



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «ІНЖЕНЕРІЯ МАШИННОГО НАВЧАННЯ У ХМАРНИХ СЕРЕДОВИЩАХ»

Компонента освітньої програми – вибіркова (3 кредити)

Освітньо-професійна програма	Інтелектуальний аналіз даних в комп'ютерних інформаційних системах
Спеціальність	122 Комп'ютерні науки
Галузь знань	12 Інформаційні технології
Рівень вищої освіти	другий (магістр)
Мова навчання	українська
Профайл викладача (-ів)	Томка Юрій Ярославович – кандидат фізико-математичних наук, доцент кафедри комп'ютерних наук https://kkn.chnu.edu.ua/about-kafedra/spivrobitnyky/tomka_yurii/
Контактний тел.	+38095-769-21-91
E-mail:	y.tomka@chnu.edu.ua
Сторінка курсу в Moodle	https://moodle.chnu.edu.ua/course/view.php?id=8317
Консультації	У відповідності до сформованого кафедрою графіків консультацій викладача, що оприлюднений на сторінці: https://kkn.chnu.edu.ua/for_student/class-schedule/

АНОТАЦІЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Навчальна дисципліна «Інженерія машинного навчання у хмарних середовищах» – вибіркова освітня компонента ОП «Інтелектуальний аналіз даних в комп'ютерних інформаційних системах» підготовки здобувачів другого (магістерського) рівня вищої освіти

Метою викладання дисципліни “Інженерія машинного навчання у хмарних середовищах” є формування у здобувачів практичних навичок проєктування, розроблення та супроводу систем машинного навчання у хмарних середовищах. Курс охоплює повний ML-lifecycle: від підготовки даних і навчання моделей до автоматизації пайплайнів, розгортання як сервісу та моніторингу у production, з основним акцентом на платформі AWS SageMaker.

Особливу увагу приділено роботі з AWS SageMaker як єдиною платформою для підготовки даних, навчання, оптимізації гіперпараметрів, побудови пайплайнів та розгортання моделей. Розглядаються також інструменти MLOps: MLflow, DVC, Airflow, Prometheus/Grafana, CI/CD для ML-систем.

Основні завдання курсу

Для досягнення поставленої мети курс передбачає вирішення таких ключових завдань:

- Опанування повного ML-lifecycle у хмарі: збір і підготовка даних, навчання та оцінка моделей, автоматизація пайплайнів, розгортання та моніторинг у production.
 - Набуття практичних навичок роботи з AWS SageMaker: SageMaker Studio, Training Jobs, Pipelines, Endpoints, Model Monitor, Feature Store.
 - Вивчення принципів MLOps: версіонування даних і моделей (DVC, MLflow), CI/CD для ML, автоматичне перенавчання та управління реєстром моделей.
- Формування навичок розгортання ML-моделей як хмарних сервісів (REST API, batch inference, real-time endpoints) з урахуванням вимог масштабованості та

надійності.

- Опанування методів моніторингу ML-систем у production: виявлення data drift та model drift, налаштування алертів, аналіз метрик якості моделей.
- Розвиток здатності проєктувати масштабовані ML-системи з урахуванням вимог безпеки, етики та відповідальності у застосуванні штучного інтелекту.
- Набуття досвіду командної розробки ML-проєктів: організація спільної роботи, документування експериментів, підготовка фінального end-to-end ML-проєкту.
-

Об'єкт вивчення

Об'єктом навчальної дисципліни є процеси та методи побудови систем машинного навчання у хмарних середовищах: підготовка та інженерія ознак, навчання та оптимізація моделей, автоматизація ML-пайплайнів, розгортання та моніторинг ML-сервісів у production-середовищах.

Практичний підхід та структура навчання

1. Проєктно-орієнтоване навчання

Кожна тема курсу передбачає практичне завдання або лабораторну роботу безпосередньо у AWS SageMaker Studio. Здобувачі працюють з реальними наборами даних, будують моделі, налаштовують пайплайни та розгортають сервіси у хмарі.

2. Наскрізний проєкт

Протягом курсу здобувачі розробляють наскрізний ML-проєкт – повноцінну систему машинного навчання у хмарі від постановки задачі та підготовки даних до розгортання як сервісу та налаштування моніторингу. Стек: AWS SageMaker, MLflow, Docker, (Python, ArchiMate 3.2, C4-PlantUML, UML, Archi, Enterprise Architect тощо).

НАВЧАЛЬНИЙ КОНТЕНТ ОСВІТНЬОЇ КОМПОНЕНТИ

Змістовий модуль 1. ДАНІ У ХМАРІ: ЗБІР, ЗБЕРІГАННЯ ТА ПІДГОТОВКА	
Тема 1	Вступ до ML-інженерії у хмарі. Огляд платформи AWS SageMaker ● ML Engineering vs Data Science: ролі та відповідальності ● Огляд ML-lifecycle: від даних до production ● AWS SageMaker: архітектура, компоненти, SageMaker Studio ● Хмарне середовище як єдиний простір для ML: переваги та обмеження ● IAM, S3, EC2, VPC: базова інфраструктура AWS для ML ● Jupyter notebooks у SageMaker Studio: налаштування середовища Практика: налаштування SageMaker Studio, перший Training Job на AWS
Змістовий модуль 2. МОДЕЛЮВАННЯ ТА НАВЧАННЯ У ХМАРІ	
Тема 2	Збір, зберігання та підготовка даних у хмарі ● Джерела даних у AWS: S3, RDS, DynamoDB, Kinesis, Glue ● AWS Glue: ETL-пайплайни та каталогізація даних ● SageMaker Data Wrangler: візуальна підготовка та трансформація даних ● SageMaker Feature Store: зберігання та управління ознаками ● Очищення та валідація даних: методи виявлення аномалій та пропусків ● Feature Engineering: кодування, нормалізація, відбір ознак Практика: підготовка набору даних через SageMaker Data Wrangler та запис у Feature Store
Тема 3	Навчання ML-моделей у AWS SageMaker ● SageMaker Training Jobs: конфігурація, instance types, spot instances ● Built-in алгоритми SageMaker: XGBoost, Linear Learner, DeepAR та ін. ● Власні контейнери (BYOC): Docker-образи для навчання у SageMaker ● Розподілене навчання: data parallelism та model parallelism у SageMaker ● Scikit-learn, PyTorch, TensorFlow у SageMaker: estimators та frameworks ● Оцінка та валідація моделей: метрики якості, cross-validation, test sets Практика: навчання класифікатора та регресійної моделі через SageMaker Training Job

Змістовий модуль 3. ML-ПАЙПЛАЙНИ ТА АВТОМАТИЗАЦІЯ	
Тема 4	Оптимізація гіперпараметрів та AutoML • SageMaker Hyperparameter Tuning: Bayesian, Random, Grid Search • SageMaker Autopilot: AutoML без написання коду • Warm pools та managed spot training: оптимізація витрат • SageMaker Experiments: відстеження та порівняння runs • MLflow Tracking: логування параметрів, метрик та артефактів • SageMaker Model Registry: версіонування та approval workflow Практика: запуск Hyperparameter Tuning Job, логування в MLflow, реєстрація моделі
Тема 5	SageMaker Pipelines: побудова ML-пайплайнів • Поняття ML-пайплайну: кроки, залежності, артефакти • SageMaker Pipelines SDK: ProcessingStep, TrainingStep, EvaluationStep, RegisterStep • Умовна логіка у пайплайнах: ConditionStep та управління потоком виконання • Параметризація та повторне використання пайплайнів • Тригери пайплайнів: EventBridge, S3 events, розклад (scheduled runs) • Apache Airflow та AWS MWAA: оркестрація складних ML-воркфлоу Практика: побудова наскрізного SageMaker Pipeline від preprocessing до реєстрації моделі
Тема 6	Версіонування даних, DVC та відтворюваність експериментів • Проблема відтворюваності у ML: причини та підходи до вирішення • DVC (Data Version Control): версіонування даних та моделей разом із Git • SageMaker Experiments: структурування та порівняння серій експериментів • Git-flow для ML-проектів: гілки, теги, pull requests • Артефакти пайплайну: зберігання, lineage, відстеження метаданих • Документування експериментів: MLflow Projects, best practices Практика: версіонування датасету через DVC, відтворення експерименту за commit-хешем
Змістовий модуль 4. РОЗГОРТАННЯ ML-МОДЕЛЕЙ	
Тема 7	Розгортання моделей у SageMaker: endpoints та inference • Типи inference у SageMaker: real-time, batch transform, async, serverless • SageMaker Real-time Endpoints: конфігурація, instance types, auto-scaling • SageMaker Batch Transform: inference на великих обсягах даних • Упакування моделей у Docker: власні inference-контейнери (BYOC) • Multi-model endpoints та multi-container endpoints: оптимізація витрат • Інтеграція endpoint з Lambda та API Gateway: побудова REST API для моделі • Практика: розгортання моделі як real-time endpoint, виклик через API Gateway + Lambda
Змістовий модуль 5. МОНІТОРИНГ ML-СИСТЕМ У PRODUCTION	
Тема 8	MLOps: CI/CD для ML-систем • MLOps: поняття, рівні зрілості (MLOps maturity levels) • CI/CD для ML: GitHub Actions, CodePipeline, CodeBuild, CodeDeploy • Автоматичне тестування ML-коду: unit tests, integration tests, model validation • SageMaker Projects: шаблони MLOps для повної автоматизації • Канаркове розгортання та A/B тестування моделей у SageMaker • Blue/Green deployment: безпечне оновлення production-моделей • Практика: побудова CI/CD пайплайну для автоматичного перенавчання та розгортання моделі
Змістовий модуль 6. СУПРОВІД, ОНОВЛЕННЯ ТА ВІДПОВІДАЛЬНИЙ ML	
Тема 9	Моніторинг якості моделей та виявлення дрейфу • Типи деградації моделей: data drift, concept drift, model staleness • SageMaker Model Monitor: налаштування baseline та моніторингових завдань • Data Quality Monitor, Model Quality Monitor, Bias Monitor, Explainability Monitor • CloudWatch: метрики endpoint, логи, налаштування алертів та дашбордів • Prometheus та Grafana для ML: побудова кастомних дашбордів якості моделей • Аналіз та звітність: автоматичні репорти щодо якості моделей • Практика: налаштування SageMaker Model Monitor, виявлення data drift на тестових

	даних
Тема 10	Версіонування, перенавчання та безперервне вдосконалення моделей • Стратегії перенавчання: scheduled retraining, trigger-based retraining • SageMaker Model Registry: approval workflow, versioning, stage transitions • A/B тестування моделей у production: трафік-спліт, оцінка результатів • Комп'ютерний зір у корпоративних рішеннях • Відповідальний AI: bias, fairness, explainability – SageMaker Clarify • Оптимізація витрат на ML у хмарі: spot instances, savings plans, rightsizing • Практика: фінальний end-to-end ML-проект – від даних до моніторингу у production

ОСВІТНІ ТЕХНОЛОГІЇ, ФОРМИ ТА МЕТОДИ МЕТОДИ НАВЧАННЯ

У процесі вивчення навчальної дисципліни використовуються інноваційні освітні технології: інформаційно-комунікаційні, технології студентоцентрованого навчання; проектна діяльність; традиційні та інтерактивні форми і методи навчання, серед яких: лекція-візуалізація, проблемна лекція, семінар-дискусія, семінар-діалог, самостійно-дослідницька робота, лабораторна робота, аналіз і рішення ситуативних професійних психолого-педагогічних задач (Case study) та ін.

ФОРМИ Й МЕТОДИ КОНТРОЛЮ ТА ОЦІНЮВАННЯ

Поточний контроль: усні та письмові (тестування) відповіді здобувача; демонстрація виконаних практичних завдань у SageMaker; захист лабораторних робіт та фінального end-to-end ML-проекту.

Підсумковий контроль – залік.

КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

Оцінювання програмних результатів навчання здобувачів освіти здійснюється за шкалою європейської кредитно-трансферної системи (ECTS).

Критерієм успішного оцінювання є досягнення здобувачем вищої освіти мінімальних порогових рівнів (балів) за кожним запланованим результатом навчання.

ПОЛІТИКА ЩОДО АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ

Дотримання політики щодо академічної доброчесності учасниками освітнього процесу при вивченні навчальної дисципліни регламентовано такими документами:

- «Етичний кодекс Чернівецького національного університету імені Юрія Федьковича» <https://www.chnu.edu.ua/media/jxdfs0zb/etychnyi-kodeks-chernivetskoho-natsionalnoho-universytetu.pdf> ;

- «Положення про виявлення та запобігання академічного плагіату у Чернівецькому національному університету імені Юрія Федьковича» https://www.chnu.edu.ua/media/f5eleobm/polozhennya-pro-zapobihannia-plahiatu_2024.pdf.

ІНФОРМАЦІЙНІ РЕСУРСИ

Основна література

1. Huyen C. Designing Machine Learning Systems. — O'Reilly Media, 2022. — 394 p. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>
2. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). — O'Reilly Media, 2022. — 861 p.

- <https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>
3. Gift N., Deza A. Practical MLOps. — O'Reilly Media, 2021. — 430 p.
<https://www.oreilly.com/library/view/practical-mlops/9781098103002/>
 4. AWS. Amazon SageMaker Developer Guide. — Amazon Web Services, 2024.
<https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>
 5. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). — O'Reilly Media, 2022. — 861 p.
<https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>
 6. AWS. Amazon SageMaker Developer Guide. — Amazon Web Services, 2024.
<https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>
- Додаткова література**
7. Lakshmanan V., Robinson S., Munn M. Machine Learning Design Patterns. — O'Reilly Media, 2020. — 408 p.
<https://www.oreilly.com/library/view/machine-learning-design/9781098115777/>
 8. Alla S., Adari S.K. Beginning MLOps with MLflow. — Apress, 2021. — 272 p.
<https://link.springer.com/book/10.1007/978-1-4842-6549-9>
 9. Winder P. Reinforcement Learning. — O'Reilly Media, 2020. — 636 p.
<https://www.oreilly.com/library/view/reinforcement-learning/9781492072386/>
 10. Kim G., Humble J. та ін. The DevOps Handbook (2nd ed.). — IT Revolution Press, 2021. — 528 p.
<https://arxiv.org/abs/2209.09125>

Електронні інформаційні ресурси

1. Domain-Driven Design Reference [Електронний ресурс] // Eric Evans. – 2015. – Режим доступу: <https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>
2. eShopOnContainers – DDD Reference Application [Електронний ресурс] // Microsoft. – 2023. – Режим доступу: <https://www.opengroup.org/archimate-forum/archimate-overview>
3. Martin Fowler's DDD Blog [Електронний ресурс] // Martin Fowler. – 2024. – Режим доступу: <https://aws.amazon.com/architecture/well-architected/>
4. Modular Monolith: A Primer [Електронний ресурс] // Kamil Grzybek. – 2020. – Режим доступу: <https://dvc.org/doc>
5. MediatR Documentation [Електронний ресурс] // Jimmy Bogard. – 2024. – Режим доступу: <https://aws.amazon.com/blogs/machine-learning/>
6. Microsoft .NET Architecture Guides [Електронний ресурс] // Microsoft. – 2024. – Режим доступу: <https://airflow.apache.org/docs/>
7. EventStoreDB Documentation [Електронний ресурс] // Event Store Ltd. – 2024. – Режим доступу:
<https://developers.google.com/machine-learning/crash-course>
8. <https://developers.google.com/machine-learning/crash-course>
9. CQRS Journey – Microsoft patterns & practices [Електронний ресурс] // Microsoft. – 2012. – Режим доступу:
<https://www.kaggle.com/learn>
10. <https://www.kaggle.com/learn>
11. Awesome Domain-Driven Design [Електронний ресурс] // GitHub community. – 2024. – Режим доступу: <https://papers.nips.cc/paper/2015/file/86df7dcfd896fcfa2674f757a2463eba-Paper.pdf>
12. Entity Framework Core Documentation [Електронний ресурс] // Microsoft. – 2024. – Режим доступу: <https://www.kaggle.com/learn>

Детальна інформація щодо вивчення курсу «Корпоративні архітектури програмного забезпечення» висвітлена у робочій програмі навчальної дисципліни, що наведена на сторінці:

https://kkn.chnu.edu.ua/for_student/course-programs/

