

Министерство образования Республики Беларусь
Учреждение образования
«Брестский Государственный технический университет»
Кафедра ИИТ

Лабораторная работа №5
По дисциплине «ИАД»
Тема: «Деревья решений»

Выполнил:
Студент 4 курса
Группы ИИ-24
Крейдич А. А.
Проверила:
Андренко К. В.

Брест 2025

Цель: На практике сравнить работу нескольких алгоритмов одиночного дерева решений, случайного леса и бустинга для деревьев решений.

Задачи

1. Загрузить датасет по варианту;
2. Разделить данные на обучающую и тестовую выборки;
3. Обучить на обучающей выборке одиночное дерево, случайный лес и реализовать бустинг для решающих деревьев (AdaBoost, CatBoost, XGBoost);
4. Оценить точность каждой модели на тестовой выборке;
5. Сравнить результаты, сделать выводы о применимости каждого метода для данного набора данных.

Задание по вариантам:

Вариант 6

- Pima Indians Diabetes
- Предсказать наличие диабета у пациенток
- Задания:
 1. Загрузите данные и стандартизируйте их;
 2. Разделите выборку в соотношении 70/30;
 3. Обучить на обучающей выборке одиночное дерево, случайный лес и реализовать бустинг для решающих деревьев (AdaBoost, CatBoost, XGBoost);
 4. Сравните производительность моделей с помощью метрики `recall` для положительного класса (наличие диабета).
 5. Обоснуйте, какой из алгоритмов предпочтительнее для данной медицинской задачи.

Результат программы:

Размер обучающей выборки: (537, 8)

Размер тестовой выборки: (231, 8)

```
Decision Tree Recall: 0.4444
Random Forest Recall: 0.5432
AdaBoost Recall:      0.4938
CatBoost Recall:       0.5309
XGBoost Recall:        0.5926
```

Итоговая таблица (сортировка по Recall):

	Model	Recall
4	XGBoost	0.592593
1	Random Forest	0.543210
3	CatBoost	0.530864
2	AdaBoost	0.493827
0	Decision Tree	0.444444

Лучшая модель по метрике Recall: XGBoost (0.5926)

Вывод: на практике сравнил работу нескольких алгоритмов одиночного дерева решений, случайного леса и бустинга для деревьев.