

**МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ЗВ'ЯЗКУ ім. О.С. ПОПОВА**

Кафедра Інформаційних технологій

ЗАТВЕРДЖУЮ

Ректор ОНАЗ ім. О.С. Попова

_____ **П.П. Воробієнко**

“ _____ ” _____ 2020 р.

Машинне навчання

ПРОГРАМА
навчальної дисципліни
підготовки *магістра*

Спеціальність: 121 Інженерія програмного забезпечення

Одеса 2020

РОЗРОБЛЕНО ТА ВНЕСЕНО:
ОДЕСЬКОЮ НАЦІОНАЛЬНОЮ АКАДЕМІЄЮ ЗВ'ЯЗКУ ІМ. О.С.ПОПОВА

РОЗРОБНИК ПРОГРАМИ: д.т.н., проф. Романюк В.В.

Програму розглянуто і схвалено на засіданні кафедри

Протокол № _____ від “_____” _____ 2020 р.

Зав. каф. _____

Програму погоджено з кафедрами:

Зав. каф. _____

Зав. каф. _____

Програму розглянуто і схвалено Вченою радою ННІ ІКПІ

Спеціальність: 121 Інженерія програмного забезпечення

Протокол № _____ від “_____” _____ 2020 р.

Директор ННІ ІКПІ _____ І.В. Стрелковська

Програму розглянуто і схвалено Вченою радою Одеської національної академії зв'язку ім. О.С. Попова

Протокол № _____ від “_____” _____ 2020 р.

Голова ради, професор _____ О.В. Бондаренко

ВСТУП

Програма вивчення навчальної дисципліни “Машинне навчання” складена відповідно до освітньо-професійної програми підготовки магістрів спеціальності 121 Інженерія програмного забезпечення.

Теорія навчання машин перебуває на стику прикладної статистики, чисельних методів оптимізації, моделювання систем, і за останні десять років сформувалася в самостійну інженерну дисципліну. Навчальна дисципліна “Машинне навчання” спрямована на успішне освоєння сучасних методів машинного навчання. Отримані в результаті освоєння дисципліни знання необхідні при вирішенні практичних завдань у сфері професійної діяльності, проектуванні і розробці інформаційних систем.

В рамках даного курсу розглядаються основні завдання навчання по прецедентах: класифікація, кластеризація, регресія, зниження розмірності. Для вирішення цих завдань студенти вивчають методи, як класичні, так і нові, створені за останні 10-15 років. Курс балансує між глибоким розумінням математичних основ та прикладним інженерним застосуванням методів машинного навчання.

Всі методи викладаються за єдиною схемою:

- базові ідеї і евристики;
- їх формалізація та математична теорія;
- опис алгоритму у вигляді слабо формалізованого псевдокоду;
- аналіз переваг, недоліків і меж застосування;
- шляхи усунення недоліків;
- порівняння з іншими методами.
- приклади прикладних задач.

Програма навчальної дисципліни складається з таких змістових модулів:

1. Методи класифікації та регресії.
2. Методи кластеризації та відбір ознак.
3. Контекстно-залежна класифікація.
4. Глибинне навчання.

1. Мета та завдання навчальної дисципліни

Метою викладання навчальної дисципліни є підготовка магістрів до ефективного застосування методів машинного навчання та оволодіння практичними навичками вирішення прикладних завдань інтелектуального аналізу даних.

Цілі курсу:

- оволодіння основними поняттями навчання машин;
- набуття фахових навичок формалізації задачі прийняття рішень та вибору відповідного методу рішення залежно від початкових даних;
- набуття практичних навичок побудови основних моделей машинного навчання;
- ознайомлення з сучасними програмними засобами розробки інтелектуальних систем;
- набуття практичних навичок по візуалізації даних та презентації отриманих результатів моделювання інтелектуальних систем.

Курс передбачає теоретичні та практичні заняття.

В межах теоретичної частини забезпечуються **знання**:

- принципи побудови векторів ознак, вирішальних правил і класифікації;
- основні види класифікаторів;
- принципи побудови лінійних та нелінійних класифікаторів;
- методи кластеризації;
- особливості вибору ознак класифікації і попередньої обробки даних.
- основні методи штучного інтелекту, засоби формалізованого подання знань, напрямки інтелектуалізації автоматизованих інформаційних та інформаційно-керуючих систем.

Проходження практичної частини курсу формує **вміння**:

- вибирати відповідний вид класифікатора в залежності від розв’язуваної задачі;
- вибирати набір ознак для класифікації і проводити попередню обробку даних;
- застосовувати алгоритми побудови і навчання класифікатора по вибірці;
- виконувати обчислення, пов’язані з навчанням і роботою класифікатора;
- розробляти структуру баз знань та баз даних;
- розробляти алгоритми та програми для обробки баз даних та знань (пошук інформації, складання логічних рівнянь);
- працювати з інструментарієм аналітичної обробки даних, який є у складі корпоративними СКБД, сховищами даних та базами даних в Internet-середовищі;
- проводити чисельні експерименти на модельних і реальних даних та інтерпретувати їхні результати;
- представляти результати досліджень в усній і письмовій формах.

Вивчення навчальної дисципліни “Машинне навчання” передбачає формування та розвиток у студентів **компетентностей**:

загальних:

- здатність до абстрактного мислення, аналізу та синтезу;
- здатність спілкуватися іноземною мовою як усно, так і письмово;
- здатність проведення теоретичних та прикладних досліджень на відповідному рівні;
- здатність мотивувати людей та рухатися до спільної мети, працювати в команді співробітників;
- здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності);
- здатність удосконалювати свої навички на основі аналізу попереднього досвіду.
- здатність генерувати нові ідеї (креативність);
- здатність управляти своїм часом та розуміти важливість дедлайнів.

фахових:

- здатність аналізувати предметні області, формувати, аналізувати та моделювати вимоги до програмного забезпечення;
- здатність ідентифікувати, класифікувати та описувати проектні завдання, знаходити раціональні методи й підходи до їх розв’язання;

– здатність проектувати програмне забезпечення, включаючи проведення моделювання його архітектури, поведінки та процесів функціонування окремих підсистем і модулів;

– здатність розвивати і реалізовувати нові конкурентоспроможні ідеї в інженерії програмного забезпечення;

– здатність оцінювати ступінь обґрунтованості застосування специфікацій, стандартів, правил і рекомендацій в професійній галузі та дотримуватися їх при реалізації процесів життєвого циклу програмного забезпечення;

– здатність розробляти і координувати процеси, фази та ітерації життєвого циклу програмних систем на основі застосування відповідних моделей, методів та технологій розробки програмного забезпечення;

– здатність використовувати методи машинного навчання та роботи з big data, засоби штучного інтелекту для дослідження та аналізу процесів життєвого циклу програмного забезпечення;

– здатність аналізувати вимоги, розробляти та тестувати хмарні застосування, реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень, обирати відповідні архітектури та проектні шаблони проектування та інтеграції таких застосувань;

– здатність використовувати знання сучасних комп'ютерних та інформаційних технологій та інструментів інженерних і наукових досліджень, розрахунків, обробки та аналізу даних, моделювання та оптимізації.

Результати навчання даної дисципліни деталізують такі **програмні результати навчання**:

– знати і системно застосовувати методи аналізу та моделювання прикладної області, виявлення інформаційних потреб і збору вихідних даних для проектування програмного забезпечення;

– оцінювати і вибирати методи і моделі розробки, впровадження, експлуатації програмних засобів та управління ними на всіх етапах життєвого циклу;

– аналізувати, оцінювати і вибирати методи, сучасні програмно-апаратні інструментальні та обчислювальні засоби, технології, алгоритмічні та програмні рішення для ефективного виконання конкретних виробничих задач з програмної інженерії;

– обґрунтовано вибирати парадигми і мови програмування для вирішення прикладних завдань; застосовувати на практиці системні та спеціалізовані засоби, компонентні технології (платформи) та інтегровані середовища розробки програмного забезпечення;

– набувати нові наукові і професійні знання, вдосконалювати навички, прогнозувати розвиток програмних систем та інформаційних технологій;

– 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;

– здобувати необхідну інформацію з іншомовної літератури, аналізувати та вибирати необхідні для вирішення фахових наукових і прикладних задач інформаційно-

довідкові та науково-технічні ресурси і джерела знань з урахуванням сучасних досягнень науки і техніки, здійснювати різні види комунікації під час спілкування;

– проектувати та створювати програмні системи зберігання та обробки великих масивів даних, розробляти високонавантаженні системи обробки даних, використовувати методи штучного інтелекту та машинного навчання у вирішенні практичних завдань;

– організовувати командну роботу, управляти проектами, підбирати команду проекту, ефективно працювати в групі, визначати та розподіляти завдання з метою вирішення різноманітних дослідницьких та практичних завдань;

– знати і застосовувати на практиці архітектури та стандарти розподілених обчислень, концепції та технології паралельної обробки інформації при розробці та експлуатації паралельного та розподіленого програмного забезпечення.

На вивчення навчальної дисципліни відводиться 270 годин / 9 кредитів ECTS.

2. Інформаційний обсяг навчальної дисципліни

Модуль 1. *Інтелектуальний аналіз даних*

Змістовний модуль 1. Методи класифікації та регресії.

Тема 1. Первинний аналіз даних з Pандас. Основні поняття і визначення. Приклади задач. Ознаки, вектори ознак. Об'єкти, класи. Класифікація. Класифікатор. Навчання, види навчання “з вчителем” і “без вчителя”.

Тема 2. Візуальний аналіз даних з Python. Лінійні методи класифікації. Розбір прикладів і рішення задач за темами: лінійна модель класифікації, метод стохастичного градієнта.

Тема 3. Класифікація, дерева рішень і метод найближчих сусідів. Метричні методи класифікації. Метод найближчих сусідів і його узагальнення. Підбір числа k за критерієм змінного контролю. Узагальнений метричний класифікатор, поняття відступу. Метод потенційних функцій, градієнтний алгоритм. Відбір еталонів і оптимізація метрики.

Тема 4. Лінійні моделі класифікації і регресії. Метод опорних векторів. Основи методу опорних векторів. Випадок лінійно розділеної вибірки. Випадок лінійно нерозділеної вибірки. Розбір прикладів і рішення задач.

Змістовний модуль 2. Методи кластеризації та відбір ознак.

Тема 5. Композиції: беггінг, випадковий ліс. Структура дерев рішень. Алгоритм побудови дерева рішень. Навчання дерева рішень. Алгоритм Random Forest. Програмна реалізація дерева рішень. Обробка пропусків. Переваги та недоліки вирішальних дерев.

Тема 6. Побудова і відбір ознак. Методи відновлення регресії. Метод найменших квадратів. Непараметрична регресія: ядерне згладжування. Лінійна регресія. Метод головних компонент.

Тема 7. Навчання без вчителя: PCA і кластеризація. Вибір ознак і підготовка даних. Вплив вибору набору ознак на результати класифікації. Попередня обробка даних. Відсутні значення. Перетворення ознак. Вибір ознак на основі перевірки гіпотез. Вибір підмножини ознак.

Модуль 2. *Наука даних*

Змістовний модуль 3. Контекстно-залежна класифікація.

Тема 8. Стохастичний градієнтний спуск. Навчання на гігабайтах з Vowpal Wabbit. Онлайн-навчання. Контекстно-залежна класифікація. Алгоритм Вітербо. Приховані марківські моделі.

Тема 9. Аналіз часових рядів з допомогою Python. Експоненціальне згладжування. Модель Хольта-Вінтерса. Крос-валідація на часових рядах. Лінійні та нелінійні моделі на часових рядах. Вилучення ознак. Алгоритм XGBoost.

Тема 10. Градієнтний бустинг. Функціональний градієнтний бустинг. Функції втрат класифікації. Функції втрат регресії.

Тема 11. Багатошарові нейронні мережі. Біологічний нейрон. Персептрон. Функції активації. Проблема повноти. Повнота двошарових мереж в просторі булевих функцій. Теорема Колмогорова, Стоуна, Горбаня (без доведення).

Тема 12. Алгоритм зворотного поширення помилок. Навчання в якості градієнтного спуску. Локальні мінімуми функції помилки. Кроки алгоритму зворотного поширення. Метод пошарового налаштування мережі. Підбір структури мережі: методи поступового ускладнення мережі, оптимальне проріджування нейронних мереж.

Змістовний модуль 4. Глибинне навчання.

Тема 13. Основи згорткових нейронних мереж. Глибокі згорткові моделі: практичні застосування. Виявлення об'єктів. Задачі класифікації та сегментації.

Тема 14. Рекурентні нейронні мережі. Нейронна мережа Хопфілда. Довга короткострокова пам'ять (LSTM). Керований рекурентний блок. Нейронний компресор історії даних.

Тема 15. Глибинне навчання з підкріпленням. Середовище та агент. Система підкріплення та її види.

Тема 16. Генеративно-змагальні мережі. Генеративна модель. Дискримінативна модель. Антагоністична гра.

Тема 17. Поліпшення глибинних нейронних мереж: тюнінг гіперпараметрів, регуляризація і оптимізація.

3. Рекомендована література

Базова

1. Géron A. Hands-on machine learning with Scikit-learn and TensorFlow. Sebastopol: O'Reilly Media, Inc., 2017. 751 p.
2. Simeone O. A brief introduction to machine learning for engineers. London: King's College London, 2017. 201 p.
3. Tagliaferri L., Morales M., Birbeck E., Wan A. Python machine learning projects. New York City: DigitalOcean, 2019. 135 p.
4. Коэльо Л. П., Ричарт В. Построение систем машинного обучения на языке Python. 2-е издание/ пер. с англ. Слинкин А. А. Москва: ДМК Пресс, 2016. 302 с.
5. Маккинли У. Python и анализ данных / пер. с англ. Слипкин А. А. Москва: ДМК Пресс, 2015. 482 с.

6. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб: Питер, 2018. 480 с.
7. Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова, Москва: ДМК Пресс, 2017. 418 с.
8. Рашид Т. Создаем нейронную сеть: пер. с англ. СПб: ООО “Альфа-книга”, 2017. 272 с.
9. Хенрик Б., Джозеф Р., Марк Ф. Машинное обучение. СПб: Питер, 2017. 336 с.
10. Шарден Б., Массарон Л., Боскетти А. Крупномасштабное машинное обучение вместе с Python. Москва: ДМК Пресс, 2018. 358 с.

Додаткова

1. Deisenroth M. P., Faisal A. A., Ong C. S. Mathematics for machine learning. Cambridge: Cambridge University Press, 2019. 411 p.
2. Romanuke V. V. Two-layer perceptron for classifying flat scaled-turned-shifted objects by additional feature distortions in training. *Journal of Uncertain Systems*. 2015. Vol. 9, No. 4. P. 286–305. <https://doi.org/10.20535/1810-0546.2017.6.110724>
3. Romanuke V. V. Optimal training parameters and hidden layer neurons number of two-layer perceptron for generalized scaled objects classification problem. *Information Technology and Management Science*. 2015. Vol. 18. P. 42–48. <https://doi.org/10.1515/itms-2015-0007>
4. Romanuke V. V. Training data expansion and boosting of convolutional neural networks for reducing the MNIST dataset error rate. *Research Bulletin of NTUU “Kyiv Polytechnic Institute”*. 2016. No. 6. P. 29–34. <https://doi.org/10.20535/1810-0546.2016.6.84115>
5. Romanuke V. V. An infinitely scalable dataset of single-polygon grayscale images as a fast test platform for semantic image segmentation. *KPI Science News*. 2019. No. 1. P. 24–34. <https://doi.org/10.20535/kpi-sn.2019.1.157259>
6. Алексеева Т. В., Амириди Ю. В., Дик В. В. Информационные аналитические системы. Москва: МФПУ «Синергия», 2013. 384 с. URL: <http://znanium.com/bookread2.php?book=451186>
7. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 3-е изд. Москва: Издательский дом «Вильямс», 2013. 1408 с.
8. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. Москва: ДМК Пресс, 2015. 400 с.

Інформаційні ресурси

1. Schmidt, M. 80 lectures on machine learning. *Mark Schmidt, University of British Columbia*. 11.01.2016 Web. 15.08.2020 Retrieved from: <https://www.cs.ubc.ca/~schmidtm/Courses/LecturesOnML/>
2. Библиотека программиста. Учебный план освоения глубокого обучения и нейросетей за 6 недель [Электронный ресурс] // <https://proglib.io/>. 2018. URL: <https://proglib.io/p/deep-learning-in-6-weeks>. Дата звернення: 31.08.2020
3. Блеканов И. С. Введение в науку о данных [Электронный ресурс] // <https://www.coursera.org/>. 2017. URL: <https://www.coursera.org/learn/vvedeniye-v-nauku-o-dannykh#instructors>. Дата звернення: 30.08.2020

4. Воронцов К. В. Машинное обучение (курс лекций, К. В. Воронцов) [Электронный ресурс] // MachineLearning.ru. 2018. URL: <http://www.machinelearning.ru/wiki/index.php?title=%> Дата звернення: 01.09.2020
5. Петрів О., Чернятевич А., Шкорінова О. Машинне навчання [Електронний ресурс] // <https://prometheus.org.ua/>. 2017. URL: https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about. Дата звернення: 23.08.2020
6. ПЗ-1.1 – Машинне навчання. Модуль 1 [Електронний ресурс] // Google Classroom. 2020. URL: <https://classroom.google.com/u/0/c/NDA0MTUyNzg4ODNa>. Дата звернення: 01.09.2020
7. ПЗ-1.2 – Машинне навчання. Модуль 2 [Електронний ресурс] // Google Classroom. 2020. URL: <https://classroom.google.com/u/0/c/NDA0MTUyNzg4ODNa>. Дата звернення: 02.09.2020
8. Машинне навчання. Навчальний курс [Електронний ресурс] // Система дистанційного навчання ОНАЗ ім. О.С. Попова URL: <https://e-learning.onat.edu.ua/course/view.php?id=172>. Дата звернення: 21.08.2020

4. Форма підсумкового контролю успішності навчання

Формою контролю є залік та курсова робота в модулі 1 та іспит в модулі 2.

5. Засоби діагностики успішності навчання

Діагностика знань студентів здійснюється за допомогою:

- усного опитування;
- тестування;
- письмових поточних контрольних робіт;
- письмових екзаменаційних завдань.