

У випадку отримання незадовільної оцінки (менше 50% від встановленого максимуму) за контрольну роботу, студенти мають право на одне перескладання цієї роботи у визначений викладачем термін із можливістю отримати не більше 80% балів від встановленої для неї максимальної кількості балів.

У випадку відсутності студентів з поважних причин відпрацювання та перездачі контрольних робіт здійснюються у відповідності до «Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу» від 1 жовтня 2010 року.

У випадку встановлення фактів порушення студентами академічної доброчесності передбачених пунктом 9.8.2 «Положення про організацію освітнього процесу у Київському національному університеті імені Тараса Шевченка» що діє від 07.05.2018, вони будуть притягнуті до відповідальності передбаченої пунктом 9.8.3 цього положення.

7.3 Шкала відповідності оцінок.

| | |
|-----------------------------|--------|
| Зараховано / Passed | 60-100 |
| Не зараховано / Fail | 0-59 |

8. Структура навчальної дисципліни. Тематичний план лекцій та лабораторних занять.

| № лекції | Назва лекції | Кількість годин | | |
|---|---|-----------------|---------------------|-------------------|
| | | Лекції | Лабораторні заняття | Самостійна робота |
| Частина 1. Задачі навчання з учителем – класифікація та прогнозування. | | | | |
| 1 | Тема 1. Види задач машинного навчання. Задача класифікації. Міри якості класифікації. Побудова моделей машинного навчання. Типи змінних. Оцінка точності, підбір параметрів. | 2 | 2 | 9 |
| 2 | Тема 2. Дерева рішень. Критерій поділу Джині та критерій ентропії. Умовні дерева рішень. Випадкові ліси. | 2 | 2 | 9 |
| 3 | Тема 3. Байесівська класифікація. Логістична регресія. Машини опорних векторів. | 2 | 2 | 10 |
| 4 | Тема 4. Задача регресії. Міри точності у задачі регресії. Лінійна регресія. Оцінки точності. Нелінійна регресія, локально-зважена регресія. | 2 | 2 | 10 |
| 5 | Тема 5. Перетворення змінних для задач машинного навчання. Регуляризовані | 2 | 2 | 10 |

| | | | | |
|--|--|-----------|-----------|-----------|
| | регресійні моделі: ridge, LASSO, Elastic-Net. Дерева регресії. | | | |
| 6 | Тема 6. Метод найближчих сусідів. Подолання перенавчання: визначення важливих змінних. Ансамблеві методи: бустинг та беггінг. | 2 | 2 | 10 |
| Контрольна робота 1 | | | | 2 |
| Всього по частині 1 | | 12 | 12 | 60 |
| Частина 2. Задачі навчання без учителя та підготовка даних. | | | | |
| 7 | Тема 7. Підготовка даних для задач машинного навчання: очистка, знаходження викидів, заповнення пропусків. Задачі навчання без учителя: кластеризація, пошук аномалій, зниження розмірності. Області їх застосування. | 2 | 2 | 8 |
| 8 | Тема 8. Кластеризація: центроїдні методи (kmeans, pam, clara, fanny). Індекси якості кластеризації — внутрішні та зовнішні. Кластеризація: методи зв'язності (agglomerative clustering). | 2 | 2 | 5 |
| 9 | Тема 9. Кластеризація: статистичні методи (gaussian mixtures). Кластеризація: методи аналізу щільності (DBSCAN, OPTICS). Кластеризація: ядерні методи (mean-shift). | 2 | 4 | 5 |
| Контрольна робота 2 | | | | 2 |
| Всього по частині 2 | | 6 | 8 | 20 |
| Консультація | | 2 | | |
| ВСЬОГО | | 18 | 20 | 80 |

Загальний обсяг – **120** год., в тому числі:

Лекцій – **18** год.

Лабораторних – **20** год.

Самостійна робота – **80** год.

Консультацій – **2** год.

Теми, винесені на самостійне вивчення.

1. Пошук аномалій: неперервні дані.
2. Пошук аномалій: категоріальні дані. Метод головних компонент.
3. Розклад невід'ємних матриць.
4. Засоби мови Python для машинного навчання: бібліотеки numpy, pandas, sci-kit learn.

9. Рекомендовані джерела.

Основні:

1. A.D. Joseph, B. Nelson, B.I.P. Rubinstein, J.D. Tygar: Adversarial Machine Learning – Cambridge University Press, 2019.
2. C.C. Aggarwal: Machine Learning for Text – Springer, 2018.
3. D. Forsyth: Applied Machine Learning – Springer, 2019.
4. E. Alpaydin: Introduction to Machine Learning, 3rd ed. – The MIT Press, 2014.
5. J.-T. Chien: Source Separation and Machine Learning – Academic Press, 2019.
6. M. Gori: Machine Learning: A Constraint-Based Approach Morgan Kaufmann, 2018.
7. M. Mohri, A. Rostamizadeh, A. Talwalkar: Foundations of Machine Learning, 2nd ed. – The MIT Press, 2018.
8. P. Larracaga, D. Atienza, J. Diaz-Rozo, A. Ogbechie, C. Puerto-Santana, C. Bielza: Industrial Applications of Machine Learning – CRC Press, 2019.
9. R.E. Neapolitan, X. Jiang: Artificial Intelligence: With an Introduction to Machine Learning, 2nd ed. – CRC Press, 2018.
10. S.W. Knox: Machine Learning: A Concise Introduction – Wiley, 2018.

Додаткові:

1. A. Geron: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems – O'Reilly Media, Inc., 2017.
2. A. Moitra: Algorithmic Aspects of Machine Learning – Cambridge University Press, 2018.
3. A. Zheng, A. Casari: Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists – O'Reilly Media, Inc., 2018.
4. A.C. Mueller, S. Guido: Introduction to Machine Learning with Python – O'Reilly Media, Inc., 2016.
5. C. Albon: Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning – O'Reilly Media, Inc., 2018.
6. G. Rebala, A. Ravi, S. Churiwala: An Introduction to Machine Learning – Springer 2019.
7. J. Unpingco: Python for Probability, Statistics, and Machine Learning, 2nd ed. – Springer, 2019.
8. M. Kubat: An Introduction to Machine Learning, 2nd ed. – Springer, 2017.
9. N. Buduma, N. Locascio: Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms – O'Reilly Media, Inc., 2017.
10. R. Fernandes de Mello, M.A. Ponti: Machine Learning: A Practical Approach on the Statistical Learning Theory – Springer, 2018.
11. W.-M. Lee: Python Machine Learning – Wiley, 2019.
12. X.-S. Yang: Introduction to Algorithms for Data Mining and Machine Learning – Academic Press, 2019.