



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего
образования

**«Дальневосточный федеральный университет»
(ДВФУ)**

**ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
(ШКОЛА)**

Департамент информационных и компьютерных систем

ОТЧЕТ

по дисциплине «системы искусственного интеллекта»

Выполнил студенты группы Б9122-
09.03.03пикд

Зверев Р. И.

Проверил преподаватель

Бочарова В. В.

зачтено/не зачтено

г. Владивосток

2025 г

Оглавление

Цель работы.....	2
Введение.....	3
Описание данных.....	4
Сравнение бустингов в зависимости от количества итераций.....	5
Сравнение BrownBoost.....	6
Сравнение классификаторов на предобработанных данных.....	7
Заключение.....	8
Список литературы.....	9

Цель работы

Целью работы является сравнение разных моделей бустинга на датасете Bioresponse.

Постановка задачи

В данной работе рассматривается задача классификации с помощью бустингов.

Необходимо реализовать следующие этапы и функции:

- Сравнить время работы и точность Ada Boost, Logit Boost и Brown Boost от количества итераций;
- Сравнить как параметры Brown Boost влияют на точность и время работы;
- Оценить точность классификаторов после предобработки данных.

Введение

В этой лабораторной работе исследуется влияние различных алгоритмов бустинга на задачу классификации на примере набора данных Bioresponse. Бустинг — мощный подход ансамблевого обучения, объединяющий простые модели в сильный классификатор; разные его реализации по-разному балансируют точность и вычислительную сложность, особенно при изменении числа итераций и внутренних параметров.

Задача работы — сравнить поведение трёх бустинговых методов (AdaBoost, LogitBoost и BrownBoost) по двум ключевым измерителям: точности классификации и времени обучения, в зависимости от числа итераций и настроек алгоритмов. Также требуется исследовать чувствительность BrownBoost к его параметрам и оценить, как предобработка данных влияет на качество моделей.

Описание данных

Датасет Bioresponse предоставляет 3751 объект (записей) и 1777 признаков (столбцов). Каждый объект — это конфигурация некоторой молекулы. Класс объекта — вызвала ли молекула реакцию или нет.

Сравнение бустингов в зависимости от количества итераций

Все модели бустинга обучались 15 раз с разным количеством итераций (1, 11, 21, 31, 41, ..., 141).

Зависимость точности от количества итераций представлена графически:

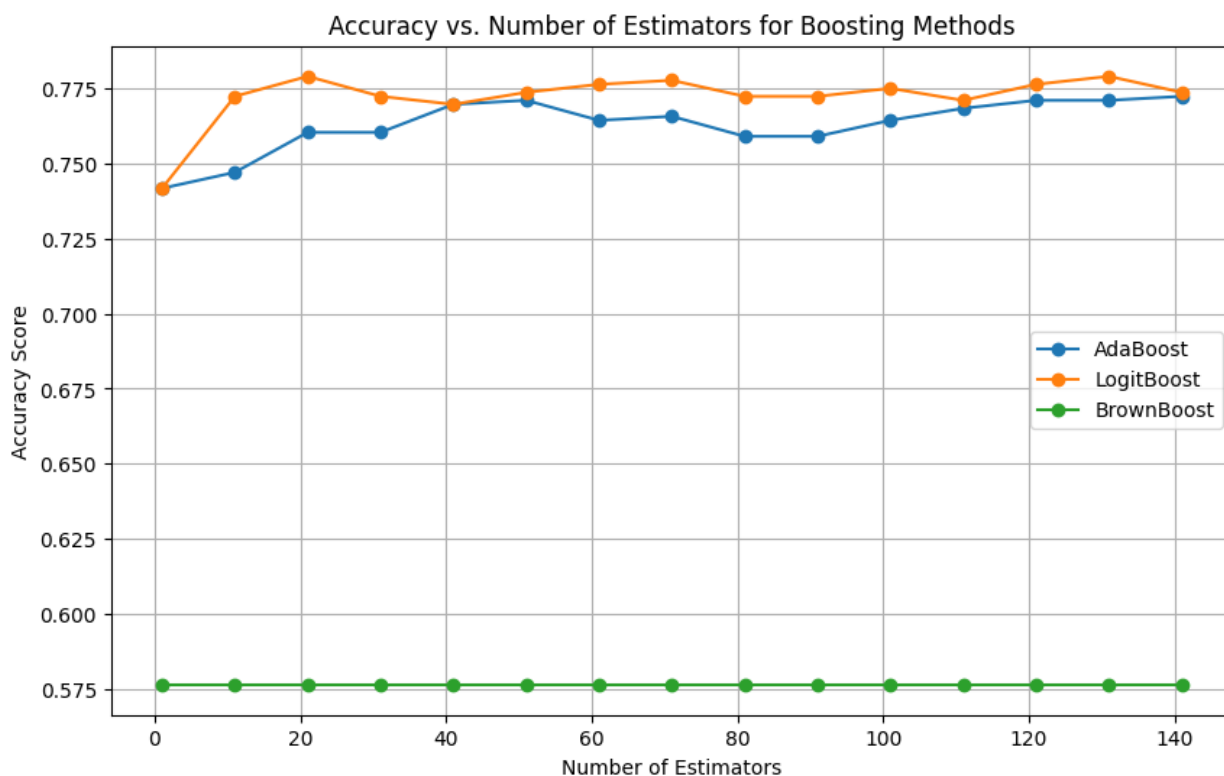


Рисунок 1. Зависимость точности от кол-ва итераций

LogitBoost показывает себя лучше в точности, чем остальные модели. BrownBoost будто просто угадывает, а не обучается.

Зависимость времени работы от количества итераций:

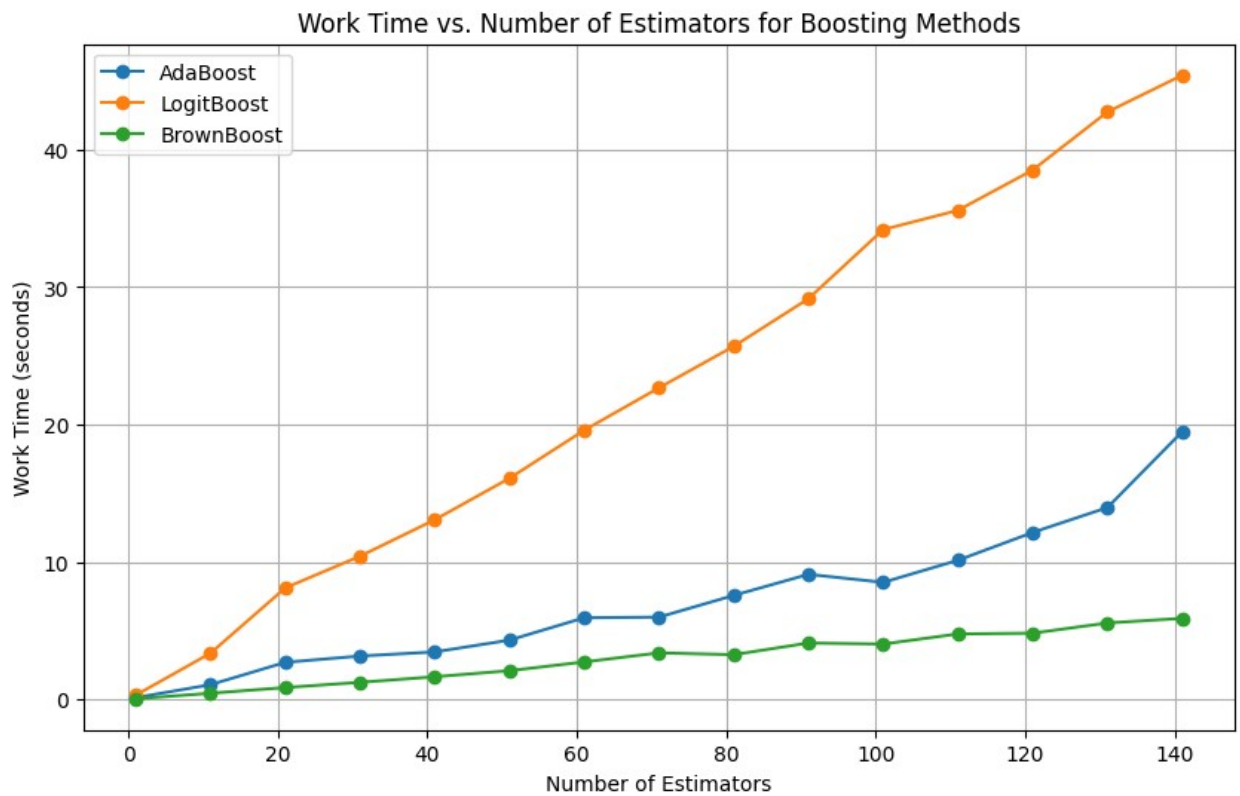


Рисунок 2. Сравнение скорости работы от кол-ва итераций

LogitBoost хоть и выигрывает в точности, но значительно проигрывает в скорости.

Сравнение BrownBoost

Построим графики зависимости точности и скорости работы BrownBoost от параметров эпсилон и ню (ϵ и ν):

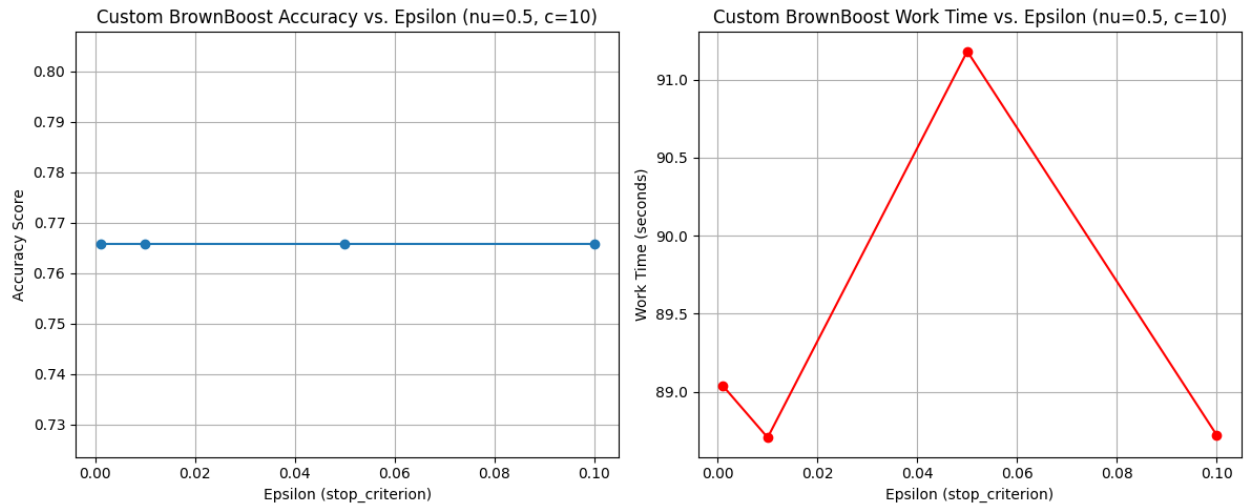


Рисунок 3. Сравнение качества и времени работы от эпсилон

Как видно, точность не изменяется абсолютно для заданных эпсилон, но время работы заметно меняется.

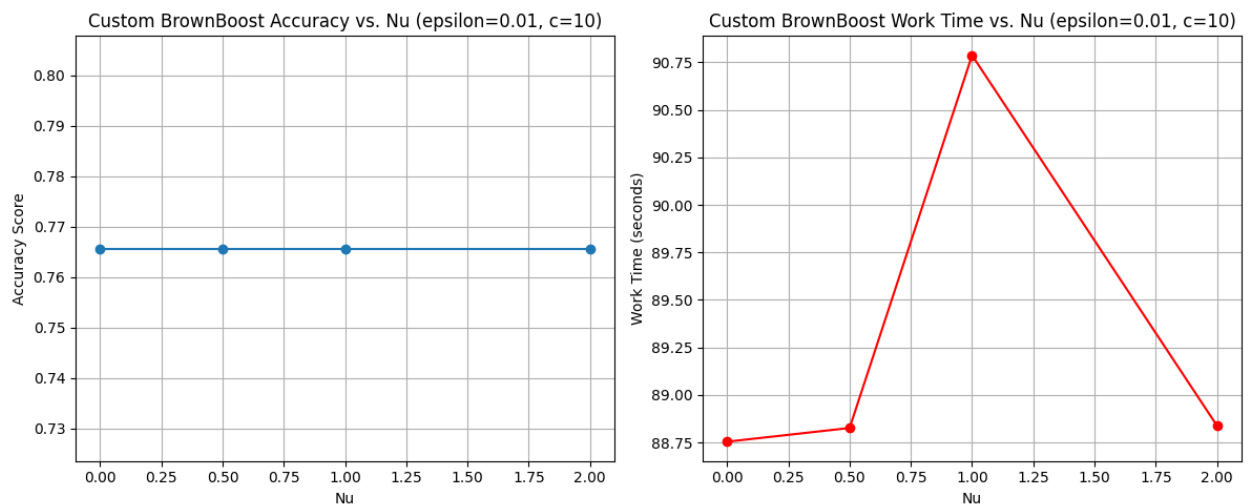


Рисунок 4. Сравнение качества и скорости работы от ню

Как видно из графиков, точность не изменяется от данных параметров, а время работы еще как изменяется. Графики очень схожи между собой.

Сравнение классификаторов на предобработанных данных

С помощью алгоритма Local Outlier Factor находим 159 аномальных записей и удаляем их, получив 3349 записей.

С помощью PCA снизим размерность данных, а то 1776 признаков много. В самом деле, 445 признаков объясняют 95% дисперсии, это почти в 4 раза меньше признаков.

В итоге тренировочная выборка имеет 2679 записей и 455 признаков, а тестовая 670 записей и 455 признаков.

Каждый бустинг обучался с ограниченным количеством итераций — 500.

В итоге сравнительный график выглядит так:

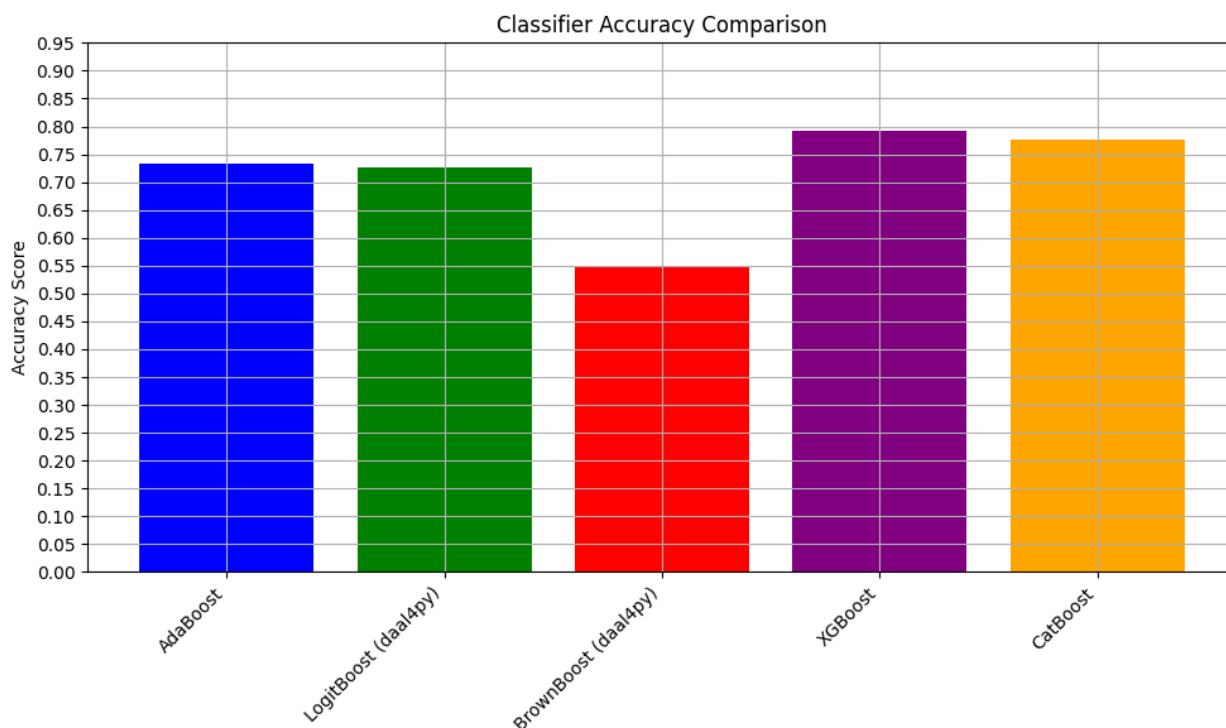


Рисунок 5. Сравнение бустингов

Как видно AdaBoost выигрывает LogitBoost в точности и, как было показано на графиках выше, в скорости работы.

Заключение

В наших экспериментах самым слабым методом оказался BrownBoost: он показал низкую и нестабильную точность (~55%), как будто «наугад» подбирает решения. Вероятные причины такой работы — чувствительность алгоритма к настройкам и шуму в данных, чувствительность к размеченности классов, особенности реализации (параметры потерь/learning rate, базовый классификатор) или нерешённые проблемы с сходимостью. Чтобы не оставлять это как загадку, полезно было бы проверить гипотезы: подобрать параметры BrownBoost (learning rate, число итераций), сравнить реализацию/параметры базового слабого классификатора, и посмотреть матрицу ошибок — возможно, алгоритм стабильно ошибается на редком классе.

Лучшей по совокупности точности и скорости в работе оказалась AdaBoost. Она показала высокую точность при относительно быстрой обучаемости — что объясняется простотой механизма пересвешивания ошибок и тем, что AdaBoost эффективно использует лёгкие базовые классификаторы (например, неглубокие деревья), быстро наращивая качество без тяжёлых вычислений. В рамках этой лабораторной задачи AdaBoost демонстрирует наилучший баланс «точность/время», поэтому его можно считать предпочтительным выбором при аналогичных условиях.

Список литературы

1. GitHub: исходный код лабораторной работы. – URL: [Лабораторная работа №3.7](#) (дата обращения: [09.10.2025]). – Текст: электронный.