

Лабораторна робота №2. Дерева прийняття рішень

Підгрупа №1

1. Відкрити та зчитати наданий файл з даними.
2. Визначити та вивести кількість записів та кількість полів у завантаженому наборі даних.
3. Вивести перші 10 записів набору даних.
4. Розділити набір даних на навчальну (тренувальну) та тестову вибірки.
5. Використовуючи відповідні функції бібліотеки *scikit-learn*, збудувати класифікаційну модель дерева прийняття рішень глибини 5 та навчити її на тренувальній вибірці, вважаючи, що в наданому наборі даних цільова характеристика визначається останнім стовпчиком, а всі інші виступають в ролі вихідних аргументів.
6. Представити графічно побудоване дерево за допомогою бібліотеки *graphviz*.
7. Обчислити класифікаційні метрики збудованої моделі для тренувальної та тестової вибірки. Представити результати роботи моделі на тестовій вибірці графічно. Порівняти результати, отриманні при застосуванні різних критеріїв розщеплення: інформаційний приріст на основі ентропії чи неоднорідності Джині.
8. З'ясувати вплив глибини дерева та мінімальної кількості елементів в листі дерева на результати класифікації. Результати представити графічно.
9. Навести стовпчикову діаграму важливості атрибутів, які використовувалися для класифікації (див. *feature_importances_*). Пояснити, яким чином – на Вашу думку – цю важливість можна підрахувати.