



ТЕХНОЛОГІЇ DATA SCIENCE ДЛЯ ЗАВДАНЬ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 Інформаційні технології
Спеціальність	123 Комп'ютерна інженерія, 121 Інженерія програмного забезпечення
Освітня програма	Комп'ютерна інженерія, Інженерія програмного забезпечення
Статус дисципліни	Вибіркова компонента ОП, циклу професійної підготовки
Форма навчання	очна (денна)
Рік підготовки, семестр	5 курс, весняний
Обсяг дисципліни	4 кредити /120 год.: лекцій 36 годин, лаб. роб. 18 год., СРС 66 год.
Семестровий контроль/ контрольні заходи	Залік
Розклад занять	http://rozklad.kpi.ua/ , http://roz.kpi.ua/
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор, лабораторні: доктор технічних наук, професор Писарчук Олексій Олександрович, rysarchuk.oleksii@kpi.ua .
Розміщення курсу	https://classroom.google.com/u/5/c/NzA5MDM3NDE4ODE1

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна «Технології Data Science для завдань електронної комерції» призначена для **набуття студентом здатності** синтезувати, верифікувати математичні моделі, розробляти спеціалізоване програмне забезпечення з обробки і аналізу даних різного типу і обсягу. Це досягається вивченням теоретичних основ синтезу математичних моделей, методології вибору методів, і алгоритмів обробки даних різного типу, верифікації отриманих результатів методами імітаційного моделювання, а також практичної реалізації обраних підходів з метою розробки спеціалізованих прикладних програм.

Метою вивчення курсу «Технології Data Science для завдань електронної комерції» є: надання комплексного ґрунтового теоретичного базису та потужних практичних навичок програмної реалізації методів, математичних моделей і алгоритмів технологічних процесів Data Science.

Специфіка курсу полягає у розгляді, поряд із класичними методологіями Data Science, передових авторських розробок, отриманих у ході реалізації прикладних R&D проектів.

Лекційні заняття розкривають теоретичні основи основних галузей Data Science, їх практичну реалізацію у програмному коді та доповнюються практикою виконання завдань лабораторних робіт.

Практична частина курсу орієнтована на застосування мови програмування високого рівня Python з опануванням функціоналу бібліотек: Numpy, Pandas, Statsmodels, Random, Scipy, Google OR-Tools, Scikit-learn, Tensorflow, Keras, Requests, Regex, OpenCV, Pillow, Matplotlib, Geopandas.

Завдання лабораторних робіт розділені за рівнями складності, побудовані за принципами нарощування функціональності в межах тем.

Зміст курсу збалансовано розкривається суть технологічних процесів Data Science: обробка даних з метою отримання інформації – обробка інформації з метою виявлення знань – маніпулювання знаннями – візуалізація результатів.

Курс розраховано на фахівців, що прагнуть опанувати знання, уміння та навички, які потребують посади: Data Scientist, Data Engineer; Data Analyst, Machine Learning Engineer – створювати та чути симфонії цифр.

Матеріали курсу може використовуватися для реалізації практичних питань із дослідження даних різного типу та об'єму для конкретних прикладних завдань, а також у реалізації завдань курсового проектування, розробки кваліфікаційних робіт тощо.

Успішне опанування дисципліни «Технології Data Science для завдань електронної комерції» потребує від студента: базових знань з програмування: принципи програмування, алгоритмізація та базові алгоритми; базових знань мови програмування Python: синтаксис, типи та структури даних, базові оператори розгалужених обчислень, функціональне та ООП програмування, робота з IDE, створення оточення; базових знань з математики: елементи теорія ймовірностей, дискретна математика, теорія матриць, дослідження функцій, аналітична геометрія, тригонометрія.

Курс включає 4 кредитів/120 год.: лекцій 36 годин, лаб. роб. 18 год., CPC 66 год.

По завершенню курсу студент матимемо експертизу у базових напрямках Data Science.

Методи та технології статистичного навчання (Statistical Learning): Етапи статистичного навчання. Дослідження та підготовка даних: модель, нормалізація, балансування, очищення. Властивості даних: статистична вибірка, похибки, види надмірності, числові характеристики. Моделювання статистичних вибірок та основних законів розподілу випадкових величин. Технології Python для методів статистичного навчання: NumPy, Pandas, Statsmodels, Random, Requests, Matplotlib. Практика генерації статистичних вибірок та визначення числових характеристик. **Підготовка даних для статистичного навчання:** Технології отримання / моніторингу даних (parsing, web scraping): HTTP протокол, методи Get/Post/Put/Delete; робота із файлами (*.txt, *.xlsx, *.json) та базами даних. Аномальні виміри: властивості, ознаки, виявлення, відкидання, відновлення. Визначення порядку (оптимізація) моделі для апроксимації даних. Практика парсингу, виявлення аномальних вимірів, оптимізації моделі. **Навчання регресійної моделі за Big Data масивом:** Статистичне навчання поліноміальної моделі (поліноміальна регресія) методом найменших квадратів МНК (Least Square Method (LSM)). Практика МНК згладжування. **Фільтрація Калмана (Kalman filter):** Сутність рекурентного згладжування / Калманівської фільтрації, застосування та властивості. Фільтр Калмана (Kalman filter): скалярні альфа-бета, альфа-бета-гамма фільтри. Практика рекурентного згладжування - Фільтрації Калмана. **Нелінійне згладжування - R&D результати:** Сутність нелінійного згладжування, класичні методи. R&D підходи до побудови нелінійних моделей. Практика нелінійного згладжування.

Технології підтримки прийняття рішень (для Decision Support System (DSS)): Сутність, завдання, застосування, методи. Основи теорії прийняття рішень, методи та моделі. Інтелектуальні системи підтримки прийняття рішень Decision Support System (DSS). Теорія розкладів та ERP (DSS) системи управління виробничими процесами. Практика планування розкладів з бібліотекою Google OR-Tools. **Багатокритеріальні методи прийняття рішень - R&D результати:** Оптимізаційні задачі. Багатофакторність та багатокритеріальність. Багатокритеріальне оцінювання, ідентифікація, розподіл ресурсів. Практика багатокритеріального прийняття рішень.

Інтелектуальний аналіз даних (Data Intelligence): Сутність, завдання, застосування, методи. Методологічні основи інтелектуального аналізу даних: OLAP, Data Mining, Text Mining.

Технології Python для інтелектуального аналізу даних. Практика інтелектуального аналізу даних: OLAP, Data Mining, Text Mining.

Машинне навчання (Machine Learning (ML)): Методи та технології кластеризації: Методи кластеризації: k-середніх, опорних векторів; найближчих сусідів. Технології Python для реалізації методів класифікації: Scipy, Scikit-learn. Приклади задач кластеризації зображень з пакетами OpenCV, Pillow. **Методи та технології класифікація:** Завдання, застосування, методи. Класифікація та ідентифікація. Ієрархічні методи кластеризації та ідентифікації: дерево рішень, випадковий ліс. Технології Python для класифікації / ідентифікації: Scipy, Scikit-learn. Практика задач класифікації / ідентифікації зображень з пакетами OpenCV, Pillow.

Технології штучних нейронних мереж (Artificial Intelligence (AI)): Штучний інтелект та штучні нейронні мережі: сутність, завдання, методи, застосування. Модель перцептрон та його навчання. Штучні нейронні мережі – базові поняття. Практика моделювання перцептрона та його навчання. **Основні типи та технології штучних нейронних мереж:** Штучна нейронна мережа: архітектура та процеси. Типи штучних нейронних мереж та їх синтез. Технології Python для реалізації штучних нейронних мереж: Tensorflow, Keras. Практика побудови та застосування штучних нейронних мереж.

Алгоритми та технології прогнозування динаміки зміни показників ефективності торговельних компаній: міні проект (в сфері електронної комерції) - макет ERP-системи: Сутність предметної галузі аналізу даних для завдань електронної комерції. Формалізація задачі та попередній аналіз даних. Розробка математичної моделі. Архітектурне проектування. Розробка програмного забезпечення.

Методологічні основи та практика SCORING – аналізу: міні проект (в банківській сфері аналізу даних) - макет ERP-системи: Поняття SCORING-гу. Математичні моделі SCORING – аналізу. Формалізація задачі та попередній аналіз даних. Розробка математичної моделі. Архітектурне проектування. Розробка програмного забезпечення.

Обробка геопросторової інформації (для Geographic Information System (GIS)): Геоінформатика, геоінформаційні технології, географічні інформаційні системи – ГІС (GIS). Технології Python для реалізації геоінформаційних технологій: Geopandas. **Практика аналізу геопросторової інформації - Макет GIS системи:** Постановка задачі та попередній аналіз даних. Розробка математичної моделі. Архітектурне проектування. Розробка програмного забезпечення.

Вибіркова навчальна дисципліна «Технології Data Science для завдань електронної комерції» підсилює фахові/загальні компетенції освітньо-професійних програм (ОПП).

123 Комп'ютерна інженерія ОПП:

ЗК 2 Здатність до абстрактного мислення, аналізу і синтезу;

ЗК 3 Здатність проводити дослідження на відповідному рівні;

ФК 2 Здатність розробляти алгоритмічне та програмне забезпечення, компоненти комп'ютерних систем та мереж, Інтернет додатків, кіберфізичних систем з ви-користанням сучасних методів і мов програмування, а також засобів і систем автоматизації проектування;

ФК14 Здатність використовувати технології штучного інтелекту;

ФК15 Здатність розробляти системи обробки великих обсягів даних;

ПРН 2 Знаходити необхідні дані, аналізувати та оцінювати їх;

ПРН17 Здійснювати дослідження і проектування інтелектуальних систем.

121 Інженерія програмного забезпечення ОПП:

ЗК01 Здатність до абстрактного мислення, аналізу та синтезу;

ЗК03 Здатність проводити дослідження на відповідному рівні;

ФК02 Здатність розробляти і реалізовувати наукові та/або прикладні проєкти у сфері інженерії програмного забезпечення;

ФК16 Здатність розробляти системи аналізу великих обсягів даних;

ПРН03 Будувані і досліджувати моделі інформаційних процесів у прикладній області;

ПРН04 Виявляти інформаційні потреби і класифікувати дані для проектування програмного забезпечення;

ПРН10 Модифікувати існуючі та розробляти нові алгоритмічні рішення детального проектування програмного забезпечення;

ПРН19 Знати і застосовувати математичні основи та технологій штучного інтелекту.

121 Інженерія програмного забезпечення ОНП:

ЗК01 Здатність до абстрактного мислення, аналізу та синтезу;

ЗК03 Здатність проводити дослідження на відповідному рівні;

ФК02 Здатність розробляти і реалізовувати наукові та/або прикладні проекти у сфері інженерії програмного забезпечення;

ФК14 Здатність до використання існуючих систем та проектування і розробки нових систем зберігання та оброблення надвеликих масивів даних;

ФК17 Здатність до математичного моделювання в економіці, розуміння прикладних задач і математичних моделей макро- і мікроекономіки, аналізу і прогнозування процесів ринкової економіки;

ПРН03 Будувати і досліджувати моделі інформаційних процесів у прикладній області;

ПРН04 Виявляти інформаційні потреби і класифікувати дані для проектування програмного забезпечення;

ПРН10 Модифікувати існуючі та розробляти нові алгоритмічні рішення детального проектування програмного забезпечення;

ПРН18 Розробляти математичне і програмне забезпечення для наукових досліджень в галузі інженерії програмного забезпечення;

ПРН25 Розробляти, реалізовувати та застосовувати різні методи інтелектуального аналізу даних Big Data, формулювати алгоритми обробки в парадигмі Map Reduce, обирати відповідну технологію зберігання і оброблення надвеликих даних, використовуючи для цього сучасні масштабовані комп'ютерні системи;

Як додатковий результат навчання вибіркова навчальна дисципліна «Технології Data Science для завдань електронної комерції» спрямована на формування здатності синтезувати, верифікувати математичні моделі, розробляти спеціалізоване програмне забезпечення з обробки і аналізу даних різного типу і обсягу.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Перереквізити:

Успішне опанування дисципліни «Технології Data Science для завдань електронної комерції» потребує від студента: базових знань з програмування: принципи програмування, алгоритмізація та базові алгоритми; базових знань мови програмування Python: синтаксис, типи та структури даних, базові оператори розгалужених обчислень, функціональне та ООП програмування, робота з IDE, створення оточення; базових знань з математики: елементи теорії ймовірностей, дискретна математика, теорія матриць, дослідження функцій, аналітична геометрія, тригонометрія.

Це підтримується навчальними дисциплінами ОПП:

123 Комп'ютерна інженерія:

Вища математика. Частина 1,2,3; Аналітична геометрія та лінійна алгебра; Програмування. Частина 1,2; Дискретна математика; Фізика; Теорія ймовірності та математична статистика; Структури даних та алгоритми; Організація баз даних; Інженерія програмного забезпечення; Алгоритми та методи обчислень.

121 Інженерія програмного забезпечення:

Математичний аналіз. Частина 1,2; Лінійна алгебра та аналітична геометрія; Теорія ймовірностей; Комп'ютерна дискретна математика; Алгоритми та структури даних. Частина 1,2; Основи програмування. Частина 1,2; Базис даних; Об'єктно-орієнтоване програмування; Методології і технології розроблення програмного забезпечення; Технології штучного інтелекту

Постреквізити:

Дисципліна відноситься до вибіркової компоненти ОПП, циклу професійної підготовки. Курс орієнтовано на прикладні аспекти проектування, синтезу та розробки математичних моделей, алгоритмів та програмного забезпечення складних розподілених інформаційних систем із властивостями інтелектуальності, що є міждисциплінарним зв'язком компонент освітньої програми.

Вибіркова навчальна дисципліна «Технології Data Science для завдань електронної комерції» є прикладною та спрямована на опанувати знання, уміння та навички - компетенцій, які потребують практичні посади: Data Scientist, Data Engineer; Data Analyst, Machine Learning Engineer.

Компетенції, отримані з курсу можуть також використовуватись для реалізації практичних питань із дослідження даних різного типу та об'єму у реалізації завдань курсового проектування, розробки кваліфікаційних робіт тощо.

3. Зміст навчальної дисципліни

МОДУЛЬ 1 МЕТОДОЛОГІЧНІ ТА ТЕХНОЛОГІЧНІ ОСНОВИ DATA SCIENCE

Тема 1.1. Методи та технології статистичного навчання (Statistical Learning).

Лекція 1. Вступ до статистичного навчання:

Про курс Data Science;

Знайомство із Data Science;

Вступ до статистичного навчання.

Лекція 2. Підготовка даних для статистичного навчання:

Модель статистичної вибірки;

Отримання реальних даних;

Дослідження та підготовка даних;

Лабораторна робота 1. Підготовка та аналіз даних для статистичного навчання.

Лекція 3. Навчання регресійної моделі за Big Data масивом:

Очищення статистичної вибірки від аномалій;

Оцінка якості та оптимізація моделі даних;

Навчання регресійної моделі за Big Data масивом з використанням МНК;

Лекція 4. Фільтрація Калмана (Kalman filter):

Рекурентна фільтрація / згладжування;

Рекурентний α - β фільтр;

Рекурентний α - β - γ фільтр;

Фільтр Калмана (Kalman filter);

Лекція 5. Нелінійне згладжування - R&D результати:

Сутність нелінійного згладжування, класичні методи;

R&D підходи до побудови нелінійних моделей.

Лабораторна робота 2. Статистичне навчання з поліноміальною регресією.

Тема 1.2. Технології підтримки прийняття рішень (для Decision Support System (DSS)).

Лекція 6. Теорія і практика підтримки прийняття рішень.

Теорія прийняття рішень (сутність, завдання, застосування, методи);

Інтелектуальні системи підтримки прийняття рішень (Decision Support System (DSS), ERP CRM системи);

Теорія розкладів (Constraint Programming) з бібліотекою

Google OR-Tools.

Лекція 7. Багатокритеріальні методи прийняття рішень - R&D результати.

Оптимізаційні задачі. Багатофакторність та багатокритеріальність (multicriteria decision analysis);

Багатокритеріальне оцінювання (SCORING), ідентифікація (identification).

Лабораторна робота 3: Макет ERP системи з підтримкою прийняття рішень.

Тема 1.3. Інтелектуальний аналіз даних (Data Intelligence).

Лекція 8. Методичні основи та технології інтелектуального аналізу даних:

Інтелектуальний аналіз даних.

Технологія OLAP.

Технологія Text Mining.

Лабораторна робота 4: Реалізація процесів інтелектуального аналізу даних: міні проекти в галузі OLAP, Data Mining, Text Mining.

Тема 1.4. Машинне навчання (Machine Learning (ML)).

Лекція 9. Методи та технології кластеризації:

Вступ до машинного навчання (Machine Learning (ML)).;

Методи кластеризації: k-середніх, опорних векторів, найближчих сусідів, ієрархічна.

Технології Python для реалізації методів кластеризації: Scipy, Scikit-learn.

Технології кластеризації зображень з пакетами OpenCV, Pillow.

Лекція 10. Методи та технології класифікації та ідентифікації:

Класифікація та ідентифікація.

Ієрархічні методи класифікації та ідентифікації: дерево рішень, випадковий ліс.

Технології Python для класифікації / ідентифікації: Scipy, Scikit-learn та інше.

Приклади задач класифікації / ідентифікації зображень з пакетами OpenCV, Pillow та інше.

Лабораторна робота 5: Реалізація методів машинного навчання (Machine Learning (ML)).

Тема 1.5. Технології штучних нейронних мереж (Artificial Intelligence (AI)).

Лекція 11. Основи штучних нейронних мереж:

Штучний інтелект та штучні нейронні мережі: сутність, завдання, методи, застосування.

Модель персептрон та його навчання.

Штучні нейронні мережі – базові поняття.

Практика моделювання персептрона та його навчання.

Лекція 12. Основні типи та технології штучних нейронних мереж:

Штучна нейронна мережа: архітектура та процеси.

Типи штучних нейронних мереж та їх синтез.

Технології Python для реалізації штучних нейронних мереж: Tensorflow, Keras.

Практика побудови та застосування штучних нейронних мереж.

Лабораторна робота 6: Реалізація штучних нейронних мереж.

МОДУЛЬ 2 ПРИКЛАДНІ АСПЕКТИ ТЕХНОЛОГІЙ DATA SCIENCE

Тема 2.1. Алгоритми та технології прогнозування динаміки зміни показників ефективності торгівельних компаній.

Лекція 13. Алгоритми та технології прогнозування динаміки зміни показників ефективності торгівельних компаній.

Сутність предметної галузі аналізу даних для завдань електронної комерції.

Формалізація задачі та попередній аналіз даних.

Розробка математичної моделі.

Архітектурне проектування.

Розробка програмного забезпечення.

Лабораторна робота 7: Розробка програмного модуля прогнозування динаміки зміни показників ефективності торгівельних компаній (міні проекти в галузі аналізу даних для завдань електронної комерції).

Тема 2.2. Алгоритми та технології визначення кредитних ризиків для банківських CRM систем.

Лекція 14. Методологічні основи SCORING – аналізу.

Поняття SCORING-гу.

Математичні моделі SCORING – аналізу.

Лекція 15. Практика SCORING – аналізу.

Формалізація задачі та попередній аналіз даних.

Розробка математичної моделі.

Архітектурне проектування.

Розробка програмного забезпечення.

Лабораторна робота 8: Макет CRM системи SCORING – аналізу (міні проекти в банківській сфері аналізу даних).

Тема 2.3. Обробка геопросторової інформації (для Geographic Information System (GIS)).

Лекція 16. Основи геоінформаційних технологій.

Геоінформатика, геоінформаційні технології, географічні інформаційні системи – ГІС (GIS).

Технології Python для реалізації геоінформаційних технологій: Geopandas.

Лекція 17. Практика аналізу геопросторової інформації.

Постановка задачі та попередній аналіз даних.

Розробка математичної моделі.

Архітектурне проектування.

Розробка програмного забезпечення.

Лабораторна робота 9: Макет GIS системи.

4. Навчальні матеріали та ресурси

4.1. Базова література:

1. Навчально-методичний комплекс з дисципліни: Технології Data Science для завдань електронної комерції
[<https://drive.google.com/drive/folders/15KnWma4iAUGtITwbOR2fy4kSZIO9FAUz>].
2. Електронний курс на освітній платформі Sikorsky «Технології Data Science для завдань електронної комерції», 2022: <https://classroom.google.com/c/NTI1NDE1NDgyMjAw?cjc=xjig4yf>.
3. Писарчук О.О. Нелінійне та багатокритеріальне моделювання процесів у системах керування рухом / Писарчук О.О., Харченко В.П. – Монографія. – К.: Інститут обдарованої дитини, 2015. – 248 с.
4. Писарчук О.О. Багатокритеріальні математичні моделі ситуаційного управління та самоорганізації у складних інформаційних системах / Писарчук О.О., Данник Ю. Г., Шестаков В.І., Соколов К.О., та інші. – Монографія. – Житомир: ПП «Рута», 2016. – 232 с.
5. Ковбасюк С.В. Метод найменших квадратів та його практичне застосування / С.В. Ковбасюк, О.О. Писарчук, М.Ю. Ракушев // Монографія. – Житомир: ЖВІ НАУ, 2008. – 228 с.
6. Писарчук О.О. Моніторинг об'єктів в умовах апріорної невизначеності джерел інформації. Теорія та практика / Бобало Ю. Я., Данник Ю. Г., Писарчук О.О., Комарова Л. О. та інші // Монографія. – Львів: Коло, 2014. – 235 с.
7. Шматок С.О. Введення до теорії дослідження операцій / Шматок С.О., Данник Ю. Г., Писарчук О.О. – Монографія. – Житомир: ЖВІ, 2015. – 316 с.
8. Писарчук О.О. Методологічні основи наукових досліджень. Математичне моделювання та оптимізація складних систем / Грабар І.Г., Данник Ю.Г., Писарчук О.О., Гуменюк М.О. та інші. Навчальний посібник. – Житомир: ЖВІ ДУТ, 2015. – 680 с. (З грифом МОН України. Лист № 1/11-10150 від 02.07.2014 р).
9. Писарчук О.О. Основи теорії систем та системного аналізу / О.О. Писарчук, М. А. Павленко, О. В. Петров, С. І. Хмелевський, та інші. Підручник – Харків: ХНУПС, 2018. – 215 с.
10. Ковбасюк С.В. Основи програмного та математичного забезпечення автоматизованих систем управління військового призначення / С.В. Ковбасюк, О.О. Писарчук, С.А. Герговський // Навчальний посібник. – Житомир: ЖВІРЕ, 2005. – 300 с.
11. Ковбасюк С.В. Програмне та математичне забезпечення обчислювальних систем АСУ / С.В. Ковбасюк, О.О. Писарчук // Конспект лекцій. – Житомир: ЖВІРЕ, 2006. – 164 с.
12. Ковбасюк С.В. Теоретичні основи автоматизації процесів вироблення рішень у системах управління / С.В. Ковбасюк, О.О. Писарчук // Навчальний посібник. – Житомир: ЖВІРЕ, 2006. – 132 с.
13. Бучик С.С. Системи підтримки прийняття рішень / С.С. Бучик, О.О. Писарчук // Конспект лекцій. – Житомир: ЖВІРЕ, 2006. – 168 с.
14. Салюк М.А. Статистична обробка даних експериментального дослідження. Методичний посібник з курсу «Експериментальна психологія» / за ред. Е.Л. Носенко. – Дніпропетровськ: Інновація, 2010. – 26 с.

15. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
16. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. – 297 с.
17. Василенко О. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. / О. А. Василенко, І. А. Сенча. – Одеса: ОНАЗ ім. О. С. Попова, 2011. – 166 с.
18. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
19. William McKinney Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. - Published by O'Reilly Media, Inc, 2013. – 453 p.
20. Plas J. Wander. Python Data Science. - Published by O'Reilly Media, Inc, 2017. – 448 p.
21. Sebastian Raska, Vahid Mirjalili. Python and machine learning. - Packt Publishing, 2017. – 656p.
22. Prateek Joshi Artificial Intelligence applications with Python. - Packt Publishing, 2017. – 446p.

4.2. Додаткова література:

1. Foster Provost and Tom Fawcett. Data Science for Business / Foster Provost and Tom Fawcett. - Printed in the United States of America. Published by O'Reilly Media, Inc, 2013. – 409 с.
2. David Dietrich. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data / David Dietrich, Barry Heller, Beibei Yang. - John Wiley & Sons, Inc., Indianapolis, Indiana, 2015. – 420 с.
3. Trevor Hastie Robert Tibshirani Jerome Friedman The Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition. – Springer, 2020. – 768 p.
4. Бахрушин В.Є. Математичне моделювання / В.Є. Бахрушин. – Запоріжжя : ГУ “ЗІДМУ”, 2003. – 138 с.
5. Бахрушин В.Є. Математичні основи моделювання систем : навчальний посібник / В.Є. Бахрушин. – Запоріжжя : КПУ, 2009. – 224 с.
6. Kishan G. Anomaly Detection Principles and Algorithms / Kishan G., Mehrotra, Chilukuri K. Mohan, HuaMing Huang. – Switzerland: Springer, 2017. – 229 с.
7. Hanna Blomquist, Johanna Möller. Anomaly detection with Machine learning. Quality assurance of statistical data in the Aid community, 2015. – 60 с. [<https://uu.diva-portal.org/smash/get/diva2:846985/FULLTEXT01.pdf>].
8. Srikanth Thudumu, Philip Branch, Jiong Jin and Jugdutt (Jack) Singh. A comprehensive survey of anomaly detection techniques for high dimensional big data. Springer, 2017. – 30 с.

4.3. Інформаційні ресурси:

<https://www.kaggle.com/>
<https://www.jetbrains.com/>
<https://github.com/PacktPublishing/Artificial-Intelligence-with-Python>
<https://pandas.pydata.org/>
<https://scapy.net/>
<https://developers.google.com/optimization>
<https://www.tensorflow.org/>
<https://keras.io/>
<https://opencv.org/>
<https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html>
<https://www.statsmodels.org/stable/examples/notebooks/generated/ols.html>
<https://scikit-learn.org/stable/modules/sgd.html#regression>

5. Методика опанування навчальної дисципліни (освітнього компонента)

Навчальна дисципліна «Технології Data Science для завдань електронної комерції» включає 4 кредитів/120 год.: лекцій 36 годин, лаб. роб. 18 год., СРС 66 год.

За дисципліною передбачено проведення таких видів аудиторних занять: лекційних занять; лабораторні роботи; модульна контрольна робота; залік.

Лекційні заняття розкривають теоретичні основи основних галузей Data Science, їх практичну реалізацію у програмному коді та доповнюються практикою виконання завдань лабораторних робіт.

Практична частина курсу (лабораторні роботи) орієнтована на застосування мови програмування високого рівня Python з опануванням функціоналу бібліотек: Numpy, Pandas, Statsmodels, Random, Scipy, Google OR-Tools, Scikit-learn, Tensorflow, Keras, Requests, Regex, OpenCV, Pillow, Matplotlib, Geopandas.

Завдання лабораторних робіт розділені за рівнями складності, побудовані за принципами нарощування функціональності в межах тем.

Тематика лекційних занять.

Лекція 1. Вступ до статистичного навчання:

Лекція 2. Підготовка даних для статистичного навчання:

Лекція 3. Навчання регресійної моделі за Big Data масивом:

Лекція 4. Фільтрація Калмана (Kalman filter):

Лекція 5. Нелінійне згладжування - R&D результати:

Лекція 6. Теорія і практика підтримки прийняття рішень.

Лекція 7. Багатокритеріальні методи прийняття рішень - R&D результати.

Лекція 8. Методичні основи та технології інтелектуального аналізу даних:

Лекція 9. Методи та технології кластеризації:

Лекція 10. Методи та технології класифікації та ідентифікації:

Лекція 11. Основи штучних нейронних мереж:

Лекція 12. Основні типи та технології штучних нейронних мереж:

Лекція 13. Алгоритми та технології прогнозування динаміки зміни показників ефективності торгівельних компаній.

Лекція 14. Методологічні основи SCORING – аналізу.

Лекція 15. Практика SCORING – аналізу.

Лекція 16. Основи геоінформаційних технологій.

Лекція 17. Практика аналізу геопросторової інформації.

Модульна контрольна робота.

Тематика лабораторних занять.

Лабораторна робота 1. Підготовка та аналіз даних для статистичного навчання.

Лабораторна робота 2. Статистичне навчання з поліноміальною регресією.

Лабораторна робота 3: Макет ERP системи з підтримкою прийняття рішень.

Лабораторна робота 4: Реалізація процесів інтелектуального аналізу даних: міні проекти в галузі OLAP, Data Mining, Text Mining.

Лабораторна робота 5: Реалізація методів машинного навчання (Machine Learning (ML)).

Лабораторна робота 6: Реалізація штучних нейронних мереж.

Лабораторна робота 7: Розробка програмного модуля прогнозування динаміки зміни показників ефективності торгівельних компаній (міні проекти в галузі аналізу даних для завдань електронної комерції).

Лабораторна робота 8: Макет CRM системи SCORING – аналізу (міні проекти в банківській сфері аналізу даних).

Лабораторна робота 9: Макет GIS системи.

6. Самостійна робота здобувача вищої освіти денної форми навчання

Види самостійної роботи (66 годин):

самостійне опрацювання матеріалів лекційних занять (1 година x 17 лекцій = 17 годин);
підготовка та оброблення проведення розрахунків за первинними даними, отриманими на лабораторних заняттях, виконання лабораторних робіт, розв'язок задач надання на перевірку (рекомендовано 2 години x 9 лабораторних робіт = 18 годин);
виконання модульної контрольної роботи (МКР = 4 години);
підготовка до заліку (8 годин);
налаштування віртуального середовища для виконання лабораторних робіт (3 години);
опрацювання тем на самостійну роботу (16 годин).

Теми на самостійне опрацювання (денна форма навчання).

МОДУЛЬ 1 МЕТОДОЛОГІЧНІ ТА ТЕХНОЛОГІЧНІ ОСНОВИ DATA SCIENCE

Тема 1.1. Методи та технології статистичного навчання (Statistical Learning).

Лекція 2. Підготовка даних для статистичного навчання:

Практика отримання реальних даних; Дослідження та підготовка даних;

Лекція 3. Навчання регресійної моделі за Big Data масивом:

Навчання регресійної моделі за Big Data масивом з використанням МНК;

Лекція 4. Фільтрація Калмана (Kalman filter):

Фільтр Калмана (Kalman filter).

Тема 1.2. Технології підтримки прийняття рішень (для Decision Support System (DSS)).

Лекція 6. Теорія і практика підтримки прийняття рішень.

Практика теорії розкладів (Constraint Programming) з бібліотекою Google OR-Tools.

Тема 1.3. Інтелектуальний аналіз даних (Data Intelligence).

Лекція 8. Методичні основи та технології інтелектуального аналізу даних:

Практика технологій OLAP, Text Mining.

Тема 1.4. Машинне навчання (Machine Learning (ML)).

Лекція 10. Методи та технології класифікації та ідентифікації:

Технології Python для класифікації / ідентифікації: Scipy, Scikit-learn та інше.

Тема 1.5. Технології штучних нейронних мереж (Artificial Intelligence (AI)).

Лекція 12. Основні типи та технології штучних нейронних мереж:

Практика побудови та застосування штучних нейронних мереж.

МОДУЛЬ 2 ПРИКЛАДНІ АСПЕКТИ ТЕХНОЛОГІЙ DATA SCIENCE

Тема 2.1. Алгоритми та технології прогнозування динаміки зміни показників ефективності торгівельних компаній.

Лекція 13. Алгоритми та технології прогнозування динаміки зміни показників ефективності торгівельних компаній.

Розробка програмного забезпечення.

Тема 2.2. Алгоритми та технології визначення кредитних ризиків для банківських CRM систем.

Лекція 15. Практика SCORING – аналізу.

Розробка програмного забезпечення.

Тема 2.3. Обробка геопросторової інформації (для Geographic Information System (GIS)).

Лекція 17. Практика аналізу геопросторової інформації.

Розробка програмного забезпечення.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Для виконання лабораторних робіт та модульних контрольних робіт встановлюються дедлайни.

Виконання лабораторних робіт поза встановлених термінів супроводжуються штрафними балами, які вираховуються із оцінки за протокол (п. 2.9. ПОЛОЖЕННЯ про систему оцінювання результатів навчання). МКР не приймається поза встановлені терміни.

Штрафні бали виставляються за: невчасну здачу лабораторної роботи. Кількість штрафних балів не більше 10 (9 лабораторних робіт + МКР). Штрафні бали та жорсткі дедлайни не запроваджуються у період військового стану.

Заохочувальні бали виставляються за: R&D результати виконання лабораторних робіт; активну участь на лекціях; виконання поточних домашніх завдань, ведення конспекту, підготовка повідомлення з презентацією по одній із тем СРС дисципліни тощо. Кількість заохочуваних балів не більше 10.

Виконанню кожної лабораторної роботи передую виконання індивідуального завдання і оформлення його у вигляді протоколу. Студент, який прийшов на заняття без оформленого протоколу до лабораторної роботи не допускається. Першим етапом студент захищає результати отримані під час виконання індивідуального завдання до лабораторної роботи, на другому етапі – захищає теорію шляхом усного опитування або тестування (за необхідності). Захист лабораторних робіт може бути проведено за результатами аналізу повноти і якості виконання протоколу. Бали отримані за виконання лабораторної роботи, за теоретичною частиною та за протокол входять в оцінку за лабораторну роботу.

Перездача лабораторної роботи у разі її позитивної оцінки з метою підвищення оцінки – не передбачено.

Виконання лабораторних робіт є обов'язковими для допуску до семестрового контролю. Умовою допуску до семестрового контролю є зарахування всіх лабораторних робіт та стартовий рейтинг не менше 60 балів.

Модульна контрольна робота пишеться на лекційному занятті без застосування допоміжних засобів (мобільні телефони, планшети та ін.); результат в електронному вигляді надаються викладачеві.

Модульна контрольна робота не переписується за умови негативної оцінки. Негативна оцінка за МКР прирівнюється до 0 балів, в цьому випадку МКР не зараховується.

Оцінка, яку студент може отримати за виконання кожної лабораторної роботи та за модульну контрольну роботу наведені в таблиці 1 оцінювання семестрових робіт, розділ 8 силабусу.

Таким чином мінімальна оцінка, яку повинен отримати студент для допуску до семестрового контролю дорівнює 60 балів, максимальна – 100 балів за виконання всіх поточних робіт за семестр.

Здобувачі, які виконали всі умови допуску до заліку (виконали всі лабораторні роботи) та мають рейтингову оцінку менше 60 балів, а також ті здобувачі, хто бажає підвищити свою рейтингову оцінку, на останньому за розкладом занятті мають можливість пройти семестровий контроль у вигляді залікової контрольної роботи.

У разі виконання залікової контрольної роботи рейтингова оцінка визначається як сума балів за залікову контрольну роботу та балів за індивідуальні семестрові завдання.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Для навчальної дисципліни «Технології Data Science для завдань електронної комерції», як прикладної дисципліни, що спрямована на отримання комплексного ґрунтовного теоретичного базису та потужних практичних навичок програмної реалізації методів, математичних моделей і алгоритмів технологічних процесів Data Science застосовується PCO-1 (п. 2.1. ПОЛОЖЕННЯ про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського, https://osvita.kpi.ua/sites/default/files/downloads/Pologennia_RSO_2022.pdf).

Семестровий рейтинг студента розраховується, виходячи із 100-бальної шкали. Рейтинг складається з балів, що студент отримав за виконання 9 лабораторних робіт РЛ та однієї модульної контрольної роботи РМКР.

Завдання лабораторних робіт розділені за рівнями складності. Високий рівень складності передбачає отримання максимум 9 балів, середній рівень – 7 балів.

Максимальна кількість балів за лабораторні роботи (РЛ) за високим рівнем складає 81 бал, за середнім рівнем - 63 балів.

Розподіл балів за виконання лабораторних робіт.

1.1. Якість / повнота оформлення протоколу з лабораторної роботи – 1 бал.

1.2. Своєчасний захист роботи – 1 бал.

1.3. Повнота аналізу отриманих результатів – 1 бал.

1.4. Якість та повнота виконання технічних умов завдання, функціональність розробленої технічної продукції (програмного скрипта) -4 бали.

1.5. Рівень теоретичної підготовки – 2 бали.

*** Для умов дистанційного навчання бали за теоретичну підготовленість (п.1.4) можуть нараховуватись за результатами аналізу вмісту протоколу з лабораторної роботи.

*** Для умов військового стану – своєчасність захисту лабораторної роботи (п.1.2) – не застосовується а додається до п.1.4.

Максимальна кількість балів за МКР - РМКР = 9.

Розподіл балів за виконання МКР.

2.1. Якість / повнота оформлення звіту з МКР – 1 бал;

2.2. Повнота розкриття суті та оригінальність відповіді на теоретичні питання – 1,5 балів за кожне питання – загалом – 3 бали;

2.3. Повнота, оригінальність та якість виконання практичного питання – 5 бали.

Максимальна кількість балів за залікову роботу складає R3 = 10 балів.

3.1. Якість / повнота оформлення звіту із залікової роботи – 1 бал;

3.2. Повнота розкриття суті та оригінальність відповіді на теоретичні питання – 2 балів за кожне питання – загалом – 4 бали;

3.3. Повнота, оригінальність та якість виконання практичного питання – 5 бали.

У разі **значних запозичень / порушення вимог доброчесності** в звітному протоколі з лабораторної роботи - робота може бути не зарахована, або повернення на переопрацювання. За таких умов МКР та залікова робота – не зараховуються.

Календарна атестація студентів (на 8 та 14 тижнях семестрів) з дисципліни проводиться за значенням поточного рейтингу студента на час атестації. Якщо значення цього рейтингу не менше 50 % від максимально можливого на час атестації, студент вважається атестованим. В іншому випадку в атестаційній відомості виставляється «неатестовано».

Таким чином, порядок визначення загального рейтингу пояснюється Таб.1 та визначається за виразом

$$R = RЛ + RМКР + R3.$$

Таблиця 1

Загальний рейтинг за дисципліною

Звітність	Лр1	Лр2	Лр3	Лр4	Лр5	Лр6	Лр7	Лр8	Лр9	МК	СУМА	Залік	Сумма+залік
Високий рівень	9	9	9	9	9	9	9	9	9	9	90	10	100
Середній рівень	7	7	7	7	7	7	7	7	7	9	72	10	82

Максимальна кількість балів за семестр не перевищує RC = 100.

З урахуванням одержаної суми балів кінцева оцінка визначається за Табл.2.

Таблиця 2

Визначення семестрової оцінки

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

9. Нормативні положення політики навчальної дисципліни (освітнього компонента)

Правила відвідування занять. Відвідування занять є вільним, бали за присутність на лекціях та практичних заняттях не додаються, відповідно штрафні бали не передбачаються. Втім, вагома частина рейтингу формується через активну участь у заходах на практичних заняттях.

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету, пов'язаних з введенням карантину або режиму воєнного стану в державі, освітній процес здійснюється у дистанційному режимі відповідно до Положення про дистанційне навчання в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), Регламенту організації освітнього процесу в дистанційному режимі (<https://profkom.kpi.ua/reglament-organizatsiyi-osvitnogo-protsesu-v-distantsiynomu-rezhimi>) та Регламенту проведення семестрового контролю в дистанційному режимі (<https://osvita.kpi.ua/node/148>). У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформах BigBlueButton, Google Meet, Zoom. Посилання на конференцію видається на початку семестру і розміщується в АС «Електронний кампус». З метою забезпечення якісної підготовки здобувачів, дистанційний курс дисципліни розміщено на Платформі дистанційного навчання «Сікорський». Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. На заняттях слід дотримуватись норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>). На території університету здобувачі мають поводити себе відповідно до Правил внутрішнього розпорядку (<https://kpi.ua/admin-rule>).

Визнання результатів навчання, набутих у неформальній/інформальній освіті. Положення про визнання результатів навчання, набутих у неформальній / інформальній освіті (<https://osvita.kpi.ua/index.php/node/179>) регламентує визнання результатів навчання, набутих у неформальній/інформальній освіті:

у разі навчання на неформальній освіті, яку студент обрав самостійно і хоче зарахувати результати навчання в рамках дисципліни, проходиться процедура валідації, що передбачає подання здобувачем заяви на ім'я декана, декларації підтверджувальних документів; рішення про визнання чи не визнання приймається комісією у складі завідувача кафедри, викладача, гаранта освітньо-професійної програми.

за рішенням викладача окремі змістовні модулі або теми дисципліни можуть бути зараховані на підставі результатів проходження здобувачем онлайн-курсу або іншого елементу неформальної освіти; перелік завдань та елементів неформальної освіти визначається викладачем заздалегідь, про що студентам повідомляються на початку семестру; у цьому випадку здобувач звільняється від виконання відповідних завдань і отримує максимальний бал згідно з рейтинговою системою оцінювання.

За даним курсом можуть бути визнані результати навчання здобуті у неформальній / інформальній освіті в обсязі, що не перевищує 10% від загального обсягу навчального курсу (п.2.6 Положення).

Політика використання штучного інтелекту. Використання штучного інтелекту регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі завдання, як під час виконання навчальних завдань з дисципліни, так і індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено.

Порушення академічної доброчесності: використання плагіату. У разі виявлення фактів використання плагіату під час виконання навчальних завдань, студент втрачає право на подальше виконання поточних завдань у межах навчальної дисципліни до з'ясування обставин порушення.

Студент зобов'язаний аргументовано пояснити причини використання плагіату, а також засвідчити готовність дотримуватись принципів академічної доброчесності в подальшому навчанні; повторно виконати завдання або, за потреби, виконати додаткове завдання відповідно до бачення викладача з урахуванням вимог щодо самостійності.

Зазначені дії регламентуються Кодексом честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>).

Робочу програму навчальної дисципліни (силабус):

Складено професором кафедри обчислювальної техніки, доктором технічних наук, професором Писарчуком Олексієм Олександровичем.

Ухвалено кафедрою обчислювальної техніки (протокол № 12 від 23.06.2025).

Погоджено Методичною комісією факультету інформатики та обчислювальної техніки (протокол № 11 від 27.06.2025).