Изображение выглядит как строительство, зарисовка, черно-белый, силуэт

Автоматически созданное описание

Міністерство освіти і науки України Національний технічний університет України

“Київський політехнічний інститут імені Ігоря Сікорського” Факультет інформатики та обчислювальної техніки Кафедра інформаційних систем та технологій

# Лабораторна робота №1

З дисципліни «Технології машинного навчання»

Тема: «Метрики якості задач класифікації»

Виконав: Перевірив:

студент групи ІА – 12 пос. Коломоєць С. О. Мельник Михайло Сергійович

Дата здачі

Київ-2024

Мета**:** Метою даної лабораторної роботи є отримання знань основних метрик якості бінарної класифікації і варіантів тонкого налаштування алгоритмів класифікації.

Завдання:

1. Завантажте файл bioresponse.csv з папки “Лаб 1” за посиланням <https://drive.google.com/drive/folders/1SYLvcaJ5XZQL2socPuGc6466AoEsrLue?usp=sharing>
2. Навчіть 4 класифікатора, щоб передбачити поле "Activity" (біологічна відповідь молекули) з набору даних "bioresponse.csv":
   1. дрібне дерево рішень;
   2. глибоке дерево рішень;
   3. випадковий ліс на дрібних деревах;
   4. випадковий ліс на глибоких деревах.
3. Розрахуйте наступні метрики, щоб перевірити якість ваших моделей:
   1. частка правильних відповідей (accuracy);
   2. точність;
   3. повнота;
   4. 𝐹1-𝑠𝑐𝑜𝑟𝑒;
   5. *log-loss*.
4. Побудуйте precision-recall і ROC-криві для ваших моделей.
5. Навчіть класифікатор, який уникає помилок II роду і розрахуйте для нього метрики якості.

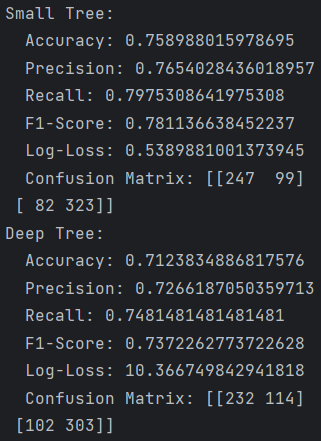
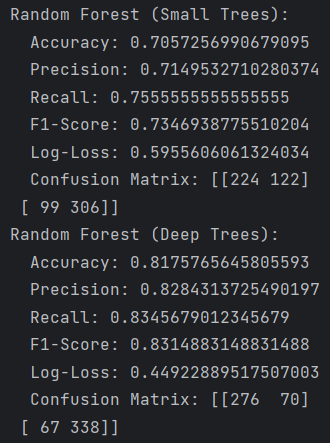
# Хід роботи

**Виконання завдання:**

1. Було навчено 4 класифікатора, взятих з бібліотеки Scikit-learn:

small\_tree = DecisionTreeClassifier(max\_depth=3)  
deep\_tree = DecisionTreeClassifier(max\_depth=None)  
rf\_small\_trees = RandomForestClassifier(n\_estimators=100, max\_depth=3)  
rf\_deep\_trees = RandomForestClassifier(n\_estimators=100, max\_depth=None)

1. Було розраховано наступні метрики, для перевірки якості отриманих моделей:

1. Було побудовано precision-recall і ROC-криві для отриманих моделей:

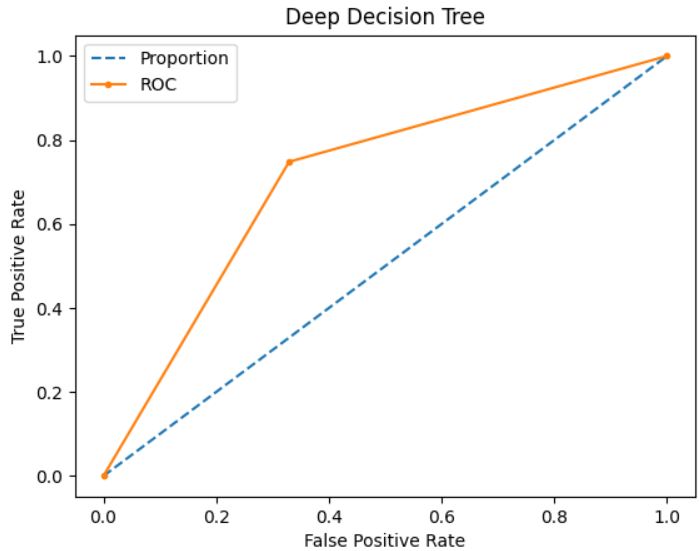
Small Decision Tree

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание Изображение выглядит как текст, линия, снимок экрана, диаграмма

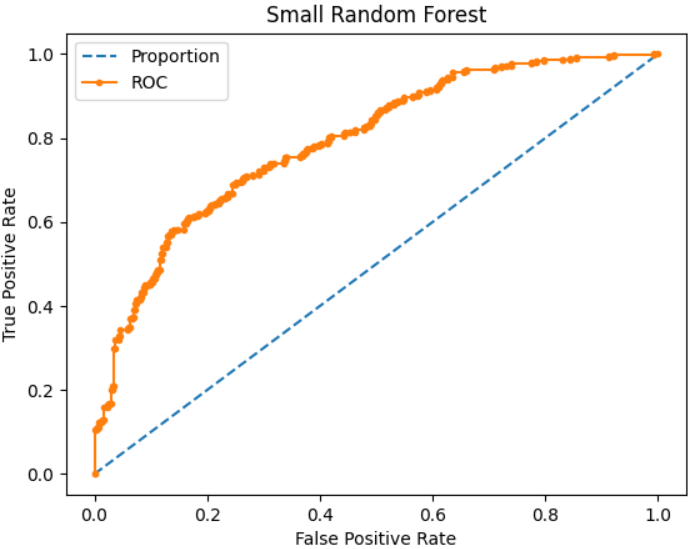
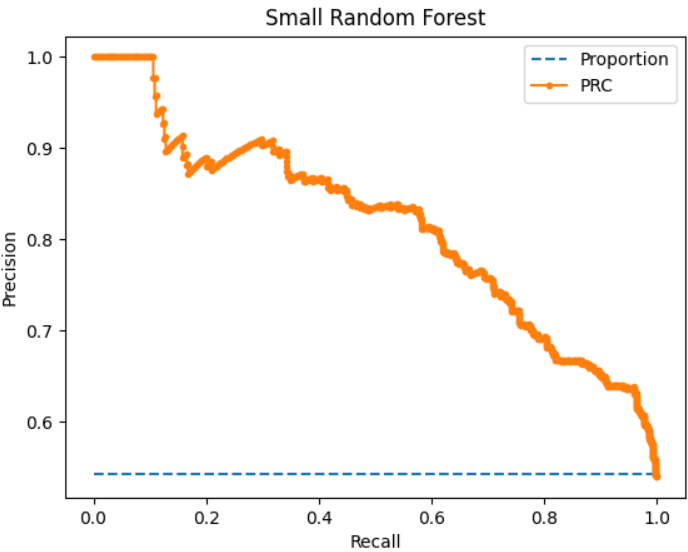
Автоматически созданное описание

Deep Decision Tree

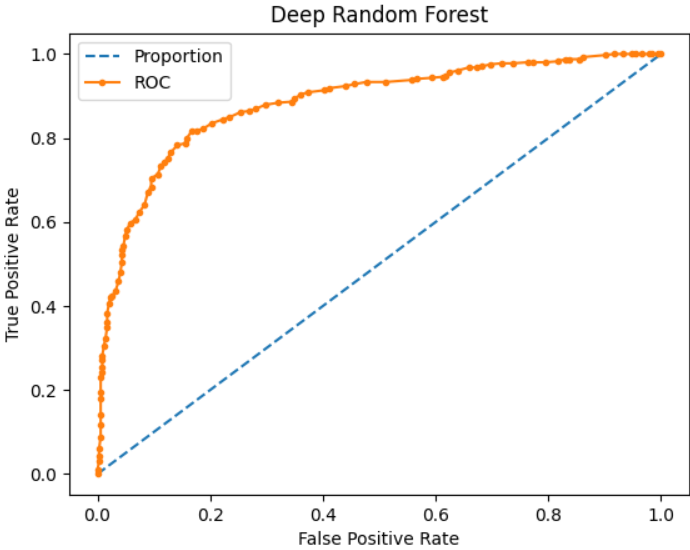
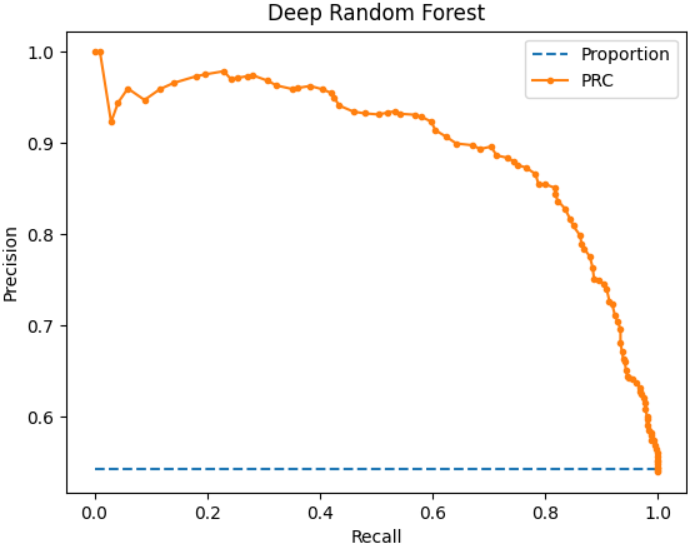
 Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание

Small Random Forest

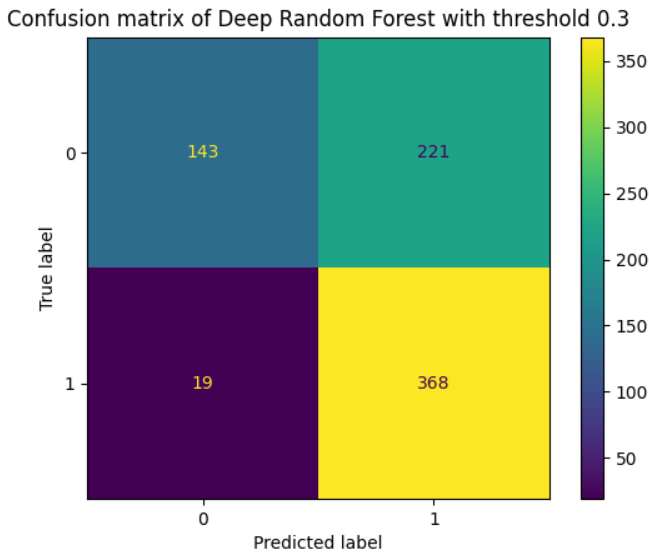
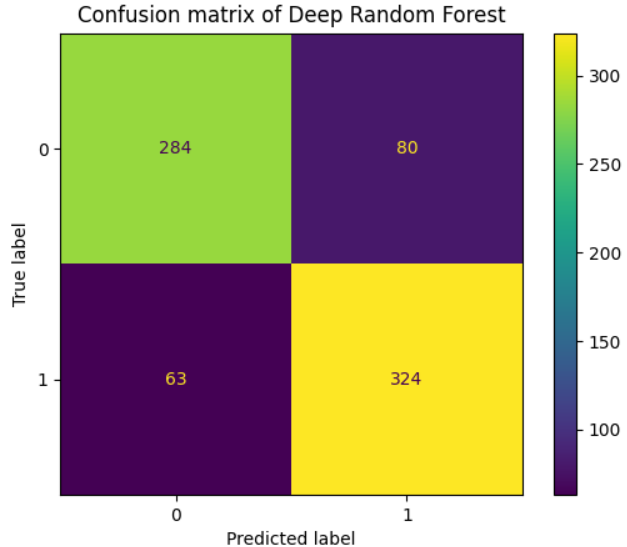
Deep Random Forest

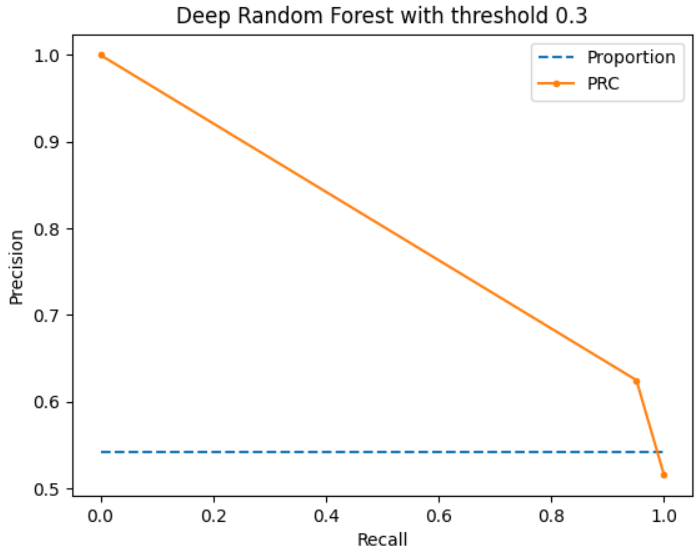
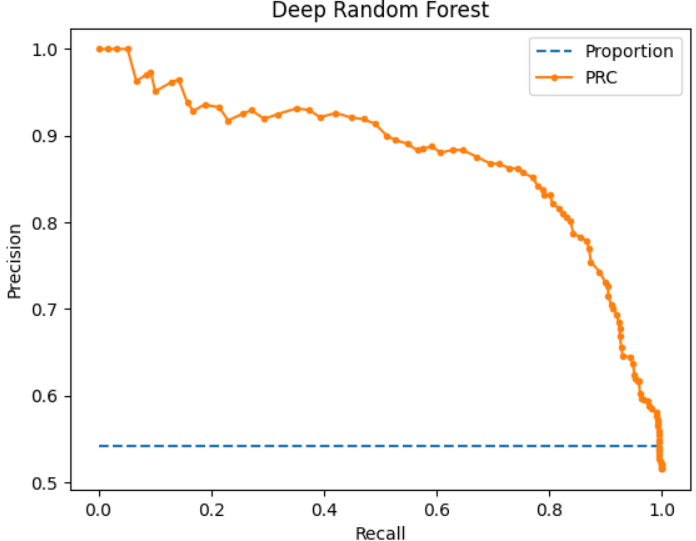
 

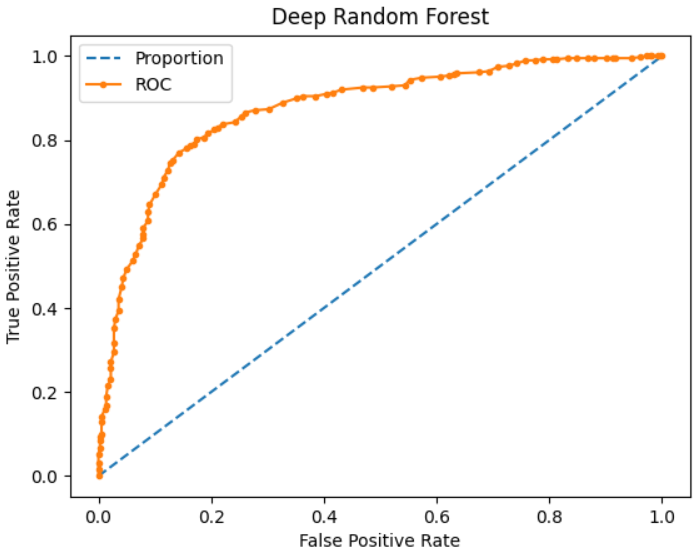
1. Було навчено класифікатор, який уникає помилок II роду, і розраховано для нього метрики якості:

Для уникнення помилки II роду було прийнято рішення зменшення порога 𝑇(ми отримуємо більше 𝐹𝑝 і менше **𝐹𝑛** помилок) для класифікаторів, що позначилося на їхніх вихідних confusion matrix та метриках якості.

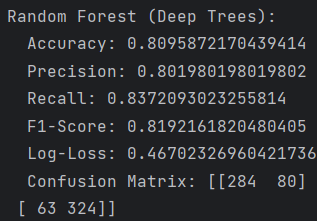
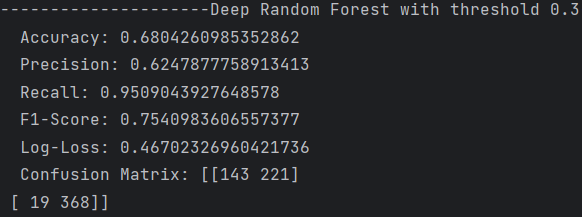
Найближчий до ідеального результат показав класифікатор Deep Random Forest зі значенням порогу 0.3:





 Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание

Висновки: в ході виконання цієї лабораторної роботи я отримав знання основних метрик якості бінарної класифікації і варіантів тонкого налаштування алгоритмів класифікації.