

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

Выполнил
Лапин В.А.,
студент группы ИИ-23

Проверила
Андренко К.В.,
преподаватель стажёр
кафедры ИИТ,
« __ » _____ 2025 г.

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

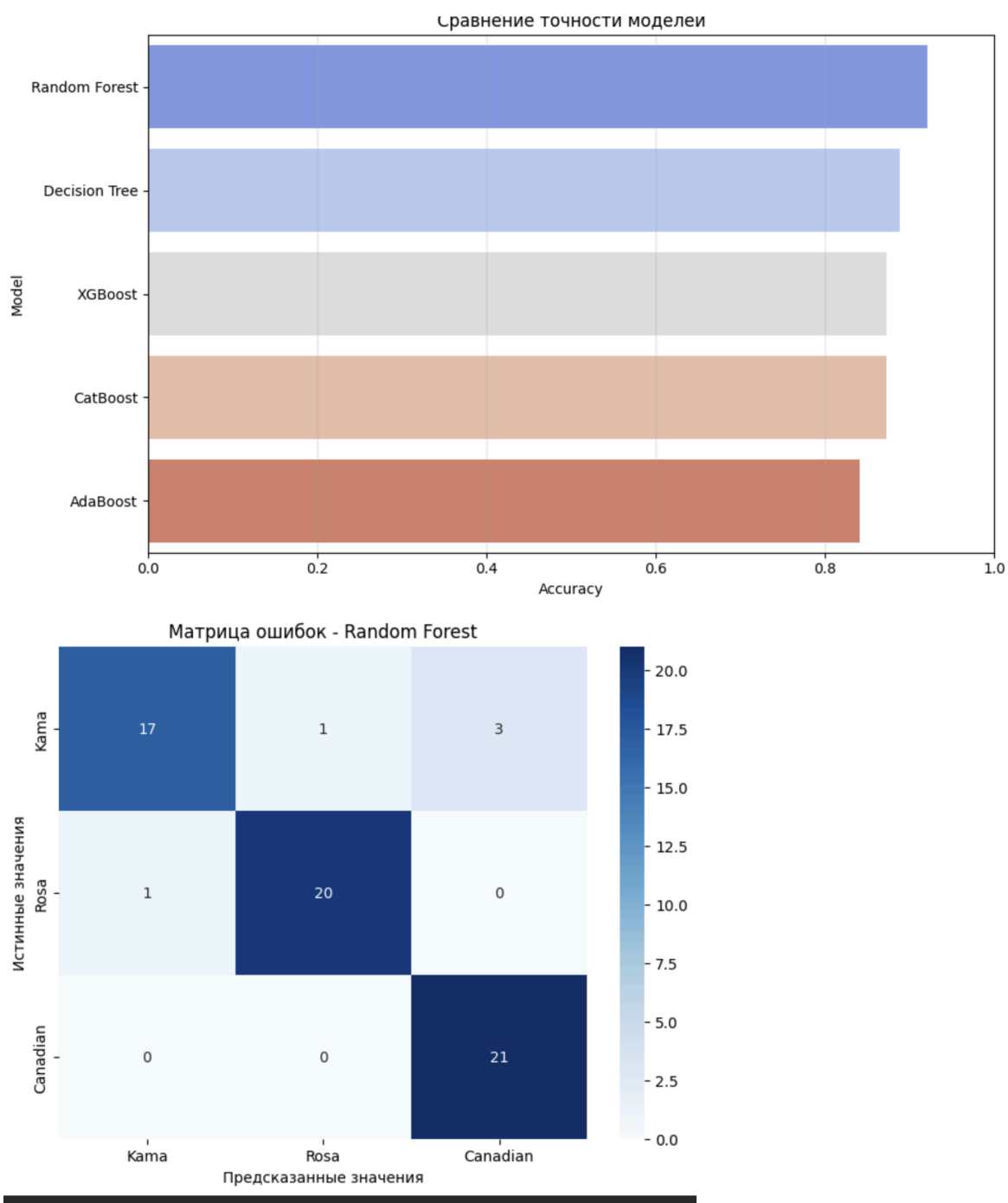
Задачи:

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 (accuracy) всех трех моделей;
 5. Визуализируйте данные в 2D (например, с помощью PCA), раскрасив точки в соответствии с предсказаниями лучшей модели.

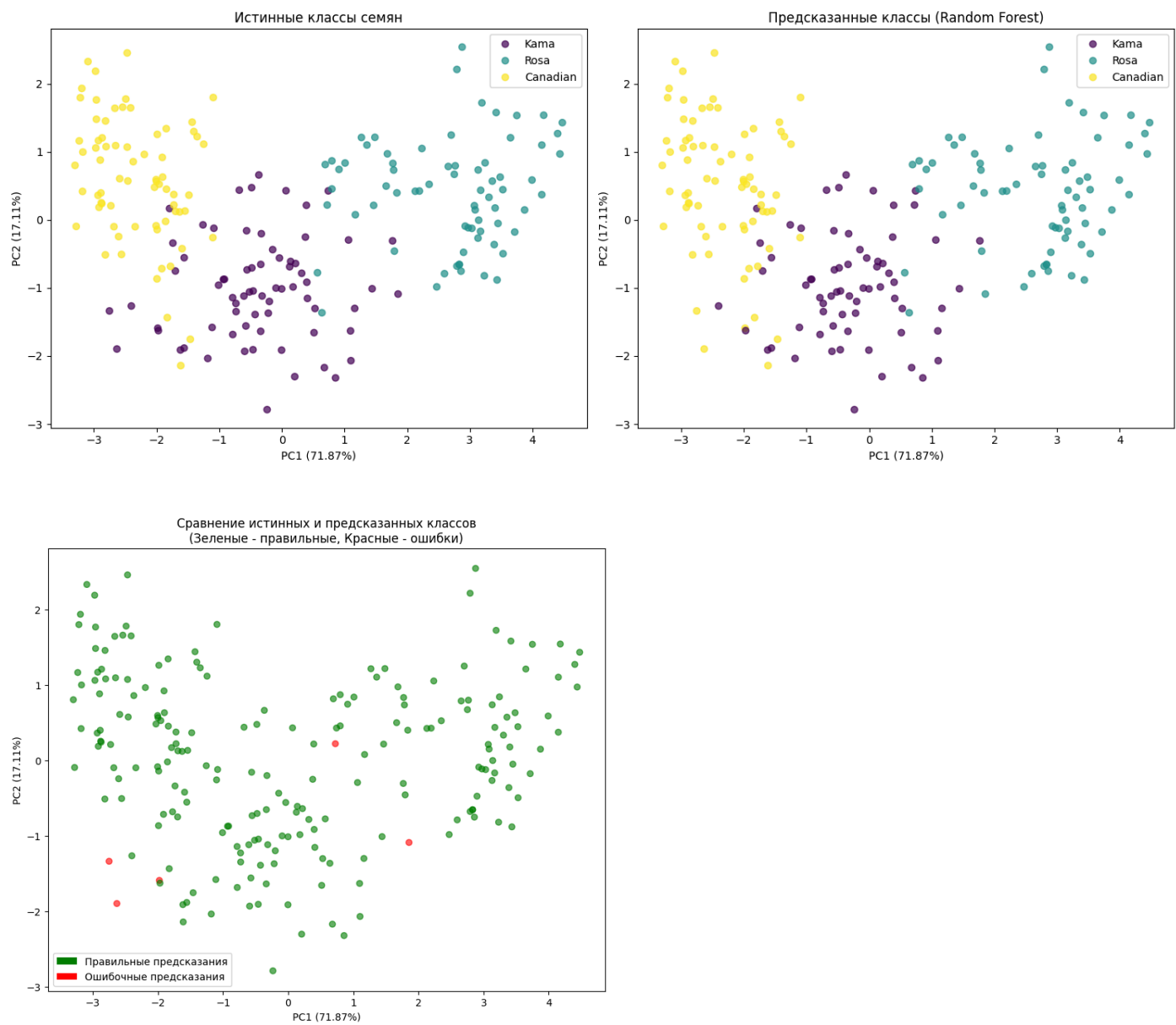
Результат:



Визуализация PCA:

Объясненная дисперсия компонент: [0.71874303 0.17108184]

Суммарная объясненная дисперсия: 0.8898



Вывод:

Лучшая модель: Random Forest с точностью 0.9206

Детальный отчет для Random Forest:

Kama	0.94	0.81	0.87	21
Rosa	0.95	0.95	0.95	21
Canadian	0.88	1.00	0.93	21
accuracy			0.92	63
macro avg	0.92	0.92	0.92	63
weighted avg	0.92	0.92	0.92	63