

Український державний університет залізничного транспорту

Рекомендовано на засіданні
кафедри інформаційних технологій
протокол № 1 від 29 серпня 2024 р.

Силабус з дисципліни
**ТЕХНОЛОГІЇ АНАЛІЗУ ВЕЛИКИХ ОБ'ЄМІВ
ДАНИХ**
2024-2025 навчальний рік, II семестр

Освітній рівень	Перший (бакалаврський) рівень вищої освіти
Галузь знань	12 Інформаційні технології
Спеціальність	126 Інформаційні системи та технології
Освітня програма	Технології штучного інтелекту

Лектор та керівник лабораторних занять:

Іванюк Олександр Ігорович (доцент кафедри ІТ, доктор філософії)

Контакти: +38 (057) 730-19-84, e-mail: ivaniuk@kart.edu.ua

Консультації за дисципліною – понеділок, 12:40-14:00.

Мета курсу – формування у студентів знань і навичок з використання сучасних технологій та інструментів для збору, зберігання, обробки та аналізу великих об'ємів даних.

Анотація

Дисципліна «Технології аналізу великих об'ємів даних» покликана ознайомити студентів з сучасними підходами та інструментами для роботи з великими даними. У процесі навчання студенти вивчатимуть архітектуру систем для обробки великих об'ємів даних, технології зберігання та аналізу, а також методи обробки потокових даних. Особлива увага приділяється використанню хмарних технологій та інструментів, таких як Hadoop, Spark та Apache Kafka, для ефективної роботи з великими наборами даних. Курс надасть студентам практичні навички в налаштуванні інфраструктури та застосуванні сучасних технологій для збору, обробки, очищення та візуалізації великих даних.

Розподіл часу

Кількість кредитів – 6.

Загальна кількість годин вивчення дисципліни – 180.

Кількість годин на проведення лекцій – 30.

Кількість годин на проведення лабораторних робіт – 30.

Кількість годин на самостійну роботу – 120.

Рік, курс та семестр навчання – 2024/2025 н. р., IV курс, II семестр.

Теми

Лекції:

1. Вступ до технологій аналізу великих даних.
2. Архітектура обробки великих даних.
3. Хмарні технології для обробки великих даних.
4. Технології збору та зберігання великих даних.
5. Обробка та очищення великих даних.
6. MapReduce та Hadoop екосистема.
7. Технології обробки потокових даних.
8. Apache Spark.
9. Методи аналізу великих даних.
10. Візуалізація великих даних.
11. Безпека та конфіденційність великих даних.
12. Індустріальні кейси використання великих даних.

Лабораторні роботи:

1. Встановлення та налаштування Hadoop і Spark для обробки великих даних.
2. Збір і підготовка великих даних з використанням NoSQL баз даних.

3. Реалізація алгоритмів MapReduce для обробки великих даних.
4. Аналіз поточкових даних з використанням Apache Kafka.
5. Моделювання та аналіз великих даних за допомогою Apache Spark.

Правила оцінювання

Порядок оцінювання результатів навчання визначається [Положенням про контроль та оцінювання якості знань студентів в УкрДУЗТ](#).

Оцінювання знань студента з дисципліни здійснюється за наступними складовими:

- виконання та захист лабораторних робіт;
- проходження модульного контролю;
- складання іспиту (обов'язкове при отриманні загальної модульної оцінки F, за вибором студента при отриманні загальної модульної оцінки D або B).

Лабораторні роботи можуть бути захищені за умови наявності самостійно та вірно виконаного завдання. Захист лабораторних робіт можливий під час проведення лабораторних занять та у години консультацій, але не більше однієї лабораторної роботи в день.

У випадку пропуску студентом більше ніж двох лекцій протягом одного модуля, поточна оцінка за відповідний модуль зменшується на три бали, за кожну пропущену лекцію понад дві. Штрафні бали можуть бути відпрацьовані, шляхом виконання додаткових індивідуальних або групових завдань, проявлення активності під час лекційних та лабораторних занять, участі у студентських конференціях (за тематикою дисципліни) або змаганнях з аналізу даних.

Модуль	Вид контролю	Вид робіт	Максимальна кількість балів	Сума		
1	Поточний контроль	Виконання та захист лабораторних робіт	ЛР №1	30	60	100
			ЛР №2	30		
	Модульний контроль		40	40		
2	Поточний контроль	Виконання та захист лабораторних робіт	ЛР №3	20	60	100
			ЛР №4	20		
			ЛР №5	20		
	Модульний контроль		40	40		

При заповненні заліково-екзаменаційної відомості та залікової книжки (індивідуального навчального плану) здобувача, оцінка, виставлена за 100-бальною шкалою, переводиться до національної шкали (5, 4, 3, 2) та шкали ЄКТС (A, B, C, D, E, FX, F).

За національною шкалою	Визначення назви за шкалою ЄКТС	За 100-бальною шкалою	За шкалою ЄКТС
ВІДМІННО – 5	Відмінно – відмінне виконання лише з незначною кількістю помилок	90-100	A
ДОБРЕ – 4	Дуже добре – вище середнього рівня з кількома помилками	82-89	B
	Добре – в загальному правильна робота з певною кількістю грубих помилок	75-81	C
ЗАДОВІЛЬНО – 3	Задовільно – непогано, але зі значною кількістю недоліків	69-74	D
	Достатньо – виконання задовольняє мінімальні критерії	60-68	E
НЕЗАДОВІЛЬНО – 2	Незадовільно – потрібно попрацювати перед тим як отримати залік (без повторного вивчення модуля)	35-59	FX
	Незадовільно – необхідна серйозна подальша робота (повторне вивчення модуля)	<35	F

Результати навчання

Дисципліна має на меті сформувати та розвинути наступні компетентності студентів:

1. Оволодіння знаннями про основні концепції та характеристики великих даних, включаючи їх обробку та ключові проблеми.
2. Володіння навичками використання сучасних інструментів для збору, зберігання та обробки великих даних, таких як Hadoop, Apache Spark та NoSQL бази даних.
3. Набуття практичних навичок налаштування та конфігурації розподілених обчислювальних систем для обробки великих об'ємів даних.
4. Здатність застосовувати методи аналізу великих даних, включаючи MapReduce та алгоритми машинного навчання.
5. Володіння навичками оцінки та візуалізації даних за допомогою сучасних інструментів для ефективного представлення результатів обробки.
6. Здатність застосовувати технології великих даних для вирішення реальних завдань у різних галузях.

Кодекс академічної доброчесності

Порушення [Кодексу академічної доброчесності УкрДУЗТ](#) є серйозним порушенням, навіть якщо воно є ненавмисним.

Зокрема, дотримання Кодексу академічної доброчесності УкрДУЗТ означає, що вся робота на іспитах та заліках має виконуватися індивідуально. Під час виконання самостійної роботи студенти можуть консультуватися з викладачами та з іншими студентами, але повинні самостійно розв'язувати завдання, керуючись власними знаннями, уміннями та навичками. Посилання на всі ресурси та джерела (наприклад, у звітах, самостійних роботах чи

презентаціях) повинні бути чітко визначені та оформлені належним чином. У разі спільної роботи з іншими студентами над виконанням індивідуальних завдань, ви повинні зазначити ступінь їх залучення до роботи.

Інтеграція студентів із обмеженими можливостями

Вища освіта є провідним чинником підвищення соціального статусу, досягнення духовної, матеріальної незалежності і соціалізації молоді з обмеженими функціональними можливостями й відображає стан розвитку демократичних процесів і гуманізації суспільства.

Для інтеграції студентів із обмеженими можливостями в освітній процес Українського державного університету залізничного транспорту створена система дистанційного навчання на основі сучасних педагогічних, інформаційних, телекомунікаційних технологій.

Доступ до матеріалів дистанційного навчання з цієї дисципліни можна знайти на [платформі дистанційного навчання УкрДУЗТ](#).

Викладачі

Лектор та керівник лабораторних занять:

[Олександр Ігорович Іванюк](#) – доктор філософії, доцент кафедри інформаційних технологій.

Контакти: +38 (057) 730-19-84, e-mail: ivaniuk@kart.edu.ua.

Ресурси

Інформаційні ресурси дисципліни розміщено на [платформі дистанційного навчання УкрДУЗТ](#).

Рекомендована література

Основна

1. Marz N., Warren J. Big data, 2015.
2. Marr B. Big data in practice, 2016.
3. Mayer-Schönberger V., Cukier K. Big data, 2013.
4. Басюк Т.М., Литвин В.В., Захарія Л.М., Кунанець Н.Е. Машинне навчання: навч. посіб. Львів: Новий Світ-2000, 2019. 315 с.
5. Hastie T., Friedman J., Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2017. URL: <https://web.stanford.edu/~hastie/ElemStatLearn>.

Допоміжна

1. MacKay D. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press, 2019. URL: <http://www.inference.org.uk/mackay/itila>.
2. Burkov A. The hundred-page machine learning book.
3. Daume H. A Course in Machine Learning (v0.9), 2017. URL: <http://ciml.info>.
4. Wasserman L. All of Statistics. New York, NY: Springer, 2013. URL: <http://www.stat.cmu.edu/~larry/all-of-statistics/index.html>.
5. Коельо Л.П., Ричарт В. Побудова систем машинного навчання на мові Python. 2016. 302 с.
6. Ng A. Machine Learning course notes. Coursera, 2011.
7. Domingos, P. The Master Algorithm: How machine learning is reshaping how we live. 2015.

Корисні інтернет-ресурси

1. Apache Hadoop Official Documentation – <https://hadoop.apache.org/docs/> Офіційна документація Hadoop, що містить детальні інструкції з налаштування та використання.
2. Apache Spark Official Documentation – <https://spark.apache.org/docs/latest/> Документація з інсталяції, конфігурації та застосування Apache Spark.
3. Apache Kafka Official Documentation – <https://kafka.apache.org/documentation/> Офіційна документація щодо налаштування та використання Apache Kafka для потокової обробки даних.
4. MongoDB University – <https://university.mongodb.com/> Безкоштовні курси та матеріали щодо роботи з базами даних MongoDB.
5. Google Cloud Big Data Solutions – <https://cloud.google.com/solutions/big-data> Офіційна сторінка Google Cloud із рішеннями для обробки великих даних.
6. AWS Big Data – <https://aws.amazon.com/big-data/> Інструменти та сервіси AWS для аналізу та обробки великих даних.
7. Coursera: Big Data Specialization – <https://www.coursera.org/specializations/big-data> Спеціалізація Coursera з великих даних, включає курси від провідних університетів.
8. Kaggle – <https://www.kaggle.com/> Платформа для змагань і навчання в галузі аналізу даних та машинного навчання.
9. Cloudera Educational Resources – <https://www.cloudera.com/about/training/certification.html> Навчальні ресурси від Cloudera для роботи з Hadoop та іншими технологіями.
10. Databricks Academy – <https://databricks.com/learn> Курси та навчальні матеріали для роботи з Apache Spark і Databricks.