Затверджено Вченою Радою ОНУ
імені І.І. Мечникова
від "20"грудня 20 16 р. № 4

Одеський національний університет імені І.І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра математичного забезпечення комп'ютерних систем

"ЗАТВЕРДЖУЮ"
Проректор з науково-педагогічної роботи
_____
"_____" _____ 20___ р.

НАВЧАЛЬНА ПРОГРАМА ДИСЦИПЛІНИ

ОК19 «Інтелектуальний аналіз даних і методи machine learning»
(назва навчальної дисципліни)

Рівень вищої освіти _____бакалавр_____

Спеціальність _____126 – Інформаційні системи та технології_____
(код і назва спеціальності (тей)

2017

Розробник:
Петрушина Т.І., к.ф.-м.н., доцент

Навчальна програма затверджена на засіданні кафедри математичного забезпечення комп'ютерних систем

Протокол № 1 від. " 31 " 08 2017 р.

Завідувач кафедри _____ (Є.В. Малахов)
(підпис) (прізвище та ініціали)

Обговорено та рекомендовано до затвердження навчально-методичною комісією (НМК) з ІТ факультету МФІТ

Протокол № 1 від. " 11 " 09 2017 р.

Голова НМК _____ (А.Л. Рачинська)
(підпис) (прізвище та ініціали)

Вступ

Навчальна програма дисципліни «Інтелектуальний аналіз даних і методи machine learning» складена відповідно до освітньо-професійної програми підготовки бакалаврів спеціальності 126 – Інформаційні системи та технології (версія **2017** р.).
Предметом дисципліни є теоретичні основи прикладного інтелектуального статистичного аналізу даних (Data Mining) і технології їх застосування в системах прийняття рішень
Місце дисципліни і значення в структурі навчального процесу.
Вивчення курсу та виконання лабораторних робіт базується на знанні програмування, базовому математичному апараті, в першу чергу лінійна алгебра, теорія ймовірності, що викладаються у дисциплінах «Алгоритмізація та програмування», «Об'єктно-орієнтоване програмування», «Вища математика», «Дискретна математика та математична логіка», «Теорія ймовірностей, ймовірнісні процеси та математична статистика».
Програма навчальної дисципліни складається з таких змістових модулів:
1. Основи інтелектуального аналізу даних.
2. Методи інтелектуального аналізу даних.
3. Нейромережеве моделювання і глибоке навчання.

1. Мета та завдання навчальної дисципліни

Метою вивчення дисципліни є формування у студентів уявлення про набори технологій та методів інтелектуального аналізу даних (Data Mining), призначених для обробки більших об'ємів вхідних даних за рахунок автоматизації процесу виведення нових, коректних та потенційно корисних знань, які можуть бути використані для подальшого прийняття рішень.
Завдання:
– вивчення методів та підходів інтелектуального аналізу даних,
– вивчення можливостей практичної реалізації моделей аналізу та обмеженості їх застосування,
– ознайомлення з способами та методами вибору найбільш відповідних методів для вирішення практичних завдань.
Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей (**у відповідності до освітньо-професійної програми для першого (бакалаврського) рівня вищої освіти спеціальності 126 «Інформаційні системи та технології» 2017року**)**:
1) загальних:
ЗК.01 Здатність до дослідницької роботи
ЗК.03 Здатність аналізувати та синтезувати науково-технічну, природничо-наукову та загальнонаукову інформацію
ЗК.07 Базові знання в області фундаментальної та прикладної математики та уміння їх застосовувати в науково-дослідній і професійній діяльності
ЗК.09 Знання психологічних закономірностей інформаційної взаємодії людини з технічними пристроями і програмним забезпеченням комп'ютерних систем, знання основ мікроергономіки та ергономічних стандартів

2) спеціальних (фахових):

ФК.01 Здатність до математичного та логічного мислення, знання основних понять, положень і методів фундаментальної математики та вміння їх використовувати під час розв'язання конкретних виробничих завдань

ФК.03 Знання закономірностей випадкових явищ і вміння застосовувати ймовірносно-статистичні методи для вирішення професійних завдань

ФК.06 Знання основних підходів, методів і технологій штучного інтелекту, уміння розробляти та застосовувати моделі відображення знань, стратегії логічного виведення, технологій інженерії знань, технологій і інструментальних засобів побудови інтелектуальних систем

ФК.11 Знання сучасних технологій та інструментальних засобів розробки програмних систем, уміння їх застосовувати на всіх етапах життєвого циклу

ФК.19 Знання концепцій сховищ даних, їх оперативної аналітичної обробки та інтелектуального аналізу; уміння виявляти в даних раніше не відомих знань, необхідних для прийняття рішень в різних сферах професійної діяльності

ФК.28 Ґрунтовна математична підготовка, а також підготовка з теоретичних, методичних і алгоритмічних основ інформаційних технологій для використання математичного апарату під час вирішення прикладних і наукових завдань в області інформаційних систем і технологій

Програмні результати навчання:

ПРН 06. Уміння використовувати основні поняття, ідеї та застосовувати методи фундаментальної та прикладної математики під час розв'язання конкретних задач в сфері інформаційних технологій.

ПРН 09. Уміння розробляти та застосовувати моделі представлення знань, стратегії логічного виведення, технологій інженерії знань, технологій і інструментальних засобів побудови інтелектуальних систем і систем штучного інтелекту

ПРН 20. Уміння моделювати системи та процеси, стани та поведінки складних об'єктів інформатизації в процесі розроблення інформаційних систем і технологій

ПРН 21. Уміння застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем

Очікувані результати навчання. У результаті вивчення навчальної дисципліни студент повинен

***знати**: основні поняття та підходи інтелектуального аналізу даних; основні стадії проведення інтелектуального аналізу даних; методи інтелектуального аналізу даних; способи підвищення ефективності прийняття рішення за рахунок застосування методів інтелектуального аналізу даних;*

***вміти**: вибирати засоби аналізу, найбільш ефективні для конкретних даних з урахуванням їх природи, погрешності, просторового та тимчасового вирішення, а також завдань дослідження; правильно розуміти та інтерпретувати отримані результати дослідження; вміти критично оцінювати можливості та обмежування використованих методів.*

На вивчення навчальної дисципліни відводиться 210 годин, що становить 7 кредити ЄКТС.

2. Зміст навчальної дисципліни

Змістовний модуль 1. Основи інтелектуального аналізу даних.
Тема 1. Основні поняття технологій збору і систематизації даних.
Література: [1, 3].
Тема 2. Передобробка даних, візуалізація, первинний статистичний аналіз.
Література: [1, 2, 4, 23].
Тема 3. Основні завдання та класифікація методів їх вирішення.
Література: [1, 3, 4].

Змістовний модуль 2. Методи інтелектуального аналізу даних.
Тема 1. Методи класифікації.
Література: [1,2, 4, 7].
Тема 2. Методи кластерного аналізу.
Література: [3, 6, 7].
Тема 3. Методи кореляційного і регресійного аналізу.

Література: [5, 6, 8].

Тема 4. Аналіз і прогнозування часових рядів.
Література: [1, 7, 9].

Змістовний модуль 3. Нейромережеве моделювання і глибоке навчання.
Тема 1. Нейронні мережі прямого поширення.
Література: [1,5, 6,8].
Тема 2. Рекурентні і згорткові нейронні мережі.
Література: [5, 9, 13].

3. Рекомендована література

3.1. Основна література

1. Чубукова И. А. Data Mining : учебное пособие / И. А. Чубукова. - Москва: ИНТУИТ, БИНОМ. Лаб. знаний, 2008.
2. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. - 3-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2009.
3. Методы и модели анализа данных: OLAP и Data Mining : учебное пособие / А. А. Барсегян [и др.]. - Санкт-Петербург: БХВ-Петербург, 2004.
4. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP : учебное пособие / А. А. Барсегян [и др.] . - СПб. : БХВ-Петербург, 2008.
5. Р. Каллан Основные концепции нейронных сетей. : Пер. с англ. - М. : Издательский дом "Вильямс", 2001. - с. : ил. - Парал. тит. англ.
6. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p.

3.2. Допоміжна література

7. А.С.Бирюков, В.В.Рязанов, А.С.Шмаков. Решение задач кластерного анализа коллективами алгоритмов. Журнал вычислительной математики и математической физики, Т.48, 2008, N 1, стр. 176-192.
8. С. Хайкин - Нейронные сети: полный курс, 2-е издание. : Пер. с англ. - М. : Издательский дом "Вильямс", 2006. - 1104 с. : ил. - Парал. тит. англ.
9. Саттон Р. С., Барто Э. Дж. Обучение с подкреплением. М.: ДМК Пресс, 2020.
10. С.Рассел, П.Норвиг Искусственный интеллект: современный подход, 2-е изд.. : Пер. С англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с.: ил.
11. Смолин Д. В. Введение в искусственный интеллект: конспект лекций. - М.: ФИЗМАТЛИТ, 2004. - 208 с. - ISBN 5-9221-0513-2.
12. Н.В.Ковшов, В.Л.Моисеев, В.В.Рязанов. Алгоритмы поиска логических закономерностей в задачах распознавания. Журнал вычислительной математики и математической физики, Т.48, 2008, N 2, стр. 329-344.
13. Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. - М.: ФИЗМАТЛИТ, 2003. - 432 с.
14. Штовба С.Д. Муравьиные алгоритмы. Exponenta Pro, 2003г.
15. Лонца А. Алгоритмы обучения с подкреплением на Python / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 286 с.: ил.
16. Анализ данных : стат. и вычисл. методы для науч. работников и инженеров / З. Брандт / Пер. с англ. О.И. Волковой – М.: МИР : АСТ, 2003 . - 686 с. : ил.

4. Електронні інформаційні ресурси

17. Data Mining Labs – Режим доступу: http://dmlabs.org/
18. Encog Machine Learning Framework - Режим доступу: https://www.heatonresearch.com/encog/
19. Neural Networks and Deep Learning - Режим доступу: http://neuralnetworksanddeeplearning.com/
20. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17341
21. Кашина О.А., Миссаров М.Д. Электронный курс "Статистический анализ данных" – Режим доступу: http://zilant.kpfu.ru/course/view.php?id=17260
22. The R Project for Statistical Computing – Режим доступу: http://www.r-project.org/
23. Курс "Data mining" – Режим доступу: http://www.intuit.ru/studies/courses/6/6/info.
24. Курс "Интеллектуальный анализ данных средствами MS SQL Server 2008" – Режим доступу: http://www.intuit.ru/studies/courses/2312/612/info.
25. Программный комплекс RStudio для интеллектуального анализа данных – Режим доступу: http://www.rstudio.com/
26. Методы интеллектуального анализа данных – Режим доступу: https://www.ibm.com/develo perworks/ru/library/ba-datamining-techniques/

5. Форма підсумкового контролю успішності навчання

Другий семестр – екзамен.

6. Методи діагностики успішності навчання

Поточне опитування, оцінка виконання лабораторних та самостійних завдань, підсумкове опитування у вигляді екзамену.

Примітки:

1. Програма навчальної дисципліни розробляється вищим навчальним закладом. Програма навчальної дисципліни визначає місце і значення навчальної дисципліни, її загальний зміст та вимоги до знань і вмінь.
2. Програма навчальної дисципліни розробляється на основі освітньо-професійної програми.
3. Форма призначена для складання робочої програми навчальної дисципліни.
4. Вищими навчальними закладами можуть вноситися зміни до форми та змістового наповнення «Програми навчальної дисципліни» залежно від специфіки та профілю вищого навчального закладу.
5. Формат бланка – А4 (210×297 мм).