

Силабуси проходять валідацію у Kharkiv IT Cluster



Шифр та назва спеціальності	123 Комп'ютерна інженерія
Назва освітньої програми	Сучасне програмування, мобільні пристрої та комп'ютерні ігри
Назва дисципліни	Машинне навчання
Викладач	д.т.н, проф. Гавриленко Світлана Юріївна
Вид (основна, вибіркова)	Основна
Блок дисципліни (залишити блок, що відповідає дисципліні)	Машинне навчання. Основи нейроком'ютингу.
Кількість студентів (поточний рік)	20
Курс/Семестр	1/1 , Рівень освіти - магістр
Навчальний рік валідації	2025

ЗАГАЛЬНА ІНФОРМАЦІЯ ПРО ДИСЦИПЛІНУ

Анотація	Курс знайомить студентів із теоретичними основами та алгоритмами машинного навчання, їх можливими практичними реалізаціями, застосуванням в процесі вирішення реальних завдань та формування відповідної компетентності.								
Мета навчальної дисципліни	Цілі навчання: - Ознайомити слухачів з основними парадигмами машинного навчання: навчання з вчителем (supervised learning), навчання без вчителя (unsupervised learning) та навчанням з підкріпленням (reinforcement learning). - Надати глибоке розуміння ключових алгоритмів та їх математичних принципів. - Навчити ефективно готувати дані для аналізу та моделювання. - Розвинути навички вибору, навчання та оцінки моделей машинного навчання. - Отримати практичний досвід роботи з популярними бібліотеками Python (NumPy, Pandas, Scikit-learn) у середовищі Google Colab.								
Типи занять та контролю	Лекції, лабораторні роботи, самостійна робота. Підсумковий контроль – іспит (необхідне зазначити)								
Загальний обсяг (кредитів)		Лекції (занять)	1	Лабораторні (занять)	1	Практичні (занять)	-	Самостійна (годин)	3
Попередні дисципліни									

МАТЕРІАЛЬНО-ТЕХНІЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДИСЦИПЛІНИ

Середовище Google Colab для Python

СТРУКТУРА ТА ЗМІСТ ДИСЦИПЛІНИ

№	Теоретична складова Назва, перелік питань або анотація лекції	Годин	Практична складова Опис та приклад завдання, а також посилання на методичні матеріали	Годин	Інструменти, засоби та технології
Тема 1 – Вступ					
1	Вступ. Інформаційні технології. Складові інформаційних технологій. Машинне навчання (МН). Історія машинного навчання. Сфери застосування МН. Алгоритми машинного навчання. Методи машинного навчання. Етапи машинного навчання. Інструменти машинного навчання. Середовище Google Colab для Python.	2	Середовище Google Colab для Python. Поняття DataFrame. Найпростіші операції: Створення DataFrame; Створення DataFrame: передавання об'єкту dictionary у конструктор DataFrame(); штучна генерація; завантаження файлів даних. Маніпуляції з даними: Фільтрація; Конкатенація; Сорткування; Вибірка; Групування; Агрегація даних та ін. Створення нових стовпців, їх перейменування, видалення, зміна типу. Створення зведених таблиць. Приклади програмного забезпечення по роботі з DataFrame. <i>Лабораторна робота 1.</i> Робота з DataFrame. (Методичні матеріали: керівництво до лабораторної роботи 1.)	4	Середовище Google Colab для Python Демонстраційна програма: 1_DataFrame_ML.ipynb)
Тема 2 – Візуалізація даних					
2	Техніки візуалізації даних: line plot, scatter plot, bar chart, pie chart, spectrogram, Box-plot, 3D-plot та ін. Бібліотеки візуалізації даних: Matplotlib, Pandas, Seaborn та ін.	4	Техніки візуалізації даних: Графіки (line plot), Стовпчасті діаграми (bar chart) і гістограми (histogram), Діаграми розсіювання (scatter plot), Секторні діаграми (pie chart), Хіт-мапа (heat map), Коробкова діаграма (boxplot), 3D графіки. Параметри налаштування зображень: властивості шрифту, ширина, колір, розмір, лінії сітки, карта кольорів. Візуальний аналіз даних: дослідження та інтерпретації інформації за допомогою візуальних уявлень. Збереження результатів візуалізації. <i>Лабораторна робота 2.</i> Методи візуалізації даних. (Методичні матеріали: керівництво до лабораторної роботи 2.)	4	Середовище Google Colab для Python Демонстраційна програма: 2_Data_Visualization_ML.ipynb)

Тема 3 – Попередня обробка даних

3	Поняття даних. Типи даних: числові та категоріальні. Структуровані та неструктуровані дані. Оцінка та розуміння наявних даних. Характеристики великих даних: Об'єм (Volume), Variety (Різноманітність), Velocity (Швидкість), Veracity (Достовірність).	2	Методи попередньої обробки даних: Обробка пропущених значень (Handling Missing Values), Масштабування та трансформація ознак (Feature Scaling and Transformation), Кодування категоріальних ознак (Encoding Categorical Features), Створення нових ознак (Feature Engineering), Обробка шумів та викидів (Handling Noisy Data and Outliers), Відбір ознак (Feature Selection), Розподіл даних (Data Splitting). <i>Лабораторна робота 3. Попередня обробка даних</i> (Методичні матеріали: керівництво до лабораторної роботи 3.)	2	Середовище Google Colab для Python Демонстраційна програма: 3_Data_Preprocessing_ML.ipynb)
4	Етапи попередньої обробки даних: Data cleaning, Data transformation, Data reduction, Data Integration. Методи попередньої обробки даних. Методи балансування даних: Undersampling, Oversampling, SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE	2	Дослідження методів балансування даних: Undersampling, Oversampling, SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE. Налаштування параметрів моделей. Порівняння та візуалізація результатів дослідження. <i>Лабораторна робота 3. Попередня обробка даних</i> (Методичні матеріали: керівництво до лабораторної роботи 3.)	2	Середовище Google Colab для Python Демонстраційна програма: 4_Imbalanced_Data.ipynb)

Тема 4 – Оцінка якості моделей

5	Матриця невідповідності (confusion matrix). Помилки першого та другого роду. Основні показники ефективності моделі (key performance indicators, KPI): Точність (Accuracy), Влучність (Precision), Повнота (Recall). Влучність та повнота: перетягування каната. Міра F1score, ROC-крива. Метрики оцінки якості моделей регресії: Середня абсолютна помилка (MAR, Mean absolute error), Середньоквадратична помилка (RMSE, Root mean squared error), Відносна абсолютна помилка (RAE, Relative absolute error), Коефіцієнт Каппа (Kappa statistic), Відносна	2			Середовище Google Colab для Python Демонстраційна програма: 5_Classification_Performance.ipynb
---	---	---	--	--	--

Силабуси проходять валідацію у Kharkiv IT Cluster

	квадратична помилка (RRSE, Root relative squared error) Бібліотеки Python для оцінки якості моделей класифікації та прогнозування даних.				
Тема 5 – Навчання з вчителем. Класифікація даних					
6	Класичні методи класифікації даних: К-найближчих сусідів, Наївний Байєс, Метод опорних векторів, Логістична регресія та ін. Бібліотеки Python для класифікації даних.	2	Побудова класифікаційних моделей на основі класичних методів. Дослідження впливу методів попередньої обробки даних на якість функціонування моделі. Дослідження впливу налаштування параметрів моделі на якість класифікації даних. Оцінка якості моделі. Візуалізація результатів дослідження. <i>Лабораторна робота 4.</i> Побудова класифікаційних моделей на основі класичних методів. (Методичні матеріали: керівництво до лабораторної роботи 4.)	2	Середовище Google Colab для Python Демонстраційні програми: 6_1_K-Nearest Neighbors.ipynb, 6_2_SVC_Iris_Kernel.ipynb, 6_3_SVM_ROC_Multi_class.ipynb,
7	Дерева рішень. Алгоритми побудови дерев рішень: ID3;C4.5; CART. Налаштування дерев рішень. Перенавчання дерев рішень. Бібліотеки Python для побудови моделей на основі дерев рішень. Налаштування моделей.	2			
8	Ансамблеві методи класифікації даних. Алгоритми формування вибірок даних: Pasting, Bootstrapping, Random Subspaces, Random Patches. Мета-алгоритм Bagging. Алгоритм Random Forest (RF). Дослідження параметрів налаштування моделі RF (criterion, max_features, min_samples_leaf, max_depth, n_estimators) на якість класифікації.	2	Побудова моделі класифікації даних на основі мета-алгоритму Bagging. Вибір базового класифікатора. Дослідження впливу налаштування параметрів ансамблю на якість класифікації даних. Візуалізація результатів дослідження. <i>Лабораторна робота 5.</i> Побудова класифікаційних моделей на основі мета-алгоритму Bagging. (Методичні матеріали: керівництво до лабораторної роботи 5.)	2	Середовище Google Colab для Python Демонстраційні програми: 7_Isolation_Forest.ipynb, 8_Baggine_K_Fold.ipynb)
9	Ансамблеві методи класифікації даних. Мета-алгоритми: Boosting, Stacking. Алгоритми AdaBoost та Blending). Бібліотеки Python по роботі з ансамблями.	2	Побудова моделі класифікації даних на основі мета-алгоритму Boosting. Вибір базового класифікатора. Дослідження впливу налаштування параметрів ансамблю на якість класифікації даних. Візуалізація результатів	4	Середовище Google Colab для Python Демонстраційні програми: 9_Adaboost.ipynb,

			дослідження. Побудова моделі класифікації даних на основі мета-алгоритму Stacking. Вибір базового класифікатора. Дослідження впливу налаштування параметрів ансамблю на якість класифікації даних. Візуалізація результатів дослідження. <i>Лабораторна робота 6.</i> Побудова класифікаційних моделей на основі мета-алгоритмів Boosting та Stacking. (Методичні матеріали: керівництво до лабораторної роботи 6.)		9_Gradient_Boosting.ipynb, 9_Stacking_1.ipynb
Тема 6 – Прогнозування даних					
10	Методи прогнозування Лінійна та множинна регресія, Дерева рішень для прогнозування, SVM алгоритм. Метрики якості для задач регресії. Бібліотеки Python для прогнозування даних.	2	Побудова прогностичної моделі. Лінійна регресія. Поліноміальна регресія. Оцінка якості прогностичної моделі. Дослідження впливу налаштування параметрів моделі на якість прогнозу. Візуалізація результатів дослідження. Дерево регресії. Оцінка якості прогностичної моделі. Дослідження впливу налаштування параметрів моделі на якість прогнозу. Візуалізація результатів дослідження. <i>Лабораторна робота 7.</i> Побудова прогностичної моделі. (Методичні матеріали: керівництво до лабораторної роботи 7.)	4	Середовище Google Colab для Python Демонстраційні програми: 10_Linregress_visualization.ipynb 10_Regr_SVM_Compare_diff_methods.ipynb
Тема 7 – Навчання без вчителя					
11	Кластеризація даних. Методи: K-середніх, DBSCAN. Бібліотеки Python з кластеризації даних. Скорочення розмірності простору ознак. Метод головних компонент (PCA), Сингулярне розкладання (SVD), Латентний розміщення Дирихле (LDA), Латентно-семантичний аналіз (LSA, pLSA, GLSA), t-SNE метод	4	Побудова моделі кластеризації даних на основі методів: K-середніх та DBSCAN. Оцінка якості кластеризації на основі показників: Силуетна оцінка, Індекс Calinski Harabaz, Індекс Девіса Боулдіна. Дослідження впливу налаштування параметрів моделі на якість кластеризації даних. Виконання порівняльного аналізу якості різних типів моделей кластеризації даних. Візуалізація результатів дослідження. Стиснення даних. Метод головних компонент (PCA). <i>Лабораторна робота 8.</i> Побудова моделі кластеризації даних (Методичні матеріали: керівництво до лабораторної роботи 8.)	4	Середовище Google Colab для Python Демонстраційні програми: 11_Clustering_K-means.ipynb, 11_K-means_visualization.ipynb, 11_2_DBSCAN_.ipynb
12	Пошук асоціативних правил. Асоціативні правила (Association Rules). Підтримка (Support). Достовірність	2			Середовище Google Colab для Python

Силабуси проходять валідацію у Kharkiv IT Cluster

(confidence). Цікавість (lift). Узагальнені асоціативні правила (Generalized Association Rules)ю Алгоритм Apriori			Демонстраційна програма: 12_Association.ipynb
---	--	--	---

Тема 8 – Виявлення аномалій в даних

13	Визначення аномалій. Типи аномалій. Детектування викидів (Outlier Detection) і «новизни» (Novelty Detection). Методики виявлення аномалій в даних: класифікація, статистичний аналіз, кластеризація, спектральні методи, гібридні методи. Алгоритм Isolation Forest.	2	Побудова моделей виявлення аномалій в даних на основі: класифікації, статистичного аналізу, кластеризації, спектральних методів, гібридних методів. Алгоритм Isolation Forest. Дослідження впливу налаштування параметрів моделі на якість прогнозу. Візуалізація результатів дослідження. <i>Лабораторна робота 9.</i> Побудова моделей виявлення аномалій в даних. (Методичні матеріали: керівництво до лабораторної роботи 9.)	4	Середовище Google Colab для Python. Демонстраційна програма: 13_Anomaly_Detection.ipynb
----	--	---	--	---	---

Тема 9 – Навчання з підкріпленням

14	Навчання з підкріпленням. Функція цінності. Методи навчання з підкріпленням. Використання навчання з підкріпленням для побудови гральних рушіїв, автономних транспортних засобів та в робототехніці. Бібліотеки Python для навчання з підкріпленням.	2			Середовище Google Colab для Python. Демонстраційна програма: 14_Reinforcement_learning.ipynb
----	--	---	--	--	--

ТЕМИ ТА ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

№	Назва та опис завдання	Методи контролю та критерії оцінювання	Годин
1	Опрацювання лекційного матеріалу.	Тестування, отримання балів за тести.	22
2	Підготовка до лабораторних занять.	Підготовка звіту у вигляді презентації, програмна реалізації та захист робіт. Отримання балів за кожну лабораторну роботу.	64

ПРОЄКТ (за наявності)

№	Назва та опис завдання	Метод контролю та захисту	Строки виконання

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ ТА НАВЧАЛЬНІ МАТЕРІАЛИ

Основні

Силабуси проходять валідацію у Kharkiv IT Cluster

№	Назва	До теми (вказати номер)
1	Гавриленко С.Ю. Машинне навчання. Конспект лекцій, Х.: НТУ «ХПІ», 2024, 232 с., https://repository.kpi.kharkov.ua/handle/KhPI-Press/80167	1-14
2	Гавриленко С.Ю. Методичні вказівки до виконання розрахункового завдання з дисципліни «Машинне навчання», Х.: НТУ «ХПІ», 2024, 28 с., https://repository.kpi.kharkov.ua/handle/KhPI-Press/80022	1-4,6-11
3	Гавриленко С.Ю., Челак В.В., Горносталь О.А., Зозуля В.Д. Машинне навчання. Лабораторний практикум, Х.: НТУ «ХПІ», 2022, 86 с. https://repository.kpi.kharkov.ua/items/c5ca20d3-d3e5-49e0-99b6-6f8a02b68b57	1-4, 6-8
4	Штовба С.Д., Козачко О.М. Machine learning: навч. посіб., Вінниця : ВНТУ, 2020, 81 с. https://www.researchgate.net/publication/338924246_Machine_Learning_startovij_kurs	2, 4, 6, 7
5	В. О. Харченко Основи машинного навчання : навч. посіб., Суми : Сумський державний університет, 2023, 264 с. https://essuir.sumdu.edu.ua/bitstream-download/123456789/92711/3/Kharchenko_osnovy.pdf;jsessionid=1DA034D950DD8F94316A568048FA12D7	1-14
6	Олещенко Л.М. Машинне навчання: комп'ютерний практикум, Київ: КПІ ім. Ігоря Сікорського, 2022, 92 с. https://ela.kpi.ua/server/api/core/bitstreams/8548df9d-941e-4284-ba2d-38006742feef/content	7,11
7	Т.А. Желдак, О.Б. Владико. Машинне навчання [Електронний ресурс] : конспект лекцій для здобувачів ступеня магістра зі спеціальності 124 Системний аналіз, Нац. техн. ун-т «Дніпровська політехніка», Дніпро : НТУ «ДП», 2024, 162 с. https://ir.nmu.org.ua/server/api/core/bitstreams/0c632381-6efa-40fe-a27a-0352138b8a43/content	6,7,8,9, 11

Додаткові

№	Назва	До теми (вказати номер)
1	S. Gavrylenko, V. Zozulia, and N. Khatsko. Methods for Improving the Quality of Classification on Imbalanced Data, Proceedings of the IEEE 4th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2023, pp. 1-5, doi: https://ieeexplore.ieee.org/document/10312879	3,4
2	S. Gavrylenko, O. Abdulov, "Improving the quality of payment fraud detection by using a combined approach of transaction analysis", Innovative Technologies and Scientific Solutions for Industries, 2024, No. 4 (30), P. 15–28. DOI: https://doi.org/10.30837/2522-9818.2024.4.031	6,7,11
3	С.Ю. Гавриленко, В.В. Зозуля Дослідження методів виявлення аномалій на етапі попередньої обробки даних. Системи управління, навігації та зв'язку, 2022, випуск 1(67), С. 52-56. doi: 10.26906/SUNZ.2022.1.052 https://journals.nupp.edu.ua/sunz/issue/view/84	3,4
4	S. Gavrylenko , V. Chelak, S. Semenov S. Development of Method for Identification the Computer System State based on the Decision Tree with Multi-Dimensional Nodes, Radio Electronics, Computer Science, Control (RECSC), 2022, No. 2, pp.113-121. doi: 10.15588/1607-3274-2022-2-11 http://ric.zntu.edu.ua/article/view/259438	5,11

Силабуси проходять валідацію у Kharkiv IT Cluster

5	S. Gavrylenko, V. Chelak, O. Hornostal . Research of Methods of Identifying the Computer Systems State based on Bagging , Proceedings of the IEEE 3rd KhPI Week on Advanced Technology (Kharkiv, Ukraine, 2022, pp.705-710 https://ieeexplore.ieee.org/document/9916439	8
6	S. Gavrylenko, V. Zozulia and N. Khatsko, "Method of Detection of Intrusions Based on Stacking Meta-Algorithm," 2024 IEEE 5th KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2024, pp. 1-5, doi: 10.1109/KhPIWeek61434.2024.10878074. https://ieeexplore.ieee.org/document/10878074	9

КОНТРОЛЬНІ ЗАХОДИ

№	Назва та опис	Методи контролю та критерії оцінювання
1	Тест 1. Візуалізація даних.	Автоматизована оцінка відповідей відповідно до завдання тесту.
2	Тест. 2. Попередня обробка та аналіз даних.	Автоматизована оцінка відповідей відповідно до завдання тесту.
3	Тест 3. Навчання з вчителем.	Автоматизована оцінка відповідей відповідно до завдання тесту.
4	Тест. 4. Навчання без вчителя.	Автоматизована оцінка відповідей відповідно до завдання тесту.
5	Тест 5. Навчання з підкріпленням.	Автоматизована оцінка відповідей відповідно до завдання тесту.
6	Захист лабораторних робіт.	За результатами захисту роботи.
7	Іспит (1 семестр)	За результатами співбесіди.

РЕЗУЛЬТАТИ НАВЧАННЯ

Розробляти та досліджувати математичні моделі та методи інтелектуального аналізу даних, алгоритмічне та програмне забезпечення, їх реалізації при розробці ІТ-проектів, мобільних пристроїв та комп'ютерних ігор.

ЗВ'ЯЗОК ІЗ РИНКОМ ПРАЦІ

Спеціальність/професія, підготовці до діяльності в якій читається курс	Аналітик даних (Data Analyst).
Посилання на вакансії (понад 3), та їх скриншоти, в яких вказані компетенції, що формується на курсі, як вимоги до кандидатів	https://robota.ua/zapros/data-analyst/ukraine https://robota.ua/company3338870/vacancy10594567 https://robota.ua/company7819611/vacancy10613694 https://robota.ua/company1020/vacancy10575819 Компетентності: Аналізувати масиви даних і знаходити цінні інсайти.

Силабуси проходять валідацію у Kharkiv IT Cluster

Виявляти патерни підозрілої поведінки користувачів.
Створювати дашборди й контролювати ключові метрики безпеки.
Виявляти аномалії в даних.
Протидіяти спаму, удосконалювати наявні та створювати нові автоматизовані логіки.
Працювати з ML-моделями.

Перелік компетентностей із вказаних як вимоги до вакансії, які набувають студенти, в процесі проходження дисципліни.

Компетентності відповідно до освітньої програми: Здатність до пошуку, оброблення та аналізу інформації з різних джерел. Здатність розробляти, обирати та досліджувати методи, моделі та інформаційні технології інтелектуального аналізу даних, для забезпечення якості прийняття проєктних рішень в предметних областях: сучасне програмування, мобільні пристрої та комп'ютерні ігри.

Інструменти оцінювання результатів навчання за дисципліною*

№	Об'єкт оцінювання (знання методів та принципів, практичні навички, командна робота тощо)	Методи контролю (тести, виконання поточних практичних завдань та їх форма: написання коду, створення діаграми Гантта, створення прототипу тощо)	Інструмент оцінювання (доступ до результатів тесту, гостьова лекція, посилання на виконані завдання, посилання на проєкт, присутність на захисті проєктів, доступ до запису захисту тощо)
1	Оцінювання знань, умінь та перевірка розуміння здобувачами навчального матеріалу з теоретичних основ та алгоритмів машинного навчання.	Тести.	Доступ до результатів тесту. Гостьова лекція.
2	Оцінювання знань, умінь та перевірка розуміння здобувачами навчального матеріалу з практичних основ та алгоритмів машинного навчання.	Виконання поточних лабораторних робіт, індивідуального завдання: написання коду, підготовка звіту та поточний захист робіт згідно графіку.	Посилання на виконані завдання, присутність на захисті робіт.

***Блок. Інструменти оцінювання результатів навчання за дисципліною** — цей блок має на меті показати можливі варіанти проведення валідації за участі експертів від ІТ-компаній. Основна умова проведення валідації це попереднє погодження всіх аспектів з експертом та викладачем. Інструменти, завдяки яким це можна зробити:

- тести за темою (аналогічно технічному інтерв'ю);

Силабуси проходять валідацію у Kharkiv IT Cluster

- поточні завдання впродовж курсу (посилання на код на GitHub-репозитарію студента тощо) ;
- підсумковий проєкт (код, презентація, скринкаст презентації тощо);
- гостьові активності (гостьові лекції, відвідування контрольних активностей тощо).