



Методи та технології аналітики даних

Робоча програма навчальної дисципліни (Силабус)

4. Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>11 Математика і статистика</i>
Спеціальність	<i>11 Прикладна математика</i>
Освітня програма	<i>Математичні методи криптографічного захисту інформації</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна(денна)</i>
Рік підготовки, семестр	<i>3 курс, осінній</i>
Обсяг дисципліни	<i>Загальна кількість: 120 годин / 4 кредити Лекційних занять: 36 годин Практичних занять: 18 годин Самостійна робота студентів: 66 годин</i>
Семестровий контроль/ контрольні заходи	<i>Залік, модульна контрольна робота</i>
Розклад занять	https://schedule.kpi.ua
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: доктор технічних наук, доцент, Смирнов Сергій Анатолійович, контактні дані e-mail: s.smirnov@kpi.ua Практичні: кандидат фізико-математичних наук, ст. викладач кафедри інформаційної безпеки, Рибак Олександр Владиславович, semperfi@ukr.net ; https://web.telegram.org/z/#1520456799</i>
Розміщення курсу	https://classroom.google.com/c/NTI3MTU3MDkyNDk5?cjc=yksviee

5. Програма навчальної дисципліни

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

Курс «Методи та технології аналітики даних» є вступом до науки про дані, а також присвячений методам і прийомам ефективної організації збору, аналізу й інтерпретації інформації, що потрібна різним сферам життєдіяльності. У першу чергу йдеться про розробку системного інструментарію інформаційно-аналітичної діяльності, спрямованого на ідентифікацію загроз та оцінку ризиків.

У курсі навчальної дисципліни «Методи та технології аналітики даних» розглядається широкий спектр питань, пов'язаних із методологією, організацією та технологіями інформаційно-аналітичної діяльності, спрямованих на забезпечення безпеки особи, суспільства та держави від загроз будь-якої природи, у тому числі й кіберзагроз.

Теоретичні матеріали курсу дають студенту знання про:

- життєвий цикл аналітики великих даних;
- концепції обробки та зберігання великих даних;
- структуровану і неструктуровану інформації;
- аналіз соціальних мереж, часові ряди і просторові змінні;
- методи машинного навчання для вирішення задач регресії;
- методи машинного навчання для здійснення класифікації даних;
- метод машинного навчання для здійснення кластеризації даних.

Зі студентами передбачено 9 практичних занять, що поглиблюють його. Передбачається їх вчасне виконання, а в разі перевищення зазначеного часу встановлений штраф: практичні роботи захищаються на мінімально позитивну оцінку. Дати дедлайнів обговорюються зі студентами на першому занятті. Унаслідок виконання практичних занять студент набуває такі уміння:

- розраховувати міри центральної тенденції;
- розраховувати міри статистичної невизначеності;
- здійснювати графічне представлення великих даних;
- застосовувати t-критерій в аналітичних дослідженнях;
- розраховувати коефіцієнти кореляції Пірсона і Спірмена;
- застосовувати модель лінійної регресії;
- застосовувати сучасні моделі ухвалення рішень.

Компетентності, набуті студентами під час вивчення курсу «Методи та технології аналітики даних», дадуть їм змогу не тільки набувати нові знання, але й поліпшувати старі, краще розуміти, пояснювати, прогнозувати і досліджувати складні явища та процеси, що відбуваються в усіх сферах життєдіяльності людини та суспільства.

Мета дисципліни: навчити студентів методам і прийомам прийняття рішень, заснованих на більш глибокому розумінні безпекових ситуацій за допомогою аналізу великих наборів даних.

Об'єкт дисципліни: методи та технології статистики, інформатики та машинного навчання для поліпшення як індивідуальної, так і колективної розумової діяльності.

Предмет дисципліни: статистичні методи і життєвий цикл великих наборів даних, алгоритми машинного навчання, спрямовані на ухвалення оптимальних управлінських рішень.

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі компетентності та програмні результати навчання за освітньою програмою:

Метою навчальної дисципліни є підсилення та закріплення у студентів наступних компетентностей.

Загальні компетентності (ЗК)

ЗК1 – Здатність учитися і оволодівати сучасними знаннями.

ЗК2 – Здатність застосовувати знання у практичних ситуаціях.

ЗК3 – Здатність генерувати нові ідеї (креативність).

ЗК4 – Здатність бути критичним і самокритичним.

ЗК6 – Здатність до абстрактного мислення, аналізу та синтезу.

Фахові компетентності (ФК)

ФК1 – Здатність використовувати й адаптувати математичні теорії, методи та прийоми для доведення математичних тверджень і теорем.

ФК2 – Здатність виконувати завдання, сформульовані у математичній формі.

ФК3 – Здатність обирати та застосовувати математичні методи для розв'язання прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень.

Програмні результати навчання

РН8 – Поєднувати методи математичного та комп'ютерного моделювання з неформальними процедурами експертного аналізу для пошуку оптимальних рішень.

РН10 – Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних.

РН11 – Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.

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

Дисципліна «Методи та технології аналітики даних» частково використовує знання та уміння, набуті під час вивчення курсів «Математичний аналіз», «Алгебра і геометрія», «Дискретна математика», «Диференційні рівняння», «Програмування» тощо, та поглиблює їх у напрямку математичного моделювання систем забезпечення кібербезпеки.

Отримані практичні навички та засвоєні теоретичні знання під час вивчення курсу «Методи та технології аналітики даних» використовуються в подальшому під час вивчення навчальних дисциплін спеціальності. Особливо в таких навчальних дисциплінах: «Системний аналіз», «Теорія керування», «Моделювання природних, економічних та соціальних процесів», «Аналіз даних», поглиблюючи знання у напрямку сучасних методів і моделей прикладної математики та

кібербезпеки. Результати вивчення даної дисципліни можуть бути застосовані для підготовки студентів як першого так і другого (магістерського) рівня вищої освіти, написання ними бакалаврських та магістерських дисертацій та у професійній діяльності за фахом.

3. Зміст навчальної дисципліни

Розділ 1 Інформатика й аналітична діяльність

- 1) Основні поняття та види аналітики
- 2) Традиційні та великі дані
- 3) Життєвий цикл аналітики великих даних
- 4) Концепції зберігання великих даних
- 5) Концепції обробки великих даних

Розділ 2 Методи машинного навчання та великі дані

- 6) Кластеризація методом k -середніх
- 7) Ієрархічна кластеризація
- 8) Асоціативні правила
- 9) Регресійний аналіз
- 10) Метод K -найближчих сусідів
- 11) Метод опорних векторів
- 12) Дерево рішень, ліс рішень
- 13) А/В-тестування, алгоритм многорукого бандита

Розділ 3. Методи статистичної обробки великих даних

- 14) Обробка неструктурованої інформації
- 15) Статистичні методи розвідувального аналізу
- 16) Аналіз соціальних мереж
- 17) Аналіз часових рядів і просторових змінних

6. Навчальні матеріали та ресурси

Базові

1. Оброблення надвеликих масивів даних (Big Data): навчальний посібник / Д. В. Ланде, І. Ю. Субач, А. Я. Гладун ; КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2021. – 168 с. <https://ela.kpi.ua/handle/123456789/46129>
2. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник / Д.В. Ланде, І.Ю. Субач, Ю.Є. Бояринова. –Київ: ІС33І КПІ ім. Ігоря Сікорського, 2018. – 300 с. <https://ela.kpi.ua/handle/123456789/45721>
3. P. Bruce, A. Bruce, P. Gedeck. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. O`Reilly Media, 363 p. <https://drive.google.com/file/d/1J4RgIgXgEIDXa5cgHm5u-NrX4HJaVdkG/view?usp=sharing>
1. W. Richert, L. P. Coelho, M. Brucher. Building Machine Learning Systems with Python: Explore machine learning and deep learning techniques for building intelligent systems using scikit-learn and TensorFlow, 3rd Edition. Packt Publishing. 2018. https://drive.google.com/file/d/1hQJU4ZMJ306aXWgDf4zijkdt_kXrQIT2/view?usp=sharing

Додаткові

2. University of California at Irvine Machine Learning Repository. <https://archive.ics.uci.edu/ml/index.php>

7. Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

У рамках дисципліни «Методи та технології аналітики даних» заплановані наступні види навчальних занять:

- лекційні заняття;
- практичні заняття;
- самостійна робота студентів.

На лекціях розкриваються найбільш суттєві теоретичні питання, що дозволяють забезпечити студентам можливість самостійного вивчення всього програмного матеріалу. Тематика та порядок самостійної роботи сформовані в логічній послідовності і цілком узгоджуються з метою дисципліни та здійснюються з використанням рекомендованої літератури та глобальної мережі Інтернет. На заняттях використовуються звичайна дошка, а також презентації лекцій з використанням мультимедіа-проектора. В дистанційному режимі використовуються засоби Zoom та відповідні слайди лекцій.

Навчання здійснюється на основі студентоцентрованого підходу та стратегії взаємодії викладача та студента для засвоєння студентами матеріалу та розвитку у них практичних навичок. Для проведення занять застосовується практичний метод. Для лекційних занять використовуються пояснювально-ілюстративний метод та метод проблемного виконання, для проведення лабораторних робіт використовується частково-пошуковий та дослідницький методи навчання, при яких викладач ставить перед студентами проблему, і ті вирішують її самостійно або під керівництвом викладача, висувачи ідеї, перевіряючи їх, підбираючи для цього необхідні джерела інформації, методи, підходи тощо.

Тематику та порядок освоєння дисципліни «Методи та технології аналітики даних» наведений нижче.

№ з/п	Тема	Кількість аудиторних годин, лекції	Кількість аудиторних годин, практичні заняття	Кількість годин, СРС
1.	8. Основні поняття та види аналітики	1	1	3
2.	Традиційні та великі дані	2	1	3
3.	Життєвий цикл аналітики великих даних	2	1	3
4.	9. Концепції обробки великих даних	2	1	3
5.	10. Концепції зберігання великих даних	2	1	3
6.	11. Обробка неструктурованої інформації	2	1	3
7.	12. Статистичні методи розвідувального аналізу	2	1	3
8.	13. Аналіз соціальних мереж	1	1	3
9.	14. Аналіз часових рядів і просторових змінних	2	1	3
10.	15. Кластеризація методом k -середніх змінних	1	1	3
11.	16. Ієрархічна кластеризація	2	1	3
12.	17. Асоціативні правила	2	1	3
13.	18. Регресійний аналіз	2	1	3
14.	19. Метод K -найближчих сусідів	1	1	3
15.	20. Метод опорних векторів	2	1	3
16.	21. Дерево рішень, ліс рішень	2	1	3
17.	22. А/В-тестування, алгоритм многорукого бандита	2	1	3
18.	23. Нейронні мережі	2	1	3
19.	Разом	32	18	54
20.	Модульна контрольна Робота	2		6
21.	Залік (залікова співбесіда)	2		6
22.	Всього 120	36	18	66

Теми практичних занять

№ з/п	Теми практичних занять	Кількість аудиторних годин
1.	Великі дані. Методи графічного представлення даних	2
2.	Концепції обробки ВД, Hadoop, MapReduce	2
3.	Кластеризація методом К-середніх	2
4.	Ієрархічна кластеризація	2
5.	Класифікація даних методом опорних векторів	2
6.	Класифікація методом KNN	2
7.	Лінійна регресія	2
8.	Оцінки невідомого розподілу	2
9.	Статистична перевірка гіпотез	2

24. Політика та контроль

6. Самостійна робота здобувача

25. Самостійна робота здобувача складається з опанування питань лекційного матеріалу, підготовки до практичних занять, модульної контрольної роботи і заліку.

7. Політика навчальної дисципліни (освітнього компонента)

26. Відвідування занять не оцінюється, але рекомендується. Контроль відвідування здійснюється викладачем систематично.

27. Завдання практичних занять виконуються та захищаються у відповідності до встановлених дедлайнів впродовж семестру.

28. Під час виконання практичних занять, а також під час контрольних заходів здобувачами повинна виконуватися політика академічної доброчесності, згідно Кодексу Честі НТУУ «КПІ».

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

Поточний контроль: виконання та захист практичних занять.

Календарний контроль проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу. Календарний контроль базується на поточній рейтинговій оцінці. Умовою позитивної атестації є значення поточного рейтингу студента не менше 50% від максимально можливого на час атестації. Бал, необхідний для отримання позитивного календарного контролю доноситься до студентів викладачем не пізніше ніж за 2 тижні до початку календарного контролю.

Практичні заняття розроблені таким чином, що їх завдання сприяють засвоєнню матеріалу за темами дисципліни «Методи та технології аналітики даних» та формуванню практичних знань і навичок.

Залік проставляється за результатами роботи здобувача у семестрі, бал формується як сума балів практичних робіт, модульної контрольної роботи. У разі коли здобувач виконав усі поточні контрольні заходи і має поточний рейтинг більший 50 балів ($RD \geq 50$), він допускається до залікової співбесіди з метою покращення балу до мінімально достатнього (60). Якщо результати співбесіди є негативними, студент отримує оцінку згідно зі своїм рейтингом. Також допускається підвищення семестрового балу у спосіб залікової співбесіди, але не більше ніж на 10 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

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

Рейтингова система оцінювання

Контрольний захід	Макс. бал	Ваговий коефіцієнт	Кількість	Всього
МКР	11	5	1	55
Практичні заняття	5	1	9	45
				100

Таблиця переведення рейтингових балів до оцінок за університетською шкалою

Рейтингові бали, RD	Оцінка за університетською шкалою	Можливість отримання оцінки «атоматом»
$95 \leq RD \leq 100$	Відмінно	-
$85 \leq RD \leq 94$	Дуже добре	-
$75 \leq RD \leq 84$	Добре	-
$65 \leq RD \leq 74$	Задовільно	-
$60 \leq RD \leq 64$	Достатньо	-
$RD < 60$	Незадовільно	-

Залік та робота у семестрі

Залік проставляється за результатами роботи в семестрі. Рейтингова оцінка роботи за семестр складається з результатів МКР та практичними заняттями. Студент може поліпшити оцінку у спосіб залікової співбесіди, однак не більше ніж на 10 балів.

Заохочувальні бали	
Критерій	Додається до семестрового рейтингу
Присутність та активність на більшості занять	+ 5 балів

9. Додаткова інформація з дисципліни (освітнього компонента)

Питання, що виносяться на МКР та залікову співбесіду цілком відповідають тим, що були зазначені в змісті дисципліни.

Робочу програму навчальної дисципліни (силабус):

Склав: доцент кафедри інформаційної безпеки, к.ф.-м.н., с.н.с., Смирнов Сергій Анатолійович

Ухвалено кафедрою ІБ (протокол № 5 від 22.06.2022)

Погоджено Методичною комісією НН ФТІ (протокол № 6__ від _30.06.2022)