Силабус освітнього компонента
Програма навчальної дисципліни

Дата майнинг

Шифр та назва спеціальності
125 – Кібербезпека та захист інформації

Інститут
ННІ комп`ютерних наук та інформаційних технологій

Освітня програма
Кібербезпека

Кафедра
Кібербезпеки (328)

Рівень освіти
Бакалавр

Тип дисципліни
Профільна підготовка, Вибіркова

Семестр
4

Мова викладання
Українська

Викладачі, розробники

МІЛОВ Олександр Володимирович
oleksandr.milov@khpi.edu.ua
Доктор технічних наук, професор кафедри кібербезпеки НТУ «ХПІ».

Автор понад 200 наукових та навчально-методичних праць. Науковий керівник з захищених кандидатських робіт, гарант освітньо-професійної програми другого (магістерського) рівня вищої освіти. Провідний лектор з дисциплін: «Математичні основи криптології та криптоаналіз», «Структури даних», «Промисловий та офісний шпіонаж» , «Цифрова криміналістика», у студентів бакалавріата та магістратури, Розділ «Методологія наукової та педагогічної діяльності в науках кіберзахисту» для аспірантів.
Детальніше про викладача на сайті кафедри

МІЛЕВСЬКИЙ Станіслав Валерійович
stanislav.milevskyi@khpi.edu.ua
кандидат економічних наук, доцент кафедри кібербезпеки НТУ «ХПІ».

Автор понад 100 наукових та навчально-методичних праць. Науковий Гарант освітньо-наукової програми другого (магістерського) рівня вищої освіти. Провідний лектор з дисциплін: «Основи математичного моделювання систем безпеки», «Англійська мова в академічних застосунках», «Моделювання кіберфізичних дій» у студентів бакалавріата та магістратури.
Детальніше про викладача на сайті кафедри

Загальна інформація

Анотація

Навчальна дисципліна "Дата майнинг" є вибірковою навчальною дисципліною. Вивчення дисципліни допомагає формуванню знань та вмінь в області аналізу даних з метою розробки повної моделі функціонування та розвитку реального бізнесу.

Мета та цілі дисципліни

Формування у студентів системи теоретичних знань і практичних навичок з основ, принципів та методів інтелектуального аналізу даних. В результаті освоєння дисципліни студенти отримають фундаментальний базис знань з основ, сучасної методології та особливостей застосування інтелектуальної обробки даних; виконають вивчення та опанування стандартів Data Mining; набудуть уміння роботи з системами Data Mining різного призначення і різної проблемної орієнтації; отримають практичні навички до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач систем та процесів захисту інформації для формування контуру безпеки бізнес-процесів в комп'ютерних системах на основі технологій data-mining та запобіганню зовнішніх кібер-загроз.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

КЗ–6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
КЗ–7. Здатність використовувати інформаційні та комунікаційні технології та формувати систему інформаційно-аналітичного забезпечення підтримки прийняття управлінських рішень щодо системи управління інформаційною безпекою.
ФК-2. Здатність аналізувати та визначати політику та стратегії забезпечення захисту інформації.
ФК-3. Проектувати системи управління та захисту інформації на підприємстві установі, організації.
ФК-5. Здатність узагальнення вітчизняного та закордонного досвіду з питань управління інформаційною безпекою.
ФК-7. Здатність організовувати та проводити аналіз оточення організації установ з метою виявлення та закриття можливих каналів витоку інформації.
ФК-8. Здатність використовувати механізми забезпечення управління інформаційною безпекою у її визначальних сферах.
ФК-9. Здатність організації реагування на загрози на об'єктах критичної інфраструктури, установах та підприємствах.
ФК-10. Здатність забезпечувати неперервність бізнесу згідно з встановленою політикою інформаційної безпеки.
ФК-11. Здатність впроваджувати та забезпечувати функціонування комплексних систем захисту інформації.

Результати навчання

ПРН-3. Вміти за допомогою абстрактного мислення, аналізу та синтезу оцінювати результати професійної діяльності та забезпечувати її якість, бути критичним і самокритичним, наполегливим щодо поставлених завдань і взятих зобов'язань.
ПРН-6. Впроваджувати процеси, що базуються на національних та міжнародних стандартах, виявлення, ідентифікації, аналізу та реагування на інциденти інформаційної безпеки.
ПРН-7. Вміти розробляти комплекс організаційних заходів щодо формування системи управління інформаційною безпекою.
ПРН-8. Вміти використовувати сучасне програмно-апаратне забезпечення інформаційно-комунікаційних технологій щодо формування системи управління інформаційної безпеки.

ПРН-10. Вміти аналізувати виклики та загрози інформаційної безпеки об’єктів критичної інфраструктури та синтезувати інформацію щодо розроблення та реалізації стратегій та політики безпеки.
ПРН-11. Вміти забезпечувати процеси захисту та функціонування системи управління інформаційною безпекою та захисту інформації на основі практик, навичок та знань, щодо інфраструктури кіберфізичних систем та інформаційних потоків.
ПРН-12. Вміти використовувати програмні та програмно-апаратні комплекси захисту інформаційних ресурсів.
ПРН-13. Вміти вирішувати задачі забезпечення та супроводу (в.т. числі: огляд, тестування, підзвітність) системи управління інформаційної безпеки згідно встановленої політики безпеки в інформаційно-комунікаційних системах.
ПРН-14. Вміти вирішувати задачі управління процедурами ідентифікації, автентифікації, авторизації процесів та користувачів в інформаційно-комунікаційних системах згідно встановленої політики інформаційної безпеки.
ПРН-16. Вміти реалізовувати заходи з протидії отриманню несанкціонованого доступу до інформаційних ресурсів і процесів в інформаційно-комунікаційних системах.
ПРН-17. Вміти розв’язувати задачі управління інфораційною безпекою в інформаційно-комунікаційних системах на основі моделей управління безпекою.
ПРН-18. Розуміти основні теоретичні поняття, застосовувати набуті практичні навички дослідження та підготовки документів, їх правильного використання в управлінській діяльності.
ПРН-19. Вміти впроваджувати заходи та забезпечувати реалізацію процесів попередження отриманню несанкціонованого доступу і захисту інформації в інформаційно-мунікаційних системах.
ПРН-20. Вміти аналізувати та проводити оцінку ефективності та рівня захищеності інформаційних ресурсів в інформаційно-комунікаційних системах згідно встановленої політики інформаційної безпеки.
ПРН-21. Вміти застосовувати теорії та методи захисту для забезпечення безпеки елементів об’єктів критичної інфраструктури, кіберфізичних систем та інформаційно-комунікаційних систем.
ПРН-22. Вміти застосовувати національні та міжнародні регулятори в сфері інформаційної безпеки щодо розслідування комп’ютерних інцидентів.

Обсяг дисципліни

Загальний обсяг дисципліни 90 год. (3 кредити ECTS): лекції – 16 год., лабораторні роботи – 16 год., самостійна робота – 58 год.

Передумови вивчення дисципліни (пререквізити)

Інформаційна безпека держави, Математичні основи криптології та криптоаналіз, Основи криптографічного захисту, Основи програмування.

Особливості дисципліни, методи та технології навчання

В ході викладання дисципліни викладачем застосовуються пояснювально-ілюстративний (інформаційно-рецептивний) та репродуктивний методи навчання. В якості методів викладання, які направлені на активізацію та стимулювання навчально-пізнавальної діяльності здобувачів, застосовуються презентації, бесіди, індивідуальні групові проєкти, майстер-класи.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Інтелектуальний аналіз даних – DataMining.
Докладно розглядаються поняття DataMining. Описано виникнення, перспективи, проблеми DataMining. Наведен погляд на технологію DataMining як на частину ринку інформаційних технологій. Докладно розглядається поняття даних. Пояснюється значення понять об'єкт та атрибут, вибірка, залежна й незалежна змінна. Докладно обговорюються типи шкал. Приводяться різні типи наборів даних. Коротко розглянуті поняття бази даних і СУБД. Описуються стадії

DataMining та дії, які виконуються в рамках цих стадій. Розглянуто відомі класифікації методів DataMining. Наведено порівняльну характеристику деяких методів, заснована на їхніх властивостях. Характеризується основна суть завдань DataMining й їхня класифікація. Докладно розглянуті поняття "інформація", "знання", а також зіставлення й порівняння цих понять.

Тема 2. Візуальний аналіз даних — VisualMining.

Розглядаються питання візуального аналізу даних. Наведені характеристики засобів візуалізації даних, методів візуалізації та методів геометричних перетворень. Порівнюються методи, орієнтовані на пікселі, а також методи аналізу ієрархічних образів та відображення іконок. Розглядаються методи й засоби візуального подання інформації, зокрема, способи подання інформації в одне-, двох-, тривимірному вимірах, а також способи відображення інформації в більш ніж трьох вимірах. Описано принципи якісної візуалізації. Викладено основні тенденції в області візуалізації.

Тема 3. Аналіз текстової інформації — TextMining.

Формулюються задачі аналізу текстів (етапи аналізу текстів, попередня обробка тексту, задачі TextMining). Розглядаються етапи аналізу текстів, такі як витяг ключових понять із тексту (загальний опис процесу витягу понять із тексту, стадія локального аналізу, стадія інтеграції й висновку понять), класифікація текстових документів (опис задач класифікації текстів, методи класифікації текстових документів), методи кластерізації текстових документів (наведення текстових документів, ієрархічні методи кластерізації текстів, бінарні методи кластерізації текстів), анотування текстів (виконання анотування текстів, методи витягу фрагментів для анотації). Порівнюються різноманітні засоби аналізу текстової інформації (засоби Oracle - Oracle Text, засоби від IBM - Intelligent Miner for Text, засоби SAS Institute - Text Miner, засоби Мега-комп`ютер Інтеліджене - Text Analyst).

Тема 4. Витяг знань з Web – WebMining.

Розглянуті проблеми аналізу інформації з Web, етапи WebMining, WebMining та інші інтернет-технології, а також категорії WebMining. Описані методи витягу Web-контенту (витяг Web-контенту в процесі інформаційного пошуку, витяг Web-контента для формування баз даних), а також методи витягу Web-структур (представлення Web-структур, оцінка важливості Web-структур, пошук Web-документів з урахуванням гіперпосилань, кластеризація Web-структур). Наведені результати досліджень використання Web-ресурсів (дослідницька інформація, етап препроцессінгу, етап витягу шаблонів, етап аналізу шаблонів та їхнє застосування).

Тема 5. Засоби аналізу процесів – ProcessMining.

Розглянуті засоби автоматизації виконання бізнесів-процесів (бізнес-процеси, формалізація бізнес-процесів, Workflow-системи, сервісно-орієнтована архітектура, проектування бізнес-процесів). Виконан аналіз процесів (технологія ProcessMining, аналіз протоколів, стандарт запису протоколів MXML, задачі ProcessMining, проблеми аналізу протоколів). Порівнюються методи ProcessMining (перші імовірнісні методи ProcessMining, метод побудови диз'юнктивної Workflow-схеми, (α-алгоритм, методи на основі генетичних алгоритмів). Описана бібліотека алгоритмів Process Mining-Pro (архітектура Pro, ProImport Framework).

Тема 6. Пошук асоціативних правил – RulesMining.

Виконана постановка задачі. Розглянуті форми подання результатів (правила класифікації, дерева класифікації, математичні функції), методи побудови правил класифікації (алгоритм побудови 1-правил, метод NaiveBayes), а також методи побудови дерев класифікації (методика "розділяй і пануй", алгоритм покриття), методи побудови математичних функцій (загальний вид, лінійні методи, метод найменших квадратів, нелінійні методи, SupportVectorMachines (SVM), регулярізаційні мережі (RegularizationNetworks), дискретизації й рідкі сітки). Розглянута постановка задачі пошуку асоціативних правил (формальна постановка задачі, секвенціальний аналіз, різновиди задач пошуку асоціативних правил), алгоритми (алгоритм Apriori, різновид алгоритму Apriori).

Тема 7. Розподілений аналіз даних.

Розглядаються системи мобільних агентів (основні поняття, стандарти багатоагентних систем, системи мобільних агентів, система мобільних агентів JADE). Продемонстровано використання мобільних агентів для аналізу даних (проблеми розподіленого аналізу даних, агенти-аналітики, варіанти аналізу розподілених даних). Побудована система аналізу розподілених даних (загальний підхід до реалізації системи, агент для збору інформації про базу даних, агент для збору статистичної інформації, агент для вирішення одного завдання інтелектуального аналізу даних, агент для вирішення інтегрованого завдання інтелектуального аналізу даних).

Теми практичних занять

Практичні роботи в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Навички роботи з програмою Deductor.
Тема 2. Візуальний аналіз даних — VisualMining.
Тема 3. Аналіз текстової інформації — TextMining. Програми Text Analyzer, VaalMini.
Тема 4. Витяг знань з Web – WebMining. Програма C5.4.
Тема 5. Засоби аналізу процесів – ProcessMining. Використання ProM.
Тема 6. Пошук асоціативних правил – RulesMining. Програма See 5.
Тема 7. Розподілений аналіз даних. Агентне моделювання за допомогою JADE та Xelopes.

Самостійна робота

Самостійна робота студента є однією з форм організації навчання, основною формою оволодіння навчальним матеріалом у вільний від аудиторних навчальних занять час. Під час самостійної роботи студенти вивчають лекційний матеріал, готуються до лабораторних робіт, контрольних робіт, заліків та іспитів. Студентам також рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення та аналізу.

Неформальна освіта

В рамках неформальної освіти згідно відповідного Положення (http://surl.li/pxssv), освітня компонента або її окремі теми можуть бути враховано у разі самостійного проходження професійних курсів/тренінгів, отримання громадянської освіти, онлайн освіти, професійного стажування тощо.
За даним компонентом врахування тем, у разі успішного завершення курсів, не передбачено.

Література та навчальні матеріали

Основна література

1 Черняк О. І. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. – К. : Знання, 2014. – 599 с.
https://moodle.znu.edu.ua/pluginfile.php/593075/mod_folder/intro/%D0%91%D0%B0%D0%B7%D0%BE%D0%B2%D0%B8%D0%B9%20%D0%BF%D1%96%D0%B4%D1%80%D1%83%D1%87%D0%BD%D0%B8%D0%BA_3%20%28%D0%A7%D0%B5%D1%80%D0%BD%D1%8F%D0%BA%20%D0%9E.%20%D0%86.%20%D0%86%D0%BD%D1%82%D0%B5%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9%20%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7%20%D0%B4%D0%B0%D0%BD%D0%B8%D1%85%29.pdf
2. Data Mining : пошук знань в даних / Гладун А. Я., Рогушина Ю. В. – К. : ТОВ «ВД «АДЕФ-Україна», 2016. – 452 с.
https://www.researchgate.net/publication/304025285_Data_Mining_Search_for_Knowledge_in_Data
3. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.
https://openarchive.nure.ua/server/api/core/bitstreams/2e55d639-52fd-48d9-b7b7-14989f49f291/content
4. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.
https://ru.scribd.com/document/460020140/%D0%9E%D0%BB%D1%96%D0%B9%D0%BD%D0%B8%D0%BA-%D0%90-%D0%9E-%D0%86%D0%BD%D1%82%D0%B5%D0%BB%D0%B5%D0%BA%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9-%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D0%B7-%D0%B4%D0%B0%D0%BD%D0%B8%D1%85
5. Прикладні задачі інтелектуального аналізу даних (DATA MINING). – К.: КНУ ім. Тараса Шевченко, 2018. – 152 с.

https://www.researchgate.net/profile/Vitalii-Akimenko/publication/325474310_DATA_MINING/links/5b1024bb4585150a0a5deaf6/DATA-MINING.pdf
6. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для студентів / В. Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
http://kist.ntu.edu.ua/textPhD/metDataManing.pdf

Додаткова література

7. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. — К.: ІСЗЗІ КПІ ім. Ігоря Сікорського», 2018. — 300 с.
http://dwl.kiev.ua/art/oiad/oiad.pdf
8. Любунь З. М. Основи теорії нейромереж / З. М. Любунь /: Текст лекцій. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2007. –142 с.
https://f.eruditor.link/file/236389/
9. Liubun Z. Hover Signal-Profile Detection / Liubun, V. Mandziy, H. Klein, O. Karpin, V. Rabyk // Proceedings of the XV International Scientific and Technical Conference “Computer Science and Information Technologies” – 2020. P. 7 – 10. (Scopus).
10. Бобало Ю.Я., Горбатий І.В. (ред.) Інформаційна безпека. Навчальний посібник. — Львів : Видавництво Львівської політехніки, 2019. — 580 с. — ISBN 978-966-941-339-0.
https://pdf.lib.vntu.edu.ua/books/2021/Bobalo_2019_580.pdf

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Бали нараховуються за наступним співвідношенням:
•лабораторні роботи: 40% семестрової оцінки;
•самостійна робота: 10% семестрової оцінки;
•контрольна робота: 10% семестрової оцінки;
•іспит: 40% семестрової оцінки.

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Погодження

Силабус погоджено	28.08.2024		Завідувач кафедри Сергій ЄВСЕЄВ
	28.08.2024		Гарант ОП Сергій ЄВСЕЄВ