

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ

«БРЕСТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» ФАКУЛЬТЕТ

ЭЛЕКТРОННО-ИНФОРМАЦИОННЫХ СИСТЕМ

Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №5

Специальность ИИ-23

Выполнила
Палто Е. С.
студентка группы ИИ-
23

Проверила
Андренко К.В.,
ст. преп.
кафедры ИИТ,
«—» ————— 2025 г.

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

Задачи:

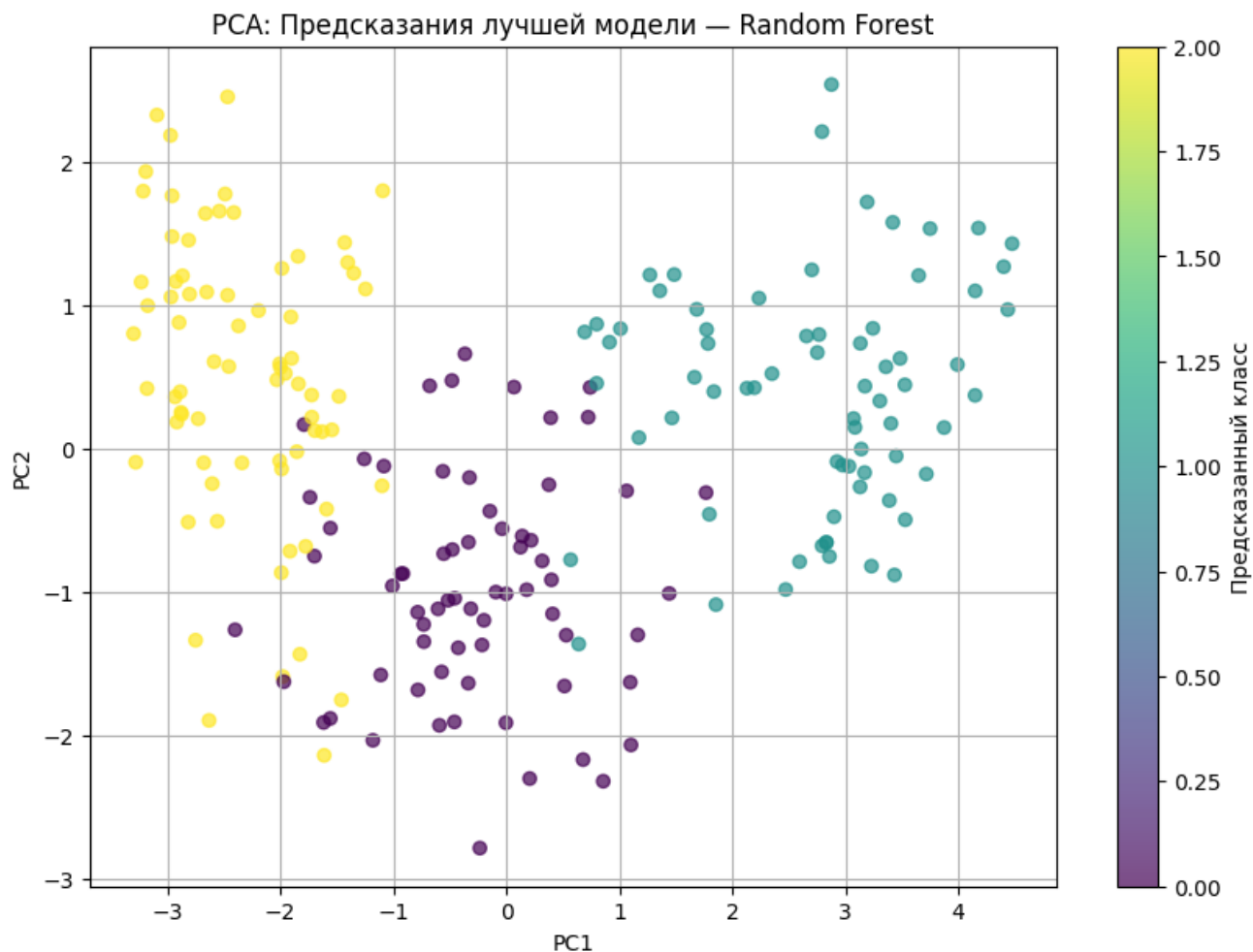
1. Загрузить датасет по варианту;
2. Разделить данные на обучающую и тестовую выборки;
3. Обучить на обучающей выборке одиночное дерево, случайный лес и реализовать бустинг для решающих деревьев (AdaBoost, CatBoost, XGBoost);
4. Оценить точность каждой модели на тестовой выборке;
5. Сравнить результаты,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

Вариант 8

- Seeds
- Классифицировать семена на три сорта пшеницы (Kama, Rosa, Canadian) на основе их геометрических параметров
- **Задания:**
 1. Загрузите и стандартизируйте данные;
 2. Разделите выборку на обучающую и тестовую;
 3. Обучить на обучающей выборке одиночное дерево, случайный лес и реализовать бустинг для решающих деревьев (AdaBoost, CatBoost, XGBoost);
 4. Сравните общую точность (ассураку) всех трех моделей;
 5. Визуализируйте данные в 2D (например, с помощью PCA), раскрасив точки в соответствии с предсказаниями лучшей модели.

```
=== Точность (Accuracy) моделей на тестовой выборке ===  
Decision Tree: 0.8889  
Random Forest: 0.9206  
AdaBoost: 0.8413  
XGBoost: 0.8730  
CatBoost: 0.9048
```

Визуализация данных в 2D с помощью PCA, раскрасив точки в соответствии с предсказаниями лучшей модели



Вывод: на практике сравнила работу нескольких алгоритмов одиночного дерева решений, случайного леса и бустинга для деревьев решений. Среди рассмотренных моделей наилучшую точность на тестовой выборке показал случайный лес (Random Forest) — 92.06%, что свидетельствует о его высокой обобщающей способности и устойчивости к переобучению на данном датасете.

Модель CatBoost также продемонстрировала хороший результат — 90.48%, уступив Random Forest менее чем на 2%.

Одиночное дерево решений и XGBoost показали сопоставимые результаты (~87–89%), в то время как AdaBoost оказался наименее точным — 84.13%, вероятно, из-за чувствительности к шуму в данных или недостаточной глубине базовых классификаторов.

Таким образом, для задачи классификации сортов пшеницы по геометрическим признакам наиболее эффективным оказался ансамблевый метод Random Forest.