Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Фізико-технічний інститут
Кафедра інформаційної безпеки

Аналіз кіберінцидентів методами машинного навчання
Робоча програма навчальної дисципліни
(Завдання до практичних робіт)

Тема 1: Логістична регресія в системах машинного навчання

Використовуючи модель логістичної регресії, виконати класифікацію програмного забезпечення на шкідливе та безпечне на даних датасету https://www.kaggle.com/datasets/piyushrumao/malware-executable-detection.

Тема 2: Дерево рішень і ліс дерев рішень в системах машинного навчання

Застосувати алгоритми дерева рішень та лісу дерев рішень до трьох датасетів https://www.kaggle.com/datasets?search=security, проаналізувати переваги та недоліки наведених алгоритмів, зробити висновки щодо якості класифікації різних даних в залежності від особливостей даних у датасетах.

Тема 3: Метод опорних векторів в системах машинного навчання

Реалізувати алгоритм методу опорних векторів, використавши один з датасетів https://www.kaggle.com/datasets?search=security, оцінити точність та фактори, від яких вона може залежати на прикладі обраного датасету.

Тема 4: Наївний байєсівський класифікатор в системах машинного навчання

Створити спам-фільтр на основі наївного баєсівського класифікатора на основі даних датасету з підбірки https://www.kaggle.com/datasets?search=spam. Дослідити проблему нульової імовірності для алгоритму Баєса.

Тема 5: Метод K-найближчих сусідів в системах машинного навчання

Реалізувати алгоритм К найближчих сусідів на одному з датасетів https://www.kaggle.com/datasets?search=security, визначити його переваги та недоліки, дослідити залежність точності алгоритму від особливостей використаного датасету.

Тема 6: Метод k-середніх, Ієрархічна кластеризація в системах машинного навчання

Виконати реалізацію алгоритмів k-середніх та ієрархічної кластеризації на одному з датасетів на вибір https://www.kaggle.com/datasets?search=security. Дослідити залежність точності методу від параметрів даних.

Тема 7: Метод чутливого розташування хешування в системах машинного навчання

Застосувати LSH до одного з датасетів https://www.kaggle.com/datasets?search=image+classification, порівняти із результатами класифікації за допомогою будь-якого іншого з раніше реалізованих алгоритмів.

Тема 8: Метод DBSCAN і A/B-тестування в системах машинного навчання

На довільному датасеті дослідити роботу DBSCAN і A/B тестування https://www.kaggle.com/datasets?search=security.

Тема 9. **Нейронні мережі в системах машинного навчання**

Створити нейронну мережу, яка може розпізнавати рукописні цифри зображені на зображеннях. Використовуйте набір даних, такий як MNIST, що містить тисячі зображень рукописних цифр разом з їх відповідними мітками. Навчіть нейронну мережу на цьому наборі даних та оцініть її ефективність у розпізнаванні нових зображень цифр.