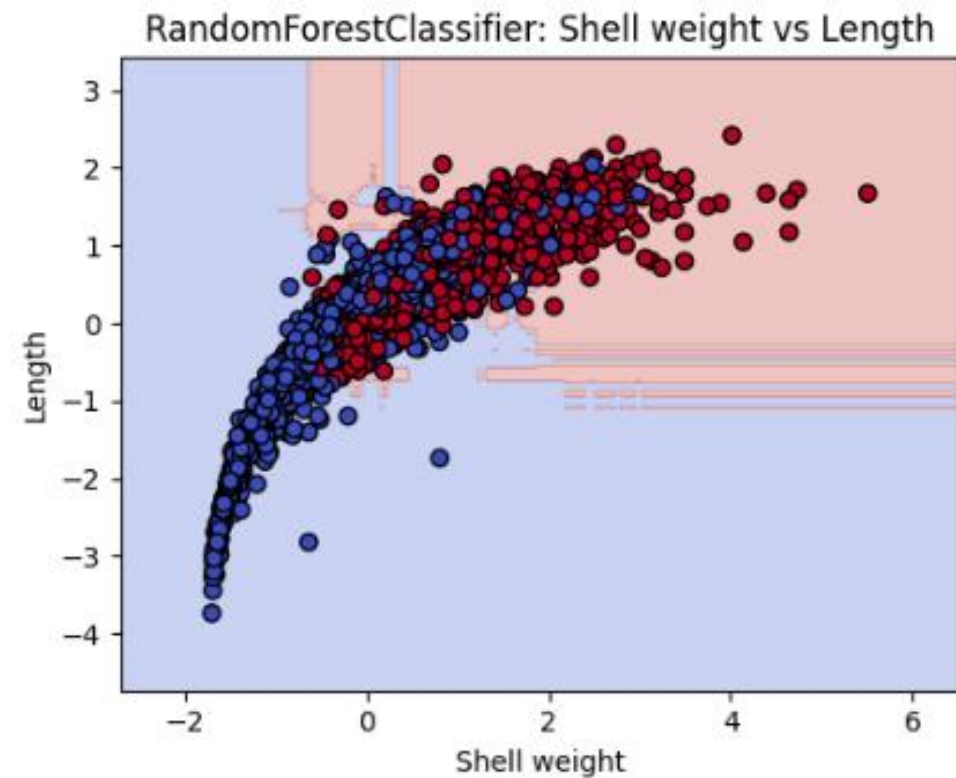
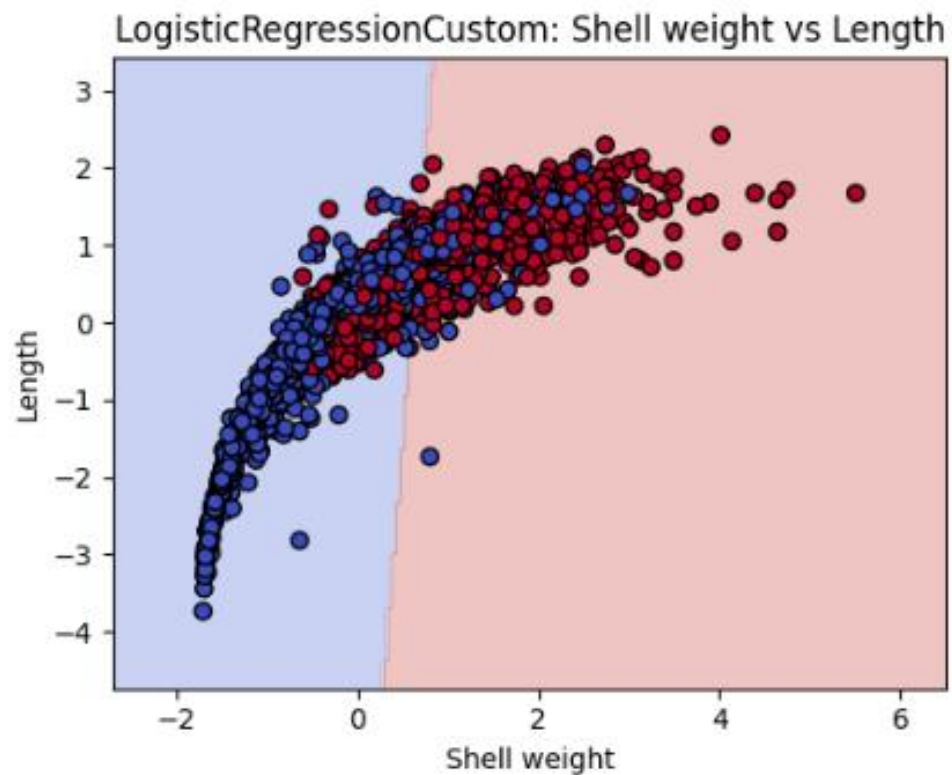


Сдача Лр. 2

Зиновьев Максим Иванович

М80-307Б-23



Random Forest vs Logistic Regression

Random Forest:

Сложные нелинейные границы

Учитывает взаимодействия признаков

Точность: 96.4%

Logistic Regression:

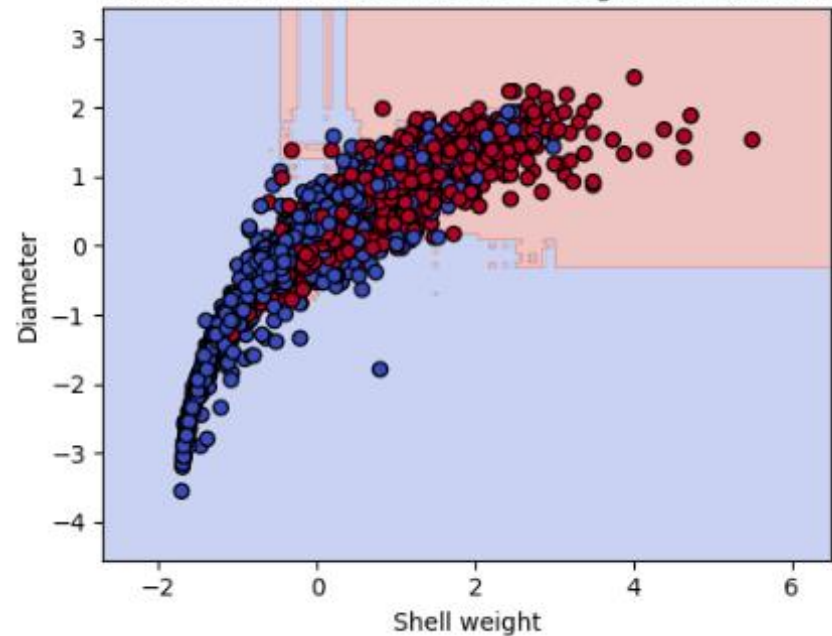
Простая прямая граница

Линейное разделение

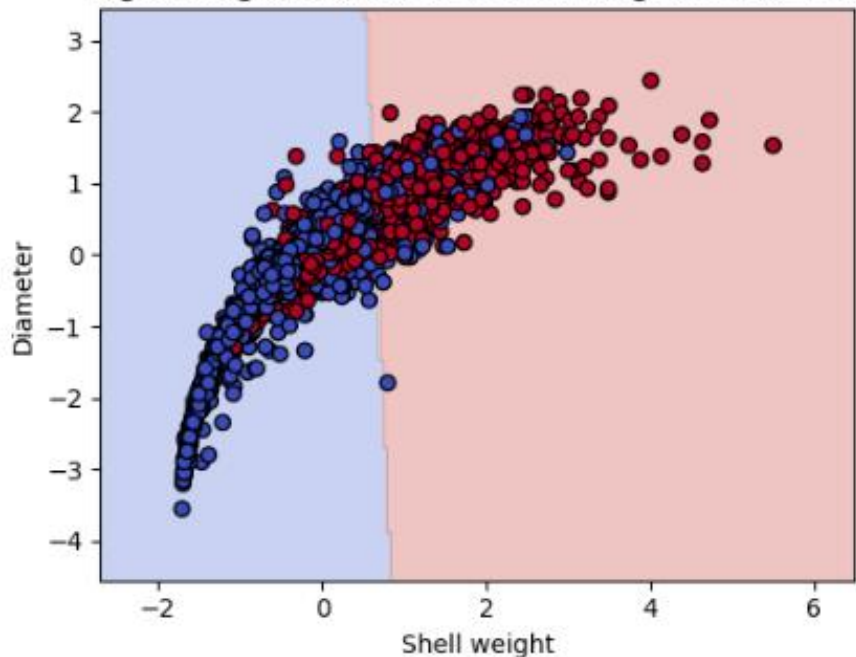
Точность: 95.4%

Итог: Random Forest на 1% точнее за счет сложных нелинейных зависимостей

RandomForestClassifier: Shell weight vs Diameter



LogisticRegressionCustom: Shell weight vs Diameter



Random Forest vs Logistic Regression

Random Forest:

Сложные нелинейные границы решений

Улавливает взаимодействия между весом раковины и диаметром

Более точное разделение классов

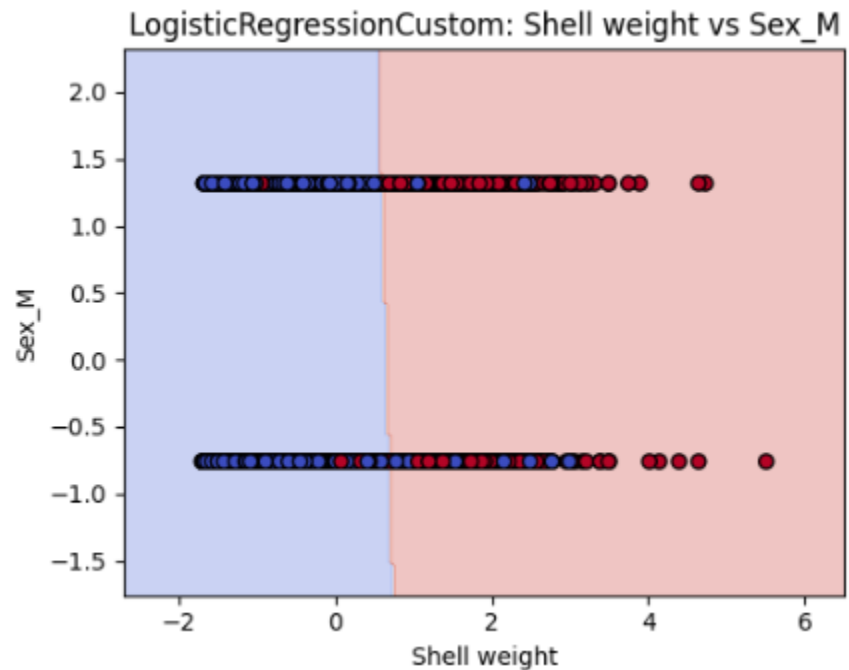
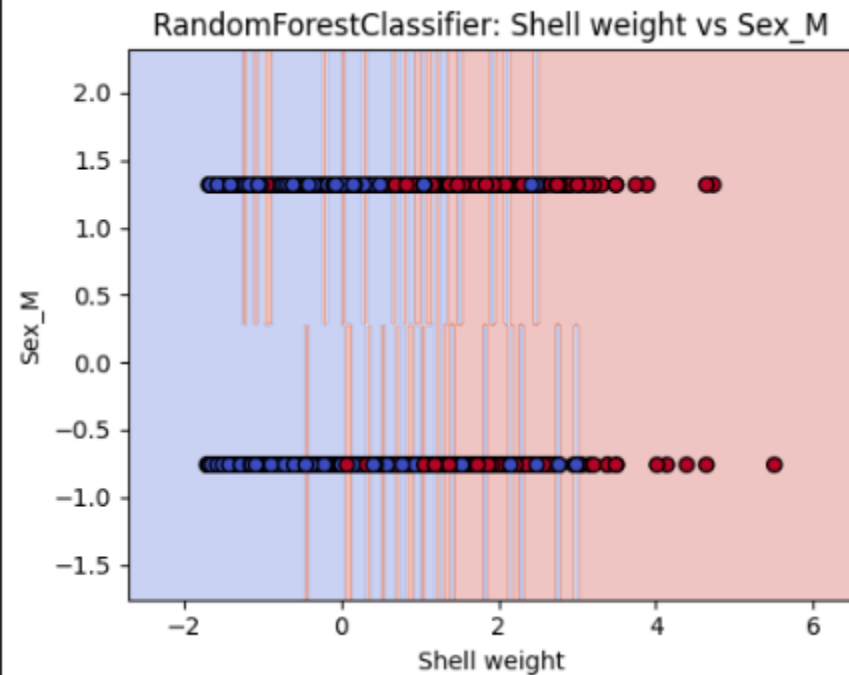
Logistic Regression:

Простая линейная граница

Прямолинейное разделение по диагонали

Менее точное в сложных случаях

Вывод: Random Forest точнее за счет способности моделировать сложные нелинейные зависимости между физическими параметрами моллюсков.



Random Forest vs Logistic Regression для категориальных признаков

Random Forest:

Эффективно работает с бинарными признаками (Sex_M)

Создает адаптивные границы для разных категорий

Учитывает взаимодействие пола с весом раковины

Logistic Regression:

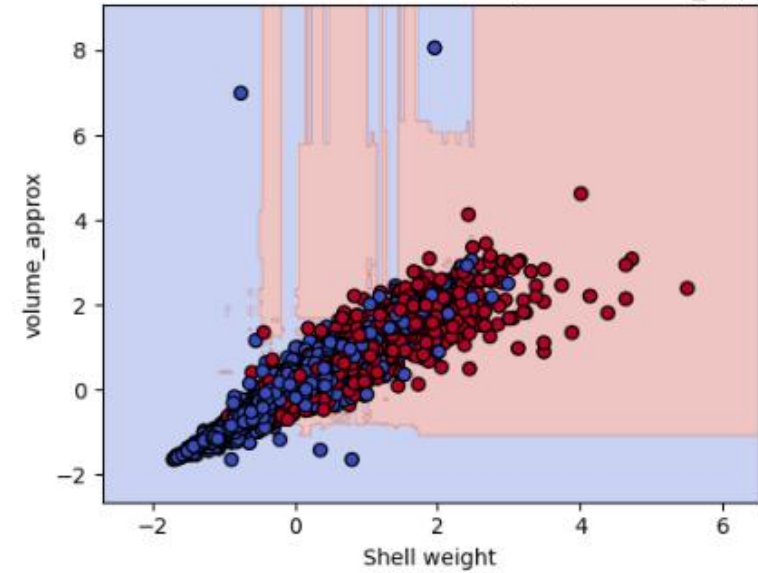
Линейная модель плохо обрабатывает one-hot кодирование

Искусственные линейные границы для категориальных данных

Менее эффективна для дискретных признаков

Вывод: Random Forest превосходит в работе с категориальными признаками, создавая более осмысленные границы решений.

RandomForestClassifier: Shell weight vs volume_approx



Random Forest vs Logistic Regression для производных признаков

Random Forest:

Сложные нелинейные границы для volume_approx

Учитывает взаимосвязь объема с весом раковины

Более точное разделение в многомерном пространстве

Logistic Regression:

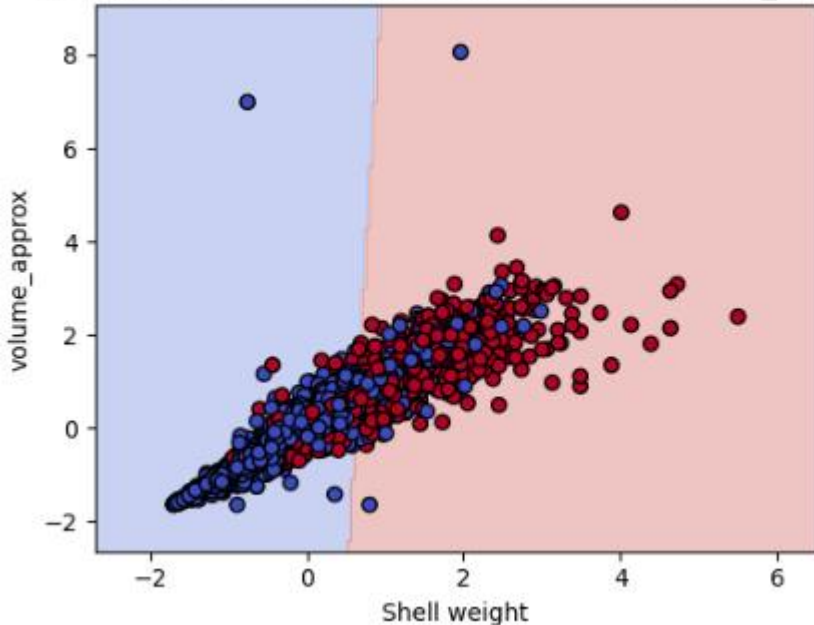
Упрощенное линейное разделение

Прямолинейная граница не отражает сложных зависимостей

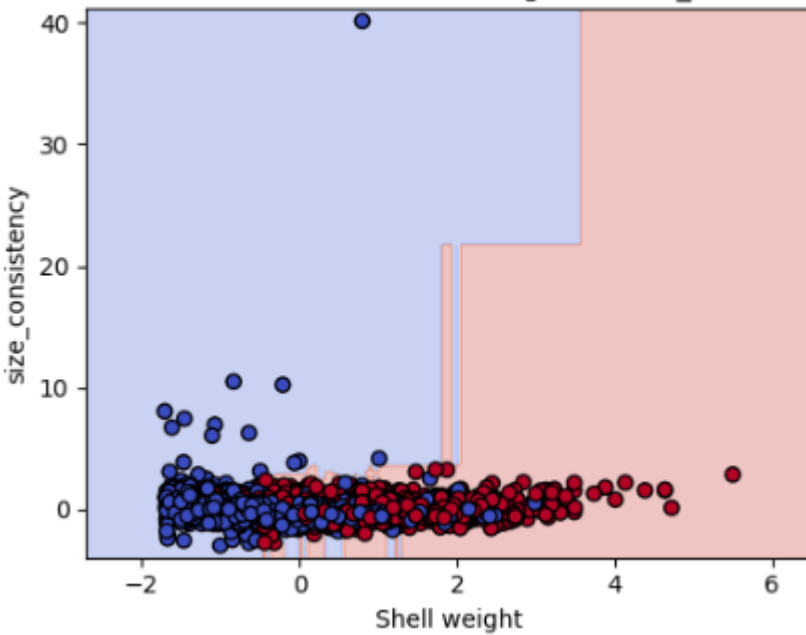
Ограничена линейными отношениями

Вывод: Random Forest лучше обрабатывает производные признаки, выявляя сложные нелинейные взаимосвязи между физическими параметрами.

LogisticRegressionCustom: Shell weight vs volume_approx



RandomForestClassifier: Shell weight vs size_consistency



Random Forest vs Logistic Regression для слабых признаков

Обе модели:

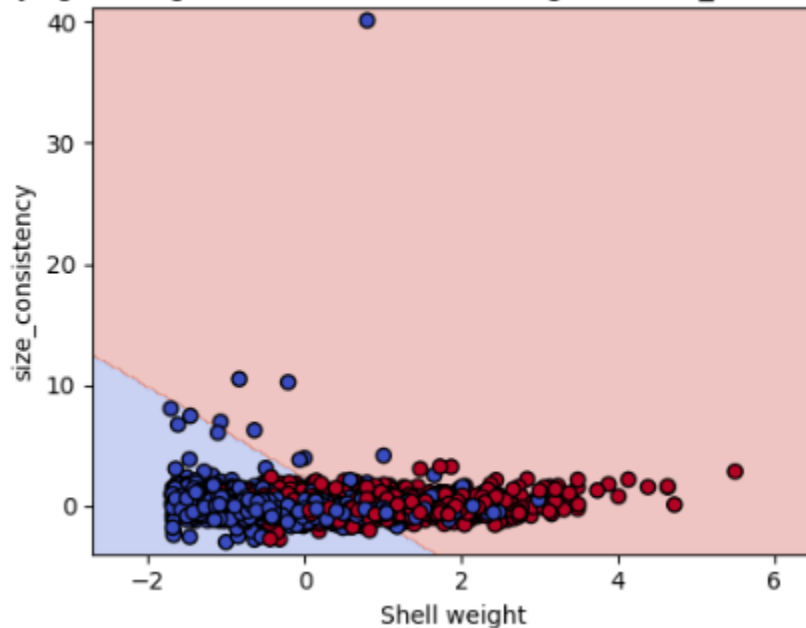
Показывают плохое разделение для size_consistency

Классы сильно перекрываются

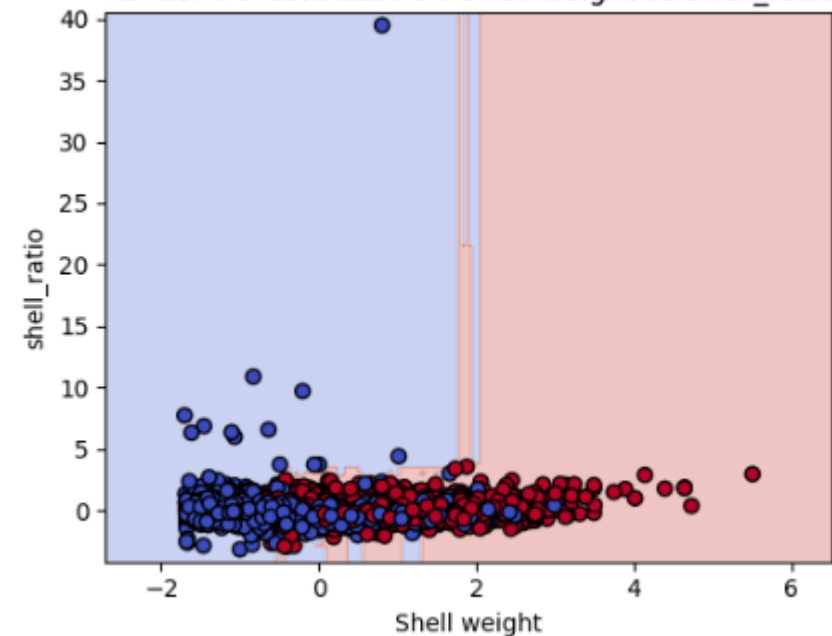
Признак имеет низкую предсказательную силу

Вывод: Обе модели подтверждают, что size_consistency - слабый признак для классификации возраста моллюсков. Это демонстрирует важность отбора признаков для качества модели.

LogisticRegressionCustom: Shell weight vs size_consistency



RandomForestClassifier: Shell weight vs shell_ratio



Random Forest vs Logistic Regression для пропорциональных признаков

Random Forest:

Создает сложные адаптивные границы для shell_ratio

Учитывает нелинейные зависимости доли раковины

Более точное разделение классов

Logistic Regression:

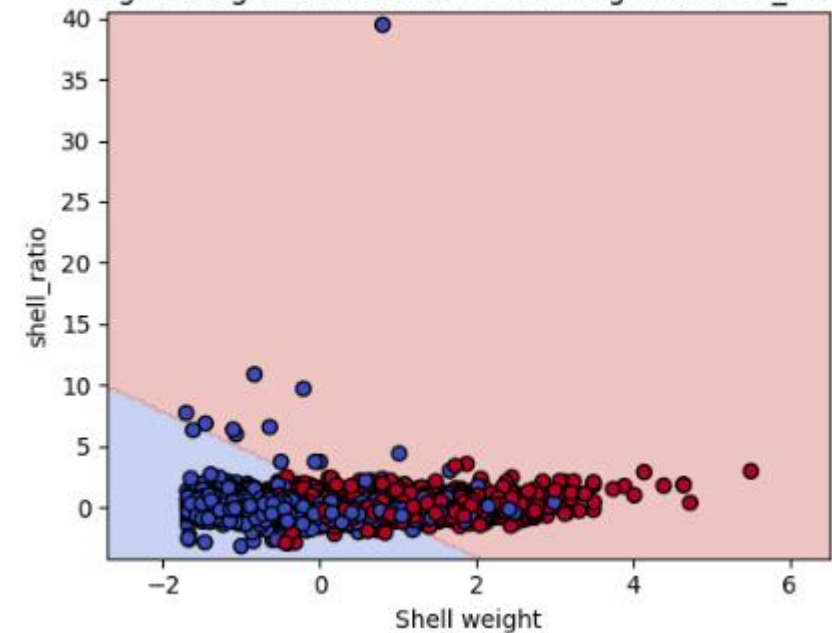
Вертикальная граница - использует в основном Shell weight

Не эффективно использует информацию о пропорциях

Упрощенное линейное решение

Вывод: Random Forest лучше использует информацию о пропорциональных признаках, создавая более точные границы решений.

LogisticRegressionCustom: Shell weight vs shell_ratio



Рейтинг моделей по точности (Accuracy):

XGBoost: 0.968 ± 0.006

Gradient Boosting: 0.967 ± 0.001

Random Forest: 0.964 ± 0.003

SVM: 0.959 ± 0.006

Logistic Regression: 0.954 ± 0.006

Decision Tree: 0.949 ± 0.005

KNN: 0.872 ± 0.004

Сравнение по всем метрикам:

XGBoost - абсолютный лидер:

Accuracy: 0.968 (лучший)

F1-score: 0.952 (лучший)

ROC-AUC: 0.994 (один из лучших)

Стабильность: ± 0.006 (хорошая)

XGBoost показал наилучшие результаты по всем ключевым метрикам