

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Інститут високих технологій

Кафедра молекулярної біотехнології та біоінформатики



«ЗАТВЕРДЖУЮ»

Заступник директора
з науково-педагогічної роботи
Галина ГРАБЧУК

« 03 » 2021 року
протокол 09

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Машинне навчання

(повна назва дисципліни)

для студентів

галузь знань	09 Біологія <i>(шифр і назва)</i>
спеціальність	091 Біологія <i>(шифр і назва спеціальності)</i>
освітній рівень	магістр <i>(молодший бакалавр, бакалавр, магістр)</i>
освітня програма	Біоінформатика та структурна біологія <i>(назва освітньої програми)</i>
вид дисципліни	<u>вибіркова</u>

Форма навчання	денна
Навчальний рік	2021/2022
Семестр	3
Кількість кредитів ECTS	4.0
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	іспит

Викладач: Старосила Сергій Анатолійович,
к.б.н., технічний директор компанії «Рецептор AI»

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

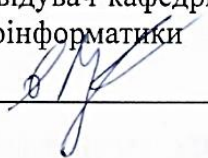
КИЇВ – 2021

Розробники:

Старосила Сергій Анатолійович, к.б.н., технічний директор компанії «Рецептор АІ»

«ЗАТВЕРДЖЕНО»

Завідувач кафедри молекулярної біотехнології та
біоінформатики


Олексій НИПОРКО


Протокол № 4 від «05» лютого 2021р.

Схвалено науково - методичною комісією

«Інституту високих технологій»

Київського національного університету імені Тараса Шевченка

Протокол від «05» 03 2021 року № 3

Голова науково-методичної комісії  (Русінчук Н.М.)

«05» 03 2021 року

ВСТУП

1. Мета дисципліни – ознайомлення студентів з методами штучного інтелекту та машинного навчання для розробки ліків та інших завдань медичної хімії і медицини; навчання базовим навичками використання сучасних типових алгоритмів та програмних середовищ у цій галузі.

2. Попередні вимоги до опанування або вибору навчальної дисципліни (за наявності):

1. Мати базові знання з молекулярної біології, біохімії та хемоінформатики.
2. Мати знання з вищої математики та математичного аналізу.
3. Володіти навичками складання алгоритмів та програмування на поширених мовах програмування.

3. Анотація навчальної дисципліни:

Предметом навчальної дисципліни є методи штучного інтелекту та машинного навчання для розробки ліків та інших завдань, що використовуються в сучасній медичній хімії і медицині з метою розробки біологічно активних хімічних сполук щодо різних молекулярних мішеней, оптимізації цих сполук до кандидатів у ліки, генерування віртуальних баз хімічних сполук, розпізнавання і класифікації біологічних чи медичних зображень, розпізнавання і упорядкування структурованих і не структурованих даних.

В курсі робиться короткий огляд програмного забезпечення та основних алгоритмів штучного інтелекту та машинного навчання, що використовуються в сучасній медичній хімії і медицині, а також детально розглядаються найбільш поширені підходи: штучні нейронні мережі, Бассова статистика, метод опорних векторів, випадковий ліс, дерева рішень, регресії, глибинне навчання, генетичні алгоритми, NLP і інші. Розглядаються основні типи задач: класифікація, регресія, кластерування, розпізнавання.

4. Завдання (навчальні цілі):

Навчання дисципліні має на меті розвинути у студентів такі компетентності:

ЗК2. Здатність генерувати нові ідеї (креативність).

СК2. Здатність застосовувати знання у професійній діяльності з урахуванням новітніх досягнень, у т.ч. для дослідницької роботи.

СК3. Здатність використовувати знання й практичні навички в галузі біологічних наук та на межі предметних галузей для виконання професійних завдань, у т.ч. для дослідження різних рівнів організації живих організмів, біологічних явищ і процесів.

СК5. Здатність аналізувати шляхи розвитку сучасної біології.

СК11. Вміння формулювати задачі моделювання, створювати моделі об'єктів і процесів у живих організмах та їхніх компонентах із використанням математичних методів й інформаційних технологій.

СК12. Вміння розробляти програмне забезпечення для обробки біомолекулярних даних.

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			
1.1	Знати основні алгоритми машинного навчання та методи збору, аналізу, класифікації і аугментації даних; обчислення і вибору дескрипторів; оцінки якості моделей машинного навчання.	Лекції	Модульна контрольна робота: 1 запитання	5%
1.2	Знати як розробляються основні моделі машинного навчання, розуміти межі їх застосовності, точність.	Лекції	Модульна контрольна робота: 2-3 запитання	10%

1.3	Знати як методами машинного навчання розробляються біологічно активні хімічні сполуки, генеруються віртуальні бази хімічних сполук, розпізнаються і класифікуються біологічні чи медичні зображення, розпізнаються і упорядковуються структуровані та не структуровані дані.	Лекції	Модульна контрольна робота: 4-5 запитання	10%
1.4	Знати перелік задач сучасної медичної хімії і медицини, для розв'язання яких застосовуються методи машинного навчання.	Самостійна робота студента	Доповідь під час інтерактивних лекцій: зміст	7%
2.1	Вміти застосовувати спеціальне програмне забезпечення по машинному навчанню для розв'язання задач медичної хімії і медицини.	Практичні роботи	Семестрова робота студента: опис результатів	32%
3.1	Вміти донести інформацію про постановку задач сучасної медичної хімії та медицини і методи машинного навчання для їх розв'язання до аудиторії.	Самостійна робота студента	Доповідь під час інтерактивних лекцій: якість представлення	5%
4.1	Прийняти і обґрунтувати рішення щодо вибору методу.	Інтерактивні лекції, практичні заняття, самостійна робота студента	Семестрова робота студента: обґрунтування методів Розв'язання задачі на модульній контрольній роботі	16% 15%

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни	1.1	1.2	1.3	1.4	2.1	3.1	4.1
Програмні результати навчання							
ПРН2. Використовувати бібліотеки, інформаційні бази даних, інтернет ресурси для пошуку необхідної інформації.				+	+	+	+
ПРН4. Представляти результати наукової роботи письмово (у вигляді звіту, наукових публікацій тощо) та усно (у формі доповідей та захисту звіту) з використанням сучасних технологій, коректно вести дискусію.	+	+	+		+		
ПРН10. Вміти моделювати основні процеси дослідження з метою вибору методів дослідження, апаратного забезпечення або створення нових методик.					+		+
ПРН11. Вміти проводити статистичну обробку, аналіз та узагальнення отриманих експериментальних даних із використанням програмних засобів та сучасних інформаційних технологій, що використовують в галузі біології.					+		+
ПРН16. Застосовувати педагогічні технології на рівні достатньому для реалізації розроблених програм навчальних дисциплін за спеціалізацією у вищих навчальних закладах.	+	+	+	+	+		+
ПРН17. Моделювати об'єкти і процеси у живих організмах та їхніх компонентах із використанням математичних методів й інформаційних технологій.	+	+	+	+	+	+	+

7. Схема формування оцінки.

7.1 Форми оцінювання студентів:

- семестрове оцінювання:

1. Самостійна семестрова робота: РН 2.1. - 48 балів/32 бали.

2. Доповідь під час лекції: РН 1.4,3.1 - 12 балів/6 балів.

- підсумкове оцінювання: іспит – 40 балів.

Усього: 100 балів/60 балів.

Оцінювання	Min	Max
Семестрове оцінювання	36	60
Іспит	24	40
Всього	60	100

7.2 Організація оцінювання:

На початку семестру студенти отримують теми для підготовки коротких (3-5 хвилин) доповідей під час проведення лекцій. Починаючи з 2 лекційного заняття студенти роблять свої доповіді із використанням будь-яких методів та форм представлення інформації: зміст доповіді оцінюється з точки зору її новизни, актуальності, науковості (використання наукових джерел інформації) та повноти викладення у 7 балів, а якість донесення інформації до аудиторії – у 5 балів. Мінімум за дане завдання – 6 балів – може бути отримано за умови підготовки доповіді у текстовому форматі з презентацією без усної доповіді під час лекції.

Після першого практичного заняття студенти отримують індивідуальні завдання до семестрової роботи. Кожне завдання складається з 6 частин. Кожна окрема частина присвячена темі окремого практичного заняття з першого по шосте. Кожна частина оцінюється в 8 балів: 5 бали за правильність розв'язку та 3 бали за обґрунтування методів розв'язку, викладене у звіті. У випадку помилок у розв'язку завдання студенту дозволяється їх виправити після перевірки. Мінімальну кількість балів студент може отримати у випадку правильного розв'язання усіх завдань без представлення їх письмового обґрунтування.

У кінці семестру після завершення вивчення тем проводиться іспит. Іспит спрямований на визначення рівня знань студентами матеріалів лекцій за весь семестр та вміння застосовувати отримані знання до розв'язання задач. Результати іспиту студенти дізнаються у той самий день. Іспит вважається складеним, якщо студент надав хоча короткі відповіді на 60 % запитань. У цьому випадку він отримує мінімально можливу оцінку за іспит – 24 бали.

Вважається курс пройденим лише за умови успішного виконання кожного з трьох оцінювань хоча б на мінімально можливий бал: підготовку доповіді, розв'язання семестрової роботи та складання іспиту.

7.3 Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59
Зараховано / Passed	60-100
Не зараховано / Fail	0-59

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

№ п/п	Назва теми	Кількість годин		
		лекції	лабораторні заняття	самостійна робота
Частина 1 Теоретичне навчання				
1	Вступ. <i>Тема 1 Вступ. Короткий історичний огляд. Поняття штучного інтелекту, машинного навчання, глибокого навчання. Огляд методів машинного навчання, їх межі застосовності, точність.. Задачі і застосування машинного навчання. Програмне забезпечення для машинного навчання.</i>	2	3	11
2	Тема 2. <i>Scikit-learn. Збір, аналіз, нормалізація і класифікація даних. Баланс класів даних у навчальній вибірці. Аугментація. Обчислення і вибір дескрипторів. Розробка моделей машинного навчання, оцінка їх якості.</i>	2	3	11
3	Тема 3. <i>Регресії. Древа рішень. Метод опорних векторів. Випадковий ліс. Генетичні алгоритми.</i>	2	3	12
4	Тема 4. <i>Градiєнтний бустинг. Баєсова статистика. Глибоке навчання. Стекінг моделей.</i>	2	3	12
5	Тема 5. <i>Використання штучного інтелекту та машинного навчання для розробки ліків. RDKit. Розробка моделей машинного навчання та віртуальний скринінг ними. Вирішення інших біо- і хемоінформатичних задач.</i>	2	3	12
6	Тема 6. <i>Штучні нейронні мережі, їх типи. PyTorch. Функції активації. Градiєнтний спуск. Застосування нейромереж для віртуального скринінгу, для генерування нових молекул. Змагальні генеративні нейронні мережі.</i>	2	4	12
7	Тема 7 <i>Трансформери. Застосування нейромереж для розпізнавання і класифікації біологічних чи медичних зображень. Машинне навчання з підкріпленням. Обробка природної мови (NLP).</i>	2	4	12
11	Модульна контрольна робота		1	
	ВСЬОГО	14	24	82

Загальний обсяг 120 год., в тому числі:

Лекцій – **14 год.**

Семінари - **24 год.**

Самостійна робота - **82 год.**

9. Рекомендовані джерела:

Основна:

1. Ching, T. et al. Opportunities and obstacles for deep learning in biology and medicine. (2018). J. R. Soc. Interface 15, 20170387
2. Chicco D "Ten quick tips for machine learning in computational biology". (2017). BioData Mining. 10 (35): 35. doi:10.1186/s13040-017-0155-3.
3. Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, et al. "Machine learning in bioinformatics". (2006). Briefings in Bioinformatics. 7 (1): 86–112. doi:10.1093/bib/bbk007
4. Pérez-Wohlfeil E, Torrenoa O, Bellis LJ, Fernandes PL, Leskosek B, Trellesa O. Training bioinformaticians in High Performance Computing.(2018). Heliyon. 4 (12): e01057. doi:10.1016/j.heliyon.2018.e01057.

Додаткова:

1. Yunyi Wu, Guanyu Wang, Machine Learning Based Toxicity Prediction: From Chemical Structural Description to Transcriptome Analysis. (2018). International Journal of Molecular Sciences 19(8):2358, DOI:10.3390/ijms19082358
2. Marwin H. S. Segler, Thierry Kogej, Christian Tyrchan, and Mark P. Waller. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. (2018). Cent. Sci. 4, 120–131.

10. Додаткові ресурси:

1. <https://www.python.org/>
2. <https://scikit-learn.org/stable/index.html>
3. <https://pypi.org/project/openbabel/>
4. <https://github.com/rdkit/rdkit/blob/master/Docs/Book/GettingStartedInPython.rst>