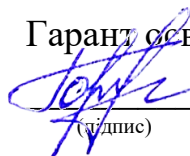


Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра геоінформаційних технологій
та космічного моніторингу Землі (№ 407)

ЗАТВЕРДЖУЮ

Гарант освітньої програми



С.І. Горелик
(ініціали та прізвище)

« 27 » _____ 08 _____ 2021 р.

СИЛАБУС ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Інтелектуальний аналіз і Big Data геоматиці

(назва навчальної дисципліни)

Галузь знань: 10 Природничі науки
(шифр і найменування галузі знань)

Спеціальність: 103 Науки про Землю
(код та найменування спеціальності)

Освітня програма: Космічний моніторинг Землі
(найменування освітньої програми)

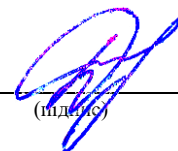
Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Силабус введено в дію з 01.09.2021 року

Харків – 2021 р.

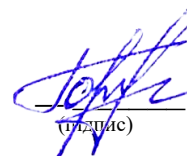
Розробник: Даншина С. Ю., доцент каф., д.т.н., доц.
(прізвище та ініціали, посада, науковий ступінь та вчене звання)


(підпис)

Силабус навчальної дисципліни розглянуто на засіданні кафедри геоінформаційних технологій та космічного моніторингу Землі. (№ 407)

Протокол № 1 від « 27 » серпня 2021 р.

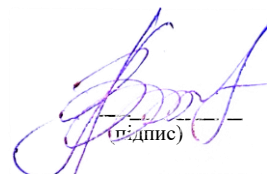
Завідувач кафедри к.т.н., доц.


(підпис)

Горелик С.І.

Погоджено з представником здобувачів освіти:

Студентка гр 435ст


(підпис)

Фомініх А.В.

1. Загальна інформація про викладача



Даншина Світлана Юріївна, д.т.н., доцент. З 2000 року викладає в університеті наступні дисципліни:

- комп'ютерні технології для ГІС-додатків;
- системний аналіз для ГІС-додатків;
- інтелектуальний аналіз даних і Bag Data в геоматиці;
- планування та управління ГІС проєктами;
- комп'ютерні системи.

Напрями наукових досліджень: геоінформаційні технології, комп'ютерні системи та технології, функціональне моделювання, Data mining & Big Data, системний аналіз, управління проєктами.

2. Опис навчальної дисципліни

Семестр, в якому викладається дисципліна – 7 семестр.

Обсяг дисципліни:

5,5 кредити ЄКТС (165 годин), у тому числі аудиторних – 64 годин, самостійної роботи здобувачів – 101 годин.

Форми здобуття освіти

Денна, дистанційна.

Дисципліна – вибіркова.

Види навчальної діяльності – лекції, практичні заняття, самостійна робота здобувача.

Види контролю – поточний, модульний і підсумковий (семестровий) контроль (залік).

Мова викладання – українська.

Необхідні обов'язкові попередні дисципліни (пререквізити) – вища математика, теорія ймовірностей та математична статистика, комп'ютерні технології для ГІС додатків.

Необхідні обов'язкові супутні дисципліни (кореквізити) – технології геоінформаційних систем, ГІС-аналіз, системний аналіз для ГІС-додатків.

3. Мета та завдання навчальної дисципліни

Мета

Вивчення навчальної дисципліни «Інтелектуальний аналіз даних і Big Data в геоматиці» полягає у придбанні студентами базових знань про інтелектуальний аналіз даних, а також навиків і вмінь вилучати корисну інформацію з даних, зокрема, великих даних при дослідженні, проектуванні, розробленні, експлуатації геоінформаційних систем і набутті практичних навичок з комп'ютерного моделювання та графічної інтерпретації даних.

Завдання

Вивчення дисципліни «Інтелектуальний аналіз даних і Big Data в геоматиці» є вивчення і засвоєння студентами основних принципів інтелектуального аналізу і Big Data для вирішення задач геоматики.

Після опанування дисципліни здобувач набуде такі **компетентності**:

ЗК7 – навички використання інформаційних і комунікаційних технологій.

ФК2 – здатність застосовувати базові знання фізики, хімії, біології, екології, математики, інформаційних технологій тощо при вивченні Землі та її геосфер.

ФК3 – здатність здійснювати збір, реєстрацію і аналіз даних за допомогою відповідних методів і технологічних засобів у польових і лабораторних умовах.

ФК5 – здатність застосовувати кількісні методи при дослідженні геосфер.

ФК6 – здатність до всебічного аналізу складу і будови геосфер.

ФК9 – здатність самостійно досліджувати природні матеріали (у відповідності до спеціалізації) в польових і лабораторних умовах, описувати, аналізувати, документувати і звітувати про результати.

ФК10 – здатність проводити комплексне оброблення результатів польових, камеральних та дистанційних досліджень з метою синтезування нових знань у сфері наук про Землю.

ФК12 – здатність до планування, організації та проведення досліджень і підготовки звітності.

ПРН4 – використовувати інформаційні технології, картографічні та геоінформаційні моделі в області наук про Землю.

ПРН7 – застосовувати моделі, методи і дані фізики, хімії, біології, екології, математики, інформаційних технологій тощо при вивченні природних процесів формування і розвитку геосфер.

ПРН9 – вміти виконувати дослідження геосфер за допомогою кількісних методів аналізу.

ПРН11 – впорядковувати і узагальнювати матеріали польових та лабораторних досліджень.

ПРН13 – уміти доносити результати діяльності до професійної аудиторії та широкого загалу, робити презентації та повідомлення.

Очікується, що після опанування дисципліни здобувач будуть досягнуті наступні **результати навчання** і він буде знати:

- принципи вилучення корисної інформації з даних для вирішення завдань геоматики;
- правила знаходження у вилученій інформації інформативних описових атрибутів об'єктів геоматики;
- алгоритми оброблення атрибутів об'єктів геоматики для вирішення завдань аналізу даних;
- методи комп'ютерного моделювання та графічної інтерпретації даних.

4. Зміст навчальної дисципліни

Модуль 1.

Змістовний модуль 1. Вступ до дата-аналітичного мислення та Big Data

Тема 1. Нова система обробки даних в геоматиці - Big Data

Форма занять: лекція, самостійна робота.

- *Обсяг аудиторного навантаження: 2 години.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Data Science, програмування й ухвалення рішень на основі даних. Обробка даних і Big Data. Джерела Big Data та їх особливості: правило «4 V». Вміння працювати з даними і Data Science як стратегічне надбання.

Обсяг самостійної роботи здобувачів: 10 годин.

Опрацювання матеріалу лекцій. Формування питань до викладача. Проблеми великих даних. Порівняльний аналіз традиційних підходів до аналізу даних і BD-аналітики.

Тема 2. Вступ до алгоритмів великих даних

- *Форма занять: лекція, самостійна робота.*
- *Обсяг аудиторного навантаження: 4 години.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Хеширування. Використання ієрархії сховищ. Поточкові й однопрохідні алгоритми. Фільтрація та вибірка. Паралелізм. Зменшення розмірності даних і виділення атрибутів.

- *Обсяг самостійної роботи здобувачів: 12 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Ознайомлення зі службами хмарних обчислень. Поняття MapReduce.

Тема 3. Інтелектуальний аналіз даних як елемент науки про дані

- *Форма занять: лекція, практичні заняття, самостійна робота.*
- *Обсяг аудиторного навантаження: 14 годин.*
- *Практичне заняття: «Формалізація текстових даних при*

інтелектуальному геоаналізі даних».

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти):* ДСТУ 34.302.2-91 (ISO/МЕК 8859-2:1999). «Інформаційні технології. Набори 8-бітних одnobайтових кодованих графічних символів. Латинський алфавіт № 2».

- *Практичне заняття:* «Репрезентація текстової інформації під час інтелектуального аналізу даних: метод мультимножини слів».

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти):* комп'ютер, пакет комп'ютерного математичного моделювання scilab.

Предмет, об'єкт і мета інтелектуального аналізу даних (Data mining). Поняття формалізації, форми формалізації даних. Особливості формалізації даних при інтелектуальному аналізі з урахуванням семантичного, синтаксичного та прагматичного підходів.

- *Обсяг самостійної роботи здобувачів: 12 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Оформлення звітів з практичних занять. Систематизація знань щодо можливого застосування методів формалізації текстової інформації під час вирішення технічних завдань.

Модульний контроль 1.

- *Форма занять:* написання модульної роботи в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).

- *Обсяг аудиторного навантаження: 2 години.*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти):* відсутні.

- *Обсяг самостійної роботи здобувачів – за необхідністю.*

Підготовка до модульного контролю.

Модуль 2.

Змістовний модуль 2. Методи інтелектуального аналізу даних в геоматиці

Тема 4. Математичні основи інтелектуального аналізу даних

- *Форма занять:* лекція, самостійна робота.

- *Обсяг аудиторного навантаження: 2 годин.*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти):* відсутні.

Моделі, індукція та прогнози. Базові моделі. Оцінювання моделей. Оцінювання середовища.

- *Обсяг самостійної роботи здобувачів: 10 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Поняття часового ряду. Задача наближення функцій.

Тема 5. Методи теорії множин в інтелектуальному аналізі даних

- *Форма занять:* лекція, практичні заняття, самостійна робота.

- *Обсяг аудиторного навантаження: 16 годин.*
- *Практичне заняття: «Способи визначення множин та операції з ними».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, пакет комп'ютерного математичного моделювання scilab.*
- *Практичне заняття: «Відображення множин».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, пакет комп'ютерного математичного моделювання scilab.*
- *Практичне заняття: «Робота з множинами під час інтелектуального аналізу даних: комбінаторика»*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Поняття множини. Особливості описання множини. Діаграми Вена як потужний елемент графічного аналізу геоданих. Типи відображень. Поняття функції як відображення множин. Основні правила відображень бієктивних множин. Комбінаторика.

- *Обсяг самостійної роботи здобувачів: 15 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Оформлення звітів з практичних занять. Можливість використання пакетів комп'ютерного моделювання для формування множини. Особливості формування відображень множин геоданих з використанням комп'ютерного моделювання.

Тема 6. Алгебра логіки в інтелектуальному аналізі даних

- *Форма занять: лекція, практичні заняття, самостійна робота.*
- *Обсяг аудиторного навантаження: 12 годин.*
- *Практичне заняття: «Методи алгебри логіки при інтелектуальному аналізі даних».*
- *Практичне заняття: «Алгебра предикатів при інтелектуальному аналізі даних».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*

Поняття висловлювання, логічних зв'язок, логічної формули, предиката. Підходи формалізації текстових геоданих за допомогою логічних операцій, зв'язок та предикатів. Методи мінімізації логічних формул. Формалізація геоданих із застосуванням кванторів та їх особливості.

- *Обсяг самостійної роботи здобувачів: 15 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Оформлення звіту з практичних занять. Закріплення на прикладах матеріалу по формалізації геоданих за допомогою логічних операцій. Вивчення випадків застосування кванторів та їх особливості.

Тема 7. Статистичні методи в інтелектуальному аналізі даних

- *Форма занять: лекція, практичне заняття, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 години.*
- *Практичне заняття: «Статистичний аналіз геоданих».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, пакет комп'ютерного математичного моделювання scilab.*

Ймовірність. Статистичний розподіл. Статистична значущість. Кореляція даних. Етапи кореляційного аналізу. Регресійний аналіз.

- *Обсяг самостійної роботи здобувачів: 17 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Оформлення звіту з практичного заняття. Класифікація та логістична регресія.

Тема 8. Маніпулювання даними

- *Форма занять: лекція, самостійна робота.*
- *Обсяг аудиторного навантаження: 2 години.*

Обрання даних. Стандартні формати даних. Збір даних. Очищення даних. Проблема відсутніх значень.

- *Обсяг самостійної роботи здобувачів: 12 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Поняття краудсорсингу як методу використання колективних знань групи.

Модульний контроль 2.

- *Форма занять: написання модульної роботи в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 2 години*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): відсутні.*
- *Обсяг самостійної роботи здобувачів – за необхідністю.*

Підготовка до модульного контролю.

5. Індивідуальні завдання

Відсутні.

6. Методи навчання

Використовуються наступні методи навчання: словесні (пояснення, розповідь, бесіда тощо), наочна (демонстрування) та практичні (практичні заняття).

7. Методи контролю

Поточний контроль (теоретичне опитування й розв'язання практичних завдань), модульний контроль (тестування за розділами курсу) та підсумковий (семестровий) контроль (залік). Форма проведення заліку – письмово-усна.

8. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
<i>Змістовний модуль 1</i>			
Активність під час аудиторної роботи	0...1	6	0...5 (максимальна кількість балів за цим показником)
Виконання і захист практичних робіт	0...8	2	0...16
Модульний контроль	0...13	1	0...13
<i>Змістовний модуль 2</i>			
Активність під час аудиторної роботи	0...1	8	0...5 (максимальна кількість балів за цим показником)
Виконання і захист практичних робіт	0...8	6	0...48
Модульний контроль	0...13	1	0...13
<i>За семестр</i>			0...100

Прийнята шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка	
	для екзамену, курсового проекту (роботи), практики	для заліку
90-100	відмінно	Зараховано
75-89	добре	
60-74	задовільно	
01-59	незадовільно з можливістю повторного складання	Не зараховано

Критерії оцінювання роботи здобувача протягом семестру

Відмінно (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та уміти застосовувати їх.

«Відмінно» – відповідає високому (творчому) рівню компетентності:

- студент виявляє особливі творчі здібності, вміє самостійно здобувати знання, без допомоги викладача знаходить та опрацьовує необхідну інформацію, вміє використовувати набуті знання і вміння для прийняття рішень у нестандартних ситуаціях, переконливо аргументує відповіді, самостійно розкриває власні обдарування і нахили;

Добре (75-89). Твердо знати мінімум, захистити всі індивідуальні завдання, виконати всі КР, здати тестування та поза аудиторну самостійну роботу.

«Добре» – студент отримує за двома рівнями оцінювання залежно від набраної кількості балів та відповідає достатньому (конструктивно-варіативному) рівню компетентності:

- студент вільно володіє вивченим обсягом матеріалу, застосовує його на практиці, вільно розв'язує справи і задачі у стандартних ситуаціях, самостійно виправляє допущені помилки, кількість яких незначна;

- студент вміє порівнювати, узагальнювати, систематизувати інформацію під керівництвом викладача; в цілому самостійно застосовувати її на практиці; контролювати власну діяльність; виправляти помилки, серед яких є суттєві, добирати аргументи для підтвердження думок;

Задовільно (60-74). Показати мінімум знань та умінь. Захистити всі індивідуальні завдання та здати тестування.

«Задовільно» – студент отримує за двома рівнями оцінювання залежно від набраної кількості балів та відповідає середньому (репродуктивному) рівню компетентності:

- студент відтворює значну частину теоретичного матеріалу, виявляє знання і розуміння основних положень; з допомогою викладача може аналізувати навчальний матеріал, виправляти помилки, серед яких є значна кількість суттєвих;

- студент володіє навчальним матеріалом на рівні, вищому за початковий, значну частину його відтворює на репродуктивному рівні

Незадовільно (0-59) – відповідає низькому (рецептивно-продуктивному) рівню компетентності:

- студент не опанував навчальний матеріал дисципліни, не знає наукових фактів, визначень, майже не орієнтується в першоджерелах та рекомендованій літературі, відсутні наукове мислення, практичні навички не сформовані.

9. Політика навчального курсу

Відпрацювання пропущених занять відбувається відповідно до розкладу консультацій, за попереднім погодженням з викладачем. Питання, що стосуються академічної доброчесності розглядаються або викладачем, або за процедурою, визначеною в Положенні про академічну доброчесність.

10. Методичне забезпечення та інформаційні ресурси

Підручники, навчальні посібники, навчально-методичні посібники, конспекти лекцій, методичні рекомендації з проведення лабораторних робіт тощо, які видані в Університеті знаходяться за посиланням:

http://library.khai.edu/library/fulltexts/doc/_001Intelektualnij.pdf.

Сторінка дисципліни знаходиться за посиланням:

<https://mentor.khai.edu/course/view.php?id=3336>.

11. Рекомендована література

Базова

1. Провост Ф., Фоусетт Т. Data Science для бізнесу. К: Наш формат, 2019. 398 с.
2. Скиена С. Наука про дані: навчальний курс. К.: Діалектика, 2020. 544 с.
3. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підручник. К.: Знання, 2010. 836 с.
4. Сергеев-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних: комп'ютерний практикум. К.: КПІ ім. Ігоря Сікорського, 2018. 73 с.
5. Мыльников Л.А. Статистические методы интеллектуального анализа данных. СПб: БХВ-Петербург, 2021. 240 с.

Допоміжна

1. Бешекнов С., Ракитина Е. Моделирование и формализация [Текст]: метод. пособие. - М.: Лаборатория Базовых Знаний, 2002. с. 336.
2. Даншина С.Ю., Луханин М.И. Информационные сети: учеб. пособ.по лаб. практикуму. Х.: Нац. аерокосм. ун-т «Харьк. авіац. ін-т», 2006. 47 с.
3. ДСТУ 34.302.2-91 (ISO/МЕК 8859-2:1999). «Інформаційні технології. Набори 8-бітних однобайтових кодованих графічних символів. Латинський алфавіт № 2».
4. Danshyna S. Yu, Nechausov A.S. Data mining for Management of development project // Polish journal of science. 2019. Vol.1, № 15 (2019). P. 34 - 39.
5. Вентцель Е.С., Овчаров Л.А. Теория вероятности и ее инженерные приложения: учеб. для вузов. М.: Высш. шк., 2000. 480 с.